

SAS ユーザー総会 2026 論文集

会場：東京大学・弥生講堂

会期：2026 年 10 月 14 日(水) ～5 日(木)

Day 0:10月13日(火)　会場:SAS Institute Japan

主催：SAS ユーザー会 世話人会

目次

論文

兵庫県立大学におけるJMPの利用　～高砂市商業調査を事例として ---13

川向肇，神棒美波，中根滉大(兵庫県立大学)，大毛佑太(小野市役所)

JMPを利用した高砂市商業調査の分析における多重対応分析 ---23

川向肇(兵庫県立大学)

生存時間解析の多重比較法の性能評価（小標本における近似精度と検出力） ---32

橋本敏夫(株式会社　タクミインフォメーションテクノロジー)，山田雅之(キッセイ薬品工業株式会社)，本田主税(小野薬品工業株式会社)，福島慎二(株式会社　タクミインフォメーションテクノロジー)

SASによるイベント駆動型臨床試験における中間解析時点予測 ---43

堀本隆一(株式会社エス・シー・エー)

AIによって再構成される解析プログラミングとその役割変化の考察 ---62

瓜生卓，小山卓己，野間基希(ファイザーR&D合同会社)

単群試験におけるノンパラメトリックベイズ流モデルに基づくマイルストン生存時間解析

吉本拓矢，澤本涼(中外製薬株式会社)，田畑耕治(東京理科大学) ---72

多次元尺度構成法による臨床データのマッピングと適合度評価 ---94

高田浩成(イーピーエス株式会社)

目次

量子計算機のいらない世界初の組み合わせLDFのRevised IP Optimal-LDF ---112

新村秀一(成蹊大学)，新村秀樹(イノベストメント(株))

Base/SASによるWOE変換を用いたスコアリングモデル構築法 ---143

坂巻英一(宇都宮大学)

SASによる傾向スコアに基づくPower 事前分布アプローチの実装 ---150

横田大輔，赤石理香子，浅野豊，伊庭克拓，黒見真央，毛利誠(大塚製薬株式会社)

In vivo QT試験の検出感度向上に寄与する平滑スプライン曲線 －ICH-E14/S7B Q&A（2022）への対応－ ---172

高橋行雄(BioStat研究所株式会社)

SASプログラムの類似度評価 ～Token化と5-gramによる構造比較～ ---182

井上貴博(協和キリン株式会社)

JADERを用いた有害事象の発現時期パターンの評価手法 ---192

玉木志織(日本メディア株式会社/久留米大学)，室谷健太，五百路徹也(久留米大学)

PROC PYTHONで広がるSASの可能性 — Excel連携からデータ活用の新しい世界へ — ---202

小澤義人，髙橋康平，高橋杏奈，稲葉洋介(千葉大学)

見た目は『.』、中身は何か — BASE SASにおける欠測表現の入力挙動と情報損失の検証 —

---207

小林邦世(イーピーエス株式会社)

目次

等張回帰入門 — 理論、実装、物理アナロジー ---231

田中真史(イーピーエス株式会社)

MIMIC-IVを用いた教育機関におけるRWD解析教育の試み ---240

小林美保，小野洋平(聖路加国際大学)

SASハッシュオブジェクトとそれを用いたデータ加工テクニック ---250

兵頭照将(イーピーエス株式会社)

データステップとPDV入門 — SASコードはどのような仕組みで動いているのか ---260

角春佳(イーピーエス株式会社)

生成AIエンジニアリングとSAS — Prompt・Context・Harness・Loop・Graph Engineeringの体系化 — ---275

小野潔

NDBオープンデータを用いた高血圧患者におけるアンメットメディカルニーズの探索 ---287

芦沢宏治，渡邊智昭，齊藤正樹(株式会社タクミインフォメーションテクノロジー)

逆ポーランド記法の実装 ---296

田中真史(イーピーエス株式会社)

抗がん剤第I相試験における3+3，BOIN及びBOIN12の用量選択特性：SASを用いたシミュレーションによる評価 ---306

牛尾英孝(株式会社ピープロ・ジャパン)

目次

報告するまでがEstimand ---319

森田祐介，太田裕二，内藤寿輝(ノーベルファーマ株式会社)

MMRMにおける分散共分散構造及びロバスト分散推定が推定結果に与える影響の検討 ---333

牛尾英孝(株式会社ピープロ・ジャパン)

『What if』の視点で読み解くPROC CAUSALTRT：IPWと標準化の理解と注意点 ---342

峯脇颯汰，浦狩保則(シミック株式会社)

ベイズ統計の漸近理論とSASによる実装 ---354

玉乃井崚太(シミック株式会社)

PROC IMLなしで実装するPattern-Mixture Multiple Imputation ---369

ビービ彩加，室永遼太郎(シミック株式会社)

置換ブロック法による被験者割付におけるブロックサイズ混合手法と固定手法の性能比較

---381

上原歩夢，室永遼太郎，菊池宏和(シミック株式会社)

MMRMにおける重みの取り扱いの検討 —SAS PROC MIXED / GLIMMIX の重み指定の違いと実務上の注意点 — ---394

江口凌平，中西豊支(シミック株式会社)

McNemar検定のサンプルサイズ設計に関する性能比較 ---405

尾崎龍成，団野友也，魚住龍史(東京科学大学)

目次

傾向スコアマッチングによる群間バランスの視覚的評価 ---415

魚住龍史(東京科学大学)

ベイズ最適区間デザインによるがん第I相試験 ---428

栗田凪，魚住龍史(東京科学大学)

Rich Kaplan-Meier Plotの提案 ---446

森岡裕(イーピーエス株式会社)

SAS/IMLで実装する推定方程式： 修正ポアソン回帰, IPW, G-computationへの応用 ---465

西尾和恭(イーピーエス株式会社)

プレゼンテーション ---481～580

SASグラフ100

森岡裕(イーピーエス株式会社)

多時点の相関を解析する - TQT試験での検討事例 -

仲家諒(武田薬品工業)

非臨床研究データの統計解析の注意点（British Journal of Pharmacology の実験計画と解析のガイドラインを中心に）

原田淳(第一三共株式会社)

RでPROC COMPAREっぽいことをやってみた

関根暁史(藤本製薬株式会社)

目次

SASは次のステージへ ～ AI・オープンソース・高速分析を実現する最新Viyaアップデート～

土生敏明(SAS Institute Japan)

非臨床統計の現在地 — 見えている課題から考える関わり方とつながり

本田主税(小野薬品工業株式会社)

OSSコミュニティに参加しよう！

仲家諒(武田薬品工業)

Robustな解析手法のSAS実装　-KL-divergenceやβ-divergence-

笹原大敬(イーピーエス株式会社)

SASによる症例数再推定を伴うアダプティブデザインの実践 ～CPZデザインの計算・評価・可視化を支えるSAS実装～

内藤寿輝，森田祐介，太田裕二(ノーベルファーマ株式会社)

小さな製薬企業における統計解析プロセスと効率化の紹介 ～試行錯誤の段階にある、一担当者によるリアルな事例共有～

太田裕二，森田祐介，内藤寿輝(ノーベルファーマ株式会社)

数式なしのPROC MCMC入門

大山暁史(イーピーエス株式会社)

製薬協 非臨床 一般毒性・DX活動紹介1　-バーチャル対照群を用いた非臨床毒性試験法の概要-

黒岡貴生(EAファーマ株式会社)

目次

製薬協 非臨床 一般毒性・DX活動紹介2　-バーチャル対照群導入に向けたSENDの活用支援-

吉池通晴(第一三共株式会社)

SASとRにおけるFirth法の実装差の整理と数値的評価

西川寛来(協和キリン株式会社)，後藤駿輔(ノバルティス ファーマ株式会社)

USDMからTLFsまで ― メタデータ駆動によるEnd-to-End SASプログラム自動生成への挑戦

魚田孝志，森岡裕(イーピーエス株式会社)

レポーティングにおけるRとSASの共存～単一ソースによる実行と文書化～

後藤駿輔(ノバルティス ファーマ株式会社)，西川寛来(協和キリン株式会社)

PROC SGPLOTのPOLYGONステートメントによるRWD可視化の実践

塩田早恵，三浦伸也(シミック株式会社)

AIエージェントとOSSから考える今後のコーディングとSAS

中松建(個人)

SASによるCDISC ARD実装とTable化の試み ～SASパッケージ化に向けて～

山野辺浩己(第一三共株式会社)

PharmaSUG 2026 U.S.レポート

堺映奈(協和キリン株式会社)

目次

正しく推定し、正しく理解するために

中川雄貴(イーピーエス株式会社)

SASによるベイズ流単群Poissonイベント発現率における症例数設計

張方紅(バイオジェン・ジャパン株式会社)

千葉大学医学部附属病院臨床試験部における統計解析業務の現状と課題　―少人数体制での効率化・標準化に向けた実践―

髙橋康平，高橋杏奈，小澤義人，稲葉洋介(千葉大学)

臨床試験のデータ解析とOSSの最近の動向

堀田真一(協和キリン株式会社)

Logistic一般化線形混合モデルにおける周辺平均の推定

王春弋(エイツーヘルスケア株式会社)

SASPy応用: 新機能の紹介から実践例まで

中松建

3群以上で使える共変量バランス評価方法

中倉章祥(久留米大学)

日本発、眼科領域におけるTAUG開発

山崎文寛(武田薬品工業株式会社)

目次

再生医療等製品の開発におけるベイズ流拡大試験の提案と適用事例

上村鋼平(東京大学)

Integrated Analysis Data Reviewer's Guide Completion Guidelinesの解説と実務におけるポイント

長又理緒(アストラゼネカ株式会社)

SASを用いた抗体医薬品の各製造工程をつなぐ統合モデルの構築と統合モデルを用いた規格設定のリスク評価

伊藤勇弥(小野薬品工業株式会社)

RMST尺度における因果媒介分析の実装

石原政佳，今村拓未，渡辺秀章，長谷川貴大(塩野義製薬株式会社)

SASプログラムの副作用を考える

平井健太(株式会社タクミインフォメーションテクノロジー)

SAS及びRによるノンコンパートメント解析（NCA）の実装とWinNonlinとの結果比較：統合的な薬物動態解析プロセスへの展開

飯泉寛(Meiji Seikaファルマ株式会社)

ダミーデータを使った時間依存性共変量の解析

矢田徹(株式会社セブントゥワン)

SASの世代データセット機能やAudit機能を利用する

森岡裕(イーピーエス株式会社)

目次

臨床試験業務に使えるオープンソースSAS Package紹介

森岡裕(イーピーエス株式会社)

Rolling joinをSASで考えてみる

森岡裕(イーピーエス株式会社)

SAS環境で始めるR活用-LSAFにおけるRパッケージ利用とpharmaverse導入事例

田中優司，堀田真一(協和キリン株式会社)

Analysis Results Standardの解説

福田裕章(MSD株式会社)

CJUG-ADaMの活動について

大内喜海(CDISC Japan User Group)

生成AIは統計プログラマーの業務をどこまで支援できるか ～CJUG-ADaMテーマ５での2025年の試行錯誤から～

栗栖章(MSD株式会社)

生成AI支援による統計解析実務の有用性と課題

中島誉也，佐藤俊太朗(長崎大学)

九州大学病院における統計解析業務の体制と今後の展望

廣瀬将行，田川公士朗，岸本淳司，田中司朗(九州大学)

目次

SAS-MCP-Server：AIエージェントとSAS Viyaの連携による分析の民主化

長縄俊生(SAS Institute Japan株式会社)

TRIにおける統計解析業務の体制と効率化への取り組み

山川茜，加藤直樹，岡田悠，岡本里香，川本篤彦(神戸医療産業都市推進機構　医療イノベーション推進センター)

AROにおける特定臨床研究解析の事例紹介（神戸大学）

今井匠(神戸大学)

SAS proc fcmpによる行列演算処理を用いた野球の打撃力評価モデルOERAの適用事例…対戦チームを考慮したOERAスコアの考察

山下敏治，荒井遠樹，小池沙織，岡野かほる，砂村渉(株式会社セルネサイエンス)

統計解析プロセス効率化の先にあるもの

魚住龍史(東京科学大学)

リバーシチャンピオンシップ

出場者：
飯島泰平(ネイチャーインサイト株式会社)
関根伸吾(ネイチャーインサイト株式会社)
伊住寿雄(ソーバル株式会社)

兵庫県立大学におけるJMPの利用

～高砂市商業調査などを事例として

○川向肇[1]、中根滉大[2]、神棒美波[2]

（[1]兵庫県立大学社会情報科学部、[2]兵庫県立大学情報科学研究科）

The Utilisation of JMP at Hyogo Prefectural University

The application of the Takasago City Commercial Survey and other case studies as exemplars

Hajime Kawamukai, Kodai Nakane, Minami Kambo

School of Social Information, and Graduate School of Information Science, University of Hyogo

要旨

兵庫県立大学社会情報科学部川向研究室では，2023 年度より JMP を利用した卒業研究など各種の研究・教育に取り組んでいる．本稿では，本学川向研究室における JMP の利活用事例および最新の研究成果の一部を紹介する．

キーワード：JMP　学部及び大学院教育　活用事例　研究事例　オープンデータ

---本文開始点---

はじめに　兵庫県立大学社会情報科学部設置以前の JMP の利用

もともとアップル社の Macintosh 上で動作する統計パッケージとして開発されてきた経緯もあり，JMP の最大の特徴は GUI ベースの対話型ソフトウェアである点にある．プログラミングの経験が浅い学生や，プログラミングへの苦手意識を持つ学生・研究者が当時の SAS や SPSS，S，R といった統計パッケージを利用する際には，まずコマンドラインでの入力作業が必要であり，特定の統計分析を実施する際に適切なコマンドを入力しなければならないという課題があった．

しかし，開発当初から JMP は現在にも継承されている優れたユーザーインターフェースを備えており，Excel などの表計算ソフトウェアからのデータ移行も，CSV ファイル等への出力を経ることなく画面上のコピー&ペーストで完結できた．

兵庫県立大学の前身の一つである神戸商科大学では，このような特徴を持つ JMP に着目し，1990 年代中葉から授業等での活用を進めてきた．統計分析の初学者やプログラミング経験のない学生であっても，筆頭著者が独自に作成した 30 ページほどのマニュアルを渡せば特段の説明がなくとも統計分析を行えるというメリットがあり，神戸商科大学管理科学科の「意思決定論演習」における標準的なデータ分析ツールとして活用してきた．

2004 年に神戸商科大学，姫路工業大学，兵庫県立看護大学の 3 大学が統合された後は，当時の有馬昌宏教授を中心として，兵庫県立大学大学院情報科学研究科において統計学の背景を持たない学生向けに JMP を利用した教育が行われてきた．

なお，本学での研究分野における利用例としては，田中・川向・有馬(2005)，藤尾・有馬・川向(2005)，繆・有馬・川向(2005)で結果を紹介している兵庫県揖保郡新宮町（現たつの市）調査，有馬・川向(2007, 2008)などでの兵庫県たつの市における住民意識調査データの分析，川向・有馬(2009)での兵庫県明石市明舞団地の住民意識調査データの解析などが挙げられる．

金融機関の経営地盤に関する分野での活用事例としては，堀江・川向(1999)や堀江・川向(2002)など，2000 年前後の我が国の金融危機下における国勢調査や政府指定統計のメッシュデータ等を活用し，金融機関の再編に伴う店舗配置の分析等を行ってきた．

官民データ活用推進基本法の施行とそれに伴う教育環境の変化

2016 年に官民データ活用推進基本法が施行されたことで，データ分析教育を取り巻く環境は大きく変化した．

従来，研究や教育でデータを利用する際には，冊子体の公刊資料から再入力作業を行ったり，所管する省庁や外郭団体へ利用申請を行ってデータを入手したりする必要があったが，同法の施行以降は幅広いデータが容易に入手できる環境が整備されていった．当初は公開が義務付けられた国の省庁や都道府県を中心にオープンデータの提供が始まり，その後，公開が努力義務とされた市町村などの基礎自治体においても順次データが公開されるようになった．

デジタル庁（2025）によれば，現在 88%の自治体がオープンデータに取り組んでいると報告されているが，多くの自治体における公開内容は公園，公共トイレ，AED の所在地，国勢調査の町丁字別人口や住民基本台帳人口などにとどまっている．しかしながら，青森県青森市，埼玉県春日部市，東京都調布市，神奈川県横浜市，愛知県豊田市，兵庫県芦屋市および加古川市など一部の自治体では，市民意識調査等の結果について個人特定ができないよう秘匿措置を施したローデータ（raw data）を，CSV 形式ないし Excel 形式の機械可読なオープンデータとして公開している．

兵庫県立大学社会情報科学部設置以降での JMP の利用

兵庫県立大学社会情報科学部の教育理念と教育環境

このようなオープンデータの利用環境が整備されていく中，兵庫県立大学神戸商科キャンパスの学部再編に伴い，2019 年に同キャンパスへ社会情報科学部が設置された．同学部では文理融合型の教育体系を構築し，人工知能，統計学，情報数理などの情報科学と社会科学の視点を組み合わせた実践的な教育を推進している．これらを通じ，現代の複雑化する社会課題の解決に向けて，デジタル技術やデータサイエンスを縦横に駆使できる人材の育成を目指してきた．

さらに，課題解決型学習（PBL）の一環として，提携企業等から提供される実データを用いた演習を行い，社会実装力やビジネス思考を兼ね備えた IT 人材の育成を図っている．こうした教育を展開するため，大半の講義および演習科目において受講時の BYOD（Bring Your Own Device）を義務化し，学生にはノート PC の持参を要請している．

社会情報科学部専門科目「社会調査法」・「卒業研究」の教育と JMP

社会情報科学部の設置に伴い，筆頭著者は従来の大学院主担当から社会情報科学部の主担当となり，担当する講義内容や教育環境が大きく変化した．2021 年以降は同学部において「社会調査法」および「地理情

報システム」の講義・演習を担当し，これらに加えて 2022 年以降は卒業研究の指導も受け持つこととなった．

後期（10 月～翌年 2 月上旬）に開講される「社会調査法」では，11 月中旬頃まで社会調査の基本概念，調査手法，調査票作成，データコーディング法などの基礎的要素について講義を行う．後期後半の 11 月下旬から 2 月上旬にかけては，クロス集計や回帰分析の基礎概念の復習，解析出力結果の読み方などについて実習を交えて指導を行っている．その際の分析素材として，公開されている自治体の市民意識調査等のローデータを取り入れてきた．

2021 年からの 2 年間は，講義を学外施設で実施することが求められた事情から，受講生がデータ解析ツールとして 2 か月程度 JMP を利用できる環境を確保できず，やむなく EZR を利用して実習指導を行った．しかし，EZR は対話型インターフェースを備えるものの，ベースがコマンドライン駆動型の統計解析アプリケーションである R であるため，時折コマンド処理や分析手法の選択を自力で行う必要が生じ，受講生がデータ分析そのものに集中する上での障壁となることが確認された．また，文理融合型教育を標榜している背景もあり，計算機処理に苦手意識を持つ学生からは，コマンドライン入力に対して強い抵抗感が示されるケースも散見された．

2023 年からは学内施設での講義が可能となったため，演習用デスクトップ PC に導入されていた JMP を活用し，当時から市民意識調査の結果を提供していた兵庫県加古川市のデータを用いた教育を開始した．社会調査の基礎分析であるクロス集計分析などの理論と実践を，講義と演習を併用するスタイルで実施するとともに，加古川市の住民意識調査ローデータを活用した卒業研究指導にも着手した．JMP の対話型インターフェースや高品質なグラフ作成機能は，学生の「社会調査法」に対する学習意欲の向上や，卒業研究におけるデータ分析の効率化に大きく貢献した．しかし一方で，BYOD での PC 利用が一般化した学生にとって，利用場所や時間，期間に制約がある大学の固定 PC でしか JMP を使用できない点に対して，学生から強い不満が示されていた．

JMP Student Edition の供給開始と社会情報科学部での JMP の利用

そうした学生の要望に応える形で，2024 年 11 月上旬には 30 日間限定の JMP トライアル版を学生の BYOD 端末にインストールさせ，兵庫県加古川市の市民意識調査データを活用した分析実習を開始した．同年の「JMP Discovery Summit Tokyo 2024」において，JMP 事業部担当者より 2025 年 1 月から「JMP Student Edition」が無償提供される旨の案内があった．これにより，以降は「社会調査法」や「卒業研究」等で JMP を全面的に活用することが可能となった．

統計解析を伴う卒業研究はもちろん，JMP のグラフィカルインターフェースを活用したデータ抽出，複数ファイルとして提供されるテーブルの特定属性項目をキーとした突合（テーブル結合），同一ファイル構造を持つデータテーブルの連結など，大規模データをスムーズに処理できる能力を生かした教育・研究を展開している．具体例としては，複数年次データの連結や，複数の 1 次メッシュ領域単位（約 80km×約 80km）ごとに提供される 5 次メッシュ（約 250m×約 250m，最大約 10 万レコード，100 フィールド前後）の統計データテーブルの連結処理などが挙げられる．

2024 年からは，社会情報科学部の 3・4 年生が JMP Discovery Summit Tokyo に参加し始め，ポスターセッションにおいて研究成果の継続的な発表を行っている． 2024 年には，学内計算機環境の JMP で分析した加古川市市民意識調査（幸福度関連項目）の解析結果，JA 兵庫六甲の POS データと会員データという 2 種類のファイルを会員番号により対応を取りテーブル結合し，来場者の属性別購買特性および併売分析の結果，学部 3 年生による加古川市における交通事故の要因分析などの成果を報告した．

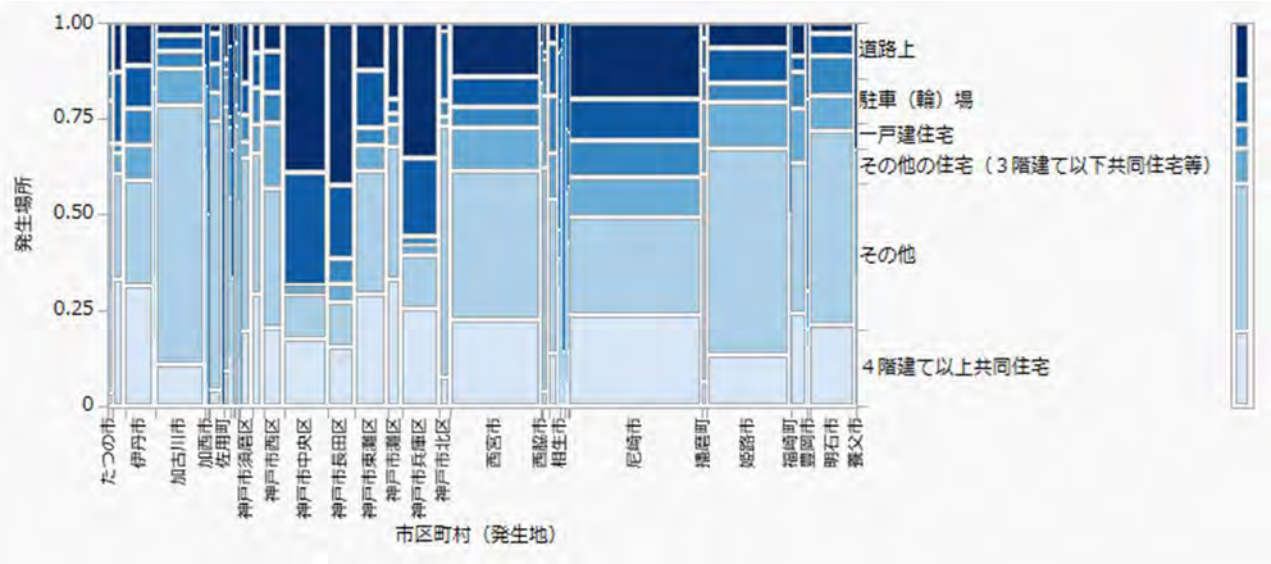


図 1　JMP による市町村別の自転車盗の被害の発生場所に関するモザイク図の例

2025 年度には，学部 4 年生の卒業研究テーマである，地域おこし協力隊導入自治体の社会経済特性分析，兵庫県北部の美方広域消防本部管内における消防庁提供データを活用した救急搬送の遅延要因分析，高砂市商業調査データを用いた多重対応分析の中間報告を行った．さらに，学内有志団体により，JA 兵庫六甲から提供を受けた農産物直売所の 260 万レコードにおよぶ POS データを活用した，付帯施設開設に伴う影響分析などの報告も行われている．

2025 年度の学部卒業研究として，會田・川向(2025)は，兵庫県南部の国勢調査および経済センサスのメッシュデータを利用した交通事故と地域の人口や産業立地についての分析，辻(2026)は，竹林の増減と地形及び地域の人口構造の変化に関する分析，川向・野田(2025)は，兵庫県下の救急搬送と消防本部の管轄域面積や域内の道路距離などに関する分析など，様々な分析を行っている．

また 2026 年前期，滋賀大学の島田貴仁教授が主催する「犯罪予防と社会安全」と題された講義・演習に本学の学部生及び大学院生が参加した．本演習には中央大学，早稲田大学，津田塾大学，名城大学，兵庫県立大学などの関係者が合同で参画し，滋賀県，神奈川県，大阪府，兵庫県，沖縄県などの各府県警察本部から提供された特殊詐欺（警察官成りすまし詐詐，ロマンス詐欺）や自転車盗に関する実データの分析と分析結果の可視化などに取り組んだ．

兵庫県立大学社会情報科学部の 4 年生および情報科学研究科の大学院生からなるグループは，JMP を活用してこれらの犯罪関連データの解析と可視化（図 1 など）を担当した．なお，図 1 からは，自治体により自転車盗の発生場所が異なることが容易に読み取れる．

複数名によるチーム分析において，JMP ファイル内に分析スクリプトを保存して容易に共有・再現できる点や，コードレスで直感的に高度な解析・可視化を行える点が分析作業の飛躍的な効率化に寄与した．演習に参加した学生からは，なぜこのような有用な分析ツールをもっと早い段階で教えてくれなかったのか，という声が寄せられるほど，その操作性と機能性が高く評価された．

現在，兵庫県立大学川向研究室では，2026 年度の卒業研究として JMP を活用した多様な研究を展開している．具体的には，五十嵐・川向(2026)では，神戸市西区における高齢化に伴う世帯減少（空き家増加）に関する分析，松位・川向(2026)は明石市における泉市政下の施策を含む子育て政策パッケージの変更効果を地価変化から検証した分析を行っている．さらに，石川・川向(2026)においては熊本県内における水災害に

よる経済的被害の推計，本田・川向(2026)では和歌山県下における救急搬送の遅延要因分析などの研究事例が挙げられる．

JMP Student Edition を利用した情報科学研究科での研究事例

2025 年からは，情報科学研究科の大学院生も修士論文作成に向けて JMP を利用した研究データの加工や分析を行っている．中根・川向(2025)では，ごみ排出量推計モデルに用いる地域分類に対して JMP の因子分析およびクラスタ分析を適用した結果について報告している．中根・川向(2026)では，JMP を活用した PLS 回帰によるごみ排出量推計モデルの予測結果や，将来の廃棄物処理施設の過不足に対応するための広域連携案について報告している．具体的には，廃棄物処理に関する将来的な一部事務組合の構成を検討するため，指定統計を活用した自治体類似度等に基づくクラスタ分析を行い，自治体間連携の可能性について検討した結果を紹介している．

神棒・川向(2026)では，株式会社 Agoop より研究用に提供を受けた約 125m×125m（6 次メッシュ）の区画領域内滞在人口データを用い，性別・年齢別・曜日別（平日／休日）における人流の地域集中特性およびその空間分布類似度の分析に JMP を活用している．具体的には，Getis and Ord (1995)が提唱した空間的集中度指標である G_i^* 統計量から計算される空間分布パターンについて，2 次元のタイル状に配列された各メッシュの G_i^*統計量を 1 次元配列に変換し，JMP を用いた通常のクラスタリング手法によって求めた類似度と， 画像品質評価手法である FSIM（Feature Similarity Index Model）を用いて評価した画像構造類似度との比較検証結果を取りまとめている．

なお，本年 11 月の JMP Discovery Summit においても，これらの JMP を活用して得られた知見について発表する予定である．

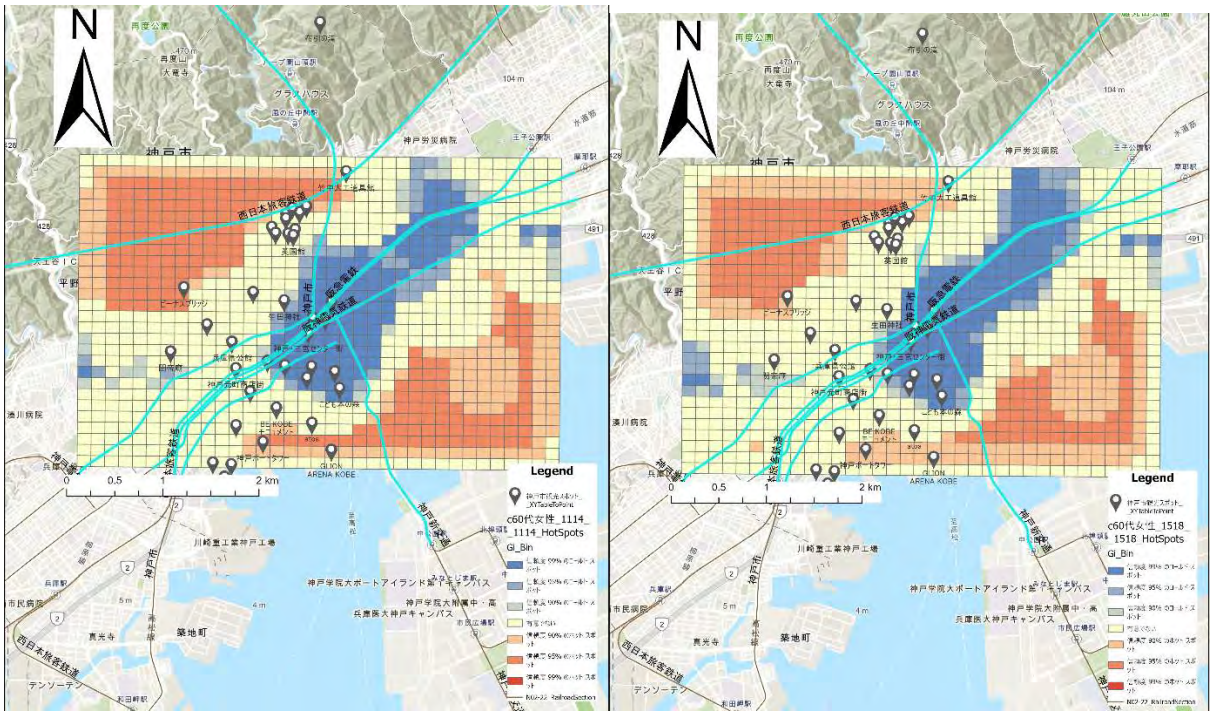


図 1 ホットスポット分析結果の類似度が上位の地図の例（クラスタ分析）

（2026 年 7 月, 60 代居住女性,11～14 時（左）及び 15～18 時（右））

兵庫県高砂市の商業実態調査での JMP の利活用

高砂市商業実態調査について

高砂市と兵庫県立大学は 2017 年に包括連携協定を，高砂商工会議所と兵庫県立大学は 2019 年に産学連携協定をそれぞれ締結している．こうした背景から，2024 年に高砂市および高砂商工会議所より，商業実態調査の実施可能性に関する照会を受けた．この照会以降，複数回の調整会議を経て，2025 年 4 月に高砂市から高砂商工会議所へ商業実態調査業務が委託され，高砂商工会議所から調査業務の大半を兵庫県立大学が正式に受託・実施する運びとなった．

受託した調査業務を充実させるため，調査票の設計に先立ち，10 回以上に及ぶ現地調査（学生を引率しての現地調査は計 4 回），前回調査票の詳細な検討，および他自治体における先進事例との比較検討を行い，調査票原案を作成した．また，高砂市からの要請もあり，回答選択肢を含めた設計段階から学生の意見を反映させた．この原案について，高砂市，高砂商工会議所，兵庫県立大学の三者による 10 余回の事前協議を重ね，調査票を確定し，商業実態調査を産官学連携により共同で実施した．設問数は，消費者調査で 50 問，商業・サービス業事業者調査で 41 問となった．2025 年調査では，社会関係資本（ソーシャル・キャピタル）の多寡が消費行動や事業活動に与える影響を検証するために関連質問項目を設定し，地域の社会構造と経済活動の双方を把握する分析を試みた．

2025 年 8 月に高砂市役所が調査票を郵送配布し，同年 9 月中旬を回収期限として実施された．消費者調査では，高砂市役所により無作為抽出された市民 3,000 名を対象に郵送配布を行い，自記式での郵送返送または Web フォームからの回答を併用する方式をとった．その結果，全回収数は 1,562 票（回収率 52.1%）となり，うち郵送回収が 1,303 票（同 43.4%），Web 回答が 259 票（同 8.6%）であった．

一方，商業・サービス業調査は，高砂商工会議所の協力のもと実施された．商工会議所の通常業務における事業所訪問を活用した機縁法（訪問配布）を採用し，回収は消費者調査と同様に郵送および Web 回答の双方を可能とした．配布数 200 票に対し，全回収数は 120 票（回収率 60.0%）であり，その内訳は郵送回収が 65 票（同 32.5%），Web 回答が 55 票（同 27.5%）であった．

2025 年調査では，事前現地調査から得られた「地域ごとに消費傾向や事業環境が大きく異なる」という知見を踏まえ，回答者の基本属性として居住地および事業所所在地の郵便番号の記載を求めることとした．これにより，従来の小学校区単位よりも詳細な位置特定が可能となっただけでなく，郵便番号と小学校区，ならびに市内の 8 行政区域との対応表を作成することで，前回調査との比較が可能な集計結果の出力を可能とした．

調査データの入力作業については専門の外部事業者に再委託し，10 月中旬に入力データの納品を受けた．

高砂市商業実態調査の分析での JMP の利用

本調査データのクリーニング，単純集計，および分析用 JMP データファイルへの加工（初学者である学生でも容易に分析が進められるよう，名義尺度や順序尺度といった各変数の尺度設定を含む）は，すべて筆頭著者が事前に実施した．

実際の分析作業，モザイク図などの作図，および解説文の作成自体は，学内の 2 つの学生団体への業務委託として実施された．その際，データ規模の大きい消費者調査は大学院生を中心とする団体へ，データ量が比較的コンパクトな商業者調査は学部生を中心とする団体へそれぞれ委託した．これらの委託作業に先立ち，学生に対しては JMP の操作手順や分析手法を取りまとめたマニュアルを作成・配布し，事前講習会を

実施した．こうした念入りな事前準備により，報告書の基礎となる集計・分析原稿は極めてスムーズに作成された．

クロス表分析の結果の取捨選択などの整理，ならびに図 3 に例示したモザイク図の作成や図 4 に例示した地図表現を踏まえた上で，現地調査などの知見を加えた政策提言等の部分について筆頭筆者が加筆・修正を行い，最終報告書としてとりまとめた．

なお，市域全体の傾向把握に重点が置かれていた前回調査の報告書に対し，2025 年調査では地域ごとの詳細な商業活性化策を検討できるよう，回答傾向の空間的特性を可視化した地図表現を含む最終報告書原案を作成し，2 月下旬までに市役所へ提示した．

JMP の利用と高砂市商業実態調査を利用した教育

一方，2025 年の「社会調査法」の講義では，11 月中旬より商業・サービス業調査の調査票およびデータクリーニング前の JMP データファイルを学習用素材として受講生に提供した．あらかじめ未加工のデータを取り扱わせることで，データクリーニングの必要性（特に郵便番号の誤記検出等）を体得させた．

その上で，単純集計，クロス集計，さらには集計結果からの客観的な含意導出に関する実習を行った．その際，名義尺度や順序尺度である変数もあえて連続尺度に設定した状態で自由に分析を行わせ，その後に本来の適切な尺度へ再設定して分析し直す実習を取り入れた．これにより，尺度に即した分析手法選択の必要性や，社会調査における変数の尺度設定の重要性を効果的に学習させる教材として活用した．

受講生からは「1 年次の『統計学』などの講義で尺度について短時間触れられた程度では実感が伴わなかったが，JMP を用いた実践的な分析実習を通じて，統計解析における尺度の重要性をはじめて体験的に理解できた」といった声が寄せられた．

なお，1 年次対象の「基礎ゼミナール」では，レポートの論理構成法や Microsoft Word・Excel を活用したレポート作成スキルの向上を図りつつ，兵庫県ないし県下自治体のオープンデータを用いた分析を行わせている．その際，分析ツールとしては Excel のピボットテーブル機能を使用させている．この段階で JMP を導入していない背景には，いくつかの理由がある．

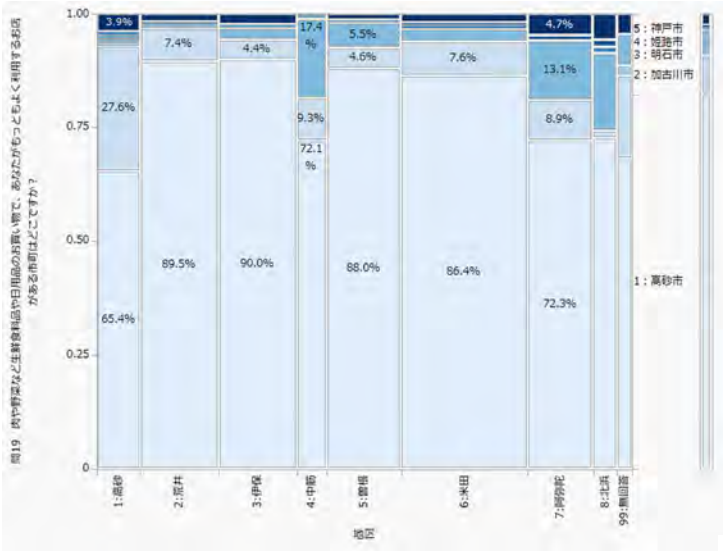


図 3　生鮮食料品や日用品購入の際に，もっともよく利用する店舗がある市町村の地域別回答結果



図 4　生鮮食料品や日用品購入の際に加古川市の店舗を利用する回答者比率の地図表現

第一に，コロナ禍以降の高校教育においてタブレット学習が普及した結果，新入生の 3 分の 1 以上が入学時点で高校や自宅での Excel 使用経験がなく，表計算ソフトウェアによる基礎的なデータ処理スキルが身についていない点が挙げられる．第二に，卒業後の企業等では表計算ソフトウェアの利用が一般的であり，社会人としての汎用的な基本スキルを早期に修得させる必要があるためである．さらに第三の理由として，早期に JMP という極めて強力な対話型ツールを与えてしまうと，1 年次後期以降に実施される Python 等を用いたプログラミング学習への意欲を低下させかねないという，教育的配慮に基づく判断が挙げられる．

データサイエンス関連学部の急増に伴う変数の尺度に応じた透析分析手法教育の重要性

本節では，名義尺度および順序尺度を多数含む社会調査データの分析例として，「高砂市商業実態調査」における消費者調査データを対象に，JMP を用いて実施した実証分析の事例を提示した． 近年の計算機性能の目覚ましい向上，ハードウェアの低価格化，ならびにユビキタス環境の定着に伴うデータ流通の円滑化は，特段の事前準備を要することなくデータ分析に着手し得る環境を社会的に広範にもたらした．また，分析ツールに関しても，Jamovi，R，EZR などの各種パッケージや，Python の Anaconda 環境における「pandas」「NumPy」「Matplotlib」「Seaborn」といった各種ライブラリの活用により，重回帰分析をはじめとする多変量解析の基礎的理論の理解を欠いた状態であっても，形式的な計算結果の算出自体は容易となっている．

さらに，これら分析ツールの利用手順を詳細に解説した Web サイトが膨大に存在し，画面推移や操作手順を参照することで所望の出力結果を安易に取得可能な状況が生じている．加えて，生成 AI 技術の発達に伴い，適切なプロンプトを入力することで，分析データの整形や具体的な処理コードの記述に至るまで，極めて精緻かつ妥当性の高い回答が即座に生成される環境が確立しつつある．

しかしながら，こうした高機能な AI 技術や分析環境の普及に伴い，単に出力結果の獲得のみを目的としたデータ処理が行われ，その背景にある統計的理論や得られた分析結果の学術的帰結を十分に吟味・考察できない学生の増加が，卒業研究や修士論文等の指導の現場において顕著となっている．筆者はこの現状に対して強い危機感を抱いており，こうした懸念は本学にとどまらず，近年のデータサイエンス系学部・学科を標榜する多くの高等教育機関においても共有される普遍的な課題であると推察される．

分析の実務家・研究者・教育者がこうした状況に直面する今日において，大学等の高等教育機関や実務者向け研修等の機会を通じ，名義尺度・順序尺度・間隔尺度・比例尺度といった各変数の属性と，それらに対応する適切な分析手法との整合性に関する体系的教育を再構築することが強く求められよう．同時に，特に初学者たる学生に対しては，多種多様な実データに触れさせることを通じ，分析手法の選定ならびにその適用妥当性に対する実践的な知見と感性を涵養させることが不可欠であると考える．

JMP はその機能的特性として，変数の尺度属性をあらかじめ適切に指定することで，対応する最適な分析手法が自動的に選択され，必要な統計量が出力される設計となっている．R や Python といったコマンドライン駆動型の分析ツールと比較した場合，分析手法の探索や該当コマンドの検索・入力といった一連の試行錯誤に伴う負担を大幅に軽減し得る．

近年の生成 AI による各種支援が可能となったことも相俟って，分析手順の設定にかかる作業的負担は格段に低減されたといえる．その意味において，JMP は現代の AI 時代における分析プロセスの効率化を先取りしたツールとして位置付けることができる．

ところで，文部科学省の方針等を背景とした大学・学部の再編に伴い，データサイエンス関連の学部・学科は急増の途にある．兵庫県立大学社会情報科学部は，国公立大学において同分野での人材育成を目指す先駆的な学部の一つとして発足した経緯を有する．

同学部における教育実践を踏まえると，学生の卒業研究や演習科目等でのデータ処理において，本来順序尺度や名義尺度として扱うべき変数を不可分に連続尺度として処理する事例が散見される．また，記述統計量の算出やクロス集計といった基本的な処理を行い，データ属性の検証を経ることなく，安易に高度な多変量解析手法を適用し，その出力をそのまま分析結果として提示しようとする傾向も学部内において観察される．さらに，高年次生であっても変数の尺度属性に対する認識を欠いたまま不適切な解析を試みる事例が一定数存在する．

学部のカリキュラムにおいては，低年次段階の講義において基礎的な分析手順が扱われているものの，全15回の講義のうち数回程度にとどまること，加えて分析手法の理論的背景や実践的感覚の修得よりも単なる単位取得を優先する学習姿勢も手伝い，背景知識を十分に定着させ得ていない学生が一定数生じている．その結果，変数の性質に応じた適切な前処理を講じることなく，名義尺度を連続尺度とみなして複雑な多変量解析を安易に適用しようとする事態がみられる．

分析手法の選定や処理手順にかかる負担が軽減されている現状に鑑みれば，本来であれば得られた出力結果の解釈や考察に対してより多くの時間とリソースが充当されるべきであろう．初学者にとって極めて重要な意味を持つのは，出力結果に対する深遠な解釈と議論のノウハウの蓄積に他ならない．そのためには，分析結果と適用手法の妥当性を巡り，対話的かつ共同的な議論を重ねる教育プロセスが極めて有効であると考える．JMP はデータファイルの共有と尺度設定によりスムーズな統計分析作業における試行錯誤を可能にする点において，こうした結果の解釈に関する対話を展開するための基盤的ツールとして優れた適性を有していると評価できよう．

まとめに代えて　これからの兵庫県立大学の JMP 教育

本稿で紹介したように，2025 年 1 月の JMP Academic ライセンス制度の導入により，兵庫県立大学における教育環境は一変した．JMP を活用することで，学生がデータ分析に主体的に取り組むことが容易になると同時に，教員にとっても指導対象の学生と分析用ファイルを共有し，分析内容を相互に確認しながら適切な指導を行える環境が整った．

今後も JMP を活用した教育を継続する予定であるが，現在の課題はその定着と展開にある．本学社会情報科学部では，1 年次後期より Python を用いた教育が行われていることもあり，大半の研究室では Python が標準的な分析ツールとして利用されている．また，指導教員の専門分野によっては Stata 等が指定される場合もある．

こうした多元的な統計パッケージが併用され，生成 AI が一般化する環境下において，学部および大学院教育の中で JMP を利用しつつ，対話的かつ共同的な議論を重ねる教育の実践を通して確立することが，筆頭著者にとっての今後の主要な課題であると認識している．

参考文献

會田安優奈・川向肇(2025), 兵庫県における交通事故の空間分析,『CSIS Days 2025 研究アブストラクト集』, A02.

有馬昌宏・川向肇(2007), 地域社会の有効なマネジメントのための住民ニーズ表出へのＡＨＰの適用可能性に関する基礎的研究－順序尺度による個別評価とＡＨＰによる総合評価の比較から－,『地域学研究』, 37(4), 931-946.

有馬昌宏・川向肇(2008), 地方自治体における行政施策評価へのコンジョイント分析適用の可能性と課題,『地

域学研究』，38(3)，629-648.
五十嵐史織・川向肇(2026)，神戸市西区における戸建て住宅の地価の空間分布，『CSIS Days 2026 研究アブストラクト集』，(forthcoming).
石川優斗・川向肇(2026)，水害被害額による産業別 GDP への影響分析―熊本県における 2011-2022 年のデータを用いた実証研究―，『CSIS Days 2026 研究アブストラクト集』，(forthcoming).
川向肇・野田晴菜(2025), 兵庫県下における救急搬送データを利用した現地到着所要時間に関する分析,『CSIS Days 2025 研究アブストラクト集』，A02.
川向肇・有馬昌宏(2009), 再生が課題となるニュータウン地域における住民の地域関与意識,『都市住宅学』, 2009(67)，98-103.
高砂市(2026)，高砂市商業実態調査，消費者・商業者サービス事業者アンケート調査回答【商工労働課】,『姫路市・播磨圏域連携中枢都市圏オープンデータカタログサイト』，https://city.himeji.gkan.jp/gkan/dataset/commerce（2026 年 9 月 4 日最終アクセス）.
田中有紀・川向肇・有馬昌宏(2005)，JMP を活用した住民参加型まちづくりのためのデータ解析と地理情報システムでの連携運用法の提案，『SAS User Forum ユーザー会 学術総会 2005 論文集』，pp.403-414.
辻栞奈(2026)，空中写真を利用した竹林の機械判別可能性と竹林への遷移要因の定量研究についての考察：淡路市の事例から，兵庫県立大学学士論文.
デジタル庁(2025)，2025 年デジタル庁活動報告，https://www.digital.go.jp/policies/report-2025（2026 年 9 月 4 日最終アクセス）.
藤尾俊幸・有馬昌宏・川向肇(2005)，JMP を利用した CVM による政策評価 ―ローカル鉄道のサービス向上の金銭的評価を事例に，『SAS User Forum ユーザー会 学術総会 2005 論文集』，pp.415-424.
堀江康熙・川向肇(1999)，小規模金融機関の経営基盤，『経済学研究』，第 66 巻第 3 号，197-226，九州大学経済学会.
堀江康熙・川向肇(2002)，信用金庫の営業基盤と合併問題，『九州大学経済学研究』，第 68 巻第 4・5 号，83-116，九州大学経済学会.
本田桃華・川向肇(2026)，和歌山県下における救急搬送データを利用した医療機関等収容所要時間に関する分析，『CSIS Days 2026 研究アブストラクト集』，(forthcoming).
松位美咲・川向肇(2026)，自治体政策の土地取引価格を介した評価可能性の検証―明石市を事例として―，『CSIS Days 2026 研究アブストラクト集』，(forthcoming).
繆青・有馬昌宏・川向肇(2005)，JMP を活用した住民意識調査データに基づく行政課題の構造に関する計量的抽出法の検討，『SAS User Forum ユーザー会 学術総会 2005 論文集』，pp.427-436．フ

高砂市商業調査の分析におけるJMPを利用した多重対応分析の利用

○川向　肇

（兵庫県立大学　社会情報科学部）

The application of Multiple Correspondence Analysis to the examination of the Takasago City Commercial Survey, utilising JMP software.

Hajime Kawamukai

Department, Corporate / University

要旨

本稿では，高砂市と兵庫県立大学との包括連携協定，ならびに高砂商工会議所と兵庫県立大学との産学連携協定に基づき 2025 年に実施した「高砂市商業実態調査（消費者調査）」のデータに対し，JMP による多重対応分析を適用した分析事例およびそこから得られた知見について報告する．

キーワード：JMP,　多重対応分析，高砂市，商業実態調査

---本文開始点---

はじめに

高砂市の商業実態調査について

高砂市と兵庫県立大学との包括連携協定，および高砂商工会議所と兵庫県立大学との産学連携協定に基づき，2025 年に以上の三者が協働して商業実態調査を実施した．本調査の実施に至る詳細な経緯については川向・中根・神棒（2026）を参照されたい．

商業実態調査のうち「消費者調査」においては，高砂市の住民基本台帳に登録されている 18 歳以上の市民から 3,000 名を無作為抽出し，調査票（一部 Web 回答を併用）を用いた居住者の意識調査を実施した．有効回収数は 1,562 票（回収率 52.1%）であり，その内訳は郵送回収 1,303 票（43.4%），Web 回答 259 票（8.6%）であった．

回答者の回答負担の軽減および回収後のデータ解析の効率化を図るため，調査票末尾に設けた自由記述 1 問を除き，すべて選択肢からの単一回答方式として設計した．単一回答式の設問数は計 49 問である．また，生鮮食料品や日用品，あるいは衣料品・家具・家電等の 1 回あたりの支出額や週あたりの購入頻度など，本来は連続尺度として取得可能な項目についても，回答者の心理的負担を考慮して階層化した順序尺度として設計・導入した．なお，具体的な調査票については高砂市（2026）において公開されている．

社会調査のデータ処理における尺度の処理方法について

自治体が実施する住民意識調査や各種企業による顧客満足度調査等においては，品質・接客・重要度・継続利用意向などの要素について，回答者に 3 段階から 10 段階程度の数値を提示した尺度形式（順序尺度）で測定されることが多い．これらの項目は数値として記録されるため一見すると連続尺度（間隔尺度・比例尺度）のように理解されることもあるが，本質的には選択肢間の順位・大小関係のみを表す「順序尺度」であり, 加減乗除等の四則演算をそのまま適用することはデータの性質上妥当ではないと考える. しかしながら，実務や一部の研究においては，分析の簡便性を優先するあまり，これらが事実上の連続尺度として扱われる事例が散見される．

近年，DX（デジタルトランスフォーメーション）が多くの企業や組織で提唱されることに伴い，組織内での各種データの二次利用が急速に拡大している．それと同時に，統計的素養を必ずしも十分にもたない分析者がデータ処理に携わる機会が増加しており，適切な尺度の変換や分析手法の吟味を経ないまま結論が導出されるリスクが高まっている．実際，高等教育の現場においても，名義尺度や順序尺度の変数に対して安易に平均値や分散を算出する事例，順序尺度間で通常の相関係数を求める事例が散見される．

これに対し統計解析ソフトウェア JMP は，各変数に対する尺度属性をあらかじめ適切に設定しておくことで, 推奨される適切な分析手法や各種統計量, 出力グラフが提示されるという優れた機能特性を備えている．分析手法を定型化しておき，JMP のデータテーブルに保存しておき，他者の分析手法とその結果を容易に参照することも可能である．より高度な分析のために，JSL（JMP Script Language）を活用した処理手順の自動化を行えば，かなり高度なデータ加工・分析手順をテンプレート化することが可能となり，分析担当者の経験によらず一定水準の出力結果を獲得できる．

データ収集が容易となった現代において，組織内に蓄積された未分析の社会調査データや定量的データを有効活用するうえで，JMP の適用は極めて有用な解となり得る．

ただし，JMP のようなツールでは，統計学の背景が薄い利用者が操作することが可能になる．それと同時に，分析手法の前提条件に対する理解を欠いたまま不適切な解釈が行われるリスクも懸念される．したがって，単なる操作手順の修得にとどまらず，変数の尺度属性，各解析手法の数学的前提，出力結果の構造的意味合いについて深く教授する教育プログラムの整備が不可欠であると考える．

高砂市の商業実態調査における多重対応分析の利用

社会調査における多重対応分析の利用のメリット

一般的な社会調査報告書においては，単純集計（JMP における「一変量の分析」）や特定の主要変数間におけるクロス集計（「二変量の分析」）が分析の中心となる．高砂商工会議所（2026）の成果物作成時においても，主にクロス集計およびモザイク図を用いた可視化によって地域消費行動の特性を記述した．しかし，設問数が 49 問に及び，各設問に 2〜8 程度の選択肢が存在する調査体系においては，作成されるクロス集計表の総数が膨大となり，多変数間に横たわる構造的な関係性を多角的に把握することは容易ではない．

これに対し，名義尺度に対する多変量解析手法として位置付けられる「多重対応分析（Multiple Correspondence Analysis: MCA）」を適用すれば，多数のカテゴリカル変数群における選択肢間の幾何学的・視覚的な関係性を低次元空間上に要約して明瞭に示すことが可能となる．

筆者はこれまで，旧揖保郡新宮町（現たつの市新宮町）の住民意識調査（繆・有馬・川向 2005）等において JMP を用いた対応分析を導入してきた．しかし，過去のバージョン（JMP 5.1 等）では二変量対応分析にとどまり，3 変数以上の解析には特殊なデータ加工を要したが，JMP 12 以降は多重対応分析の UI が標準

化され，多変量のカテゴリカルデータの解析が著しく容易となった．以下では，JMP Student Edition 19.1.5 を用いて得られた分析結果を示す．

高砂市商業環境実態調査のデータでの分析例

オープンデータとして提供されている消費者調査データ（高砂市 2026）より，無回答（欠損値）を除外した 1,277 サンプルを抽出し，以下の主要変数について多重対応分析を実施した．

名義尺度： 生鮮食料品・日用品の最頻購入市町，ならびに主要購入店舗形態

順序尺度： 購買時の重視要素（価格，品質，自宅からの近さ等の 4 段階評価），および高砂市の商業環境全般に対する満足度（4 段階評価）

図 1 における赤色破線の矩形領域に着目すると，「品質の良さ」「品揃えの豊富さ」「価格」「接客態度」「自宅からの近さ」といった一般的な購買決定要因について「重視しない」と回答した消費者層のクラスタが形成されていることが確認できる．注目すべきは，当該クラスタが次元 1・次元 2 の空間上において「個人商店」を選択した回答者群と比較的近接した領域の北東方向に位置している点である．

このような結果となった背景としては，市内の個人商店や伝統的商店街において昭和期からの店舗形態や営業様式がそのまま維持されているという現地調査での知見と整合的であると解釈できる．

次に青の一点鎖線で囲った部分を拡大した図 2 について検討する．

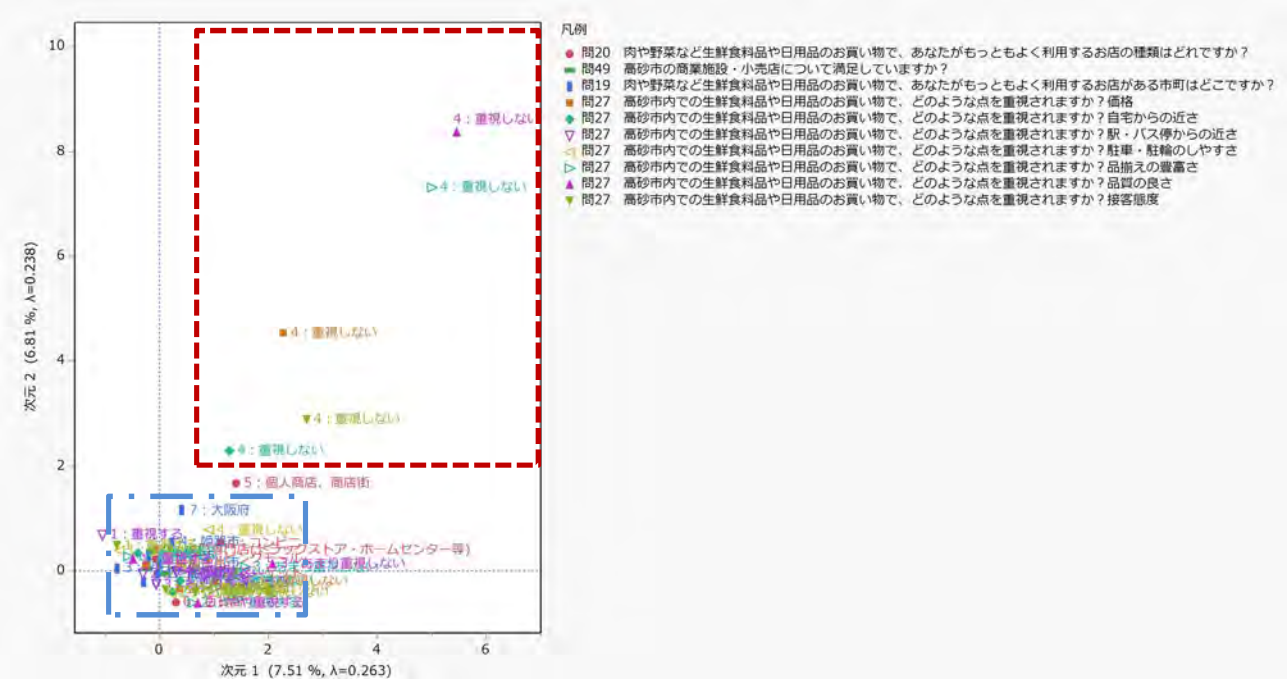


図 1　対応分析の結果　（次元 1　と　次元 2）

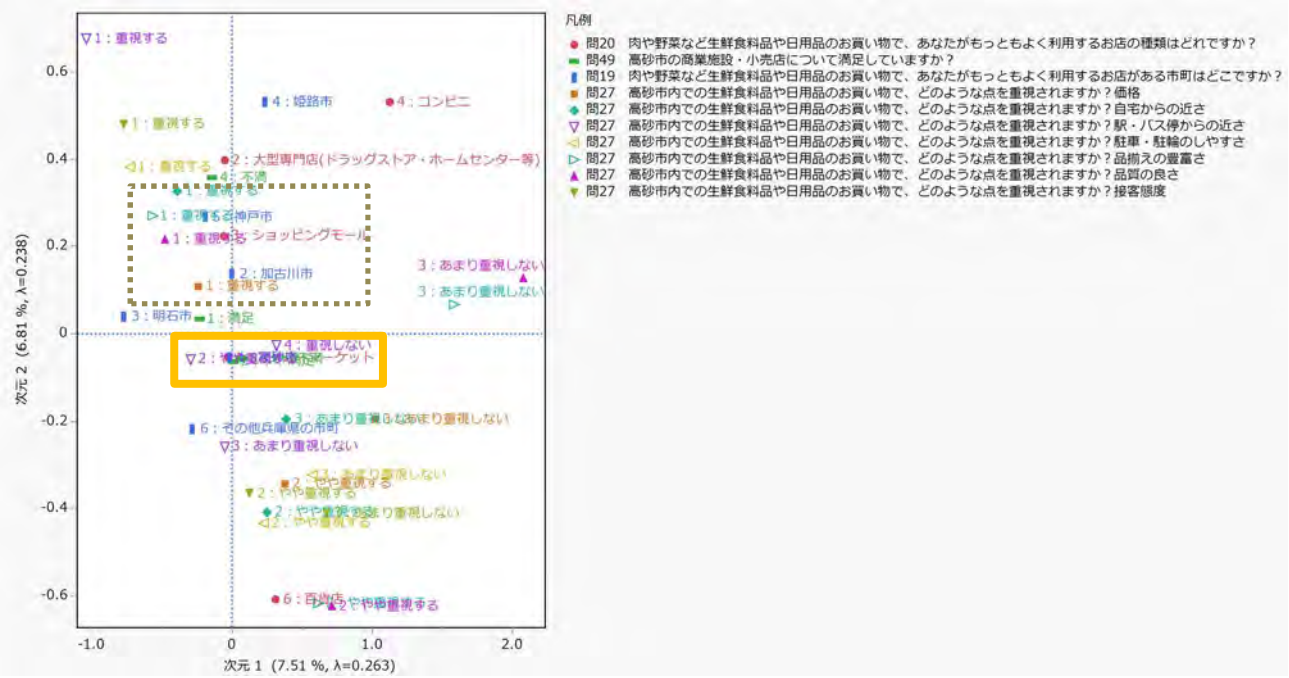


図 2　図 1 の青の一点鎖線部分を拡大図

図 2（図 1 の青色一点鎖線に対応する領域）のうち，ベージュ色破線の矩形領域について検討する．の領域内では，生鮮食料品・日用品の購入先として「神戸市」や「加古川市」を選択し，かつ店舗形態として「ショッピングモール」を選択した回答者群と，「品揃えの豊富さ」「品質の良さ」「価格」を「重視する」と回答した群が明確に近接している．これは，品質や品揃え，価格競争力を強く求める層が，大規模商業施設が集積し競合環境が整っている周辺主要都市（神戸市・加古川市）へ流出している状況を如実に物語っている．

一方で，図 2 の黄色実線の矩形領域には，購入先を「高砂市内」，店舗形態を「スーパーマーケット」とし，商業環境への満足度を「やや満足」「やや不満」とする層が配置されている．

図 3 のモザイク図からは，高砂市内で日常買いを行う消費者の 90%以上が「スーパーマーケット」に依存していることが読み取れる．また図 4 からは，高砂市内を中心に購買を行う回答者の満足度評価が，「満足」「不満」の両極端へは過度に偏重せず，中間的な評価としている回答者が多数存在している構造が示されている．

このように，多重対応分析を起点として個別のクロス分析表（モザイク図）の検証を組み合わせるアプローチは，回答者の潜在的な購買傾向や意識構造を迅速かつ包括的に把握するうえで極めて有効である．個別のクロス表を 1 つずつ積み上げて全体像を把握する従来手法に比べ，調査結果をより効果的・効率的に解釈できる分析ツールとして，自治体の住民意識調査や企業の顧客満足度調査等において一層活用されることが期待される．

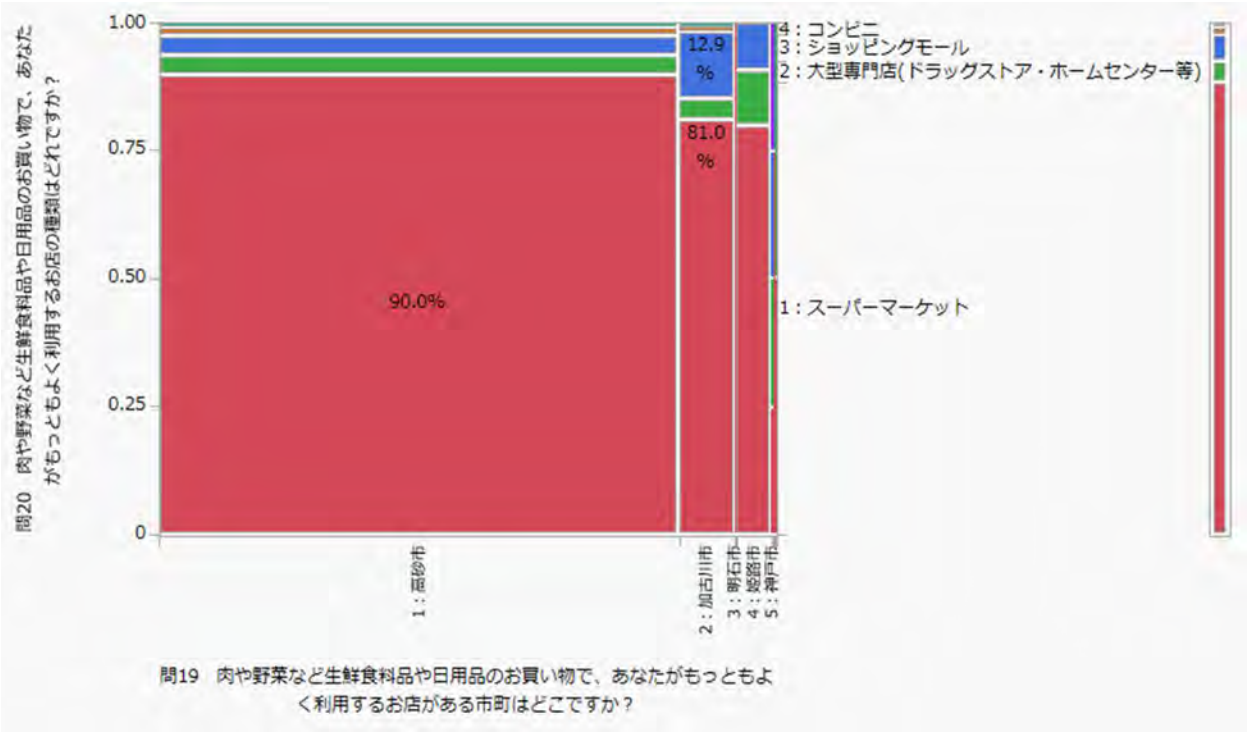


図 3　生鮮食料品や日用品についての購入先市町と購入先の商業施設種別についてのモザイク図

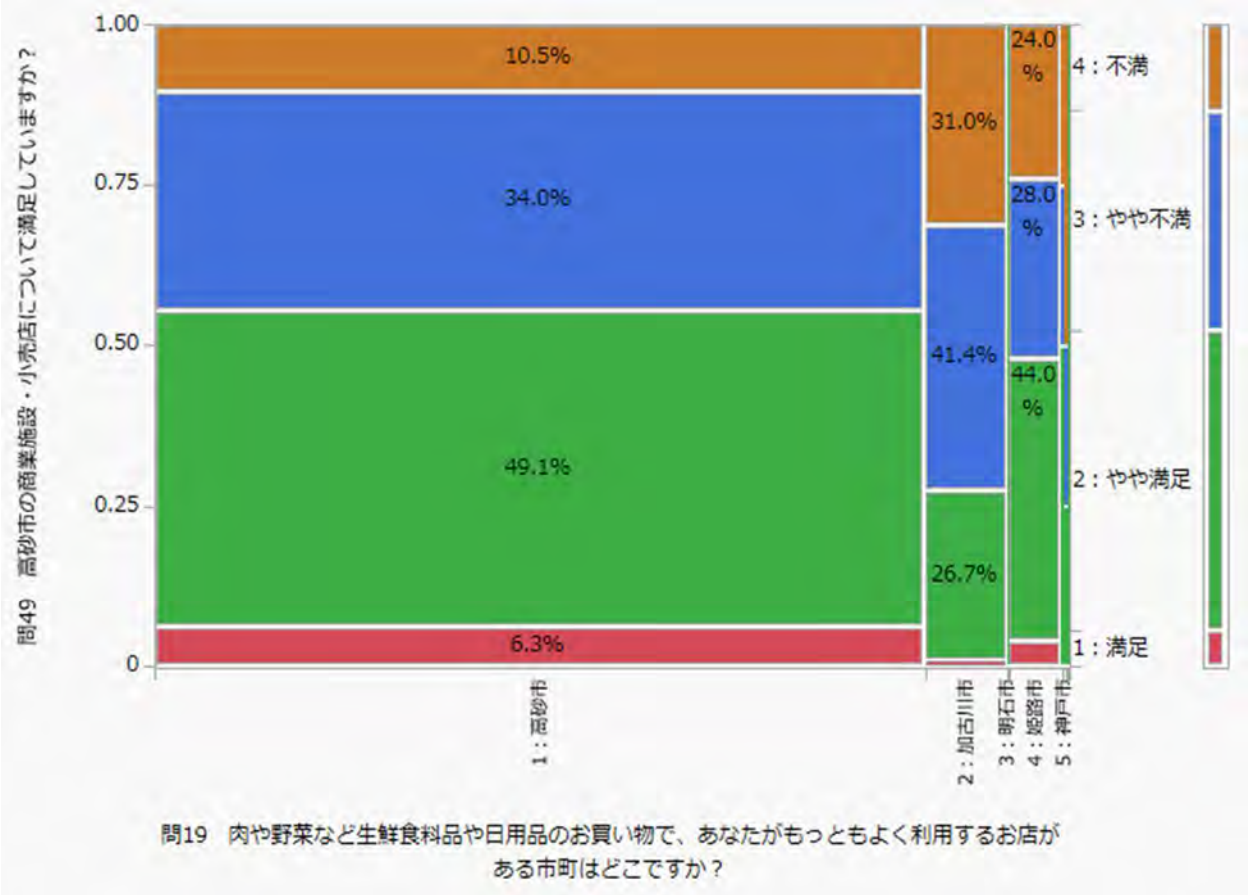


図 4　生鮮食料品や日用品についての購入先市町と高砂市内での商業環境に関する 4 段階評価での満足度に関するモザイク図

生鮮食料品や日用品の購入先自治体・小学校区・市内の商業環境に関する満足度の分析

続いて，分析対象項目を「高砂市内商業環境への全体満足度（4 段階）」「日常品の最頻購入市町（7 自治体群）」「居住小学校区（10 区）」の 3 変数に絞り込み，空間的・地域的特性に焦点を当てた多重対応分析を実施した結果（図 5）を紹介する．

図 5 の赤色の一点鎖線による楕円部には，小学校区では，「１：高砂小学校区」，生鮮食料品や日用品の購入先商業施設所在地については「２：加古川市内」，高砂市内の満足度に関しては「４：不満」が隣接して配置されている．

このような配置の背景には，歴史的・地理的な商業構造の変容が存在する．旧高砂町域（現高砂小学校区周辺）は，かつて加古川水運と瀬戸内海運の結節点たる旧高砂湊を抱え，東播地方における物資集散の中心的商業地として繁栄していた．近代以降も大企業工場の立地や通勤流動を背景とする十分な消費の存在，1976 年の「サンモール高砂」開業等により，高砂小学校区は市内の核的商業機能の位置を保ち続けてきた．

しかし，日本国有鉄道高砂工場の閉鎖，近年の企業での生産設備投資に伴う省力化や，産業構造の変化に伴う事業所・通勤人口の減少，JR 高砂線の廃止といった交通体系の変化に加え，核店舗（西友）の撤退や建物の老朽化・耐震性不足により，2017 年 12 月にサンモール高砂は閉鎖を余儀なくされた．これに伴い周辺商店街においても商業者・サービス事業者の休廃業が相次ぎ，いわゆるシャッター街化が急速に進行した．現在，同校区内においては生鮮三品のうち，精肉を除く青果・鮮魚を専門に扱う小売店舗の営業は極めて限定的な状況にある．

2024 年～2025 年にかけた現地調査においても，サンモール跡地に近接するドラッグストア等が一部生鮮食料品を供給する商業施設としての機能を担っているものの，住民の日常的購買需要を十分に充たし得ていない可能性がある実態が確認されている．

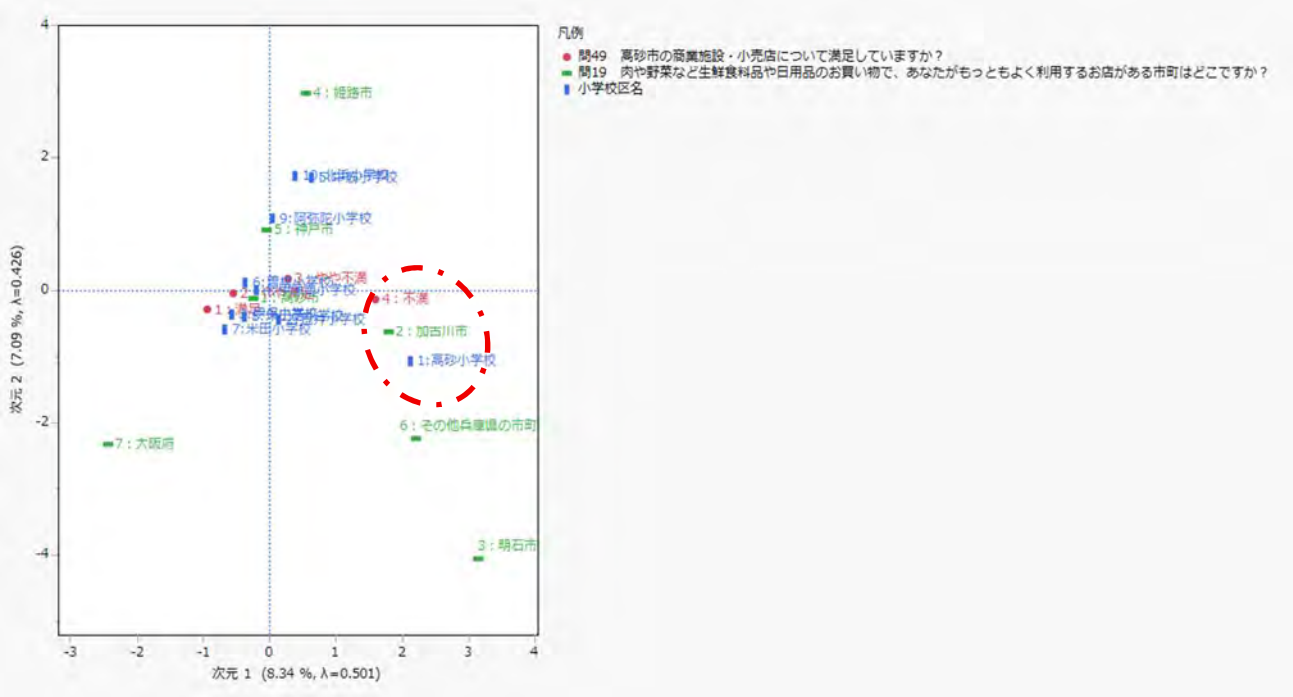


図 5　商業環境全体に関する 4 段階評価での満足度，生鮮食料品や日用品を最も頻繁に利用する商業施設がある市町（7 自治体群），小学校区による多重対応分析の結果

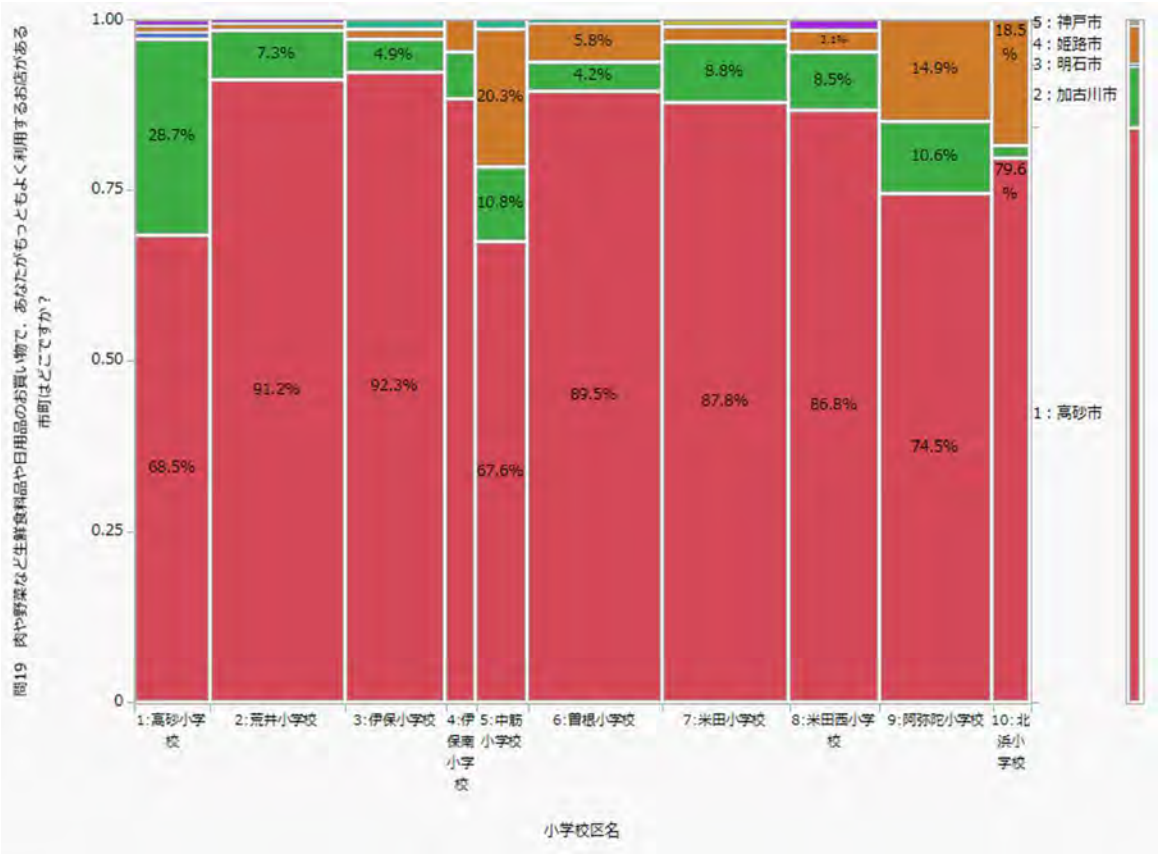


図 6　小学校区別の生鮮食料品，日用品の購入先の商業施設がある自治体についてのモザイク図



図 7　高砂市内の小学校区と高砂市内の商業環境に「不満」・「やや不満」と回答者した回答者比率の和に関する題図

図 6 に示すように，高砂小学校区の住民が生鮮食料品・日用品の購入先を加古川市とする割合は 28.7%に達し，市内他地区と比較して突出して高い．図 7 に示している地理的関係からも明らかなように，市域を接する加古川市の大型ショッピングモールへの交通利便性（加古川バイパス・JR 神戸線・山陽電鉄等）が存在することが，こうした購買行動の流出につながっていることが考えられる．
一方，高砂市北西部に位置し姫路市と境界を接する中筋・北浜・阿弥陀の各小学校区においては，姫路市内を最頻購入先とする割合がそれぞれ 20.2%，18.5%，14.9%と高い値を示している．これは，加古川バイパス・姫路バイパスや JR 山陽本線を利用して姫路市内の中心市街地や，姫路市内の大型ショッピングセンターへのアクセスが容易であること，ならびに市境界付近に立地する大型スーパーセンター等の存在が大きく寄与しているものと推測される．

高砂市の商業実態調査データを用いたその他の分析事例について

紙幅の都合上本稿での詳細な算出結果の提示は割愛するが，川向（2026）では高砂市内の商業環境満足度を被説明変数とし，「性別」「年齢層」「日用品購入市町」「買回り品購入市町」「職業」「個人商店利用有無」「居住小学校区」の計 7 変数を説明変数とした順序ロジスティック回帰分析を実施した結果を示している．
なお，オッズ比に基づく検討からは，以下の知見が得られている：

地域差： 基準とした高砂小学校区以外のすべての校区において，不満傾向の相対的な緩和（オッズ比の低下）が確認される．
世代間ギャップ： 若年・子育て世代ほど現状の商業環境に対して厳しい評価（不満傾向）を示し，高齢層ほど相対的に満足度が高い．
購買行動の影響： 買回り品および生鮮品・日用品の購入先を主要周辺都市群（加古川市・姫路市・神戸市）とする回答者群において，不満度の抑制・緩和傾向が統計的に有意に認められた．

まとめに代えて

本稿では，高砂市商業実態調査の実データを事例として，名義尺度や順序尺度を多数含む社会調査データに対する JMP を用いた多重対応分析およびモザイク図等の併用による調査結果についての構造的可視化手法の有効性を示した．
現在，多くの自治体や民間企業において多様なアンケート調査が実施され，政策立案や経営意思決定の基礎資料として活用されている．しかしながら，取得されたデータの尺度的特性を無視した不適切な解析が実施されている例にも，時折遭遇する．また，二次元のクロス集計分析のみによる解析結果の利用にとどまることで，多変数間に存在する潜在的な相互関連性が見過ごされている事例は少なくないのではないかと考える．
多重対応分析の体系的な適用およびその解釈手法が社会的に広く普及することは，調査対象が抱える多次元的な構造を極めて効果的かつ包括的に把握するうえで極めて重要であるといえるのではないだろうか．
我が国における少子高齢化の進行に伴い，自治体行政の税収減，企業経営における売上数量や利益の減少など，我が国における多くの組織を取り巻く環境は一層厳しさを増している．

そのような現代において，社会調査の手法によって得られた実証データに対し，多重対応分析などを用いることで，客観的かつ多面的な実態把握は極めて有効であろう．さらに，データの共有に加え，各種の分析結果の共有が，より実効性の高い政策形成ならびにエビデンスに基づく事業推進に資することを期待したい．

参考文献

川向肇（2026），地方都市における商業環境の変化と住民の商業行動，高砂市商業実態調査の結果から，GIS 学会第 35 学術研究発表大会 予稿集(forthcoming)

川向肇・中根滉大・神棒美波（2026）．兵庫県立大学における JMP の利用 ～高砂市商業調査などを事例として SAS ユーザー総会 2026 論文集(forthcoming)

大毛佑太・川向肇(2025)，兵庫県高砂市の商業調査の分析結果について 『ディスカバリーサミットジャパン - 配布資料』，https://community.jmp.com/kvoqx44227/attachments/kvoqx44227/abstracts/2284/5/P8_Yuta-DAIMOM_Poster.pdf?profile.language=ja（最終アクセス日 2026 年 9 月 10 日）

高砂市(2026)，高砂市商業実態調査 消費者・商業者サービス事業者アンケート調査回答【商工労働課】，『姫路市・播磨圏域連携中枢都市圏オープンデータカタログサイト』, https://city.himeji.gkan.jp/gkan/dataset/commerce（最終アクセス日 2026 年 9 月 10 日）

高砂商工会議所（2026）．高砂市商業実態調査 最終報告書 (PDF ファイルは https://www.city.takasago.lg.jp/soshikikarasagasu/sangyoshinkoka/sangyoshinko/2/6124.html から入手可能

繆青・有馬昌宏・川向肇，JMP を活用した住民意識調査データに基づく行政課題の構造に関する計量的抽出法の検討，『SAS User Forum ユーザー学術総会 2005 論文集』pp.425-436.

生存時間解析における多重比較法の性能評価

○橋本 敏夫[1]、山田 雅之[2]、本田 主税[3]、福島 慎二[1]

([1]株式会社タクミインフォメーションテクノロジー、

[2]キッセイ薬品工業株式会社、[3]小野薬品工業株式会社)

Evaluation of the Performance of Multiple Comparison Methods in Survival Analysis

Toshio Hashimoto[1], Masayuki Yamada[2], Chikara Honda[3], Shinji Fukushima[1]

[1]Takumi Information Technology Inc., [2]KISSEI PHARMACEUTICAL CO., LTD.,
[3]ONO PHARMACEUTICAL CO., LTD

要旨

薬効薬理試験では、対照群、低用量群、中用量群および高用量群の 4 群を比較し、薬理作用を評価することが一般的である。薬理作用の評価指標としてイベント発生までの時間を用いる場合には、log-rank 検定や一般化 Wilcoxon 検定が広く利用されているが、これらの検定に対する多重比較法の性能については十分な検討がなされていない。本研究では、これらの検定とその多重比較法（Dunnett 法、Holm 法、固定順序法）及び解析手順（4 群を同時に解析する joint analysis、対照群と各用量群を個別に解析する separate analysis）を対象としてシミュレーション評価を行い、帰無仮説下における実質有意水準および対立仮説下における検出力を評価した。

帰無仮説条件下では、log-rank 検定における Holm 法と固定順序法の separate analysis において、少数例かつイベント発生割合が高い条件で実質有意水準のインフレーションが認められた。一方、log-rank 検定の joint analysis 及び一般化 Wilcoxon 検定では、いずれの条件においても実質有意水準は概ね 5%に維持された。Dunnett 法は、少数例かつイベント発生割合が高い条件で非収束が認められた。

対立仮説条件下では、薬理試験で一般的に実施される単調な用量反応関係が仮定できる条件（直線的に変化、中用量飽和）を評価した。対照群と各用量群の差の検出力は、固定順序法が Holm 法に比し同程度又は高い検出力を示した。一方、Dunnett 法は固定順序法、Holm 法よりも検出力が低かった。

また、joint analysis は、log-rank 検定の少数例かつイベント発生割合が高い条件を除き、separate analysis と同程度またはそれ以上の検出力を有していた。

本研究の結果から、生存時間解析における薬効評価で、単調な用量反応関係が仮定できる場合には固定順序法が有用で、joint analysis の利用が望ましいことを示した。

キーワード：
生存時間解析、log-rank 検定、一般化 Wilcoxon 検定、joint analysis、固定順序法、Holm 法

1. はじめに

薬理試験（および臨床研究）では、イベント発生までの時間を評価指標とする生存時間解析が広く利用されている。薬効薬理試験における用量反応試験では、対照群と複数の投与群を比較するため、多重比較法の適切な選択が必要となる。 生存時間解析の検定としては log-rank 検定が広く利用されているが、イベントの多くが観察期間前半に発生するため後半での差が小さくなる場合には一般化 Wilcoxon 検定が利用される。

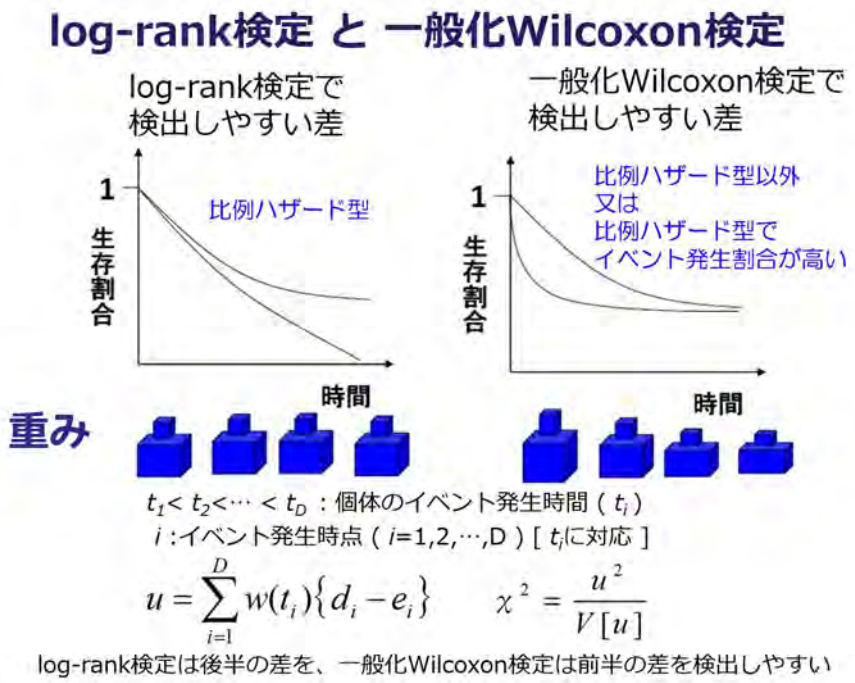
多重比較法としては Dunnett 法、Holm 法、固定順序法などが利用可能であるが、生存時間解析においてそれらの性能を比較した報告は少ない。また、ノンパラメトリックな多重比較法においては、4 群全体で解析する joint ranking による多重比較と、比較群ごとと 2 群比較を行う separate ranking による多重比較がある。対照群との比較においては、separate ranking による Steel の多重比較法 [1]にのみ明確な出典が存在する。一方、生存時間解析においては、4 群全体で解析する（joint analysis）と比較する 2 群ごとに解析する（separate analysis）の性能比較を行った研究 [2]は存在するが、論文化はされていない。

そこで本研究では、生存時間解析における各種検定法、多重比較法、及び比較する単位（joint/separate）について、コンピュータシミュレーションの方法を用いて、実質有意水準と検出力を評価し、薬理試験（および臨床研究）における適切な統計解析方法を提案することを目的とした。

2. 方法

2.1. 生存時間解析の検定手法

生存時間解析の検定方法として、log-rank 検定と一般化 Wilcoxon 検定を評価することとした。いずれの検定も、イベント発生時点ごとに、群とイベント有無に関する分割表を作成し、観測値と期待値の差の重み付け和を求め、その分散を用いて χ^2 統計量を構成する。log-rank 検定は各時点の重みを 1、一般化 Wilcoxon 検定はイベント発生時点の総例数を重みとしている。そのため、log-rank 検定は後期イベントの差に感度が高く、一般化 Wilcoxon 検定は前期イベントの差に感度が高い検定として知られている。検定の有意水準は両側 5%に設定した。



2.2. 生存時間解析の多重比較法

多重比較法として、対照群と低用量群、中用量群、及び高用量群の 3 回の検定の多重性を調整する方法として、Dunnett-Hsu の多重比較[3)]（以下 Dunnett 法と略す）、simulation 法、Holm の多重比較法[4)]（以下 Holm 法と略す）、高用量から順に検定を行う固定順序法（以下 固定順序法と略す）を評価することとした。以下については、log-rank 検定と一般化 Wilcoxon 検定に共通した記述である。

Dunnett 法は、PROC LIFETEST で adjust=dunnett と指定することで実施できる。Simulation 法は、Dunnett 法が収束しない場合の代替法であり、adjust=simulate を指定して実行する。Holm 法は、3 つの p 値を小さい順にならべ、最も小さな p 値を α/3、次に小さな p 値を α/2、最も大きな p 値を α で検定する方法で、検定が有意でなくなった時点で検定を終了する方法である。固定順序法は、高用量群→中用量群→低用量群と検定の順番を固定して有意水準 α=0.05 の検定を実施して、有意でなくなった時点で検定を終了する方法である。Holm 法および固定順序法は PROC LIFETEST で直接実行できないため、4 群全体の解析（joint analysis）または 2 群ごとの解析（separate analysis）で得た各対比較の p 値を用い、SAS マクロで検定を実施した。何れの検定も両側有意水準を 5%とした。

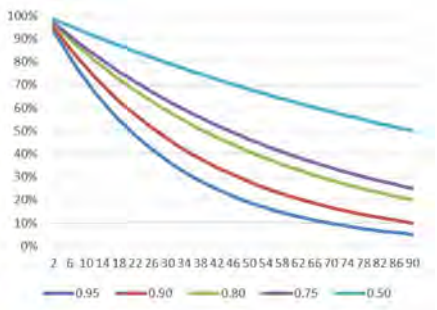
2.3. シミュレーションの方法

シミュレーションでは約 3 か月（90 日間）の試験で、対照群、低用量群、中用量群、高用量群の 4 群比較を行う試験を想定した。1 群の例数は n= 8, 12, 16, 20, 30, 40, 50、シミュレーション回数を 10 万回とした。試験期間中のハザードを一定と仮定して、90 日後のイベント発生割合（95%～25%）を元に各条件でのハザードを$\lambda = \frac{-log(1-S(t))}{t}$ により群ごとに計算し、 各個体の生存時間は、群ごとに定めた λ を母数とする指数分布に互いに独立に従うものとして発生させた（平均は 1/λ）。なお、SAS の RAND 関数は指数分布を尺度母数（平均）で指定するため、実装上は RAND("EXPONENTIAL", 1/λ)と記述した。観察期間を 90 日間とし、生存時間が 90 日を超えたデータは、90 日での打切りデータとした。生存時間が 90 日以下のデータについては、生存時間とは独立な一様乱数により一定割合（0%, 20%）で打切りデータとした。

2.3.1. 帰無仮説条件下のシミュレーション

帰無仮説条件下のイベント発生割合とハザードの関係を以下に示す。帰無仮説条件下の実質有意水準は、「3.1. 帰無仮説条件下のシミュレーション結果」に、図の上段に対照群と何れかの用量群で有意差がついた割合を、図の下段に対照群と高用量群で有意差がついた割合として示した。

	観察期間（日）と イベント非発生割合				イベント発生割合
λ	10	30	60	90	Day90
0.0333	71.7%	36.8%	13.6%	5.0%	95%
0.0256	77.4%	46.4%	21.5%	10.0%	90%
0.0179	83.6%	58.5%	34.2%	20.0%	80%
0.0154	85.7%	63.0%	39.7%	25.0%	75%
0.0077	92.6%	79.4%	63.0%	50.0%	50%



2.3.2. 対立仮説条件下のシミュレーション

対立仮説条件では薬効により用量反応依存的にイベント発生割合が低下するケース（直線的に変化）と、薬効が中用量で頭打ちとなるケース（中用量飽和）についてシミュレーションを行った。対照群に対して、低用量群、中用量群、高用量群が有意となった割合をそれぞれの用量群の検出力とした（Dunnett の非収束数

は検出力計算の分母に含めた)。

【90 日後のイベント発生割合が直線的に変化する条件】



【90 日後のイベント発生割合が中用量で頭打ちとなる条件】



3. 結果

3.1. 帰無仮説条件下のシミュレーション結果

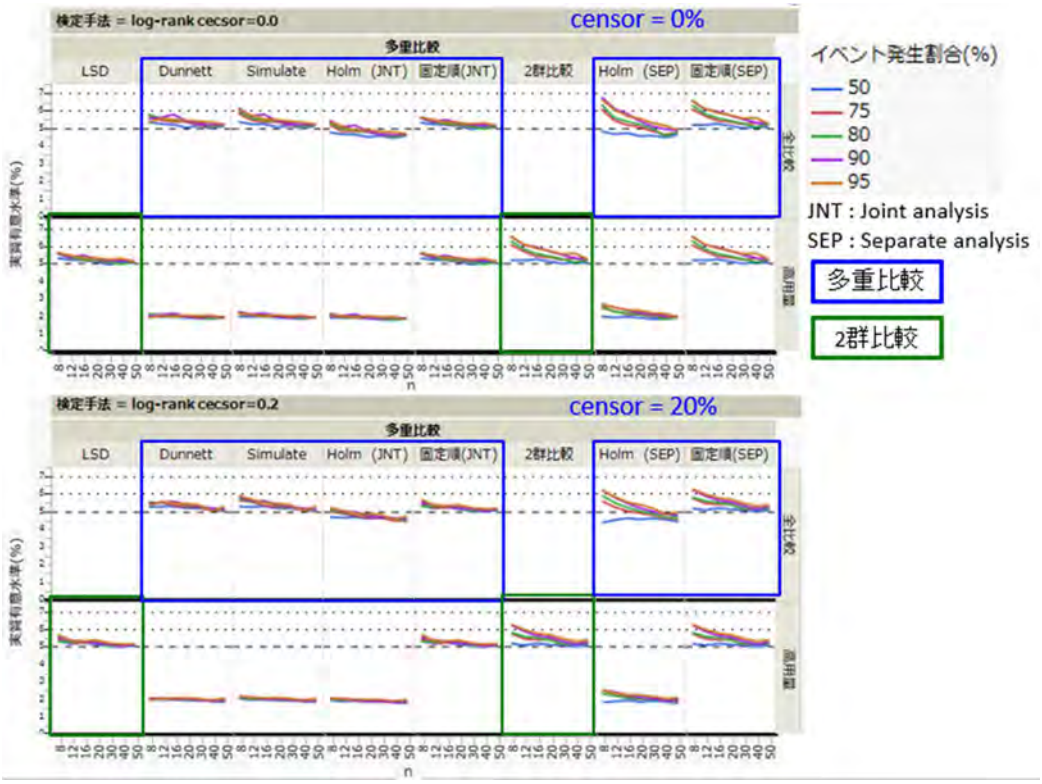
3.1.1. log-rank 検定

帰無仮説条件での log-rank 検定のシミュレーション結果を以下に示す。図の上段が、打切り割合 0%、下段が打切り割合 20%のシミュレーション結果である。

上側の青枠が多重比較法である Dunnett 法、simulation 法、Holm 法（JNT）及び固定順序法(JNT)の実質有意水準は、概ね名目有意水準 5%を維持していたが、Holm 法（SEP）と、固定順序法（SEP）の実質有意水準は、例数が少ない程、イベント発生割合が高い程、名目有意水準（5%）を超えていた（以下は実質有意水準のインフレーションと略す)。

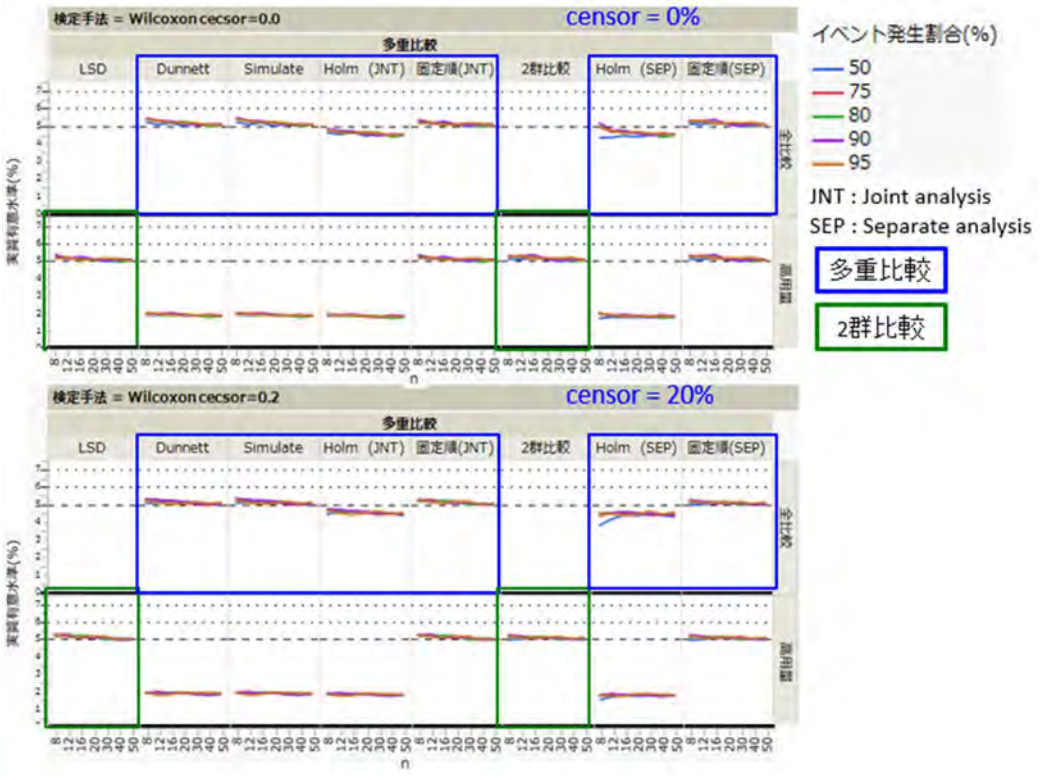
下側の緑枠が多重性を調整しない場合である。実質有意水準が 5%を超えた原因が多重比較に起因するかを確認するために追加したシミュレーションである。Separate analysis による実質有意水準のインフレーション

は、多重性を調整しない方法で発生していた。打切り割合 20%も同様な結果であった。



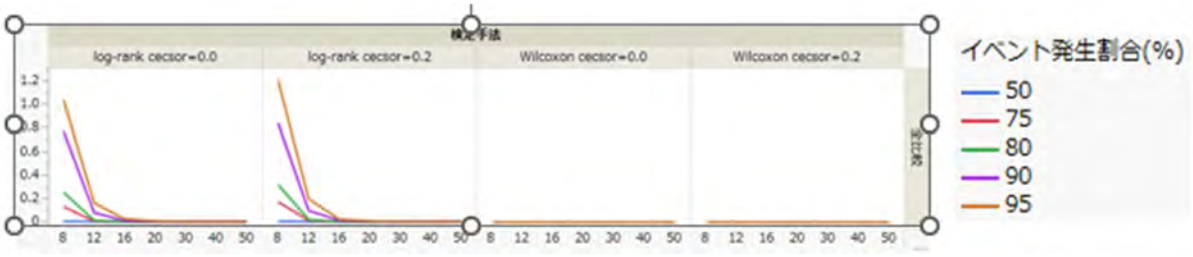
3.1.2. 一般化 Wilcoxon 検定

帰無仮説条件での一般化 Wilcoxon 検定のシミュレーション結果を以下に示す。一般化 Wilcoxon 検定では、各種多重比較法の実質有意水準が概ね 5%に維持されていた。



3.1.3. Dunnett 法の非収束割合

log-rank 検定の Dunnett 多重比較は、少数例かつイベント発生割合が高い条件で非収束の割合が増加した。一般化 Wilcoxon 検定では非収束はほとんど認められなかった。打切り割合 20%も同様の結果であった。

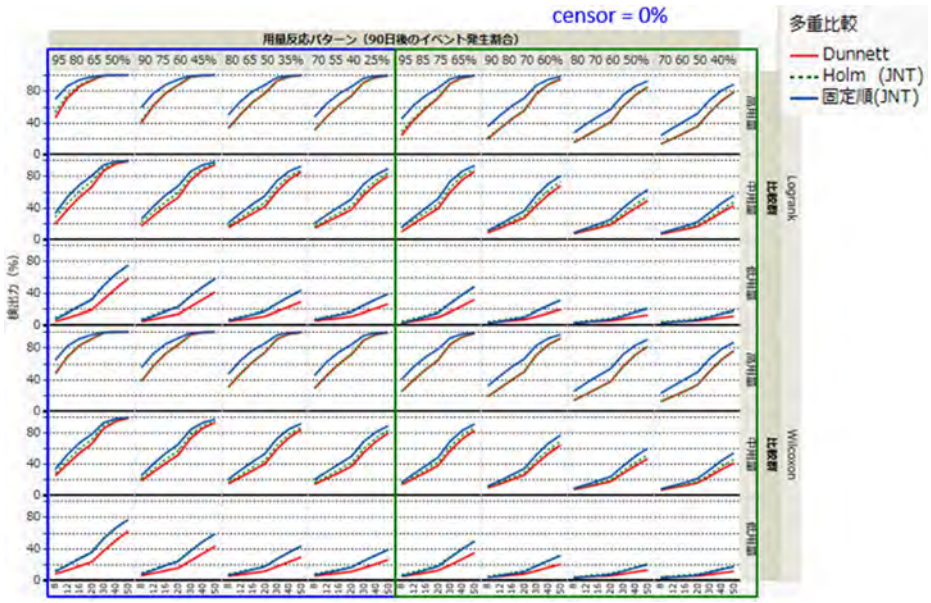


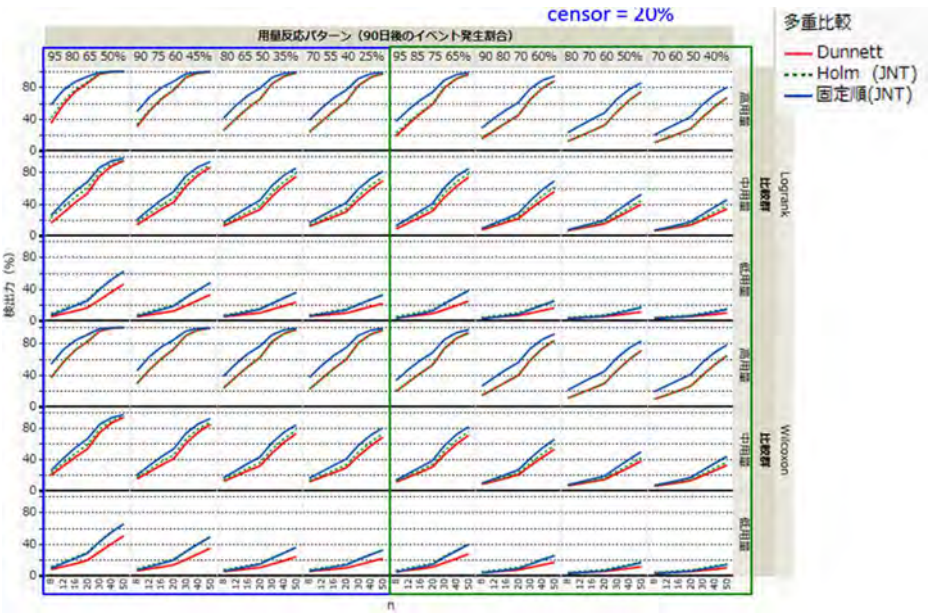
3.2. 対立仮説条件下のシミュレーション結果

3.2.1. イベント発生割合が直線的に変化する場合

3.2.1.1. joint analysis の多重比較法間の検出力比較

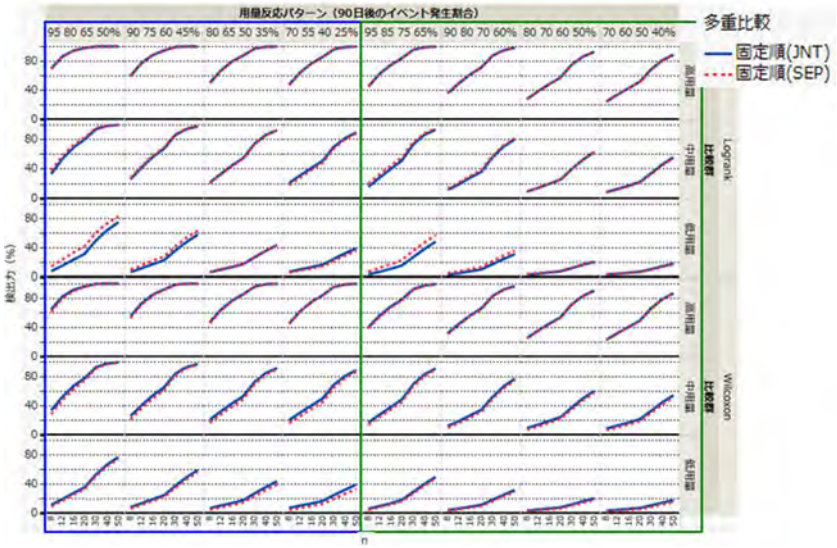
検出力は固定順序法、Holm 法、Dunnett 法の順であった（Holm 法の検出力は高用量で Dunnett と重なり、低用量では固定順序法と重なっている）。





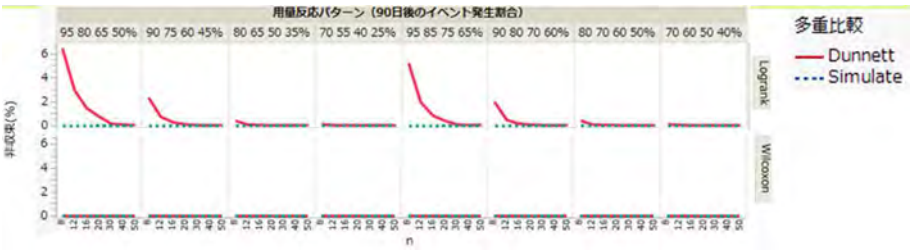
3.2.1.2. joint analysis と separate analysis の検出力比較

固定順序法について、joint analysis と separate analysis の検出力を比較した。log-rank 検定の対照群のイベント発生割合が高い条件（95%,90%）では、低用量、中用量の separate の検出力が joint よりも高かったが、対照群のイベント発生割合が高くない条件（70%）では joint の検出力が高かった（打切り割合 0%）。一般化 Wilcoxon 検定では、joint の検出力が separate よりも高かった。打切り割合 20%も同様の結果であった。



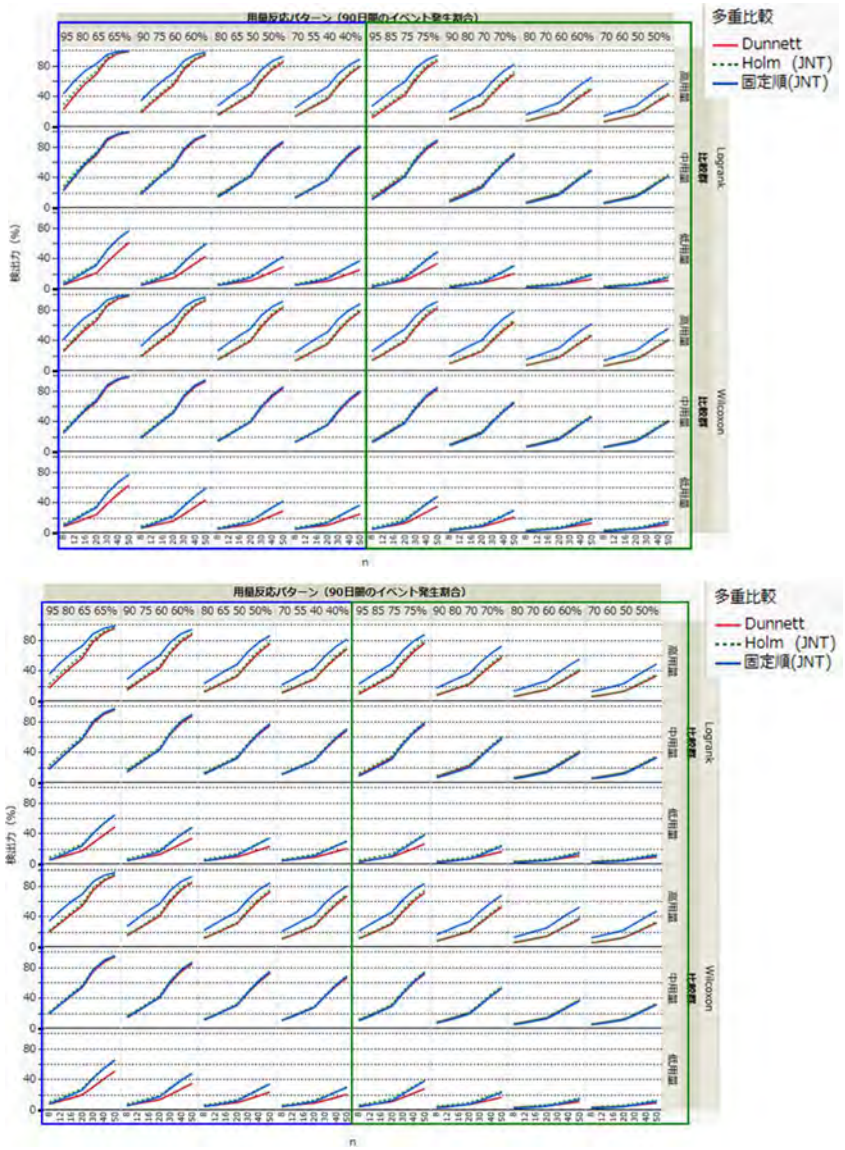
3.2.1.3. Dunnett 法の非収束割合

Dunnett 法は少数例かつ対照群のイベント発生割合が高い条件で、非収束が認められたが、Simulation 法で回避された（打切り割合 0%）。打切り割合 20%も同様な結果であった。



3.2.2. イベント発生割合が中用量で飽和する場合

検出力は、高用量 : 固定順序法＞Holm 法≒Dunnett 法、中用量 : Holm ≒固定順序法≒Dunnett、低用量 : Holm ≒固定順序法＞Dunnett であった。下段の打切り割合 20%は、上段の打切り割合 0%よりやや検出力が低下したが、多重比較法間の相対関係は同様であった。



4. 考察

本研究では、生存時間解析における複数の多重比較法の性能を比較した。

帰無仮説条件下では、log-rank 検定の separate analysis で実質有意水準のインフレーションが認められた。log-rank 検定の実質有意水準のインフレーションについて報告した先行研究の概要を以下に示す。本研究の結果は、小標本条件における log-rank 検定の非保守性を指摘した Kellerer ら [5]、Heinze ら [6]、Wang ら [7]、及び水間ら [8]の報告と整合した。

Kellerer, A. M., Chmelevsky, D. (1983)

log-rank 検定は、小標本（特に例数がアンバランスな試験）で非保守的になり得ることを示した。

Heinze, G., Gnant, M., Schemper, M. (2003)

標本の不均衡があると漸近（χ^2）版の log-rank 検定は名目水準を保てないことを示した。
“exact logrank”を提案し、 実質有意水準が制御されることを示した。

Wang, Z., Zhang, A., Chen, Y., et al. (2019)

log-rank 検定は、小標本や例数がアンバランスな試験で、第I種の過誤を過大評価する場合がある。原因は検定統計量の分布が標準正規分布から乖離するためである。正規近似の代わりに approximate permutation log-rank test を提案し、実質有意水準の制御が改善することを示した。

水間 浩太郎，杉本 知之(2023).

log-rank 検定は、小標本や群サイズの偏りが大きい状況では χ^2近似の精度が悪化するという指摘を整理し、計算代数統計（分割表のマルコフ基底）を用いた MCMC による正確検定を提案し、シミュレーションで有意水準が適切に保持されることを示した。

Joint analysis と separate analysis の検出力を比較した結果、log-rank 検定の低用量群で、対照群のイベント発生割合が高い条件（95%および 90%）で separate analysis の検出力が joint analysis を上回った。しかし、これらの条件は帰無仮説下で実質有意水準のインフレーションが認められていることから、その結果は χ^2近似の精度低下の影響によるものと考えられた。したがって、イベント発生割合が高く、かつ各群の例数が限られる状況においては、separate analysis の適用および結果の解釈には慎重な対応が必要と考えられる。

Dunnett 法の非収束については、log-rank 統計量間の相関行列が特異に近い状態となったことが一因と考えられた。一方、一般化 Wilcoxon 検定では、早期イベントに大きな重みが付与されることにより有効情報量が増加し、相関行列の推定が安定化した可能性がある。

以上の結果から、単調な用量反応関係が仮定できる条件（直線的に変化、中用量飽和）においては、固定順序法が有用であると考えられた。薬理試験では、一般に用量依存的な薬効が期待される範囲で用量設定が行われ、高用量群または中用量群以上における有効性の検証が主要な目的となる場合が多い。そのため、本研究で想定した条件は実際の試験においても適用される場合が多いと考えられる。一方、我々にはほとんど経験がなく、今回のシミュレーションでは評価していないが、高用量群における薬効の減弱があらかじめ想定され中用量群の有効性検証を主目的とする試験を計画することがあれば、Holm 法も選択肢となり得るかもしれない。

また、Dunnett 法については、一部の条件下で収束しない事例が認められたことに加え、本研究では固定順序法および Holm 法と比較して検出力が低かった。したがって、少なくとも本研究で検討した条件下においては、薬理試験等への適用上の利点は限定的であると考えられた。

5. まとめ

生存時間解析における用量反応試験を対象としたシミュレーション研究を行った。その結果、薬理試験等の少数例の試験における統計解析の選択方針を以下のように提案する。

1. 単調な用量反応関係（中用量飽和を含む）が想定される場合は、多重比較法として固定順序法を推奨する。
2. 高用量群における薬効の減弱があらかじめ想定され、中用量群の有効性検証を主目的とする場合には、Holm 法が選択肢となるかもしれない。
3. 固定順序法および Holm 法を用いる場合には、joint analysis が望ましい。
4. Dunnett 法は一部の条件で非収束が発生し、固定順序法、Holm 法と比較して検出力が低い。
5. 生存曲線の早期から群間差が認められる場合や、少数例かつイベント発生割合が高い場合には、一般化 Wilcoxon 検定が有用である。

本研究では固定順序法を推奨したが、現時点では SAS の LIFETEST PROCEDURE において固定順序法を直接実行することはできない。一方、非臨床試験向け統計解析パッケージ **Stat Preclinica version 4.2**（㈱タクミインフォメーションテクノロジー） は、多重比較法として固定順序法を実装しており、joint analysis と separate analysis の双方に対応している。

なお、Wang らで検討された permutation log-rank test は R の coin パッケージに実装[9]されているようである。今後、SAS ユーザー総会等での議論を通じて、その必要性が判断されれば、permutation log-rank test についても SAS/STAT に実装されることを期待する。

6. 謝辞

本研究にあたり、田辺ファーマ株式会社のデータサイエンス部統計解析グループの皆様より、貴重なご助言とご支援を賜りました。深く感謝申し上げます。

7. 参考文献

1. Steel, R. G. D. (1959). A multiple comparison rank sum test: Treatments versus control. Biometrics, 15(4), 560-572.
2. 島村 文也, 浜田 知久馬, 佐野 雅隆. (2014). 生存時間解析におけるノンパラメトリック検定の多重比較に関する研究. *SAS ユーザー総会講演論文集*.
3. Hsu, J. C. (1992). The factor analytic approach to simultaneous confidence intervals for multiple comparisons with the best. *Journal of Computational and Graphical Statistics*, 1, 151-168.
4. Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70.
5. Kellerer, A. M., Chmelevsky, D. (1983). Small-sample properties of censored-data rank tests. *Biometrics*, 39, 675-682.
6. Heinze, G., Gnant, M., Schemper, M. (2003). Exact log-rank tests for unequal follow-up. *Biometrics*, 59, 1151-1157.
7. Wang, Z., Zhang, A., Chen, Y., et al. (2019). Type I error inflation of log-rank test with small sample size: A permutation approach and simulation studies. *Journal of Statistical Research*, 53(2), 93-109.
8. 水間 浩太郎, 杉本 知之. (2023). 計算代数統計を用いたログランク検定. 日本統計学会誌, 52(2), 355-371
9. Hothorn, T., Hornik, K., van de Wiel, M. A., Zeileis, A. (2008). Implementing a Class of Permutation Tests: The coin Package. *Journal of Statistical Software*, 28(8), 1-23.

SAS によるイベント駆動型臨床試験における中間解析時点予測

○堀本　隆一
（株式会社エス・シー・エー）

Prediction of Interim Analysis Timing in Event-Driven Clinical Trials Using SAS

Ryuichi Horimoto
System Communication Assist Co., Ltd.

要旨

イベント駆動型臨床試験では、所定イベント数への到達時点が中間解析および最終解析の実施時期を規定する。試験開始前に想定した登録率、イベント発生率および打ち切り率は実測値と乖離し得るため、進行中のデータに基づいて予測到達時点を更新することが試験運営上重要となる。

本研究では、指数分布モデル、ワイブル分布モデル、ノンパラメトリックモデルおよび Flexible parametric model（FPM）の 4 手法を SAS で実装した。シミュレーションでは、目標 200 イベントへの到達時点を予測した。症例数 300 例、登録期間 24 か月の仮想試験を生成し、真のイベント時間分布を指数分布、ワイブル分布および区分指数分布とした。予測は 6、12、18 か月時点で実施し、外側反復として仮想試験を 200 回生成し、Bias、MAE（平均絶対誤差）、95%予測区間の被覆割合（Coverage）および平均予測区間幅を評価した。なお、各仮想試験では、内側反復として将来経過を 1,000 回生成し、得られた目標イベント数到達時点の分布の中央値を点予測、2.5 パーセンタイルおよび 97.5 パーセンタイルを 95%予測区間とした。

指数分布シナリオでは指数分布モデル、ワイブル分布シナリオではワイブル分布モデルの MAE が各予測時点で最小であった。区分指数分布シナリオでは 6 か月時点の MAE はワイブル分布モデルが最小であり、柔軟なモデルが一貫して最良とはならなかった。全般的に予測時点が遅くなるほど点予測精度は改善したが、Coverage は真の分布、モデルおよび予測時点に依存した。以上から、本研究の設定では、柔軟なモデルの採用のみで予測性能が向上するとは限らず、イベント成熟度や外挿仮定を踏まえて、複数の手法の結果を比較する必要があると考えられた。

キーワード：イベント駆動型臨床試験、中間解析、予測、生存時間解析、SAS

1. はじめに

イベント駆動型臨床試験では、暦上の固定日ではなく累積イベント数を基準として中間解析や最終解析を実施する。したがって、目標イベント数に到達する時期の見込みは、データカット、データクリーニング、解析、データモニタリング委員会の開催および試験資源の配分に直接関係する。試験計画時には登録速度、イベント発生分布および追跡不能・中止による打ち切りを仮定して試験期間を見積もる。しかし、実際の登録率とイベント率が計画時の仮定から乖離すれば、当初想定した解析時期も変化する。Heitjan らは、進行中

の試験データを用いて登録数やイベント数の予測を逐次更新する方法を整理し、試験運営におけるリアルタイム予測の意義を示している[1]。

目標イベント数到達時点の予測では、予測時点までのイベント・打ち切り済み患者に加え、追跡中患者と未登録患者の将来を扱う必要がある。先行研究では、登録、イベント発生および打ち切りを組み合わせ、将来の試験経過を条件付きでシミュレーションする枠組みが提案されてきた[2-5]。パラメトリック法は外挿が明確である一方、分布仮定の誤りが予測に影響する。ノンパラメトリック法は観察範囲内の分布仮定を弱められるが、観察範囲外には追加の仮定が必要となる。柔軟なパラメトリックモデルは複雑なハザード形状を表現できるが、試験初期の少数イベントでは推定が不安定となる可能性がある。

本研究の目的は、進行中のイベント駆動型試験における目標イベント数への到達時点を指数分布モデル、ワイブル分布モデル、ノンパラメトリックモデルおよび Flexible parametric model（FPM）の 4 手法で予測する SAS 実装を示し、シミュレーションにより性能を比較することである。主要評価指標は Bias と MAE とし、95%予測区間の Coverage と平均幅を補足的に評価した。

2. 方法

2.1. 予測の枠組み

予測時点において、各患者をイベント発生済み、打ち切り済み、追跡中、未登録の 4 状態に分類した。イベント発生済み、打ち切り済みおよび追跡中の登録済み患者のデータから、指数分布モデル、ワイブル分布モデル、ノンパラメトリックモデルおよび Flexible parametric model（FPM）の 4 手法によって、イベント時間分布を推定した。さらに、その分布から不確実性を反映したイベント時間分布を生成し、その分布に基づいて将来経過を生成した。

将来経過の生成では、イベント発生済み患者は既に観察された暦時間上のイベント時点を固定し、打ち切り済み患者は以後イベントは発生しないものとした。追跡中患者は、予測時点までイベントが発生していないことを条件とするイベント時間分布から将来イベント時間を生成した。未登録患者は、推定したイベント時間分布から登録後のイベント時間を生成した。また、追跡中および未登録患者は将来の打ち切り時間も生成し、それぞれを暦時間に変換した上で、将来のイベント時点が将来の打ち切り時点以前であった場合のみイベントとして計上した。なお、登録時間は登録期間内の一様分布、打ち切り時間は一定ハザードを仮定した指数分布に従うものとした。最後に、生成された暦時間上のイベント時点を昇順に並べ、D 番目のイベント発生時点を目標イベント数到達時点とした。

この不確実性を反映したイベント時間分布の生成から将来経過の生成、目標イベント数到達時点の抽出までを B 回反復し、得られた到達時点の分布の中央値を点予測、2.5 パーセンタイルおよび 97.5 パーセンタイルを 95%予測区間とした。

2.2. SAS による将来イベント時間予測

2.2.1. 指数分布モデル

指数分布モデルでは、イベントハザードが時間によらず一定であると仮定する。予測時点までに登録済みの患者データ pred_fit に対して、PROC LIFEREG の DIST=EXPONENTIAL を適用し、切片μを推定した。指数分布のハザードは$\lambda = \exp(-\mu)$として求められる。また、切片μの標準誤差から$\mathrm{Var}(\hat{\mu}) = SE(\hat{\mu})^2$を求めた。

```
ods output ParameterEstimates=_exp_pe;
proc lifereg data=pred_fit;
  model followup*event_pred(0)= / dist=exponential;
run;

data _null_;
  set _exp_pe;
  if upcase(Parameter)='INTERCEPT' or Parameter='切片' then do;
    call symputx('exp_mu_hat', Estimate, 'g');
    call symputx('exp_var_mu', StdErr**2, 'g');
    call symputx('lambda_hat', exp(-Estimate), 'g');
  end;
run;
```

followup　：予測時点で観察されている登録後追跡時間

event_pred　：予測時点でイベント発生済みであれば 1、打ち切り済みまたは追跡中であれば 0

パラメータ推定の不確実性を将来予測に反映するため、各反復$b = 1, \ldots, B$において、切片μを

$$\mu^{(b)} = \hat{\mu} + \sqrt{\widehat{\mathrm{Var}}(\hat{\mu})}\, Z, \qquad Z \sim N(0,1)$$

から生成し、$\lambda^{(b)} = \exp\{-\mu^{(b)}\}$とした。

```
data _exp_draws;
  call streaminit([seed value]);
  do b=1 to &B;
    mu_draw=&exp_mu_hat + sqrt(&exp_var_mu)*rand('normal');
    lambda_draw=exp(-mu_draw);
    output;
  end;
run;
```

追跡中患者については、予測時点までイベントが発生していないことを考慮し、予測時点までの登録後追跡時間をt_0として条件付き将来イベント時間を生成した。指数分布の無記憶性により、一様乱数Uを用いて、

$$T = t_0 - \frac{\log U}{\lambda^{(b)}}$$

とした。一方、未登録患者では$t_0 = 0$とし、登録後の将来イベント時間を同じ式から生成した。

```
if status_pred='ONGOING' then t0=followup;
else t0=0;
u=rand('uniform');
event_since_entry=t0 - log(u)/lambda_draw;
```

追跡中患者では、生成した将来イベント時間を暦時間へ変換し、将来の打ち切り時点以前にイベントが発生した場合のみイベントとして採用した。未登録患者では、予測時点から登録終了時点までの一様分布から将来登録時点を生成し、登録後のイベント時点が打ち切り時点以前であった場合に、登録時点と将来イベント時間の和を暦上のイベント時点とした。

2.2.2. ワイブル分布モデル

ワイブル分布モデルでは、指数分布モデルと同様に、予測時点までに登録済みの患者データ pred_fit から、PROC LIFEREG によりイベント時間分布を推定した。ワイブル分布の生存関数は

$$S(t) = \exp\left\{-\left(\frac{t}{a}\right)^{k}\right\}$$

である。ここで、aは尺度パラメータ、kは形状パラメータである。$k > 1$の場合はハザードが時間とともに増加し、$k < 1$の場合は減少する。$k = 1$の場合は指数分布に一致する。

SAS 実装では PROC LIFEREG の DIST=WEIBULL を適用し、切片μと Scale パラメータσを推定した。尺度パラメータと形状パラメータは、

$$a = \exp(\mu), \qquad k = \frac{1}{\sigma}$$

として求められる。さらに、パラメータ推定の不確実性を将来予測に反映するため、COVB オプションにより(μ, σ)の分散共分散行列を取得した。

```
ods output ParameterEstimates=_wb_pe
           CovB=_wb_cov;
proc lifereg data=pred_fit;
  model followup*event_pred(0)= / dist=weibull covb;
run;

data _null_;
  set _wb_pe;
  if upcase(Parameter)='INTERCEPT' or Parameter='切片' then do;
      call symputx('wb_mu_hat',Estimate,'g');
  end;
  if upcase(Parameter)='SCALE' or Parameter='尺度' then do;
    call symputx('wb_sigma_hat',Estimate,'g');
  end;
```

```
run;
```

各反復$b = 1, \ldots, B$では、(μ, σ)の推定分散共分散行列をデルタ法により$(\mu, \log\sigma)$尺度へ変換し、

$$\begin{pmatrix} \mu^{(b)} \\ \log\sigma^{(b)} \end{pmatrix} \sim N_2\left[\begin{pmatrix} \hat{\mu} \\ \log\hat{\sigma} \end{pmatrix}, \begin{pmatrix} \widehat{\mathrm{Var}}(\hat{\mu}) & \dfrac{\widehat{\mathrm{Cov}}(\hat{\mu}, \hat{\sigma})}{\hat{\sigma}} \\ \dfrac{\widehat{\mathrm{Cov}}(\hat{\mu}, \hat{\sigma})}{\hat{\sigma}} & \dfrac{\widehat{\mathrm{Var}}(\hat{\sigma})}{\hat{\sigma}^2} \end{pmatrix} \right]$$

からパラメータを同時生成した。各反復におけるワイブル分布の尺度および形状パラメータは、

$$a^{(b)} = \exp\{\mu^{(b)}\}, \qquad k^{(b)} = \exp\{-\log\sigma^{(b)}\} = \frac{1}{\sigma^{(b)}}$$

とした。

```
proc iml;
  call randseed([seed value]);
  mean={&wb_mu_hat,%sysfunc(log(&wb_sigma_hat))};
  cov=j(2,2,0);
  cov[1,1]=&wb_v11;
  cov[1,2]=&wb_v12/&wb_sigma_hat;
  cov[2,1]=cov[1,2];
  cov[2,2]=&wb_v22/(&wb_sigma_hat##2);
  draw=randnormal(&B,mean,cov);
  b=t(1:&B);
  wb_scale_draw=exp(draw[,1]);
  wb_shape_draw=exp(-draw[,2]);
  create _wb_draws var {b wb_shape_draw wb_scale_draw};
  append;
  close _wb_draws;
quit;
```

追跡中患者については、予測時点までの登録後追跡時間をt_0として条件付き将来イベント時間を生成した。ワイブル分布では、一様乱数Uを用いて、

$$T = a^{(b)}\left[\left(\frac{t_0}{a^{(b)}}\right)^{k^{(b)}} - \log U\right]^{1/k^{(b)}}$$

とした。この式により、$T > t_0$という条件の下での将来イベント時間を生成できる。一方、未登録患者では$t_0 = 0$であるため、上式は推定したワイブル分布からの将来イベント時間の生成に一致する。

```
if status_pred='ONGOING' then t0=followup;
else t0=0;
```

```
u=rand('uniform');
event_since_entry=wb_scale_draw*
  ((t0/wb_scale_draw)**wb_shape_draw - log(u))**(1/wb_shape_draw);
```

追跡中患者および未登録患者の暦時間上の将来イベント時点は、指数分布モデルと同様の方法で求めた。

2.2.3. ノンパラメトリックモデル

ノンパラメトリックモデルでは、登録済み患者のイベント時間分布を Kaplan-Meier 法に基づいて推定した。さらに、推定された生存関数の不確実性を将来予測に反映するため、各反復$b = 1, \ldots, B$において、登録済み患者iに独立な指数乱数$w_i^{(b)} \sim \mathrm{Exp}(1)$を付与し、Bayesian bootstrap に基づく重み付き Kaplan-Meier 生存関数を生成した。イベント時点t_jにおける重み付きリスク集合を$R_j^{(b)}$、重み付きイベント数を$D_j^{(b)}$とすると、生存関数は

$$\hat{S}^{(b)}(t) = \prod_{t_j \le t} \left(1 - \frac{D_j^{(b)}}{R_j^{(b)}}\right)$$

として求められる。

```
data _np_bb_long;
  call streaminit([seed value]);
  do b=1 to &B;
    do _i=1 to &n_registered;
      set pred_fit point=_i;
      w=rand('exponential');
      event_w=w*event_pred;
      tail_pt=w*max(followup - &np_tail_start,0);
      tail_ev=w*event_pred*(followup > &np_tail_start);
      output;
    end;
  end;
  stop;
run;

proc summary data=_np_bb_long nway;
  class b followup;
  var w event_w;
  output out=_np_by_time sum=w_total d_total;
run;
```

```
proc sort data=_np_by_time;
  by b descending followup;
run;
data _np_risk_steps;
  set _np_by_time;
  by b descending followup;
  retain risk;
  if first.b then risk=0;
  risk + w_total;
  if d_total > 0 and risk > 0 then do;
    km_t=followup;
    factor=1 - d_total/risk;
    output;
  end;
run;

proc sort data=_np_risk_steps;
  by b km_t;
run;
data _np_steps;
  set _np_risk_steps;
  by b km_t;
  retain km_s j;
  if first.b then do; km_s=1; j=0; end;
  km_s=km_s*factor;
  j + 1;
  output;
run;
```

&n_registered　：予測時点までに登録済みの患者数

&np_tail_start　：観察イベント時間の 75 パーセンタイル

Kaplan-Meier 曲線のみでは観察範囲を超えたイベント時間を生成できないため、観察イベント時間の 75 パーセンタイルを指数尾部ハザードの推定開始点τとし、それ以後のデータから指数尾部ハザードを推定し、最後のイベント時点以後は推定したハザードによる外挿を接続した。各反復bにおける尾部ハザードは、τ以後の重み付きイベント数を重み付き総観察時間で除した

$$\hat{\lambda}_{\text{tail}}^{(b)} = \frac{\sum_i w_i^{(b)} \delta_i I(T_i > \tau)}{\sum_i w_i^{(b)} \max(T_i - \tau, 0)}$$

により推定した。ここで、T_iは観察追跡時間、δ_iはイベント指示変数である。尾部ハザードを推定できない場合は、指数分布モデルから得たハザード推定値を代替値として用いた。以上より、最後のイベント時点をt_L、その時点の生存確率を$\hat{S}^{(b)}(t_L)$とすると、$t > t_L$に対する生存関数は

$$\hat{S}^{(b)}(t) = \hat{S}^{(b)}(t_L)\exp\left\{-\hat{\lambda}^{(b)}_{\text{tail}}(t - t_L)\right\}$$

となる。

```
proc summary data=_np_bb_long nway;
  class b;
  var tail_pt tail_ev;
  output out=_np_tail_b sum=tail_pt tail_ev;
run;

data _np_tail_b;
  set _np_tail_b;
  if tail_pt > 0 and tail_ev > 0 then lambda_tail=tail_ev/tail_pt;
  else lambda_tail=&lambda_hat;
run;
```

追跡中患者については、予測時点までの登録後追跡時間をt_0として条件付きイベント時間を生成した。推定生存関数を$\hat{S}^{(b)}(t)$とすると、

$$P(T > t \mid T > t_0) = \frac{\hat{S}^{(b)}(t)}{\hat{S}^{(b)}(t_0)}, \qquad t \geq t_0$$

である。したがって、一様乱数Uを用いて、$\hat{S}^{(b)}(T) = U\hat{S}^{(b)}(t_0)$を満たす$T$を生成した。Kaplan-Meier 曲線上では、目標生存確率$U\hat{S}^{(b)}(t_0)$以下となる最初のイベント時点を探索した。目標生存確率に対応するイベント時点が Kaplan-Meier 曲線の観察範囲を超える場合は、指数ハザードによる尾部外挿を用いて、

$$T = t_L - \frac{\log\left[U\hat{S}^{(b)}(t_0)/\hat{S}^{(b)}(t_L)\right]}{\hat{\lambda}^{(b)}_{\text{tail}}}$$

としてイベント時間を生成した。未登録患者については、$t_0 = 0$、すなわち$\hat{S}^{(b)}(0) = 1$として同じイベント時間分布から登録後イベント時間を生成した。

```
if status_pred='ONGOING' then t0=followup;
else t0=0;

if t0=0 then s0=1;
else if t0 <= last_t then do;
  s0=1; lo=1; hi=nstep;

  if nt[1] <= t0 then do;
```

```
    do while(hi - lo > 1);
      mid=floor((lo + hi)/2);
      if nt[mid] <= t0 then lo=mid; else hi=mid;
    end;
    if nt[hi] <= t0 then lo=hi;
    s0=ns[lo];
  end;
end;
else s0=s_last*exp(-lambda_tail*(t0 - last_t));

target=rand('uniform')*s0;

if nstep > 0 and target >= s_last then do;
  lo=1; hi=nstep;
  if ns[1] <= target then event_since_entry=nt[1];
  else do;
    do while(hi - lo > 1);
      mid=floor((lo + hi)/2);
      if ns[mid] > target then lo=mid; else hi=mid;
    end;
    event_since_entry=nt[hi];
  end;
end;
else event_since_entry=last_t - log(target/s_last)/lambda_tail;
```

nt ：観察イベント時点t_j

ns ：観察イベント時点t_jに対応する生存関数$\hat{S}(t_j)$

追跡中患者および未登録患者の暦時間上の将来イベント時点は、指数分布モデルおよびワイブル分布モデルと同様の方法で求めた。

2.2.4. Flexible Parametric Model（FPM）

Flexible parametric model（FPM）では、PROC ICPHREG の BASEHAZ=SPLINES(DF=[自由度])を用いて、対数累積ベースラインハザードを自然 3 次スプラインで表現した。登録後イベント時間をt、累積ハザードを$H(t)$とすると、本研究のモデルは概念的に$\log H(t) = s(\log t;\gamma)$と表される。ここで、$s(\cdot)$は自然 3 次スプライン関数、$\gamma$はスプライン係数ベクトルである。したがって、生存関数は

$$S(t) = \exp\{-H(t)\} = \exp[-\exp\{s(\log t;\gamma)\}]$$

として得られる。

イベント発生例では区間端点を観察イベント時間とし、打ち切り例では右端点を欠測とした。ODS OUTPUT によりスプライン係数、分散共分散行列、ノット情報および収束情報を取得した。

```
data _fpm_in;
  set pred_fit;
  left=followup;
  if event_pred=1 then right=left;
  else right=.;
  keep left right;
run;

ods output ParameterEstimates=_fpm_pe
           CovB=_fpm_cov
           HazardParms=_fpm_hp
           ConvergenceStatus=_fpm_conv;
proc icphreg data=_fpm_in;
  model(left,right) = / basehaz=splines(df=&FPM_DF) covb;
  baseline out=_fpm_base
           timelist=0.01 to &EVENT_HORIZON by &GRID_STEP
           survival=surv;
run;
```

&FPM_DF　　：自由度（3）

&EVENT_HORIZON　：生存曲線を計算する最大時間（120 か月）

&GRID_STEP　　：生存曲線を計算する時間間隔（0.1 か月刻み）

パラメータ推定の不確実性を将来予測に反映するため、各反復$b = 1, \ldots, B$において、スプライン係数ベクトルγを推定値とその分散共分散行列に基づく多変量正規分布$\gamma^{(b)} \sim N\left(\hat{\gamma}, \hat{\Sigma}_{\gamma}\right)$から生成した。

```
proc iml;
  call randseed([seed value]);
  use _fpm_theta;
  read all var {&fpm_coef_list} into mean;
  close _fpm_theta;
  use _fpm_cov;
  read all var {&fpm_coef_list} into cov;
  close _fpm_cov;
  draw=randnormal(&B,mean,cov);
quit;
```

&fpm_coef_list　：スプライン係数

生成した係数ベクトルを用いて各反復に対応する生存関数$S^{(b)}(t)$を再構成した。実装では、0.1 か月刻みで生存曲線を計算し、数値計算上の微小な増加を除くため単調非増加となるように補正した。

```
fs[1]=1;
prev=1;

do k=2 to &_fpm_ng;
  tt=(k-1)*&GRID_STEP;
  x=log(tt);

  eta=coef1_draw + coef2_draw*x;

  %do _q=2 %to %eval(&FPM_P-1);
    %let _ci=%eval(&_q+1);

    ej=(&&fpm_k&FPM_P - &&fpm_k&_q) /
       (&&fpm_k&FPM_P - &&fpm_k1);

    vv=max(0,x-&&fpm_k&_q)**3
       - ej*max(0,x-&&fpm_k1)**3
       - (1-ej)*max(0,x-&&fpm_k&FPM_P)**3;

    eta=eta + coef&_ci._draw*vv;
  %end;

  ss=exp(-exp(min(eta,700)));
  ss=min(prev,ss);

  fs[k]=ss;
  prev=ss;
end;
```

fs　　：再構成した生存関数$S(t)$

&_fpm_ng　：生存曲線を構成するグリッド点数（1201）

&fpm_p　　：スプライン係数・ノットの数（4）

&fpm_k[x]　：ノットの位置

追跡中患者については、予測時点までイベントが発生していないことを考慮し、条件付き将来イベント時間を生成した。予測時点までの登録後追跡時間をt_0とすると、

$$P(T > t \mid T > t_0) = \frac{S(t)}{S(t_0)}, \qquad t \geq t_0$$

となる。したがって、一様乱数Uを用いて、$S^{(b)}(T) = U\,S^{(b)}(t_0)$を満たす$T$を生成した。

FPM ではこの逆変換を閉形式で求めることが困難であるため、推定生存曲線上で数値的に求めた。まず、$S^{(b)}(t_0)$を線形補間により算出し、目標生存確率$S_{\text{target}} = U\,S^{(b)}(t_0)$を設定した。次に、この生存確率に対応する時点を二分探索で探索し、隣接するグリッド間を線形補間してイベント時間を求めた。未登録患者では$t_0 = 0$であるため、$S^{(b)}(t_0) = 1$とし、$S^{(b)}(T) = U$を満たす登録後イベント時間Tを同じ数値的逆変換により生成した。

```
if status_pred='ONGOING' then t0=followup;
else t0=0;
u=rand('uniform');

if t0=0 then s0=1;
else if t0 >= &EVENT_HORIZON then s0=fs[&_fpm_ng];
else do;
  pos=t0/&GRID_STEP + 1;
  lo=max(1,floor(pos));
  hi=min(&_fpm_ng,lo + 1);
  frac=pos - lo;
  s0=fs[lo] + frac*(fs[hi] - fs[lo]);
end;

target=u*s0;

if target < fs[&_fpm_ng] then event_since_entry=.;
else do;
  lo=1; hi=&_fpm_ng;
  do while(hi - lo > 1);
    mid=floor((lo + hi)/2);
    if fs[mid] > target then lo=mid;
    else hi=mid;
  end;

  if fs[lo]=fs[hi] then
    event_since_entry=(hi - 1)*&GRID_STEP;
```

```
  else event_since_entry=(lo - 1)*&GRID_STEP
    + &GRID_STEP*(fs[lo] - target)/(fs[lo] - fs[hi]);
end;
```

追跡中患者および未登録患者の暦時間上の将来イベント時点は、他の 3 モデルと同様の方法で求めた。

3. シミュレーション

表 1　シミュレーション設定

項目	設定
症例数	300 例
登録期間	24 か月 一様分布
目標イベント数	200
予測時点（か月）	6、12、18
1 年で打ち切りとなる割合	10%
真のイベント時間分布	指数分布 ワイブル分布 区分指数分布
予測モデル	指数分布モデル ワイブル分布モデル ノンパラメトリックモデル Flexible parametric model（FPM）
外側反復（仮想試験生成回数）	200
内側反復（1 試験当たりの将来経過生成回数）	1,000

指数分布シナリオでは、中央値が 12 か月となるハザードを設定した。ワイブル分布シナリオでは、中央値を 12 か月、形状パラメータを 1.5 とし、時間とともに増加するハザードを設定した。区分指数分布シナリオでは、ハザードを登録後 0-6 か月は 0.035、6-18 か月は 0.080、18 か月超は 0.030 と設定した。この設定は、単一の指数分布またはワイブル分布では表せない非単調なハザードを与える。

仮想試験を繰り返し生成する反復を外側反復、各仮想試験において将来経過を繰り返し生成する反復を内側反復とした。外側反復として真のイベント時間分布 3 種類ごとに 200 試験を生成し、各試験について三つの予測時点、四つの予測モデルで解析した。各試験では、内側反復として将来経過を 1,000 回生成した。乱数シードは真のイベント時間分布、外側反復、予測時点および予測モデルから決定論的に構成した。

各試験における真の 200 イベント到達時点を$T_{D,r}^{\mathrm{true}}$、内側反復の中央値による点予測を$\hat{T}_{D,r}$とし、予測誤差e_r、Bias および MAE を下式で定義した。正の Bias は真の到達時点よりも遅く、負の Bias は早く到達時点を予測したことを表す。

$$e_r = \hat{T}_{D,r} - T_{D,r}^{\text{true}}, \qquad \text{Bias} = \frac{1}{R}\sum_{r=1}^{R} e_r\,, \qquad \text{MAE} = \frac{1}{R}\sum_{r=1}^{R} |e_r|$$

主要評価指標は Bias および MAE とした。また、Coverage（95%予測区間が真の到達時点を含む外側反復の割合）および平均予測区間幅（上限と下限の差の平均）を補足的に評価した。

4. 結果

表 2　Bias / MAE

真のイベント時間分布	予測時点（か月）	EXP	WEIBULL	NP	FPM
指数分布	6	0.994 / 6.563	-0.584 / 7.652	1.297 / 9.545	-2.686 / 7.869
	12	0.116 / 3.151	0.778 / 4.643	-0.433 / 4.936	1.673 / 5.945
	18	0.017 / 2.101	0.285 / 2.708	-1.069 / 2.891	0.912 / 3.503
ワイブル分布	6	26.900 / 26.924	-0.592 / 5.340	10.965 / 13.925	-2.708 / 6.121
	12	12.018 / 12.018	0.391 / 2.700	1.696 / 3.899	0.640 / 3.590
	18	6.938 / 6.938	0.055 / 1.460	0.025 / 1.717	0.284 / 1.778
区分指数分布	6	21.815 / 22.363	3.619 / 7.897	17.325 / 20.139	1.891 / 10.050
	12	8.814 / 9.228	0.890 / 5.547	-3.156 / 5.194	-4.883 / 5.918
	18	2.814 / 3.685	-2.809 / 3.454	-4.137 / 4.416	-4.437 / 4.632

指数分布シナリオでは指数分布モデル、ワイブル分布シナリオではワイブル分布モデルの MAE が 6、12、18 か月時点の全てで最小であった。区分指数分布シナリオでは 6 か月時点でワイブル分布モデル、12 か月時点でノンパラメトリックモデル、18 か月時点でワイブル分布モデルの MAE が最小であり、ノンパラメトリックモデルまたは FPM が一貫して最小とはならなかった。ワイブル分布シナリオに対する指数分布モデルの Bias は 6 か月時点で 26.900 か月、12 か月時点で 12.018 か月、18 か月時点で 6.938 か月であり、予測時点が遅くなっても正の Bias が残った。

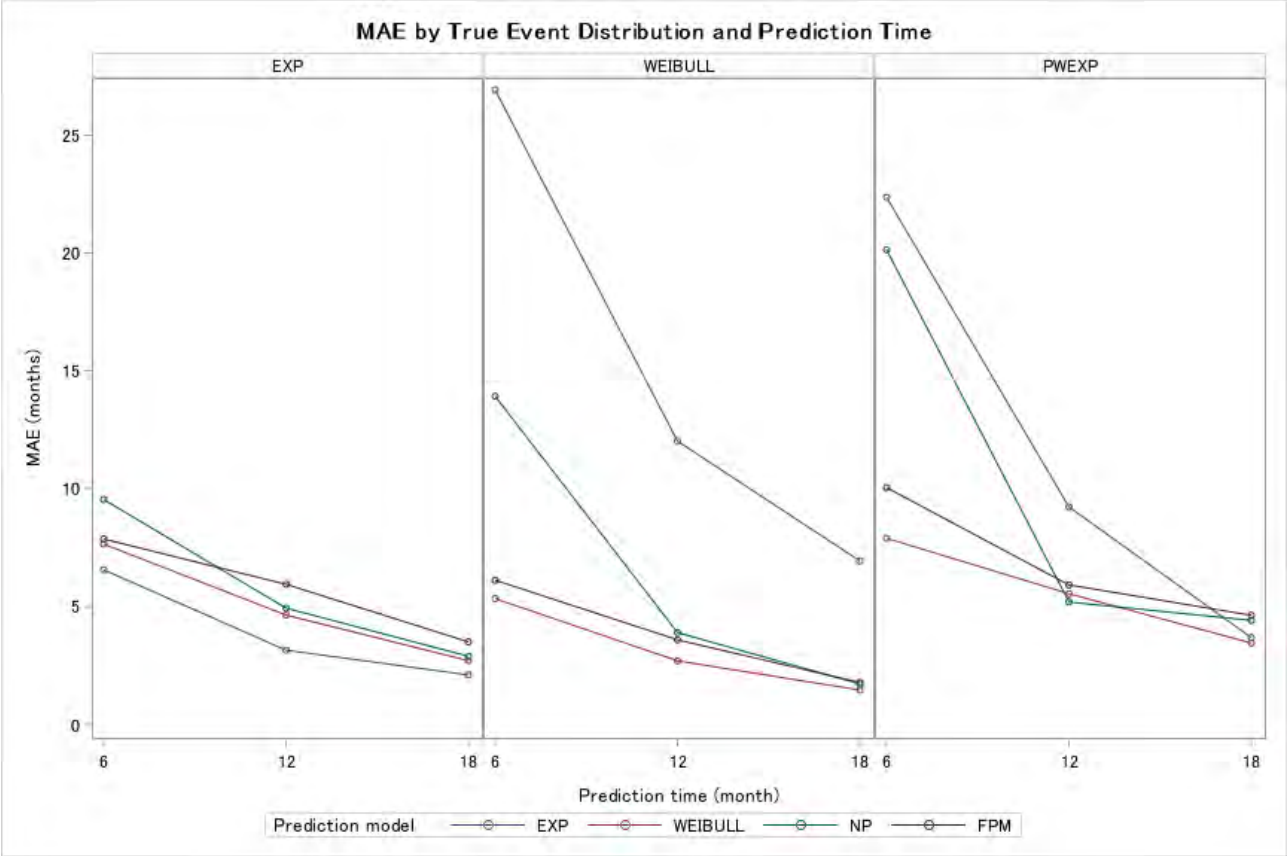


図 1　真のイベント時間分布別・予測時点別の MAE

真のイベント時間分布と予測モデルの全ての組み合わせで、予測時点が 6 か月から 18 か月へ進むにつれて MAE が低下した。指数分布シナリオでは MAE のモデル間差は比較的小さかった。ワイブル分布シナリオでは指数分布モデルの MAE が他のモデルよりも大きく、一定ハザード仮定の影響が明瞭であった。

区分指数分布シナリオでは 6 か月時点で大きなモデル間差が見られた。この時点までに観察された範囲は真のハザードが 0.035 に設定された区間であり、6 か月時点以後のハザード上昇はまだ観察されていない。したがって、モデルが柔軟であっても未観察の変化をデータのみから捉えることは難しい。12 か月時点以後は変化後のイベント情報が加わり、モデル間差が縮小した。なお、6 か月時点でワイブル分布モデルが最小の MAE であった結果を一般的な優位性として解釈すべきではない。本研究の設定の下で得られた結果である。

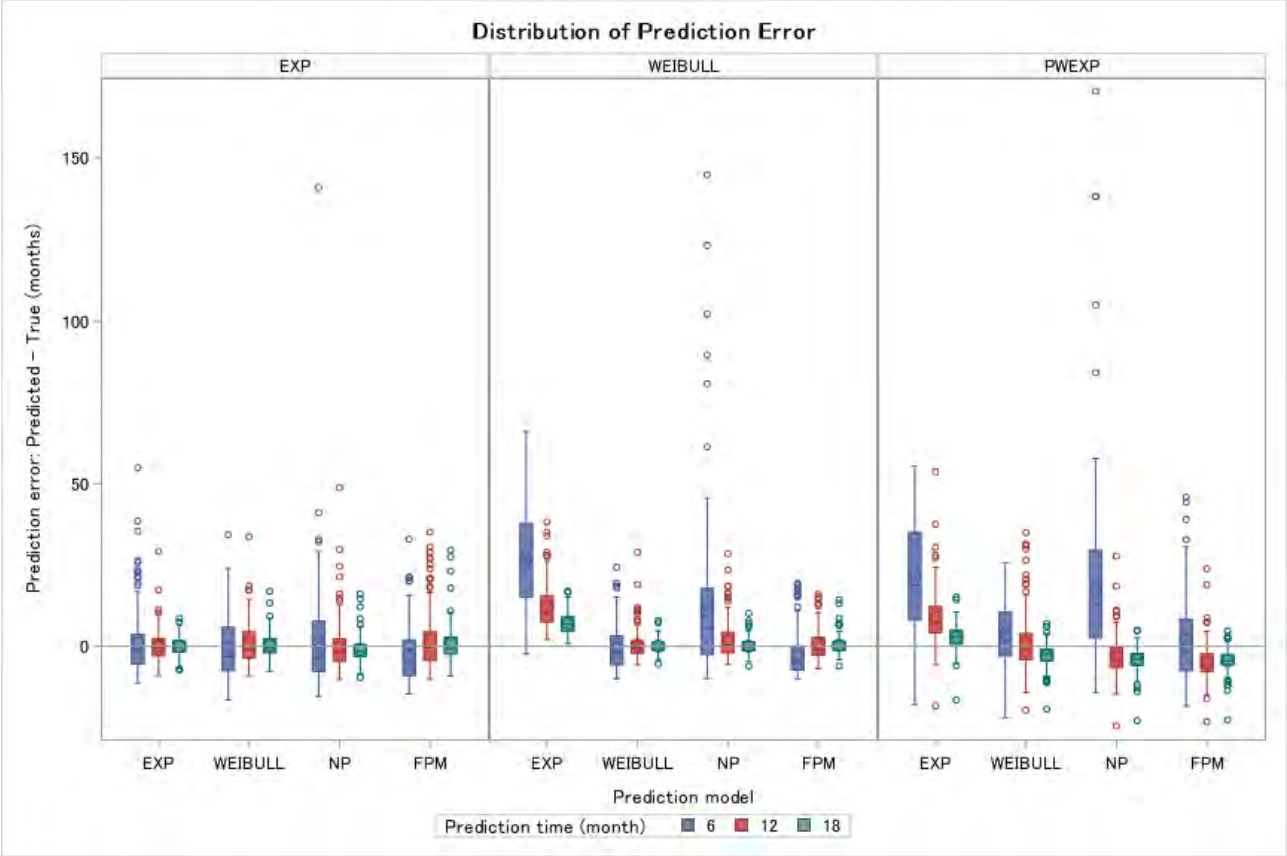


図 2　予測誤差の分布

6 か月時点では予測誤差の分布が広く、特にワイブル分布シナリオと区分指数分布シナリオに指数分布モデルを適用した場合、正の方向への偏りが大きかった。12 か月時点および 18 か月時点では予測誤差のばらつきが概して縮小し、観察情報の蓄積に伴う安定化が確認された。

表 3　Coverage / 平均予測区間幅

真のイベント時間分布	予測時点（か月）	EXP	WEIBULL	NP	FPM
指数分布	6	93.5% / 40.619	89.5% / 81.126	97.0% / 85.098	98.5% / 67.527
	12	95.0% / 16.820	96.0% / 30.379	94.0% / 32.728	96.5% / 52.868
	18	94.5% / 10.941	94.0% / 14.708	91.5% / 15.172	96.5% / 28.676
ワイブル分布	6	54.5% / 116.098	87.5% / 61.720	92.0% / 111.316	95.2% / 57.918
	12	28.5% / 29.860	95.0% / 15.737	93.5% / 29.336	96.0% / 37.997
	18	26.5% / 13.809	97.0% / 7.487	91.0% / 9.256	92.5% / 13.018
区分指数分布	6	87.5% / 113.968	96.5% / 127.332	91.5% / 135.912	99.4% / 78.513
	12	72.0% / 33.026	92.0% / 33.604	89.0% / 28.632	82.0% / 29.635
	18	85.5% / 14.379	75.5% / 10.145	70.0% / 10.223	72.5% / 11.891

区分指数分布シナリオの 6 か月時点では Coverage が 87.5-99.4%であった一方、予測区間幅は 78.513-135.912 か月と広かった。予測時点が進むと予測区間幅は縮小したが、18 か月時点の Coverage は 70.0-85.5%であった。また、ワイブル分布シナリオに指数分布モデルを適用した場合、Coverage は 6、12、18 か月時点でそれぞれ 54.5%、28.5%、26.5%であり、予測区間幅は予測時点が進むにつれて縮小したが、モデルの誤指定による系統的な偏りが残存したため、真の到達時点を十分に包含できず、Coverage は低下した。点予測精度と区間予測性能は別に評価する必要がある。

5. 考察

5.1. 結果の解釈

本研究では、目標イベント数への到達時点の予測性能が、モデルの柔軟性だけではなく、真のハザード形状、予測時点までのイベント成熟度および必要な外挿量に強く依存した。指数分布シナリオとワイブル分布シナリオでは、それぞれ正しく指定されたパラメトリックモデルが全予測時点で最小の MAE を示した。特に単調増加するハザードを仮定したワイブル分布シナリオに指数分布モデルを適用すると、早期の低いハザードを将来も維持すると仮定するため、到達時点を遅く予測する正の Bias が残った。区分指数分布シナリオでは、ノンパラメトリックモデルや FPM の柔軟性が常に小さな MAE に結び付く結果とはならなかった。6 か月時点までのデータには、その直後に設定されたハザード上昇の情報が含まれない。FPM は観察された範囲を曲線で柔軟に表現できても、未観察の構造変化を自動的に予見するわけではない。ノンパラメトリックモデルも観察範囲外では指数尾部に依存し、極端に長い予測を生じる反復があった。一方、単純なモデルが常に適切という結果でもない。ワイブル分布シナリオに対する指数分布モデルでは明確なモデル誤指定の影響があり、区分指数分布シナリオでも 6 か月時点の指数分布モデルの MAE は 22.363 か月と比較的大きかった。本研究の結果は、複雑なモデルが常に最良ではないことを示すと同時に、単純さだけを理由に単調ハザードを仮定する危険も示している。実務では、イベント成熟度、Kaplan-Meier 曲線、推定ハザードの形状、予測期間の長さおよびモデル間の予測差を併せて確認する必要がある。

予測時点が遅くなるほど MAE と予測区間幅は概して縮小したが、Coverage は単調に改善しなかった。モデルが誤指定されている場合、データが増えると推定分散が小さくなり予測区間幅が狭くなる一方、構造的な Bias は残り得る。ワイブル分布シナリオに対する指数分布モデルで 12 か月時点以後の Coverage が約 30%以下となった結果はこの点を表している。区分指数分布シナリオの 6 か月時点で高い Coverage を示したモデルの一部は、非常に広い予測区間幅を伴った。したがって、Coverage のみでモデルを評価すると、実務上利用しにくい広い予測区間幅を伴うモデルを高く評価する可能性がある。予測モデルの実用性を評価するには、Bias と MAE を主要評価指標としつつ、Coverage と予測区間幅を対として確認することが適切である。

5.2. 試験運営への応用

目標イベント数への到達時点を予測することは、中間解析の実施日をある一点に決めるためではなく、準備を開始する時期と不確実性の範囲を共有するために用いることができる。例えば、点予測を中心的な予定として、95%予測区間をデータクリーニング、解析プログラムの固定、盲検下レビュー、委員会日程および外部委託先の資源確保の時間幅として示すことが考えられる。

実際の運用では、試験開始時の予測を固定せず、イベント成熟に合わせて同じ SAS プログラムを再実行する。例えば、6 か月時点でモデル間差や予測区間が大きい場合は、単一の予定日に固定せず、12 か月時点で再評価する。複数モデルの予測が近づけば予定の確度が増したと判断できる一方、差が残る場合は外挿仮定への感度として関係者に提示できる。

SAS 実装上は、患者状態の分類、各モデルによるイベント時間分布の推定、将来イベント時間の生成、目標イベント数到達時点の予測および予測性能の評価をマクロで分離した。これにより、他の処理を維持しながらモデルの推定部分だけを変更できる。PROC LIFEREG、PROC ICPHREG、PROC IML および DATA STEP の配列・HASH OBJECT を組み合わせることで、モデルの推定から到達時点の予測までを SAS 内で完結できる。

5.3. 限界

本研究にはいくつかの限界がある。第一に、登録は 24 か月間の一様分布、打ち切りは既知の一定ハザードをもつ独立指数分布にそれぞれ従うと仮定した。実際の試験では、施設間差、登録速度の時間変化および報告遅延などが予測へ影響し得る[1]。第二に、治療群、患者背景、予後因子、競合リスクおよび情報依存打ち切りを扱っていない。第三に、真のイベント時間分布は 3 種類、区分指数分布シナリオは一つのパターンに限られ、予測時点も 3 時点のみである。第四に、ノンパラメトリックモデルの外挿は指数ハザードによる最終イベント時点以後の尾部、FPM は自由度 3、0.1 か月刻み、最大 120 か月に固定した。これらの設定に対する感度は評価していない。第五に、FPM の不確実性は推定共分散行列に基づく正規近似によって反映しており、有限標本での不確実性を完全に表すとは限らない。最後に、本研究はシミュレーションによる比較であり、実データでの予測妥当性や計算時間を検証していない。

今後は、進行中データから登録率と打ち切り率も更新するモデル、施設別登録、共変量を含むモデル、ノンパラメトリックモデルの外挿および FPM の自由度の感度分析、実データでの逐次予測を検討する必要がある。

6. まとめ

イベント駆動型臨床試験の進行中データから、目標イベント数への到達時点と 95%予測区間を求める SAS プログラムを指数分布モデル、ワイブル分布モデル、ノンパラメトリックモデルおよび Flexible parametric model の 4 手法で実装した。また、各手法をシミュレーションで比較した結果、正しく指定されたパラメトリックモデルは高い精度を示した一方、モデルの誤指定は Bias と Coverage へ影響した。区分指数分布シナリオでは、初期予測時点にノンパラメトリックモデルや Flexible parametric model が常に最小の MAE を示すわけではなかった。

本研究の範囲では、モデルの柔軟性だけで予測性能は決まらず、イベント成熟度と外挿仮定の影響が大きかった。試験運営では、単一モデルの点予測に依存せず、複数モデルの Bias、MAE、予測区間幅を比較し、試験の進行に合わせて予測を更新することが有用と考えられる。

参考文献

1. Heitjan DF, Ge Z, Ying GS. Real-time prediction of clinical trial enrollment and event counts: A review. Contemporary Clinical Trials. 2015;45:26-33.
2. Bagiella E, Heitjan DF. Predicting analysis times in randomized clinical trials. Statistics in Medicine. 2001;20:2055-2063.
3. Ying GS, Heitjan DF, Chen TT. Nonparametric prediction of event times in randomized clinical trials. Clinical Trials. 2004;1:352-361.
4. Ying GS, Heitjan DF. Weibull prediction of event times in clinical trials. Pharmaceutical Statistics. 2008;7:107-120.
5. Ying GS, Heitjan DF. Prediction of event times in the REMATCH Trial. Clinical Trials. 2013;10:197-206.
6. Rubin DB. The Bayesian bootstrap. The Annals of Statistics. 1981;9:130-134.
7. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Statistics in Medicine. 2002;21:2175-2197.
8. SAS Institute Inc. SAS/STAT User's Guide: The LIFEREG Procedure. Cary, NC: SAS Institute Inc.
9. SAS Institute Inc. SAS/STAT User's Guide: The ICPHREG Procedure. Cary, NC: SAS Institute Inc.

AIによって再構成される解析プログラミングとその役割変化の考察

瓜生　卓、小山　卓己、〇野間　基希
（ファイザーR&D合同会社 バイオメトリクス・データマネジメント統括部 スタティスティカルデータサイエンス・アナリティクスグループ）

A Consideration of the Changing Role of Statistical Programming Reconfigured by AI

Taku Uryu, Takumi Koyama, Motoki Noma
Statistical Data Sciences & Analytics Group, Biometrics & Data Management Dept., Pfizer R&D Japan G.K.

要旨

AIの台頭により、解析プログラム開発はウォーターフォール型からアジャイル型を経て、AIが要件整理・実装・検証・文書化を支援する段階へ移行しつつある。本稿では、この変化が解析プログラミングに与える影響と、求められる役割の変化を整理する。

キーワード：AI、解析プログラミング、ソフトウェア開発ライフサイクル（SDLC）、役割変化

1. はじめに

医薬品開発における解析プログラミングは、長年にわたりSASを中核的な基盤として発展し、仕様書に基づく正確な実装と再現性の確保を重視する形で成熟してきた。近年は解析に用いる言語や環境の選択肢が広がりつつあるが、要件を解釈し、実装し、結果の妥当性を検証するという本質は言語を問わず共通している。本稿で論じるAI導入に伴う役割の変化も、SASを含む解析プログラミング全般に共通するテーマである。

解析プログラミングは、ProtocolやStatistical Analysis Plan（SAP）などに基づく解析要件の整理および仕様書作成を起点として、データ処理、解析、出力生成、そして検証へと至る一連の活動から構成される。現在、AIの急速な進展により、仕様書作成、コード生成、検証支援など、各工程でAIが活用され、個別工程における効率化が進んでいる。

しかし近年では、個別工程におけるAI活用にとどまらず、こうした一連の作業手順そのものが見直され、開発プロセス全体の構造あるいは進め方がAIの利用を前提とした形で最適化されつつある。

本稿では、まず開発モデルの変遷を整理し、AI導入による変化の内容を明らかにしたうえで、解析プログラミングにおける役割の変化について考察する。

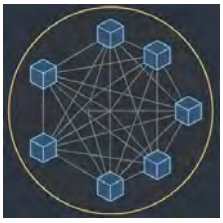
2. 解析プログラム開発モデルの変遷

従来の解析プログラム開発は、ウォーターフォール型に代表される仕様書駆動型であった。すなわち、事前に仕様書を作成・確定し、その内容に従って後工程として実装・検証を進める形である。この進め方では、要件を事前に確定することにより、後工程における手戻りを抑制しやすく、仕様に基づく一貫した結果を得やすい点に特徴がある。また、工程ごとに成果物と責任範囲が明確であるため、規制対応や品質管理の観点においても適用しやすい枠組みであった。

これに対して、近年ではアジャイル型の進め方を採用する例もみられる。このアプローチでは、開発を進めながらフィードバックを取り込み、仕様を段階的に具体化・確定していく。初期段階での詳細な仕様確定を前提とせず、実装と確認を反復することで、迅速に開発を進められる点に特徴がある。これにより、要件の変化や曖昧さに柔軟に対応できる。一方で、工程全体の整合性を維持するためには、継続的なレビューと調整が求められる。

さらに現在は、AI を前提とした解析プログラム開発の段階に入りつつある。AI はコード生成や文書作成といった個別作業の支援に加え、要件整理、仕様定義、処理設計、検証といった複数の工程を横断的に支援することが可能となっている。これにより、従来は工程ごとに分断されていた情報や判断を工程横断的に共有しながら開発を進めることが可能となり、仕様整理、実装、検証といった工程が相互に連携しながら並行的に進められ、必要に応じて往復的な見直しが行われるようになっている。その結果、各工程を独立したものとして捉える枠組みから、相互に連携した活動として捉える枠組みへ移行しつつある（表 1）。

表 1　解析プログラム開発モデルの変遷

ウォーターフォール型 （仕様書駆動）	アジャイル型 （反復とフィードバック）	AI 前提型 （工程横断的協調）
		
各工程が明確に分離。手戻りを抑制し一貫した結果を得る従来のアプローチ	迅速な開発とフィードバックの反復。仕様を段階的に具体化	AI が複数工程を横断的に支援。各工程の境界線が融合

こうした流れを背景として、近年では AI を前提にソフトウェア開発ライフサイクル（SDLC）そのものを再設計する試みが提案されている。これらは、AI の関与の程度と人との役割分担の観点から、以下の 3 つの段階に整理することができる。

1)　AI-Assisted（AI 支援型）：局所的な支援

現在普及しつつあるアプローチであり、従来の枠組みを維持したまま、コード生成や文書作成の局所的な「補助」として AI を利用する。これは、人間と AI がリアルタイムに協調してコードを書き進める、いわば「AI

とのペアプログラミング」と呼べる形態である。ここでは依然として人が実装の実行主体であり、AI は作業を効率化するためのツールにとどまっている。

2) AI-Native / AI-Driven SDLC：実行主体の移行

AI を単なる補助ツールではなく、開発プロセスにおける主要な「実装エンジン（実行主体）」として位置付ける枠組みである。

- AI の役割：プロセスの中心的な実行主体となる。与えられた仕様に基づき、コード生成やテストを主体的に実行する。
- 人の役割：実装を中心とした役割から離れ、AI に対する「仕様の定義」と、生成された結果の「検証・意思決定」へと役割を明確にシフトさせる。
- SDLC の特徴：第一の段階（AI 支援型）が従来プロセスの延長線上に位置するのに対し、本モデルは「AI が実装すること」を前提（Native）として、開発プロセスそのものを再設計・システム化している点に本質的なパラダイム転換がある。

3) Agentic SDLC：自律性と役割の拡張

AI の関与が「実行主体」からさらに「自律的実行」へと拡張した形態として位置付けられる。

- AI の役割：単なる実行にとどまらず、役割の異なる複数の AI エージェントが自律的に協調し、要件理解から実装、検証、修正に至るまでの一連のタスクを連続的に遂行する。すなわち、「解析要件を満たすために、どのような手順で実装し、どのように検証し、問題が発生した場合にどのように修正するか」といった一連の判断および処理を、AI エージェント自身が計画し、遂行する構造となる。
- 人の役割：個別の作業や逐次的な判断に直接関与するのではなく、AI エージェントの振る舞い（動作および判断）の設計、運用における監督、およびアウトプットの評価・統制を担う。
- 規制適合上の課題： AI が自律的に判断・行動するプロセスは、意思決定過程の不透明化（いわゆるブラックボックス化）を伴う可能性がある。そのため、再現性、トレーサビリティ、説明可能性といった規制要件（例：FDA や EMA の関連ガイダンス）との整合性を確保することが課題となる。したがって、本形態は現時点で直ちに実務へ適用可能なものではなく、AI の関与がさらに高度化した場合に想定される将来的なモデルとして位置付けられる（表 2）。

表 2　AI 活用レベル別の実行構造と人・AI の役割

	AI-Assisted	**AI-Native/AI-Driven SDLC**	**Agentic SDLC**
実行構造	人間が主体的に実装。AI が補助	AI が実務を主体的に遂行。人が意思決定	複数の AI エージェントが自律的に連携・遂行
AI の役割	ツール（局所的支援・補助）	開発の実行主体	開発プロセス全体の自律的遂行
人の役割	実装の実行主体	評価・判断と意思決定	AI の振る舞いの設計・監督・統制（アーキテクト）
AI の介入度	低	高	極めて高

このように、AI の関与は「局所的支援」から「自律的実行」へと段階的に拡張し、それに伴い人の役割も実装中心から評価・判断・統制へとシフトしている。これは単なる効率化にとどまらず、解析プログラム開発におけるプロセスの構造と役割分担の根本的な再構成を促す要因となっている。

3. 解析プログラミングへの影響

解析プログラミングは複数の工程から構成されるため、AI の影響は特定の工程に限定されず、広範囲に及ぶ。特に、以下の領域では AI による効率化が進みやすい。

- 仕様やルールに基づくコード生成
- 文書の作成支援
- ルールに基づく検証
- 差分検出や整合性確認

これらは、あらかじめ定義されたルールやパターンに基づく処理であり、AI との親和性が高い領域である。入力と出力の関係が明確で、規則に従って記述可能な作業ほど、AI による支援が有効に機能する。

このような特性は、具体的な実務プロセスにおいても同様の構造で適用されると考えられる。例えば、AI はコード生成に加えて、CDISC 標準に準拠したデータ構造の解釈や、define.xml や仕様書との整合性確認、さらにはトレーサビリティの確保といった複数の関連するプロセスを横断的に支援することが可能である。具体的には、仕様書に記載された変数定義と実装コードとの差分検出において、変数属性や導出ロジックの不一致を抽出することが考えられる。また、SDTM から ADaM への変換過程における派生ロジックの整合性確認として、入力変数や計算手順の対応関係を検証することや、解析結果と仕様との整合性評価として、統計量や集計結果が仕様定義と一致しているかを確認することが挙げられる。

このように、AI はルールに基づく処理や整合性確認の領域において高い有効性を示す。一方で、これらの処理は仕様やルールに基づく形式的な整合性の確保には有効であるものの、解析意図やデータの文脈を含めた意味的な適合性までを保証するものではない。

解析プログラミングにおいては、要件や仕様に対する関係者間の解釈のズレにより、認識の不一致が生じ、エラーの発生要因となる場合がある。AI は仕様や要件の構造化や要約、観点整理を支援することで、関係者間の共通理解の形成を促進し、こうした解釈のズレに起因するコミュニケーション上の齟齬の低減に寄与しうる。特に、要件の可視化や論点の明確化を通じて、従来は個人の解釈に依存していた部分を共有しやすくする点において、有効に機能することが期待される。

一方で、こうした理解は入力された要件や文脈に依存するため、前提が不十分または曖昧である場合には、AI 自身が要件を誤って解釈し、新たな解釈のズレを生じさせる可能性がある。このように、AI の導入により解釈のズレは完全に解消されるのではなく、その発生および検出の在り方が変化するため、出力結果をどのように評価し、どのように解釈するかの重要性が相対的に高まる。

このような評価および解釈の重要性の高まりを踏まえると、解析プログラマの価値は、出力結果をどのよう

に解釈し、その妥当性をどのように評価するかにあると整理できる。AI は結果の生成や形式的な検証を支援することができるが、出力が解析意図に適合しているかの評価や、想定外のデータに起因する結果の取扱いに関する判断といった側面では、人の関与が不可欠である。

したがって、出力の妥当性を評価し、その結果をどのように解釈・採用するかを判断する役割は、AI の導入後においても本質的に重要であり続ける。このように、AI の導入は単なる効率化にとどまらず、解析プログラミングにおける役割の再定義を促す要因となっている（図 1）。



図 1　AI 導入に伴う解析プログラミングにおける役割変化の概念図

4. 評価・判断が求められる背景

前章で示したように、解析プログラミングにおいては、定型的な処理（仕様やルールに基づく処理）や形式的な検証（あらかじめ定義された仕様や条件に基づく検証）は AI により効率化されつつある。しかしながら、これらの処理が正しく実行されていることと、その結果が解析として適切であることとは必ずしも一致しない。

解析においては、仕様に基づいて算出された結果であっても、データの特性や前提条件によっては、解析意図と整合しない場合がある。特に、想定外のデータが含まれる場合には、仕様どおりに処理された結果であっても、その妥当性について追加的な検討が必要となる。

このような状況では、単にルールを適用するだけでは十分ではなく、結果をどのように解釈し、どのように扱うかを判断することが求められる。例えば、仕様とコードの一致確認やデータ間の整合性確認といった形式的な検証処理自体は AI により効率化可能であるが、外れ値の取扱いや試験間での不整合の解釈といった点については、単純なルール適用では対応できない場合が多い。このようなケースでは、解析の目的やデータの背景を踏まえたうえで、結果をどのように扱うべきかを人が判断する必要がある。

近年は、臨床試験データを取り巻く環境の二極化が進んでおり、これが評価・判断の重要性を一層高める要因となっている。ICH M11（臨床プロトコルの電子的構造化）や Unified Study Definitions Model（USDM）といった構想の進展により、厳格に管理された試験においてプロトコルから報告書作成まで一気通貫でデータフローが整備される環境が整いつつある。このような標準化が進む領域においては、形式的なエラーや予見可能なイレギュラー（想定外のデータ）は減少し、AI による自律的な処理が容易になる。

しかし他方で、リアルワールドデータ（RWD）など、多様かつ非定型なデータソースを活用する機会が急増している。RWD は臨床研究の目的の下で計画的に収集・品質管理されたデータではないことから、臨床研究においては「想定外」と解釈され得るデータが含まれやすいが、これらを解析要件にどう適合させるかは単純なルール適用では解決できない。加えて、試験デザインの複雑化、あるいはウェアラブルデバイス等を用いた収集データの多様化も、過去のパターンには当てはまらない新たな「想定外」を生み出す要因となっている。

特に、このような多様なデータを扱う過程においては、データ間の不整合や矛盾、あるいは想定されていなかったパターンが顕在化しやすい。これらについては、AI を活用することで不整合や想定外のパターンの検出自体は効率化・高度化が期待される一方で、検出された事象をどのように解釈し、どのように対処・判断するかについては、解析の目的やデータの背景を踏まえた人の関与が不可欠である。したがって、データの多様化が進むほど、AI による検出とそれに基づく人の判断との役割分担の重要性は一層高まる。

すなわち、一部の定型的な領域で標準化が進むからこそ、逆に非定型なデータソース（RWD 等）や複雑な試験デザインにおいて生じる「想定外のデータ」に対する、データの特性や解析意図を踏まえた「解釈」と「判断」の価値は相対的に、そして飛躍的に高まるのである。

また、AI による生成結果は必ずしも正確であるとは限らず、誤った出力や不完全な理解に基づく結果が含まれる可能性がある。さらに、生成過程の再現性や判断根拠の説明可能性といった観点においても課題が残されている。このため、AI の出力をそのまま受け入れるのではなく、その妥当性を適切に評価することが不可欠である。

さらに、こうした評価・判断の重要性は、近年の規制当局による AI 関連文書においても一貫して示されている。FDA および EMA が共同で公表した Guiding Principles of Good AI Practice in Drug Development では、医薬品開発における AI 利用に関して、人間中心の設計（Human-Centric by Design）、リスクベースアプローチ、標準への準拠、使用目的（Context of Use）の明確化、データガバナンスと文書化、モデル開発・性能評価、ライフサイクル管理など、10 の原則が示されている。これは、AI の利用を単なる技術導入としてではなく、使用目的、リスク、データ、モデル、および運用管理を含む包括的な品質管理の枠組みとして捉える考え方である。

この考え方は個別の規制文書にも具体化されている。例えば、FDA の 2025 年ドラフトガイダンスでは、AI を活用したシステムに対して、ライフサイクル全体を通じた設計、文書化、リスク管理、透明性の確保、バイアスへの対応、および継続的な性能監視が求められている。また、EMA が公表した医薬品ライフサイクルにおける AI 利用の見解書（Reflection paper）においても、臨床試験をはじめとする各プロセスでの人間による監視の重要性や、データ・モデル・意思決定プロセスおよびトレーサビリティの確保の必要性が強調されている。これらに共通する考え方は、AI の利用にあたって、その使用目的（Context of Use）を明確に定義し、出力の信頼性をリスクに応じて評価・文書化することであり、最終的な判断や責任は人間が担うという点にある。

EMA の Annex 22 においても、継続的に学習・適応する動的な AI モデルについては、GMP の重要な用途において使用すべきではないとされており、人間による確認と承認（Human-in-the-loop）が必須とされている。

このような再現性や監査証跡を重視する規制当局のスタンスは、製造領域（GMP）にとどまらず、臨床試験や製造販売後調査などの GxP 規制下の業務において、医薬品の信頼性確保という観点から本質的に共通する要求（考え方）であると言える。

これらの規制は、再現性、トレーサビリティ、および監査可能性を担保できる範囲においてのみ AI の利用を許容するという前提に立っている。そのため、実行中に挙動が変化するような自律型 AI の運用は、これらの要件と整合しにくく、結果として適用が強く制約される構造にある。このように、AI の自律性が高まるほど、その利用可能範囲は拡大するのではなく、むしろ規制上の制約との関係において慎重に設計される必要がある。

これは、AI が高性能化するほど、その出力を前提として業務を進めることのリスクが高まると同時に、その出力が規制要件に照らして妥当であることを説明・証明する責任もまた増大することを意味している。すなわち、AI の自律性が高まるほど、人の関与が不要になるのではなく、その利用範囲をどのように定義し、その結果に対してどのように責任を持つかを設計する役割が強く求められる。

したがって、AI による処理支援が進む中においても、何をもって妥当とするかを定義し、その結果を受け入れるか否かを決定する役割は、従来以上に重要な意味を持つ。特に実務においては AI の活用と人の判断のバランスを適切に設計し、その妥当性を説明可能な形で継続的に維持することが重要となる。

5. 評価・判断に関わる役割の変化

解析プログラミングにおいて、整合性確認や形式的な検証は AI による支援が進む領域である。コードと仕様の差分検出やルールへの適合確認などは、あらかじめ定義された条件に基づいて判定可能であり、AI が得意とする処理である。このため、これらの領域では効率化が進みやすく、形式的な誤りや不整合の検出において AI の役割は今後さらに大きくなると考えられる。

一方で、解析プログラマに求められる役割は、こうした形式的な検証では捉えきれない側面にある。すなわち、以下のような観点では、判断の質が重要となる。

- 出力結果が解析意図に照らして適切であるかの評価
- 想定外のデータに起因する結果の取扱いに関する判断
- 結果の解釈に基づく判断基準の整理
- 判断根拠を説明可能とするための整理
- 品質およびリスクに対する最終的な判断

これらは、単にルールを適用することで導かれるものではなく、データの特性や解析の目的、前提条件といった文脈を踏まえた総合的な判断を必要とする。そのため、AI が処理を支援できる範囲が広がるほど、何をもって妥当とするかを定義し、その判断結果に責任を持つ役割がより重要となる。

このように、AI の導入は確認作業の効率化をもたらす一方で、解析プログラマに求められる役割を、単なる検証から評価・判断・統制へとシフトさせるものである。

6. 解析プログラミングの役割再定義

これまでに示した解析プログラム開発モデルの変化および AI の導入による影響を踏まえると、解析プログラミングに求められる役割は従来とは大きく異なるものとなる。AI の導入により、定型的な実装や形式的な検証の多くは効率化されつつある。その結果、単に仕様どおりに実装することに加えて、得られた結果をどのように評価し、どのように扱うかを判断することの重要性が高まっている。

近年提案されている AI-Native / AI-Driven SDLC といった枠組みにおいても、AI が実装や検証の多くを担う一方で、人は生成された結果の評価、妥当性確認、重要な意思決定を担う役割へと移ることが示されている。また、Agentic SDLC に関する議論では、AI エージェントが複数工程を横断して委任された実行を担い、人は監督、評価、統制を担う構図が強調されている。このように、AI の導入により、実装中心であった従来の役割から、評価・判断・統制を中心とする役割への転換が起こりつつある点は、分野を問わず共通した傾向である。

例えば、仕様に基づいて算出された結果であっても、想定外のデータにより解析意図と整合しない場合がある。このような場合には、その結果をそのまま採用するのか、修正するのか、あるいは仕様自体を見直すのかといった判断が求められる。ここでは、AI が出力を生成することよりも、その出力をどのように受け止め、どこまで受容し、何を追加で確認すべきかを見極めることが重要となる。

このような変化を踏まえると、解析プログラマに求められる役割は以下のように整理できる。

- AI 活用を前提とした解析プログラム開発の進め方および役割分担の設計
- 出力結果の妥当性の評価とその解釈
- 結果の取扱いに関する判断
- 品質および手順の管理と統制

ここでいう設計には、人と AI の役割分担を定めることに加えて、AI に解析意図や文脈を適切に伝えるための要件整理や、AI が必要な情報を適切に参照できるように仕様・ルール・関連情報を整備することも含まれる。このような設計により、AI に伝達される情報の曖昧性は低減され、より一貫した処理が可能となる。しかしながら、解析においては完全に形式化できない文脈や、複数の解釈が生じ得る。特に、AI の関与によって仕様整理、実装、検証といった工程が並行的かつ反復的に実行される環境下では、事後的に計画や仕様が変更される「結果に基づく意図的な選択（後出し）」と疑われやすいプロセス上の課題が発生し得る。また、複数の生成結果や解釈の選択肢の中から、無意識に都合の良い処理を採用してしまう可能性も否定できない。これらの課題は、試験結果やその解釈に偏り（バイアス）の影響を継続的に監視し、解析プロセスの透明性と説明可能性を適切に管理・統制することが求められる。従来のように完成した成果物を事後的に確認する品質管理（QC）に過度に依存するのではなく、AI に対する解析要件の構造化や、生成から検証に至るトレーサビリティの確保など、人と AI が協調するプロセスそのものに品質を内包させる「Quality by Design（QbD）」のアプローチが不可欠となる。

また、AI の活用が前提となる環境においては、これらの評価・判断を適切に行うための基盤として、プログラミング能力を維持・強化することも不可欠である。この能力は単にコードを記述するためのものではなく、

AI の生成結果を読み解き、その妥当性を判断するための基盤として再定義されるべきものである。さらに、AI の自律的な処理過程がブラックボックス化するのを防ぐためにも、生成された「コード」は AI の処理内容を可視化し、人と AI の間の意図をすり合わせるための極めて重要なコミュニケーション手段としての役割を担うことになる。

すなわち、解析プログラマの役割は、仕様に基づく実装中心の役割から、開発全体を設計し、結果を評価し、その取扱いを判断し、品質を統制することへとシフトしている。また、AI の導入により、結果の妥当性や適用範囲をどのように定義し、その責任をどのように担保するかを設計する役割がより重要性を増している。AI の導入は、この変化を新たに生み出すというよりも、従来から重要であった評価・判断・統制の役割を、より明確に可視化する契機になっていると捉えることができる。

7. おわりに

解析プログラム開発は、ウォーターフォール型（仕様書駆動型）からアジャイル型を経て、AI を前提とした開発段階へと変化しつつある。この変化により、定型的な実装や確認作業では効率化が進む一方で、結果の評価や取扱いに関する判断といった役割の重要性が相対的に高まっている。さらに、近年提案されている AI-Native / AI-Driven SDLC や Agentic SDLC といった概念は、AI の導入が開発モデル全体と人の役割分担そのものを再構成する方向へ進んでいることを示している。

本稿で示したとおり、解析プログラミングにおいては、仕様に基づく正確な実装に加えて、得られた結果をどのように解釈し、どのように受け入れるかを決定することが重要な役割である。これらは従来から求められてきたものであるが、AI の導入により定型的な処理が効率化されることで、その重要性がより明確になっている。今後の解析プログラミングにおいては、実装に加えて、AI を活用しながら開発全体を設計し、結果を評価し、その取扱いを判断し、品質を統制する役割がより中心的な意味を持つようになる。

一方で、AI による支援の高度化に伴い、どの範囲までを AI に委ね、どの判断を人が担うべきかといった役割分担の整理や、判断基準の明確化といった課題も残されている。特に、AI の自律性が高まる一方で、その利用をどのように定義し、どの範囲を統制し、どの結果に責任を持つのかという問題は、今後の解析プログラミングにおける中心的な論点となると考えられる。

さらに重要なのは、このような変化を踏まえ、解析プログラマ自身がどのような役割を担っていくのか、何を強化し、何を手放していくのかについて、継続的に議論していくことである。AI は強力な支援手段である一方で、その出力を無批判に受け入れることはできない。むしろ、その内容を理解し、解析として意味のある結果であるかを判断する責任は、これまで以上に人に依存する。

今後の解析プログラミングにおいては、AI を活用しながらも、その振る舞いを理解し、意味のある結果として成立させるための判断を担う存在としての役割が、これまで以上に重要なものとなる。この点について、実務の現場においても継続的に議論を深めていくことが必要であるとともに、その中で解析プログラマとしての専門性をどのように位置付けていくかが、今後の重要な課題となる。

なお、本稿で論じたこうした解析プログラミングの役割の重要性は、将来的な規制当局へのデータ提出の在

り方の変化を見据えると、さらに一段と高まることが予想される。例えば、近年ではR Consortium と FDA が合同で進める「R Submissions Pilot Project」に代表されるように、従来の静的な図表に加えて、探索的なデータ確認を可能とするインタラクティブな解析アプリケーションを規制当局へ提出する議論が進展している。これに伴い、今後の解析プログラミングにおいては「インタラクティブな解析環境の設計者」としての役割への拡張が求められるようになると考えられる。このような出力結果の動的化（システム化）と AI による支援の高度化が同時に進む未来の環境下においては、どの範囲までを AI に委ね、システムの妥当性をどのように人が保証するかといった課題について、さらなる検討を重ねていく必要がある。

謝辞

本稿の作成にあたり、貴重なご助言をいただいたファイザーR&D 合同会社クリニカル・ファーマコロジー＆バイオアナリティクス部の土綿 慎一氏に深く感謝申し上げます。

参考文献

1. Hymel, C. (2024). *The AI-Native Software Development Lifecycle: A Theoretical and Practical New Methodology*. arXiv preprint arXiv:2408.03416.
2. Raja, S. P. (2025, July 31). *AI-Driven Development Life Cycle: Reimagining Software Engineering. AWS DevOps & Developer Productivity Blog*.
3. Bhati, H. (2026). *Agentic AI in the Software Development Lifecycle: Architecture, Empirical Evidence, and the Reshaping of Software Engineering*. arXiv preprint arXiv:2604.26275.
4. U.S. Food and Drug Administration, & European Medicines Agency. (2026, January). *Guiding Principles of Good AI Practice in Drug Development*.
5. U.S. Food and Drug Administration. (2025, January). *Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products: Draft Guidance for Industry and Other Interested Parties*.
6. European Medicines Agency. (2024, September). *Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle. EMA/CHMP/CVMP/83833/2023.*
7. European Commission. (2025, July). *EU GMP Annex 22: Artificial Intelligence (Draft). EudraLex Volume 4.*

単群試験におけるノンパラメトリックベイズ流モデルに基づくマイルストン生存時間解析

○吉本 拓矢[1]，澤本 涼[1]，田畑 耕治[2]
（[1]中外製薬株式会社 バイオメトリクス部，[2]東京理科大学 創域情報学部）

Milestone Survival Analysis based on Non-Parametric Bayesian model for Single-arm Study

Takuya Yoshimoto[1], Ryo Sawamoto[1], Kouji Tahata[2]
[1]Biometrics Dept., Chugai Pharmaceutical Co., Ltd.,
[2]Faculty of Information Science and Technology, Tokyo University of Science.

要旨

がん臨床試験において，第II相試験はより短期的なエンドポイントとしてRECIST v1.1による奏効の有無に基づく単群試験として実施されることが多い。しかしながら，高い奏効割合が得られ，かつ長期予後が期待できるがん種においては，従来の奏効の有無を主要評価項目とした単群試験における意思決定戦略の実効性は限定的である。他方で，第III相試験で設定される無増悪生存時間及び全生存時間等のtime to event型アウトカムを主要評価項目としたランダム化第II相試験の実施は，有意水準を緩めたスクリーニングデザイン（Green et al., 2013）を前提としても，エンリッチメント戦略を要する場合等には登録可能性の観点から実施は困難を極める。このような状況において，Chen（2015）は特にがん免疫療法に焦点を当て，マイルストン生存時間解析の有用性を主張している。また，検証的な第III相試験の実施可否の意思決定に必要となるエビデンスの創出を目的とした場合，当該意思決定は規制当局による薬事承認判断ではなく開発者側の判断を前提とするため，頻度流デザインよりも柔軟なベイズ流デザインが適合していると考えられる。そこで，本稿ではtime to event型アウトカムを評価項目とした単群試験において，ノンパラメトリックベイズ流モデルに基づくマイルストン生存時間解析の方法を紹介し，SASで実装する方法を解説する。

キーワード： 確率過程，ガンマ過程，ノンパラメトリックベイズ，ポアソン過程，区分指数モデル

1　はじめに

がん臨床試験において，第 II 相試験は抗腫瘍効果について新薬のスクリーニングを行い，検証的な第 III 相試験を立案するための基となるエビデンスを創出することを目的としている（Green et al., 2013）。伝統的に，より短期的なエンドポイントとして RECIST v1.1 による奏効の有無に基づく単群試験として実施される

ことが多い。一方，近年のドライバー遺伝子異常を有する患者に対する分子標的薬，或いはがん免疫療法等により，高い奏効割合が得られ，かつ長期予後が期待できるがん種も見られている。このようながん種を目標とした新薬の開発において，従来の奏効の有無を主要評価項目とした単群試験における意思決定戦略の実効性は限定的である。他方で，第 III 相試験で設定される無増悪生存時間及び全生存時間等の time to event 型アウトカムを主要評価項目としたランダム化第 II 相試験の実施は，有意水準を緩めたスクリーニングデザイン（Green et al., 2013）を前提としても，エンリッチメント戦略を要する場合等には登録可能性の観点から実施は困難を極める。このような状況において，Chen（2015）は特にがん免疫療法に焦点を当て，マイルストン生存時間解析の有用性を主張している。この解析は，臨床的に関心のある時点に対する生存割合を評価する接近法であり，評価時点の選択には注意が必要であるものの，time to event 型アウトカムに基づく早期の意思決定を可能とし得る。これらの点を考慮して，検証的な第 III 相試験の実施可否の意思決定に必要となるエビデンスの創出を目的とした場合に，単群試験における time to event 型アウトカムに基づくマイルストン生存時間解析に注目する。また，上述の意思決定は，規制当局による薬事承認判断ではなく開発者側の判断を前提とするため，頻度流デザインよりも柔軟なベイズ流デザインが適合していると考えられる。加えて，ノンパラメトリック法はより柔軟な分布の特徴を捉えることが可能であり，後続の検証試験で適用され得るノンパラメトリックな生存時間解析を踏まえると，ノンパラメトリック手法に基づく意思決定がより適していると考える。

以上を踏まえ，本稿では time to event 型アウトカムを評価項目とした単群試験において，ノンパラメトリックベイズ流モデルに基づくマイルストン生存時間解析の方法及びその意思決定方法を紹介する。具体的には，確率過程を用いて分布関数を表現し，得られた累積ハザード関数の増分に対する事後分布の導出方法を紹介する。さらに，累積ハザード関数の増分に対する事後分布から，ハザード関数の事後分布を導出し，区分指数モデルに基づくハザードの事後分布との関連性を整理する。その上で，これらを SAS で実装する方法を解説する。なお，使用する SAS のバージョンは，SAS/STAT 15.3（SAS 9.4 TS1M8）である。

2　ノンパラメトリックベイズ流モデル

2.1　計数過程表現

単群試験の総被験者数をnとし，i番目の被験者に対してイベントが観察されるまでの時間をt_i，打切りまでの時間をc_iとする（$i = 1, \dots, n$）。$N_i(t) = I(t_i \le t, t_i \le c_i)$を時間$t$までイベントを起こしていない場合に$N_i(t) = 0$，イベントが生起している場合に$N_i(t) = 1$となる階段関数とする。ここに，$I(\cdot)$は指示関数である。微小時間$dt$に対する$N_i(t)$の増分を$dN_i(t) = N_i(t^- + dt) - N_i(t^-)$とすると，$dN_i(t)$は 0 か 1 の値のみを取る確率変数となる。ただし，t^-は時点tの直前を表す。また，$Y_i(t) = I(t_i \ge t, c_i \ge t)$を at risk 関数とする。すなわち，$Y_i(t)$は時点$t$の直前までに打切りとならず，かつイベントが生起していない場合に 1 をとり，そうでなければ 0 をとる関数である。さらに，区間$[t, t + dt)$の間でイベントを起こす瞬間の発生率として，ベースラインハザード関数（単群試験におけるハザード関数）を$\lambda(t)$とする。このとき，以上の表記に対する具体例を図 1 に示す。このとき，微小時間dtの間にイベントが生起する確率は次式で与えられる（図 1 の被験者 4 のケース）。

$$\Pr(dN_4(t) = 1 | N_4(t^-) = 0, Y_4(t^-) = 1) = \lambda(t)dt$$

同様に，時間tの直前までにイベントを起こしている場合（図 1 の被験者 1，2 のケース）並びにi番目の被験者が打切りである場合（図 1 の被験者 3 のケース）ではそれぞれ

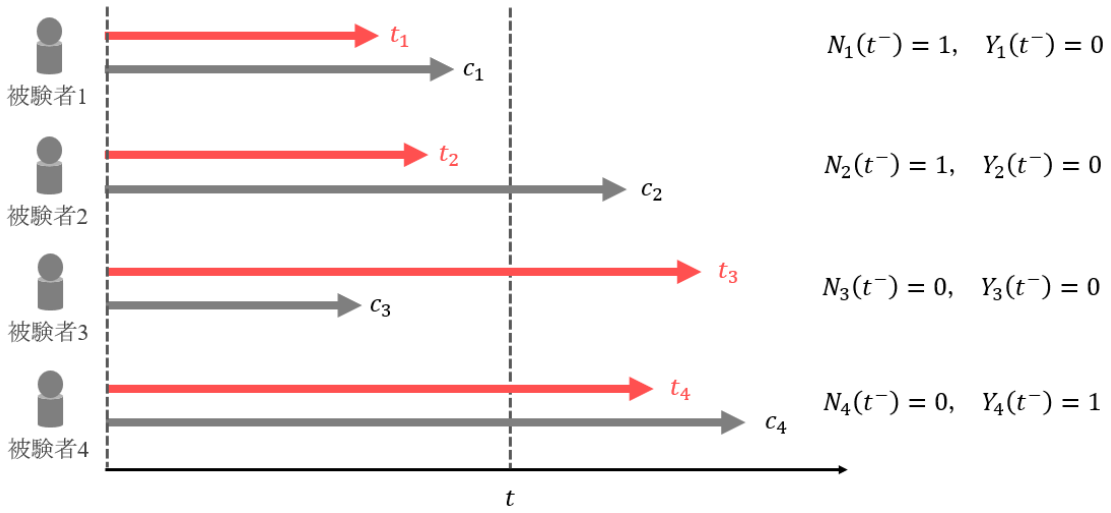


図 1 時間tにおける具体例（総被験者数が 4 例の場合）

$$\Pr(dN_1(t) = 1|N_1(t^-) = 1, Y_1(t^-) = 0) = 0,$$
$$\Pr(dN_2(t) = 1|N_2(t^-) = 1, Y_2(t^-) = 0) = 0,$$
$$\Pr(dN_3(t) = 1|N_3(t^-) = 0, Y_3(t^-) = 0) = 0$$

となる。これらは，時間tの直前にリスク下にない被験者においては微小時間dtの間にイベントが生起する確率は 0 を取ることを意味する。ここで，$N_i(t)$，$Y_i(t)$及び$\lambda(t)$の$t = 0$からtの直前までの値の経過の履歴に対応する完全加法族を$\mathcal{F}(t^-)$で示すことにすると，上式は次式のように整理できる。

$$\Pr\big(dN_i(t) = 1|\mathcal{F}(t^-)\big) = Y_i(t)\lambda(t)dt$$

また，この条件付き確率は大きさが 1 の二項分布（ベルヌーイ分布）であるから，

$$E\big(dN_i(t)|\mathcal{F}(t^-)\big) = Y_i(t)\lambda(t)dt = Y_i(t)d\Lambda(t)$$

である。ここに，$d\Lambda(t)(= \lambda(t)dt)$は累積ハザード関数の増分であり，$Y_i(t)\lambda(t)$は値が大きければイベントが生起しない確率は小さくなることから強度関数とも呼ばれる。これらの加算過程表現の詳細は，Hobbs and Carlin（2008），中村（2001）及び矢田，浜田（2016）等を参照されたい。

2.2 ポアソン過程モデル

観測されたn人のデータを$D = \{N_i(t), Y_i(t)\}_{i=1}^n$とし，データ$D$が与えられた下での尤度を考える。なお，$\Lambda(t)$は累積ハザード関数である。Clayton（1991）では，無情報打切りの仮定のもと，加算過程表現を用いて，i番目の被験者に対する尤度を以下のように表現している。

$$\left[\prod_{t\geq 0} I_i(t)\right]^{dN_i(t)} \exp\left[-\int_{t\geq 0} I_i(t)dt\right]$$

ただし，$I_i(t) = Y_i(t)\lambda(t)$は強度関数とする。上記より，尤度は本質的に時点$t$ごとに平均$I_i(t)dt$を持つポアソン分布の核（$dN_i(t) \sim Poisson(I_i(t)dt)$）からなり，$i$番目の被験者に対する尤度が多くのデータ“atoms（単位）”の独立した寄与によって構成されるイベント発現若しくは打切りまでのポアソン過程とみなす（Clayton, 1988; Clayton, 1991）。ただし，各“atoms（単位）”は，強度関数が区間内で一定とみなせるような非常に短い区間での個々の観察を対象とする。

いま，Ibrahim et al.（2001）に従い，対象となるイベント発生時刻を表す確率変数列を$0 < t_1 < t_2 < \cdots < t_J$として，観察されたイベントに対して$J$個の区間，$(0, t_1], (t_1, t_2], \cdots, (t_{J-1}, t_J]$について時点$t_j$までの累積のイベ

ント発生件数を考える。また，時点t_jの累積ハザード関数を$\Lambda(t_j)$とし，同様に累積ハザード関数の増分を$d\Lambda(t_j)(=\Lambda(t_j)-\Lambda(t_{j-1})=\lambda(t_j)dt)$とする。尤度は$J$個の区間及び$n$人の被験者に対する独立な寄与の積として表現できるから，次式のように書ける。

$$\prod_{i=1}^{n}\left[\prod_{j=1}^{J}\frac{\exp\left(-Y_i(t_j)d\Lambda(t_j)\right)\left(Y_i(t_j)d\Lambda(t_j)\right)^{dN_i(t_j)}}{\left(dN_i(t_j)\right)!}\right] \tag{1}$$

ここで，$r_j=\sum_i Y_i(t_j)$，$d_j=\sum_i dN_i(t_j)$として，それぞれ区間$(t_{j-1},t_j]$におけるリスクセット数及び区間$(t_{j-1},t_j]$における死亡数とすると，(1)式は下記のように書き表すことができる。

$$\prod_{i=1}^{n}\left[\prod_{j=1}^{J}\frac{\exp\left(-Y_i(t_j)d\Lambda(t_j)\right)\left(Y_i(t_j)d\Lambda(t_j)\right)^{dN_i(t_j)}}{\left(dN_i(t_j)\right)!}\right]=\prod_{j=1}^{J}\frac{\exp\left(-r_j d\Lambda(t_j)\right)\left(r_j d\Lambda(t_j)\right)^{d_j}}{d_j!}$$

$$\propto\prod_{j=1}^{J}\left[\exp\left(-r_j d\Lambda(t_j)\right)\left(d\Lambda(t_j)\right)^{d_j}\right]=\exp\left(-\sum_{j=1}^{J}r_j d\Lambda(t_j)\right)\prod_{j=1}^{J}\left(d\Lambda(t_j)\right)^{d_j} \tag{2}$$

(2)式より，各区間で独立にポアソン分布に従うことから，ポアソン過程となることがわかる。

2.3　ガンマ過程モデル

本項では，Kalbfleisch（1978）及び Ibrahim et al.（2001）に従い，ガンマ過程（Gamma Process）を考える。$t\geq 0$に対して，$Z(t)$を以下の性質を満たす確率過程とするとき，$\{Z(t):t\geq 0\}$をガンマ過程と呼ぶ（Ibrahim et al., 2001; 矢田，浜田，2016）。

(i). $Z(0)=0$

(ii). $Z(t)$は排反な区間において独立増分を持つ

(iii). $\alpha(0)=0$を満たす左連続な増加関数$\alpha(t)$と$c>0$が存在するとき，任意の$t>s$に対して

$$Z(t)-Z(s)\sim Ga(\alpha(t)-\alpha(s),c)$$

である。$Ga(\alpha,\beta)$は，形状パラメータαと尺度パラメータβのガンマ分布を表す。

ガンマ過程$\{Z(t):t\geq 0\}$は，$GP(\alpha(t),c)$と表現される。いま，2.2 項で示した累積ハザード関数に関してガンマ過程を導入する。$\Lambda^*(0)=0$を満たす左連続な増加関数$\Lambda^*(t)$と$c>0$を用いて，$\Lambda(t)\sim GP(c\Lambda^*(t),c)$を仮定する。このとき，確率変数列$t_j\ (j=1,\ldots,J)$について，累積ハザード関数の増分$d\Lambda(t_j)$は性質(iii)より以下のガンマ分布に従う。

$$d\Lambda(t_j)\sim Ga\left(cd\Lambda^*(t_j),c\right) \tag{3}$$

ただし，$d\Lambda^*(t_j)=\Lambda^*(t_j)-\Lambda^*(t_{j-1})$である。(3)式は，平均$d\Lambda^*(t_j)$，分散$d\Lambda^*(t_j)/c$のガンマ分布であり，分散から読み取れるとおり，$c$は$d\Lambda^*(t_j)$への確信度を表す。

2.4　事後分布

イベント発生件数にポアソン過程を導入し，累積ハザード関数に対してガンマ過程を導入した。本項では，①ノンパラメトリックベイズ流モデルに基づく累積ハザード関数の増分に対する事後分布を導出し，②得られた結果から時点t_jにおける当該区間のハザード関数の事後分布の導出を試みる。

まず，目的の 1 つ目である累積ハザード関数の増分に関する事後分布を導出する。観測されたn人のデータDのうち，j番目の区間のデータをD_jとするとき，$d\Lambda(t_j)$の事後分布は下記のベイズの定理より得られる。

$$p(d\Lambda(t_j)|D_j) \propto p\left(D_j|d\Lambda(t_j)\right) p\left(d\Lambda(t_j)\right)$$

ここに，$p(d\Lambda(t_j)|D_j)$は$d\Lambda(t_j)$の事後分布である。また，$p\left(D_j|d\Lambda(t_j)\right)$は尤度であり，(2)式より$j$番目の区間に対して次式で与えられる。

$$p\left(D_j|d\Lambda(t_j)\right) \propto \exp\left(-r_j d\Lambda(t_j)\right)\left(d\Lambda(t_j)\right)^{d_j} \tag{4}$$

$p\left(d\Lambda(t_j)\right)$は$d\Lambda(t_j)$の事前分布であり，(3)式のとおりガンマ分布に従う。具体的には以下の確率密度関数を持つ。

$$p\left(d\Lambda(t_j)\right) \propto d\Lambda(t_j)^{cd\Lambda^*(t_j)-1} \exp\left(-cd\Lambda(t_j)\right) \tag{5}$$

(4)と(5)式より，事後分布は以下のように導かれる。

$$\begin{aligned} p(d\Lambda(t_j)|D_j) &\propto p\left(D_j|d\Lambda(t_j)\right) p\left(d\Lambda(t_j)\right) \\ &= \exp\left(-r_j d\Lambda(t_j)\right)\left(d\Lambda(t_j)\right)^{d_j} d\Lambda(t_j)^{cd\Lambda^*(t_j)-1} \exp\left(-cd\Lambda(t_j)\right) \\ &= \exp\left(-d\Lambda(t_j)(c+r_j)\right) d\Lambda(t_j)^{cd\Lambda^*(t_j)+d_j-1} \end{aligned}$$

ガンマ分布とポアソン分布は共役の関係にあり，事後分布もガンマ分布に従う。具体的には下記のパラメータを持つガンマ分布に従うことがわかる。

$$d\Lambda(t_j|D_j) \sim Ga\left(cd\Lambda^*(t_j) + d_j, c + r_j\right) \tag{6}$$

ここまでの操作で，累積ハザード関数の増分が(6)式で与えられる事後分布に従うことを示した。しかしながら，結果の解釈にあたっては累積ハザード関数の増分ではなく，ハザード関数そのものの事後分布を得ることが望ましいと考えられる。そこで，以降では 2 つ目の目的である時点t_jにおける当該区間のハザード関数の事後分布を考える。区間内でハザードが一定とみなせる場合，ハザード関数は累積ハザード関数の増分に対する平均変化率で定義されるため，以下の変数変換を施すことで対応する事後分布を得ることができる。

$$\lambda(t_j|D_j) = \frac{d\Lambda(t_j|D_j)}{dt_j}$$

ただし，$dt_j = t_j - t_{j-1}$であり，j番目の区間幅に対応する。$d\Lambda(t_j|D_j)$を$1/dt_j$倍するという尺度変換に対応するため，上記の変数変換を施してもガンマ分布に従うことがわかる。ここに，$p\left(\lambda(t_j|D_j)\right)$は$\lambda(t_j)$の事後分布であり，ヤコビアンが$dt_j$となることに注意すると

$$\begin{aligned} p\left(\lambda(t_j|D_j)\right) &\propto \exp\left(-\lambda(t_j|D_j)(c+r_j)dt_j\right)\left(\lambda(t_j|D_j)dt_j\right)^{cd\Lambda^*(t_j)+d_j-1} \times dt_j \\ &\propto \exp\left(-\lambda(t_j|D_j)(c+r_j)dt_j\right)\left(\lambda(t_j|D_j)\right)^{cd\Lambda^*(t_j)+d_j-1} \end{aligned}$$

となり，下記のガンマ分布に従うことがわかる。

$$\lambda(t_j|D_j) \sim Ga\left(cd\Lambda^*(t_j) + d_j, (c+r_j)dt_j\right) \tag{7}$$

事前情報とデータの対立を考慮して客観的事前分布を設定する場合，$\Lambda^*(t_j)$への確信度であるcに十分小さい値を設定すればよい。この場合にc及びcを係数とする項が極めて小さい値をとるものとし，無視可能とする。ここに，確率変数列$t_j\ (j = 1, \dots, J)$をイベントが観察された時刻とするとd_jは必ず 1 を取ることに注意すれば，(6)，(7)式で示される累積ハザード関数の増分及びハザード関数の事後分布はそれぞれ以下の(8)式，(9)式のよ

うに表現できる。

$$d\Lambda(t_j|D_j) \sim Ga(1, r_j) \tag{8}$$

$$\lambda(t_j|D_j) \sim Ga(1, r_j dt_j) \tag{9}$$

さらに，(9)式の尺度パラメータの$r_j dt_j$は，当該区間における総人年に対応している。以上より各区間に対応するJ個の事後分布を解析的に得ることができる。以降，(9)式のハザード関数の事後分布を用いる。

3　区分指数モデルに基づくベイズ流モデル

ここまで，ノンパラメトリックベイズ流モデルに基づく事後分布の導出を試みた。本項では，(9)式で得られるハザード関数の事後分布が，区分指数モデルを適用したベイズ流モデルにおけるハザードの事後分布と密接な関係があることを示す。

被験者iに対するイベントが生起する或いは打切りとなるまでの時間をy_iとし，2 項と同様にJ個の区間，$(0, t_1], (t_1, t_2], \cdots, (t_{J-1}, t_J]$を考える。このとき，各区間のハザードを$\lambda_j\ (j = 1, \ldots, J)$とし，$\lambda_j$の事前分布に以下のガンマ分布を与える。

$$\lambda_j \sim Ga(c\lambda_j^*, c)$$

λ_j^*とcは超パラメータであり，平均がλ_j^*，分散がλ_j^*/cのガンマ分布である。ここに，δ_{ij}を被験者iがj番目の区間でイベント若しくは打切りとなる場合に 1 をとり，それ以外で 0 をとる指示関数とし，v_iを被験者iがイベントの場合に 1 をとり，それ以外で 0 をとる指示関数とする。このとき，j番目の区間におけるハザードは$\lambda_j\ (t \in (t_{j-1}, t_j])$で一定であるから，観測終了が$j$番目のときの累積ハザードは

$$\lambda_j(y_i - t_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(t_g - t_{g-1})$$

とかける。したがって，被験者iの尤度は次式で与えられる。ただし，v_iによりイベントの場合に確率密度関数を，打切りの場合にはその生存関数をそれぞれ尤度に寄与させる操作をまとめて表現している。

$$\prod_{j=1}^{J} \lambda_j^{\delta_{ij} v_i} \left[\exp\left\{ -\delta_{ij} \left(\lambda_j(y_i - t_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(t_g - t_{g-1}) \right) \right\} \right]$$

以上より，区分指数モデルに基づく尤度は次式で与えられる。詳細は，Ibrahim et al.（2001）及び Hobbs et al.（2013）を参照されたい。

$$\prod_{i=1}^{n} \prod_{j=1}^{J} \lambda_j^{\delta_{ij} v_i} \left[\exp\left\{ -\delta_{ij} \left(\lambda_j(y_i - t_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(t_g - t_{g-1}) \right) \right\} \right]$$

また，上式の尤度は以下のように整理できる。

$$\prod_{j=1}^{J} \left[\prod_{i=1}^{n} \lambda_j^{\delta_{ij} v_i} \left[\exp\left\{ -\delta_{ij} \left(\lambda_j(y_i - t_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(t_g - t_{g-1}) \right) \right\} \right] \right]$$

$$= \prod_{j=1}^{J} \left[\lambda_j^{\sum_{i=1}^{n} \delta_{ij} v_i} \left[\exp\left\{ \sum_{i=1}^{n} \left(-\delta_{ij} \left(\lambda_j(y_i - t_{j-1}) + \sum_{g=1}^{j-1} \lambda_g(t_g - t_{g-1}) \right) \right) \right\} \right] \right]$$

ここに，$d_j = \sum \delta_{ij} v_i$を第$j$区間で発生したイベント数とし，

$$y_{ij} = (t_j - t_{j-1}) \sum_{m=j+1}^{J} \delta_{im} + \delta_{ij}(y_i - t_{j-1})$$

とすると，全体の尤度は

$$\prod_{j=1}^{J} \lambda_j^{d_j} \exp\left(-\lambda_j \sum_{i=1}^{n} y_{ij}\right)$$

となる。このとき，j番目の区間における総人年を

$$s_j = \sum_{i=1}^{n} y_{ij}$$

とすると，全体の尤度は

$$\prod_{j=1}^{J} \lambda_j^{d_j} \exp(-\lambda_j s_j) = \exp\left(-\sum_{j=1}^{J} \lambda_j s_j\right) \prod_{j=1}^{J} \lambda_j^{d_j}$$

となる。導出の詳細は付録 1 を，s_jがj番目の区間における総人年であることを補足するためのy_{ij}の別表現については，付録 2 を参照されたい。これはj番目の区間におけるイベント発生件数の強度（期待値）が，当該区間の総人年とハザード（率）の積で表現されたポアソン過程である。ここに，(2)式も参照されたい。以上より，j番目の区間におけるハザードの事後分布$p(\lambda_j|D_j)$は

$$p(\lambda_j|D_j) \propto p(D_j|\lambda_j) p(\lambda_j)$$

$$\propto \lambda_j^{d_j} \exp(-\lambda_j s_j)\, \lambda_j^{c\lambda_j^*-1} \exp(-c\lambda_j) = \lambda_j^{d_j + c\lambda_j^* - 1} \exp(-(c + s_j)\lambda_j)$$

となり，

$$\lambda_j|D_j \sim Ga(c\lambda_j^* + d_j, c + s_j) \tag{10}$$

を得る。(10)式が区分指数モデルに基づくハザードの事後分布である。2.4 項と同様にcに十分小さい値を設定し，J個の区間を観察されたイベント時点で分割することに注意すると，(11)式の事後分布を得る。

$$\lambda_j|D_j \sim Ga(1, s_j) \tag{11}$$

さらに，打切りを含まない場合に限り，s_jは$r_j dt_j$と同値であることに注意すると，打切りを含まない場合にはハザードの事後分布は次式に帰する。これは(9)式の事後分布に等しい。

$$\lambda_j|D_j \sim Ga(1, r_j dt_j)$$

図 2 に，以上の数理について簡易的に 4 例のデータを用いた概要を示す。図 2 では，被験者 3 が区間②で打切りとなっている。このため，区間①と③のハザード関数の事後分布と区分指数モデルに基づくハザードの事後分布は一致するが，区間②はそれぞれ異なる点に注意されたい。

4　マイルストン生存時間解析に基づく意思決定基準

2.4 項で示したハザード関数の事後分布を用いて，生存関数の事後分布を導出する。累積ハザード関数と生存関数の関係性より

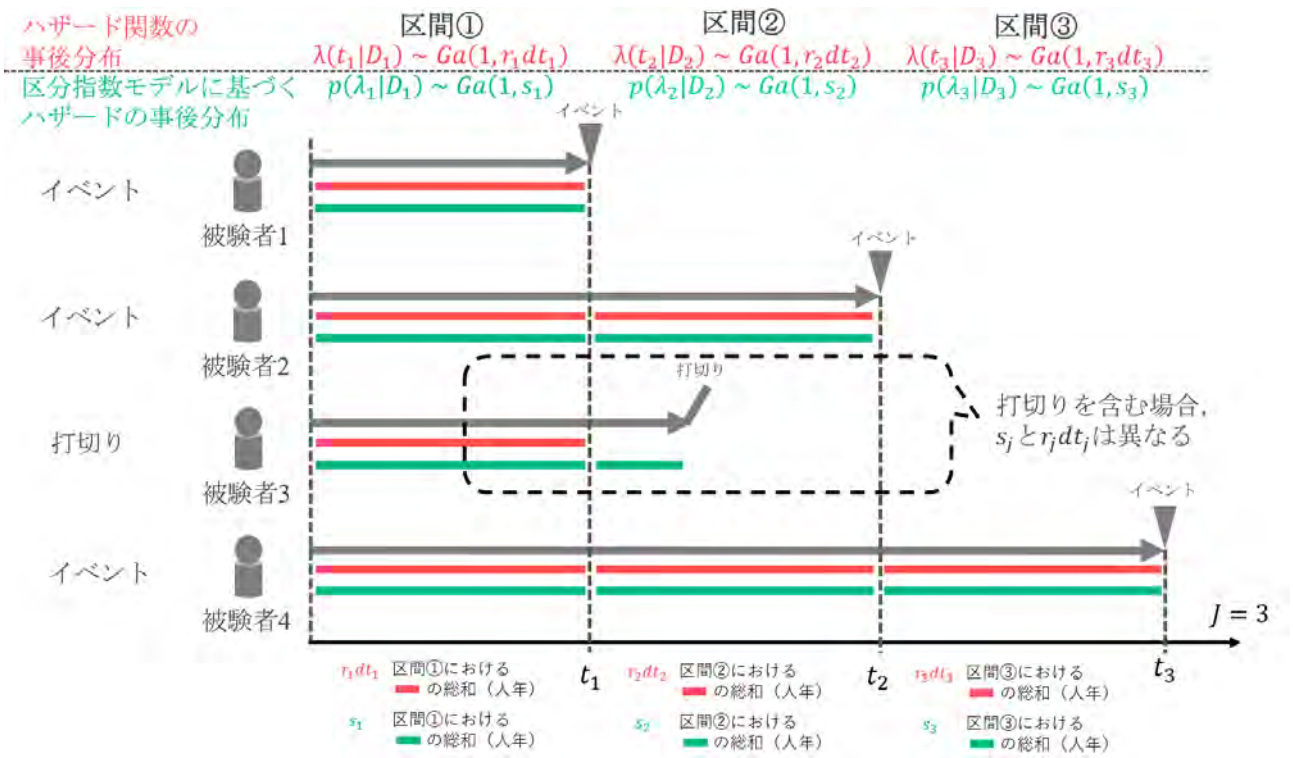


図 2 ハザード関数の事後分布及び区分指数モデルに基づく
ハザードの事後分布の導出過程（総被験者数が 4 例の場合）

$$\mathrm{S}(t_j|D_1,\ldots,D_j) = \exp\left(-\Lambda(t_j|D_1,\ldots,D_j)\right) = \exp\left(-\sum_{g=1}^{j} d\Lambda(t_g|D_g)\right) = \exp\left(-\sum_{g=1}^{j} \lambda(t_g|D_g)dt_g\right) \quad (12)$$

が得られる。(12)式は，(9)式のハザード関数の事後分布から得られるモンテカルロサンプリングを用いて導出する。Kaplan Meier 法に基づく生存関数は，イベントが観察された時点以外ではハザードを 0 とする仮定に基づき，階段上の曲線と推定される（Kalbfleisch and Prentice, 2002）。本稿では，生存関数の事後分布において同様にイベントが観察された時点以外ではハザードを 0 とし，当該事後分布に対して，関心のある評価時点t_{j^*}に対する生存確率の事後分布を得ることとする。このとき，成功の定義として例えば下記の意思決定基準を設定する。

$$\Pr\left(\mathrm{S}(t_{j^*}|D_1,\ldots,D_{j^*}) > \phi_{j^*}\right) > \theta$$

ただし，ϕ_{j^*}は関心のある評価時点j^*における生存確率の目標値であり，θは意思決定のための事後確率の目標値である。また，本稿では$\Pr\left(\mathrm{S}(t_{j^*}|D_1,\ldots,D_{j^*}) > \phi_{j^*}\right)$を成功確率と呼ぶこととする。

5　仮想データに基づく成功確率の算出

4 項までの理論に基づき，仮想データを用いて成功確率を算出する（ファイル名: 01_ data_GP.csv）。本項では，特に 2 項及び 4 項で紹介したノンパラメトリックベイズ流モデルを用いる。本データは，以下の設定に基づいて疑似的に生成された仮想データである。このとき，単群試験における評価時点の成功確率を算出する。

- 登録期間　　：12 カ月
- 観察期間　　：24 カ月
- 参照ハザード　　：$-\log(0.700)/12$
  （12 カ月時点の生存確率が 0.700 における指数分布のハザードに対応）
- 期待ハザード　　：$-\log(0.700)/(12/0.65)$
  （参照ハザードに対してハザード比 0.65 の効果を期待する設定に対応）
- 年次脱落率　　：5%
- 総被験者数　　：50 例
- 評価時点　　：6 カ月，12 カ月，18 カ月，24 カ月
- 各評価時点における生存確率の目標値：0.837，0.700，0.589，0.490
  （各評価時点に対応する参照ハザードの指数分布に基づく生存確率に対応）

具体的なデータの構造を表 1 に示す。ただし，表 1 はデータセットの一部である。このデータは，被験者番号（patient），生存時間（t），イベントの有無（event: イベント 1，打切り 0）の 3 つの変数を有し，総イベント数は 20 件である。

表 1 仮想データ（データセットの一部）

patient	t	event
1	3.31195	1
2	18.69304	1
3	27.72718	0
4	34.93170	0
5	16.47486	1

表 1 の仮想データに対して，$J = 20$の区間についてノンパラメトリックベイズ流モデルに基づいて，J個のハザード関数の事後分布及びモンテカルロ法に基づく生存関数の事後分布を導出する。このための計算過程を以下に示す。

1. 観察されたイベントに基づき，J個の区間を生成する。
2. (9)式に基づき，ハザード関数の事後分布の尺度パラメータに相当する各区間の総人年を計算する。
3. 2 で得られた尺度パラメータを持つガンマ分布に従う疑似乱数を 100,000 個生成し，(12)式からモンテカルロ法に基づく生存関数の事後分布を生成する。
4. 3 で得られた生存関数の事後分布に対して，マイルストン生存時間解析を行う評価時点を特定する。
5. 4 の評価時間における成功確率（生存確率の目標値を超える確率）を計算する。

以下に，ハザード関数及び生存関数の事後分布を導出するための SAS のマクロプログラムを示す。SAS プログラムは本稿の別添として添付（ファイル名: 02_GP_analysis.sas）しているが，プログラムが 2 項の数理を補足することを期待し，ここでは全スクリプトを本文中に掲載する。ただし，前処理として IMPORT プロシジャを用いて data_GP.csv を SAS データセット「data_GP」に読み込ませておく。

```
/*評価時点*/
%let TP1 = 6; /*6カ月*/
%let TP2 = 12; /*12カ月*/
%let TP3 = 18; /*18カ月*/
%let TP4 = 24; /*24カ月*/

/*評価時点における参照生存確率*/
%let REF1 = 0.837; /*6カ月時点*/
%let REF2 = 0.700; /*12カ月時点*/
%let REF3 = 0.589; /*18カ月時点*/
%let REF4 = 0.490; /*24カ月時点*/

%macro post_survival_func;
proc sort data = data_GP; by t; run;
/*----------------------------
1. 観察されたイベントに対するJ個の区間幅の算出
----------------------------*/
data data_int;
          set data_GP(where = (event = 1)) ;
          t2 = lag(t);
          if t2 = . then t2 = 0;
          interval = t - t2;
run;

/*出力結果のマクロ変数化*/
data _null_; set data_int end = _EOF;
    if _EOF then call symputx("neve", _N_); /*総イベント数*/
    call symputx(cats("t",_N_) , t, "L"); /*各イベント時点*/
    call symputx(cats("interval",_N_) , interval, "L"); /*各区間の幅*/
run;

/*----------------------------
2. ハザード関数の事後分布の尺度パラメータに相当する各区間の総人年の計算
----------------------------*/
/* j番目の区間のリスクセットrjの計算*/
data riskset_perinterval;
          set data_GP;
          /*不動小数点の誤差を考慮*/
```

```
        %do i = 1 %to &neve.;
                if (t - &&t&i.) >= -1e-8 then R&i. = 1; else R&i. = 0;
        %end;
run;

/*j番目の区間の人年rjtjの計算*/
data postparam_;
        set riskset_perinterval;
        %do i = 1 %to &neve.;
                scale&i. = &&interval&i.*R&i.;
        %end;
run;

/*各区間の総人年の計算*/
proc means data = postparam_ noprint;
    var scale1-scale&neve.;
    output out = postparam sum = scale1-scale&neve.;
run;

/*尺度パラメータのマクロ変数化*/
data _null_; set postparam;
        %do i = 1 %to &neve.;
                call symputx("scale&i.", scale&i., "L");
        %end;
run;

/*----------------------------
3. モンテカルロ法に基づく生存関数の事後分布の生成
----------------------------*/
data post_survival_;
        call streaminit(4989);
        do n = 1 to 100000;
                s0 = 0;
                %do i = 1 %to &neve.;
                        s&i. = s%eval(&i - 1) + rand('gamma', 1) * (1/&&scale&i.) * &&interval&i.;
                %end;
                output;
        end;
```

```
run;

data post_survival;
        set post_survival_;
        %do i = 1 %to &neve.;
                posts&i. = exp(-s&i.);
        %end;
run;

proc transpose data = post_survival out = post_survival_long;
        by n;
    var posts1-posts&neve.;
run;

proc sort data = post_survival_long; by _NAME_; run;

/*生存関数の事後分布の事後平均, 95%信用区間の算出*/
proc means data = post_survival_long noprint;
    var COL1;
    by _NAME_;
    output out = post_survival_summary mean = p5= p95 = / autoname;
run;

data post_survival_summary;
        set post_survival_summary;
        rename _NAME_ = timepoint COL1_Mean = mean COL1_P5 = P5 COL1_P95 = P95;
        drop _FREQ_ _TYPE_;
run;

/*生存関数の事後分布の描画のためのデータ成形*/
proc sort data = post_survival_summary out = post_survival_summary
         sortseq = linguistic(numeric_collation=on);
    by timepoint;
run;

data post_survival_summary_add;
        if _N_ = 0 then set post_survival_summary; /*元データのメタデータの保持*/
        if _N_ = 1 then do;
```

```
            timepoint = "posts0";
            mean = 1; P5 = 1; P95 = 1;
        row_num = 1;
        output;
        end;
        set post_survival_summary;
        row_num = _N_ + 1;
        output;
run;

data time;
        if _N_ = 0 then set data_int; /*元データのメタデータの保持*/
        if _N_ = 1 then do;
                t = 0;
                row_num = 1;
                output;
        end;
    set data_int;
    row_num = _N_ + 1;
    output;
    keep row_num t t2 interval;
run;

data result;
    merge time post_survival_summary_add;
    by row_num;
    drop row_num;
run;

/*----------------------------
4. マイルストン生存時間解析を行う評価時点の特定
----------------------------*/
data prob;
        set result;
        if t2 < &TP1. && (t2 + interval) > &TP1. then FL&TP1. = 1;
        else if t2 < &TP2. && (t2 + interval) > &TP2. then FL&TP2. = 1;
        else if t2 < &TP3. && (t2 + interval) > &TP3. then FL&TP3. = 1;
        else if t2 < &TP4. && (t2 + interval) > &TP4. then FL&TP4. = 1;
```

```
run;

%do i = 1 %to 4;
data _null_;
        set prob;
        where FL&&TP&i. = 1;
        call symputx("milestone&i." , timepoint, "G");
run;
%end;

/*----------------------------
5. 評価時点における生存確率の目標値を超える成功確率の計算
----------------------------*/
data prob_success;
        set post_survival(keep = &milestone1. &milestone2. &milestone3. &milestone4.);
        if &milestone1. > &REF1. then FL&TP1. = 1; else FL&TP1. = 0;
        if &milestone2. > &REF2. then FL&TP2. = 1; else FL&TP2. = 0;
        if &milestone3. > &REF3. then FL&TP3. = 1; else FL&TP3. = 0;
        if &milestone4. > &REF4. then FL&TP4. = 1; else FL&TP4. = 0;
run;
%mend;

%post_survival_func;

/*表2: ハザード関数の事後分布の尺度パラメータの出力*/
proc print data = postparam(drop = _TYPE_ _FREQ_);
```

以上より，得られたハザード関数の事後分布を表 2 に示す。同様に，モンテカルロ法を用いて評価時点における生存確率の事後分布を導出した結果を図 3 に示す。生存関数は，観察されたイベント時点までを描画している。なお，SGPLOT プロシジャを用いた作図の詳細は，大橋ほか（2016）を参照されたい。また，各評価時点の生存関数の事後分布を描画するための SAS プログラムを以下に示し，その出力結果を図 4 に示す。なお，SAS プログラムは 6 カ月時点の生存関数に対応するスクリプトのみを示している。ここでは，イベントが観察された時点以外ではハザードを 0 とする仮定を前提として，各評価時点を含む区間の生存時間の事後分布に注目している。なお，ヒストグラムに対してカーネル密度推定により平滑化した結果も同時描画している。また，各評価時点における生存確率の目標値を点線で示した。

表 2 ハザード関数の事後分布

区間	事後分布	区間	事後分布
(0,0.28]	$Ga(1,14.132)$	(10.25,16.47]	$Ga(1,236.436)$
(0.28,2.87]	$Ga(1,124.128)$	(16.47,17.34]	$Ga(1,31.918)$
(2.87,3.16]	$Ga(1,13.815)$	(17.34,18.53]	$Ga(1,42.796)$
(3.16,3.31]	$Ga(1,6.871)$	(18.53,18.69]	$Ga(1,5.836)$
(3.16,4.22]	$Ga(1,41.030)$	(18.69,21.83]	$Ga(1,103.603)$
(4.22,7.88]	$Ga(1,160.858)$	(21.83,22.71]	$Ga(1,28.315)$
(7.88,7.93]	$Ga(1,2.219)$	(22.71,22.89]	$Ga(1,5.493)$
(7.93,9.63]	$Ga(1,71.184)$	(22.89,23.94]	$Ga(1,31.495)$
(9.63,9.90]	$Ga(1,11.257)$	(23.94,28.41]	$Ga(1,62.565)$
(9.90,10.25]	$Ga(1,14.091)$	(28.41,33.75]	$Ga(1,26.693)$

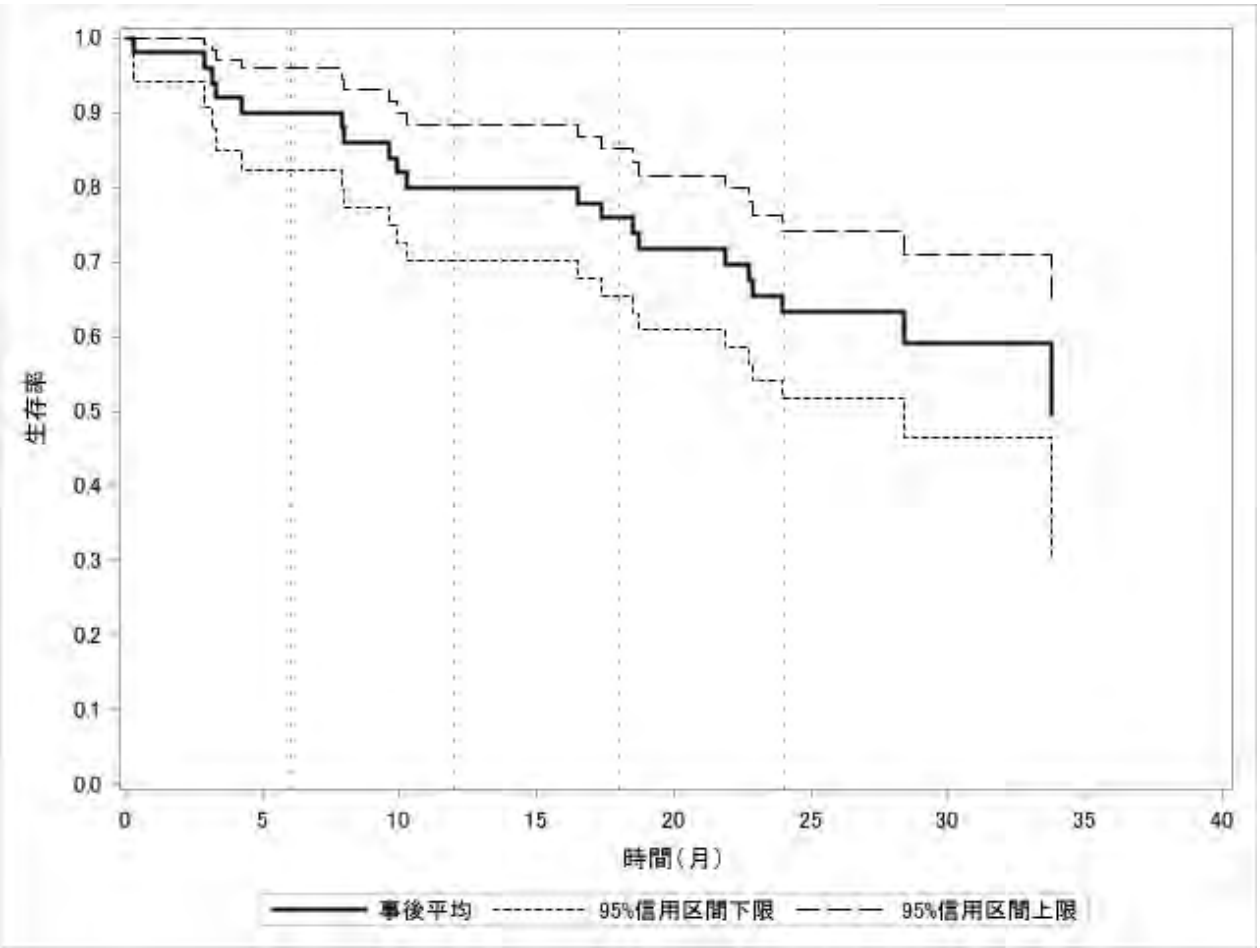


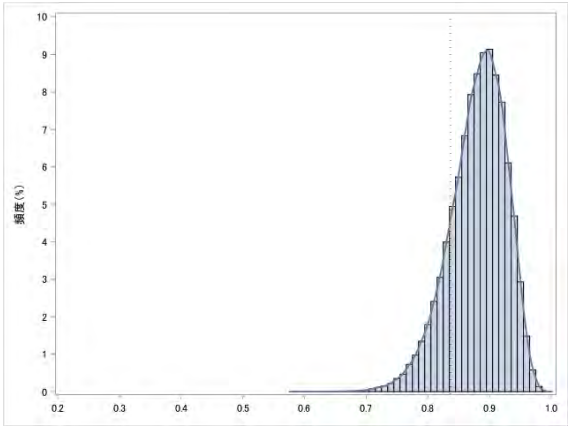
図 3 生存関数の事後分布

```
/*図3: 生存関数の事後分布の描画*/
proc sgplot data = result;
        step x = t y = mean  / lineattrs = (thickness = 2px color = black) legendlabel='事後平均';
```

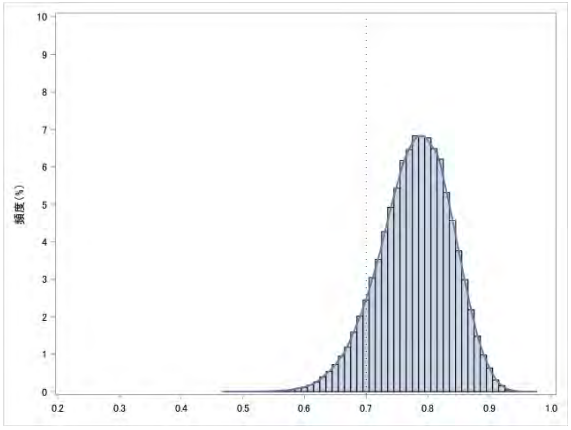
```
        step x = t y = P5 / lineattrs = (color = black pattern = 3) legendlabel='95%信用区間下限';
        step x = t y = P95 / lineattrs = (color = black pattern = 4) legendlabel='95%信用区間上限';
        refline &TP1. &TP2. &TP3. &TP4. / axis = x lineattrs = (color = black pattern = 35);
        yaxis values = (0 to 1 by 0.1) label = '生存率';
        xaxis values = (0 to 40 by 5) label = '時間（月）';
run;

/*図4(a): 6カ月時点生存割合の事後分布の描画*/
proc sgplot data = post_survival noautolegend;
    histogram &milestone1. / binwidth = 0.01;
    density &milestone1. / type = kernel;
    refline &REF1. / axis = x lineattrs = (color = black pattern = 35);
    yaxis values = (0 to 10 by 1) label = '頻度（%）';
        xaxis display = (nolabel) values = (0.2 to 1.0 by 0.1) label = "";
run;
```

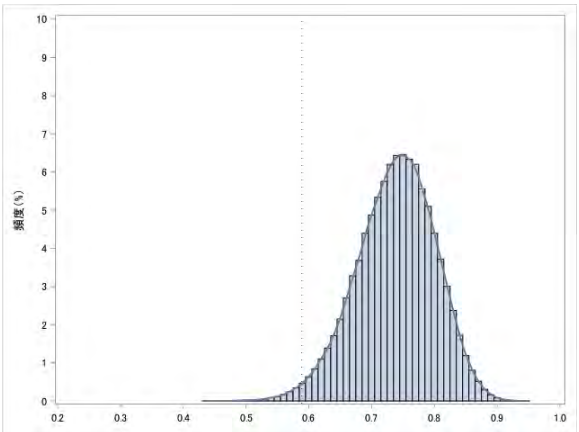
(a) 6カ月生存割合



(b) 12カ月生存割合



(c) 18カ月生存割合



(d) 24カ月生存割合

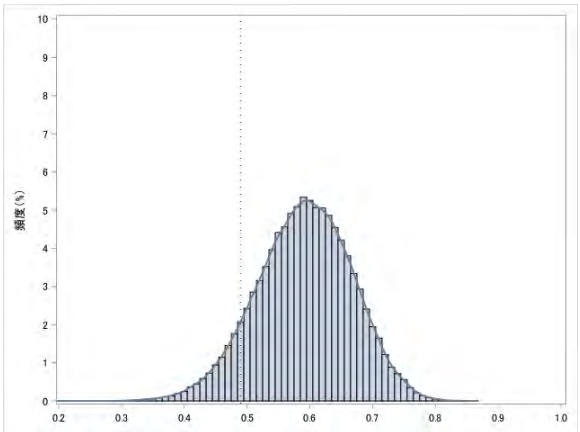


図 4 各評価時点における生存確率の事後分布

最後に，期待ハザード及び評価時点における成功確率を算出する SAS プログラムを下記に示す。また，表 3 にその結果を示す。本項で用いた仮想データは，time to event 評価項目の確率分布は指数分布に従うものとして生成している。このため，より早期の評価時点においては相対的に生存確率の目標値に対して差を検出しにくく，6 カ月時点の成功確率は比較的低い結果が得られた。

```
/*表3: 各評価時点における成功確率の計算*/
proc freq data = prob_success;
        table FL&TP1. FL&TP2. FL&TP3. FL&TP4.;
run;
```

表 3 各評価時点における成功確率

	評価時点			
	6カ月	12カ月	18カ月	24カ月
成功確率	83.33%	90.68%	98.74%	90.67%

6 意思決定基準の性能評価

5 項で検討した意思決定基準について，モンテカルロ・シミュレーションにより頻度論的動作特性を評価する。疑似データの生成には，5 項で示した仮想データの設定値を踏襲する。ここに，動作特性の評価に必要となる追加の設定を示す。ただし，time to event 型アウトカムの確率分布は指数分布とし，そのパラメータには下記の期待ハザードを真値として 1 つのみを取る分布を考える。すなわち，デザイン事前分布に退化分布を与えることとし，事前予測分布は指数分布に従うものとする。

- 期待ハザード　　：$-\log(0.70)/(12/0.65)$（対立仮説），$-\log(0.70)/12$（帰無仮説）
- 評価項目　　：12 カ月時点の生存割合（評価時点を 12 カ月とする）
- 事後確率の目標値　　：0.80
- 評価例数　　：50 例，60 例，70 例，80 例
- シミュレーション回数　：10,000 回

デザイン事前分布については手良向（2023），ベイズ流試験デザインにおけるシミュレーションの利用法については，製薬協報告書（2025）を参照されたい。ここでは，以上の設定を踏まえて試験の成功を「12 カ月時点の生存関数の事後分布が閾値生存割合である 0.700 を超える確率が 80%以上であること」と定義する。なお，12 カ月時点の閾値生存割合は，参照ハザードである$-\log(0.70)/12$をハザードに持つ指数分布から算出される（すなわち，70%）。以上の設定の下で，指定したシミュレーション回数の臨床試験を仮想的に実施し，得られた結果に基づいて第一種の過誤確率，検出力を評価する。特に，第一種の過誤確率は帰無仮説下で成功の定義を満たす確率，検出力は対立仮説下で成功の定義を満たす確率と定義する。表 4 に頻度論的動作特性の結果を示す。例えば，70 例規模の試験を計画できれば，設定した意思決定基準は第一種の過誤確率を 20%以下，検出力を 80%以上確保できる。第 II 相試験の位置づけを踏まえ，ある程度の第一種の過誤確率を許容できるものとすれば，高い検出力を有することから設定に足る基準であると考えられる。

表4 頻度論的動作特性の結果

例数	帰無仮説下 （第一種の過誤確率: %）	対立仮説下 （検出力: %）
50例	16.18%	69.56%
60例	18.23%	77.74%
70例	18.35%	81.56%
80例	17.78%	83.82%

表 4 を得るための SAS のマクロプログラムは，別添（ファイル名: 03_GP_OC_simulation.sas）を参照されたい。5 項で示したプログラムを繰り返し適用することで極めてシンプルに記述できることから，本文中の掲載は控え，考え方のみの提示に留める。具体的には，以下の手順に従って設定ごとに繰り返し処理を実行している。

1. 疑似乱数を用いて単群臨床試験データを生成する。
2. 5項で示した手順に基づき，成功確率を算出する。なお，ここでは生存関数の事後分布の導出のためのモンテカルロサンプルは10,000個とする。
3. 2で得られた成功確率に対して，成功の定義を満たすか否かを判断し，結果をフラグ変数に格納する。
4. 1～3 を 10,000 回繰り返し，結果フラグを頻度集計する。

7 まとめ

ノンパラメトリックベイズ流モデルに基づく生存時間解析として，確率過程を用いて分布関数を表現し，得られた累積ハザード関数の増分に対する事後分布からハザード関数の事後分布を導出した。具体的には，時間軸をイベントが発現した時点で分割し，各区間の累積ハザード関数がガンマ過程に従うものとして事後分布を推定した。これらの解析方法は，Hobbs and Carlin（2008）及び矢田，浜田（2016）に詳しい。さらに，ハザード関数の事後分布は，打切りが存在しない，かつ事前情報を無視可能とする状況下において，区分指数モデルに基づくハザードの事後分布に帰することも示した。なお，矢田，浜田（2016）でも言及しているとおり，ノンパラメトリックとはパラメータを置かないことではなく，無限個のパラメータを持つという意味である点に注意したい。一方で，得られるデータは有限であるため，実際の解析にあたっては有限個のパラメータに対して事後分布を計算することができる。

本稿では，以上のノンパラメトリックベイズ流モデルを踏まえ，推定した生存関数の事後分布を用いてマイルストン生存時間解析を行う方法を紹介した。しかしながら，本稿で取り扱ったような単群臨床試験において time to event 型アウトカムを評価指標とする点については，薬剤が奏効した場合に分子に寄与させる奏効割合とは異なり，腫瘍の増悪や死亡等をイベントとする特性上，集団の選択バイアスの影響をより強く受ける等の注意すべき課題もある。例えば，matching adjusted indirect comparison 法（Signorovitch, et al., 2010）に代表される母集団調整法により，観測された重要な予後因子の分布を調整する必要もあるかもしれない。また，マイルストン生存時間においては，Chen（2015）で言及されているとおり，評価時点の決定には極めて慎重な議論が必要であり，少なくとも最終登録例が設定した評価時点まで観察できるよう十分なフォロー

アップ期間を設定する等のデザイン上の工夫は必須である。一方で，マイルストン生存時間解析は event-driven ではなく事前に定めた評価時点で解析できることから，臨床開発計画全体の中間を担う第 II 相試験の位置づけを踏まえると，この点は魅力的であろう。以上より，考慮すべき課題は残るものの，特に従来の奏効割合での評価が困難である場合等の開発が複雑化する領域において，本稿で紹介した単群試験におけるマイルストン生存時間解析に基づく意思決定は，後続の第 III 相試験の実施可否の検討に資すると考える。

最後に，本稿がノンパラメトリックベイズ流モデルに基づくマイルストン生存時間解析の理解及び SAS による解析の一助になれば幸いである。

謝辞

本稿の SAS プログラムの作成にあたり，ご助言いただいた中外製薬株式会社の生井伴幸氏に感謝申し上げます。

参考文献

[1] Chen, T. T. (2015). Milestone Survival: A Potential Intermediate Endpoint for Immune Checkpoint Inhibitors. *JNCI: Journal of the National Cancer Institute*, **107**(9).

[2] Clayton, D. (1988), The analysis of event history data: A review of progress and outstanding problems. *Statistics in Medicine*, **7**, 819–841.

[3] Clayton, D. G. (1991). A Monte Carlo Method for Bayesian Inference in Frailty Models. *Biometrics*, **47**(2), 467–485.

[4] Green, S., Benedetti, J. and Smith, A. et al. [著]．JCOG データセンター[訳]．米国 SWOG に学ぶ がん臨床試験の実践 -第 2 版-．医学書院．2013．

[5] Ibrahim, J. G., Chen, M. H. and Sinha, D. (2001). Bayesian Survival Analysis. Springer-Verlag, New York.

[6] Hobbs, B. P. and Carlin, B, P. (2008). Practical Bayesian Design and Analysis for Drug and Device Clinical Trials. *Journal of Biopharmaceutical Statistics*, **18**(1), 54–80.

[7] Hobbs, B. P., Carlin, B, P. and Sargent D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials*, 10, 430–440.

[8] Kalbfleisch, J. D. (1978). Non-Parametric Bayesian Analysis of Survival Time Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **40**(2), 214–221.

[9] Kalbfleisch, J. D. and Prentice, L. R. (2002). The Statistical Analysis of Failure Time Data. Wiley, New Jersey.

[10] Signorovitch, J. E., Wu, E. Q. and Yu, A. P. (2010). Comparative Effectiveness Without Head-to-Head Trials A Method for Matching-Adjusted Indirect Comparisons Applied to Psoriasis Treatment with Adalimumab or Etanercept. *Pharmacoeconomics*, **28**, 935–945.

[11] 大橋靖雄，浜田知久馬，魚住龍史．（2016）．生存時間解析 応用編 SAS による生物統計．東京大学出版会．

[12] 手良向聡．（2023）．ベイズ流臨床試験における標本サイズ設定：2 つの事前分布を用いた推論上の性能に基づく接近法．計量生物学，**44**(1)，35–51．

[13] 中村剛．（2001）．Cox 比例ハザードモデル．朝倉図書．

[14] 日本製薬工業協会 医薬品評価委員会 データサイエンス部会.（2025）．ベイズ流試験デザインにおけるシミュレーションの利用法.

[15] 矢田真城，浜田知久馬.（2016）．ガンマ過程を用いた比例ハザードモデルによる解析．SAS ユーザー総会 2016 論文集.

付録1

3 項で与えた区分指数モデルに基づくベイズ流モデルにおける全体の尤度の導出の詳細を示す。

$$\prod_{i=1}^{n}\prod_{j=1}^{J}\lambda_j^{\delta_{ij}v_i}\left[\exp\left\{-\delta_{ij}\left(\lambda_j(y_i-t_{j-1})+\sum_{g=1}^{j-1}\lambda_g(t_g-t_{g-1})\right)\right\}\right]$$

$$=\left(\prod_{i=1}^{n}\prod_{j=1}^{J}\lambda_j^{\delta_{ij}v_i}\right)\prod_{i=1}^{n}\prod_{j=1}^{J}\left[\exp\left\{-\delta_{ij}\left(\lambda_j(y_i-t_{j-1})+\sum_{g=1}^{j-1}\lambda_g(t_g-t_{g-1})\right)\right\}\right]$$

$$=\left(\prod_{j=1}^{J}\lambda_j^{d_j}\right)\exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{J}\delta_{ij}\left(\lambda_j(y_i-t_{j-1})+\sum_{g=1}^{j-1}\lambda_g(t_g-t_{g-1})\right)\right\}$$

$$=\prod_{j=1}^{J}\lambda_j^{d_j}\exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{J}\delta_{ij}\left(\lambda_j(y_i-t_{j-1})\right)-\sum_{i=1}^{n}\sum_{j=1}^{J}\delta_{ij}\left(\sum_{g=1}^{j-1}\lambda_g(t_g-t_{g-1})\right)\right\}$$

$$=\prod_{j=1}^{J}\lambda_j^{d_j}\exp\left\{-\sum_{i=1}^{n}\sum_{j=1}^{J}\delta_{ij}\left(\lambda_j(y_i-t_{j-1})\right)-\sum_{i=1}^{n}\sum_{j=1}^{J}\delta_{ij}\left(\sum_{g=1}^{J}\lambda_g(t_g-t_{g-1})I(g\le j-1)\right)\right\}$$

$$=\prod_{j=1}^{J}\lambda_j^{d_j}\exp\left\{-\sum_{j=1}^{J}\lambda_j\left(\sum_{i=1}^{n}\delta_{ij}(y_i-t_{j-1})\right)-\sum_{g=1}^{J}\lambda_g(t_g-t_{g-1})\left(\sum_{i=1}^{n}\sum_{j=1}^{J}\delta_{ij}I(g\le j-1)\right)\right\}$$

$$=\prod_{j=1}^{J}\lambda_j^{d_j}\exp\left\{-\sum_{j=1}^{J}\lambda_j\left(\sum_{i=1}^{n}\delta_{ij}(y_i-t_{j-1})\right)-\sum_{g=1}^{J}\lambda_g(t_g-t_{g-1})\left(\sum_{i=1}^{n}\sum_{j=g+1}^{J}\delta_{ij}\right)\right\}$$

$$=\prod_{j=1}^{J}\lambda_j^{d_j}\exp\left\{-\sum_{j=1}^{J}\lambda_j\left(\sum_{i=1}^{n}\delta_{ij}(y_i-t_{j-1})\right)-\sum_{j=1}^{J}\lambda_j(t_j-t_{j-1})\left(\sum_{i=1}^{n}\sum_{m=j+1}^{J}\delta_{im}\right)\right\}$$

$$=\prod_{j=1}^{J}\lambda_j^{d_j}\exp\left\{-\sum_{j=1}^{J}\lambda_j\left((t_j-t_{j-1})\left(\sum_{i=1}^{n}\sum_{m=j+1}^{J}\delta_{im}\right)+\sum_{i=1}^{n}\delta_{ij}(y_i-t_{j-1})\right)\right\}$$

$$=\prod_{j=1}^{J}\lambda_j^{d_j}\exp\left\{-\sum_{j=1}^{J}\lambda_j\sum_{i=1}^{n}\left((t_j-t_{j-1})\sum_{m=j+1}^{J}\delta_{im}+\delta_{ij}(y_i-t_{j-1})\right)\right\}$$

$$= \left(\prod_{j=1}^{J} \lambda_j^{d_j} \right) \exp\left\{ -\sum_{j=1}^{J} \lambda_j \sum_{i=1}^{n} y_{ij} \right\}$$

$$= \left(\prod_{j=1}^{J} \lambda_j^{d_j} \right) \exp\left\{ -\sum_{j=1}^{J} \lambda_j s_j \right\}$$

$$= \left(\prod_{j=1}^{J} \lambda_j^{d_j} \right) \left(\prod_{j=1}^{J} \exp\{-\lambda_j s_j\} \right)$$

$$= \prod_{j=1}^{J} \lambda_j^{d_j} \exp\{-\lambda_j s_j\}$$

付録2

s_jがj番目の区間における総人年であることを補足するため，y_{ij}の別表現について考える。具体的には，下記の表現が同値であることを示す。

$$y_{ij} = (t_j - t_{j-1}) \sum_{m=j+1}^{J} \delta_{im} + \delta_{ij}(y_i - t_{j-1}) \qquad (A1)$$

$$y_{ij} = \max\{0, \min(y_i, t_j) - t_{j-1}\} \qquad (A2)$$

ここで，

$$\delta_{ik} = I(y_i \in (t_{k-1}, t_k]), \qquad \sum_{k=1}^{J} \delta_{ik} = 1$$

いま，被験者iの観察終了時間がk番目の区間にあるとする。すなわち，

$$\delta_{ik} = 1, \qquad \delta_{im} = 0 \ (m \neq k), \qquad y_i \in (t_{k-1}, t_k]$$

である状況を考える。このとき，(A1)式を指示関数を用いて表現することを考える。δ_{im}は$m = k$のとき1であるので，

$$\sum_{m=j+1}^{J} \delta_{im} = \begin{cases} 1, & (k > j) \\ 0, & (k \leq j) \end{cases}$$

より指示関数を用いて

$$\sum_{m=j+1}^{J} \delta_{im} = I(k > j)$$

とかける。また，

$$\delta_{ij} = \begin{cases} 1, & (j = k) \\ 0, & (j \neq k) \end{cases}$$

であるから，同様に

$$\delta_{ij} = I(j = k)$$

とかける。以上より，(A1)式は次式と同値である。

$$y_{ij} = (t_j - t_{j-1})I(k > j) + (y_i - t_{j-1})I(j = k) \qquad (A1')$$

以降，jとkの大小で分けて(A1')式と(A2)式が同値であることを示す。

- $k < j$の場合: このとき，$I(k > j) = 0$，$I(j = k) = 0$であるから，(A1')式は

  $$y_{ij} = (t_j - t_{j-1})I(k > j) + (y_i - t_{j-1})I(j = k) = 0$$

  となり，$y_i \leq t_{j-1}$であるから(A2)式についても

  $$y_{ij} = \max\{0, \min(y_i, t_j) - t_{j-1}\} = 0$$

  となる。以上より，$k < j$の場合に(A1')式と(A2)式は同値である。

- $k = j$の場合: このとき，$I(k > j) = 0$，$I(j = k) = 1$であるから，(A1')式は

  $$y_{ij} = (t_j - t_{j-1})I(k > j) + (y_i - t_{j-1})I(j = k) = y_i - t_{j-1}$$

  となる。また，$y_i \in (t_{j-1}, t_j]$であるから，(A2)式は

  $$y_{ij} = \max\{0, \min(y_i, t_j) - t_{j-1}\} = y_i - t_{j-1}$$

  となるため，$k = j$の場合においても(A1')式と(A2)式は同値である。

- $k > j$の場合: このとき，$I(k > j) = 1$，$I(j = k) = 0$であるから，(A1')式は

  $$y_{ij} = (t_j - t_{j-1})I(k > j) + (y_i - t_{j-1})I(j = k) = t_j - t_{j-1}$$

  である。また，$y_i > t_j$であるから，(A2)式は

  $$y_{ij} = \max\{0, \min(y_i, t_j) - t_{j-1}\} = t_j - t_{j-1}$$

  となるため，$k > j$の場合においても(A1')式と(A2)式は同値である。

以上より，jとkのすべての大小関係において(A1')式と(A2)式は同値である。したがって，すべての被験者に対するy_{ij}の総和をs_jとするとき，(A2)式からs_jはj番目の区間における総人年を表すことがわかる。

多次元尺度構成法による臨床データのマッピングと適合度評価

高田 浩成

（イーピーエス株式会社）

Mapping and Goodness-of-Fit Assessment of Clinical Data Using Multidimensional Scaling

Hiroshige Takata
EPS Corporation

要旨

多次元尺度構成法(MDS)は、高次元データの対象間の非類似度を低次元空間に配置して表現する技法である。元来は社会科学で活用され、近年は自然科学において複雑構造の視覚化のためにも再注目されている。非類似度を距離とし、要素ごとの集合から距離行列を構成し、次元削減を行う。

本稿では、臨床試験を想定したテストデータを用いて MDS の適用例を示す。各領域の情報について、複数変量からなるデータを分析対象とし、目的別に座標配置を視覚化する。SAS においては、DISTANCE プロシジャにより距離行列に変換し、それを MDS プロシジャによりマッピングすることになる。なお、連続値データに対してはユークリッド距離、二値データに対してはジャッカード距離を適用し、目的やデータ構造によって計量／非計量の処理を使い分ける。

さらに、適合度評価においても SAS は有効な機能をもつ。実装理論は数学的な古典的 MDS ではなく、反復法による数値最適化手法の計量 MDS／非計量 MDS となる。ストレス関数の初期化・評価・更新を繰り返すことにより座標の配置を改善し、この過程での次元とストレス値の推移を観測し、モデルの収束状況や適合度も評価する。

キーワード：計量 MDS、非計量 MDS、ユークリッド距離、ジャッカード距離、ストレス関数、ストレス値

1. 背景

多次元尺度構成法(Multidimensional Scaling: MDS)は、対象間の距離や非類似度を低次元空間上に可視化する手法として、社会科学分野で発展してきた。その後、計算機性能やアルゴリズムの発展に伴い、近年では自然科学分野にも適用範囲が拡大している。とくに遺伝距離や化学構造など、対象間の距離情報や構造類似度を扱う様々な分野で利用されている。

MDS は距離行列または非類似度行列を入力データとして利用し、対象間の関係性を低次元空間上に表現することを目的とする。この特徴により、変数値そのものだけでは把握しにくい対象間の類似性や関連性を直感的に可視化することが可能である。多変量データの可視化手法としては主成分分析(Principal Component

Analysis: PCA)も広く利用されているが、PCA が変数の分散構造に着目するのに対し、MDS は距離構造そのものを可視化できる点に特徴がある。PCA と MDS の主な違いを Table 1 に示す。

	主成分分析	多次元尺度構成法
入力	変数値行列	距離・非類似度行列
主目的	分散の要約	距離構造の可視化
次元選択	寄与率	ストレス値
出力	主成分得点	マッピング座標

Table 1　次元削減としての PCA と MDS の比較

なお、医療分野では選好評価を目的としたコンジョイント分析も用いられるが、本研究では対象間の距離構造を可視化できる MDS に着目した。

2. 手法

2.1 距離尺度と距離行列

距離尺度は、とくに本研究で用いるユークリッド距離(Euclidean distance)およびジャッカード距離(Jaccard distance)について説明する。

ユークリッド距離は多変量空間における 2 対象間の直線距離を表し、連続値データに対する代表的な距離尺度である。n を変数の数、x_i を対象 x の i 番目変数値、y_i を対象 y の i 番目変数値として Equation 2.1.1 の式で示される。

$$d = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Equation 2.1.1　ユークリッド距離の計算式

ジャッカード距離の計算式について Equation 2.1.2 に示す[1]。ジャッカード係数(Jaccard coefficient)は、存在･非存在を表す二値データに対する代表的な非類似度指標である。対象 A および B を要素の集合とすると、ジャッカード係数は両集合の共通部分の要素数を和集合の要素数で除した値として定義される(a)。ここで、A∩B は両集合の共通部分、A∪B は両集合の和集合、|･|は集合の要素数を表す。ジャッカード距離は、ジャッカード係数から次式により求められる(b)。

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

(a)ジャッカード係数

$$d(A,B) = 1 - J(A,B)$$

(b)ジャッカード距離

Equation 2.1.2　ジャッカード係数によるジャッカード距離の計算式

ジャッカード係数は、両対象がともに存在しない状態(0, 0)を類似度計算に含めない特徴を有する。そのため、存在情報そのものに着目したい場合や、非存在データが多数を占める疎な二値データの解析に適している。一方、少なくとも一方が存在する組合せ(1, 0)、(0, 1)、(1, 1)のみが類似度計算に寄与する。

000095

MDS では、対象間の距離または非類似度を表す行列を入力として用いる。対象数を n とすると、距離行列 D は Equation 2.1.3 で表される。

$$D = \begin{pmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1n} \\ d_{21} & 0 & d_{23} & & d_{2n} \\ d_{31} & d_{32} & 0 & & d_{3n} \\ \vdots & & & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \cdots & 0 \end{pmatrix}$$

Equation 2.1.3　距離行列式

ここで d_{ij} は対象 i と対象 j の距離を表す。距離行列は一般に対称行列であり、$d_{ij}=d_{ji}$ が成り立つ。また、対象自身との距離は 0 であるため、$d_{ii}=0$ となる。

一方、対象間距離が既に距離行列または非類似度行列の形式で与えられている場合には、距離変換を行わずにそのまま MDS の入力として利用することが可能である。

2.2 多次元尺度構成法の種類

MDS は、対象間の距離または非類似度情報を低次元空間上に表現する解析手法であり、古典的 MDS(Classical MDS)、計量 MDS(Metric MDS)および非計量 MDS(Nonmetric MDS)に大別される(Table 2.2.1)[2]。

古典的 MDS は Torgerson (1958)により体系化され、距離行列に対して二重中心化(double centering)を適用し、固有値分解(eigenvalue decomposition)によって低次元空間上の座標を推定する手法である。一方、計量 MDS および非計量 MDS では、元の距離構造と低次元空間上の距離構造との乖離を表すストレス関数(stress function)を最小化することにより配置座標を求める。そのため、古典的 MDS とは異なり反復計算を必要とし、解の適合度をストレス値(stress value)によって評価できる特徴を有する。計量 MDS は距離そのものを保持することを目的とするのに対し、非計量 MDS は距離の順位関係を保持することを目的とする。

	古典的 MDS	**計量 MDS**	**非計量 MDS**
入力	距離行列	距離行列	非類似度行列
保持情報	距離構造	距離値	距離順位
計算方法	二重中心化・固有値分解	ストレス最小化	ストレス最小化
反復計算	不要	必要	必要
適合度評価	固有値	ストレス値	ストレス値

Table 2.2.1　多次元尺度構成法の種類の比較

古典的 MDS は、対象間距離から得られた距離行列 D に対して二重中心化を適用し、内積行列へ変換した後に固有値分解を行うことで低次元空間上の座標を推定する手法である(Equation 2.2.1)[3]。古典的 MDS は固有値分解に基づくため反復計算を必要とせず、PCA と数学的に近い性質を有する。対象 i および j の距離 d_{ij} と内積 $\mathbf{x}_i^T\mathbf{x}_j$ の間には(a)の関係が成り立つ。この関係を利用して距離行列から内積行列を求める操作を二重中心化という。さらに、(b)に示す二重中心化により得られた内積行列に対して、(c)のように固有値分解を適用することで低次元空間上の配置座標を推定する。二重中心化により得られた内積行列 B に対して固有値分解を行い、固有ベクトル行列 U および固有値対角行列 $U\Lambda U^T$ を求める。さらに、(d)に示すように固有ベクトルと固有値の平方根を用いて配置座標行列 X を算出する。古典的 MDS では、上位の固有値に対応する固有ベ

クトルを選択することで、元の距離構造を保持した低次元空間上の座標を得ることができる。

$$d_{ij}^2 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2 = \|\boldsymbol{x}_i\|^2 + \|\boldsymbol{x}_j\|^2 - 2\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$$

$$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)^{\mathrm{T}}$$

(a)距離と内積の関係

$$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} d_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} d_{ij}^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^2 \right)$$

$$\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}} = -\frac{1}{2}\left(I_n - \frac{1}{n}J_n\right) D \left(I_n - \frac{1}{n}J_n\right)$$

(b)二重中心化の計算過程

$$|A - \lambda I| = 0$$

$$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$$

(c)固有値分解による座標推定

$$B = XX^{\mathrm{T}} = U\Lambda U^{\mathrm{T}}$$

$$X = U\Lambda^{\frac{1}{2}}$$

(d)エッカート・ヤング分解

Equation 2.2.1　古典的 MDS における二重中心化・固有値分解

古典的 MDS が二重中心化および固有値分解により座標を推定するのに対し、計量 MDS および非計量 MDS では反復最適化法が用いられる。代表的なアルゴリズムとして SMACOF(Scaling by MAjorizing a COmplicated Function)が知られている[4]。SMACOF は、初期配置から開始し、各反復において対象の座標を更新することでストレス関数を最小化し、元の距離構造を低次元空間上にできるだけ忠実に再現する座標配置を求める。

MDS 解の適合度はストレス値により評価される。ストレス値は値が小さいほど適合度が高いことを示す。計量 MDS におけるストレス関数を Equation 2.2.2 に示す。

$$S = \sqrt{\frac{\sum_{i \neq j}\left(\tilde{d}_{ij} - d_{ij}\right)^2}{\sum_{i \neq j} d_{ij}^2}}$$

Equation 2.2.2　計量 MDS におけるストレス関数

ここで、d_{ij} は元の距離、$\hat{d}_{ij}$ は低次元空間上で再現された距離、S はストレス値を表す。SMACOF では、初期配置から配置後距離 $\hat{d}_{ij}$ を計算し、ストレス値を評価した後、ストレスが減少するよう座標を更新する。この処理で、収束条件を満たすまで繰り返すことで最終解を得る。SMACOF による反復最適化の概略を Table 2.2.2 に示す。

Step	処理	出力
1	元の距離を入力	d_{ij}
2	低次元空間上に初期配置	初期配置

3*	配置後距離を計算	$\hat{d}_{ij}^{(t)}$
4*	ストレス関数を計算	$S^{(t)}$
5*	ストレス値が減少するよう配置を更新	$\hat{d}_{ij}^{(t+1)}, S^{(t+1)}$
6	収束まで Step 3 から Step 5 を反復	最終配置、最終ストレス値

*：反復 Step、t：反復回数

Table 2.2.2　SMACOF アルゴリズムによる反復最適化の手順

さらに、非計量 MDS では、元の距離に対して順位情報を保持する単調変換を施した後、同様の考え方に基づいてストレスを最小化する。非計量 MDS にはクラスカル法（Kruskal's nonmetric MDS）を用いる。クラスカル法では、元の距離そのものではなく距離の順位関係を保持することを目的とし、単調増加変換を適用した後の disparity と低次元空間上の距離との乖離が最小となるよう配置を推定する。適合度はストレス関数により評価し、ストレス値が小さいほど元の距離構造が良好に再現されていると解釈した。

MDS 解の適合度は ストレス値により評価した。ストレス値は、元の距離構造と低次元空間上で再現された距離構造との乖離を表す指標であり、値が小さいほど適合度が高いことを示す。一般に、ストレス値が 0.05 未満の場合は Excellent、0.05 以上 0.10 未満の場合は Good、0.10 以上 0.20 未満の場合は Fair、0.20 以上の場合は Poor と評価される(Table 2.2.3)[5]。この基準を参考に各次元数における MDS 解の適合度を評価する。本研究では、各次元数において算出されたストレス値を用いて適合度を評価し、マッピングの解釈性とのバランスを考慮して適切な次元数を検討した。

ストレス値	適合度評価
< 0.05	Excellent
0.05～0.10	Good
0.10～0.20	Fair
> 0.20	Poor

Table 2.2.3　ストレス値と適合度評価の対応例

ただし、ストレス閾値は経験則であり、対象数・データ構造・MDS の定義に依存する。

3. SAS による実装

3.1 DISTANCE プロシジャ

DISTANCE プロシジャは、多変量データから対象間の距離または非類似度を算出し、距離行列を作成するためのプロシジャである[6]。MDS では対象間の距離行列を入力データとして利用する。Program 3.1 に DISTANCE プロシジャの基本形式を示す。

```
proc distance data=【多次元データ】 out=【距離行列】 method=[距離尺度] shape=[行列表示] ;
    id [基準変数] ;
    var [尺度]([次元 1] -- [次元 n]) ;
run ;
```

Program 3.1　DISTANCE プロシジャの基本形式

DISTANCE プロシジャでは、METHOD=オプションにより距離尺度を指定する。代表的な距離尺度として、連続値データに適したユークリッド距離を表す EUCLID（デフォルト）や、存在・非存在を表す二値データ

に適したジャッカード距離（非類似度）を表す DJACCARD が利用できる（JACCARD の場合は類似度として1－非類似度）。また、マンハッタン距離の CITYBLOCK やコサイン類似度の COSINE など、多様な距離尺度も選択可能である。SHAPE=オプションでは出力する距離行列の形式を指定する。TRIANGLE（デフォルト）は距離行列の片側のみを出力し、SQUARE は完全な対称行列として出力する。

ID ステートメントでは対象を識別する変数を指定する。VAR ステートメントでは距離計算に用いる変数および尺度水準を指定する。尺度水準としては、本研究では、ユークリッド距離に対して連続尺度を表す INTERVAL や順位尺度を表す ORDINAL、ジャッカード距離に対して ANOMINAL を用いて距離行列を作成した。

3.2 MDS プロシジャ

MDS プロシジャは、DISTANCE プロシジャで作成した距離行列または非類似度行列を入力として、MDS による低次元空間へのマッピングを行うためのプロシジャである[7]。SAS は古典的 MDS ではなく、計量 MDS または非計量 MDS を実装している。Program 3.2 に MDS プロシジャの基本形式を示す。

```
proc mds data=【距離行列】 out=【マッピング座標】 level=[計量水準] dimension=[圧縮次元] oiter pfit ;
    id [識別変数] ;
    var [距離 1] -- [距離 n] ;
run ;
```

Program 3.2　MDS プロシジャの基本形式

MDS プロシジャでは、LEVEL=オプションにより計量水準を指定する。主な指定として、距離そのものを保持する計量 MDS を表す ABSOLUTE と、距離順位を保持する非計量 MDS を表す ORDINAL がある。LEVEL=ABSOLUTE では、入力された距離値をそのまま再現するように配置座標を推定する。一方、LEVEL=ORDINAL では距離順位を保持することを目的とし、単調変換後距離(disparity)と配置後距離との乖離が最小となるよう座標を推定する。DIMENSION=オプションでは出力するマッピングの次元数を指定する（デフォルトは 2)。ID ステートメントでは出力結果に表示する識別変数を指定する。MDS プロシジャでは、SMACOF アルゴリズムを用いてストレス関数の最小化を行うために、OITER オプション(Output Iterations)を指定することで反復過程および収束状況を確認できる。また、OUT=オプションによりマッピング座標を出力し、OUTFIT=オプション(Output FIT statistics)によっても適合度評価結果を出力できる。PFIT オプション(Print FIT statistics)はその最終結果をアウトプット画面に出力できる（小数点以下 2 桁までであるが適合度の評価には十分といえる)。

3.3 SGPLOT プロシジャ

MDS プロシジャでは低次元空間上の座標を算出できるが、マッピング結果の解釈や可視化を行うためにはグラフ化が必要である。本研究では、MDS プロシジャで出力された座標データセットを利用し、SGPLOT プロシジャによってマッピング図およびストレス推移図を作成した。Program 3.3 に SGPLOT プロシジャの基本形式を示す。

```
proc sgplot data=[MDS 結果（加工後）] ;    /* グループ変数は必要に応じて追加 */
    scatter x=Dim1 y=Dim2 / datalabel=[識別変数] group=[グループ変数] ;
    xaxis label='Dimension 1' display=(noticks novalues) offsetmin=0.1 offsetmax=0.1 ;
```

```
    yaxis label='Dimension 2' display=(noticks novalues) offsetmin=0.1 offsetmax=0.1 ;
run ;
```

(a)マッピング座標

```
proc sgplot data=[MDS 結果（加工後）] ;
    scatter x=_ITER_ y=Dim1 ;
    series x=_ITER_ y=Dim1 ;
    xaxis label='Iteration' ;
    yaxis label='Stress' ;
run ;
```

(b)ストレス推移

Program 3.3　SGPLOT プロシジャの基本形式

SGPLOT プロシジャでは、(a)のように SCATTER ステートメントにより MDS の配置座標を散布図として描画できる DATALABEL=オプションを指定することで対象名を表示でき、GROUP=オプションを指定することでカテゴリ別のシンボル・カラーの設定が可能である。MDS では距離構造を低次元空間上へ写像するため、各対象の位置関係から類似性や関連性を視覚的に把握できる。一般に 2 次元表示は視認性に優れ、最も利用される形式である。一方、1 次元表示は距離構造を単一軸上で表現できるため解釈が容易であり、3 次元表示はより多くの情報を保持できるが視覚的な解釈が複雑になる。図の作成にあたっては、MDS 座標の絶対値よりも対象間の相対的な位置関係が重要であるため、軸目盛や補助線は必要に応じて簡略化した。MDS マッピングでは、プロット領域の端に配置されたラベルやシンボルをグラフ内で調和させるために、OFFSETMIN=および OFFSETMAX=オプションを用いて軸の両端に余白を設けることが望ましい。

また、OITER オプションによって出力された反復履歴データを用いてストレス推移図を作成し、SMACOF アルゴリズムの収束過程を確認した。ストレス推移図は(b)のように作成できる。TYPE='CRITERION'のレコードでは Dim1 変数がストレス値を表す。

4. 臨床データにおける実装

4.1 テストデータ

臨床データを想定し、簡略化した架空のテストデータを 2 パターン作成した。本研究では、患者背景データおよび有害事象データを対象として MDS を適用した。患者背景データでは症例間の類似性評価を目的とし、年齢、BMI、重症度および QOL から算出したユークリッド距離を用いて計量 MDS を実施した。なお、例えば性別(SEX)は名義尺度であり、男性=1、女性=2 のように数値化されることが多い。しかし、この数値には大小関係や間隔の意味はなく、1 と 2 との差に定量的な意味を持たせることは適切ではない。そのため、名義尺度データはカテゴリとして扱うべきであり、距離計算においては尺度水準に応じた適切な指定が必要である。

一方、有害事象データでは事象間の関連性評価を目的とし、発現有無データに基づくジャッカード距離を用いて非計量 MDS を実施した。

両データの特徴を Table 4.1.1、患者背景データを Table 4.1.2、有害事象データを Table 4.1.3 に示す。

	患者背景	有害事象
評価対象	症例	事象
評価目的	類似性	関連性（共発現）
観測単位	1 症例 1 レコード	1 事象複数症例
データ形式	連続値	離散値(0/1)
距離尺度	ユークリッド距離	ジャッカード距離
計量水準	計量 MDS	非計量 MDS

Table 4.1.1　テストデータの特徴

SUBJECT	AGE	BMI	SEVERITY	QOL
P001	34	21	Very Mild	Excellent
P002	42	27.5	Mild	Excellent
P003	47	23	Very Mild	Fair
P004	53	29	Moderate	Excellent
P005	58	24.5	Mild	Poor
P006	61	31.5	Severe	Good
P007	67	26	Moderate	Poor
P008	72	32	Very Severe	Good
P009	75	28.5	Severe	Fair
P010	80	34	Very Severe	Poor

(a)データ本体

SEVERITY		QOL	
1	Very Mild	4	Excellent
2	Mild	3	Good
3	Moderate	2	Fair
4	Severe	1	Poor
5	Very Severe		

(b) 順序尺度カテゴリの定義

Table 4.1.2　患者背景のテストデータ

SUBJECT	SEQ	AE_CATEGORY	AE_TERM
P001	1	Headache	Neurological
P001	2	Dizziness	Neurological
P002	1	Headache	Neurological
P002	2	Rash	Dermatological
P003	1	Headache	Neurological

P003	2	Dizziness	Neurological
P003	3	Fatigue	General
P004	1	Nausea	Gastrointestinal
P004	2	Vomiting	Gastrointestinal
P005	1	Nausea	Gastrointestinal
P005	2	Vomiting	Gastrointestinal
P006	1	Nausea	Gastrointestinal
P006	2	Fatigue	General
P007	1	Fatigue	General
P007	2	Vomiting	Gastrointestinal
P008	1	Rash	Dermatological
P009	1	Headache	Neurological
P009	2	Dizziness	Neurological
P009	3	Rash	Dermatological
P010	1	Nausea	Gastrointestinal
P010	2	Vomiting	Gastrointestinal
P010	3	Rash	Dermatological

(a)データ本体

AE_CATEGORY	AE_TERM
Neurological	Dizziness
	Headache
Gastrointestinal	Nausea
	Vomiting
Dermatological	Rash
General	Fatigue

(b)カテゴリ対応の辞書

Table 4.1.3　有害事象のテストデータ

4.2 マッピングおよび適合度評価

MDS の解析結果は、選択する距離尺度および計量水準によって大きく異なる。そのため、データ構造や解析目的に応じて、DISTANCE プロシジャおよび MDS プロシジャの設定を適切に選択する必要がある。

本研究では、まず DISTANCE プロシジャにより対象間距離を算出し、距離行列を作成した。患者背景データについては、年齢および BMI を連続変数として用いた。重症度および QOL は本来順序尺度であるが、本研究では MDS の実装方法を例示することを目的として、各カテゴリに順序に沿った数値を割り当て、隣接するカテゴリ間の差を等間隔と仮定したうえで、便宜的に量的変数として扱った。さらに、変数間の単位およびばらつきの違いが距離計算に与える影響を調整するため、STDIZE プロシジャを用いて、各変数を平均 0、標準偏差 1 となるよう標準化した。その後、標準化した各変数を用いてユークリッド距離を算出した。なお、この標準化は変数間の尺度の大きさを調整する処理であり、順序尺度そのものを連続尺度へ変換するもので

はない。有害事象データについては発現有無を表す二値データに適したジャッカード距離を採用した。

続いて、作成した距離行列を MDS プロシジャへ入力し、低次元空間へのマッピングを行った。患者背景データに対しては距離情報そのものを保持する計量 MDS(LEVEL=ABSOLUTE)を適用した。一方、有害事象データに対しては距離順位を保持する非計量 MDS(LEVEL=ORDINAL)を適用した。ジャッカード距離は共発現症例の割合に基づいて類似性を評価するため、有害事象間の発現パターンの比較に適した距離尺度である。

得られた座標は SGPLOT プロシジャを用いて可視化した。また、MDS プロシジャの PFIT オプションおよび OITER オプションを利用し、ストレス値による適合度評価および反復過程の収束状況を確認した。

DISTANCE プロシジャおよび MDS プロシジャの周辺について、テストデータ別の SAS プログラムを Program 4.2.1 に示す。

```
proc stdize data=DM out=DM_STD ;     /***  標準化  ***/
     var AGE -- QOL_N ;
run ;

proc distance data=DM_STD out=DM_DST method=euclid ;     /***  ユークリッド距離行列変換  ***/
     id SUBJECT ;
     var interval(AGE -- QOL_N) ;     /*  順位尺度は数値に置き換えて使用（標準化のため interval で統一）  */
run ;

proc mds data=DM_DST out=DM_MDS level=absolute dimension=2 oiter pfit ;     /***  計量 MDS 実施  ***/
     id SUBJECT ;
     var P001 -- P010 ;
run ;
```

(a)患者背景データ（ユークリッド距離行列・計量 MDS）

```
proc freq data=AE ;     /***  集計  ***/
     table SUBJECT * AE_TERM / out=AE_FRQ ;
run ;
proc sort data=AE_FRQ ;
     by AE_TERM ;
run ;
proc transpose data=AE_FRQ out=AE_TRNS ;     /***  転置  ***/
     by AE_TERM ;
     id SUBJECT ;
     var COUNT ;
run ;
data AE_CMP ;     /***  欠測補完  ***/
     set AE_TRNS ;
     array X _numeric_ ;
     do over X ;
```

```
        if missing(X) then X = 0 ;
    end ;
run ;

proc distance data=AE_CMP out=AE_DST method=djaccard ;    /*** ジャッカード距離行列変換 ***/
    id AE_TERM ;
    var anominal(P001-P010) ;    /* 転置後のため最初と最後の変数を指定する場合は順序に注意 */
run ;

proc mds data=AE_DST out=AE_MDS level=ordinal dimension=2 oiter pfit ;    /*** 非計量 MDS 実施 ***/
    id AE_TERM ;
    var Dizziness -- Vomiting ;
run ;
```

(b)有害事象データ（ジャッカード距離行列・非計量 MDS）

Program 4.2.1　テストデータの実装（DISTANCE プロシジャおよび MDS プロシジャ）

有害事象データについては、被験者ごとの発現有無を 0/1 形式で集計した後、各有害事象を行、被験者を列とする行列形式へ転置した(Table 4.1.4)。ジャッカード距離は存在・非存在を表す二値データを前提としているため、発現記録が存在しない被験者については欠測値を 0（非発現）として補完した。欠測値を残した場合には対象間距離を適切に算出できないため、すべての被験者について発現有無が明示された二値行列へ変換した。このように整形したデータを用いてジャッカード距離行列を作成し、非計量 MDS の入力データとした。

VIEWTABLE: Work.Ae_cmp

	AE_TERM	P001	P002	P008	P006	P007	P009	P003	P004
1	Dizziness	1	1	1	0	0	0	0	0
2	Fatigue	0	0	1	1	1	1	0	0
3	Headache	1	1	0	0	1	0	1	0
4	Nausea	1	0	0	0	0	1	1	1
5	Rash	0	0	0	1	0	0	0	0
6	Vomiting	0	0	0	0	0	0	1	1

Table 4.1.4　事前処理（集計・転置・欠測補完）した有害事象データ

マッピング座標とストレス推移のグラフ描画の過程は有害事象データを例に、Program 4.2.2 に示す。カテゴリはシンボルやカラーを変えて表示し、多い場合は上位のいくつかに絞る方法もある。

```
/* 反復過程における最終データの抽出 */
proc sort data=AE_MDS ;
    by AE_TERM _ITER_ ;
run ;
data AE_MDS_WHR ;
    set AE_MDS ;
    where _TYPE_ = 'CONFIG' ;
```

```
    by AE_TERM ;
    if last.AE_TERM ;
    Y0 = 0 ;    /* 1 次元マッピング用の固定軸 */
run ;
data AE_MAPPING ;
    merge
      AE_MDS_WHR
      AE_DICTIONARY    /* 辞書データセットとマージしてカテゴリを取得 */
    ;
    by AE_TERM ;
run ;

ods graphics / attrpriority=none border=off ;
/*** 2 次元マッピング座標の描画 ***/
proc sgplot data=AE_MAPPING ;    /* MDS プロシジャで demension=2 で出力したデータを使用 */
    styleattrs datasymbols=(circlefilled squarefilled trianglefilled diamondfilled) ;    /* シンボル指定 */
    scatter x=Dim1 y=Dim2 / group=AE_CATEGORY datalabel=AE_TERM datalabelattrs=(size=12pt)
markerattrs=(size=12 color=black) ;
    xaxis label='Dimension 1' display=(noticks novalues) offsetmin=0.1 offsetmax=0.1 ;
    yaxis label='Dimension 2' display=(noticks novalues) offsetmin=0.1 offsetmax=0.1 ;
run ;
/*** 1 次元マッピング座標の描画 ***/
proc sgplot data=AE_MAPPING ;    /* MDS プロシジャで demension=1 で出力したデータを使用 */
    refline 0 / axis=y lineattrs=(color=black) ;    /* 参照線の追加 */
    styleattrs datasymbols=(circlefilled squarefilled trianglefilled diamondfilled) ;
    scatter x=Dim1 y=Y0 / group=AE_CATEGORY datalabel=AE_TERM datalabelattrs=(size=12pt)
markerattrs=(size=12 color=black) ;
    xaxis label='Dimension 1' display=(noticks novalues) offsetmin=0.1 offsetmax=0.1 ;
    yaxis display=(nolabel noticks novalues) offsetmin=0.1 offsetmax=0.1 ;
run ;

/* 反復過程におけるストレス値レコードの抽出 */
data STRESS ;
    set AE_MDS ;
    where _TYPE_ = 'CRITERION' ;
run ;
/***  ストレス推移の描画 ***/
proc sgplot data=STRESS noautolegend ;
```

```
    scatter x=_ITER_ y=Dim1 / markerattrs=(symbol=circlefilled color=black size=10) ;
    series x=_ITER_ y=Dim1 / lineattrs=(color=black) ;
    xaxis label='Iteration' labelattrs=(size=18) valueattrs=(size=12pt) ;
    yaxis label='Stress' labelattrs=(size=18) valueattrs=(size=12pt) ;
run ;
```

Program 4.2.2　マッピング座標とストレス推移のグラフ描画（有害事象データの例）

5. 結果

5.1 距離行列

DISTANCE プロシジャを用いて作成した距離行列を Table 5.1 に示す。

(a)の患者背景データでは、年齢、BMI、重症度および QOL を用いてユークリッド距離を算出した。距離行列は対称行列となり、対角要素は対象自身との距離を表すため 0 となった。また、症例間の距離は一様ではなく、患者背景の違いに応じて異なる値を示しており、対象間の類似性の差異が反映されていることが確認された。

一方、(b)の有害事象データでは発現有無情報からジャッカード距離を算出した。ジャッカード距離は 0 から 1 の範囲をとり、値が小さいほど類似した発現パターンを示し、値が大きいほど発現パターンが異なることを意味する。有害事象間で異なる距離値が得られており、共発現パターンの違いが距離行列として表現されていることが確認された。

これらの距離行列を入力データとして MDS を適用し、対象間の関係性を低次元空間上に可視化した。

VIEWTABLE: Work.Dm_dst

	SUBJECT	P001	P002	P003	P004	P005	P006	P007	P008	P009	P010
1	P001	0	.	.	.	.	.	.	.	.	.
2	P002	1.7848086113	0	.	.	.	.	.	.	.	.
3	P003	1.8613259374	2.0538485162	0	.	.	.	.	.	.	.
4	P004	2.6674132448	1.0570264788	2.5556654554	0	.	.	.	.	.	.
5	P005	3.0489373313	2.6914756779	1.3183302439	2.7059476682	0	.	.	.	.	.
6	P006	3.7817915256	2.2243681497	3.1197087645	1.3104413682	2.6769401595	0	.	.	.	.
7	P007	3.6969690715	2.9898505133	2.171933208	2.6415690888	0.9694683738	2.2017652495	0	.	.	.
8	P008	4.6094910373	3.1365664532	3.9106676343	2.1317349066	3.2659177075	1.0006252137	2.5461011383	0	.	.
9	P009	4.1537370878	3.0284454384	3.0468123853	2.2570426634	2.1519986939	1.4185802795	1.3104413682	1.3496090129	0	.
10	P010	5.6550110783	4.3003828166	4.4353875465	3.4742849442	3.3814080057	2.2128687736	2.5026786415	1.7316941463	1.7137404514	0

(a)患者背景データ（ユークリッド距離）

VIEWTABLE: Work.Ae_dst

	AE_TERM	Dizziness	Fatigue	Headache	Nausea	Rash	Vomiting
1	Dizziness	0	.	.	.	.	.
2	Fatigue	0.75	0	.	.	.	.
3	Headache	0.3333333333	0.8	0	.	.	.
4	Nausea	1	0.8	1	0	.	.
5	Rash	1	1	0.75	1	0	.
6	Vomiting	1	0.8	1	0.5	1	0

(b)有害事象データ（ジャッカード距離）

Table 5.1　距離行列に変換したデータ

5.2 マッピング

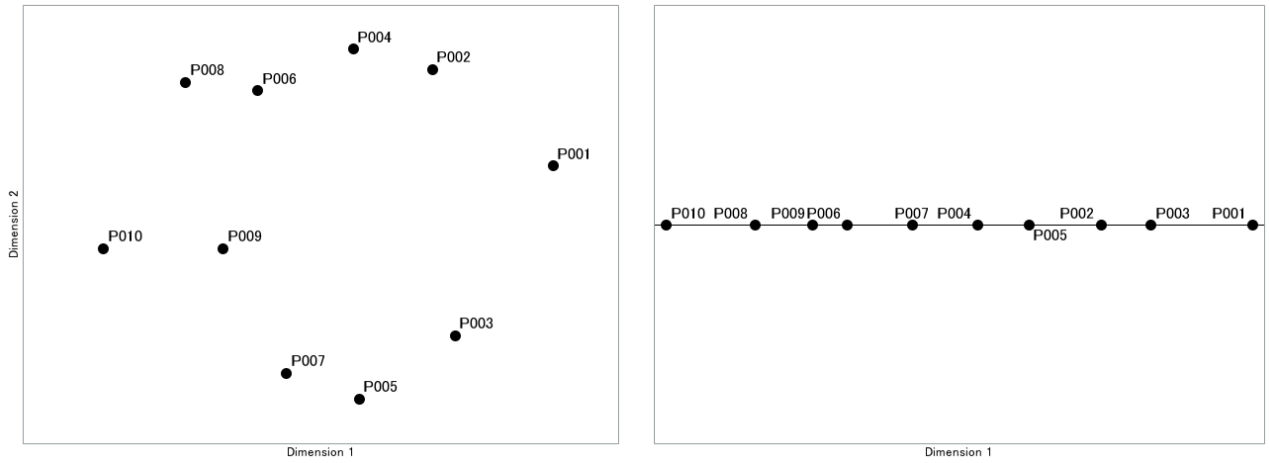
距離行列を入力した MDS プロシジャによる結果例を Table 5.2 に示す。OITER オプションにより出力された反復履歴における収束時を抽出している。

VIEWTABLE: Work.Mapping

	ITER	_DIMENS_	_MATRIX_	_TYPE_	AE_TERM	_NAME_	Dim1	Dim2
1	11	2	.	CONFIG	Rash	Dizziness	0.4602283014	1.2726796526
2	11	2	.	CONFIG	Nausea	Fatigue	-1.341352486	0.9654566948
3	11	2	.	CONFIG	Vomiting	Headache	1.0927828445	0.4009942994
4	11	2	.	CONFIG	Fatigue	Nausea	1.0286688664	-0.53343085
5	11	2	.	CONFIG	Dizziness	Rash	-1.516317113	-0.781657389
6	11	2	.	CONFIG	Headache	Vomiting	0.2759895865	-1.324042408

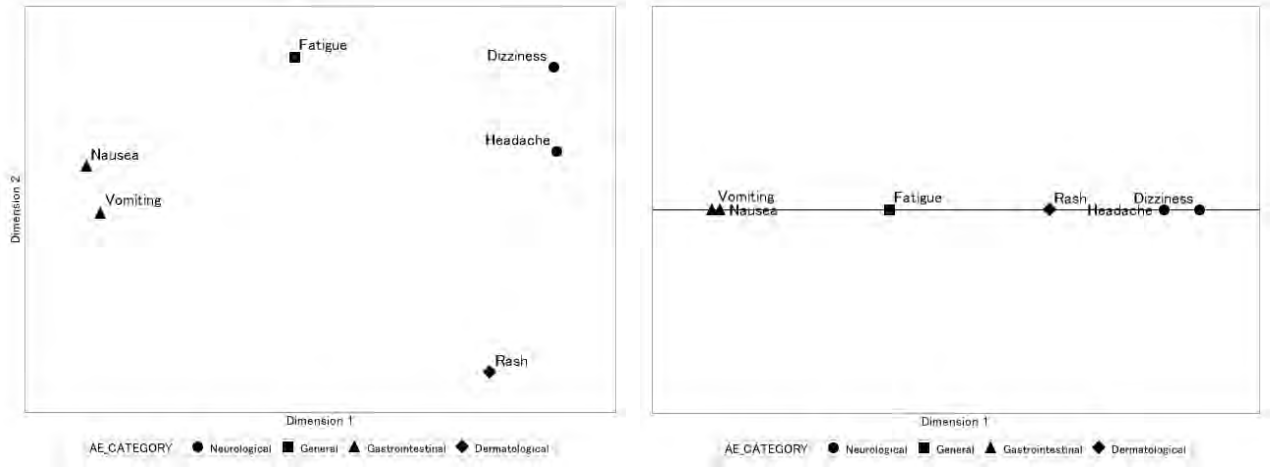
Table 5.2　最終マッピング座標のデータセット例（有害事象・2 次元の場合）

Figure 5.2.1 は患者背景データ、Figure 5.2.2 は有害事象データについて、マッピング座標のデータセットを SGPLOT プロシジャでグラフ化したものである。



(a)2 次元　　(b)1 次元

Figure 5.2.1　計量 MDS による患者背景データのマッピング結果



(a)2 次元　　(b)1 次元

Figure 5.2.2　非計量 MDS による有害事象データのマッピング結果

患者背景データの計量 MDS 結果では、1 次元解では症例がおおむね年齢順に配置されており、年齢による患者背景の差異が距離構造に反映されていることが示唆された。一方、2 次元解では左右方向の配置に加えて上下方向への分離が認められた。全体的に、左右に年齢順に並んでいるが、上下では重要度や QOL の違いが示された。例えば、P008、P006、P004 および P002 は上部に配置されているのに対し、P007、P005 および P003 は下部に配置されており、重症度や QOL などの患者背景因子の違いが反映されたものと考えられる。本解析では各変数を事前に標準化しているため、マッピング結果は年齢のみの影響ではなく、BMI、重症度および QOL を含む複数の患者背景因子を総合的に反映した結果である。

有害事象データの非計量 MDS 結果では、1 次元解では各有害事象が単一軸上に配置され、発現パターンに基づく相対的な近さを把握できる。しかし、非計量 MDS では距離そのものではなく距離順位を保持することを目的とするため、1 次元解では複数の関係性を十分に表現できない場合がある。一方、2 次元解では有害事象間の位置関係がより明瞭に表現された。例えば、Neurological カテゴリの Headache と Dizziness は近接して配置されており、両事象が類似した共発現パターンを有することが示された。同様に、Gastrointestinal カテゴリの Nausea と Vomiting も近接して配置されており、発現傾向の類似性が示唆された。一方、Rash は他の有害事象から離れて配置されており、比較的独立した発現パターンを有することが示された。また、Fatigue は両群の中間付近に位置しており、複数の有害事象群との共発現を反映した結果と考えられた。このように、類似した発現パターンを有する事象や異なる発現パターンを有する事象を視覚的に区別しやすくなった。

なお、MDS で得られる座標配置は距離構造の再現を目的としており、軸は回転・反転・平行移動に対して一意ではない。そのため、Dimension 1 および Dimension 2 を特定の臨床的潜在因子として解釈することは適切ではなく、対象間の相対的な位置関係に着目して解釈する必要がある。また、近接して配置された対象は距離行列に基づく類似したパターンを有することを示すが、因果関係、医学的関連性、または発現頻度の高さを意味するものではない。

5.3 適合度評価

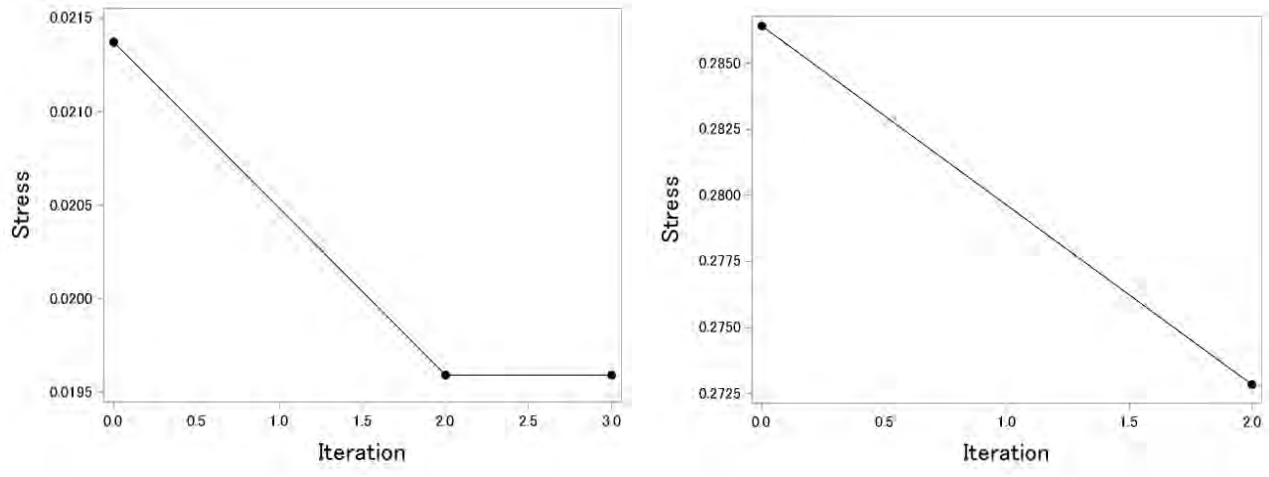
Table 5.3.1 は、OITER オプションにより出力された反復履歴データセットの一例である。_ITER_は反復回数を表す SAS 内部変数であり、0 は初期配置を示す。本解析では_ITER_の値が 0, 2, 3, 4, …と出力されており、1 は表示されていない。これは PROC MDS 内部での初期評価および反復処理の実装によるものであり、実際の収束状況の評価には影響しない。本研究では、CRITERION レコードに格納されたストレス値の推移を用いて SMACOF アルゴリズムの収束状況を評価した。

VIEWTABLE: Work.Stress

	ITER	_DIMENS_	_MATRIX_	_TYPE_	AE_TERM	_NAME_	Dim1	Dim2
1	0	2	.	CRITERION			0.1679673432	.
2	2	2	.	CRITERION			0.0680653515	.
3	3	2	.	CRITERION			0.0621276691	.
4	4	2	.	CRITERION			0.0618088605	.
5	5	2	.	CRITERION			0.0513541006	.
6	6	2	.	CRITERION			0.051285457	.
7	7	2	.	CRITERION			0.048583521	.
8	8	2	.	CRITERION			0.0485041145	.
9	9	2	.	CRITERION			0.0479556954	.
10	10	2	.	CRITERION			0.0472315274	.
11	11	2	.	CRITERION			0.0472315274	.

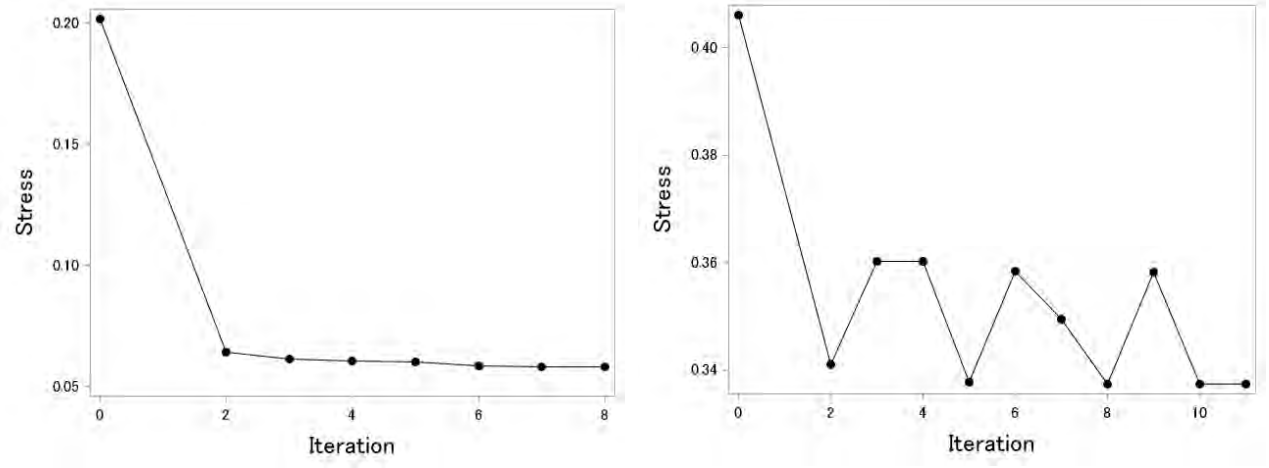
Table 5.3.1　反復法によるストレス推移のデータセット抽出例（有害事象・2 次元の場合）

Figure 5.3.1 は患者背景データ、Figure 5.3.2 は有害事象データについて、ストレス推移のデータセットを SGPLOT プロシジャでグラフ化したものである。



(a)2 次元　　(b)1 次元

Figure 5.3.1　計量 MDS による患者背景データにおけるストレス値の推移



(a)2 次元　　(b)1 次元

Figure 5.3.2　非計量 MDS による有害事象データにおけるストレス値の推移

患者背景データのストレス値はいずれの次元数においても少数回の反復で収束した。2 次元解は 1 次元解よりも低いストレス値を示し、患者背景データの距離構造をより良好に保持していることが確認された。

有害事象データに対する非計量 MDS では、反復初期にストレス値が大きく低下し、その後は小幅な増減を繰り返しながら収束した。非計量 MDS では距離そのものではなく順位情報を保持するため、各反復で単調変換後の disparity と配置後距離との適合が同時に調整される。そのため、ストレス値が単調に減少するとは限らず、収束過程において一時的な変動が認められる場合がある。本解析においても、反復後半ではストレス値はほぼ一定の範囲内で推移しており、安定した解へ収束していることが確認された。

さらに、次元別に最終反復番号、最終ストレス値およびその適合度評価の結果一覧を Table 5.3.2 に示す。

データ	計量水準	次元数	最終反復番号	最終ストレス値	適合度評価
患者背景	計量 MDS	3	3	0.0015	Excellent
		2	3	0.0196	Excellent
		1	2	0.2729	Poor
有害事象	非計量 MDS	3	13	0.0215	Excellent
		2	8	0.0583	Good
		1	11	0.3374	Poor

Table 5.3.2　データ・次元別の結果比較

本検討では、患者背景データに対する計量 MDS および有害事象データに対する非計量 MDS について、次元数を変化させた際の適合度をストレス値により評価した。その結果、両データセットとも 3 次元解は最も低いストレス値を示したが、2 次元解の時点で十分良好な適合度が得られた。さらに、2 次元解は視覚的な解釈が容易であることから、適合度と解釈性の両面を考慮すると実用的な選択肢であると考えられた。これに対し、1 次元解ではストレス値が大幅に増加し、距離構造の再現性が著しく低下した。

患者背景データでは、ストレス値は 3 次元解の 0.0015 から 2 次元解の 0.0196 へわずかに増加したのみであったが、1 次元解では 0.2729 まで上昇した。有害事象データにおいても同様に、3 次元解の 0.0215 から 2 次元解の 0.0583 への増加は限定的であった一方、1 次元解では 0.3374 まで増加した。これらの結果は、2 次元解が元の距離構造を概ね保持しながら、1 次元解では表現できない情報を保持できることを示している。

また、3 次元解は最も高い適合度を示したものの、論文や報告書における可視化および解釈の容易さを考慮すると実用性には限界がある。一方、2 次元解は散布図として容易に表現可能であり、対象間の関係性を直感的に把握できる利点を有する。したがって、本検討の結果から、2 次元解は適合度と視覚的解釈性のバランスに優れた実用的な次元数であることが示唆された。

さらに、有害事象データに対する非計量 MDS は、患者背景データに対する計量 MDS と比較して収束までに多くの反復を要する傾向が認められた。これは、非計量 MDS では距離そのものではなく順位関係の保持を目的としてストレス関数を最小化するため、最適化過程がより複雑となることに起因すると考えられる。

6. まとめ

SAS の DISTANCE プロシジャおよび MDS プロシジャを用いて、多次元尺度構成法の適用方法と適合度評価について検討した。ユークリッド距離に計量 MDS、ジャッカード距離に非計量 MDS を適用し、それぞれのデータ構造に応じて、類似性や共発現の解析方法を示した。適合度評価にはストレス値を用い、いずれのデータにおいても、1 次元解と比較して 2 次元解で大幅な改善が認められた。一方、2 次元解から 3 次元解への改善は限定的であり、解釈性と適合度の観点から 2 次元解が実用的であることが確認された。

本稿では、理解を容易にするため少数例のテストデータを用いたが、計算負荷、ラベル重複、可視化の困難さの影響を受けつつ、MDS の有用性は対象数や変数数が増加する大規模データにおいてより顕著になることも考えられる。特に、多数の症例、多数の事象、あるいは高次元のデータなど、対象間の距離構造を直感的に把握することが困難なデータに対して、MDS は探索的解析および可視化の有効な手法として活用できると考えられる。

参考文献

[1] Chung N, Miasojedow B, Startek M, Gambin A. (2019). Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. BMC Bioinformatics. 20(Suppl 15), 644.

[2] Kruskal J. B., Wish M. (1978). Multidimensional Scaling. Sage Publications.

[3]日本統計学会(2020)、統計学実践ワークブック、学術図書出版社、232-238.

[4] de Leeuw J., Mair P. (2009). Multidimensional Scaling Using Majorization: SMACOF in R. Journal of Statistical Software, 31(3), 1-30.

[5] Kruskal J. B. (1964). Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. Psychometrika, 29(1), 1-27.

[6] SAS Institute Inc. (2025), SAS/STAT®User's Guide -The DISTANCE Procedure.

[7] SAS Institute Inc. (2025), SAS/STAT®User's Guide -The MDS Procedure.

量子計算機のいらない世界初の組み合わせ最適化 LDF(**Revised IP Optimal-LDF)**

新村秀一、 Dr. Sci.　成蹊大学名誉教授

新村秀樹、イノベストメント株

概要: Fisher は n 件* p 変数の 2 群が平均だけ異なる正規分布と仮定し LDF を定義し判別理論を創設した。そして癌以外の医学診断や Altman スコア等の各種評価システムに用いられたが、判別結果を Number of Misclassifications や Error Rate で評価できない問題があった。そこで 2^n の組み合わせで Minimum-NM= k を見つける Optimal-LDF の RIP を開発した。k=0 であれば Linearly Separable Data であり、 k が 1 以上であればデータに固有な k 例の誤判別される $\mathbf{x}_i$ を特定できる。この誤判別理由を見つける事が判別分析で重用である。そして k 例を省くと LSD になる。山中 4 因子を見つけた変数減少法で、(2^p-1)の判別モデルから最小変数のモデルの Basic Gene Set を見つける。このモデルは正しい k 重 CV の検証標本の平均誤分類確率の M2 が最小のオッカムの剃刀をみたす Best なモデルである。これを検証することが判別分析の目的である。

以上のことを Iris、スイス銀行紙幣の真贋判定、試験の合否判定、普通車と小型車の判別と 169 の遺伝子の発現データで癌と正常が LSD である事を実証研究し、表 1（PP 参照）の LSD の判別理論[1-23]を体系化した。また量子計算機や物流システムの配送問題に用いられる Integer Programming ソフトの信頼性の検証に RIP と LSD 研究のデータを用いる事を提案した[24]。そして 53 年の成果をヨーロッパ医学会誌に分かり易い論文にまとめた[25]。

1.　判別理論の深刻な間違いと組みあわせ最適化 O LDF の RIP による解決

—LSD が判別理論の基本で MNM が真の判別理論の統計量—

Gauss が 2 地点間の距離 y を n 回測定し釣り鐘上の Gauss 分布を見つけ、関連する最小自乗法も考えた。これを統計が採用し、単に集計する記述統計から重回帰分析や分散分析等の推測統計学に発展した。

推測統計学の発展に貢献した Fisher は、n 件 p 変数の 2 群が平均の異なる 2 つの正規分布と仮定し、2 群の正規分布の比を log 変換して LDF を定義し判別理論を創設した。しかし LDF を Iris データで評価したのは、恐らく判別理論の欠点を示し後世に正しい判別理論の再構築を期待した良心からと考えるが誰も指摘していない。

LDF に次の問題点がある。

1）Fisher の仮説を満たす分散共分散行列を用いて多くの判別関数が開発され利用されたが、仮説を満たす実データは Iris 以外ないくらい厳しい条件であり判別結果も役に立たない。

2）判別結果を評価する NM は、図 1 で説明するとおり 0 から(n-1)の n 個の整数値の組みあわせ問題になる。一方重回帰分析や分散分析は正規分布を仮定した役に立つ統計手法であり、判別分析を同一視するのは間違いである。また判別結果を評価する NM あるいは ER は同じデータで異なり、全く信頼できず役にたたないという判別分析の Problem1 がある。各データに一意に決まる MNM が正しい統計量で、MNM=0 であれ

ば LSD であり LSD が判別理論の基本である。MNM=k の場合、多くの判別関数の NM は k 以上でバラつくので評価に値しない。

3）医学診断の 2 群を考えれば、y_i = 1(癌/疾病)/-1 (正常)は医師が種々の情報から決めた外的基準である。判別は、この外的基準が正しいか否かを検討するのが第 1 目的である。誤分類がなぜ起きたか検討する必要があり、RIP しかデータ固有に決まる誤分類を特定できない。多群判別は正常のようなコントロール群と他の群の 2 群判別で検討すべきである。多群の判別やクラスター分析は、分析対象を把握するのが目的で、それが最終目的であってはいけない。

4）y_i を重回帰分析の目的変数とする Plug-in 規則で、Goodnight 創業社長の開発した SAS の掃き出し演算子[30]を用いた RSQUARE で児頭骨盤不均衡の CPD（240＊19）データで$(2^{19}-1)$の組みあわせから 16 変数のモデルを選んだが、RIP の選んだ 14 変数の Best なモデルに敵わなかった。

Vapnik(1995)はデータ空間で 2 つの Support Vector で 2 群を分け、SV 間にケース x_i がないことで LSD を示す H-SVM を提案した。しかし誤判別データに使えず誰も利用しない。ChatGPT はこの事実で SVM 研究が LSD を唯一研究しているというが間違いである。筆者が世界で初めて LSD 理論に成功した。次に 2 つの SV で x_i が誤判別される事を認め、これらの SV からの誤判別の距離の和を Penalty c で 2 次計画法で最小化する S-SVM を提案した。実証研究で c＝1000 とすれば、LSD と誤判別データの両方に対応できることを確認した。しかし理論研究家や統計ソフトは、S-SVM の LDF でなく、理論の斬新さから Kernel－SVM だけが今日用いられている。しかし正確に LSD か否かが分かるかを誰も実証研究していない。また S-SVM に比べて優れていることを示すべきである。多くの研究者は Kernel－SVM や、非階層クラスター分析等の難しい手法を高尚と考えている。しかし、RIP や階層型クラスター分析等の基本的な分析結果を基準として比較し、何が優れているのか明確に示すべきである。筆者の予想では、S-SVM の方が Kernel－SVM より役に立つと考える。

開原東大医学部名誉教授[26]が企画した介護保険システムは、東京都の副病院長が厚生省が億単位で集めた「I分間タイム Study データ」を難しく役に立たない SPSS の手法で分析し失敗した。そこで私が分類木を用いることを提案した。彼は 3 ヶ月で SPSS の Chaido で分析し[秀新1. 1]、それを C コードにして完成させた。国会の決めた実施時期に遅れていたため、分析に用いていないデータで検証しなかったので、実施当初に順位の逆転の問題が指摘された。筆者も統計分析の取得当初は、判別、重回帰、数量化などの理論から出発したが、名義尺度、順序尺度、連続尺度の尺度分類は成蹊大学の経済学部の学生に統計を教えるために初めて注目した。彼は医師向けの統計書を出版しているが、尺度分類の重要性を考えなかったと推測する[秀新2. 1]。

LSD に関しては、JMP の最尤推定法を用いたロジスティック回帰は、判別境界を ROC で変えた複数の 2＊2 の分割表から NM＝0 を捜せば、多くの場合 LSD になることを確認した。非線形判別関数の NM は正体不明である。もし NM=0 の場合、RIP で LSD であるか確認すべきである。理論研究家は、非線形判別関数や非階層クラスター分析を好むが、RIP と階層型クラスター分析の基本情報と比較研究すべきである。

一方 RIP で次の事が分かる。1）MNM=0 が LSD で、MNM=k がデータに固有な k 個の誤分類 x_i を特定できる。この k 個がなぜ誤分類されたかを検討するのが真の判別分析の目的である。2）そして k 個を省けば全判別データが LSD になる「ケースの選択法 4」ができ、判別理論を限りなく有用な理論に革新できる。3) 表 1 の LSD の判別理論（PP 参照）で、これ迄考えられなかった種々のデータで驚く程簡単で役にたつ判別

結果が得られた。分析対象を群に分けることで、判別分析はアイデア次第でこれまで考えてもみなかった成果が得られる。Fisher は恐らく判別分析のこの利点に期待し無理な仮説で判別理論を提案したと考える。4）既に優良企業と倒産企業を問題の多い LDF で判別した Altman スコアが経営分析で用いられている。このような過去の成功している成果を RIP で見なおせば驚く成果が得られる。医学診断で判別分析が用いられている場合、RIP で再検討すべきである。

2　RIP：世界初の組み合わせ最適化 OLDF による量子計算機のいらない LSD の判別理論[24]

2.1　判別理論の最重要な「NM と LDF の驚く関係(Fact1)」:一種の双対性

図 1 の 3 件 2 変数の LSD の判別を考える。$\mathbf{x}_i$ を代入した LDF＝$\mathbf{b}*x_i+1$ で LDF＝0 とすれば、2 次元判別係数空間を判別得点が正と負に分ける 3 組の線形判別超平面 Hi になる。ここで 定数項を 1 に固定して、$\mathbf{x}_i$ というデータ空間の値を判別係数 $\mathbf{b}$ と入れ替えることで判別係数と NM の関係が分かる一種の**双対な関係**を利用している。定数項を 1 で固定することで可能になる。図の右下がりの x_1 に対応する H1 の下側の＋半平面の任意の点の RIP は、判別得点が LDF＞0 の正で x_1 を正しく判別する RIP になる。上の負の半平面の点は、LDF＜0 の負で x_1 を誤判別する RIP になる。各凸体を囲む Hi の-半平面の和が NM になり、0 から 2 の 3 個の整数値になる。RIP は三角形の MNM＝0 の LSD を見つける。他の LDF は、7 凸体の一つを選び NM＝0 を見つけないことが多い。定数項 c を考えた $b*x_i$ ＋ c ＝0 を Hi とすれば、8 個の凸体は原点を共有し $2^3=8$ の LDF から意味のある情報は得られない。

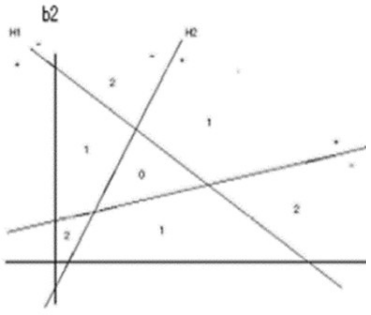


図 1　NM と LDF の関係

この Fact1 は LDF の定数項 c を 1 とする事で、1978 年に見つけた判別係数と NM の関係が分かる LSD 理論の基本である。NM は 0、1、2 の 3 整数値で、NM=3 は NM=0 と同じである 。この図は正規分布を仮定するのでなく、判別理論が「組み合わせ理論」である事を意味する。判別結果の評価は、ER の確率表現より NM の方が良い。NM=0 は MNM=0 と同じ LSD になり、科学技術で重要な 2 群が分かれているか否かを示す確率でなく 2 値の確定情報になる。他の LDF は現実を反映しない正規分布を当てはめるため、どの NM を選ぶか分からない。19 変数の CPD は、LINGO の IP ソルバーで作った RIP の Program1 で、240 症例の 2^240 組みの凸体から MNM=2 の最適凸体を見つける。この 2 例を省くと全判別データが LSD になる。この「ケースの選択法 4」は判別理論の応用範囲を限りなく拡大する。掃き出し演算子は 19 変数の分散共分散行列から$(2^{19}-1)$組みの任意の重回帰モデルの R2 値の大きな(18*19+1)個のモデルを選ぶ。そして変数増加法と減少法で AIC 最小の 16 変数を選んだ。重回帰分析ではこの方法は有用であるが、判別モデルで用いるべきでなかった。

2.2　RIP：n 件 p 変数の連立不等式で H_i と e_j の 2 組みの 2^n の組み合わせ最適化 OLDF

$\mathbf{x}_i$ が H_i：${}^t\mathbf{x}_i\mathbf{b}+1=0$ で LDF>0 と LDF＜0 で分かれれば LSD である。2 群を示す外的基準 $y_i=1/-1$ で、 y_i *LDF＞0 で LSD が定義できる。数値計算で 0 判定は避けたいので y_i *LDF＞1 とし、(-1、1)に $\mathbf{x}_i$ が無いことが LSD になる。そして RIP で誤分類を特定するため、判別境界の 1 を $e_j=1$ で（ $1-1000e_j$ ）＝−999

に変更する。これで e_j ＝0 であれば x_i が判別され、1 であれば誤判別データを y_i *LDF＞－999 で特定できる。整数計画法は e_j ＝1/0 で x_i を 2 者択一でき組み合わせ最適化ができる。そして、MNM=MIN(Σ e_j); で e_j の和を最小化する。MNM＝k であれば、誤分類される k 例が特定できる。高校数学で数理計画法の入門の線形計画法の「領域の最大/最小問題」を習うが、IP は単に LP の凸領域で整数値しか取らないので理解しやすい。しかし計算時間がかかる。

$$MNM=MIN(\Sigma e_i); \quad y_i(\overset{t}{\mathbf{x}_i}\mathbf{b}+1) \geq 1-1000^* e_i;$$

p 次元線形判別係数空間を n 個の H_i で分割すると各凸体は 0 から(n-1) の NM をもつ(2^n-1)組み以下の凸体になる。LSD の場合、RIP は MNM=0 の最適凸体の Optimal CP を見つける。LSD でない場合、MNM=k の誤分類を唯一特定できる。誤分類を省く「ケースの選択法 4」で全判別データが LSD になり、限りなく判別理論の適用範囲が広がる。即ち、RIP は H_i と e_i の 2 種の 2^n の組み合わせ最適化で、MNM＝k の OCP を特定できる。後で紹介するが、更に p 変数の（2^p-1）のモデルから最小の LSD である Basic Gene Set を iPS 研究で山中 4 因子を見つけた変数減少法で見つける。BGS が 10 重 CV の検証標本の ER の平均の M2 が最小で「Best なオッカムの剃刀を満たすモデル」になる多くの結果を示す。iPS 研究は万能細胞ができるか否か、LSD 判別は MNM=0 か否かの明白な事実であり、判別分析は決して確率現象ではないことの理解が重用だ。

2.3　判別手法の 4 分類

ChatGPT は SVM が LSD 研究していると言うが間違いである。筆者が RIP で始めて実用的な LSD 理論を体系化した[24,25]。

以下で既存の判別手法を 4 分類した。H-SVM は誤分類に対応できず誰も使わない。S-SVM は 2 次計画法の QP の 2 目的最適化で、ペナルティ c を 1000 にすれば両データに対応できるが、データ固有の誤分類が分からない中途半端な LDF である。kernel SVM は、LSD を必ず特定できるかを研究者が確認する必要がある。RIP と JMP のロジスティック回帰が、169 の Array と SMall Matryoshka と BGS を判別し LSD を確認した。その他の LASSO、正則化判別関数、SVD を使った高次元 LDF、機械学習等の分類手法は、LSD の判別に役に立たない。

分類 1：理論的に LSD が特定できるのは RIP と H-SVM だけである。しかし H-SVM は LSD しか判別できず、誤分類に対応できないので誰も利用しない。

分類 2：JMP のロジスティック回帰は Array、 SM、 BGS で LSD を判別できる事を確認したが、筆者が作った表 19 の 3 件 2 変数の簡単な LSD で、データ件数が少ないためか収束計算がうまくできず NM=1 になる。このデータは、IP ソフトや量子計算機の瑕疵を LSD という分かり易い基準で判定できる。量子計算機の評価問題に「巡回セールスマン問題」が良いとされているが間違いである。確かにプログラムが正しければ組み合わせ最適化で最適解が得られるが、誰もが容易に LSD のように正しいか否か直ぐに判断できない。

分類 3：Decision Tree は筆者が関係した大阪国際癌センターの多群の ECG 診断や、決定木分析は介護保険システムの実用システムの実績がある。RIP の欠点は、多群判別に対応するのに手間がかかる。2 群から 7 群の Microarray を 169 の 2 群にして全て LSD であったのは、遺伝子の出す蛋白量の発現量が、癌と正常あるいは異なった癌腫を分ける最高品質の計測値であることを示す。

分類 4：他の判別手法は研究目的であり、適切な分析対象で実証研究すべきである。

2.4　数理計画法ソルバーLINGO を用いた 4 つのプログラム

シカゴ大学発の LINDO 社が開発した LINGO[4]で、LSD 判別のための RIP の 4 プログラムを開発した。これらは H-SVM、S-SVM、LP-OLDF 等も選択し実行できるので、これら全てで比較実証研究をして RIP の優れていることを確認した。

Program1 は、普通のデータ（n≥p）で MNM=k の p 変数以下の OCP を見つける。k が 0 の LSD を SM と呼ぶ。k が 1 以上の場合、データ固有の k 個の誤分類を特定でき、これを省けば全判別データが LSD になる画期的な判別手法である。LSD である事が分かった 169 の高次元 Array のような幅広データ（n<p）は、全 SM は n 個以下の RIP の判別係数になる。残りの（p-n）個の係数は 0 になる。即ち RIP は Array を判別するだけで、高次元の p 個からほぼ n 個の遺伝子の SM を選択できる。恐らく早く見つかる SM ほど癌診断に役立つ。

Program3 は、LSD を複数の SM に分割する。山中教授は iPS 研究で、遺伝子 DB を用いて 3 万個以上の遺伝子からラットの ES 細胞で活性化する 32 遺伝子を特定した。SM はこの 32 遺伝子に対応する。どちらも理工学研究が考えていない効率の良い FS である。

Program4 は、p 変数の組みあわせでできる（2^p-1）個の判別モデルから最小変数の BGS1 を見つける。BGS1 を手動で省き次の BGS2 を見つける。山中教授は 4 遺伝子を見つけるために、医学研究に反する 32 個から 1 個を省いた 32 組の 31 遺伝子を同時培養した。多能性幹細胞が形成された 31 個の複数の組の同時培養を繰りかえし最後に 4 遺伝子を特定した。この FS は、万能細胞ができるか否かの二者択一であり、LSD であるか否かの FS に利用できる。即ち統計や ML 研究で考えていない FS であるが、Program3 に比べて計算時間がかかる。

Program2 は、判別データを k 回コピーして擬似母集団の検証標本にする。これをランダムにソートし、k 個の学習標本に分割する。この k 個の ER の平均の M1 は、変数が増えると単調減少しフルモデルが最小になり FS には適さない。k 個の RIP で検証標本を判別し、k 個の ER の平均の M2 を計算する。M2 = 0%になるモデルは、MNM = 0 の LSD（検証法 1）よりも厳しい検証法 2 で評価されたことになる。これが母集団と標本の関係を示す正しい k 重 CV である。一般に用いられる LOO 法を利用した k 重 CV は母集団と標本の関係を無視している。

3　Fisher が LDF の評価に用いた Iris の誰も指摘していない LSD の重要な事実

(1)　Iris(150*4)の 3 組の LSD と iPS 研究の山中 4 因子と同じ役割の 3 組の BGS

Fisher が LDF の評価に用いた Iris は、各 50 例の Setosa(G1:赤)、Versicolor(G2:青)、Virginica(G3：緑)の 4 変数(花弁と萼片の長さと幅)の普通のデータ（n≥p）である。

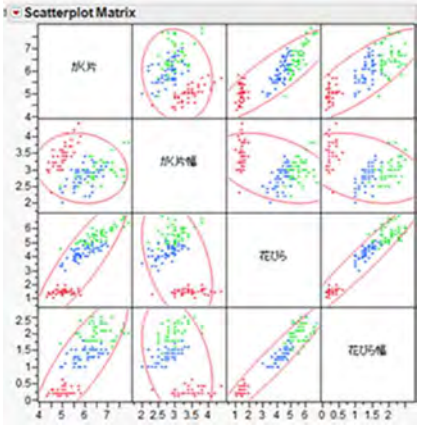


図 2　Iris の行列散布図

図 2 の散布図から、(X1,X2)が G1 と G23 を LSD に分ける最小の 2 変数の BGS で、X3 と X4 は 3 組の LSD (G1 と G2、 G1 と G3、 G1 と G23) を分ける最小の 2 組の 1 変数の BGS である。3 組の BGS は、恐らく山中 4 因子のように遺伝子が関係していると考える？同様に猿などの形態で種が分かる場合、遺伝子が関係していると考えられる。

Fisher の時代 DT は唯一の判別手法で、X3 あるいは X4 の NM は 0 になる。癌の遺伝子データも 1 変数の発癌遺伝子があり Program1 は SM を 1 秒程度で求まるようだ。しかし種々の LDF で Iris を判別しても、NM が 1 以上である。Fisher はこの事実を知っていたと考えるが、広い応用分野のある判別理論の重要性を優先したようだ。彼は組み合わせ最適化技術を知らなかったので、正しい LSD の判別理論を後世の研究に委ねたようだ。

（2） 4 変数の Iris は LSD の Nest 構造をもつ 14 体の Matryoshka 人形

[7]で Iris が LSD で Nest 構造になることを指摘した。4 変数の LSD は、その中により少ない変数の LSD を含み、3 組の BGS のどれかで終わる Matryoshka 人形になる。もし山中 4 因子が 1 変数であれば、膨大な Matryoshka 人形になり、規則正しい万能細胞ができないであろう。これが LSD 判別に重要なデータ構造である。RIP の Program1 で普通の LSD(n≥p) を判別すると、ほぼ p 変数以下の LSD の SM になる。Iris の 4 変数の SM は次の 14 体の Matryoshka 人形になる。

Matryoshka1: (X1、X2、X3、X4) ⊃(X2、X3、X4)⊃(X3、X4)⊃(X3)、

Matryoshka2: (X1、X2、X3、X4) ⊃(X2、X3、X4)⊃(X2、X3)⊃(X3)、

Matryoshka3: (X1、X2、X3、X4) ⊃(X1、X3、X4)⊃(X3、X4)⊃(X3)、

Matryoshka4: (X1、X2、X3、X4) ⊃(X1、X3、X4)⊃(X1、X3)⊃(X3)、

Matryoshka5: (X1、X2、X3、X4) ⊃(X1、X2、X3)⊃(X2、X3) ⊃(X3)、

Matryoshka6: (X1、X2、X3、X4) ⊃(X1、X2、X3)⊃(X1、X3) ⊃(X3)、 6 体の人形は X3 を含む

Matryoshka7 : (X1、X2、X3、X4) ⊃(X1、X2、X4)⊃(X2、X4)⊃(X4)、…..

Matryoshka12: (X1、X2、X3、X4) ⊃(X1、X2、X4)⊃(X1、X4)⊃(X4)、 6 体の人形は X4 を含む

Matryoshka13: (X1、X2、X3、X4) ⊃(X1、X2、X4)⊃(X1、X2)、…..

Matryoshka14: (X1、X2、X3、X4) ⊃(X1、X2、X3)⊃(X1、X2)、 2 体の人形は(X1、X2)を含む

Array のような高次元 LSD (n << p) の横長データ(n<p)は、Program3 で n 変数以下の SM に分割できる。各 SM は通常の LSD (n ≥ p) と同じ Matryoshka 人形になり小標本として分析できる。この事実を知らない理工学研究は、高次元データ分析は 1)分析手法がない、2)高次元の FS は Np ハードである、3)高次元の雑音から信号を見つけることは困難と、3 つの言い訳した。

（3）Program4 による山中 4 遺伝子と同じ最小の LSD である BGS の見つけ方の概要

4 変数の(X1、X2、X3、X4) から X1 を省いても(X2、X3、X4)は LSD である。次に X2 を省いた(X3、X4)も LSD である。そして X3 を省いた X4 は LSD で 1 変数の LSD の BGS 1 になる。次に (X1、X2、X3) から X3 が BGS2 になる。最後 に(X1、X2) から X1 を省くと LSD でないので残す。X2 を省いても LSD でないので 2 変数の BGS3 になる。この様な試行錯誤で Program4 は BGS を見つけるので時間がかかる。

(4) X1 と X2 だけが LSD でない事実は遺伝子が長さを決めたと考えれば癌研究の入門になる

LOO に習った間違った k 重 CV で、判別結果の ER で評価する方法が用いられている。筆者は判別データを k 回コピーし疑似母集団を表す検証標本とする。そして乱数で並べ替えて、k 組みの学習標本を作る。これを RIP で判別し、得られた k 個の ER の平均を M1 と呼ぶ。この k 個の RIP で検証標本を判別し、ER の平均を M2 と呼ぶ。これが正しい k 重 CV である。MNM=0 が判別結果の評価 1 で、M2 が最小のモデルを Best モデルとして選ぶのが評価 2 である。評価 2 は LSD を更に M2=0 であるか否かで差別化できる。

表 2 は、3 種の LSD のセトサとバーシクルの結果である。他の 2 群も同様である。Iris の 15 の判別モデルで、X1 と X2 だけが LSD でない。当然検証標本の 10 重 CV の M2 も 0%でない。他の 13 モデルは、LSD で、M2 も 0%になる。これは 4 変数、SM、そして 3 つの BGS の学習標本と検証標本全てが LSD である事を示す。恐らく 4 変数の計測値は、腫をきめる遺伝子と関係していて良い結果になったと考える。京大の猿研究でも、中国の隔離された地域の猿も体長と遺伝子の関係を指摘している。動植物の種は遺伝子情報が分かる以前から特長のある計測値で種が決められていたので当然の発見である。

表 2　15 モデルの 10 重 CV の驚く結果

	min	Max	mean	Min	Max	Mean	min	max	Mean	MIN	MAX	MEAN
1	7	13	10.9	7	13	10.9	110	140	113	11.00	14.00	11.30
2	13	17	15.6	13	17	15.6	160	200	170	16.00	20.00	17.00
3	0	0	0	0	0	0	0	0	0	0	0	0
4-14	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0

4.　3 種の通常の LSD

4.1　世界初の LSD 研究：スイス紙幣の真贋判定

Flury ら (1988) は、 真札/偽札各 100 枚の 6 変数(X1:幅、X2:左縦長、X3:右縦長、X4:下枠長、X5:上枠長、X6:対角線長)を測定し判別分析の本を出版した。RIP で 63 モデルを判別し、MNM の単調減少 ($MNM_k \geq MNM_{(k+1)}$)の Fact2 を発見した。(X4,X6)が BGS の MNM=0 で、BGS を含む 16 モデルだけが LSD で、残り 47 モデルは LSD でなく棄却できる。LSD か LSD でないかで強力な FS ができる。これが世界初の LSD 研究である。LSD 理論は Fact2 とべき乗計算が重用で、重回帰分析のような FS と検定による曖昧な確率による判断は必要ない。

表 3 は BGS を含む 16 モデルの LSD を示す。RIP 列は MNM=0 で、他の 4 個の判別関数の NM は 1 から 3 でバラつく。M1 は 100 重 CV の学習標本の ER の平均である。16 個の LSD で BGS1 のみが M2 = 0 で、他の 15 モデルは 0.05%から 2%と大きい。M2 = 0 は他の LSD と差別化できる。しかし紙幣の摩耗度は使用法で異なり、摩耗に影響される BGS は真贋判定に適さない。これは Iris の 13 の LSD が全て M2=0 であるのと大きな違いがある事でも分かる。真贋判定は、長さよりもインクと紙の質等が重要である。また X5 でなく X4 が選ばれる理由が分からないが 200 枚の紙幣に固有な結果と考える。著名な日本の計量経済学者である

竹内啓教授も、私の日本語の著書[34]の書評で LSD の可能性を疑ったようだ。このデータの日本での人気の説明ができないが、Fact2 等の最初の LSD の成果を得て、世界初の LSD 研究の出発点である。

RIP 以外の判別関数の NM は 1 以上で、誰も LSD を発見していない。有名な成蹊大学の統計の教授は 5%の間違いを認めるので数例の違いは誤差の範囲と批判するのに驚いた。RIP で判別し(X4,X6)が 2 変量の BGS1 である。M1 は(X2,X4,X6)以外 0 であるが、M2 は BGS だけが 0%で他は 0.05%から 2%である。即ち最小の 2 変数が他の変数の不要な情報に影響されないコンパクトで判別に有効なオッカムの剃刀を満たす具体的な Best なモデルである。即ち簡素なモデルで説明力のある BGS が一番重用であることを示す。統計の FS と検定ではこのような分かり易い判別結果は得られない。従来の判別分析は意味の無い ER を議論してきた。BGS に選ばれた変数の組みの意味を考えることが重要だ。表 1 で LSD であるか否かが検証法 1 で、その中から更に M2 が最小のモデルを良いと考えるのが検証法 2 である。

表 3　16 の LSD で M1 と M2 と 5 種の判別関数の NM

SN	P	var.	M1	M2	RIP	SVM1	LDF	QDF	RDA
1	6	1-6	0	0.2	0	1	1	1	1
2	5	2-6	0	0.1	0	1	1	1	1
3	5	1,3-6	0	0.05	0	1	1	1	1
4	5	1,2,4-6	0	0.15	0	1	1	1	1
5	5	1-4,6	0	0.2	0	2	1	1	1
8	4	3-6	0	0.2	0	1	1	1	1
9	4	2,4-6	0	0.05	0	1	1	1	1
10	4	1,4-6	0	0.3	0	1	1	1	1
11	4	2-4,6	0	0.1	0	2	1	1	1
12	4	1,3,4,6	0	0.2	0	2	1	1	1
13	4	1,2,4,6	0	0.2	0	2	2	1	1
23	3	4-6	0	0.1	0	1	1	1	1
24	3	3,4,6	0	0.2	0	2	1	1	1
25	3	1,4,6	0	0.4	0	2	2	2	1
26	3	2,4,6	0.9	2	0	2	1	1	1
43	2	4,6	0	0	0	2	3	1	1

4.2　試験の合否判定：自明な LDF が自明な LSD になる判別分析の優等生

2 得点の合格点が 50 点の場合、自明な LDF = T1 + T2 – 49.5 は自明な LSD になる。大学入試センターから 3 年間 13 科目の研究用データを借用し、得点分布の 3 水準(90%、50%、10%) で分析した。理系の数学 2B の 90%で、LDF の ER は 0%でなく 32%である。また 90%が合格点の場合、全合格者が特定の問題に正解し不合格者がバラつく場合、一般化逆行列の瑕疵で QDF は全合格者を不合格群に誤判別する。

国語は 80%ぐらいの設問数の 1 組の BGS に、数学は小数の設問の数組の BGS になる。他の教科は、これらの間の BGS の組み数になる。2011 年のある私大における試験用紙配布ミスのため、センターは筆者の 3 研究論文の撤回を求めたので自分の統計学入門の 10 択 100 問の中間試験と期末試験で研究を再開した。

000119

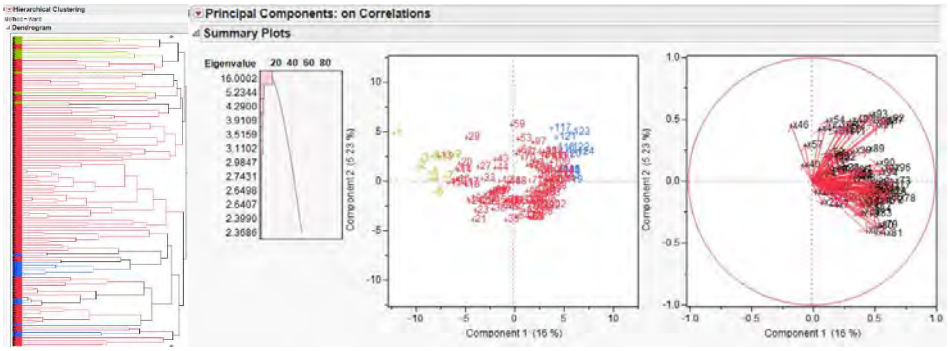


図 3　124 人の受講生の 100 問の Ward 法と PCA の結果

図 3 は 2019 年の中間試験の 90%以上を青で、10%以上を赤で、未満を緑で示した Ward 法と PCA の出力である。赤の 80%の学生の中に青と緑が幾つかのクラスターに分かれている。散布図から分かるように、合否判定は自明な LDF が自明な LSD になるが、僅か 1 点差で人為的に決めた判別超平面の近くに両群の学生がいる特殊なデータ構造である。このため判別境界で向かい合った 1 点差の学生同士の方が、同じ群の他の学生より距離的に近いという特長がある。

表 4 で 90%点の 5 組の BGS を 100 問や SM1 と比較した。5 つの BGS は 63 問あり、BGS1 は 8 問で最小で、M2 は唯一 0%でオッカムの剃刀をみたす Best モデルである。BGS3 の 4.4%でも SM1 の 9.2%や、100 問の 10.5%より小さい。以上からたった 8 門の BGS1 が合否判定に一番有効である。設問数の多い 100 問や 77 問の SM1 や 4 個の BGS は、LSD であっても合否判定にそれ程役に立たないことが出題者の筆者に初めて分かった。以上から合否判定の研究は、評価する側とされる側にとって社会的に必要であり、LSD 研究に良いデータを提供してくれる。なぜ評価する教員や教育機関が、今になっても取り組まないのか理解できない。RatioSV は、RIP の SV 間の距離を 2 群の範囲で割って求める。T2 が 50%と 2 群を一番分けているが、M2 は 9.2%と大きい。

表 4　　90%点の SM 分割と BGS 分割を検証標本の 10 重 CV で検証した結果

Type	var	min	max	mean	MIN	MAX	MEAN(M2)	Ratio	MNM
100var	100	0	0	0	6.5	18	10.5	25.24	0
T1（基礎統計）	29	0	1.6	0.6	6.5	16	9.1	5.02	2
T2（相関）	12	7.3	9.7	8.9	8.9	9.7	9.2	50.00	11
T3（分割表）	19	0	0	0	3.2	9.7	6.0	33.33	11
T4（単回帰）	40	0	2.4	0.8	7.3	11	8.7	33.33	11
SM1	77	0	0	0	6.5	12	9.2	23.98	0
SM2	23	2.4	7.3	4	8.9	14	11.0	5.06	6
BGS1	8	0	0	0	0	0	0.0	5.88	0
BGS2	11	0	0	0	0.8	4	2.7	2.00	0
BGS3	13	0	0	0	1.6	10	4.4	3.33	0
BGS4	13	0	0	0	1.6	7.3	4.1	3.33	0
BGS5	18	0	0	0	0.8	7.3	4.3	1.67	0
bgs6	37	1.6	6.5	3.5	8.9	17	12.0	10.9	6

4.3 標準車 29 台（y_i=1）と小型車 15 台（y_i=-1）の判別

(1) 法律で定義された 2 つの最高品質の BGS

法律で、小型車は排気量 2000cc 未満（X1）と 4 人乗り（X3）と定められている。標準車と小型車は X1 と X3 の 2 つの BGS で LSD になる。表 5 は、Plug-in 規則で従属変数 y_i=1/-1 を用いた重回帰分析の変数増加法の結果を示す。変数は p の順に追加される。t 値が大きい X1 と X3 が 2 つの BGS であるが、FS は p＝2 で X1 と X3 の 2 変数を選択できない。LDF の NM は 2、1、1、1、0、0 と単調減少するが、2 つの BGS を特定できない。QDF は、2 群が異なる正規分布を前提としていて p=1 と p=2 の NM は 0 です。しかし 3 変数に X3 が含まれると、QDF の SVD の欠陥で小型車が標準車に誤分類される。JMP は対応策がなく小型車に全車の分散共分散行列を用い、3 変数以降も 0 にした。RIP は X1 が BGS1 であり、BGS1 を含む 6 つの MNM 値は 0 です（事実 2)。正則化判別分析は、2 つの λ と γ を[0,1]の範囲で 10×10 の 100 個の格子点を探索し、λ = γ = 0.1 で LSD を見つけた。例えば 0.8 の場合、6 つの NM は表に示すように悪い。筆者の実証研究で S-SVM のペナルティ c は、1000 にすればほぼ LSD を発見できる。以上から面倒な 正則化判別分析より S-SVM が優れている。しかし RIP で LSD である SM と BGS が直接分かるので、重回帰分析の FS や t 検定、そして他の判別手法は不要になる。

表 5　判別理論の瑕疵（Problem3）の説明

p	Var.	t	LDF	QDF	MNM	λ=γ =0.8
1	Emission (X1)	11.37	2	0	0	2
2	Price (X2)	5.42	1	0	0	4
3	Capacity (X3)	8.93	1	29	0	3
4	CO_2 (X4)	4.27	1	29	0	4
5	Fuel (X5)	-4.00	0	29	0	5
6	Sales (X6)	-0.82	0	29	0	5

表 6 は 10 重 CV で 63 モデルを直接評価し、次の結果を得た。

SN=1 と 2 は RIP=5.92＊X1-4.893 になるが、 6 変数の M2 は 0.023 で 5 変数より悪い。X6 が無駄であり、変数が少ない方が良いモデルであることを示す。また一定値の X3 より X1 の方が選ばれる。SN=2 は X6 を省くと SN=1 と同じモデルになるが M2=0 になる。即ち X6 という営業成果を示す変数は、車自体の特性値より M2 で販売台数の品質が悪いことが分かる。このような事実は、表 5 の重回帰分析の統計分析で分からない。

SN が 3 から 48 までの 46 モデルは、X1 を含む 28 の RIP = 5.92*X1 -4 と X3 を含む 20 の RIP = 2*X3 -9 に分かれ全て M2=0 である。この詳細な理由は分からないので、後世の研究にまかせたい。

2 つの BGS を含まない SN が 49 から 63 までの 15(=2^4-1)モデルは、LSD でなく M2 は 1.295 以上であり棄却できる。

以上から法律が決めた 2 つの BGS が、同じ判別分析あっても統計の判別分析では分からない新しい事実が分かった。1）変数の少ないモデルの方が良い、2）値がバラつく BGS の方が一定値の BGS より説明力がある、3）自動車の特性値の方が営業結果などの販売台数より判別に影響する、4）正規分布を仮定した無限の判別関数は重要な情報を提供せず、役にたたない。

以上より、法律で決めた BGS の方が遺伝子より高品質と考えた。しかし Iris のような単純な結果でないのは、価格、販売台数等の各車の営業的な変数があるためと考える。法律が決める BGS は多くあり、将来の統

計研究に良い研究テーマである。

表 6　RIP の 6 変数からできる 63 モデルの判別係数と BGS と M2 の関係

SN	Models	Var.	BGSs	MNM	M2	x1	x2	x3	x4	x5	x6	C
1	Full	1-6	1	0	0.023	5.92	0	0	0	0	0	-4.893
2		1-5	1	0	0	5.92	0	0	0	0		-4.893
12		1,3,6	1	0	0	5.92		0			0	-4.893
13		1-3	3	0	0	0	0	2				-9
14		1,3,4	3	0	0	0		2	0			-9
15		1,3,5	3	0	0	0		2		0		-9
16		1,3	3	0	0	0		2				-9
17		3,5,6	3	0	0			2		0	0	-9
23	SM1/BGS1	3	3	0	0			2				-9
24		2-6	3	0	0		0	2	0	0	0	-9
32		3-5	3	0	0			2	0	0		-9
33		1,2,4-6	1	0	0	5.92	0		0	0	0	-4.893
40	SM2	1,2,6	1	0	0	5.92	0				0	-4.893
48	BGS2	1	1	0	0	5.92						-4.893
49	bgs3	2,4-6	0	3	1.295		0		-46	-199	0	5342.8
60	SM3	4,5	0	10	2.841				3.54	10.8		-601.3
63		6	0	13	3.205						-0.7	6773.5

5　CPD による医学診断に革命的な「ケース選択法 4」

5.1　多重共線性のある 240 例 19 変数の児頭骨盤不均衡データ[30]

日本医大の鈴村名誉教授は、表 7 の 17 計測値から作った 2 変数 X9＝X7-X8 と X12=X13-X14 で、自然分娩 180 例と帝王切開 60 例の手術法を決める鈴村氏法を提案した。筆者が 1978 年に統計的に検証した。X9 と X12 は、多重共線性を示す VIF が 21.9 と 1484 と大きい。SAS の RSQUARE は、y_i を重回帰分析の目的変数にした(2^19-1)の 52 万以上の 1 変数から 18 変数の重回帰モデルで、決定係数の大きい上位 19 モデルの 343　(= 19*18＋1)個の R^2 を掃き出し演算子で出力する。この 343 モデルから、図 4 のように変数増加法 F と減少法 B で、AIC と Cp 統計量等で 16 変数（M2＝3.55）を選び、数編の医学と統計論文を出したが、 この 16 変数が良いか否かの確かな根拠がなかった[秀新3. 1]。

表 7　多重共線性のある 19 変数の CPD240

Var.	Description	VIF
X1	age of the pregnant woman	1.2
X2	the number of times delivery	1.3
X4	the utero-posterior distance at the pelvic inlet	24.6
X7	the shortest Anteroposterior distance	57.0
X8	Biparietal fetal diameter	5.3
X9	X9=X7-X8	21.0

X10	the utero-posterior distance at the pelvic inlet	3.7
X11	biparietal diameter at the pelvic inlet	1.7
X12	X12=X13-X14	1484
X13	the area at the pelvic inlet	1466
X14	the area of the fetal head	638
X15	the area at the bottom length of the uterus	1.4
X19	lateral conjugate	1.4

図 4 の左は 1 変数から 19 変数の JMP の NM の折れ線グラフで、右は SAS の折れ線グラフである。赤が F で青が B で両方の NM は 6 変数から 12 変数で大きく異なるので、多重共線性のないモデル選択を最優先にした。決定係数の大きい 343 モデルを選び、その中から FS で AIC の小さい 16 変数を選び多重共線性がないことを期待したが自信が無かった。しかし 2025 年に、RIP が 14 変数の多重共線性のない 3 組みの BGS を選んだ。BGS は最小変数の LSD であるので、多重共線性のある無駄な変数の組みを選ばないことが分かった。X12 を含む BGS1 は最小の M2=2.44 になり X13 と X14 を含む BGS2 と BGS3 より良いので、VIF の大きな X12 が X13 と X14 の両方の情報を含んでと重用であり、VIF の小さな X9 は不用である事が分かった。

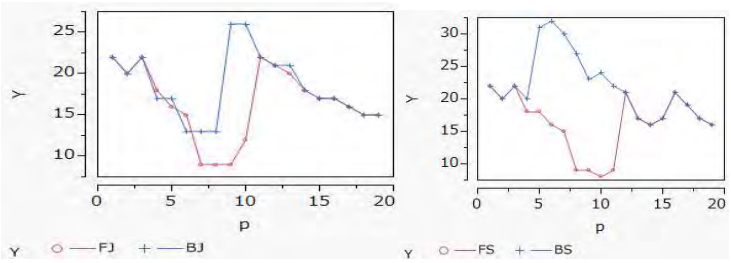


図 4　JMP (左) と SAS (右)の変数増加法（赤）と減少法（青）の NM

5.2　ケースの選択法 4：判別理論の福音

Springer[23]を執筆中、4 種の普通の LSD を見直し、CPD の扱いに苦慮した。突然 RIP の MNM が k の場合、データに一意な k 個の誤分類を見つける重要性に気づき、「ケースの選択法 4」を開発した。

Program1 で判別し、 2 ^ 240 乗の Hi と e_i の 2 種の組み合わせ最適化で、MNM=2 の凸体を選ぶ。この 2 例を省くと、全判別データが LSD になる。238 例の LSD から、Program4 で 1 変数から 19 変数の(2 ^ 19 – 1)の判別モデルから、山中 4 因子を選ぶ変数減少法で最小の 14 変数の BGS1 を選んだ。X12 を含み X13 と X14 を含まないので、多重共線性が自然に解消された。BGS は冗長性のない最小の LSD であり、多重共線性のある無駄な 3 変数を選ばないことに気づいた。そこで X13 だけを含む BGS2 と、X14 だけを含む BGS3 の 2 組の BGS がある事を無駄な計算をしないで確認した。RIP で BGS が簡単に見つかり、判別データを Plug-in 規則で重回帰分析する必要がない。2024 年に、Program1 で MNM=2 を見つけ、ケースの選択法 4 で 238 例の LSD から Program4 で多重共線性のない 14 変数の BGS を見つけた。そして Prigram2 の 10 重 CV で M2 が最小の Best なモデルであることを確認し、53 年かけて全ての医学診断に役立つ診断法を完成した。

これ迄多くの医学診断で LDF 等の判別手法が用いられてきた。現在も利用されておれば、RIP で見直すべきである。表 8 は 1978 年に検討した 19 変数の 52 万モデルから、各次元で R^2 が大きい上位 19 モデルから変数減少法 B と増加法 F で 38 モデルを選んだ。そして AIC、BIC、Cp で検定して、多重共線性のない 16

変数の重回帰モデルを選んだ。表の MNM 列は単調減少する。OCP は最適凸体の個数である。悪戦苦闘して多重共線性のない 16 変数モデルを選んだが当然 LSD でない[30]。しかし 5.3 で示すように RIP は 2025 年に数時間で多重共線性がない Best な 14 変数を選んだ。

表 8　52 万以上の判別モデルの FS

P	Rank	F/B(JMP)	AIC	BIC	C_p	RIP	OCP	JF	JB	F/B(SAS)	SF	SB
1	1	FB12	110	120	21	20	1	22	22	FB12	22	22
2	1	FB9	94	108	5	13	3	20	20	FB9	20	20
3	1	FB18	93	110	3	12	2	22	22	FB18	22	22
4	1	F15	93	113	3	10	2	18		FB15	18	20
	3	B2	93	114	3	9	1		17			
14	1	B11	102	155	11	3	10		18	FB11	16	16
	2	F4	102	155	11	3	16	18				
15	1	B19	104	160	12	2	1		17	FB16	17	17
	2	F11	104	160	12	3	1	17				
16	1	F19B16	106	165	14	2	2	17	17	FB8	21	21
17	1	FB6	108	171	16	2	8	16	16	FB10	19	19
18	1	FB3	111	176	18	2	15	15	15	FB6	17	17
19	1	FB8	113	182	20	2	16	15	15	FB3	16	16

5.3　M2 が最小の 14 変数の BGS1

Program1 で 240 人 19 変数の CPD を判別すると、MNM＝2 の最適凸体を B&B 法で 1 秒で見つけた。表 9 は 2 例を省いた 238 人の LSD の結果である。19 変数の 10 組の学習標本の平均 ER の M1 は 0%だが、検証標本の M2 は 2.86%である。Program1 で判別すると、実際は 1 変数の係数が 0 で 18 変数の SM1（M2=2.9%）が見つかる。SM1 の M1=0 であるが M2=2.9%で 19 変数より悪くなるので、19 変数より良いとは判断できない。Program4 は X12 を含む 14 変数の BGS1 を選んだ。M2＝2.44%で恐らく 524,287 モデルの中で最少値と考える。残りの 5 変数の bgs2 は LSD でなく、M2=6.97 と大きい。即ち M2 は 19 変数を意味のある 14 変数の BGS1 と役に立たない 5 変数の bgs2 に分けることで、19 変数の M2=2.86 が、14 変数の最小の M2=2.44 と、5 変数の最大の M2=6.97 に分解できた。これは BGS1 が、哲学的な「オッカムの剃刀」を満たす Best なモデルである事を具体的に示す、世界初の事例である。多重共線性の関係から、X13 と X14 を BGS に固定すると、他の 2 変数を含まない BGS2 と BGS3 が見つかる。しかし BGS2 と BGS3 の M2 は BGS1 より大きい。以上から、BGS1 は最小変数の冗長性のないモデルで、自然に多重共線性を排除する。半年以上かけた重回帰分析の研究は、16 変数（M2＝3.55）を選び、結果が悪い。以上から M2 最小モデルを選ぶ検証法 2 が、遺伝子診断でも最適な判別モデルになると考えた。検証法 3 の Ward 法と PCA で、2 群が綺麗に分かれる事を確認した。しかし遺伝子診断は未知の事が多く、M2=0 であっても散布図で 2 群が 2 クラスターに綺麗に分かれない。

表 9　10 重 CV の CPD238 の明快な結果

Model	Var.	min	max	mean	MIN	MAX	MEAN	min	max	mean	MIN	MAX	MEAN
19Var	19	0	0	0	0	0	0	40	100	68	1.68	4.20	2.86

SM1	18	0	0	0	0	0	0	50	100	69	2.10	4.20	2.90
BGS1	14	0	0	0	0	0	0	30	120	58	1.26	5.04	2.44
ｂｇｓ2	5	4	14	9.5	1.68	5.88	3.99	140	190	166	5.88	7.98	6.97
16Var	16	0	0	0	0	0	0	30	110	67	1.26	4.62	2.82

6. 癌の遺伝子診断の基礎研究（2015 年～2018 年）[14]

6.1 米国の 6 Microarray 研究にみる医学と理工学の遺伝子研究の間違い

2015 年 10 月 28 日に Higgins 教授の HP から 2004 年迄に計測された Microarray を入手した。Jeffery ら（2006）は t 検定や相関等で FS を行い遺伝子の組みを選んだが、SM や BGS のような「多変量の癌遺伝子の候補」ではない。表 10 は症例数と遺伝子数と間違った FS の WVM 等や非階層型クラスター分析の SOM 等の後、筆者が JMP が開発した SVD を使った高次元 LDF の ER が 2.34%以上で役にたたない事、SM と BGS の個数を示す。統計や Machine Learning 等の研究者も同じ間違った研究をして、癌研究に悪影響を与えた。更に「癌研究が既に見つけた 1 変量の発癌遺伝子と、Array で見つけた新しい多変量の癌遺伝子の候補の関係」を誰も検討していない。

Chiaretti らだけが、既に見つけた発癌遺伝子を組み合わせて研究し NIH の研究費を貰った。他の研究は問題のある FS で、50 個程度の一組の遺伝子の組を選ぶ致命的な間違いをした。たった一組の多変量の発癌遺伝子で不均質な癌を解明できないことは素人の私でも分かる。また既に見つかっている 1 変量の癌遺伝子との関係を無視している。1970 年頃から Array 研究をしていて誰もが第 1 人者と認める Golub らが NIH の研究費を貰わなかったことで Golub ショックが走り、2004 年以降の第 2 世代の Array 研究では、データ解析的な方法論が行われなくなった。

表 10　第一世代の古い Microarray の概要

	Golub (1999)	Shipp (2002)	Singh (2002)	Chiaretti (2004)	Alon (1999)	Tien (2003)
3 原則	—	—	1，2，3	3	1	1
G1	25 AML	19 Follicular	50 Normal	33 T-cell	22 Normal	36 False
G2	47 All	58 DLBCL	52 Prostate	95 B-cell	40 Colorectal	137 True
n*p	72*7129	77*7129	102*12625	128*12625	62*2000	173*7129
ER by LDF	11.11%	3.89%	5.58%	2.34%	7.81%	16.76%
FS[1]	WVM	WVM	WVM	legacy Oncogene?	Correlation	t-test,
Method	SOM	SOM	SOM	ML と LDF	Aneeling	logistic
SM2	69	213/237	179/139	270	64	159/104
BGS					130	204
Type(SM/Signal)	Type-a / -A	Type-a / -A	Type-a / -A	Type-a / -A	Type-a /-A	Type-a /-A

6.2　54 日間の快挙（2015 年 10 月 28 日～12 月 20 日）

10 月 28 日に RIP で Shipp の 2 種の白血病を判別した。驚くことに 1 s で LSD が分かり 32 個の遺伝子の SM1 の判別係数が非零で、残り 7097 個が 0 になった。IP のアルゴリズムの B&B 法は DT と同じく、1 個の発癌遺伝子を含む SM1 をすぐに判別したと考える。図 1 の Fact1 から 77 個以下の遺伝子が選ばれるが、計測値に冗長性があるため 32 個になったと考える。当日に Research Gate に最初の Report をあげ、54 日間で 15 の報告をした。29 日以降、SM1 を省いて手作業で 10 組み目の SM10 を見つけ、Array が多くの SM に

分割できる事が分かり LSD を SM に分割する Program3 を開発した。そして 2016 年に 山中 4 因子を見つける変数減少法で Program4 を開発し BGS を見つけた。最初は SM 毎に BGS をすぐに見つけ、残りの遺伝子をまとめて判別する事を繰り返した。この方法は計算は速いが煩雑である。しかし医師は個々の BGS を医学的に慎重に検証して、次に進めるので結果を確認しながら分析をくり返せばよい。筆者はこの方法を途中でやめ、10 個の SM に含まれる約 2000 個程度の遺伝子から BGS を見つけることにした。計算時間がかかるが、PC の結果を待てばよい。これ以降の 7 章で BGS の結果を示す。

6.3 6 研究の問題点

Singh が、50 例の正常と 52 例の Prostate のほぼ同数で研究している。このようにコントロール群の正常と癌の比較をほぼ同数で行う医学診断の原則が、多くの癌研究で無視されている。さらに Chiaretti 以外の 5 研究は、間違った FS で、たった 1 組の 50 個程度の遺伝子の組を多変量の発がん遺伝子と考え、間違った非線形クラスター手法で評価した。最近の多くの癌研究でも非線形クラスター手法で多群のケースと遺伝子のカラーマプで特定の癌が分かれているという報告が多い。これらの結果は、注目する 2 群に限定し LSD を確認する必要がある。米国の医学プロジェクトは統計学者を雇い分析をまかせている。彼等は自分の興味のある手法に限定し研究していることに問題がある。筆者のように可能な手法を比較して良い手法を選んでいない。NEJM に掲載された Tien らは唯一ロジスティク回帰を用い、クラスター分析と PCA で分析しているが LSD を見つけていない。Chiaretti らだけが医学的な知識で複数の発がん遺伝子を組み合わせて研究し、ML 研究で判別などをしたようだが詳細は不明であるが NIH の研究費を貰った。他の 5 研究は貰わなかったので、第 2 世代以降ではデータ解析的な研究が行われなくなったようだ。

7 バイオ工学で品質管理された CuMiDa の 78Array (n=[12、357]、 p=[12,621、54,676]) [23]

7.1 RIP は Machin Learning で選ばれた 8 Classifiers より優れている

2019 年に RG から、ブラジルのバイオ工学の Buruno ら（2019）が、2004 年以降の 1 万遺伝子以上の第 2 世代の新しい Array を登録した CuMiDa という遺伝子 DB を開発したメールが届いた。バイオ工学技術で品質試験し、問題のある症例を修正/削除している。統計分析で行えない工学的な一種のケース選択である。ML 研究が使うフリーソフトの Weka の 8 種の Classifier で、高次元 Array を直接分析している。これは数学や統計研究で SVD を用いて高次元の LDF や PCA を開発し、遺伝子診断を行うのと同じ間違いである。1 万変数の LDF をどうやって医学診断に用いるのか考えていない。CT のように高次元で一たを画像診断しているのとおおきな違いがある。

5 種の 1 群、正常のある 57 の 2 群、3 群から 7 群の 16 の Array である。57Array だけがほぼ同数の正常との 2 群判別で分析に値する。1 群は例数も少ないの分析しない。3 群以上は例数が少なく正常がないものもあるが、これらから 106 組の 2 群データを作り RIP の結果と比較した。

表 11 のように 3 重 CV で各 Array の手法毎の正信率の優劣を評価している。前代未聞の 3 重 CV の利用は、間違いなく 20 例以下の Array に合わせたと考える。これは間違いで、症例数の少ない Array は単に記述手法の分析に限定すべきである。さらに ML 研究は、判別手法とクラスター分析等を同一視していて問題があるがクラスター分析の結果が明らかに悪くて参考になる。そして多くの Array で、kernel-SVM、ランダム・フォレスト（RF）、MLP`が良いとしている。RIP で SM の M2 を求め最小の M2 が 0 であれば、正信率が 1 の結果より遥かに良いことを示す次の結論を得た。

1．1 群の 5Array は省く。57 の 2 群は正常例があり、医学診断に適している。3 群以上の 16Array から 106 組の 2 群データを作り、第 1 世代と併せた 169 の Array で LSD を確認した。3 群以上から機械的に 2 群を作ったのは問題であるが、取りあえず 2 種の異なった癌種でも LSD である事を示せた。この結果が遺伝子診断の重要な事実であるが、筆者以外誰も指摘していない。

2．169Array は症例数が 12 件から 357 件と大きく異なる。少ない Array の評価のため、ゆるい 3 重 CV の評価を用いたのは間違いである。

3．Classifier は、教師ありの判別手法と教師無しのクラスター分析等を同一視している問題がある。一方、高次元 Array の散布図で視覚的に 2 群が分かれているのに、LSD の可能性を指摘していない。

4．多くの Array で、kernel-SVM と MLP の正信率が 52%から 100%、ランダム・フォレスト（RF）が 35%から 100%で他の 5 手法より良いと結論している。この結論は納得できる。多くの ML 研究が行う Array 毎にどの手法が良いかという研究は、その理由が分からなければ意味が無い。

5．一方 RIP は、169 の Array 全てが LSD である事を示した。その上で 10 重 CV で全 SM の M2 の最少値が 0%から 12.24%である。表 11 の多くの Array は 0%である。169 の Array と SM が LSD であっても、M2 が 0%と 0%以上の Array に分かれる。0%以上の Array の症例設計に、問題があることを見つけた。

表 11　2 群の第 2 世代 57 Microarray

TYPE	n	GENE	SVM	MLP	RF	MinM2	TYPE	n	SVM	MLP	RF	MinM2
Max	357	54676	1	1	1	12.24	Max	357	1	1	1	12.24
Min	12	12621	0.52	0.52	0.35	0	Min	12	0.52	0.52	0.35	0
Pancreatic	51	54676	0.86	0.78	0.82	0.45	Prostate3	31	0.52	0.52	0.35	5.81
Breast1	16	36623	1	0.88	0.94	0	Prostate4	49	0.94	0.96	0.92	0
Breast5[47]	139	33580	1	1	0.99	0	Prostate9	124	0.67	0.65	0.69	9.76
Breast6	116	54676	0.99	0.99	0.97	0	Ovary2	24	1	1	1	0
Breast8	289	35982	0.93	0.7	0.86	6.51	Ovary3	15	1	1	1	0
Breast10	43	33580	0.95	0.95	0.98	0	Bladder1	85	0.64	0.58	0.55	12.24
Liver1	91	54676	1	0.99	0.95	0	Lung1	51	0.96	0.96	0.98	1.37
Liver3	357	22278	0.97	0.8	0.96	0.11	Lung3	114	0.93	0.85	0.92	1.75
Liver10	41	22278	1	0.98	1	0	Renal3	143	0.83	0.83	0.85	0.04
Throat1	103	54676	0.87	0.83	0.87	0	Renal4	20	0.85	0.8	0.85	0
Throat4	35	32784	1	0.97	0.97	0	Colorectal5	52	0.96	0.94	0.96	0
Leukemia1	71	33580	1	1	1	8.01	Colorectal6	63	1	1	1	0
Leukemia8	52	22284	0.98	0.98	0.92	1.51	Colorectal8	33	1	1	0.97	0
Prostate2	115	12647	0.64	0.63	0.66	9.91						

7.2　100 例以下の判別結果が良い驚く発見（Fact4）

表 11 の正常群がある 2 群の 57Array から表 12 の驚く事が分かった。100 件以下の 41Aray は SM1 が M2=0 は 31 個で比率 1 は 76%と多く、Classifier の複数の正信率が 1 の比率 2 が 75.6%と多い。一方、100 例未満の比率 1 は 31%で比率 2 が 12.5%と少ない。推測統計学では、症例数が多いほど結果が良くなると教育されてきたが、これが否定された！

最終的に「発現量は医学診断で用いる最高品質の計測値であり、100 例以下の症例数の研究通い結果であった。しかし白血病 1(71 例)の 3 つの正信率が 1 で、M2 は 8.01%と大きい Array もある。これは 3 重 CV という緩い評価法の問題であり、この結果で 3 手法が優れていると考えられず、10 重 CV で再評価すべきである。RIP で 3 重 CV で評価する必要はない。

表 12　57 の驚く結果

	Array 件数	M2=0	比率 1	複数の正信率が 1	比率 2
n<100	41	31	76	31	75.6
n>=100	16	5	31	2	12.5
	57	36	100%	11	

これ迄の医学診断は、正常（コントロール群）と疾病群の 2 群(Principle1)を、ほぼ同数(Principle2)で、100 例以上の標本数 (Principle3)が望ましいという「症例設計の 3 原則」を考えてきた。しかし発現量に限って 100 例以下の中規模の標本でも良い(新しい Principle3)という事実を得た。これを「癌症例設計の 3 原則」と呼ぶ。これは恐らく遺伝子の出す発現量が、医学診断が出会った最高品質の計測値であり、多くの症例を必要としないと考えた。また癌研究で多くの症例を集めた Array 程、医師は研究目的の違う質の違った症例を研究費の範囲内で集める傾向があり、例数が多いと結果が悪くなると考える。これを説明するために、最良のデータと思われる Liver3、不均衡なデータで分析前から悪い結果が想定できる Breast5 、50 例程度の中規模な癌と正常の Colorectal5 と 6 に焦点を当て、3 つの重要な研究を行った。

しかし自動車のように法律で決めた分類が、M2 で評価して最高であり、Iris の 4 変数は恐らく遺伝子が関係していて、Array と同じく高品質の第 2 順位であると考えた。

7.3　5 つの Array による 3 つの重要な研究

7.3.1 正常群を含む 2 群の Array による新しい「癌症例設計の 3 原則」の発見

表 13 の 5 つの Array で「癌症例設計の 3 原則」を確認する。Breast6 のみが正常例が少なく不均衡なデータで原則 2 を無視して結果が悪いことは事前に分かる。Liver3 は、ほぼ同数の 357 例で「症例設計の 3 原則」を満たし、データ分析の観点から最良のデータである。しかし M2 は他の Array が 0%であるのに、0.11%以上と悪い。恐らく多くの Genome Cohort（検証 5）で評価しているので、研究目的の違う症例の集合と考えられる。症例数が一番多くても、 Liver3 の結果は悪い。

表 13 正常患者と癌患者の 5 つの Microarray

TYPE	GSE	N (Cancer: Normal)	GENE	SVM 正診率 (ER)	RF	MiNM2
Liver3	GSE14520	357 (181:176)	22278	0.97 (3%)	0.96 (4%)	0.11%
Breast6	GSE42568	116 (101:15)	54676	0.99 (1%)	0.97 (3%)	0%
Colorectal6	GSE8671	63 (31:32)	54676	1 (0%)	1 (0%)	0%
Colorectal5	GSE25070	52 (26:26)	24527	0.96 (4%)	0.96 (4%)	0%
Renal1	GSE66270	28 (14, 14)	54676	1 (0%)	1 (0%)	0%

以下の 3 研究で、Array に限って症例数が 50 件前後の中規模標本でも良いという仮説を確認する。また Renal1 のように 28 例と少なすぎると、M2=50%という停止則が機能することが偶然分かった。以上から次の癌症例の収集指針を提案する。1）各 15 例程の症例で M2 が 0 を確認し、2）結果がある程度良ければ同

000128

じ症例収集の方針で 10 例ずつ増やして分析を繰り返して M2=0%になる 50 例程度の良い結果を得る。3）一気に 100 例以上を集めて M2=0 を確認する研究は効率的で無い。

この仮説を以下の 3 研究で検討する。SM1 の M2 と散布図のみを示す。

7.3.2　研究 1　NIH の Liver3　(181,176）：Bruno らが品質検査で削除したので同数でない

（1）一番望ましい Array が M2 の検証 2 や散布図の検証 3 の結果が悪い理由

図 5 の Liver3 は、肝癌と正常がほぼ同数で、357 例と例数が多く、統計的に「症例設計の 3 原則」を満たす一番良いデータである。しかし Ward 法で、赤の癌と青の正常が白で多くの小さなクラスターに分かれた。これは小数の性格が少し異なる多数のグループがあることを示す。PCA の散布図は症例数がほぼ同じであるが 2 群が Prin1 の正と負で綺麗に分かれていない。これは 2 群が異質の亜種の集合の可能性を示す。

Genome コホートの検証データ（検証 5）が幾組かあり、研究費が潤沢で研究目的の異なる症例群を集めたと考える。研究費の少ない研究でも、Microarray の計測費用が高いので研究目的の異なる癌や正常例を集める傾向がある。それ以上に癌研究者は、色々な技術を習得する必用があり、Golub らの第 1 世代の失敗の悪影響で、データ分析の重要性を考えないで症例を集めているようだ。

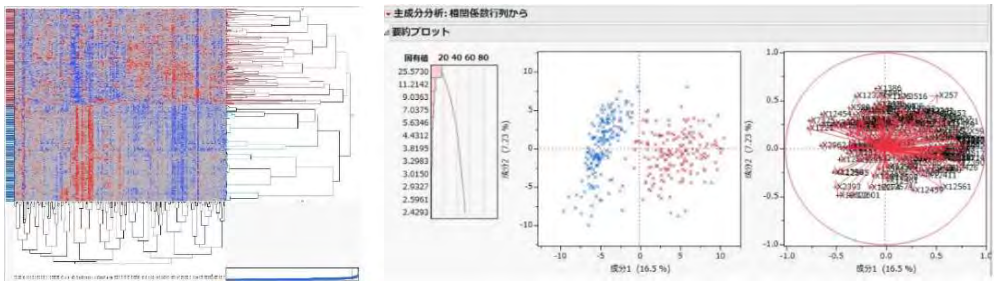


図 5　Liver3 の Ward 法と PCA

（2）各 20 例の Liver40 は 42 の M2=0 があり、Liver30 で M2=50%の停止則が働く

典型的な癌患者と正常患者は、RIP 得点や PCA の Prin1 得点の絶対値が大きいと考えた。RIP の絶対値が小さい症例を順次省くと良い結果が得られるのは当然である。そこで判別が目的でないデータのバラツキを評価する Prin1 で表 14 の 357 人から Prin1 の値の小さい症例を順次省いて 7 データを選択した。

Liver60 の SM1 に含まれる 85 個の BGS に 21 個の M2=0 があり、M2 の範囲が[0%, 9.7%]である。Liver40 は 99 個の BGS に 42 個の M2＝0 があり範囲は[0, 44.5]である。最大値は 44.5%と大きくなり、50%の停止則に近づく。Liver30 の 2 群の散布図は綺麗に分かれるが、M2 は全て 50%で 30 例の小標本で停止規則が働く。

Liver40 と Live30 の患者 10 名の違いは、M2 の評価に大きな違いになる。また遺伝子数が 1 個の BGS は Liver40 で 22 個あり、これが既に発見されている 1 変量の発がん遺伝子か調べる事が今後の癌研究に重用だ。一方、症例数が 357 で BGS は 20 個から 47 個の遺伝子を含むので、Golub らが 72 例の白血病の 2 群で約 50 個の遺伝子を見つけたのも仕方が無いかもしれない。筆者は、肝癌の 2 種の亜種を識別する固有の 1 変量の発癌遺伝子があるのかいなかの知識がない。

NIH の研究者は、RIP と PCA で症例を省く方法が妥当か否かを正しく評価してほしい。それ以前に研究目的の異なった症例を識別するコードを添付すれば、癌研究が正しい方向に発展できる。理工学研究者も、種々の方法でこのデータを研究すべきだ。

000129

表 14　8 つの Microarray の概要

	Liver357	Liver300	Liver200	Liver100	Liver60	Liver40	Liver30
Type	A	A	A	A	A	A	A
n	357	300	200	100	60	40	30
p	254	306	199	198	298	296	300
BGSs in SM1	7	25	106	16	85	99	106
M2 Range	[1.6, 5.2]	[0, 0.17]	[0, 3.4]	[0, 10.9]	[0, 9.7]	[0, 44.5]	[50, 50]
M2=0	0	24	1	1	21	42	0
BGS genes	[20, 47]	[6, 26]	[1, 12]	[4, 19]	[1, 12]	[1, 40]	[1, 17]
gene=1	0	0	1	0	11	22	37

7.3.3　研究 2:　Breat6 (101,15)

Breat6 は LSD（検証法 1）であるが、2 群を赤と青で表示しなければ、図 6 の青い 15 人の正常者が赤い 101 人の癌患者の外れ値と見なされ（検証法 3）、M2 も 0 でない（検証法 2）。標本数がほぼ同じ場合、一般的に両方の症例は Prin1 で綺麗に 2 群に分かれる事が望ましい。乳癌は同数でない標本数でも他の癌より良い結果になり、乳癌の研究者は少数の正常例しか計測しない悪い傾向があるようだ。癌研究は、癌腫毎に先行研究のやり方を踏襲する傾向もある。しかし、「癌症例設計の 3 原則」と 3 つの検証法を満たす Array で研究すべきだ。

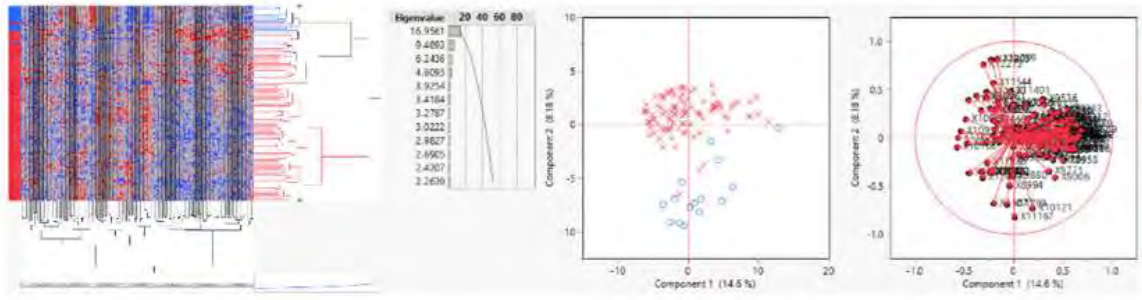


図 6　乳癌 6 の結果

7.3.4 研究 3：結腸 6（31,32）

表 15 の SM1 から SM540 までが LSD で、SM2 の M2＝1.58%を除き、539 個の M2 は全て 0%です。SM2 で悪影響を及ぼすケースは特定しなかった。RatioSV は 2 群間の SV の幅とケースの最大幅の比で 44.14%と SM3 の 52.92%に次いで大きい。M2 と RatioSV の不一致の理由は説明できない。RatioSV が 50%であれば SV が全体のバラツキの 50%幅の空白で 2 群を分けている。44.14%の空白幅があるのに M2=1.54 と悪いのは数個の遺伝子が極端な値をとるなどが考えられる。SM540 までの 539 組の Array と SM の全てが M2=0%であるので、乳癌が示す LSD よりより条件の良い LSD である事を示す。sm541 から sm544 は LSD でないので小文字の sm で示す。3 つの M1 は 0%であるが、M2 は 12.222%を超えている。これで M1 が信頼できないことと、M1＝0 が LSD の中でも特に良い状態の LSD である事が分かる。

表 15　　Method1 による Type-1 543 の SM と Type-2 SM544

SM	Gene	MIN	MAX	M1	MIN	MAX	M2	RatioSV
1	13	0	0	0	0	0	0	37.36

2	18	0	0	0	0	4.76	1.58	44.14
3	16	0	0	0	0	0	0	52.92
20	11	0	0	0	0	0	0	39.02
50	11	0	0	0	0	0	0	16.62
106	18	0	0	0	0	0	0	32.22
141	11	0	0	0	0	0	0	24.24
165	14	0	0	0	0	0	0	28.34
SM541	45	0	0	0	17.46	25.397	21.270	-
SM542	38	0	0	0	12.698	26.984	20.794	-
SM543	41	0	0	0	15.873	22.222	19.206	-
SM544	12	0	3.17	1.11	4.7619	23.81	12.222	-
MIN	6	0	0	0	0	0	0	-
MAX	48	0	3.17	2.06	17.46	31.746	21.27	-
Mean	23.2	0	0.03	0.01	2.5152	10.361	5.8252	-

図 7 の Ward 法と散布図は、SM1 の 2 群を明確に分けている。大腸癌の研究者は「癌患者設計の 3 原則」を守り、多くの良い結果になる。約 60 例の 2 群が慎重に選択されたと考える。上記のように SM1 を BGS に分割し、MNM=0、M2=0、そして Ward 法と PCA で 2 群が明確に分れるかどうかの 3 検証した後で、医師は医学的に BGS をより正確に研究できる。

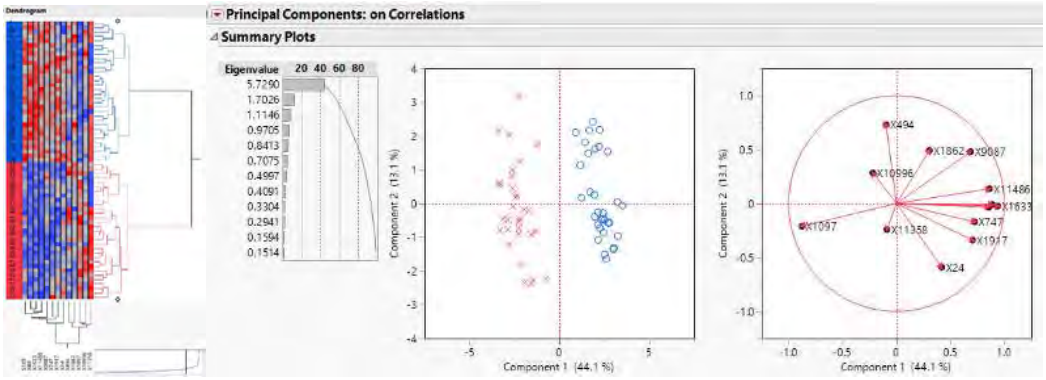


図 7　　大腸癌 6 の SM1 の Ward と PCA の散布図

7.4 結腸 5（26,26）で研究 3 の確認：南カリフォルニア大学エピジェネティクスセンター

結腸 5 の各 26 例の癌と正常は患者ではなく、癌細胞と正常細胞のペアのようだ。図 8 の SM1 は、ケース番号の問題のある 4 例のため M2=0 ではない。この 4 細胞を省いた修正結腸 5（24,24）は、後の表 18 で M2=0 で遺伝子数が 5 個以下の「vital BGS」が 1700 組み以上あった。Hinoue 博士ら(2012)は、エピジェネティクス研究の目的でこれらを含めたと考える。私の主張が正しいか否かを確認したい。

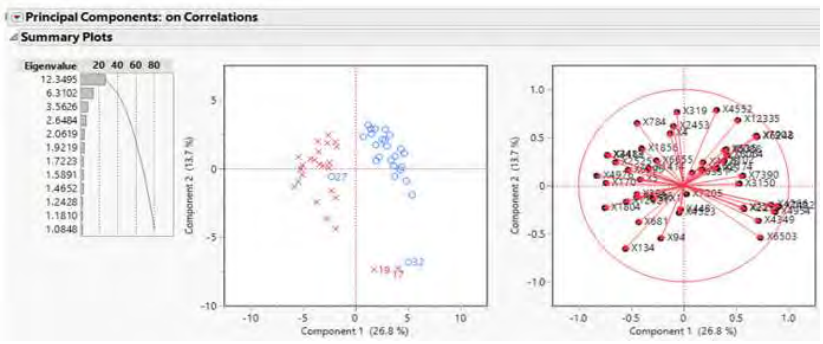


図 8　大腸癌 5 の 4 例の症例選択

vital BGS が、癌研究で既に発見されている 400 以上の 1 変量の癌遺伝子の幾つかを含んでいる場合、それは間違いなく新しい多変量の癌遺伝子の候補となる。初期の癌研究や機械学習などの科学技術研究が、過去の研究結果との関連性を考慮しなかったことは誤りです。従って約 50 例の症例が理想的と再確信した。これは、高品質のアレイに限定した「癌患者設計の新原則 3」です。この結果を南カリフォルニア大学医学部で評価してほしい。

PCA で修正大腸癌 5 を分析し、Prin1 で 2 群が分かれれば Type－A、重なれば Type－a とした。Type－B は Prin2 で、Type－C は斜めの軸で分かれる。図 9 の Type のヒストグラムで斜線の A を選び、遺伝子数、Prin1 と Prin2 の固有値と累積固有値の斜線部分を検討した。Type-A は 284 個中 41 個の約 14%で、誤分類のある Type-a は 75 個で約 26%で、両方で 136 個(40%)である。Type-B と-b は 5 個（1.7%）と 59 個(20.5%)で 64 個（22.2%）である。2 群を分ける軸が傾いている Type-C と- c は、55 個（9.1%）と 39 個(13.6%）で 94 個（32.8%）である。4 象限で重なる Type- d は、13 個（4.5%）である。医師は優先的に Type A と Type B を、癌遺伝子診断に選んだ方が良いと考える。

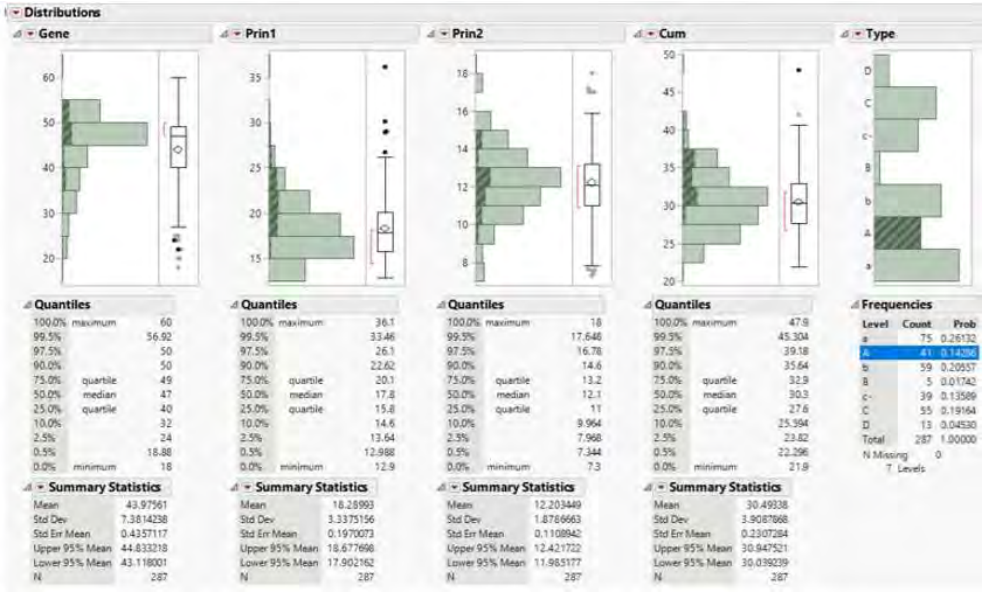


図 9　286 個の SM：遺伝子数、２つの寄与率、累積寄与率、散布図の 7 タイプの数

表 15 は 21 個の vital BGS の概要を示す。遺伝子数は 7 から 2 に減少する。1 個の遺伝子を持つ 7 BGS は、RatioSV の範囲が [0.38%、22.04%]と、2 群を一番分けている。RatioSV は、BGS の小さな症例数では

SV を固定するケースの違いで結果が大きく変わり Robust な統計量でない。これが表 14 の SM2 の原因かもしれない。RatioSV は、通常のデータでそれなりの意味はある。これらの 7BGS は、間違いなく多変量の癌遺伝子の候補に含まれていると考える。

表 15　M2=0 の 21 の BGS の概要

Number of Genes	Number of BGSs	RatioSV range
1	7	[0.38, 22.04]
2	6	[0.003, 16.95]
3	4	[0.161, 1.05]
4	2	[2.108, 6.364]
5	2	[0.27,1.294]

8　癌の遺伝子診断のための「実用的な 2 段階スクリーニング法 (Method3) 」

8.1　量子計算機なしで、医師が PC で癌の遺伝子データ解析する提案

LSD 判別理論は、5 つの問題(信頼できない NM、LSD 研究がない、分散共分散行列の瑕疵、間違った k 重 CV、量子計算機や IP ソフトを LSD で検証)を、4 つの事実（正規分布でなく組み合わせ論、MNM の単調減少性で FS ができる、LSD は Matryoshka 人形、高品質な発現量で 169 の Array が LSD）と 4 手法（正しい k 重 CV, SM 分割、2 段階スクリーニング法の Method3、ケースの選択法）で解決した。そして RIP を種々の用途で用いる 4 プログラムを LINGO で作成し、4 種の普通の LSD で実証研究した。そして 2 例の誤分類のある CPD で誤分類を省けば LSD になり、完璧な医学診断を確立した。その集大成として、169 の Array が LSD である事を初めて示した。そして表 1 の LSD 理論を体系化した。

医師がこの成果を簡単に利用できる簡便で強力な 2 段階スクリーニング法の Method3 を以下で示す。癌研究者が PC で n 症例 p 個の遺伝子データから、2 ^ n の症例の組みあわせで SM を見つける。そして p 個の遺伝子 の(2^p-1) 個の全変数の組みあわせで、M2=0 の vital BGS を見つける。

Step1：Program3 で Corolectal6 のように SM に分割する。検証法 1 で各 SM が MNM=0 を確認し、検証法 3 で Ward 法と散布図で問題がないかを確認する。

Step2：問題のない SM 毎に、 Program4 で BGS に分割する。Program2（検証法 2）で、M2=0 である vital BGS を見つける。多くの場合、遺伝子数が 1 から 5 個程度である。この中に癌研究が既に見つけた 1 変量の発癌遺伝子が含まれておれば、新しい多変量の発癌遺伝子になる（検証 4）。含まれていなければ、Genome　Cohort や生存時間解析で、新しい発癌遺伝子か否かの検討を行う（検証 5）。

8.2　癌患者の症例設計の 3 原則

RIP で疾病群と正常群(コントロール群)の医学診断を行う事が望ましい（原則 1）。正常がない場合は、より軽症か性格の分かっている方を仮のコントロール群と考える。そして症例数はほぼ同数である事が望ましく（原則 2）、症例数が 100 例以上と多いほど推定精度が上がると考える（原則 3）。これを「症例設計の 3 原則」と呼ぶ。しかし 169 の癌と正常の判別で、Colorectal6 のように 63 例の判別結果が良く、357 例の Liver3 が悪かった。Liver3 は多くの Genome Cohort を提供していて研究目的の異なる症例を含むと考えた。そして PCA の Prin1 のスコア値が小さな症例を順次省いた。40 例が一番多い 42 組の vital BGS に分割できた。しかし 30 例にすると、M2=50%の停止則が働いた。そこで症例数が 50 例ぐらいが良いとする新原則 3 に置きかえた「癌症例設計の 3 原則」を考えた。

000133

表 17 で 2 群の正常を含む 78Array を、この新 3 原則を満たす順位 1 の 26Array と、原則 2 を満たさない症例数が不均衡な順位 2 の 16Array 等の 7 順位に分けて、8.3 で vital BGS の評価した。

表 17 癌患者の 3 原則と 7 つの優先事項

Principle と Priority	Definition
Principle-1	Analysis of 2 groups with Cancer and Normal patients.
Principle-2	Almost the same sample size.
Principle-3	More sample size, better accuracy of estimation
New Principle-3 of expression data	Less than 100 patients can find many Vital BGSs.
Priority-1	26 data satisfy Principle-1, Principle-2, and new Principle-3.
Priority-2	16 data satisfy Principle-1 and new Principle-3.
Priority-3	9 data satisfy Principle-1 and Principle-2 and have over 100 patients.
Priority-4	6 data satisfy Principle-1 and have over 100 patients.
Priority-5	10 data consist of over 3 groups, including the Normal group.
Priority-6	6 data consist of over 2 groups without the Normal group
Priority-7	5 data consist of 1 cancer group.

8.3 44 Array の分析結果

表 18 の C/N は癌と正常の症例数、範囲 2 は BGS の遺伝子数、SM と BGS は M2=0 の個数を示す。予測は、遺伝子数が 5 個以下で M2=0 の vital BGS の個数を示す。SN=3 の修正大腸 5（24／24）は、SN=45 の最下位の大腸 5（26／26）から各 2 例を省いた結果である。vital BGS は予測値の 0 が 1714 個に増え、45Array で第 3 位になった。2020 年の日本癌学会で 400 個程度の 1 変量の発癌遺伝子と抑制遺伝子があるという発表があった。筆者はこれを遺産的癌遺伝子と呼ぶ。SN=20 の胃癌 1 の予測値が 408 個の境目である。上位の 20Array の 400 個以上の vital BGS に、遺産遺伝子が 1 個以上含む Array と含まない Array がある。含む場合は、1 変量の遺産的遺伝子の特徴で多変量の発癌遺伝子の特徴を研究できる。含まない場合は新しい発癌遺伝子になる。第I世代の癌研究では、このような遺産的遺伝子と新しい多変量の発癌遺伝子の検討がない。

癌研究の検証法は、生存時間分析が定番である。それに加えて Genome Cohort 分析が提案された(検証法 5)。一方第 1 世代の Array 研究では、1 変量や 2 変量の統計手法を用いて FS を行い約 50 個程度の一組を選んだ。Golub らは加重投票法で遺産遺伝子の ε 近傍にある約 50 個の遺伝子を一組選んだ。Chiaretti らだけが、恐らく医学的な遺産遺伝子の知識から複数の組を選んで NIH の研究資金を得た。

しかし 169 の Array が検証法 1 で LSD になる。この事実を示した研究はない。さらに結腸 5 は検証法 3 で見つけた 4 例を省くだけで、結果が極端に良くなった。以上の分析を、医師が PC でデータ解析すれば、困難な癌研究の前段階で助けになる。これが可能になるのは、量子計算機の得意とする組み合わせ最適化の RIP が、PC で 4 つのプログラムで可能になる Method3 を体系化したからである。そして分析結果を癌研究の評価に用いられる生存分析やゲノムコホートの評価の前に、新しい 3 検証法を開発した。Pancreatic の Detail の 5/1 は、1 個の BGS に 5 個の遺伝子が含まれることを示す。多くても 8 個なので、仮に 5 個以下を vital BGS とした。

表 18 スクリーニング方法による 44 個の Microarray の結果 (Method5)

順位	Cancer	C/N	SM	範囲 1	10 SM	BGS	範囲 2	SM	BGS	予測	Detail

								M2=0	M2=0		
2	Pancreatic	36/15	252	[38,58]	510	48	[5,18]	0	1	25	5/1
1	Breast1	8/8	-	-	149	33	[1,9]	0	5	424	1/1,2/1,4/3
1	Breast3	6/6	1083	[6,12]	120	30	[1,8]	0	8	836	1/1,2/3,3/2,5/1,
2	Breast4	19/8	509	[18,30]	260	41	[2,16]	0	5	243	2/1, 3/2, 5/2
5->2	Breast9	18,21/8	292	[28,48]	438	108	[1,9]	0	25	820	1/4, 2/8, 3/8, 4/2, 5/4, 8/1
2	Breast10	30/13	308	[22,50]	420	45	[2,14]	0	2	60	
1	Colorec-5	26/26	286	[10,22]	152	28	[2,10]	0	0	0	
1->1	Colorec-5	24/24	295	[23,42]	152	60	[1,45]		21	1844	1/6,2/6,3/5,4/2,5/2
1	Colorec-6	32/31	211	[1,11]	512	105	[1,8]	0	31	864	1/13,2/8,3/5,4/1, 5/2,8/2,9/1
1	Colorec-8	16/18	138	[9,18]	285	39	[1,8]	0	5	461	1/2,2/1,3/1,6/1

9. まとめ

9.1 169 の Microarray で分かったこと

RIP で判別するだけで、1 s で高次元の Microarray（n<< p）が、ほぼ n 遺伝子以下の MNM=0 の SM になり普通のデータとして分析できる。これが幅広データ（n< p）の特徴である。RIP で普通のデータ（n≥ p）を判別すると、ほぼ p 変数の SM になる。LSD 分析は、SM と BGS の分析である。初めに見付けた SM には 1 変量の発癌遺伝子が含まれることが多いようで、枝別れ法で Iris の X3 か X4 で LSD が簡単に分かれたように、これが IP の B&B 法で高速で SM を発見できた理由と考える。即ち DT と B&B 法は、似たアルゴリズムである。

LSD である SM は、マトリョーシカ人形のように小さな LSD を含み、最終的に最小次元の BGS で終わる Nested 構造をもつ。 iPS 研究で人の 3 万以上の遺伝子から Rat の ES 細胞で活性する 32 遺伝子を遺伝子 DB で見つけた。これが SM に対応する。そして 32 遺伝子から同時培養を繰りかえし山中 4 因子を見つけたと同じ方法で、BGS を見つけた。

癌と正常の SM は、多変量の癌遺伝子の候補発見の出発点になる。そして山中 4 因子と同じ位置づけの BGS が、癌研究の中心になる。多くの M2=0 の BGS は、1 個から 5 個程度の遺伝子を含む「vital BGS」になる。その中に癌研究が既に見つけた 1 変量の発癌遺伝子を含めば、多変量の発癌遺伝子と考えられる。含まない場合、Genome Cohort や生存時間解析などで新しい多変量の発癌遺伝子の候補として検討できる。しかし Liver3 のように異なった研究目的の癌種を含む場会、結果は複雑で悪くなる。

発現量は医学診断の最高品質の計測値である。Iris の 3 個の BGS は、同様に植物の遺伝子の発現量で決まるようだ。2 ^ 4 の 16 モデルで、X1 と X2 だけが LSD でなく、残り 14 モデルは LSD になり M2=0 である。恐らく隔離された地域の野生の猿でも、体長で猿の種別が LSD になる。京大の猿学で人類の系統樹の分岐点で 2 種が LSD になる BGS を調べれば、分岐の現員を示す BGS を特定できると考える。

2 種の異なった癌種の Microarray は、正常と 2 組の癌の BGS の違いを事前に比較研究する必要がある。単に異なった癌種でも LSD になることが分かった。

9.2 CPD(240＊19)：ケースの選択法で 53 年間で医学診断の完成

Program1 の RIP で CPD を判別すると、データに一意に 2 例の誤判別例が特定できる。この 2 例を省いた「ケースの選択法（Method4）」で、CPD238 は LSD になる。この 2 例が誤分類された理由を、医師は真剣に検討し、外的基準の決定の改善ができる。このような事は、他の判別手法で信頼のできない NM を見つけても実現できない。 Method4 は、判別理論に革命的な成果をもたらす。Program4 で、14 変数の BGS1 が見つかる。他のモデルの M2 と比べても最小である。Program2 で M2 を比較すると、BGS1 の M2 が 19 変数、18 変数の SM1、他の 2 つの 14 変数の BGS と比べて最小の Best なモデルである。情報科学で言われる哲学的な Occam の剃刀を具体的に初めて実現できた。

全判別データを LSD にして医学診断を行えば、誤判別例の理由と BGS1 に選ばれた変数の詳細な研究が可能になる。

9.3 BGS の M2 による評価で LSD の 4 分野の順位

紙幣の真贋判定で、2 変数が BGS である事と MNM の単調減少性（Fact2）が分かった。これから BGS を含む 2 ^ 4 の 16 モデルが LSD で、残り 48 モデルは MNM が 1 以上で棄却できる。検定統計量による FS は、万能細胞ができるか否かや、LSD であるか否かの二者択一の判別では意味が無い。

試験の合否判定は、得点 2 問が 50 点以上であれば合格とする場合、自明な LDF(=T1+T2-49.5)が正であれば合格、負であれば不合格になる。LDF=0 の線形超平面上にケースがないことで自明な LSD になる。合格点は、例えば得点分布の 90%、50%、10%と自由に設定し研究できる。そして 90%だけで M2=0 が見つかった。また数学IIb で Fisher の LDF の ER が 32%もあった。また QDF は一般化逆行列の瑕疵で、合格者がある設問を全員正解し不合格者がばらつく場合、合格者全員が誤判別された。以上のように分散共分散行列は未解決の問題が多い。SVD で分散共分散行列を高次元に拡張した LDF と PCA の研究が行われている。少なくとも高次元の LDF は CT のように診断に役に立たない。

小型車と普通車は、法律が決めた 2 つの 1 変量の BGS がある。63 モデルのうち 48 モデルが LSD で、残り 15 モデルは棄却できる。43 の LSD は M2=0 で、5 個の LSD は M2 は 0 でなかった。明らかに遺伝子より BGS の説明力は高い。

以上の M2 の結果で LSD を順位化すれば、次の 3 分類になる。1）法律がきめた BGS、2）遺伝子による癌診断や同植物の種が形態で分類できれば恐らく遺伝子が関係している。3）試験の合否判定は自明な LDF があり自明な LSD になるが、90%等を選ぶ必要がある。4）スイス銀行紙幣のような汎用的でない LSD。

更に Method4 で全ての判別データは LSD になる。一番大きな応用分野は以下で示す医学診断である。

9.4 LSD 判別理論の応用

以上で 5 種の普通の LSD と 169 の Array の実証研究を示した。しかし判別理論は、分析対象を幾つかのグループに分けて知りたいテーマを考えるだけで、これ迄想定できなかった驚く結果が得られる。

既に多くの医学診断は判別分析を利用している。しかしそれらの結果を RIP の判別で再検討するだけで、症例に固有な MNM=k を特定できる。この k 例をなぜ間違ったのかを検討する事が、判別分析の質を限りなく向上する。その後で、k 例を省いて LSD として BGS を見つけ M2 を始めとした検証法で評価する。そして簡単に p 値で結果を相殺するのでなく、BGS に選ばれた変数の組みの意味を検討する。

考えられる新規の応用分野としては次のテーマがある。

000136

1）経営分析で倒産企業と優良企業を LDF で判別し、Altman スコアが提案された。しかし LDF そのものに問題があり見なおすべきであろう。種々の分野で確立している評価法も同様である。

2）新しいテーマとしては、不良品の出る工程と正常の工程、実験などを繰り返す場合に成功した実験と問題の実験など、これ迄科学のに当たっていない未検討な項目が山ほどある。

3）現在、量子計算機が実用化に向けて注目されている。そして計算結果の信頼性を検証するために、OR の巡回セールスマン問題が提案されているが間違いである。この最適解を求める方法は正しいが、利用されるソフトが常に正しい結果を出すか否かの問題がある。また結果を万人が確認し理解することは容易でない。一方、RIP のアルゴリズムは連立不等式で分かり簡潔で、結果も MNM=0 か否かで分かり易い。そして世界中の研究者が利用できる Microarray が公開されている[23]。

しかしその前に表 19 は、4 種の LSD と 2 種の MNM=1 の 6 データで検証を薦める。x1 と x2 の 2 変数が説明変数で、y_i=1 が群 1 を、y_i=-1 が群 2 を表す識別子である。RIP 行は 6 データの RIP の判別係数で、MNM は最小誤分類数を示す。MNM=0 の 4 データは LSD で、MNM=1 は 1 個の誤分類を示す。2 変数をプロットすれば、この結果は確認できる。そして使用する IP と統計ソフトで RIP と同じ結果が出れば、安心して世界最先端の研究に参画できる。世界最大規模の物流システムに用いられている IP ソフトが表 19 のデータで正しい解を出さなかった。恐らく多くの統計と IP ソフトでも正しい解を出さないソフトがあると考える。

表 19　2 変数の 6 判別データ

		x1	x2	y_i	SCORE			x1	x2	y_i	SCORE
Data	RIP1	4	1	-6	MNM=1	Data	RIP4	-1	-2	7	MNM=0
A	1	1	3	1	1	D	1	0	3	1	1
	2	2	2	1	4		2	1	1	1	4
	3	2	-1	1	1		3	4	1	1	1
	4	1	1	-1	1		4	3	3	-1	2
	5	2	1	-1	-3		5	4	2	-1	1
	RIP2	2	2	-7	MNM=0		RIP5	-4	-1	0	MNM=0
B	1	1	3	1	1	E	1	-1	-1	1	5
	2	2	2	1	1		2	0	1	-1	1
	3	3	1	1	1		3	1	-3	-1	1
	4	1	1	-1	3		RIP6	-0.5	0.5	1	MNM=1
	5	2	1	-1	1	F	1	-1	-1	1	1
	RIP3	1	1	-1	MNM=0		2	1	1	1	1
C	1	1	1	1	1		3	0	1	-1	-1.5
	2	1	-1	-1	1		4	1	-3	-1	1
	3	-1	1	-1	1						

Fisher の仮説による判別理論は役に立たない。Fisher は恐らく判別理論の応用の広さを予見し、間違った仮説で判別理論を提案したと考える。

1. Miyake A, Shinmura S.: Error rate of linear DF. In: Dombal FT, Gremy F (ed) North-HollandPublishing Company, The Netherland, pp. 435–445. (1976).
2. Bruno CF, Edardo BC,Bruno IG, Marcio D (2019) CuMiDa: an extensively curated microarray database for benchmarking and testing of machine learning approaches in cancer research. J Comput Biol 26:1-11.
3. Jeffery IB,Higgins DG, Culhane C (2006) Comparison and evaluation of methods for geneating differentially expressed gene lists from microarray data. BMC Bioif 7 (1) :359 https://doi.org/10.1186/1471-2105-7-359
4. Schrage L (2006) Optimization modeling with LINGO. LINDO Systems Inc. (Shinmra S. translate it into Japan.
5. Shinmra S.: (2000) A new algorithm of the linear discriminant functions using integer programming. New Trends in Probaility and Statistics 5: pp. 133–142
6. Shinura S.: Four serious problems and new facts of the discriminant analysis. In: Pinson E, Valente F, Vitoriano B (ed) Opeations study and enterprise systems. Springer, Berlin, pp. 15–30. (2015).
7. Shimura SMatryoshka feature selection method for microarray data. Study Gate (15). pp. 1–16. (2015)
8. Shnmura S.: The best model of Swiss bank note data. Stat Optim Inf Comput 4: pp. 118–131. (2016).
9. Sinmura S.: Discriminant analysis of the linearly separable data—Japanese 44 cars. J Stat Sci Appl 4 (7–8): pp.165–78. (2016).
10. Theory1: Shinmura S.: New theory of discriminant analysis after R. Fisher. First edn. Springer, Singapore. (2016).
11. Shinmura SMatryoshka feature selection method for microarray data. Biotechnol : pp. 1–8. (2016). (Best PaperAward)
12. Shinura S.: First success of cancer gene analysis by microarrays. In: Biocomp'18, pp 1–7. (2018).
13. Shimura S.: Release from the Curse of High-Dimensional Data Analysis. Big Data, Cloud Computing, and Data Scence Engineering (Studies in Computational Intelligence 844): pp. 173-196. (2019).
14. Theory2: Sinmura S.: High-dimensional Microarray Data Analysis – Cancer Gene Diagnosis and Malignancy Indexes by Microarray-. First edn. Springer, Singapore (2019)
15. Shinmura S.: Cancer Gene Diagnosis of 84 Microarrays by Rank using 100-fold Cross-Validation. Transactions on Computational Science & Computational Intelligence, Springer Nature. (2020)
16. Sinmura S.: Cancer Diagnosis of 78 Microarrays registered on GSE froM2007 to 2017. Transactions on omputational Science & Computational Intelligence, Springer Nature. (2020).
17. Shinmura S.: First Success of Cancer Gene Data Analysis of 69 Microarrays for Medical Diagnosis. CSCI-ISCB: COMPUTATIONAL BIOLOGY1-7. Transactions on Coputational Science & Computational Intelligence, Springer Nature. (2020)
18. Shinmura S.:Twenty-three SeriouSMistakes of Cancer Gene Data Analysis Because 1995. In: Arabnia HR et al. (eds.), Advances inComputer Vision and Computational Biology, Transaction on Computational Science and Computational Intelligence Springer Nature Switzerland (2021), in Press.
19. Shinmura .: First Theory of Cancer Gene Data Analysis of 169 Microarrays and Four Universal Data Structures for Big Data.CSCI-ISCB: COMPUTATIONAL BIOLOGY1-14. Transactions on Computational Science & Compuuttional Intelligence, Springer Nature, (2021) in Press.
20. Shinmra S.: Practical Screening Method for Cancer Gene Diagnosis – Four Universal Data Structures of Discriminant Data - BIOLOGY1-16. Transactions on Computational Science & Computational Intelligence, Springer Nature, (202) in Press.

21. Shinmra S.: Practical Screening Method for Cancer Gene Diagnosis-2 –Four Patients Design Principles by Method1 and Satter Plot-. BIOLOGY1-17. Transactions on Computational Science & Computational Intelligence, Springer Natue, (2022) in Press.
22. Shnmura S.: Screening Results by Colon5 of 169 Microarrays. CSCI22 1-7. IEEE Book and Proceedings. (2022) .
23. Theory3: Sinmura S.: The First DT of Linearly Separable Data – From Exams and Medical Diagnosis with Misclassifications to 169 Microarrays for Cancer Gene Diagnosis. First edn. Springer, Singapore (2024) .
24. Shimura S.: The World's First LSD Discriminant Theory using RIP. 1-22. Springer, Singapore Nature Switzerland (2027)
25. Shinmra S.: First Study of 169 LinearlySeparable Expression Data. 1-21. The European Society of Medicine. Medical Research Archives, Vol14 Issue 8.
26. 川村昇、新村秀一、開原成允（1984）.医学データ解析のための統計プログラムパッケージ、医学のあゆみ、110/113, 871-878. 注：SAS を医薬や医学で使うことの開原先生の意思を示す。
27. 森村・牧野他編（分担執筆）（1984）。統計・OR 活用事典。東京書籍。注：経営科学 OR 用語大事典（浅倉）やマルチメディ統計百科事典に発展。OR と統計の融合を先人は考えられていた。新村は統計は「データで表される問題」を、OR は「数式で表される問題」を解決する学問として大学教育に取り入れる事を提案したが、賛同者はいなかった。簡単にオバマのデータサイエンスを採り入れる軽薄さを愁う。
28. 新村秀一・高森寛 （1987） 統計処理エッセンシャル、丸善。SAS で学生データを各種法で分析。
29. Schrge,L.： 青沼龍雄・新村秀一訳(1980) GINO によりモデリングと最適化、共立出版。
30. Sall,J. (1986) SAS による回帰分析の実践、朝倉書店。注：SAS の 2 つのテクにあるレポート（SAS Regression Application と Swee p Operate r ）の紹介。
31. Schrge,L.；新村秀一(1982) 実践数理計画法。朝倉書店。
32. 新村秀一 （校正）JMP を用いた統計およびデータ分析入門. SAS Japan. ：Sall,J. et al.(1999) JMPStart Statistics, SAS Institute.
33. 新村秀一 （2004） JMP 活用統計学とっておき勉強法。講談社。注：4 件 2 変数で JMP の分析法の紹介。
34. 新村秀一 （2010） 最適線形判別関数、日科議連。 注：RIP の解説。
35. 新村秀一 （2010） 数理計画法により問題解決法、日科議連。 注：LINGO の重要な雛形モデルの紹介。
36. 新村秀一 （2012） コラム「SAS/JMP との歩み」、SAS Technical News. 春、夏、秋、冬号。

謝意

A. SAS との思い出

1971 年に京大の数学科の大学院に落ち、10 月頃進路を考えた。駸々堂横の本屋で偶然 NHK 出版の近藤次郎著の「オペレーションズ・リサーチ入門」という数学で種々の問題を解決する本を見つけた。著者は数学科の先輩で東大の航空工学科に再入学され、YS11 の開発や公害や原子力等の多方面で活躍され、理工学研究の知の巨人のような先生である。大学院に進んで OR 的な問題を解決する一流の応用数学者になぜなられなかったのか疑問に思った。きっと敗戦で奪われた日本の空を取り戻そうと考えられたと勝手に判断した。

そこで当時情報処理が注目されていたので、NEC と SCS の役員面接を受けて採用された。できたばかりの SCS に断りにいくと、NEC から出向の津田専務に「秀才きらめく NEC にいっても芽が出ないので SCS の方がずっとおもしろい」と諭されて入社した。しかし 5 月初めに NEC に出向し大阪成人病センター（現大阪国際癌センター）の野村先生の研究室に出向した。先生から NEC と共同開発中の「心電図の診断論理

を東大出版会の計量診断学を読んで開発するように」と言われた。大卒が翌年から学会発表でき頑張った。しかし 4 年間の成果は先生の枝分かれに敵わなかった。これが 53 年間医学診断に役立つ判別理論開発の動機である。NEC の計算機室で知り合った鈴木先生と大阪府の疫学データを分析し発表した。

4 年後に SCS に戻って IBM で稼働する統計ソフトとして当時 No1 の世界で 2000 ユーザー以上ある SPSS の導入を考えた。しかしユーザー数 800 前後の日本に代理店を探していた SAS の薄い SAS82 のマニュアルを読んで Proc と Data ステップを分離した斬新さに注目し Assist 社長とノースカロライナに行き Goodnight 社長と代理店交渉したが、日商エレに決まっていたので計算サービスを始めることにした。

一方仕事に直接関係しない OR 学会に入った。しかし森村、牧野両先生から「統計・OR 活用事典[27]」で統計手法の紹介を行う機会を得た。両先生の意思を受け継ぎ、「統計はデータで表される問題解決」を、「OR は数式で表される問題を解決する学問」で、両方を習得することを提案をしたが誰も賛同しなかった。それがオバマ大統領がデータサイエンティストを提案すると、追従する研究者が現れた。私は医学診断の最初のデータサイエンティストの生き残りであろう。

暫くして IBM よりもシカゴ大学で開発された数理計画法ソフトの LINDO 製品の方が良いと判断しシカゴ大学を訪問し代理店になった。私の統計の勉強法は、4 件 2 変数の整数値データで分析し、簡単な行列演算でアルゴリズムの理解をする［33］。関係する英文の資料を翻訳あるいは校正する［32］。そして SAS や LINDO の解説書や記事を書くという方法である。

これによって大橋講師が奥野先生の代理で、日科議連で行われていた品質管理の定例勉強会で SAS の紹介を行った。奥野、芳賀先生は日科議連のソフトから SAS へ方針変更された。また東大医学部の伝説の三秀才と言われ厚生省の医療行政の責任者の開原先生の月例の勉強会に参加させてもらった。また SAS の紹介を医学誌に連名で出させてもらった［26］。恐らくこの時点で SAS を新薬の承認に用い世界標準に対応する事を考えられたと思う。そして大橋先生を助教授にむかい入れ、SAS ユーザー会が東大で行われるようになったと推測する。勉強会で小学校から大学まで同級生の三宅日本医大数学科教授の知己を得た。先生は数学的にヒューリスティック OLDF を研究されていたが日本の学会誌でリジェクトされていたので、私に実証研究の共同研究を頼まれた。後年これを IP で RIP を開発する事になった。先生は同大学の医師からデータ解析を依頼されていたが、統計ソフトを使わないので私にまわされた。重用なのは CPD の分析と丸山ワクチン 32 万症例の厚生省申請のための分析である。全てが投与群であるので、「SSM は水のように無害であるが治療効果もない」という帰無仮説で、手術後の 15 万例を 3 ヶ月毎の 4 期に分け、3 ヶ月以内に投与した患者の生存日数が重回帰分析で長いことを示した。大橋先生が編集長の開原先生が設立に貢献された医療情報学会に投稿した論文は「新村は帰無仮説の意味を分かっていないという指摘を押さえて採択しました」と聞かされた。この後、大橋先生はある特殊な癌で厚生省の認可を取り、SSM はI年ごとに審査する治験薬を利用者も多いので審査無しで継続となった。これは開原先生の英断と考える。この後、介護保険制度も企画されたようだ。筆者の高校の 1 年先輩の D 医師は東京都の副院長であった。趣味が SPSS で統計分析し医師向けの統計の本も出している。厚生省が特養等の介護施設で、看護師が医師や医療スタッフとペアになり、各患者にどれだけ対応したかの 1 分間タイムスターディ調査をした。D 医師はそのデータを用いて介護保険のランクを決める分析をした。頭が良いので主成分回帰等の難しい手法で分析し、国会が決めた日程に間に合わなかった。SPSS の営業から頼まれて彼に会い、私は JMP の Partitioning の分類木を薦めた。彼は SPSS で分析したいというので CHAID になった。わずか 3 ヶ月で分析し、結果を C コードのプログラムを開発し全国の特養と在宅で実施した。その時点で成蹊大学の経済学部教授になっていた。4 月頃に田原総一

朗のサンデープロジェクトに彼が登場し、「僕の作ったシステムはボロボロなんです」と告白したのに驚いた。話しから考えられる失敗は、1）順序尺度にすべき変数を名義尺度で分析したと考える。2）特養患者データの結果を在宅まで適用した。3）更に問題なのは分析に用いていないデータで実施前に検証が行われていない点である。この点を SAS ユーザー会等で指摘した。数日後に厚生省の課長から国立大蔵病院で行われる開原院長の会合の委員になることの依頼を受けた。恐らく科研費などの公的研究費を貰っていないので厚生省の研究費でも貰えると思ったが、D 医師が省かれているので断った。現在この委員になっておれば、癌の遺伝子データ解析を日本の医学者に普及する手がかりになったと後悔している。山上会館で行われた SAS の懇親会に、開原先生が現れ入り口で私を捜してウイスキーグラス片手においでになった。そして 1 口のんで「介護保険は私が企画し全責任がある」とおっしゃった。私はその経緯や SSM 認可の裏話を聞きたかったが黙ってしまった。その数ヶ月後に翌年国際医療福祉大学の学長になられる予定であったが、教授会の座長をつとめながら亡くなられたことを後で知った。

B.　なぜ前人未踏の Fisher の判別理論の書き直しができたののか？

SAS や JMP という世界最高水準の使い易い統計ソフトを用いて、多くの実際のデータを幅広い手法で分析した。そして SAS や JMP を個人家庭教師として統計では 4 件 2 変数の整数値で分析し、アルゴリズムの理解に努めた[32]。この他、Goodnight 博士や Sall 博士の SAS テクニカルレポートを出版した[30]。また JMP の解説書の校正を行った[32]。一方シカゴ大学の Linus 教授の開発した LINDO 製品の代理店になった。もちろん技術資料は 2 人の英語のできる女性課員に飜訳させ私が手を入れた。また彼の著書を翻訳した[29,31]

日商エレクトロニクス、アシストと 4 年ごとに SAS の代理店が代わった。その後にアシストの担当者を中心に SAS Japan ができた。そして担当営業が汎用機の SAS しか売らないので、私が VAX 版の代理店になり 32 の製薬企業の動物実験から第 4 相比較臨床試験を SAS で開発し、厚生省に新薬請求のための部門コンピュータを SAS の医薬部門の営業と協力し 32 社に販売した。その後に東洋信託銀行（現東京 UFJ 信託）に I億円前後の VAX8000 と SAS、Oracle、日経 NEED と LINDO 製品を金融ポートフォリオ・システムとして販売した。DEC の金融部隊に高く評価され、日本で最初の SWIFT の代理店になった。しかし金融担当の営業は、SAS/IML で投資分析を提案するコンサルタントくみ、この分野を開拓するということで代理店が解消された。私はユーザー会で、IML は研究者のちょいプロのためのソフトでシステム開発用でないと指摘した。和光証券の 2 名の SAS ユーザーが、東京三菱銀行に転職した。億円以上の開発費をかけたシステムが、多くの場合で株式の 100%の組み入れ比率が 0%になる問題の解決を依頼された。LINDO で幾つかのモデルを流すと正しい解が得られるので、単に収束計算が実システムの要求に対応していないという 100 万円程度の報告書で終わり、LINDO で再開発の依頼はなかった。SAS、LINDO、日経 Need、Oracle という組み合わせで日本の銀行と信託銀行に投資分析システムの普及に失敗した。また Swift の日本の最初の代理店になったが、SAS の計算サービスでリクルートや日本気象協会等の SAS 営業を担当した課員を金融の部長に推薦して金融との繋がりをたったのは間違っていたかもしれない。

ｃ，　まとめ

近藤先生の生き様に感銘し、津田専務に言われるまま NEC にいかず新天地の SCS に入社した。しかし NEC は SCS に月 40 万円の出向料を払い大阪成人病センターの野村医長の研究室に身を寄せた。当時は第 1 次統計ブームで、医療データを統計分析すれば計量診断学と呼ばれていた。学卒で翌年から野村先生と ME

学会や循環器学会で多数の医師の前で発表した。しかし 4 年間頑張ったが野村先生の枝分かれ論理に敵わなかった。そこで医学診断に真に役立つ判別理論の開発が 53 年間の研究の動機になった。

統計的判別理論の問題は次のようである。

1）3 群以上の判別関数もあったが使われなくなった。その理由は各 2 群の組み合わせで最適な説明変数の組みと多群で最適な組みは異なるので、多群を 2 群の組み合わせで判別する必要がある。

2）ロジスティック回帰が LDF より遙かに優れている。

3）心電図診断のような実用システムは、単純に 1 回の判別で結論が出ない。

4）また筆者も判別分析した場合、長らく NM から計算した ER の報告で終わっていた。しかし CPD で紹介したように、判別の目的は医師がきめた外的基準の y_i が誤判別された場合、その理由を検討する事が一番重用だ。そして誤分類を省いた LSD で最小変数のモデルの BGS を見つける。BGS は M2 が最小な Best なモデルであるので、BGS の変数の組みの意味を総合的に検討すべきである。

Base SASによるWOE変換を用いたスコアリングモデル構築法

○坂巻英一[1]

([1]国立大学法人宇都宮大学)

Scoring Model Development Using Weight of Evidence Transformation in Base SAS

Yoshikazu SAKAMAKI
School of Data Science and Management / Utsunomiya University

要旨

Weight of Evidence（WOE）変換は、信用スコアリング分野で広く利用されている変数変換手法であり、カテゴリ変数を目的変数との関連性に基づく数値尺度へ変換できる特徴を持つ。本稿では、Base SAS を用いてWOE 変換を実装し、2 項ロジットモデルによるスコアリングモデルを構築する方法を紹介する。また、Kaggle で公開されている銀行のダイレクトマーケティングデータを用いて従来のダミー変数を用いたモデルと比較を行い、対数尤度、AIC および Accuracy Ratio（AR 値）の観点から性能を評価した。その結果、WOE 変換を用いることで説明変数および推定パラメータ数を削減しながらモデル適合度および識別性能を改善できることを確認した。本稿では、Base SAS による WOE 変換からモデル構築までの一連の実装手順について紹介する。

キーワード： scoring model, Weight of Evidence (WOE), Base SAS, PROC LOGISTIC

１．はじめに

消費者選択行動をはじめとする社会科学の多くの分野において、二値選択問題のモデル化は古くから重要な研究課題の一つである。その代表的な手法が 2 項ロジットモデルである。二値事象の生起確率をモデル化する手法として 1950 年代頃から研究が進められ、信用リスク計測やダイレクトメールの反応予測など、スコアリングをはじめとするさまざまな分野で広く利用されている。

とりわけ、モデル内でカテゴリ変数を扱う際には、各カテゴリにダミー変数を割り当ててパラメータを推定する方法が一般的である。しかし、カテゴリ数の増加に伴い説明変数およびパラメータ数が増加するため、過学習や推定分散の増大を招き、モデルの予測性能や解釈性が低下する可能性がある。

こうした課題を改善する手法の一つが、Weight of Evidence（WOE）変換である。WOE 変換では、各カテゴリを目的変数との関連性を表す数値尺度に変換できるため、金融分野の信用スコアリングを中心に広く利用されてきた（大勢待，2008）。一方で、消費者選択行動やダイレクトマーケティングの分野では、その適用事例は必ずしも多くない。

そこで本稿では、Base SAS を用いた WOE 変換によるスコアリングモデルの構築方法を概説するとともに、ダイレクトメールの配信効率化を目的とした顧客ターゲティングへの適用方法を紹介する。

２．先行研究概要

２．１　２項ロジットモデルについて

まず２項ロジットモデルの概要について説明する。今、

k　　顧客番号 ($k = 1, 2,\ldots,K$)

m　　変数番号 ($m = 1, 2,\ldots,M$)

x_{km}　　顧客 k の変数 m に対する属性値

y_k　　顧客 k において事象が生起した時１そうでない時０をとる２値変数

β_m　　変数 m に対するパラメータ

とする。選択肢に対する線形予測子を

$$v_k = \sum_{m=1}^{M} \beta_m x_{km}$$

とする時、事象が生起する確率は

$$p_k = \frac{1}{1 + \exp(-v_k)} \tag{1}$$

によって定式化される。対数尤度関数を

$$LL = \sum_{k=1}^{K} y_k \log p_k \quad + (1 - y_k) \log(1 - p_k) \tag{2}$$

により定式化し最尤推定法によりパラメータ推定を実施する。

２．２　Weight of Evidence (WOE)について

次に Weight of Evidence (WOE) について説明する。今、

m　　変数番号 ($m = 1,2,\ldots,M$)

j　　カテゴリ番号 ($j = 1,2,\ldots,J_m$)

x_{mj1}　事象に対する反応があった顧客のうち変数 m のカテゴリ j に含まれる人数

x_{mj0}　事象に対する反応がなかった顧客のうち変数 m のカテゴリ j に含まれる人数

とする。このとき、変数 m のカテゴリ j に対する WOE (Weight of Evidence)を

$$w_{mj} = log\left(\frac{\frac{x_{mj1}}{\sum_{j'=1}^{J_m} x_{mj'1}}}{\frac{x_{mj0}}{\sum_{j'=1}^{J_m} x_{mj'0}}}\right) \tag{3}$$

によって定式化する。

なお WOE の算出に先立ち、連続量データは、あらかじめビンを設定してカテゴリ変数へ変換した上で分析に使用する。またカテゴリ変数についてはそのまま WOE を算出する。

２．３　本稿における提案手法

ここで WOE 変換後の変数を説明変数としたスコアリングモデルについて説明する。今、

k　　顧客番号 (k=1,2,….,K)

$d^{(m)}_{kj}$　顧客 k が変数 m においてカテゴリ j に属する場合には 1 それ以外は 0 となるダミー変数

w_{mj}　変数 m のカテゴリ j に対応する WOE

β_m　変数 m に対するパラメータ

y_k　顧客 k がダイレクトメールに反応するとき 1 それ以外は 0 をとる 2 値変数

とし、ある事象に対して顧客 k が持つ効用を

$$v_k = \sum_{m=1}^{M} \beta_m \left(\sum_{j=1}^{J_m} d^{(m)}{}_{kj} w_{mj} \right) \tag{4}$$

により定式化する。その上で事象が生起する確率を(1)式で定式化する。また対数尤度を(2)式により定式化し最尤推定法によりパラメータ推定を実施する。

２.４　比較対象手法

本稿では、提案手法の有効性を検証するため、比較対象モデルとして、WOE 変換を適用せず、元の変数をそのまま説明変数として用いた 2 項ロジットモデルを使用する。カテゴリ変数については基準カテゴリを一つ設定し、ダミー変数を用いてモデルを構築した。

２.５　分析結果の検証

本稿ではモデルの性能を評価する指標として対数尤度、AIC 統計量および Accuracy Ratio（AR 値）を使用する。AR 値は、スコアリングモデルの識別性能を評価する指標であり、金融工学分野で広く利用されている。ROC 曲線に基づいて算出され、値が大きいほど識別性能が高いことを示す。

３．実データへの適用

３.１　データの概要

本節では、Kaggle で公開されているオープンデータを用いて提案手法の検証を行った。検証に使用したデータソースは以下の通りである。

データ出典元　Kaggle (データ分析用のオープンデータを提供するプラットフォームである)

データセット名　Bank Direct Marketing

URL　https://www.kaggle.com/datasets/psvishnu/bank-direct-marketing

利用ファイル　bank-full.csv

分析ツール　本稿では Base SAS および SAS /STAT を使用した。

備考　本データはポルトガルの銀行が実施したダイレクトマーケティング・キャンペーンに関する結果をまとめたものである。このマーケティング・キャンペーンは定期預金の契約促進に関するものであり 2008 年 5 月から 2010 年 11 月にかけて電話によって実施された。

分析に使用したデータについてテーブル定義書を表 1 に示す。なお、Kaggle で配布されているデータセットは 17 個の変数から構成されていたが、欠損値が多いことや、目的変数との情報漏洩（リーケージ）が懸念される変数を除外し、本稿では表 1 で示す 8 個の変数および契約促進キャンペーンの結果のみを使用してモデルを構築する。

変数名	役割	型	変数名（日本語）	説明	欠損値
age	説明変数	数値	年齢	年齢	なし
job	説明変数	カテゴリ	職業	職業の種類（'admin.'：管理職、'blue-collar'：肉体労働、'entrepreneur'：起業家、'housemaid'：家事労働、'management'：管理職、'retired'：退職者、'self-employed'：自営業、'services'：サービス業、'student'：学生、'technician'：技術職、'unemployed'：失業者、'unknown'：不明）	なし
marital	説明変数	カテゴリ	婚姻状況	婚姻状況（'divorced'：離婚・死別、'married'：既婚、'single'：未婚、'unknown'：不明）	なし
education	説明変数	カテゴリ	学歴	教育水準（'primary'：初等 'secondary'：中等 'tertiary'：高等 'unknown'：不明）	なし
balance	説明変数	数値		年間平均口座残高	なし
housing	説明変数	２値		住宅ローンの有無	なし
contact	説明変数	カテゴリ		連絡手段（'cellular'：携帯電話、'telephone'：固定電話）	あり
campaign	説明変数	整数		今回のキャンペーン中に当該顧客へ行った接触回数（最後の接触を含む）	なし
y	目的変数	２値		定期預金を契約したかどうか	なし

表１　テーブル定義書

３.２　WOE の算出

ここで 2.2 項に基づき WOE を職業および年齢の例について算出した結果を表２に示す。

	JOB	NO	YES	TOTAL	RATIO(%)	WOE
1	admin.	4,540	631	5,171	12.2%	0.0479
2	blue-collar	9,024	708	9,732	7.3%	-0.5239
3	entrepreneur	1,364	123	1,487	8.3%	-0.3847
4	housemaid	1,131	109	1,240	8.8%	-0.3182
5	management	8,157	1,301	9,458	13.8%	0.1856
6	retired	1,748	516	2,264	22.8%	0.8012
7	self-employed	1,392	187	1,579	11.8%	0.0139
8	services	3,785	369	4,154	8.9%	-0.3067
9	student	669	269	938	28.7%	1.1102
10	technician	6,757	840	7,597	11.1%	-0.0636
11	unemployed	1,101	202	1,303	15.5%	0.3256
12	unknown	254	34	288	11.8%	0.0103
	TOTAL	39,922	5,289	45,211	11.7%	

AGE	min	max	NO	YES	TOTAL	RATIO(%)	WOE
1	18	27	2,377	529	2,906	18.2%	0.5187
2	27	30	3,357	529	3,886	13.6%	0.1735
3	30	33	4,004	518	4,522	11.5%	-0.0238
4	33	35	4,070	452	4,522	10.0%	-0.1764
5	35	37	4,026	496	4,522	11.0%	-0.0727
6	37	41	4,080	442	4,522	9.8%	-0.2012
7	41	44	4,143	379	4,522	8.4%	-0.3703
8	44	48	4,108	414	4,522	9.2%	-0.2735
9	48	53	4,114	408	4,522	9.0%	-0.2896
10	53	95	5,643	1,122	6,765	16.6%	0.4060
TOTAL			39,922	5,289	45,211	11.7%	

表２　WOE を算出した結果（左表：職業　右表：年齢）

分析を行った結果、

職業：

〇サービス業では WOE が低く、契約率が平均より低い傾向が確認された。

〇退職者や学生では WOE が高く、契約率が高い傾向が確認された。

年齢：

〇若年層およびシニア層では WOE が高く、ミドル層では WOE が低い傾向が確認された。

３.３パラメータ推定結果

次に 2.3 項および 2.4 項に基づきパラメータ推定を実施した結果を表３および表４にまとめて示す。

変数	パラメータ推定値	WaldChiSq	p値
Intercept	-2.0539	9633.7949	<.0001
age	-0.5235	54.1738	<.0001
balance	-0.8565	180.0071	<.0001
campaign	-1.0197	186.8330	<.0001
job	-0.4499	69.5349	<.0001
marital	-0.6404	41.9847	<.0001
education	-0.3156	11.8840	0.0006
housing	-0.6250	168.6595	<.0001
contact	-0.8585	363.2230	<.0001

表３　本稿における提案モデルを基にパラメータ推定を実施した結果

変数		パラメータ推定値	WaldChiSq	p値
Intercept		-3.1216	115.7084	<.0001
age		0.0069	8.1825	0.0042
balance		0.0000	26.5021	<.0001
campaign	1	0.2717	17.9976	<.0001
	2	-0.1102	2.5511	0.1102
	4	-0.3246	11.2324	0.0008
	5++	-0.6679	55.6947	<.0001
job	admin.	0.2484	0.9897	0.3198
	blue-collar	-0.0418	0.0282	0.8667
	entrepreneur	-0.3515	1.6274	0.2021
	housemaid	-0.1626	0.3475	0.5555
	management	0.0444	0.0321	0.8578
	retired	0.6281	6.1005	0.0135
	self-employed	0.0475	0.0325	0.8569
	services	-0.0850	0.1114	0.7386
	student	0.8281	9.9530	0.0016
	technician	-0.0127	0.0026	0.9591
	unemployed	0.3862	2.1743	0.1403
marital	married	-0.1132	3.0851	0.079
	single	0.2520	11.8666	0.0006
education	primary	-0.2765	5.9673	0.0146
	secondary	-0.1319	1.7441	0.1866
	tertiary	0.1017	0.9340	0.3338
housing	yes	-0.5828	185.4625	<.0001
contact	cellular	1.2499	385.1473	<.0001
	telephone	0.9639	97.4812	<.0001

表４　比較対象モデルを基にパラメータ推定を実施した結果

算出した対数尤度および AIC の値は表 5 および表 6 の通りである。

	-2logL	AIC
本研究における提案手法	17,467.0	17,485.0
従来型の手法	17,644.0	17,696.0

表 5　対数尤度および AIC

	平均	標準偏差	最小値	最大値
本研究における提案手法	0.4485	0.0088	0.4279	0.4691
従来型の手法	0.4269	0.0085	0.4086	0.4450

表 6　AR 値の平均値

４．まとめ

本稿では、Weight of Evidence（WOE）変換を用いたスコアリングモデルの構築方法について紹介した。2 項ロジットモデルを用いてスコアリングモデルを構築する際、カテゴリ変数をダミー変数として扱う従来手法では、カテゴリ数の増加に伴い説明変数および推定パラメータ数が増加し、モデルの複雑化や推定の不安定化を招く可能性がある。本稿では、この課題を改善するため、説明変数に対して WOE 変換を適用し、カテゴリ情報を数値尺度へ変換する手法を検討した。

実データを用いた検証の結果、WOE 変換を用いることでモデル内の説明変数およびパラメータ数を削減しながら、従来手法と比較して良好なモデル適合度および識別性能が得られることを確認した。

本稿で紹介した WOE 変換を用いたスコアリングモデル構築手法が、ダイレクトマーケティングをはじめとするデータ分析業務に携わる実務家の一助となれば幸いである。

参考文献

1. 大勢待利明（2008）, デフォルト確率推定モデル作成における WOE 変換の役割とその利用方法, SAS ユーザー総会アカデミア/テクノロジー&ソリューションセッション論文集(2008 年 07 月 23 日開催), pp.298-305

付録　WOE を算出するための SAS コード

```
/* 【ステップ 1】 職種ごとに、y="no" と y="yes" の件数をそれぞれ数え上げる */
proc sql ;
    create table work.step1_counts as
    select
        job,
        sum(y = "yes") as good_cnt,
        sum(y = "no")  as bad_cnt
    from Mylib.BANK
    group by job
    ;
quit ;

/* 【ステップ 2】 データ全体の Good の総数と Bad の総数をそれぞれ計算する */
proc sql ;
    create table work.step2_totals as
    select
        sum(good_cnt) as total_good,
        sum(bad_cnt) as total_bad
    from work.step1_counts
    ;
quit ;

/* 【ステップ 3】 ステップ 1 とステップ 2 を合体させ、1 行ごとに全体総数を持たせる */
proc sql ;
    create table work.step3_combined as
    select
        a.job,
        a.good_cnt,
        a.bad_cnt,
        b.total_good,
        b.total_bad
    from work.step1_counts as a
    cross join work.step2_totals as b
    ;
quit ;

/* 【ステップ 4】 データステップを使って、1 行ずつ WOE の数式を計算する */
data work.step4_woe_map;
    set work.step3_combined;
    /* 0 除算を防ぐための確率（シェア）の計算 */
    good_share = (good_cnt + 0.5) / (total_good + 0.5);
    bad_share  = (bad_cnt  + 0.5) / (total_bad  + 0.5);
    /* WOE の算出 */
    WOE_JOB = log(good_share / bad_share);
run ;

/* 【ステップ 5】 完成した WOE マッピング表を元のデータ「Mylib.BANK」に結合する */
proc sql ;
    create table work.BANK_WOE as
    select
        a.*,
        b.WOE_JOB
    from Mylib.BANK as a
    left join work.step4_woe_map as b
    on a.job = b.job
    ;
quit ;
```

SASによる傾向スコアに基づくPower事前分布アプローチの実装

○横田　大輔、赤石　理香子、浅野　豊、伊庭　克拓、黒見　真央、毛利　誠

（大塚製薬株式会社 新薬開発本部 バイオメトリックス部 統計解析室）

Implementation of Propensity Score-integrated Power Prior Approach using SAS

Daisuke Yokota, Rikako Akaishi, Yutaka Asano, Katsuhiro Iba, Mao Kuromi, Makoto Mouri
Office of Biostatistics, Department of Biometrics, Headquarters of Clinical Development, Otsuka Pharmaceutical Co., Ltd.

要旨

希少疾患などを対象とする医薬品のより効率的な評価を可能にする試験デザインへの関心が高まっている。方法論の１つとして，他試験のプラセボ群やレジストリデータなどを外部対照として活用し，無作為化比較試験（RCT）の対照群と組み合わせることによって効率を高めるハイブリッド対照アプローチが注目されている。外部データを利用する統計手法はこれまでいくつか提案されており，その 1 つにベイズ流アプローチにより外部データを事前分布に組み込む power 事前分布を用いた方法がある。Power 事前分布は，初期事前分布と情報量を割り引かれた外部データの尤度に基づいて構築され，外部データの寄与度を定数もしくは新規データとの類似性に応じて動的に決めることができる。外部データの寄与度については，新規データと外部データの傾向スコアの類似性に応じて決める方法も提案されている。本稿では，傾向スコアに基づく power 事前分布アプローチに着目し，ハイブリッド対照を用いた RCT での SAS による実装を試みるとともに簡単な性能評価を行う。

キーワード：Power 事前分布，傾向スコア，MCMC プロシジャ，外部対照

1. はじめに

医薬品開発において有効性および安全性に関するエビデンスを得るための基本的な手段は，同一試験内で被験治療を対照治療と比較するランダム化比較試験（RCT）である。しかしながら，希少疾病用医薬品などの開発では，有効性の検証に必要な例数の確保が困難，さらには RCT 自体の実施が困難な状況もある。これに対するアプローチとして，近年，外部データを用いて対照群を設定し被験治療と比較する方法への関心が高まっている。外部データを活用するアプローチの 1 つとして，RCT での対照群データに既存試験やリアルワールドデータなどの外部対照群のデータを加えて補強するハイブリッド対照アプローチがある（武田ら, 2015; 野村ら, 2022; 野村ら, 2026）。限られた症例数での有効性評価が必要な希少疾患での開発や迅速な意思決定が必要な状況において，外部データの適切な統合は，試験効率の向上や倫理的負担の軽減に寄与す

る可能性がある（製薬協, 2022; 製薬協, 2024）。ハイブリッド対照アプローチに用いられる統計的方法論として，外部データから事前分布を構成するベイズ流の解析手法がよく用いられている。医薬品開発でのベイズ流アプローチの利用に関しては，近年，ガイドラインなどが発出されており（AMED, 2023; EMA, 2026; FDA, 2026），解析での事前情報の活用や意思決定の柔軟性などの観点からその有用性が期待されている。

一方で，ハイブリッド対照アプローチには解析上留意する点がいくつかある。その 1 つが RCT データと外部データとの間の差異（prior-data conflict）によるバイアスや第 1 種の過誤確率の上昇の懸念である。この課題に対しては，外部データとして臨床試験の公表情報などの要約データを用いる場合（武田ら, 2015; 野村ら, 2022）および外部データとして個体レベルのデータを用いる場合（野村ら, 2026）について，外部データからの情報借用量を調整したり，交絡因子を調整したりするための種々の方法が提案されている。その 1 つとして，外部データの尤度に対して情報量を割り引くためのパラメータ（power パラメータ）を導入することで，外部データからの情報借用量を調整する power 事前分布がある。従来の conditional power 事前分布および normalized power 事前分布では，RCT データと外部データの異質性に対して個体レベルの共変量データを用いた調整はできない。一方で，傾向スコアに基づく power 事前分布（Wang et al, 2019; Lu et al, 2022）は，個体レベルのデータが利用できることを前提とし，異質性への対処のために傾向スコアを用いることで RCT データと外部データの共変量分布の差異を調整しつつ，power 事前分布の枠組みで power パラメータによる外部データからの情報借用量の調整を行う手法である。これにより，外部データの活用に伴うバイアスの低減と効率的な情報活用の両立が期待される。

本稿では，この外部対照データを RCT に取り込んだハイブリッド対照デザインでの傾向スコアに基づく power 事前分布アプローチの活用に着目し，理論的枠組みを概説する。またこの手法は SAS の PSMATCH プロシジャおよび MCMC プロシジャなどを用いることで実装可能である。PSMATCH プロシジャは傾向スコアの推定および傾向スコアに基づく，マッチング，重み付けおよび層別が実装可能であるため（SAS Institute Inc., 2026; 魚住ら, 2020），傾向スコアに基づく power 事前分布の実装に適している。また，マルコフ連鎖モンテカルロ（MCMC）法に基づいて一般のベイズ流解析を行うことができる MCMC プロシジャ（SAS Institute Inc., 2026）を用いることで，power 事前分布に基づくベイズ推定を柔軟に実行することができる（武田, 2014）。次節以降，傾向スコアに基づく power 事前分布アプローチの概要を説明していく。順を追って理解できるように，外部データを用いない RCT でのベイズ流アプローチから始め，傾向スコアを用いない基本的な power 事前分布（conditional および normalized power 事前分布）について説明した後で，傾向スコアに基づく power 事前分布を説明する。説明に用いるアウトカムは 2 値データ（反応あり，反応なし）とする（ただし，連続データなどの他のアウトカムにも適用可能）。3 節では各アプローチの SAS での実装方法を説明し，4 節ではシミュレーションによる性能評価を行い，本手法の特性および有用性を検討する。

2. 傾向スコアに基づく power 事前分布アプローチの概要

2.1 RCT でのベイズ流アプローチ

ベイズ流の解析では，関心のある治療効果に関するパラメータθの試験開始前の情報を反映した事前分布$\pi_0(\theta)$と試験で得られたデータの情報である尤度$L(\theta|Data)$をベイズの定理によって組み合わせることで，試験後の更新された情報を反映したθに関する事後分布$\pi(\theta)$を形成し，この事後分布に基づいてθに対する推

論を行う。関心のある治療群と対照群の反応率に関するパラメータを$\theta^{(1)}$と$\theta^{(0)}$とすると，それらの事後分布は

$$\pi(\theta^{(1)}|D_{RCT}^{(1)}) \propto L(\theta^{(1)}|D_{RCT}^{(1)})\pi_0(\theta^{(1)})$$
$$\pi(\theta^{(0)}|D_{RCT}^{(0)}) \propto L(\theta^{(0)}|D_{RCT}^{(0)})\pi_0(\theta^{(0)})$$

によって表現される。$D_{RCT}^{(1)}$および$D_{RCT}^{(0)}$は，RCT の治療群および対照群の測定データである。尤度は，2 値データの場合ベルヌーイ分布が用いられる。また，事前分布には共役事前分布であるベータ分布$Beta(a,b)$が用いられることがあり，パラメータa,bによってθに関する事前情報を調整することができる（入江ら, 2022）。外部データの事前情報を考慮しない RCT でのベイズ流の解析では，無情報事前分布として$Beta(1,1)$が用いられることがある。$\theta^{(1)}$と$\theta^{(0)}$のそれぞれの事後分布から，治療群間差$\delta = \theta^{(1)} - \theta^{(0)}$の事後分布を求めることが可能であり，$\delta$の事後分布に基づいて，$\Pr(\delta > 0) > 0.975$などの成功基準を設定し，治療効果に関する意思決定を行うこともできる（FDA, 2026）。本稿では治療群は RCT データのみによる推測を行うため，以降は power 事前分布アプローチを適用した対照群の推測に焦点を当てて説明する。

2.2 Conditional power 事前分布

Conditional power 事前分布は，外部データが事後分布に過度に影響しないように power パラメータα(ただし$0 \le \alpha \le 1$)を導入し，外部データの尤度をα乗することにより外部データの情報を割引く。Conditional power 事前分布ではαは解析者が設定する定数である。このとき conditional power 事前分布は，α乗した外部データの尤度とパラメータの事前分布$\pi_0(\theta^{(0)})$（power 事前分布と区別して初期事前分布と呼ばれる）から

$$\pi(\theta^{(0)}|D_{EXT}^{(0)},\alpha) \propto \left[L(\theta^{(0)}|D_{EXT}^{(0)})\right]^{\alpha}\pi_0(\theta^{(0)})$$

となる。$D_{EXT}^{(0)}$は外部データの対照群の測定データである。この conditional power 事前分布と RCT データの尤度をベイズの定理によって組み合わせることにより対照群の事後分布が求まる。

$$\pi(\theta^{(0)} \mid D_{RCT}^{(0)}, D_{EXT}^{(0)}, \alpha) \propto L(\theta^{(0)}|D_{RCT}^{(0)})\pi(\theta^{(0)}|D_{EXT}^{(0)},\alpha)$$

Power パラメータが$\alpha = 1$の場合はすべての外部データを用いることを，$\alpha = 0$の場合は外部データを全く利用しないことを意味する。また，外部データの例数を$N_{EXT}^{(0)}$とすると，power パラメータαとの積$\alpha N_{EXT}^{(0)}$が解析に用いた情報量（外部データの例数）と解釈できる。

2.3 Normalized power 事前分布

Normalized power 事前分布（modified power 事前分布と呼ばれることもある）は joint power 事前分布の改良版である（製薬協, 2022）。Conditional power 事前分布は power パラメータを定数として扱っていたが，joint power 事前分布は power パラメータを確率変数として推定する。Power パラメータの事前分布を$\pi(\alpha)$としたとき joint power 事前分布は

$$\pi(\theta^{(0)},\alpha|D_{EXT}^{(0)}) \propto \left[L(\theta^{(0)}|D_{EXT}^{(0)})\right]^{\alpha}\pi_0(\theta^{(0)})\pi(\alpha)$$

となる。$0 \le \alpha \le 1$であるため，$\pi(\alpha)$としてベータ分布が用いられることが多い。Joint power 事前分布では尤度原理を満たさないなどの問題が指摘されており（武田ら, 2015; Duan et al, 2006），これに対処するため joint power 事前分布を基準化した以下の事前分布を用いる。

$$\pi(\theta^{(0)},\alpha|D_{EXT}^{(0)}) = \frac{\left[L(\theta^{(0)}|D_{EXT}^{(0)})\right]^{\alpha}\pi_0(\theta^{(0)})}{\int\left[L(\theta^{(0)}|D_{EXT}^{(0)})\right]^{\alpha}\pi_0(\theta^{(0)})d\theta^{(0)}}\pi(\alpha)$$

この事前分布を normalized power 事前分布と呼ぶ。Normalized power 事前分布と RCT データの尤度によって対照群のパラメータの事後分布を求める。Normalized power 事前分布の分母の基準化定数は，一般には数値積分などによって算出する必要があるが，2 値アウトカムなどの一部の変数型では解析的に求めることができる（Wang et al, 2019）。2 値アウトカムで初期事前分布に$Beta(1,1)$を用いた場合（$\pi_0(\theta^{(0)})=1$），基準化定数は

$$\int \left[L\left(\theta^{(0)}\middle|D_{EXT}^{(0)}\right)\right]^{\alpha}\pi_0\left(\theta^{(0)}\right)d\theta^{(0)} = B(\alpha N_{EXT}^{(0)}\bar{Y}_{EXT}+1, \alpha N_{EXT}^{(0)}-\alpha N_{EXT}^{(0)}\bar{Y}_{EXT}+1)$$

となる（一般の$Beta(a,b)$でも求めることが可能である）。ここで$B(\cdot)$はベータ関数であり，$N_{EXT}^{(0)}$は外部データの例数，$\bar{Y}_{EXT}$は外部データのアウトカムの平均値（反応例の割合）を表す。

2.4 傾向スコアに基づく Power 事前分布

外部データを利用する場合は，対象者の背景因子の分布が RCT データと外部データで異なる可能性があるため，治療群と対照群（RCT データの対照群と外部データを併合したハイブリッド対照）との間の比較可能性を考慮する必要がある。RCT データの背景因子の分布が群間で類似していても，RCT データと外部データで背景因子が異なる場合，外部データをそのまま RCT データにプールして解析することは適切ではない。傾向スコアに基づく power 事前分布アプローチはこの問題に対処する方法の 1 つである。傾向スコアは，交絡調整を行い他の統計手法と組み合わせることで，比較可能性を高めた解析を行うことができる汎用性の高い統計手法である。本稿で着目した傾向スコアに基づく方法は，RCT データと共変量の値が類似した外部データを傾向スコアに基づき，選択・層別化したうえで情報借用を行うことにより，RCT データと外部データの共変量の分布の違いによるバイアスを軽減し，外部データを活用した推論を可能にする。

傾向スコア$e(\boldsymbol{X})$は，観測された共変量$\boldsymbol{X}$が与えられたときに，特定の処置（治療，曝露など）に割り当てられる条件付き確率として知られている（Rosenbaum and Rubin, 1983）。本稿で説明する傾向スコアに基づく power 事前分布アプローチでは，共変量$\boldsymbol{X}$が与えられたときに特定のデータソース$Z\left(Z=0:\text{RCT データ}, Z=1:\text{外部データ}\right)$に属する条件付き確率となり，ここでは RCT データに属する条件付き確率を傾向スコアと定義する。

$$e(\boldsymbol{X}) = \Pr(Z=0|\boldsymbol{X})$$

RCT データの傾向スコアの範囲外にある外部データの対象者は，RCT 集団とは背景因子が異なるとして解析対象から除外（トリミング）し，傾向スコアに基づき対象者をいくつかの層に分ける。ここでは層の数をKとする$(k=1,\cdots,K)$。傾向スコアに基づき対象者を層別化することにより，共変量がより類似する対象者のグループに分類することができる。層別に用いる傾向スコアのカットオフ値は，各層の RCT データの例数がほぼ均等となるように設定することが考えられる（Lu et al, 2022）。層kでの治療群および対照群の反応率のパラメータを$\theta_k^{(1)}$および$\theta_k^{(0)}$とすると，層kの対照群のパラメータ$\theta_k^{(0)}$の power 事前分布は

$$\pi\left(\theta_k^{(0)}|D_{k,EXT}^{(0)},\alpha_k\right) \propto \left[L\left(\theta_k^{(0)}\middle|D_{k,EXT}^{(0)}\right)\right]^{\alpha_k}\pi_0\left(\theta_k^{(0)}\right)$$

となる。$D_{k,EXT}^{(0)}$は層kに属する外部データの対照群の対象者の測定データである。層kの power パラメータα_kを

$$\alpha_k = \lambda_k / N_{k,EXT}$$

$$\lambda_k = \min\left(\frac{A\, r_k}{\sum_{k'=1}^{K} r_{k'}}, N_{k,EXT}\right)$$

と定義する。$N_{k,EXT}$は層kの外部データの例数である。Aは外部データ全体から借用したい名目上の対象者数

であり解析者側で事前に規定する。r_kは層kにおける RCT データと外部データの共変量の類似度を示す指標であり，RCT データおよび外部データの層kの傾向スコアの確率密度関数を$g_{k,RCT}$および$g_{k,EXT}$とすると，

$$r_k = \int_0^1 \min\left[\hat{g}_{k,RCT}(u), \hat{g}_{k,EXT}(u)\right] du$$

となる。これは 2 つの傾向スコアの確率密度関数の曲線下面積が重なる領域となる。$Ar_k / \sum_{k'=1}^{K} r_{k'}$は，RCT データと外部データの傾向スコアの分布の重なりの程度に応じて重み付けられた各層から借用したい例数であり，傾向スコアが類似している対象者の多い層からより多くの情報を借用することになる。ただし，この重み付き例数が$N_{k,EXT}$以上の場合は$\alpha_k = 1$となり，その層のすべての外部データを解析に用いるようλ_kを設定している。本稿ではこの power パラメータを用いた方法を傾向スコアに基づく power 事前分布（固定配分）と呼ぶ。この傾向スコアに基づく power 事前分布（固定配分）と RCT データの尤度から，対照群の各層のパラメータの事後分布が求まる。

$$\pi\left(\theta_1^{(0)}, \cdots, \theta_K^{(0)} | \boldsymbol{D}_{RCT}^{(0)}, \boldsymbol{D}_{EXT}^{(0)}, \alpha_1, \cdots, \alpha_K\right) \propto \prod_{k=1}^{K} \left[L\left(\theta_k^{(0)} | D_{k,RCT}^{(0)}\right)\right] \pi\left(\theta_1^{(0)}, \cdots, \theta_K^{(0)} | \boldsymbol{D}_{EXT}^{(0)}, \alpha_1, \cdots, \alpha_K\right)$$

ここで，$\boldsymbol{D}_{RCT}^{(0)} = \left(D_{1,RCT}^{(0)}, \cdots, D_{K,RCT}^{(0)}\right)$および$\boldsymbol{D}_{EXT}^{(0)} = \left(D_{1,EXT}^{(0)}, \cdots, D_{K,EXT}^{(0)}\right)$は RCT および外部データの対照群の各層に属する対象者の測定データである。また，治療群の各層のパラメータの事後分布は以下となる。

$$\pi\left(\theta_1^{(1)}, \cdots, \theta_K^{(1)} | \boldsymbol{D}_{RCT}^{(1)}\right) \propto \prod_{k=1}^{K} \left[L\left(\theta_k^{(1)} | D_{k,RCT}^{(1)}\right)\right] \pi_0\left(\theta_1^{(1)}, \cdots, \theta_K^{(1)}\right)$$

$\boldsymbol{D}_{RCT}^{(1)} = \left(D_{1,RCT}^{(1)}, \cdots, D_{K,RCT}^{(1)}\right)$は RCT の治療群の各層に属する対象者の測定データである。関心のある治療群間差δ（治療群と対照群の反応率の差）は，各層の RCT データの例数の割合に応じた重み付けにより算出するが，各層の RCT データの例数がほぼ等しくなるように傾向スコアのカットオフ値を設定している場合は以下となる（Lu et al, 2022）。

$$\delta = \frac{1}{K} \sum_{k=1}^{K} \left(\theta_k^{(1)} - \theta_k^{(0)}\right)$$

各層の power パラメータα_kの決め方について，固定配分では層kの配分（v_k）を$r_k / \sum_{k'=1}^{K} r_{k'}$としたが，これを確率変数とするベイズ流のアプローチも提案されている。$v_k \geq 0$かつ$\sum_{k=1}^{K} v_k = 1$であるため，v_kの事前分布として，r_kをパラメータとした以下のディリクレ事前分布を用いることができる。ここで，Rはv_kの事前分散を制御する定数であり（$R \to 0$とすると固定配分），解析者が事前に設定する（Wang et al, 2019）。

$$\pi(v_1, \cdots, v_K) = Dir\left(\frac{r_1}{R}, \cdots, \frac{r_K}{R}\right)$$

このとき power パラメータは定数ではないことから，2.3 項と同様に基準化した以下の power 事前分布を用いる（Duan et al, 2006）。初期事前分布$\pi_0\left(\theta_1^{(0)}, \cdots, \theta_K^{(0)}\right)$を無情報事前分布とすると，傾向スコアに基づく power 事前分布は以下となり，本稿では傾向スコアに基づく power 事前分布（ベイズ配分）と呼ぶ。

$$\pi\left(\theta_1^{(0)}, \cdots, \theta_K^{(0)}, v_1, \cdots, v_K | \boldsymbol{D}_{EXT}^{(0)}\right) = \frac{\prod_{k=1}^{K} \left[L\left(\theta_k^{(0)} | D_{k,EXT}^{(0)}\right)\right]^{\alpha_k}}{\int \prod_{k=1}^{K} \left[L\left(\theta_k^{(0)} | D_{k,EXT}^{(0)}\right)\right]^{\alpha_k} d\theta_1^{(0)} \ldots d\theta_K^{(0)}} \pi_0\left(\theta_1^{(0)}, \cdots, \theta_K^{(0)}\right) \pi(v_1, \ldots, v_K)$$

分母の積分は，2 値アウトカムに対しては 2.3 項と同様に解析的に求めることができ，各層の外部データの例数，アウトカムの平均値（反応例の割合），$\theta_k^{(0)}$の初期事前分布のパラメータを用いて算出できる。この外部データの情報を含めた事前分布と RCT データの尤度によって，対照群の各層のパラメータの事後分布が求まる。

$$\pi\left(\theta_1^{(0)},\cdots,\theta_K^{(0)},v_1,\cdots,v_K|\boldsymbol{D}_{RCT}^{(0)},\boldsymbol{D}_{EXT}^{(0)}\right)\propto\prod_{k=1}^{K}\left[L\left(\theta_k^{(0)}|D_{k,RCT}^{(0)}\right)\right]\pi\left(\theta_1^{(0)},\cdots,\theta_K^{(0)},v_1,\cdots,v_K|\boldsymbol{D}_{EXT}^{(0)}\right)$$

このように，傾向スコアに基づく power 事前分布には，α_kを解析者が決める定数，もしくは確率変数とする 2 つの方法があり，power パラメータの決め方以外については固定配分では層ごとに conditional power 事前分布，ベイズ配分では層ごとに normalized power 事前分布と同様の形式の power 事前分布を用いることによって外部データの借用を行っていると考えることができる。

3. SAS による実行

本節では，2 節で説明した解析方法を SAS の MCMC プロシジャなどを用いて実装する方法を説明する。解析に用いるデータセットについて簡単に説明した後，2 節で説明した解析方法と同じ順で説明する。解析に用いるデータセットを表 1 に示す。データセットに含まれる変数は，SOURCE（データソース，0:RCT データ，1:外部データ），TRT（群，1:治療群，0:対照群），SUBJ（対象者番号），X1-X3（共変量 1~3），Y（アウトカム，1:反応あり，0:反応なし）である。本稿では，RCT の治療群および対照群の例数を 200 例および 100 例，外部対照のデータは 1000 例とした。このデータでは RCT の治療群と対照群の反応率に群間差があり，また，外部対照データは RCT と共変量の分布が異なり，結果として反応率も RCT の対照群と異なる。パラメータの設定値を含むデータ生成方法の詳細は 4.1 項で説明する。各手法の実行結果は 3.5 項にまとめて示す。

表 1 解析データセット（EXAMPLE）

SOURCE	TRT	SUBJ	X1	X2	X3	Y
0	0	1	7.57	8.16	10.07	0
:	:	:	:	:	:	:
0	0	100	4.75	5.61	6.82	0
0	1	1	6.79	6.37	8.04	1
:	:	:	:	:	:	:
0	1	200	5.81	8.36	7.62	0
1	0	1	5.95	7.44	8.00	0
:	:	:	:	:	:	:
1	0	1000	6.84	8.66	9.04	0

3.1 RCT

RCT データのみで無情報事前分布を用いたベイズ流の解析を行うための SAS コードを以下に示す。なお，共役事前分布であるベータ分布を用いた場合は事後分布を解析的に求めることもできるが（入江ら，2022），本稿では MCMC プロシジャによってパラメータの事後分布を近似する。

```
proc mcmc data=EXAMPLE nmc=50000 nbi=2000 thin=5 seed=9999 outpost=OUT1
monitor=(p_t p_p dif);
```

```
  parms p_t 0.5;
  parms p_p 0.5;
  prior p_t ~ beta(1,1);
  prior p_p ~ beta(1,1);

  if TRT = 1 then do;
    loglik = logpdf('bernoulli', Y, p_t);
  end;
  if TRT = 0 then do;
    loglik = logpdf('bernoulli', Y, p_p);
  end;

  model general(loglik);

  beginnodata;
    dif = p_t - p_p;
  endnodata;
  where SOURCE = 0;
run;
```

ここでは外部データは利用しないため，WHERE 文で RCT データ（SOURCE=0）に限定している。PROC MCMC 文では，MCMC の反復回数（nmc），バーンインの反復回数（nbi），事後サンプルの間引き間隔（thin），疑似乱数のシード値（seed），事後サンプルを格納するデータセット（outpost）を指定することができ，monitor=で指定した変数もデータセットに含まれる。また，diag=all や plot(smooth)=all といったオプションを追加することで，トレースプロットや収束診断の出力を制御できる。PARMS 文ではモデルのパラメータおよびその初期値を定義し，上記では治療群の反応率（p_t）および対照群の反応率（p_p）を定義している。なお，1 つの PARMS 文で複数のパラメータを指定することも可能であるが，同一の PARMS 文で指定されたパラメータは基本的にブロック化され，MCMC の過程で多変量の提案分布から得られた疑似乱数の組に対して採択または棄却が決定されるため，状況によってはサンプリングが非効率になる可能性がある。一方で，多数のパラメータのブロックを逐次更新する場合は計算コストが増加するなどの問題もある。このように，PARMS 文でのパラメータの指定方法は MCMC の挙動や性能に影響する可能性があるため，（最適なブロック化の特定は容易ではないかも知れないが）特に多数のパラメータを扱う場合は注意が必要である。詳細は MCMC プロシジャのマニュアル（SAS Institute Inc., 2026）を参照されたい。PRIOR 文では，各パラメータの事前分布を指定し，上記では p_t および p_p に対して無情報事前分布として，$Beta(1,1)$を指定している。MODEL 文では目的変数と目的変数が従う分布（尤度関数）を指定する。2 値データの場合は通常 model Y ~ binary(p_t)のように指定を行うが，power 事前分布への拡張を考慮し，プログラム文で定義した一般の対数尤度を指定できる general(対数尤度)を用いた。対数尤度（上記のコードでは loglik）の計算には確率（密度）関数の対数値を返す logpdf 関数を用いることができる。本稿では個別対象

者データが利用可能な状況を想定していることからベルヌーイ分布を指定している。反応率の群間差（dif）の算出のように解析データセットのオブザベーションに依存しない計算は，BEGINNODATA/ENDNODATA 文で囲むことにより，データセットのすべてのオブザベーションを処理するステップをスキップし，各 MCMC 反復における最初と最後のオブザベーションの 2 回のみ実行されることで計算時間を短くできる。dif のように PARMS 文で定義していない変数は，PROC MCMC 文の monitor オプションで指定しないと，事後サンプルを格納するデータセットに出力されないので注意が必要である。

MCMC プロシジャを実行すると各パラメータの事後分布の要約（事後平均，事後標準偏差（SD），95% 最高事後密度信用区間（HPD CI））や MCMC の収束診断結果が得られると共に，各パラメータの事後サンプルを格納したデータセット（OUT1）が出力される。このデータセットを用いて，dif＞0 となるオブザベーションの割合を求めることによって，群間差δが 0 よりも大きい確率$\Pr(\delta > 0)$を求めることができる。

3.2 Conditional power 事前分布

RCT データに対して外部データを利用するハイブリッド対照デザインにおいて，外部データの借用量を解析者が事前に規定する conditional power 事前分布アプローチの実装方法を示す。

```
proc mcmc data=EXAMPLE /*nmc, nbi, thin, seed は 3.1 RCT と同様*/  monitor=(p_t p_p dif);
  parms p_t 0.5;
  parms p_p 0.5;
  prior p_t ~ beta(1,1);
  prior p_p ~ beta(1,1);

  begincnst;
    alpha = 0.1;
  endcnst;

  if TRT = 1 then do;
    loglik = logpdf('bernoulli', Y, p_t);
  end;
  if TRT = 0 then do;
    loglik = logpdf('bernoulli', Y, p_p);
    if SOURCE = 1 then do;
      loglik = alpha * loglik;
    end;
  end;

  model general(loglik);

  beginnodata;
```

```
    dif = p_t - p_p;
  endnodata;
run;
```

3.1 項の RCT のベイズ流解析と異なる点としては，借用量を規定する power パラメータαの導入である。Conditional power 事前分布では，この power パラメータを解析者が事前に規定することになり，本稿では$\alpha = 0.1$とした。これは，外部データが 1000 例ある場合に 100 例分(= 1000 × 0.1)の情報量を借用することに相当する。αの指定は MCMC の反復計算を行う前に 1 度だけ処理されればよいため，BEGINCNST/ENDCNST 文の中で設定することで計算時間を短くできる。尤度については，治療群と対照群で設定方法が異なる。対照群については，IF 文により，RCT データの場合（SOURCE=0）は 3.1 項と同じように尤度を設定し，外部データ（SOURCE=1）の場合は，外部データの尤度を power パラメータによってα乗（対数尤度なのでα倍）することで対照群の尤度を構成し，general 関数でこの尤度をモデル化する。

3.3 Normalized power 事前分布

Conditional power 事前分布では power パラメータを解析者が事前に規定（$\alpha = 0.1$）したが，ここでは power パラメータαを確率変数とし，RCT データと外部データの類似性によって動的にαを決定する。

```
proc mcmc data=EXAMPLE2 /*nmc, nbi, thin, seed は 3.1 RCT と同様*/  monitor=(p_t p_p alpha
dif);
  parms p_t 0.5;
  parms p_p 0.5;
  parms alpha 0.5;
  prior p_t ~ beta(1,1);
  prior p_p ~ beta(1,1);
  prior alpha ~ beta(1,1);

  if TRT = 1 then do;
    loglik = logpdf('bernoulli', Y, p_t);
  end;
  if TRT = 0 then do;
    loglik = logpdf('bernoulli', Y, p_p);
    if SOURCE = 1 then do;
      loglik = alpha * loglik;
      if SUBJ = 1 then do;
        loglik = loglik - logbeta(alpha * n_ext * p_ext + 1, alpha * n_ext * (1 - p_ext) + 1);
      end;
    end;
  end;
```

```
  model general(loglik);

  beginnodata;
    dif = p_t - p_p;
  endnodata;
run;
```

Power パラメータαを PARMS 文でパラメータ（確率変数）として定義し，本稿では PRIOR 文で無情報の事前分布$Beta(1,1)$を設定する（情報を与えて借用量を調整することもできる（野村ら, 2022））。基準化定数の計算については外部データ全体の尤度に対する基準化であるため，MCMC プロシジャで読み込む解析データセットのオブザベーションが複数であってもこの引き算（前述した式では割り算の分母に相当するが，対数尤度のため引き算となる）は 1 回のみ行われるように指定する必要があり（武田, 2014），ここでは SUBJ=1 のときのみ計算が行われるよう条件をつけている。なお，この引き算の部分については，本稿では 2 値アウトカムを用いているため，ベータ関数を用いて解析的に計算が可能である（2.3 項参照）。対数尤度に変換するためにベータ関数で計算した値を log 関数の引数として用い対数変換することも可能であるが，引数がとても小さい値のときに 0 と扱われて log 関数では計算不能となる場合がある。これを回避するため，log 関数を使わずにベータ関数の対数値を返す logbeta 関数を用いた。この計算には外部データの例数および反応例の割合を求める必要がある。コードは省略したが，これらの変数についてはあらかじめ FREQ プロシジャなどで計算しておき，MCMC プロシジャで読み込む解析データセットにマージしておく。したがって，3.1 および 3.2 項で用いたデータセット（EXAMPLE）に対して外部データの例数（N_EXT）および反応例の割合（P_EXT）の変数をマージした別のデータセット（EXAMPLE2）を MCMC プロシジャで読み込ませていることに注意が必要である。上述したように SUBJ=1 のオブザベーションのときのみこれらの変数が用いられるが，処理上すべてのオブザベーションに含めておいたままでも問題はない。外部データの例数および反応例の割合の値を参照する別の方法としては，read_array 関数を用いることで，これらの値を含んだデータセットを配列として読み込むことも可能である。

3.4 傾向スコアに基づく power 事前分布

傾向スコアに基づく power 事前分布（固定配分）について実行例を示す。ここでは傾向スコアの算出に PSMATCH プロシジャを用い，各層の power パラメータはデータステップなどで算出した。算出した power パラメータも含めた解析データセットを MCMC プロシジャで読み込むことで解析を行う。

```
ods output StrataInfo = out_Num;
proc psmatch data = EXAMPLE region = treated(extend = 0.0);
  class SOURCE;
  psmodel SOURCE(treated = '0') = X1 X2 X3;
  strata nstrata = 5 key = treated;
  output out(obs = region) = OUT_PS ps = PS strata = STRATA ;
```

```
run;

proc sort data = OUT_PS;
  by SOURCE STRATA;
run;

proc kde data = OUT_PS;
  univar ps / out = out_kde gridl = 0.0 gridu = 1.0 ngrid = 1001;
  by SOURCE STRATA;
run;

proc sort data = out_kde;
  by strata value SOURCE ;
run;

proc transpose data = out_kde out = out_kde_1 prefix = density_;
  by strata value;
  id SOURCE;
  var density;
run;

data out_kde_2;
  set out_kde_1;
  by strata;
  length overlap prev_x prev_y area 8.;
  overlap = min(density_0, density_1);
  prev_x = lag(value);
  prev_y = lag(overlap);
  if not first.strata then area = (value - prev_x)*(overlap + prev_y) / 2;* 台形法 ;
run;

data out_kde_3;
  set out_kde_2;
  output; * 層別集計用 ;
  strata = 0; * 全体集計用 ;
  output;
run;
```

```
proc sort data=out_kde_3;
  by strata;
run;

proc means data = out_kde_3 sum ;
  var area;
  by strata;
  output out = OUT_R sum = R; * RCT データと外部データの共変量の類似性を示す指標;
run;
*ここまでがベイズ配分と同じ;

data r_total;
   set OUT_R(where=(STRATA=0) ); * r の合計値のみ抽出 ;
   keep r;
   rename r=r_total;
run;

data alpha;
   if _n_=1 then set r_total;
  merge OUT_R(where=(STRATA^=0)) OUT_Num(rename=(StratumIndex= strata));
  by strata ;
  A = 100;
  lambda = min(A*r/r_total ,  ControlFreq);
  if ControlFreq ^= . then alpha = lambda / ControlFreq;* alpha(Power Parmeter) ;
  SOURCE = 1;*マージキーとするため変数を補足;
  keep SOURCE strata alpha;
run;

data EXAMPLE3;
  merge out_ps alpha;
  by SOURCE strata;
run;

proc mcmc data= EXAMPLE3 /*nmc, nbi, thin, seed は 3.1 RCT と同様*/  monitor=(p_t1-p_t5
p_p1-p_p5 p_trt p_plc dif);
  array p_t[5];
  array p_p[5];
  parms p_t: 0.5;
```

```
  parms p_p: 0.5;
  prior p_t: ~ beta(1, 1);
  prior p_p: ~ beta(1, 1);

  if TRT = 1 then do;
    loglik = logpdf('bernoulli', Y, p_t[strata]);
  end;
  if TRT = 0 then do;
    loglik = logpdf('bernoulli', Y, p_p[strata]);
    if SOURCE = 1 then do;
      loglik = alpha * loglik;
    end;
  end;

  model general(loglik);

  beginnodata;
    p_trt = (p_t1 + p_t2 + p_t3 + p_t4 + p_t5) / 5;
    p_plc = (p_p1 + p_p2 + p_p3 + p_p4 + p_p5) / 5;
    dif = p_trt - p_plc;
  endnodata;
run;
```

PSMATCH プロシジャの主な設定について説明する。プロシジャの詳細は魚住ら（2020）を参照頂きたい。region オプションで treated を指定しているが，PSMODEL 文でデータソース（SOURCE）を指定しているため，ここでは RCT データを意味する。SAS のデフォルトでは，region オプションで指定した層別の対象となる領域は，傾向スコアのロジットの共通標準偏差のプールした推定値の 0.25 倍分が拡張されるが，RCT データの傾向スコアの範囲外にある外部データの対象者を除外（トリミング）するために extended=0.0 としている。STRATA 文の nstrata オプションにより層の数を指定する（$k = 5$）。key オプションで treated を指定した場合，各層の RCT データの例数がほぼ均等となるように傾向スコアのカットオフ値が設定されるが，外部データも同じカットオフ値で層別されるため，層によって外部データの例数は異なりうる。PSMATCH プロシジャから出力される各対象者の層および傾向スコアの推定値を含んだデータセットから，いずれの層にも含まれなかった（RCT データの傾向スコアの範囲外であった）対象者を除外したデータセット（OUT_PS）を用いて，カーネル密度推定を行う KDE プロシジャで傾向スコアの確率密度関数の推定値を算出する。KDE プロシジャから出力されたデータセット（OUT_KDE）をもとに，データステップと基本的なプロシジャを用いて各層の RCT と外部データの分布の重なりを算出する（データセット OUT_R）。PSMATCH プロシジャで算出される各層の対象者数を含むデータセット（OUT_NUM）と OUT_R をマージして各層の power パラメータを求める。本稿では，外部データ全体から借用したい名目上の対象者数は$A =$

100と設定した。MCMC プロシジャで読み込む解析データセットには，層別後のデータセット（OUT_PS）に各層の power パラメータを含むデータセット（ALPHA）をマージしたものを用いる（EXAMPLE3）。MCMC プロシジャでは，各層の反応率のパラメータを定義するため，ARRAY 文で層の数に対応するパラメータを配列（p_t および p_p）として指定する。各群の反応率パラメータ（p_trt および p_plc）については，各層の反応率のパラメータの重み付き平均とし，それらの差として反応率の群間差を求める。

最後に各層の配分をベイズ流で算出する場合の SAS コードを以下に示す。なお，傾向スコアの算出から各層の RCT および外部データの傾向スコアの分布が重なる領域の算出までは固定配分と同様である。ベイズ配分については，層kの配分v_kを確率変数とし，その事前分布のディリクレ事前分布のパラメータとして用いるr_k（固定配分で作成したデータセット OUT_R が該当）を MCMC プロシジャで読み込ませる解析データセットにマージしておく（EXAMPLE4）。また，ベイズ配分では power パラメータも確率変数として扱うことから，normalized power 事前分布と同様に基準化定数の計算が必要となる。そのため，事前に各層の例数（Y_N），反応例数（N_R），非反応例数（N_NR）も算出しておき，併せて MCMC プロシジャで読み込む解析データセット（EXAMPLE4）にマージしておく。

```
*傾向スコアの算出，各層の RCT と外部データの傾向スコアの分布の重なり r1,...r5 の算出までは固定配分と同様;

proc mcmc data= EXAMPLE4 /* nmc, nbi, thin, seed は 3.1 RCT と同様*/ monitor=(p_t1-p_t5
p_p1-p_p5 p_trt p_plc dif v1-v5 alpha1-alpha5);
  array p_dir[5]; *ディリクレ事前分布のパラメータ用の配列;
  array v[5];
  array alpha[5];
  array p_t[5];
  array p_p[5];
  parms v: 0.5;
  parms p_t: 0.5;
  parms p_p: 0.5;
  prior p_t: ~ beta(1, 1);
  prior p_p: ~ beta(1, 1);

  begincnst;
    A = 100;
    R=0.1; *事前分散パラメータ;
    p_dir[1] = r1/R;
    p_dir[2] = r2/R;
    p_dir[3] = r3/R;
    p_dir[4] = r4/R;
    p_dir[5] = r5/R;
  endcnst;
```

```
  prior v~ dirichlet(p_dir);
  alpha[strata] = min(A*v[strata]/Y_N,1);

  if TRT = 1 then do;
    loglik = logpdf('bernoulli', Y, p_t[strata]);
  end;
  if TRT = 0 then do;
    loglik = logpdf('bernoulli', Y, p_p[strata]);
    if SOURCE = 1 then do;
      loglik = alpha[strata] * loglik;
      if FIRST = 1 then do;     /*基準化定数，外部データの各層で 1 回だけ実行*/
        loglik = loglik - logbeta(alpha[strata]*N_r+1, alpha[strata]*N_nr+1);
      end;
    end;
  end;

  model general(loglik);

  beginnodata;
    p_trt = (p_t1 + p_t2 + p_t3 + p_t4 + p_t5)/5;
    p_plc = (p_p1 + p_p2 + p_p3 + p_p4 + p_p5)/5;
    dif = p_trt - p_plc;
  endnodata;
run;
```

層kの配分v_kに対するディリクレ事前分布のパラメータを設定するために ARRAY 文により配列 p_dir を指定する。PARMS 文でv_kを指定し，PRIOR 文でディリクレ事前分布を指定することでベイズ流に配分を決めることができる。本稿では，v_kの事前分散を制御するパラメータを$R = 0.1$と設定した。v_kに基づいて算出される power パラメータα_kの配列も指定する。基準化の処理（3.3 項参照）については，normalized power 事前分布と同様に各層で 1 回だけ基準化定数を算出するために，各層の最初のオブザベーションを特定するためのフラグ変数（FIRST）を導出しておく必要がある。

3.5 各手法の解析結果

各手法について表 1 のデータを使って解析した結果を以下に示す。治療群，対照群の事後要約（事後平均，事後 SD）と群間差の事後平均，95%HPD CI および群間差が 0 を上回る確率$\Pr(\delta > 0)$を表 2 に示した。治療群はいずれの手法も RCT のデータのみを用いているため，事後平均および事後 SD はほぼ同じである。ただし，傾向スコアに基づく power 事前分布については，傾向スコアで層別した各層のパラメータの重

み付き平均であるため，他の手法の事後分布とわずかに異なっている。一方で対照群については外部データを利用するため，事後平均は手法間で異なっており，事後 SD は RCT と比較して他の手法で小さかった。

表 2 各方法の解析結果

	治療群（$\theta^{(1)}$）		対照群（$\theta^{(0)}$）		群間差（δ）		
方法	事後平均	事後 SD	事後平均	事後 SD	事後平均	95%HPD CI	$\Pr(\delta > 0)$
RCT	0.525	0.035	0.333	0.046	0.192	(0.073, 0.304)	0.999
Conditional	0.525	0.035	0.287	0.032	0.238	(0.144, 0.330)	1.000
Normalized	0.525	0.035	0.268	0.032	0.256	(0.166, 0.350)	1.000
PS (Fixed)	0.511	0.031	0.362	0.030	0.148	(0.065, 0.236)	1.000
PS (Bayes)	0.510	0.032	0.361	0.031	0.149	(0.063, 0.237)	1.000

Conditional: Conditional power 事前分布，Normalized: Normalized power 事前分布，PS（Fixed）: 傾向スコアに基づく power 事前分布（固定配分），PS（Bayes）: 傾向スコアに基づく power 事前分布（ベイズ配分）

4. シミュレーションによる性能評価

本節では，シミュレーションにより傾向スコアに基づく power 事前分布アプローチの性能評価を行う。比較対象として，RCT データに対するベイズ流の解析（外部データは利用しない），傾向スコアを用いない基本的な power 事前分布による外部データを利用した解析（conditional および normalized power 事前分布）もシミュレーションに含めた。各手法での解析者が規定するパラメータなどの設定は 3 節と同じとした。

4.1　シミュレーションデータの設定

評価項目は 2 値アウトカムY（1: 反応あり，0: 反応なし），共変量は 3 つ（$X1, X2, X3$）とし，いずれも多変量正規分布に従う連続量とした。RCT データの共変量の分布は群間で同じとし，3 つの共変量と投与群を説明変数としたロジスティック回帰モデルで求めた各個体の反応率のベルヌーイ分布に従うアウトカムを生成した。外部データの共変量は RCT とは異なる分布を設定し，外部データのアウトカムの反応率は RCT データと同じロジスティック回帰モデルで求めた。そのため RCT の対照群と外部対照群の反応率は異なる。関心のある推定対象は，アウトカムの反応率の治療群間差δである。頻度論での統計的仮説検定の帰無仮説に相当するδの真値を定数 0 とした場合（帰無仮説下）およびδの真値を定数 0.15 とした場合（対立仮説下）の 2 つのシナリオでシミュレーションを行い, 動作特性を評価した。RCT データおよび外部データの各群の例数，共変量の平均値および分散共分散，各シナリオ（帰無仮説下および対立仮説下）での各群の真の反応率の設定を表 3 に示す。疑似乱数を発生させ，各シナリオで 10000 回分のシミュレーションデータを生成した。なお，3 節で用いたデータセットは，このシミュレーションデータ（対立仮説下）の 1 回分のデータを用いた。

4.2 評価方法

評価指標として，δのバイアス（事後平均と真値の差）の平均値および平均 2 乗誤差（MSE，バイアスの 2 乗の平均値），$\Pr(\delta > 0) > 0.975$を満たす割合（帰無仮説下では第 1 種の過誤確率に相当し，対立仮説下では検出力に相当する），被覆確率（95%HPD CI が真値を含む割合）および 95%HPD CI の幅の平均値を用

いた（FDA, 2026）。加えて，動的に外部データの借用量を決定する normalized power 事前分布および傾向スコアに基づく power 事前分布では，事前分布の情報量をサンプルサイズの単位であらわした事前有効サンプルサイズ（ESS）も評価した。一般の事前分布に対して事前 ESS を求めるための方法はいくつか提案されているが（野村ら, 2022），本稿では power 事前分布を対象としているため，外部データの例数$N_{EXT}^{(0)}$，power パラメータαおよび初期事前分布のベータ分布の事前 ESS（$a+b$）から，事前 ESS を$\alpha N_{EXT}^{(0)}+a+b$とした。ただし，傾向スコアに基づく power 事前分布では層ごとに power パラメータが異なることから層ごとに ESS を計算した。

表 3 シミュレーションデータの設定

		RCT データ	外部データ
症例数		治療群：200 対照群：100	対照群：1000
共変量	平均値	$E\left[\begin{pmatrix}X1\\X2\\X3\end{pmatrix}\right]=\begin{pmatrix}6.0\\7.0\\8.0\end{pmatrix}$	$E\left[\begin{pmatrix}X1\\X2\\X3\end{pmatrix}\right]=\begin{pmatrix}6.6\\7.6\\8.6\end{pmatrix}$
	分散共分散	$Var\left[\begin{pmatrix}X1\\X2\\X3\end{pmatrix}\right]=\begin{pmatrix}1 & 0.5 & 0.5\\0.5 & 1 & 0.5\\0.5 & 0.5 & 1\end{pmatrix}$	$Var\left[\begin{pmatrix}X1\\X2\\X3\end{pmatrix}\right]=\begin{pmatrix}1 & 0.5 & 0.5\\0.5 & 1 & 0.5\\0.5 & 0.5 & 1\end{pmatrix}$
シナリオ	真の反応率 （H_1：対立仮説下）	治療群：0.50 対照群：0.35	対照群：0.25
	真の反応率 （H_0：帰無仮説下）	治療群：0.35 対照群：0.35	対照群：0.25

4.3 結果

第 1 種の過誤確率および検出力以外の評価指標は，帰無仮説下と対立仮説下でほぼ同じであったため，その他の評価指標については，対立仮説下の結果のみを示す。バイアスの平均値は，conditional および normalized power 事前分布で正のバイアス（外部データの反応率が RCT データよりも小さいため）が認められたが，それ以外ではバイアスが認められなかった（図 1 左）。MSE は，conditional および normalized power 事前分布では RCT と比較して大きかったが，傾向スコアに基づく power 事前分布では固定配分とベイズ配分ともに RCT と比較して小さかった（図 1 右）。第 1 種の過誤確率は，RCT および傾向スコアに基づく power 事前分布（固定配分およびベイズ配分）では 2.5%に近かったが，conditional および normalized power 事前分布では 15.7%および 23.8%だった（図 2 左）。検出力は，RCT では 69.7%だったが，power 事前分布アプローチではいずれも 90%以上だった（図 2 右）。被覆確率は，RCT および傾向スコアに基づく power 事前分布（固定配分およびベイズ配分）では 95%に近かったが，conditional および normalized power 事前分布では 95%を下回った（図 3 左）。95% HPD CI の幅の平均値は，RCT と比較していずれの power 事前分布でも小さかったが，conditional および normalized power 事前分布よりも傾向スコアに基づく power 事前分布（固定配分およびベイズ配分）が相対的に小さかった（図 3 右）。

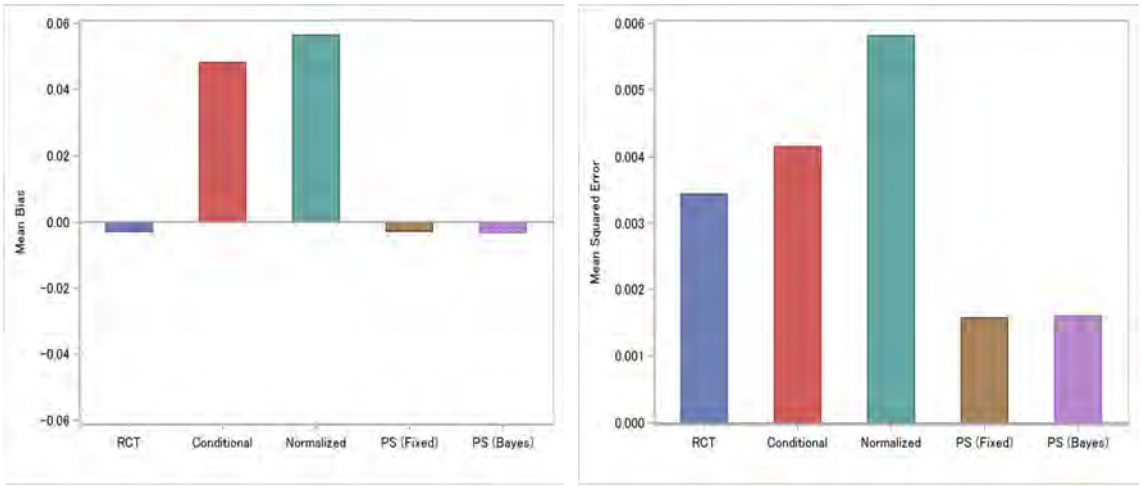


図 1 バイアスの平均値（左）および MSE（右）

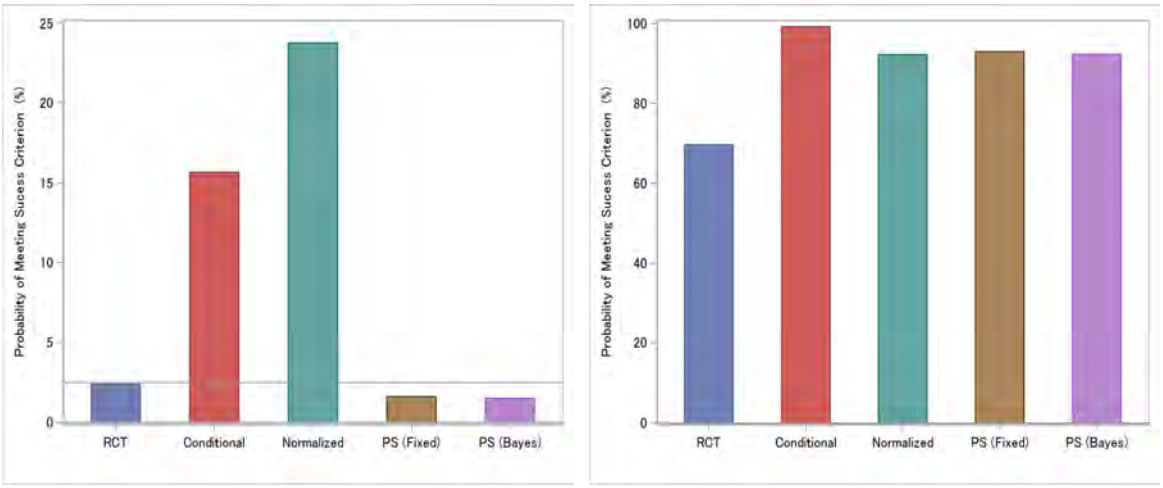


図 2 第 1 種の過誤確率（左）および検出力（右）

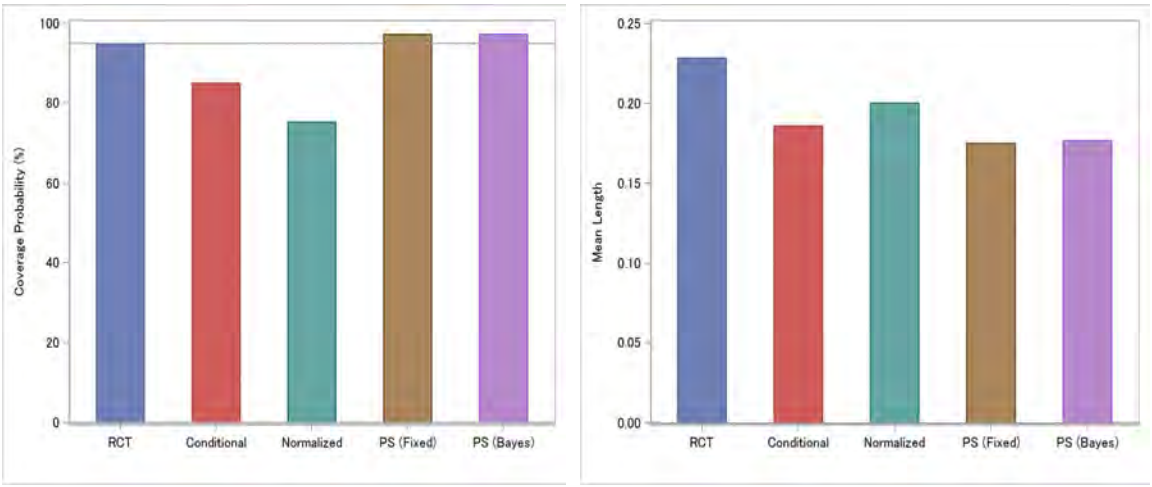


図 3 95% HPD CI の被覆確率（左）および幅の平均値（右）

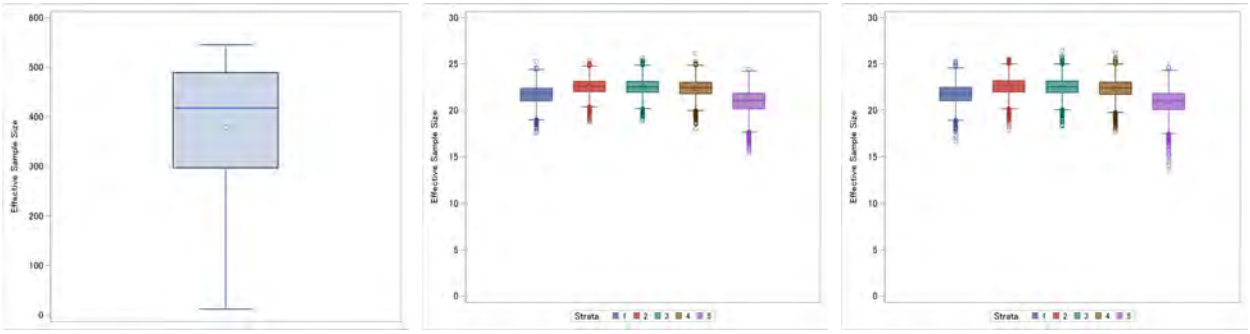


図 4 事前 ESS（左：normalized power 事前分布，中央：傾向スコアに基づく power 事前分布（固定配分），右：傾向スコアに基づく power 事前分布（ベイズ配分））

Normalized power 事前分布の事前 ESS の中央値および 4 分位範囲は，417.5（296.4～488.8）であり（図 4），conditional power 事前分布で$\alpha = 0.1$と設定した事前 ESS（0.1×1000+1+1=102）よりも大きくなる傾向があった。傾向スコアに基づく power 事前分布の各層の事前 ESS の中央値は，固定配分およびベイズ配分ともに 21.1～22.6 であり，4 分位範囲は 21.0～22.6 の範囲内だった（図 4）。また，5 つの層を合わせた事前 ESS は，外部データ全体から借用したい名目上の対象者数に初期事前分布の事前 ESS を加えた値（100＋1×5+1×5=110）に近かった。

5. 考察とまとめ

本稿では，RCT でハイブリッド対照を用いる場面において，傾向スコアに基づく power 事前分布アプローチによる解析を行うための SAS の実装方法を示した。簡単のため外部データのソースが 1 つの場合を想定したが，外部データのソースが複数の場合にも適用可能である（Lu et al, 2022）。外部データのソースが複数の場合は，外部データのソースごとに power パラメータを設定する必要があるが，実装方法は本稿で説明した方法と大きく異ならないため，本稿の SAS コードを参考に実装が可能と考えられる。また，本稿ではアウトカムがベルヌーイ分布に従うと仮定する 2 値データの場合で説明したが，アウトカムが他の分布の場合でも傾向スコアに基づく power 事前分布は適用可能であり，正規分布を仮定した連続値データの場合の SAS での実装方法については本稿の付録 1 に補足した。Power パラメータを確率変数とする normalized power 事前分布およびベイズ配分での基準化定数の算出には，一般の場合は数値積分による算出が必要であり，MCMC の反復の中で数値積分を行うと，層ごとの基準化定数が必要なベイズ配分では特に計算負荷が大きくなる。そのため，計算負荷を低減するためのプログラムの工夫が必要である（武田, 2014）。しかし，本稿で示したように基準化定数を解析的に求めることができる場合は，計算負荷を大幅に低減できる。

傾向スコアに基づく power 事前分布の性能評価を行うため，外部データを用いない RCT を想定したベイズ流の解析，傾向スコアを用いない power 事前分布アプローチとともに，簡単なシミュレーションを行った。Conditional および normalized power 事前分布アプローチのように単純に外部データを解析に含めるアプローチではバイアスがあり，MSE は RCT データのみの解析よりも大きく，第 1 種の過誤確率のインフレ，被覆確率の低下といった問題が生じていた。一方で，個体レベルの共変量の情報を考慮し，外部データのうち RCT データと共変量の値が同じような対象者のみを解析に含める傾向スコアに基づく power 事前分布では，バイアスがなく，RCT データのみの解析よりも MSE は小さく，第 1 種の過誤確率のインフレおよび被覆確率の低下も認められず，検出力については RCT データのみの解析より高かった。これらの結果より，傾向スコア

に基づく power 事前分布は外部データを借用する際の共変量の分布の差異に起因するバイアスを抑え，かつ外部データの利用により推定精度を向上させる方法として有用であると考えられる。ただし，4 節のシミュレーション実験はアウトカムに影響する共変量がすべて特定され，正しく推定された傾向スコアに基づく結果であることに注意が必要である。未観測の共変量がある場合は，観測された共変量だけに基づく傾向スコアで調整してもバイアスが調整しきれない（野村ら, 2026）。この点について，4 節のシミュレーションで用いた共変量のうち$X3$を未測定の共変量とし, $X1$および$X2$だけで傾向スコアを求めた追加シミュレーション実験（反復回数 5000 回）を行った。その結果，固定配分およびベイズ配分ともに若干の正のバイアスが生じ，第 1 種の過誤確率のインフレ（固定配分：3.8%，ベイズ配分：3.5%）が生じていた（付録 2）。一般には共変量が完全に特定されているとは限らないため，計画段階でのシミュレーションによる未測定共変量の影響の程度や感度分析による結果の頑健性の確認が重要と考えられる。

本稿では，傾向スコアに基づく power 事前分布として固定配分とベイズ配分の 2 つのアプローチを検討した。上記の追加シミュレーション実験ではベイズ配分のバイアスが若干小さかったものの，2 つのアプローチの性能に大きな違いは認められなかった。Wang et al（2019）では単群試験に傾向スコアに基づく power 事前分布を用いる状況でシミュレーション実験を行い, 固定配分と比べてベイズ配分でバイアスが小さく, MSE は若干大きかったと考察している。

本稿のシミュレーション実験は，傾向スコアに基づく power 事前分布の特徴を示すために限定的なシナリオでのみ実施した。アウトカムや各データの例数，解析に利用可能な共変量などは適用する状況ごとに異なると考えられるため，実際に傾向スコアに基づく power 事前分布を適用する際には，試験の計画段階で想定している条件下で，ガイドライン（FDA, 2026）などを参考にシミュレーションによる試験の動作特性の評価を行う必要がある。また，実際の適用では統計解析上の問題だけでなく，外部データの信頼性担保なども考慮する必要がある（医薬品医療機器総合機構, 2025）。

本稿が SAS で傾向スコアに基づく power 事前分布を実装する際の参考になれば幸いである。

参考文献

Duan Y, Ye K, Smith EP (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics*, 17, 95-106.

European Medicines Agency (2026). Concept Paper for the Development of a Reflection Paper on the use of Bayesian methods in clinical development. https://www.ema.europa.eu/en/use-bayesian-methods-clinical-development-scientific-guideline [2026/9/16 アクセス]

Food and Drug Administration (2026). Use of Bayesian Methodology in Clinical Trials of Drug and Biological Products Guidance for Industry. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-bayesian-methodology-clinical-trials-drug-and-biological-products [2026/9/16 アクセス]

Lu N, Wang C, Chen WC, Li H, Song C, Tiwari R, Yue LQ (2022). Propensity score-integrated power prior approach for augmenting the control arm of a randomized controlled trial by incorporating multiple external data sources. *Journal of Biopharmaceutical Statistics*, 32(1), 158–169.

Rosenbaum PR, Rubin DB (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

Wang C, Li H, Chen WC, Lu N, Tiwari R, Xu Y, Yue LQ (2019). Propensity score-integrated power prior approach for

incorporating real-world evidence in single-arm clinical studies. *Journal of Biopharmaceutical Statistics*, 29(5), 731–748.

SAS Institute Inc. (2026). SAS/STAT User's Guide. SAS Institute Inc., Cary, NC, USA.

入江 薫, 菅澤 翔之助, 橋本 真太郎 (訳), ピーター・D・ホフ (著) (2022). 標準 ベイズ統計学. 朝倉書店.

医薬品医療機器総合機構 (2025). 外部対照試験に関する留意事項. https://www.pmda.go.jp/files/000274653.pdf [2026/9/16 アクセス]

魚住 龍史, 矢田 真城 (2020). SAS による傾向スコアの活用：COVID-19 患者に対する観察研究をひも解く. 日本 SAS ユーザー会フォーラム 2020. https://www.sas.com/content/dam/SAS/documents/event-collateral/2020/ja/japan-sas-users-group-forum/2-02-uozumi.pdf [2026/9/16 アクセス]

AMED 医薬品等規制調和・評価研究事業 (2023). 希少疾患領域の臨床試験におけるベイズ流アプローチの適用に関する基本指針. https://www.jpma.or.jp/information/evaluation/symposium/jtrngf0000001cap-att/DS_202302_hirakawa1.pdf [2026/9/16 アクセス]

武田 健太朗, 大庭 真梨, 柿爪 智行, 坂巻 顕太郎, 田栗 正隆, 森田 智視 (2015). 臨床試験におけるヒストリカルコントロールデータの利用. 計量生物学, 36(1), 25-50

武田 純 (2014). MCMC プロシジャを用いた Normalized Power Prior の実用的な実装. SAS ユーザー総会 2014. https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups14-a-06.pdf [2026/9/16 アクセス]

日本製薬工業協会 (2022). Rare disease の治療効果の推測法. https://www.jpma.or.jp/information/evaluation/results/allotment/jtrngf000000085m-att/ds_202212_rare.pdf [2026/9/16 アクセス]

日本製薬工業協会 (2024). Complex Innovative Trial Design の現状と事例の整理 - FDA ガイダンス邦訳付き. https://www.jpma.or.jp/information/evaluation/results/allotment/tcjmdm00000010pp-att/DS_202404_CID.pdf [2026/9/16 アクセス]

野村 尚吾, 大東 智洋, 澤本 涼 (2022). Hybrid control アプローチを用いるランダム化比較試験の計画と解析：外部データが要約統計量の場合. 計量生物学, 43(1), 63-96.

野村 尚吾, 澤田 航太, 大東 智洋, 澤本 涼 (2026). Hybrid control アプローチを用いるランダム化比較試験の計画と解析：外部データが個人レベルデータの場合. 計量生物学, 47(1), 1-30.

付録 1：アウトカムが正規分布に従うと仮定する連続値の場合

アウトカムが正規分布に従うと仮定する連続値の場合でも，本稿で紹介した SAS コードを修正することで傾向スコアに基づく power 事前分布を実装できる。本文の SAS コードでの層kの反応率のパラメータの定義と初期事前分布を，平均のパラメータに対応した設定に置き換える必要がある。なお，層kのアウトカムの分散σ_k^2は既知とみなし，実務的に層kの標本標準偏差$\hat{\sigma}_k$などを用いる（Wang et al, 2019）。SAS コードは以下のようになる。

```
array mu_t[5];
array mu_p[5];
parms mu_t: 0;
```

```
parms mu_p: 0;
prior mu_t: ~ general(0);
prior mu_p: ~ general(0);
```

ここでは平均のパラメータの初期事前分布としてフラット事前分布$\pi_0(\theta_k^{(0)}) \propto 1$を用いた。また，対数尤度の計算については logpdf 関数を正規分布に変更する。

```
loglik = logpdf('normal', Y, mu_t[strata], SIGMA);
```

SIGMA は各層の標本標準偏差を格納した変数である（事前に MEANS プロシジャなどで算出した値を解析データセットにマージしておく）。対照群の場合は，mu_t[strata]を mu_p[strata]に置き換える。ベイズ配分で用いる各層の基準化定数は，正規分布の場合でも解析的に算出することができる。初期事前分布にフラット事前分布を用いた場合は以下となる（Wang et al, 2019）。

$$\int \left[L(\theta_k^{(0)}|D_{k,EXT}^{(0)})\right]^{\alpha_k} d\theta_k^{(0)} = \left(\frac{1}{\sqrt{2\pi\sigma_k^2}}\right)^{\alpha_k N_{k,EXT}} \sqrt{\frac{2\pi\sigma_k^2}{\alpha_k N_{k,EXT}}} \exp\left[-\frac{\sum_{i=1}^{N_{k,EXT}}(y_i - \bar{Y}_{k,EXT})^2}{2\sigma_k^2/\alpha_k}\right]$$

したがって，本文の SAS コードでの基準化定数の対数変換値を引く部分を次のように修正する。Y_N は$N_{k,EXT}$，SS は$\sum_{i=1}^{N_{k,EXT}}(y_i - \bar{Y}_{k,EXT})^2$を格納した変数である（事前に算出してマージする）。

```
    loglik = loglik - (1-alpha[strata]*Y_N)*log(2*constant('pi')*SIGMA**2)/2
  + log(alpha[strata]*Y_N)/2 + SS/(2*SIGMA**2/alpha[strata])
```

付録 2：未測定の共変量がある場合の追加シミュレーション実験結果

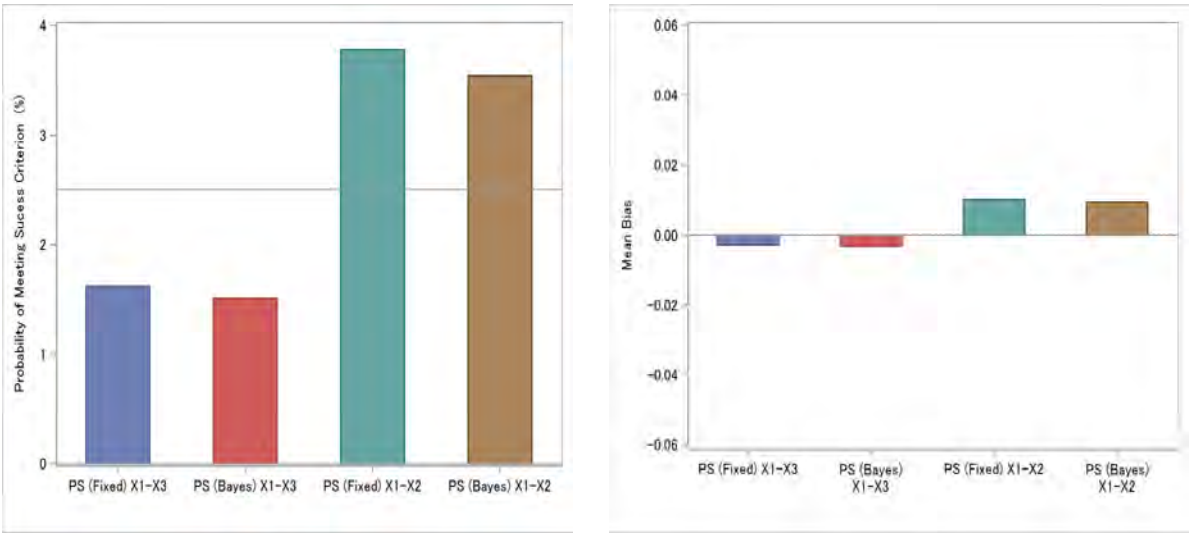


図 A1 すべての共変量を用いた場合（*X1-X3*）と*X3*を未測定とした場合（*X1-X2*）の第 1 種の過誤確率（左）およびバイアスの平均値（右）

In vivo QT 試験の検出感度向上に寄与する平滑スプライン曲線
― ICH-E14/S7B Q&A(2022)への対応 ―

高橋 行雄
BioStat 研究所(株)

Smoothing Spline Curves Contributing to Improved Detection Sensitivity in *in Vivo* QT Studies
― Addressing the ICH-E14/S7B Q&A (2022) ―

Yukio Takahashi
BioStat Research Co.,Ltd.

要旨: *In vivo* QT 試験は, Williams 型ラテン方格での実験が一般的であり, その結果は, 4×4 のノイズまみれの超多時点データとして得られる. 一定の時間区分で中央値を求めても揺らぎが取り切れず, さらに投与・給餌・点灯/消灯などの影響も受ける. 横断的パラダイムによるラテン方格 ANOVA を適用するために, 複数の時間区分データをさらにまとめたスーパーインターバルが使用されている. 薬物濃度 C-QT 解析のためには, 時間マッチの QT データが必要となり, 一定の時間区分データが使われている. スーパーインターバルに対するラテン方格 ANOVA から得られる RMSE(Root Mean Square Error)の分布に比べ, 一定の時間区分での RMSE の分布は, 右に大きく裾を引き, 分析感度が不安定となる. この超多時点データに対し平滑スプライン曲線をあてはめ, 時点ごとに ANOVA を適用した場合の RMSE は, スーパーインターバルによる RMSE と同等の分布が得られた. C-QT 解析に際し, 一定の時間区分データではなく平滑スプラインデータを使うことにより, より精密な C-QT 解析が行える.

キーワード: ICH-E14/S7B, テレメトリーQT 試験, 超多時点データ, ラテン方格, 平滑スプライン曲線

1. はじめに

In vivo QT 試験は, Williams 型ラテン方格での実験が一般的であり, その結果は, 4×4 のノイズまみれの超多時点データとして得られる. 一定の時間区分で中央値を求めても揺らぎが取り切れず, さらに投与・給餌・点灯/消灯などの影響も受ける. Rossman ら(2023)により, 標準とすべき統計解析法, 解析結果のまとめ方法が, 新たに提言された. 高橋(2025)は, イヌを用いた *in vivo* QT 試験から得られた HR, QT 間隔, RR 間隔で補正した QTci 間隔データを用いて, Rossman 論文に準じた解析を行い, その提言の有用性を確認した.

In vivo QT 試験では, テレメトリーシステムにより薬物投与後のバイタルサイン, RR 間隔, QT 間隔などの測定に特化したフェイズを先行し, 薬物濃度測定に特化したフェイズを分けて実施している. これは, 人間とは異なりイヌなどでは, 採血時の心電図に無視できない乱れが生じてしまうため PK/PD(C-QT)解析のためのデータを得るための対処法である.

In vivo QT 試験では, 無拘束の非げっ歯類を対象にしており, テレメトリーシステムで得られるバイタルサイン, 心電図解析で得られる RR 間隔, QT 間隔などが, 安定することなく, ノイズまみれの超多時点データとして観測される. RR 間隔で補正した QT 間隔(QTc 間隔)を用いて用量反応関係を浮き彫りにするた

000172

めに，スーパーインターバルと称する 60 分以上の区間での QTca 間隔の平均値などが用いられている．このことが，PK/PD（C-QT）解析のために採血時間にマッチした QTca 間隔を求め難くしている．

テレメトリーシステムから標準的に得られる RR 間隔で補正した QTca 間隔は，連続した 1 分間の QTca 間隔の平均値であり，サルの場合ならば，1 分間に 100 回程度の心拍数なので，100 回の平均値となっている．十分な数なので，安定していると思いたいのであるが，常に揺れ動き安定しない．さらに，採血時刻の前後 15 分間に広げた 30 分の平均値でも，滑らかな QTca 間隔が得られない．PK/PD（C-QT）解析の精度を上げるために，前後 30 分の 60 分の平均値の必要性も示唆されている．このように，安定した QTca 間隔を得るために投与 15 分後，30 分後の採血時刻マッチした安定した QTca 間隔を得ようとしても，時間幅を広げることができない．

そこで，ノイズまみれの超多時点データに対し，何らかの平滑化の方法に，採血時間とマッチし安定した QTca 間隔を取り出し，PK/PD（C-QT）解析の精度を上げたい．SAS では，一般化加法モデル GAM プロシジャにより各種の平滑化が行えることは，田中ら（2009），田中ら（2013），古川（2023），古川（2024）により詳細に示されているが，単一の時系列データを対象にしており，PK/PD（C-QT）解析で必要となる多数の時系列データに対し一括した平滑化については，言及されていない．

SAS の姉妹版の JMP には，GAM プロシジャに相当する機能はないが，時系列データに対する平滑化がいくつかのプラットホーム（プロシジャに相当）で平滑スプライン曲線のあてはめなどが可能であり，*in vivo* QT 試験の時系列データに対して活用できるか検討を行った．

2. *In vivo* QT 試験の超多時点データ

In vivo QT 試験の共同研究グループ（J-ICET : Japan activity for Improvement of Cardiovascular Evaluation by Telemetry system）で得られたデータの一部を用いる．J-ICET については，小松（2016）に詳しく示されている．ここでは，Moxifloxacin 投与の TestSite＝C のデータの結果を用い，超多時点データの特徴を示す．図 1 に示すのは，4×4 Williams 型ラテン方格実験の Vehicle 群（0 mg/kg）について 4 頭について 1 分区分で得られた QTca 間隔の 26 時間連続の経時データである．

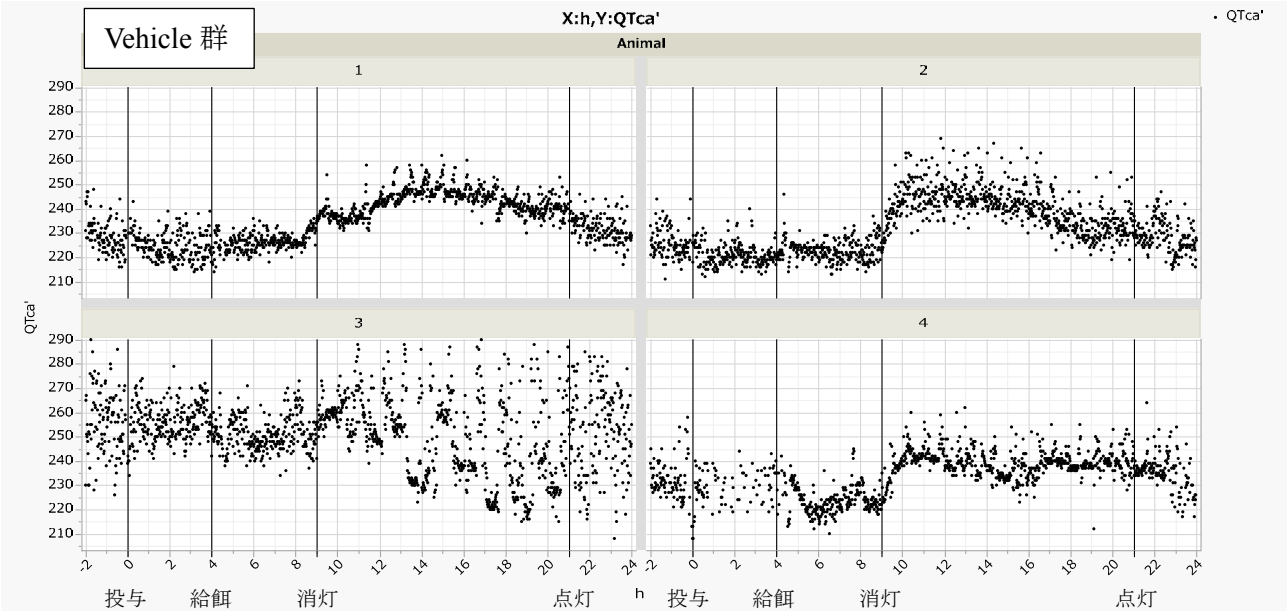


図 1．TestSite=C：Vehicle 群：サル 4 頭：1 分区分の 26 時間にわたる QTca 間隔の経時変化

サルの心拍数は，100 回程度なので，1 分区分の QTca 間隔は，100 回分の平均値となっている．さらに，60 分×26 時間＝1560 分の経時データとなっているが，上下に常に変動を繰り返し，安定している時間帯が全くない．ここに示したのは，4 頭の Vehicle 群であるが，Moxifloxacin の 25 mg/kg，50 mg/kg，および，100 mg/kg 投与による QTca 間隔の延長をラテン方格に対する ANOVA を適用して評価したい．しかしながら，このようなノイズまみれの 1 分区分のデータをそのまま使うことは絶望的である．

3. JMP による平滑スプライン曲線

JMP では，「二変量の関係」および「グラフビルダー」で平滑スプライン曲線をあてはめることができる．「モデルのあてはめ」では，モデル効果の属性の設定に「B-スプライン効果」があるが，超多時点データには不向きである．操作性に優れているのは「グラフビルダー」であり，デフォルトで平滑スプライン曲線のあてはめが行われる．図 1 は，デフォルト描かれる平滑スプライン曲線を意図的に消去した結果である．図 1 にデフォルトの平滑スプライン曲線を加えると図 2 の結果から Animal ごとのサーカディアンリズム（日内変動）が浮き彫りになる．Animal=3 は，点灯時に QTca 間隔が長く，消灯時に短くなり，他の 3 頭とは異なる昼夜逆転のサーカディアンりズムが示されている．

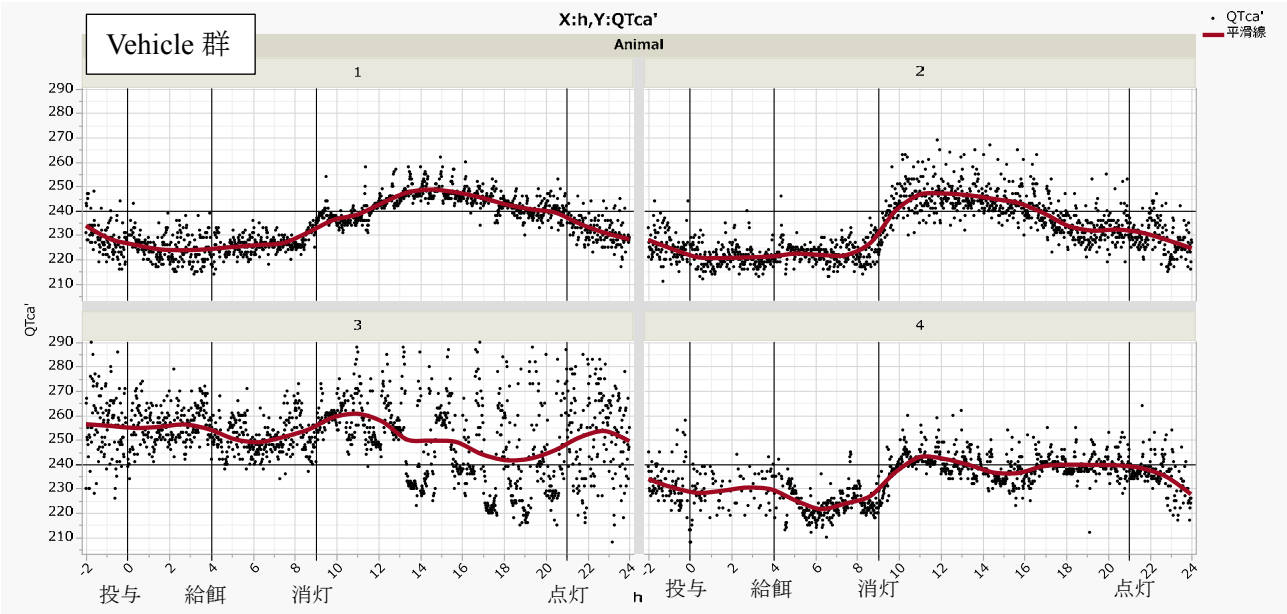


図 2. Vehicle 群：サル 4 頭に対するデフォルトのスプライン曲線のあてはめ

「グラフビルダー」で描かれたデフォルトの平滑スプライン曲線の 4 頭の Vehicle 群へのあてはめは，どのような基準で平滑パラメータ λ を設定したのか簡潔に説明することができないが，上々の結果と思われる．「グラフビルダー」では，平滑パラメータ λ をスライダーで自由に動かすことができるようになっているので，平滑スプライン曲線の形状の変化を連続的に目視できる．平滑パラメータ λ を大きくすると回帰直線となり，小さくすると短時間的な変化を追従する曲線となる．実際にスライダーでの設定を左端にした結果，右端に設定した結果を図 3 に示す．このように，平滑パラメータ λ を変化させることにより，平滑スプライン曲線は劇的に変化する．どのような平滑スプライン曲線が望ましいか，あらかじめ基準を定める必要がある．たとえば，Vehicle 群の場合であれば，サーカディアンリズムを考慮したコサイン曲線らしい滑らかな曲線となるように平滑パラメータ λ を設定するといった方法である．

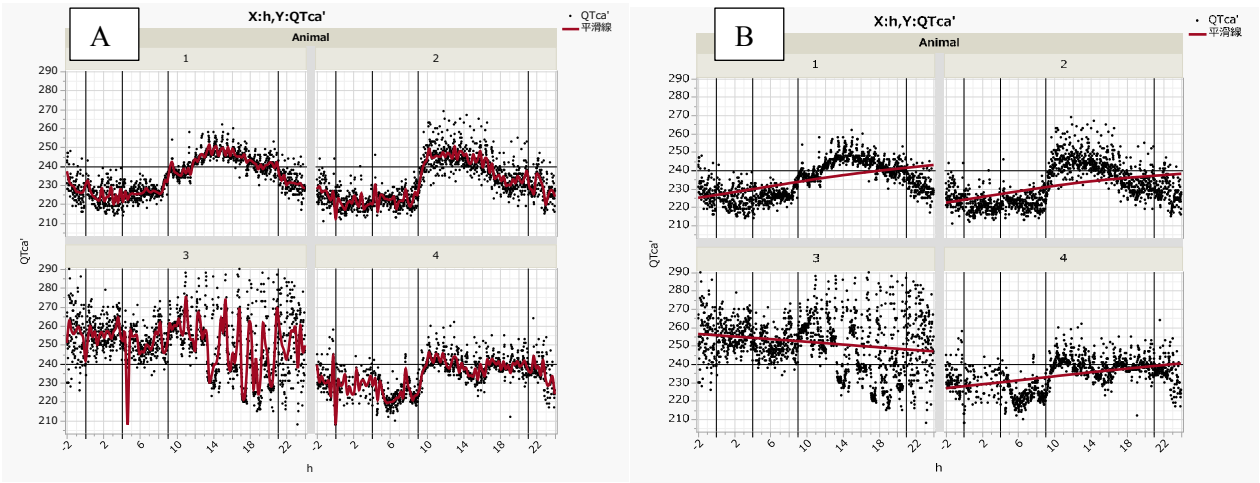


図 3. Vehicle 群：平滑パラメータ λ 変化させた平滑スプライン曲線（A：λ 最小化，B：λ 最大化）

Vehicle 群に対しサーカディアンリズムを考慮したコサイン曲線らしい滑らかな曲線が望ましいとの判断は，投与後に QT 延長がある代表的な Moxifloxacin 投与後の経時データへ適用することは難しい．ベースとなる Vehicle 群のサーカディアンリズムに Moxifloxacin 投与後の QTca 間隔の増加が上乗せされると想定される．

図 4 に示すのは．Moxifloxacin 100 mg/kg 投与後の平滑スプライン曲線のあてはめである．投与後に QTca 間隔が上昇している．その影響を受けて投与前 -2 h から 0 h の QTca 間隔が上昇傾向となり，図 2 の Vehicle 群と異なる平滑スプライン曲線となっている．

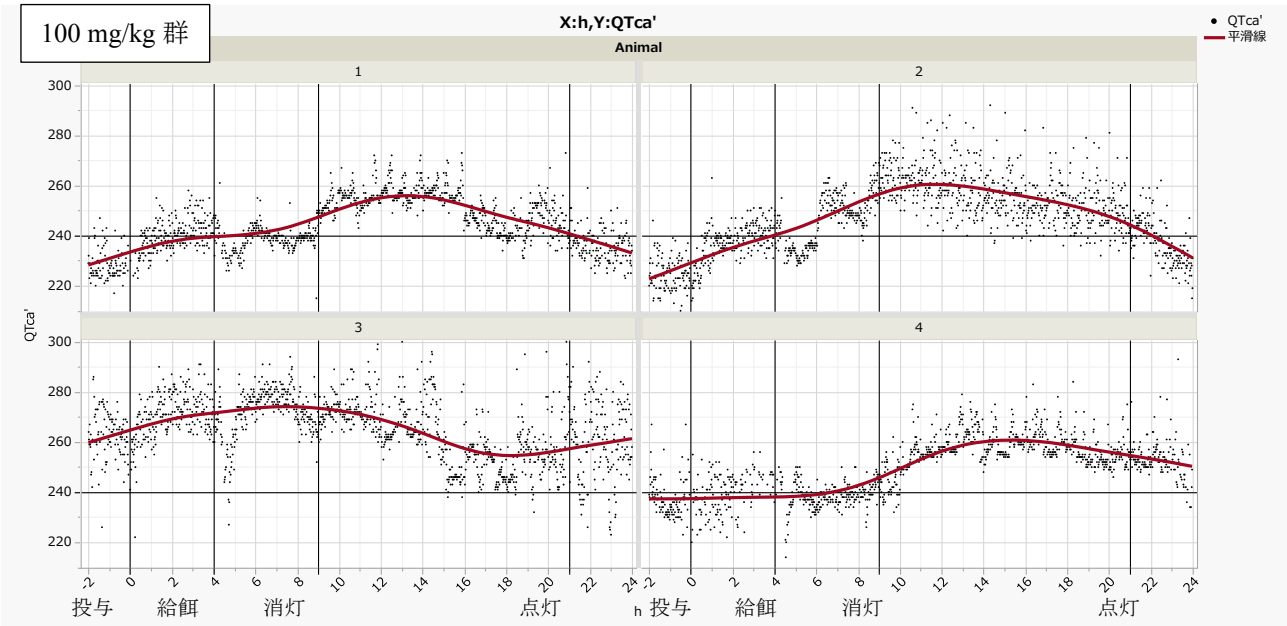
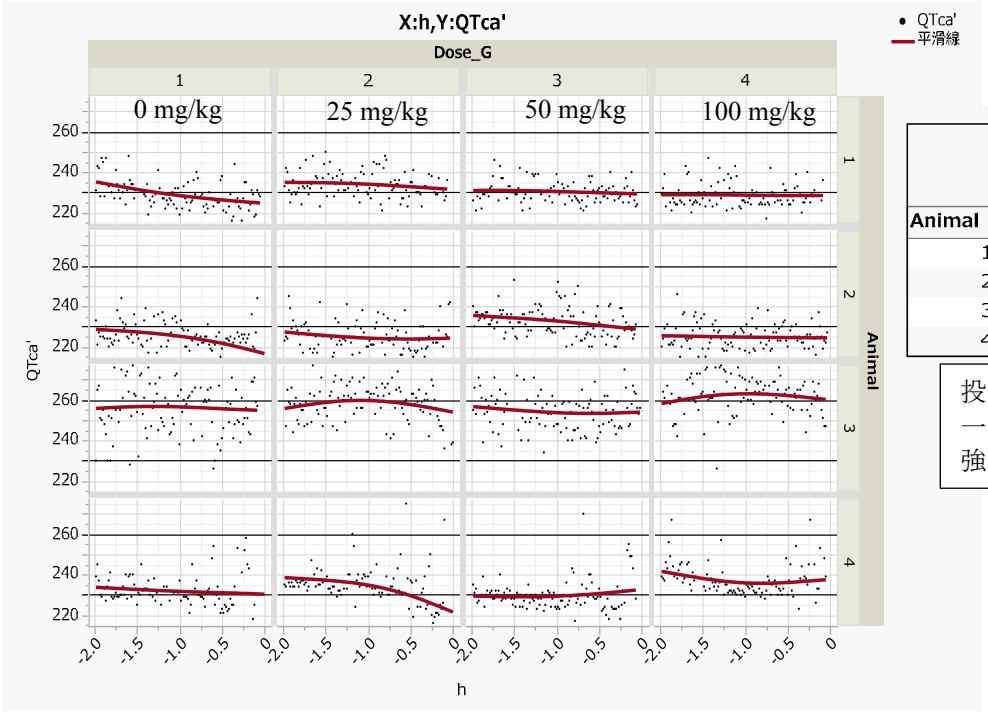


図 4. Moxifloxacin 100 mg/kg 群の QTca 間隔に対する平滑スプライン曲線のあてはめ

クロスオーバー法によるラテン方格実験では，同一の Animal に対し，Vehicle，Moxifloxacin の 3 用量が投与され，投与前 2 h から 0 h までの QTca 間隔が同程度であることを前提にしている．図 5 に示すように，投与前の QTca 間隔は，Animal により差異があるが，用量 Dose_G による差異は，ほとんどなく，クロスオーバー法で期待される同一 Animal 内で安定していることが示されている．図 2 の Vehicle 群の投与前

2 h の平滑スプライン曲線と図 5 の投与前値のみの平滑スプライン曲線は，同様の挙動と読み取れるが，図 4 の 100 mg/kg 群の投与前は，上昇傾向であり，図 5 の平行的な推移と場合と異なる．



	QTca' 平均 Dose_G			
Animal	1	2	3	4
1	228.9	233.7	230.3	228.6
2	224.4	224.7	232.4	224.8
3	256.1	258.0	254.2	261.6
4	231.9	233.3	229.7	237.2

投与前の QTca には，同一 Animal 内で 5.0 msec 強の揺らぎがある．

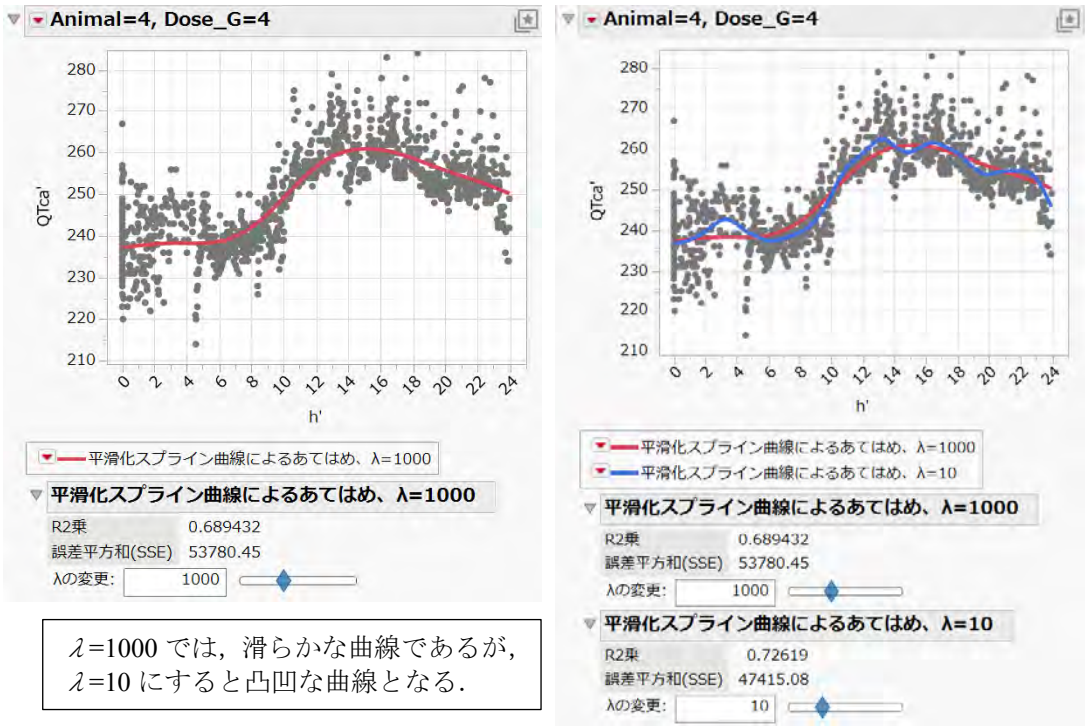
図 5. 投与前 2 時間の QTca 間隔に対する平滑スプライン曲線のあてはめ

このように，投与前 2 h の QTca 間隔は，ほぼ平行に推移することから，すべてを含めて投与前 0 h とみなすことにする．このように，0 h のデータ数を 2h×60 分＝120 個とすることにより，平滑スプライン曲線のあてはめる際に Moxifloxacin 投与後の QTca 間隔の上昇の影響を打ち消し，0 h のデータを安定化させることができる．

4. 平滑パラメータ λ

複数の経時データに対し「グラフビルダー」では，平滑パラメータ λ をスライダーで動かすことにより，複数の平滑スプライン曲線の形状を連続的に変化させることができるが，平滑パラメータ λ を明示的に設定することができない．「二変量の関係」では，平滑パラメータ λ をスライダーで自由に動かすこともできるが，明示的に平滑パラメータ λ を設定することもできる．

図 6 左は，Animal=4，Dose_G=4 の場合について，平滑パラメータ λ=1000 と明示的に設定した場合の平滑スプライン曲線である．図 6 右は，平滑パラメータ λ=1000 が設定されている状態で，平滑パラメータ λ=10 の場合を追加した結果が重ね書きされている．ただし，どの程度の平滑パラメータ λ が適切なのかを統計的に判断するための統計量が存在しない．統計量として示されている誤差平方和（SSE）は，λ が小さい場合に小さくなり，平滑化した場合には，大きくなるので統計量として使えない．平滑スプライン曲線の 0 h の平均値が変わらない範囲で，かつ，凸凹が消える最小の λ で，切のいいのが λ=1000 であった．このように，平滑パラメータ λ を明示的に設定し，それらを重ね書きすることにより，平滑パラメータ λ が平滑スプライン曲線の形状に与える影響を可視化し，自らの責任で平滑パラメータ λ を設定する必要がある．



λ=1000 では，滑らかな曲線であるが，
λ=10 にすると凸凹な曲線となる．

図 6.「二変量の関係」による複数の平滑スプライン曲線のあてはめ

「二変量の関係」は，多機能であり，by変数 で Animal および Dose_G を設定し，レイアウト オプションの 複数行に配置 を使い 4×4 の「二変量の関係」をマトリックス状に表示し，それらに共通の平滑パラメータλの同時あてはめが可能となる．基本は，マトリックス状の 4×4 の「二変量の関係」に対し，個別に平滑スプライン曲線を描くのであるが，Ctrl キーを押しながら平滑パラメータλを設定すると 4×4＝16 枚の「二変量の関係」に対し一括して平滑スプライン曲線のあてはめることができる．

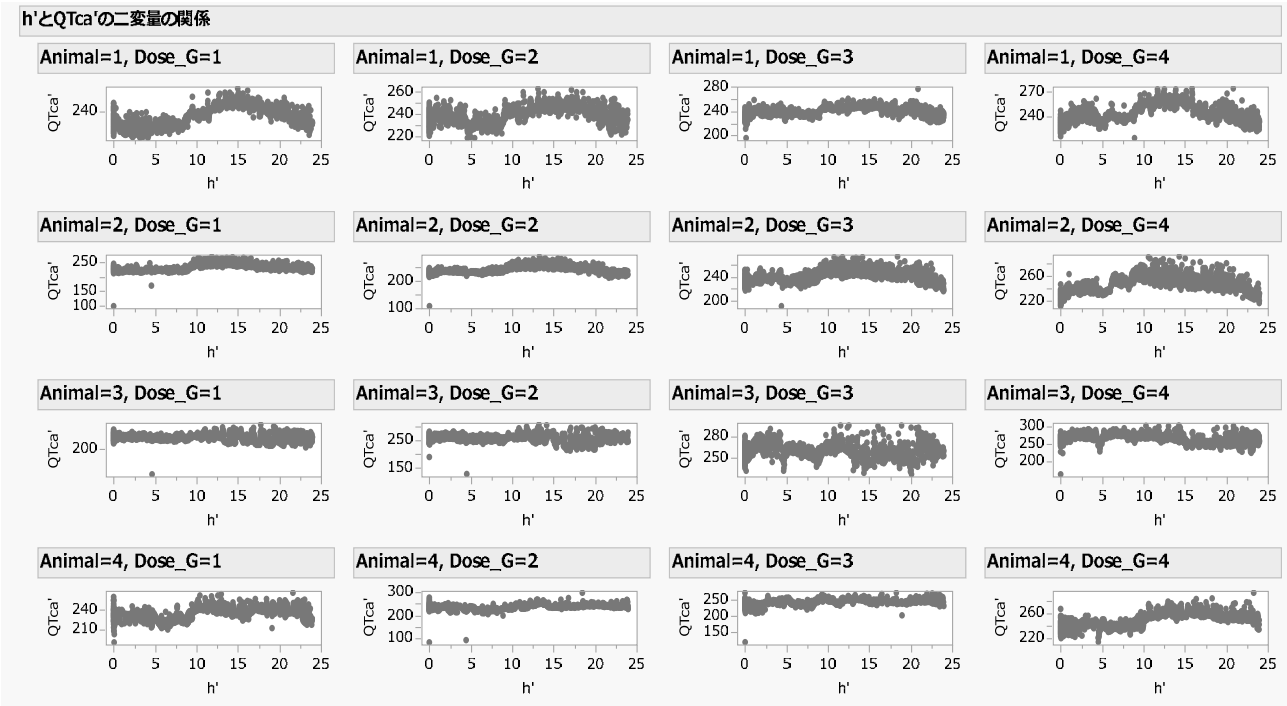


図 7. 平滑スプライン曲線あてはめ前の「二変量の関係」による 4×4 の散布図行列

あてはめた平滑スプライン曲線の予測値を JMP データテーブルに保存したい．Ctrl キーを押しながら ▼—— 平滑スプライン曲線 のプルダウンメニューから 予測値の保存 を選択すると，全ての Animal の Dose_G に対し，予測値が JMP データテーブルに出力される．薬物濃度の測定時間とマッチを念頭に（0, 0.25, 0.5, 1, 2, 4, 8, 12, 16, 24）時間の予測値を抜き出し，コンパクトな JMP データテーブルとする．

C-QT 解析では，Vehicle 群のサーカディアンリズムによる変動を取り除くために，各用量群の QTca 間隔から時間マッチの Vehicle 群の QTca 間隔を差し引く必要がある．その際に，投与前値 0 h から各時点の差 Δpre を求め，さらに時間マッチの Vehicle 群の Δpre を差し引く ΔΔVehicle が好まれている．統計的な観点から，クロスオーバーによるラテン方格実験では，同じ Animal 内の用量群間の投与前値は，同程度で違いが統計的には検出できない．そのために Δpre を差し引かない Δvehicle が望ましいと思えるが，どちらが望ましいのか一律に判断することができない．

図 8 左は，平滑パラメータ λ=1000 で得られた平滑スプライン推定値の Animal 別の用量群についての重ね書きプロットである．図 8 右は，投与前値 0 h から各時点の差 Δpre を求めた場合の重ね書きプロットである．図 8 右の Δpre は，投与前 0 h がすべてゼロ点を通り，望ましいようにも思われるが，用量反応関係が逆転している時点が多数発生している．

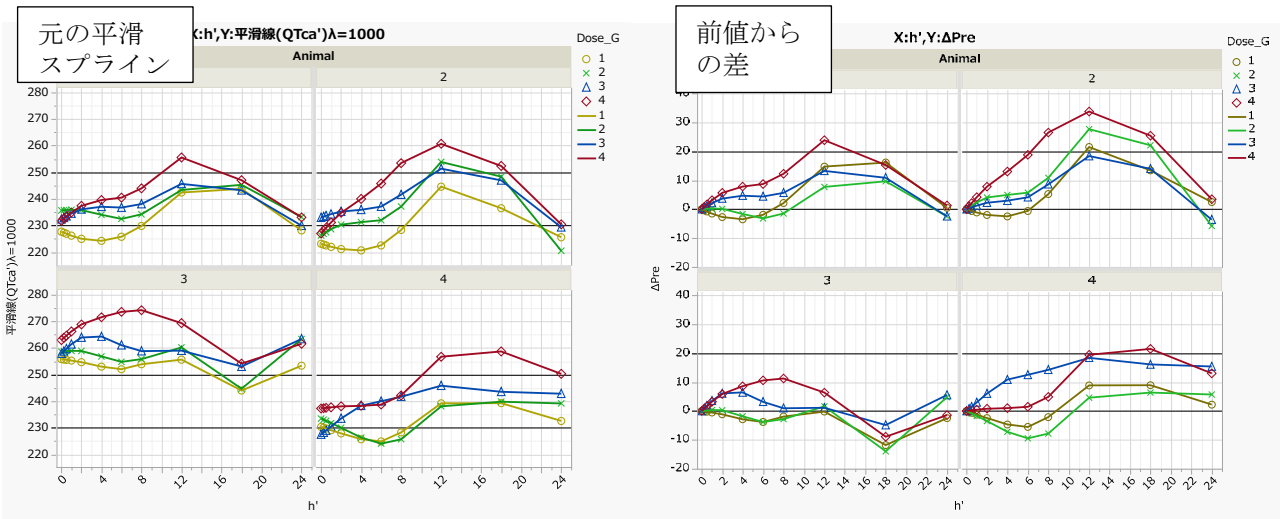


図 8. 平滑パラメータ λ=1000 とした QTca 間隔に対する平滑スプラインの推定値

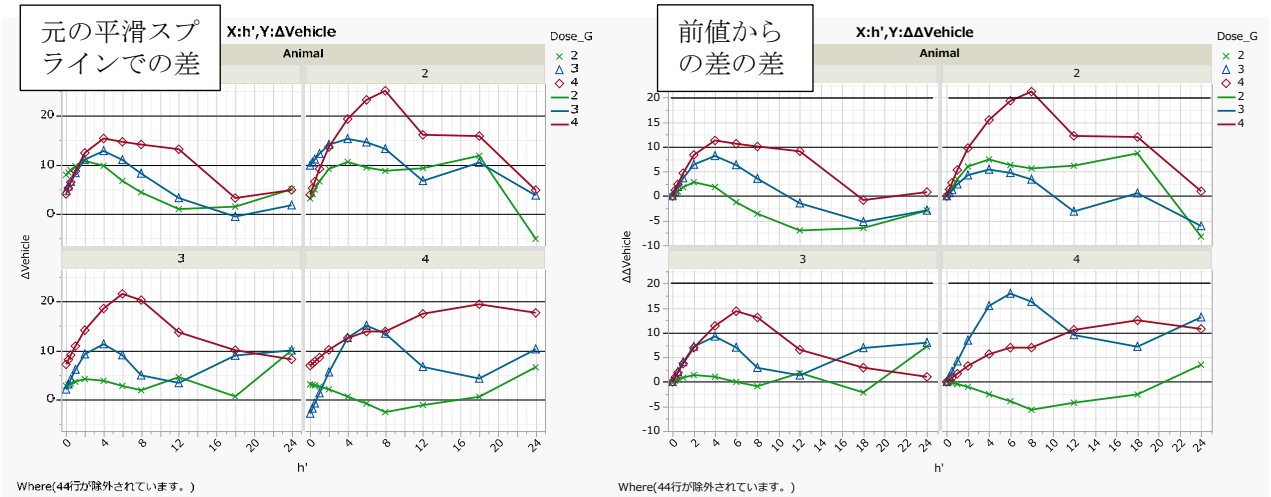


図 9. Vehicle 群との時間マッチの差 QTca 間隔に対する平滑スプライン推定値

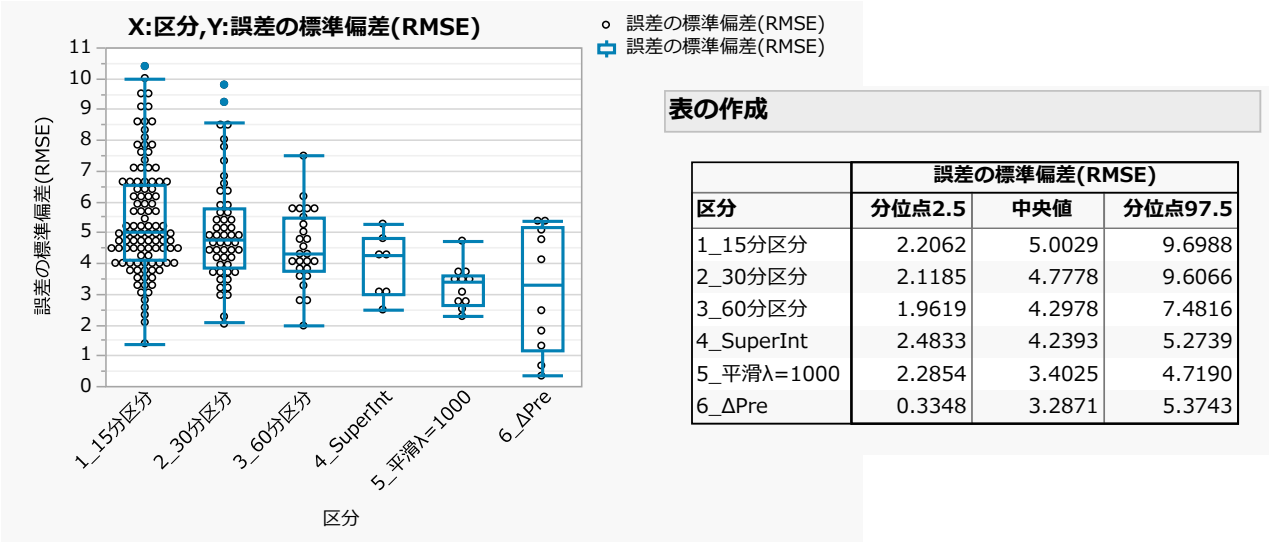
図 9 に示すのは，Vehicle 群の時間マッチした差し引いた QTca 間隔である．図 9 左は，元の平滑スプライン推定値での Vehicle 群からの差 Δvehicle である，図 9 右は，投与前からの差 Δpre の平滑スプライン推定値に対し，Vehicle 群からの差 $\Delta\Delta$vehicle である．Δvehicle の場合は，用量反応関係が比較的保たれ，反応が負になることは少ないが，$\Delta\Delta$vehicle の場合は，用量反応関係が Animal=2 で低用量と中用量で逆転し，Animal=4 で中用量と高用量で逆転するだけでなく，反応が負となる場合が多数発生している．

このような用量反応関係の逆転が多発するのは，測定された Animal ごとの投与前値が，用量群ごとに個体内変動の結果としてランダムに揺れ動き，その投与前値を差し引くとにより，その後の用量反応関係に対し，乱れを生じさせた結果と推測される．

5. スーパーインターバルの RMSE との比較

ノイズまみれの超多時点データの QTca 間隔に対し，30 分区分の中央値を算出し，さらに複数の区分の平均を取ったスーパーインターバルとすることにより，横断的パラダイムによるラテン方格 ANOVA での RMSE が相対的に小さく安定することにより分析感度が向上する．C-QT 解析に際し，採血時間にマッチした QTca 間隔を必要とするのであるが，採血時間（0.25 h，0.5 h，1 h）に対し，（15 分区分，30 分区分，60 分区分）などを用いたくなる．ただし，図 10 に示すように，それらの区分データでの QTca 間隔はスーパーインターバルに比べ不安定であり，したがって，C-QT 解析の結果も不安定になる．

平滑スプライン曲線から時間マッチで得られた QTca 間隔をスーパーインターバルで求めた QTca 間隔のラテン方格 ANOVA を適用した際の RMSE と比較検討する．図 8 に示した時間ごとの QTca 間隔に対しラテン方格 ANOVA を適用し RMSE を求め，図 10 に示すように（15，30，60）分区分の RMSE も含めてボックスプロットで示す．平滑パラメータ λ =1000 での平滑スプライン曲線から時間マッチの Qtca 間隔の RMSE の中央値は，3.4025 とスーパーインターバルの 4.2393 よりも小さく，また 95%信頼区間も小さめであり，スーパーインターバルに対し勝るとも劣らないことが示されている．投与前値を差し引いた ΔPre は，RMSE の中央値が 3.2871 と同程度であるが，時点により RMSE が大きめになったり，小さ目になったり，不安定な RMSE の分布となっている．



表の作成

	誤差の標準偏差(RMSE)		
区分	分位点2.5	中央値	分位点97.5
1_15分区分	2.2062	5.0029	9.6988
2_30分区分	2.1185	4.7778	9.6066
3_60分区分	1.9619	4.2978	7.4816
4_SuperInt	2.4833	4.2393	5.2739
5_平滑λ=1000	2.2854	3.4025	4.7190
6_ΔPre	0.3348	3.2871	5.3743

図 10.　各種の区分および平滑スプライン QTca 間隔に RMSE の比較

C-QT 解析に用いる QTca 間隔は，図 9 に示すように Vehicle 群との時間マッチの差にする必要がある．そこで，投与前値を差し引かない⊿Vehicle と差し引いた⊿⊿Vehicle の Vehicle 群を除いた 3 用量について欠測値があるラテン方格 ANOVA を適用し，RMSE の比較を行う．

表 1 に示すように，投与前値を差し引いた⊿⊿Vehicle は，0.25 h から 1 h までは，⊿Vehicle に比べ RMSE がかなり小さくなっている．投与前の QTca 間隔については，図 5 に示したように，Animal 間では明らかな個体間差があるが，同じ Animal 内では，誤差的な変動となっている．ただし，投与後 1 h ぐらいまでの QTca 間隔の増加量⊿⊿Vehicle とした場合に，投与前の誤差的変動の影響を強く受けることにより，用量反応関係が⊿Vehicle に比らべ RMSE が小さくなっている．

表1　RMSEの比較

時間 h	⊿Vehicle	⊿⊿Vehicle
0	5.1310	
0.25	4.7776	0.3854
0.5	4.4315	0.7696
1	3.7766	1.5290
2	2.7622	2.9677
4	2.6875	5.1917
6	3.4602	6.2014
8	3.6107	6.5790
12	3.2069	6.5897
18	3.0723	2.3909
24	4.0738	3.1785
分位点2.5%	2.6875	0.3854
中央値	3.6108	3.0731
分位点97.5%	5.1310	6.5897

投与 2 時間目以後は，⊿Vehicle は，⊿⊿Vehicle に比べ，RMSE が一貫して小いか同等レベルであり，C-QT 解析には，⊿Vehicle 方が望ましいとも言えるのであるが，投与後 1～2 時間までは，⊿⊿Vehicle が望ましいとの結果である．このようなことからも，⊿Vehicle と⊿⊿Vehicle のどちらが望ましいか判断が付かない．

6. 薬物濃度 C-QT 解析

C-QT 解析では，X 軸に薬物濃度，Y 軸に QTca 間隔とする散布図の作成が基本であり，薬物濃度と薬理反応（QTca）の関係が，直線のあてはめで収まるか，あるいは，反応が頭打ちとなる場合には Emax 曲線のあてはめにするかは，得られたデータに依存する．Moxifloxacin 投与後の薬物濃度と薬理反応（QTca）の関係を論ずる際に⊿Vehicle か⊿⊿Vehicle か，どちらを用いたら良いのだろうか，悩ましい問題である．

図 11 左は，⊿Vehicle と⊿⊿Vehicle 別々に Animal 別の回帰直線をあてはめた場合で，⊿Vehicle では，薬物濃度が 0 の場合に切片がプラス側に振れすぎているので，⊿⊿Vehicle が望ましいと思えるが，高濃度域での頭打ち目立ち，回帰直線が低めに推定されてしまう．

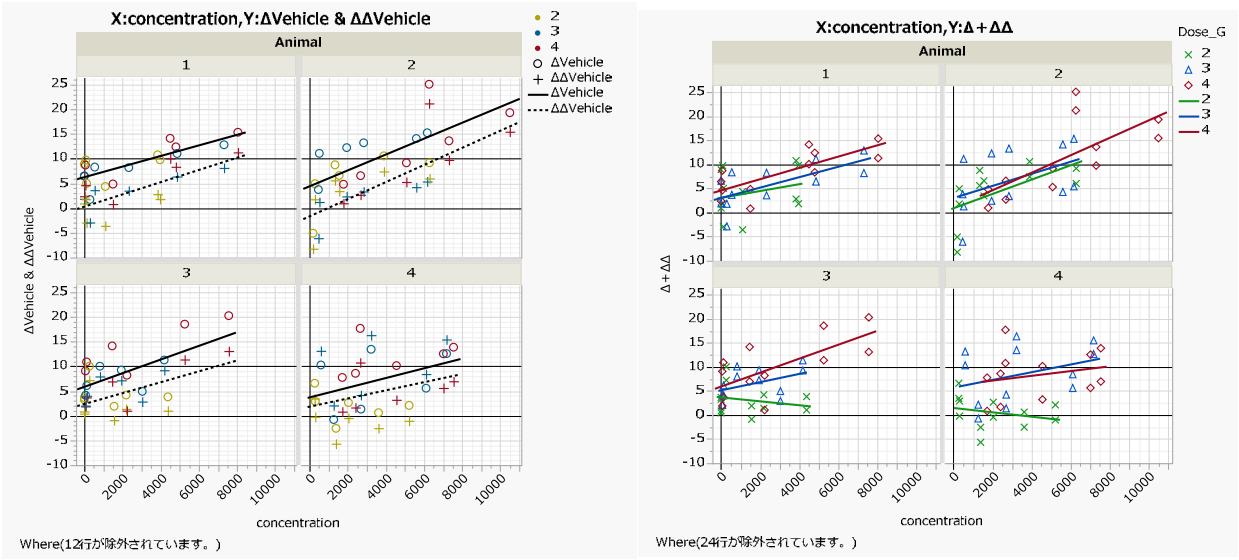


図 11.　⊿Vehicle と⊿⊿Vehicle 別々（左）および⊿Vehicle＋⊿⊿Vehicle 合算（右）の回帰直線

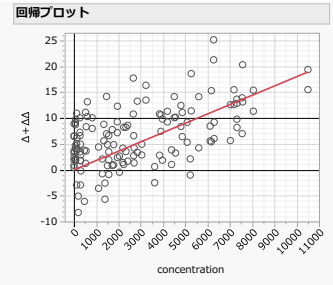
図 11 右は，⊿Vehicle および⊿⊿Vehicle 全体に対し Animal ごとに用量群 Dose_G それぞれに回帰直線をあてはめた場合である．Animal=(3, 4)の低用量×印に負の用量反応関係がみられるものの，全体としては，重なり合うきれいな回帰直線となっている．

7. 考察

In vivo QT 試験おける薬物濃度 C-QT 解析では，ノイズまみれの超多時点 QTca 間隔データに対し，薬物濃度の測定時点にマッチした QTca 間隔データを得るために，1 分区分の元データから 30 分区分の中央値を用いてきた．ただし，ノイズが取り切れないために 60 分区分に広げたいとの思いもあるが，0.25 h あるいは 0.5 h の血中濃度測定にマッチする QTca 間隔には，適用しがたい．他の時間について区分幅を広げたスーパーインターバルの活用も考えられるが，ややこしくなるだけであり避けたい．そこで，平滑スプライン曲線をあてはめることにより，ノイズまみれの超多時点データからスムーズに平滑化された時間マッチの QTca 間隔を得ることができ，精度の高い C-QT 解析が行えると期待される．

平滑スプライン曲線は，一つの平滑パラメータλによって規定されるが，あてはまりの良さを規定する統計量が存在しないために，目視によるあてはまり具合を判断し，平滑パラメータλを設定する必要がある．そのために，Vehicle 群に着目し，滑らかなサーカディアンリズムとなるような平滑スプライン曲線となる平滑パラメータλを求め，設定されたλを用いて他の用量群全てに対し平滑スプライン曲線のあてはめ，一意な平滑スプライン曲線を得ることができる．

得られた平滑スプライン曲線から，薬物濃度の測定時間にマッチした QTca 間隔の予測値を取り出し，ラテン方格 ANOVA を適用した際の RMSE は，スーパーインターバルの場合に勝るとも劣らないことが確認され，精度の高い C-QT 解析が行えると期待される．



投与前値を差し引き，さらに Vehicle 群との時間マッチの⊿⊿Vehicle が標準的に使われているが，投与前の個体内変動に関連して，低めになったり，高めになったり，用量反応関係が乱れることもあり，投与前値を差し引かない⊿Vehicle の活用も視野にいれた解析方法を例示した．また，切片を持つ回帰直線のあてはめではなく，原点を通る回帰直線が望ましいとも考えられるが，更なる検討が必要である．

文献

Rossman E.I., Wisialowski T.A., VargasH.M., et al.(2023),Best practice considerations for nonclinical in vivo cardiovascular telemetry studies in non-rodent species: Delivering high quality QTc data to support ICH E14/S7B Q&As, Journal of Pharmacological and Toxicological Methods 123 (2023) 107270.

高橋行雄(2025)，超多時点のラテン方格 QT 試験データに関する統計解析の新たな方向性，－ICH-E14/S7B Q&A(2022)への対応－，SAS ユーザー総会 2025 論文集，277-286.

田中祐輔，伊庭克拓，辻谷将明(2009)，GAM とその周辺，SAS ユーザー総会 2009 論文集，183-193.

田中祐輔，辻谷将明(2013)，一般化加法モデル GAM によるイベントヒストリー解析，SAS ユーザー総会 2013 論文集，203-212.

古川敏仁(2023)，ビッグデータに対する GAM の活用の仕方，SAS ユーザー総会 2023 論文集，76－86.

古川敏仁(2023)，アウトカムを説明するモデル(非直線的な関係も含むモデル)への特定変数の調整後の寄与や直線性を可視化する，―SAS GAM を用いて－，SAS ユーザー総会 2024 論文集，6－28.

小松竜一(2016)，実データに基づいたサル QT 延長試験結果の臨床との比較，https://biostat.jp/archive_teireikai_2_download.php?id=138.

SASプログラムの類似度評価

～Token化と5-gramによる構造比較～

井上 貴博

（協和キリン株式会社）

A Token-Based Similarity Metric for SAS Programs

Using 5-gram Features and Jaccard Similarity for Structural Comparison

Takahiro Inoue

Biometrics Department, Development Division, Kyowa Kirin Co., Ltd.

要旨

臨床試験の統計解析業務では、解析結果の信頼性を確保するため、Production Program とは独立した Validation Program(QC Program)を作成するダブルプログラミングが広く実施されている。しかしながら、QC Program の独立性を客観的に評価する指標はほとんど存在しない。

本稿では、SAS プログラムを Token 化し、5-gram および Rolling Hash を用いて構造的特徴を抽出し、Jaccard 係数によりプログラム間の類似度を定量化する手法を紹介する。コメントや書式、データセット名などの表面的な差異の影響を低減し、プログラム構造に着目した類似度評価を実現した。

本手法は QC Program の妥当性や解析結果の正しさを保証するものではないが、Production Program と QC Program の独立性を評価するためのサロゲート指標として活用できる可能性がある。実装方法と評価事例を紹介するとともに、類似度評価の有用性と限界について考察する。

キーワード：SAS Program Similarity, Double Programming, Independent Programming, Tokenization, 5-gram, Rolling Hash, Jaccard Similarity

1　はじめに

臨床試験の統計解析業務では、解析結果の品質を保証するため、ダブルプログラミングが広く利用されている。一般的には、異なる担当者が同一仕様書に基づいて Production Program と Validation Program(QC Program)を独立に作成し、それぞれの成果物を比較することで解析結果の品質を保証する。

ダブルプログラミングの本質的な価値は、異なる視点から同一仕様を実装することにある。独立した実装により、プログラムミスや仕様解釈の誤りを発見しやすくなる。一方で、Production Program を参照して作成された QC Program では、同じ誤りを再現してしまう可能性があり、品質保証機能が低下する。したがって、成果物の一致のみならず、実装過程の独立性も重要である。

しかしながら、実際の業務では過去プログラムの流用、Production Program の参照、納期やリソースの制約などにより、独立性が十分に確保されない可能性がある。これらは個々の担当者の意識だけで解決できるものではなく、業務量、組織文化、開発プロセス、レビュー方針などの影響を受ける。

本稿では、ダブルプログラミングにおける独立性を直接評価するのではなく、Production Program と QC Program の構造的類似性を定量化することで、独立性に関する客観的な議論を支援することを目的とする。

2　ダブルプログラミングにおける独立性の課題

2.1　独立性が重要となる理由

ダブルプログラミングでは、Production Program と QC Program が独立した発想で作成されることにより、同一仕様に対して異なる実装経路から結果を検証できる。これは単なる二重作業ではなく、仕様解釈、データハンドリング、例外処理、集計ロジックなどにおける誤りを検出するための重要な品質保証活動である。

一方で、QC Program が Production Program のコード構造に強く依存している場合、Production Program に含まれる誤りを QC Program が再現する可能性がある。この場合、両プログラムの出力が一致しても、それは必ずしも解析結果の正しさを意味しない。

2.2　独立性が懸念されるケース

要因	具体例	懸念点
過去プログラムの流用	過去試験プログラム、テンプレート利用	担当者は異なっていても、構造が過去資産に強く依存する可能性がある。
Production Program の参照	実装内容の閲覧、コード共有、相談時の参照	QC Program が Production Program と同じ実装方針に収束する可能性がある。
納期・リソース制約	短納期、要員不足、複数試験の並行対応	効率優先となり、既存ロジックや本番コードへの依存が高まる可能性がある。

これらの要因は、必ずしも不適切な行為を意味するものではない。過去プログラムやテンプレートの活用は、効率化や標準化の観点では有効である。一方で、ダブルプログラミングの目的が独立した検証である以上、実装構造の過度な類似は品質保証上の議論対象となり得る。

3　独立性評価の現実的な課題

独立性を確認する方法として、担当者へのヒアリング、開発記録の確認、プログラムレビュー、差分比較ツールによる確認などが考えられる。これらは個別ケースの確認には有用である。

しかし、実際の臨床試験では Production Program と QC Program がそれぞれ数百本に及ぶこともある。例えば、1 試験あたり Production Program が 500 本以上、QC Program が 500 本以上存在する場合、比較対象ペアは 500 組以上となる。すべてのペアを人手で詳細に確認することは、工数面でも品質の均一性の面でも現実的ではない。

つまり、独立性の有無を人手で一つずつ判断するのではなく、多数のプログラムから「構造的に非常に似ている組」を客観的に抽出し、レビュー対象の優先順位付けに利用できないか、という点が本稿の出発点である。

4　類似度評価手法の検討

SAS プログラムの類似度を評価する方法として、複数の候補を検討した。本稿では、目視レビュー、テキスト比較、AST 解析、Token + N-gram 手法を比較し、目的に対する適合性を評価した。

要因	Pros	Cons
目視レビュー	人が内容まで理解できる。実装意図や仕様解釈も含めて確認できる。	500 組の比較は非現実的。品質は評価者の経験や判断に依存する。
テキスト比較	既存ツールが多数存在し、導入が容易である。差分を視覚的に確認できる。	整形、改行、変数名変更の影響を強く受ける。構造的類似性の評価には限界があり確認時間が必要である。
AST 解析	プログラム構造を最も正確に比較できる。構文レベルの差異を扱える。	SAS Parser が必要であり、実装コストが高い。多数プログラムへの適用時に処理負荷も大きい。
Token + N-gram	構造情報を一定程度保持しつつ、高速に大量プログラムへ適用できる。	プログラムの意味理解までは行えない。Token 設計に結果が依存する。

4.1　目視レビュー

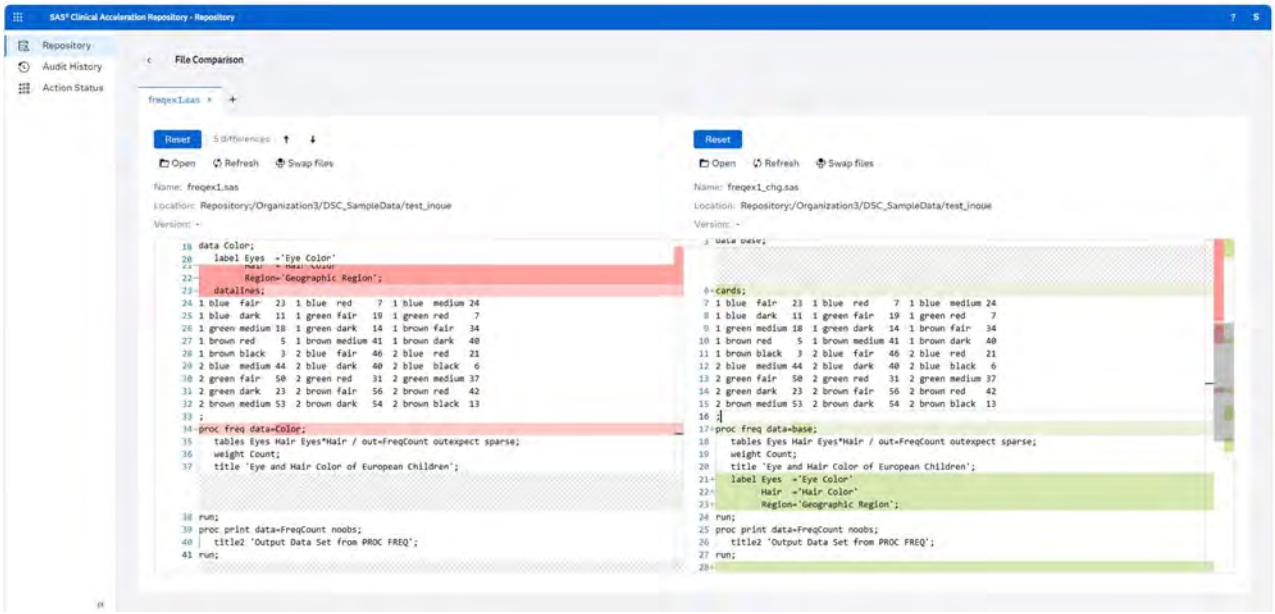
目視レビューは、Production Program と QC Program を並べて確認し、実装内容、処理順序、例外処理、コーディングスタイルなどの類似性を人が判断する方法である。人間が内容を理解しながら確認できるため、実装意図や仕様解釈の差異まで把握できるという利点がある。

一方で、大量プログラムには適用しづらい。500 組のプログラムペアをすべて目視で比較する場合、品質は評価者の経験、集中力、確認基準に大きく依存する。また、評価結果の再現性を確保することも難しい。

4.2　テキスト比較

テキスト比較は、既存ツールを用いてプログラム差分を確認する方法である。既存ツールを利用できるため導入が容易であり、1 ファイル単位の確認には有用である。

参考例として、SAS Clinical Acceleration Repository(SAS CA Repository)でもプログラム比較機能が提供されており、プログラム間の差分を視覚的に確認できる。



しかし、本稿の目的はコードの文字列一致ではなく、実装構造の類似性を評価することである。例えば、データセット名や変数名を変更しただけで差分は大きく見える可能性がある。一方で、表面的な差分が存在しても処理構造はほぼ同一である場合もある。したがって、単純なテキスト比較は構造的類似度評価には十分ではない。

4.3 AST 解析

AST(Abstract Syntax Tree、抽象構文木)解析は、プログラムを構文木として表現し、その木構造を比較する方法である。例えば、SAS コードの if 文を条件部と処理部に分けて木構造として表現できれば、変数名や表面的な書式に左右されず、構造レベルでの比較が可能となる。

```
if sex = 'M' then output;
```

このコードを AST として表現すると、概念的には IF ノードの下に条件ノードと処理ノードが配置される。Token 列が一次元の並びであるのに対し、AST は階層構造を持つため、プログラム構造をより厳密に表現できる。

AST 解析の最大の利点は、構文を理解した比較が可能である点である。変数名やデータセット名の違い、改行やインデントの違いに影響されにくく、実装構造の類似性を精密に評価できる可能性がある。

一方で、SAS プログラムの AST を生成するには SAS Parser が必要となる。SAS は DATA Step、PROC、Macro Language が混在する言語であり、構文解析の難易度が高い。Macro は実行前に展開されるため、Macro を含むプログラムをどの段階で解析するかという課題も存在する。さらに、AST 生成後は Tree 構造同士の比較が必要となり、実装量と計算負荷が増大する。

AST 解析は理論的には最も精密な手法である。しかし、本稿の目的は多数の SAS プログラムを実務的に評価することであり、SAS Parser の構築コストおよび運用負荷を考慮すると、初期実装としては過度に複雑であると判断した。

4.4 Token + N-gram 手法

本稿では、構造比較と実装容易性のバランスを重視し、Token + N-gram 手法を採用した。Token 化により変数名やデータセット名などの表面的な違いを抽象化し、N-gram により局所的な処理構造を特徴量として抽出する。

この方法は AST 解析ほど厳密ではないが、構造情報を一定程度保持しつつ、高速に大量プログラムへ適用できる。ダブルプログラミングの独立性評価を支援するためのサロゲート指標としては、実用性の高い選択肢であると考えた。

5 提案手法

本手法は、SAS Program を入力とし、Lexer による Token 化、5-gram 生成、Rolling Hash による特徴量化、Jaccard Similarity による類似度算出を順に行う。

SAS Program → Lexer → Token List → 5-gram → Rolling Hash → Jaccard Similarity → Similarity Score

5.1 Tokenization

Tokenization では、SAS プログラムを構造比較に適した Token 列へ変換する。ここでは、データセット名や変数名そのものを比較するのではなく、それらを識別子として抽象化する。例えば、adsl、trtp、aval といった具体的な名前は D_IDENTIFIER として扱う。

一方、DATA 文、SET 文、IF 文、演算子、Macro 文など、プログラム構造を示す要素は区別して Token 化す

る。これにより、名前や書式に左右されず、処理構造に着目した比較が可能となる。

以下に Token の一例を示す。

Token	例	意味
D_KEYWORD	data, set, merge, if	DATA Step 中の主要キーワード
D_FUNCTION	scan(), substr()	DATA Step 関数
D_IDENTIFIER	adsl, trtp, aval	データセット名、変数名などの識別子
OPERATOR	=, +, -, eq, ^=	演算子
NUMBER	1, 123, 0.5	数値
M_STATEMENT	%if, %do, %let	Macro 文
M_FUNCTION	%scan, %substr	Macro 関数
M_VAR	&var, &var.	Macro 変数
EOS	;	ステートメント終端
BOUND_END	run, quit	処理ブロック終端

5.2　コメントと書式の扱い

コメント、空白、改行、インデントは、構造的類似性を評価する上ではノイズとなる場合が多い。そのため本手法では、Lexer の前処理としてコメントや空白の影響を除去する方針を採用した。

ただし、SAS ではアスタリスクがコメント開始記号としても演算子としても利用されるため、単純な文字列置換では誤削除が発生する可能性がある。例えば、`TABLE A * B` や `TIME variable * censor(list)` のようなコードでは、アスタリスクをコメントとして扱ってはならない。このため、コメント除去は SAS の記法を考慮して慎重に設計する必要がある。

5.3　5-gram による局所的な類似性評価

N-gram とは、連続する N 個の Token を 1 つの特徴量として扱う手法である。1-gram では単一 Token のみを扱うため、DATA、SET、IF などの出現頻度は把握できるものの、処理の流れや構造的特徴を十分に表現することは難しい。一方、2-gram や 3-gram では隣接する Token 間の関係を捉えることが可能となるが、SAS プログラムにおける条件分岐やデータ加工などの局所的な処理パターンを表現するには情報量が不足する場合がある。

例えば、IF 条件文による分岐や、変数に対する演算結果の代入といった一連の処理を特徴として捉えるためには、より長い連続 Token 列を扱う必要がある。逆に、10-gram のような長い N-gram を採用すると、局所構造を詳細に表現できる一方で、軽微な修正や 1 行の追加による影響を受けやすくなる。N が大きくなるほど、わずかな変更によって多数の N-gram が変化するため、実質的には同一の処理を実装しているプログラムであっても類似度が過度に低下する可能性がある。

N-gram	特徴	本手法での評価
1-gram	単一 Token の出現を評価する。	処理の流れを表現しにくい。
2〜3-gram	隣接 Token の関係を捉える。	局所構造の表現としてはやや短い。
5-gram	複数 Token からなる局所的な処理パターンを保持する。	構造情報と頑健性のバランスが良い。
10-gram 以上	長い構造パターンを保持する。	軽微な変更の影響を受けやすい。

本稿では、構造情報量と変更に対する頑健性のバランスを考慮し、5-gram を採用した。字句解析によって得られたトークン列から、連続する 5 個の Token を抽出し、これを比較の基本単位とすることで、DATA→SET、IF→THEN、変数→演算子→値といった局所的な処理パターンを保持しながらプログラム間の類似性を評価できる。

具体的には、トークン列が

```
A, B, C, D, E, F, G, H, I, J, ...
```

である場合、5-gram として

```
[A, B, C, D, E]
[B, C, D, E, F]
[C, D, E, F, G]
[E, F, G, H, I]
[F, G, H, I, J] ...
```

のように部分列を、開始位置を 1Token ずつ移動させながら生成する。これにより、プログラム全体を単一の特徴として扱うのではなく、局所的な構造パターンの集合として表現することが可能となる。

また、5-gram はプログラム変更の影響を局所化できるという特徴を有する。例えば、トークン D と E の間に新たなトークン X が挿入された場合、挿入位置を含む、または挿入位置をまたぐ 5-gram は変更される。一方で、変更箇所から十分に離れた部分では、変更前と同一の 5-gram が維持される。具体的には、

```
[A, B, C, D, X]
[B, C, D, X, E]
[C, D, X, E, F]
[D, X, E, F, G]
[X, E, F, G, H]
[E, F, G, H, I]
[F, G, H, I, J] ...
```

のように[E, F, G, H, I]以降の 5-gram は、変更前後で一致する。

このように、5-gram による比較では、変更箇所周辺の構造パターンのみが影響を受け、変更箇所から離れた構造パターンは保持される。そのため、1 行追加や軽微な修正が行われた場合でも、プログラム全体の類似度が極端に低下することを抑制できる。特に、独立して作成されたプログラム間の構造的類似性を分析する本研究の目的に対して、5-gram は局所構造の保持と変更耐性を両立する特徴表現として有効である。

5.4 Rolling Hash による特徴量化

生成された 5-gram は、そのまま文字列として比較することも可能である。しかし、大量プログラムを対象とする場合、5-gram 文字列を保持・比較することはメモリと処理時間の点で不利になる。

そこで本手法では Rolling Hash を用いて 5-gram を数値化し、比較しやすい特徴量集合として扱う。Rolling Hash では、隣接する 5-gram 間で共通する Token を再利用しながら Hash 値を計算できるため、効率的な特徴量生成が可能となる。

Hash 化には衝突の可能性が存在するが、本手法は QC の妥当性を直接判定するものではなく、構造的類似性を把握するための支援指標である。そのため、実装容易性、速度、大量適用性を優先し、Rolling Hash を採用した。

5.5 Jaccard Similarity による類似度算出

2 つのプログラムから得られた Hash 集合を比較し、共通する特徴量の割合を Jaccard Similarity として算出する。Jaccard Similarity は以下で定義される。

$$Similarity = \frac{|A \cap B|}{|A \cup B|}$$

値域は 0 から 1 であり、1 に近いほど類似度が高い。本研究では、類似度評価の対象を「プログラムに含まれる特徴的な処理パターンの有無」とし、それぞれの 5-gram の出現回数ではなく、出現したかどうかに着目した。そのため、Hash 値を集合（Set）として扱う通常の Jaccard Similarity を採用した。

Multiset Jaccard を用いて出現回数まで考慮した場合、同一処理を繰り返すプログラムやデータ件数に依存し

たコード構造の影響を受けやすくなり、実装上の差異よりも処理量の差異が類似度に反映される可能性がある。一方、本研究の目的は、ダブルプログラミングにおける実装アルゴリズムの類似性評価であり、特徴パターンの共有率を評価することが重要であるため、Set ベースの Jaccard Similarity を採用した。

6　実装と実装方式の検討

本手法は複数の実装方式を試行しながら開発を進めた。最初に PROC GROOVY を用いた Groovy 版を作成し、その後、クラウド系 SAS 環境での実行制約を踏まえて Python 版を作成した。さらに、PROC PYTHON を利用できる環境が比較的新しい SAS に限定されることを考慮し、外部環境に依存しない Base SAS 版も作成した。

最終的には、速度面と管理性のバランスを考慮し、Python による一括マッチ型(Single-pass Regex Lexer)を代表実装として選択した。

6.1　Lexer 実装方式の比較

Lexer 実装方式として、多重スキャン型(Multi-pass Regex Scanning)、一致マッチ型(Single-pass Regex Lexer)、逐次走査型(Sequential Lexical Analysis)の 3 方式を検討した。

方式	概要	長所	短所
Multi-pass Regex Scanning	Token ごとに全文検索し、最後に位置順へ並べる。	実装が簡単。Token 追加が容易。	同じソースを何度も走査する。Token 競合制御が弱い。
Single-pass Regex Lexer	複数 Token 定義を統合した正規表現で全文を 1 回走査する。	高速。Token 管理と性能のバランスが良い。	統合 Regex の設計とデバッグに慣れが必要。
Sequential Lexical Analysis	先頭から順に Token を確定しながら進む。	Token 優先順位を明示でき、競合に強い。	コード量が多く、Token 追加時の保守負荷が高い。

Python では 3 方式のいずれも比較的容易に実装できる。一方、Groovy では Single-pass Regex Lexer との親和性が高く、Base SAS では Sequential Lexical Analysis が自然である。

本検討では、Groovy で一括マッチ型、Base SAS で逐次走査型も作成しており、同様のスコア傾向が得られた。コメント削除、改行・空白変更、データセット名変更などに対する挙動は概ね一致しており、どの実装方式であっても、類似度スコアの解釈に大きな問題はないと考えられた。

本稿では、可読性、管理性、実行速度のバランスを考慮し、一括マッチ型(Single-pass Regex Lexer)を代表実装として選択した。

6.2　Python 版の特徴

Python 版は、正規表現、List、Set、Dictionary などの標準的なデータ構造を用いて簡潔に実装できる。特に、Token List、5-gram 集合、Hash 集合、Jaccard 計算を自然に表現できる点が利点である。

最初の試作では Multi-pass Regex Scanning を採用したが、Token ごとに全文を複数回走査するため、パターン数が増えると効率面で不利となる。そこで、最終的には Single-pass Regex Lexer を採用した。

Python 版の主な課題は、PROC PYTHON が利用できる SAS 環境に依存する点である。SAS Viya や SAS 9.4M9 以降など、利用可能な環境では有力な選択肢となるが、すべての SAS 環境で利用できるわけではない。

6.3　Groovy 版の特徴

Groovy 版は、PROC GROOVY を利用して SAS から Groovy コードを実行する実装である。Java の正規表現エンジンやデータ構造を利用できるため、Single-pass Regex Lexer や Rolling Hash との親和性が高い。

一方で、PROC GROOVY は実行環境の制約を受けやすい。特に NOXCMD 環境や LOCKDOWN 環境では利

用できず、クラウド系 SAS 環境では実行できないケースがある。この点は、実務システムに導入する上で大きな制約となる。

6.4 Base SAS 版の特徴

Base SAS 版は、DATA Step、PRX 関数、PROC SQL、PROC SORT などを用いて外部言語に依存せず実装した。外部環境に依存しないため、厳格なシステム制約下でも導入しやすい利点がある。

ただし、Lexer を DATA Step で記述するため、Token 定義の追加や修正時にコード量が増えやすい。また、5-gram 生成を SQL の自己結合で実装する場合、大量 Token に対する処理負荷が高くなる可能性がある。

6.5 実装方式の Pros / Cons

実装	Pros	Cons
Python 版	可読性が高い。正規表現、集合演算、Hash 処理を簡潔に実装できる。Single-pass 方式により性能と管理性のバランスが良い。	PROC PYTHON が必要であり、利用できる SAS 環境に制約がある。
Groovy 版	Single-pass Regex Lexer との親和性が高い。Java 資産を利用でき、Rolling Hash も実装しやすい。	PROC GROOVY が必要。NOXCMD や LOCKDOWN 環境では実行できない可能性がある。保守人材が限られる。
Base SAS 版	外部言語に依存せず Base SAS のみで実行可能。SAS Programmer には導入しやすい。	コード量が多い。保守負荷が高い。実行速度など性能面で不利となる可能性がある。

7 評価例

サンプルプログラムを用いて、提案手法の挙動を確認した。評価では、コメント除去、タブ除去、データセット名変更、軽微なロジック変更、異なるサンプルプログラムとの比較を行った。

比較対象	gramsA	gramsB	common	union	similarity	解釈
コメント除去	57	57	57	57	1.000	コメント除去の影響を受けない。
タブ除去	57	57	57	57	1.000	タブや空白の影響を受けない。
データセット名変更	57	57	57	57	1.000	データセット名変更の影響を受けない。
軽微なロジック変更	57	65	54	68	0.794	一部修正により類似度は低下するが構造の多くは共通している。
異なるサンプルプログラム	57	128	41	144	0.285	異なるサンプルプログラムでは類似度が低くなる。

これらの結果から、本手法はコメント、空白、改行、単純な名称変更の影響を受けにくく、プログラム構造の類似性を反映することが確認できた。一方で、異なる処理構造を持つプログラムでは類似度が低下した。

なお、本評価は手法の挙動を説明するためのサンプル評価であり、Similarity Score の閾値を定義するものではない。実運用では、試験、プログラム種別、標準テンプレートの有無などを踏まえて解釈する必要がある。

8 考察

8.1 Similarity Score の意味

Similarity Score は、2 つの SAS プログラムがどの程度共通する 5-gram 特徴量を持つかを示す指標である。高い Similarity Score は、2 つのプログラムが構造的に類似していることを意味する。

ただし、類似度が高いことは直ちに問題を意味しない。独立に作成されたプログラムでも類似度が高くなる場合がある。

8.2　独立性のサロゲート指標としての活用

本手法は、独立性そのものを直接測定するものではない。しかし、Production Program と QC Program の構造的類似性を客観的に示すことで、独立性に関する議論の出発点を提供できる。

特に、多数のプログラムを対象とする場合、Similarity Score が高いペアを優先的に確認することで、レビューの効率化に寄与する可能性がある。これは、人によるレビューを代替するものではなく、レビュー対象の優先順位付けを支援するものである。

8.3　組織運用上の注意

類似度評価ツールは、担当者を監視するための道具として利用すべきではない。独立性が低下する背景には、納期、リソース、業務負荷、レビュー方針など、組織的な要因が存在する場合が多い。

したがって、本ツールは個人を責めるためではなく、品質向上のための対話のきっかけとして活用することが望ましい。独立性を求めるのであれば、独立性を維持できる環境を同時に提供する必要がある。

9　本手法の限界

本手法には以下の限界がある。

- プログラムの正しさを評価するものではない。
  Similarity Score が低くても QC Program が正しいとは限らず、Similarity Score が高くても誤りがあるとは限らない。
- 独立性を直接評価するものではない。
  Similarity Score は独立性のサロゲート指標であり、類似度のみから独立性の有無を断定することはできない。人による評価を代替するものではなく、補助する目的のものである。特に、多数のプログラムを対象とする場合、Similarity Score が高いペアを優先的に確認することで、レビューの効率化に寄与する可能性がある。
- 意味理解までは行わない。
  Token + N-gram 手法は構造的な類似性を評価するが、プログラムの意味や統計的妥当性を理解するものではない。
- Token 設計に結果が依存する。
  どの SAS 構文要素をどの Token として扱うかは、スコアに影響する可能性がある。

10 今後の課題

今後の課題として、実試験データを用いた Similarity Score 分布の確認、プログラム種別ごとの傾向分析、閾値設定の検討が挙げられる。また、5-gram 以外の N-gram 長との比較や、Token 設計の妥当性評価も重要である。

さらに、レビュー支援ツールとして活用する場合には、Similarity Score だけでなく、どの構造パターンが共通しているかを可視化する機能が有用と考えられる。これにより、単なる数値ではなく、レビュー担当者が判断しやすい説明可能な出力を提供できる可能性がある。

11 結論

本稿では、SAS プログラムを Token 化し、5-gram、Rolling Hash、および Jaccard Similarity を用いて構造的類似度を定量化する手法を提案した。

本手法により、コメント、空白、改行、データセット名変更の影響を受けにくい比較が可能となり、プログラム構造の類似性を Similarity Score として可視化できることを示した。

Similarity Score はプログラムの正しさや独立性そのものを評価するものではない。しかし、Production Program と QC Program の独立性について客観的に議論するためのサロゲート指標として活用できる可能性がある。

また、本ツールは独立性を作り出すものではなく、独立性を維持するための教育、意識付け、業務環境の整備と組み合わせて利用されるべきである。人を追い込むためではなく、品質向上のための対話を促すツールとして活用することが望ましい。

参考文献

[1] Base SAS Software - SAS 9.4 & SAS Viya 3.4
https://documentation.sas.com/doc/ja/pgmsascdc/9.4_3.5/pgmsashome/home.htm
(参照: 2026-9-1)

[2] SAS/STAT 15.1 User's Guide Example Programs (Sample Library)
https://support.sas.com/documentation/onlinedoc/stat/ex_code/151/index.html
(参照: 2026-9-1)

[3] Karp RM, Rabin MO.
Efficient randomized pattern-matching algorithms.
IBM Journal of Research and Development. 1987;31(2):249-260.
doi:10.1147/rd.312.0249.

[4] Jaccard P.
Étude comparative de la distribution florale dans une portion des Alpes et du Jura.
Bulletin de la Société Vaudoise des Sciences Naturelles. 1901;37:547-579.

[5] Manning CD, Raghavan P, Schütze H.
Introduction to Information Retrieval.
Cambridge University Press; 2008.
https://nlp.stanford.edu/IR-book/information-retrieval-book.html
(参照: 2026-9-1)

[6] Leskovec J, Rajaraman A, Ullman JD.
Mining of Massive Datasets. 3rd ed.
Cambridge University Press; 2020.
http://www.mmds.org/
(参照: 2026-9-1)

JADERを用いた有害事象の発現時期パターンの評価手法

○玉木志織[1,2] 、室谷健太[3,4]、五百路徹也[4]

([1]日本メディア株式会社、[2]久留米大学大学院医学研究科、[3]久留米大学医学部医療検査学科、[4]久留米大学バイオ統計センター)

Evaluation of Time-to-Onset Patterns of Adverse Events Using the Japanese Adverse Drug Event Report Database

Shiori Tamaki[1,2], Kenta Murotani[3,4], Tetsuya Ioji[4]

[1]Nihon Media Co., Ltd. , [2]Graduate School of Medicine, Kurume University,

[3]Department of Medical Technology, School of Medicine, Kurume University,

[4]Biostatistics Center, Kurume University

要旨

Japanese Adverse Drug Event Report database（JADER）に収載された薬剤投与日および有害事象発現日の情報を用いて、有害事象の発現までの時間的特徴を評価した。ニボルマブ単剤が投与された非小細胞肺癌の症例を対象に、有害事象ごとの Restricted Mean Time Lost（RMTL）を算出し、365 日時点の RMTL で正規化した経時的推移に有限混合回帰モデルを適用した。解析対象となった 777 種類の有害事象は 6 クラスに分類され、ROR による報告不均衡の程度が近い有害事象であっても、異なる時間的特徴を示す場合が認められた。ROR と発現までの時間情報を併せて評価することで、ROR のみでは同程度に見えるシグナル候補を時間的特徴に基づいてさらに特徴づけられる可能性が示唆された。

キーワード：JADER、有害事象、発現時期、RMTL、有限混合回帰モデル

1. 背景

承認前に実施される臨床試験のみでは、医薬品に関連するすべての有害事象を把握することは困難である。臨床試験では、対象患者の背景、症例数、併用薬および観察期間等が限定されるため、発現頻度が低い有害事象、長期投与後に発現する有害事象、ならびに特定の患者集団で生じる有害事象を十分に把握できない場合がある。そのため、製造販売後に収集される安全性情報を継続的に監視し、医薬品と有害事象との新たな関連、または既知の関連に関する新たな側面を示唆する情報を検出・評価する医薬品安全性監視が重要となる。

製造販売後の安全性情報源の一つとして、自発報告データベースがある。自発報告制度は、医薬品の使用後に認められた有害事象について、医療従事者、患者および製薬企業等から報告を収集する制度であり、臨床試験では捉えにくい安全性上の問題を探索するうえで重要な役割を担っている。一方、自発報告データには、未報告、重複報告、欠測、報告内容の不正確さ、ならびに社会的関心や注意喚起に伴う報告傾向の変化

等が存在する。また、医薬品を使用した患者総数および各患者の正確な観察期間を把握できないため、自発報告データのみから有害事象の発生率またはリスクを直接推定することはできない。
日本では、独立行政法人医薬品医療機器総合機構（PMDA）が Japanese Adverse Drug Event Report database（JADER）を公開している。JADER は、国内で報告された副作用が疑われる症例を収載した自発報告データベースであり、2004 年 4 月以降に収集された症例情報が含まれ、毎月更新されている。JADER は無償で利用可能な公開データであるため、企業内の安全性データベースや商用の診療データベースへのアクセスが限られる研究者においても利用できる。
自発報告データベースを用いたシグナル探索には、Reporting Odds Ratio（ROR）や Proportional Reporting Ratio（PRR）、Bayesian Confidence Propagation Neural Network（BCPNN）、Multi-item Gamma-Poisson Shrinker（MGPS）等の不均衡分析が用いられている。これらは計算方法が異なるものの、特定の医薬品と有害事象の組合せが、データベース内の他の報告と比較して相対的に高い頻度で報告されているか、すなわち報告の不均衡が認められるかを評価する方法である。ROR および PRR は 2×2 分割表に基づいて報告不均衡を評価する。一方、BCPNN および MGPS は、それぞれベイズ法または経験ベイズ法による縮約を用いることにより、特に報告数が少ない組合せにおける推定値の不安定性を抑制する。各手法の判定基準を満たした医薬品と有害事象の組合せは、報告不均衡を示す組合せとして、詳細な評価を行う候補となる。
ただし、不均衡指標は、自発報告データベース内における相対的な報告の偏りを示すものであり、医薬品と有害事象との因果関係、実臨床における発生率、リスクの大きさ、またはシグナルの確定を示すものではない。したがって、不均衡分析によって抽出された医薬品と有害事象の組合せについては、個別症例の医学的情報、既知の安全性情報、関連文献、他のデータソース等に基づく追加評価が必要となる。
ROR 等の不均衡分析では、解析対象期間内の報告を一つの 2×2 分割表に集約する。そのため、対象医薬品と有害事象の組合せに報告不均衡が認められるかを評価できる一方、医薬品の投与開始から有害事象発現までの期間およびその経時的な分布を直接表現しない。したがって、同程度の ROR を示す複数の有害事象が存在する場合でも、投与開始後早期に発現報告が集中する有害事象、一定期間にわたって継続的に発現が報告される有害事象、一定期間経過後に発現報告が顕在化する有害事象を、ROR のみから区別することは困難である。
有害事象の発現時期は、医薬品と有害事象との時間的関係を特徴づけるための情報の一つである。近い ROR を示す有害事象であっても、一方の発現報告が投与開始後早期に集中し、他方が一定期間経過後に増加する場合には、両者は異なる時間的特徴を有すると考えられる。このような違いを把握することにより、ROR のみでは区別できない有害事象間の時間的な異質性を示し、個別症例等の追加評価において検討すべき時間的仮説を、有害事象ごとに具体化できる可能性がある。
JADER に収載された医薬品の投与開始日および有害事象発現日を利用することで、投与開始から有害事象発現までの期間を算出し、医薬品と有害事象との時間的関係を検討することができる。
ただし、JADER は医薬品を使用した患者を継続的に追跡するコホートデータではないため、得られる結果は実臨床における有害事象の発生率や累積発生確率を示すものではなく、日付情報が利用可能な報告症例における発現時期の特徴として解釈する必要がある。
以上より、ROR による報告不均衡と JADER に記録された発現までの時間情報を併せて評価することで、ROR 単独では直接表現されない有害事象の時間的特徴を記述できる可能性がある。特に、近い ROR を示す有害事象間で発現までの時間的特徴が異なる場合、ROR では同程度のシグナル候補に見える有害事象を区別し、候補ごとの追加評価において検討すべき時間的仮説を具体化できると考えられる。

2. 目的

本研究では、JADER に収載された医薬品の投与開始日および有害事象発現日の情報を活用し、ROR による報告不均衡と発現までの時間的特徴を併せて評価する方法について探索的に検討した。これにより、ROR では同程度に見えるシグナル候補を時間的特徴によって区別し、追加評価における検討仮説を具体化することを目的とした。

3. 方法

3.1. 研究デザインおよびデータソース

本研究は、Japanese Adverse Drug Event Report database（JADER）を用いた後ろ向きの探索的研究である。独立行政法人医薬品医療機器総合機構（PMDA）が公開する JADER データを使用した。JADER は、症例一覧テーブル（demo）、医薬品情報テーブル（drug）、有害事象テーブル（reac）および原疾患テーブル（hist）の 4 テーブルから構成される。各テーブルを症例識別番号により連結し、解析用データセットを作成した。有害事象名は、Medical Dictionary for Regulatory Activities/Japanese version（MedDRA/J）の基本語（Preferred Term、PT）を用いて集計した。

3.2. 解析対象

対象薬剤はニボルマブとし、対象疾患は非小細胞肺癌とした。drug テーブルにおいてニボルマブを抽出した。さらに、hist テーブルまたは drug テーブルの使用理由に基づき、非小細胞肺癌に対してニボルマブが使用された症例を特定した。

本研究では、ニボルマブ単剤投与例を対象とし、次のいずれかに該当する症例を除外した。

1. ニボルマブの投与開始日が欠測している症例
2. 有害事象発現日が欠測している症例
3. 有害事象発現日がニボルマブ投与開始日より前である症例

同一症例において同一の有害事象が複数回登録されている場合は、最初の発現日のみを採用した。

3.3. データクリーニングおよび重複処理

JADER には同一症例に対する追加報告が含まれる可能性があるため、症例識別番号および報告回数等を用いて重複処理を行った。医薬品名については、一般名、販売名からニボルマブおよびオプジーボに該当する記録を抽出した。日付情報については、完全日付のみを採用した。

3.4. 投与開始から有害事象発現までの期間の定義

時間の起点は、各症例におけるニボルマブの初回投与日とした。対象有害事象が報告された症例では、当該有害事象の初回発現日をイベント日とし、次式により発現までの日数を算出した。

$$T_{ij} = \textit{対象有害事象}\ j\textit{の初回発現日} - \textit{症例}\ i\textit{のニボルマブ初回投与日}$$

解析対象期間は投与開始後 365 日までとし、365 日を超えて発現した有害事象は解析対象外とした。対象有害事象が報告されていない症例については、当該症例に記録された他の有害事象の最終発現日を打ち切り日

とした。ただし、この日は通常のコホート研究における観察終了日とは異なり、JADER 上で確認可能な最終の日付情報に相当する。
したがって、本研究で推定する曲線およびその要約指標は、実臨床における有害事象の未発現確率、累積発生確率または発生リスクを推定するものではなく、解析可能な日付情報を有する JADER 報告症例における時間的特徴を記述するための指標として用いた。

3.5.　発現までの時間的特徴の推定

各有害事象について、ニボルマブ初回投与日を起点として、Kaplan-Meier 法によりイベント未発現曲線$S_j(t)$を推定した。さらに、次式により累積イベント曲線を求めた。

$$F_j(t) = 1 - S_j(t)$$

j：有害事象
t：ニボルマブの初回投与日からの有害事象発現までの経過日数

前述のとおり、JADER は追跡型のコホートデータではないため、$S_j(t)$および$Fj(t)$は真の生存関数または累積発生関数として解釈せず、本研究で定義した打ち切り規則の下で得られる時間的推移として取り扱った。

3.6.　Restricted Mean Time Lost の算出

有害事象ごとの時間的推移を複数の評価時点までの面積として要約するため、Restricted Mean Time Lost（RMTL）を算出した。有害事象 j の評価時点τにおける RMTL は、次式で定義した。

$$RMTL_j(\tau) = \int_0^\tau \{1 - S_j(t)\}dt = \int_0^\tau F_j(t)dt$$

RMTL は、評価時点までの累積イベント曲線の下面積に相当する。本研究では、投与開始後 30、60、90、120、150、180、210、240、270、300、330 および 365 日の 12 時点について RMTL を算出した。
単一の評価時点における RMTL のみではなく、複数の評価時点における RMTL の推移を用いることで、投与開始後のどの時期から曲線が増加し、その後どのように推移するかを数値的に表現した。

3.7.　RMTL の正規化

有害事象間で RMTL の絶対的な大きさではなく、365 日までの経時的な推移を比較するため、各評価時点における RMTL を 365 日時点の RMTL で除して正規化した。

$$NRMTL_j(\tau) = \frac{RMTL_j(\tau)}{RMTL_j(365)}$$

ここで、$NRMTL_j(\tau)$は有害事象jの評価時点τにおける正規化 RMTL を表す。
365 日時点の RMTL が 0 であった有害事象については、正規化 RMTL を定義できないため分類対象から除外した。正規化によって各有害事象の 365 日時点の値は 1 となるため、本解析で比較する対象は、報告数または RMTL の絶対的な大きさではなく、365 日までの相対的な時間的推移の形状である。

3.8.　発現までの時間的特徴に基づく分類

有害事象ごとの正規化 RMTL の経時的推移を分類するため、有限混合回帰モデルを適用した。有限混合回帰モデルは、観測対象が異なる回帰構造を有する有限個の潜在クラスから構成されると仮定し、クラスごとの回帰曲線、混合比率および各対象のクラス所属確率を推定するモデルである。
有害事象 iの 12 評価時点における正規化 RMTL をy_i、評価時点をt、潜在クラス数を Kとしてモデルを次式で表した。

$$f(\boldsymbol{y}_i \mid \boldsymbol{t}) = \sum_{k=1}^{K} \pi_k \, f_k(y_i \mid t, \theta_k)$$

π_k：クラスkの混合比率
θ_k：クラス
k：固有の回帰係数および残差分散等のパラメータ

各有害事象について、観測された正規化 RMTL の経時的推移が各潜在クラスの回帰曲線に適合する程度を、モデルに基づく事後所属確率として算出した。各有害事象は、事後所属確率が最も高いクラスに分類した。

4. 結果

4.1. 解析対象症例および有害事象

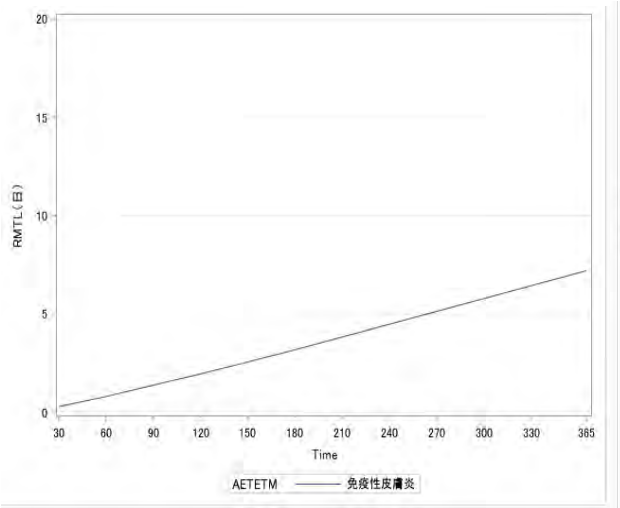
2026 年 1 月に公開された JADER データを使用した。ニボルマブ単剤が投与された非小細胞肺癌 4,993 症例を解析対象とした。対象症例には 817 種類、延べ 8,631 件の有害事象が抽出された。817 種類の有害事象のうち、365 日時点の Restricted Mean Time Lost（RMTL）が 0 であった 40 種類は、正規化 RMTL を算出できないために除外した。この結果、777 種類の有害事象を有限混合回帰モデルによる分類対象とした。また、817 種類の有害事象のうち、報告症例数が 1 件のみであった有害事象は 362 種類であり、全体の 44.3%を占めた。

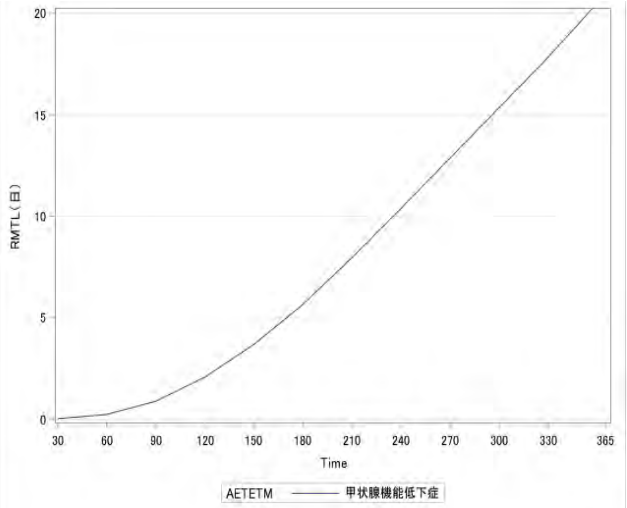
4.2. RMTL の経時的推移

各有害事象について、投与開始後 30 日から 365 日までの 12 評価時点における RMTL を算出した。有害事象間で RMTL の絶対値ではなく、365 日までの相対的な経時的推移を比較するため、各評価時点の RMTL を 365 日時点の RMTL で除して正規化した。

基本語(PT)	発症期間ごとのRMTL											
	30日	60日	90日	120日	150日	180日	210日	240日	270日	300日	330日	365日
1型糖尿病	0.00553801	0.086230329	0.265812139	0.496872073	0.808366842	1.243027468	1.734785307	2.351889787	3.158637841	4.066559735	5.242930688	6.938552577
2型糖尿病	0.000285551	0.016089551	0.045420902	0.075561157	0.112764221	0.177351166	0.284235517	0.441753498	0.599271479	0.756789461	0.914307442	1.193001971
A型インフルエンザウイルス検査陽性	0	0	0	0.001974984	0.021724819	0.041474654	0.06122449	0.080974325	0.100724161	0.120473996	0.140223832	0.163265306
B型肝炎再活性化	0.002095338	0.009952855	0.017810372	0.025667889	0.033525406	0.041382923	0.04924044	0.057097957	0.064955474	0.072812991	0.080670508	0.089837611
C-反応性蛋白増加	0.034428865	0.084283401	0.160003961	0.250835431	0.353332829	0.475029137	0.596725445	0.718421753	0.840118061	0.961814368	1.083510676	1.225489702
CD4リンパ球減少	0	0	0	0	0.018906145	0.039162728	0.059419311	0.079675895	0.099932478	0.120189061	0.140445645	0.164078326
COVID-19	0.0137407	0.035723677	0.066679422	0.097635167	0.128590911	0.159546656	0.251441634	0.347689414	0.490955672	0.634221931	0.86048827	1.374989899
COVID-19肺炎	0.003567182	0.010701546	0.01783591	0.024970274	0.032104637	0.039239001	0.046373365	0.053507729	0.060642093	0.067776457	0.074910821	0.083234245

365 日より前の評価時点では、有害事象によって正規化 RMTL の推移が異なり、比較的早い時点までに値が大きく増加する有害事象（例：免疫性皮膚炎）と、前半の増加が小さく、後半の増加が大きい有害事象（例：甲状腺機能低下症）が認められた。



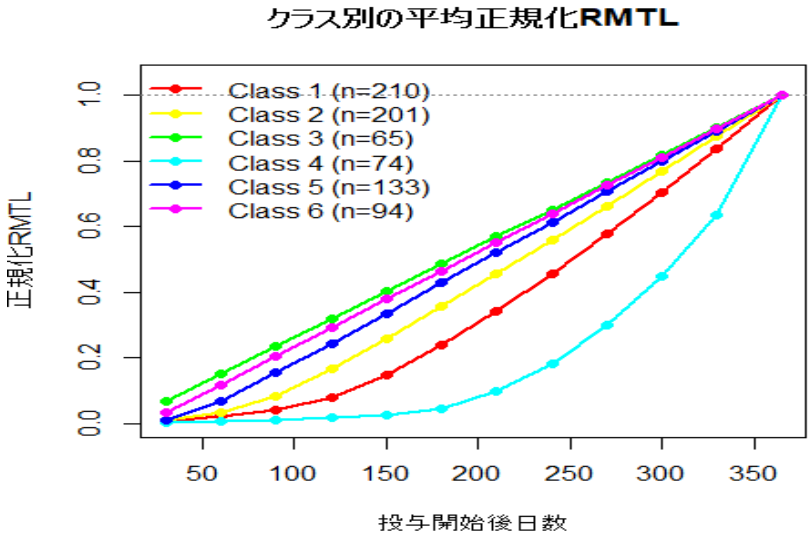


4.3. 有限混合回帰モデルによる分類

777 種類の有害事象について、12 評価時点の正規化 RMTL の経時的推移に有限混合回帰モデルを適用した。クラス数を 2 から 6 まで設定してモデルを比較した結果、検討した範囲では 6 クラスモデルの Bayesian Information Criterion（BIC）が最小であった。このため、6 クラスモデルを採用した。

各有害事象について、正規化 RMTL の経時的推移が各潜在クラスに適合する程度を、モデルに基づく事後クラス所属確率として算出した。各有害事象は、事後クラス所属確率が最も高いクラスに分類した。

本研究では、各クラスの特徴を示すため、正規化 RMTL の相対的な経時的推移に基づく記述的名称を付与した。採用した 6 クラスモデルでは、比較的早い時点から正規化 RMTL が増加するクラス 3（早期増加型）およびクラス 6（準早期増加型）、中間的な推移を示すクラス 5（中間増加型）、前半の増加が比較的小さく中盤以降に増加するクラス 2（中後期増加型）およびクラス 1（後期増加型）、ならびに前半ではほとんど増加せず後半に大きく増加するクラス 4（晩期増加型）が認められた。



4.4. ROR と有害事象発現時期クラスとの対応

各有害事象について ROR および 95%信頼区間を算出し、有限混合回帰モデルによる分類結果との対応を検討した。ROR の 95%信頼区間下限が 1 を上回った有害事象は、特定のクラスに限定されず、6 クラスすべてに認められた。また、各クラスには ROR の値が異なる有害事象が含まれていた。さらに、近い ROR を示しながら異なるクラスに分類された有害事象も認められた。以上より、ROR による報告不均衡の程度が近い有害事象であっても、投与開始から発現までの時間的特徴が異なる場合があることが示された。

5. 考察

5.1. 主要な結果

本研究では、JADER に収載されたニボルマブ単剤投与の非小細胞肺癌症例を対象として、有害事象ごとの投与開始から発現までの時間的特徴を検討した。各評価時点の Restricted Mean Time Lost（RMTL）を 365 日時点の RMTL で除して正規化し、その経時的推移に有限混合回帰モデルを適用した。その結果、検討した 2 から 6 クラスの範囲では 6 クラスモデルの Bayesian Information Criterion（BIC）が最小であった。得られたクラス間では、比較的早い時点までに正規化 RMTL が大きく増加する推移や、前半の増加が小さく後半にかけて増加する推移等、異なる時間的特徴が認められた。

また、ROR の 95%信頼区間下限が 1 を上回る有害事象は、特定のクラスに限定されず、複数の時間クラスに分布していた。さらに、ROR による報告不均衡の程度が近い有害事象であっても、異なるクラスに分類される例が認められた。例えば、肺炎は報告件数 960 件、ROR 1.48 でクラス 1 に分類されたのに対し、尿細管間質性腎炎は 179 件、ROR 1.47 でクラス 2 に分類された。また、食欲減退は 654 件、ROR 1.68 でクラス 1 に分類された。一方、発疹は 760 件、ROR 1.66 でクラス 5 に分類された。

これらの結果から、ROR による報告不均衡の程度が近い有害事象であっても、投与開始から発現までの相対的な時間推移は異なる場合があり、発現までの時間的特徴は ROR の大きさのみからは把握できないことが示唆された。ROR と発現までの時間情報を併せて評価することにより、ROR のみでは同程度に見えるシグナル候補を、時間的特徴に基づいてさらに区別できる可能性がある。

5.2. ROR と時間情報を併せて評価する意義

ROR は、特定の医薬品と有害事象の組合せが、自発報告データベース内で相対的に偏って報告されているかを示す指標である。本研究で算出した ROR は、解析対象期間内の報告を一つの 2×2 分割表に集約したものであり、投与開始から有害事象発現までの時間を算出過程に含まない。

一方、正規化 RMTL に基づく分類は、投与開始から 365 日までの発現報告の相対的な時間推移を表す。このため、ROR の値が近いことは報告不均衡の程度が近いことを意味するが、発現までの時間的特徴が近いことを意味しない。本研究で、近い ROR を示す有害事象が異なるクラスに分類されたことは、同程度の報告不均衡を示す候補の中に、異なる時間的特徴が存在することを示している。

したがって、ROR と発現までの時間情報を併せて評価することにより、ROR のみでは同程度に見えるシグナル候補を、時間的特徴に基づいてさらに区別できる可能性がある。また、ROR によって抽出された候補について、投与開始後のどの時期に発現した症例に注目して評価するかを具体化するための情報となり得る。

ただし、時間クラスは新たなシグナルを確定するものではなく、医薬品との因果関係や臨床的重要性を示すものでもない。また、本研究の分類から、臨床上の監視期間や治療中止時期を直接決定することはできない。ROR と時間情報は、報告不均衡の程度と発現報告の時間的推移をそれぞれ示す、補完的な情報として解釈する必要がある。

5.3. 正規化 RMTL とクラス分類の解釈

本研究では、有害事象間で RMTL の絶対値ではなく、365 日までの相対的な経時的推移を比較するため、各時点の RMTL を 365 日時点の RMTL で除して正規化した。この操作により、報告数や RMTL の大きさが異なる有害事象についても、時間的推移の形状を共通の尺度で比較できる。

一方、正規化によって RMTL の絶対的な大きさに関する情報は、クラス分類に直接反映されない。そのため、同じクラスに分類された有害事象であっても、正規化前の RMTL365 や報告症例数が大きく異なる場合がある。したがって、クラス分類は、報告不均衡の強さ、有害事象の発生頻度、重篤度または臨床的重要性を表すものではない。

有限混合回帰モデルによって得られた 6 クラスは、医学的に確立された 6 種類の発現型が存在することを示すものではない。これらは、本研究の解析対象、評価期間、正規化方法およびモデル設定の下で、正規化 RMTL の経時的推移が類似する有害事象を統計的に要約したものである。このため、早期増加型等のクラス名称は、クラス別曲線の形状を表す記述的名称として解釈する必要がある。

5.4. 本研究の限界

本研究にはいくつかの限界がある。第一に、JADER は自発報告データベースであり、未報告、重複報告、欠測、報告内容の不正確さ、ならびに社会的関心や注意喚起に伴う報告傾向の変化等の影響を受ける。また、医薬品を使用した患者総数を把握できないため、本研究で算出した ROR はデータベース内における相対的な報告の不均衡を示すものであり、実臨床における有害事象の発生率、リスクの大きさ、または医薬品との因果関係を示すものではない。

第二に、本研究の発現時期に関する解析は、医薬品の投与開始日および有害事象発現日の完全日付が利用可能な報告症例に限定されている。そのため、日付情報が欠測している症例との間に系統的な違いが存在する可能性がある。また、JADER からは正確な観察終了日を把握できないため、対象有害事象が報告されていない症例については、当該症例に記録された他の有害事象の最終発現日を打ち切り日として用いた。しかし、この日付は通常のコホート研究における観察終了日を表すものではなく、打ち切りが有害事象の発現と独立であることも保証されない。したがって、本研究で推定した曲線および RMTL は、実臨床における未発現確率や累積発生確率を推定するものではなく、本研究で定義した打ち切り規則の下で得られた記述的な指標として解釈する必要がある。

第三に、本研究で解析対象となった有害事象には、報告症例数の少ないものが多数含まれていた。特に、817 種類の有害事象のうち 362 種類（44.3%）は報告症例数が 1 例のみであり、このような有害事象では、正規化 RMTL の経時的推移およびクラス分類が少数症例の発現日に大きく依存する可能性がある。したがって、報告症例数の少ない有害事象のクラス所属については慎重に解釈し、少数の報告に基づく分類が不安定となり得る点に留意する必要がある。

第四に、本研究で得られたクラス分類は、365 日という評価期間、RMTL の正規化方法、評価時点および有限混合回帰モデルの設定に依存する。本研究では 2 から 6 クラスまでを比較し、BIC が最小となった 6 クラスモデルを採用したが、異なるクラス数やモデル設定によって別の分類が得られる可能性がある。さらに、

本研究の結果は JADER に収載されたニボルマブ単剤投与の非小細胞肺癌症例に基づくものであり、他の医薬品、疾患またはデータベースに対する一般化可能性は明らかではない。今後、異なる解析条件および他のデータを用いて、本研究で得られた発現時期パターンの再現性を検討する必要がある。
以上の限界を踏まえ、本研究で得られたクラス分類は、医学的に確立された有害事象の発現型を示すものではなく、JADER 上で観察された発現報告の時間的特徴を探索的に整理した結果として解釈する必要がある。そのうえで、本研究の方法は、ROR による報告不均衡に加えて、投与開始から発現までの時間的特徴を併せて検討するための一つの探索的アプローチとなる可能性がある。

6. 結論

本研究では、JADER に記録された投与開始日および有害事象発現日の情報を用い、正規化 RMTL の経時的推移に基づいて有害事象を 6 クラスに分類した。ROR による報告不均衡の程度が近い有害事象であっても、投与開始から発現までの時間的特徴が異なる場合が認められた。これらの結果から、発現までの時間的特徴は ROR の大きさのみからは把握できず、ROR と時間情報を併せて評価することで、シグナル候補を時間的観点からさらに特徴づけられる可能性が示唆された。

参考文献

1）久保田潔：自発報告からのシグナル検出－英国 MCA，米国 FDA，WHO の新しい方法－．薬剤疫学，6（2），101－108，2001．　（最終閲覧日：2026 年 9 月 18 日）

2）厚生労働省：承認後の安全性情報の取扱い：緊急報告のための用語の定義と報告の基準について．　（最終閲覧日：2026 年 6 月 13 日）

3）独立行政法人医薬品医療機器総合機構：データセットのダウンロードページで提供する情報について．（最終閲覧日：2026 年 5 月 23 日）

4）酒井隆全：JADER を用いた研究発表の際に留意すべきチェックリストの提案．薬剤疫学，25（2），64-73，2020．

5）広岡禎，山田雅之：PMDA の「医薬品副作用データベース」を用いた副作用リスクの評価－自社医薬品副作用リスクのポジショニングへの活用－．SAS ユーザー総会論文集，2012．

6）Grün B, Leisch F. FlexMix version 2: Finite mixtures with concomitant variables and varying and constant parameters. Journal of Statistical Software. 2008;28(4):1-35.

7）Grün B, Leisch F, Sarkar D, Mortier F. flexmix: Flexible Mixture Modeling. R package version 2.3-20. https://cran.r-project.org/package=flexmix　（最終閲覧日：2026 年 9 月 18 日）

PROC PYTHONで広がるSASの可能性

― Excel連携からデータ活用の新しい世界へ ―

○小澤　義人[1]、髙橋　康平[1]、高橋　杏奈[1]、稲葉　洋介[1]

（[1]千葉大学 医学部附属病院 臨床試験部生物統計室）

Expanding the Possibilities of SAS with PROC PYTHON
From Excel Integration to a New World of Data Utilization

Yoshihito Ozawa, Kohei Takahashi, Anna Takahashi, Yosuke Inaba

Biostatistics section, Clinical Research Center

Chiba University Hospital

要旨

SAS 9.4M9 では、2026 年 7 月 8 日公開の Hot Fix（R7Z001） により PROC PYTHON が利用可能となり、Python が提供する豊富なライブラリや外部機能を SAS の処理フローに組み込むことが可能となった。これは、従来 SAS で行ってきた処理を単純に Python へ置き換えるものではなく、SAS が得意とするデータ処理・統計解析を基盤としながら、Excel 操作、Figure 作成、各種データ形式の処理、外部システムとの連携などへ処理範囲を拡張するものと位置付けられる。本稿では、PROC PYTHON 以前に用いられてきた SYSTASK および PROC FCMP Python object との違いを整理したうえで、Excel テンプレートへの出力、Heatmap 作成、CDISC Dataset-JSON の読み込みという 3 つの実例を紹介する。また、Hot Fix、Python 実行環境、requirements.txt など、導入時に必要となる設定と運用上の課題について述べる。

キーワード：PROC PYTHON、SAS、Python、Excel、Dataset-JSON、データ可視化

1. はじめに

Python の大きな特徴の一つは、言語そのものの機能に加えて、多様な分野のライブラリ、パッケージ、ツールから構成される豊富な**Python Ecosystem**を利用できる点にある。Python Package Index（PyPI）には多数の公開パッケージが登録されており、NumPy や SciPy による数値計算・科学技術計算、pandas によるデータ加工、Matplotlib などによる可視化、scikit-learn による機械学習、openpyxl や win32com による Excel を含む Office 文書の操作など、目的に応じた機能を組み合わせて利用することができる。この Ecosystem はデータ解析に限定されず、JSON、XML、CSV など多様なデータ形式の処理、フォルダやサブフォルダを対象とした再帰的なファイル処理、REST API や HTTP 通信を利用した外部システムとのデータ交換など、データの取得から加工、解析、可視化、成果物の作成まで幅広い処理を Python 上で構築し、SAS のネイティブデータセット

として連携ができる。PROC PYTHON の意義は、単に SAS から Python コードを実行できることだけではなく、この Python Ecosystem を SAS から利用できるようになることにある。既存の SAS プログラム、SAS データセットおよび解析業務の流れを維持しながら、必要な処理に応じて Python のライブラリや外部機能を組み込むことが可能となる。すなわち、PROC PYTHON は SAS を Python に置き換えるための機能ではなく、SAS が持つデータ処理・統計解析機能と Python Ecosystem を接続するインターフェイスとして捉えることができる。これにより、SAS を解析の中心に据えながら、その周辺に Excel 連携、Figure 作成、多様なデータ形式への対応、ファイル処理、Web/API 連携などの機能を拡張することが容易になることが期待される。

2. SAS から Python を利用する方法の変遷

PROC PYTHON が提供される以前にも、SAS から Python を利用する方法は存在した。第一の方法は、x コマンドや SYSTASK COMMAND ステートメントにより Python を外部プロセスとして起動する方法である。既存の Python プログラムをそのまま実行できる反面、データ交換には CSV、JSON、Excel、SAS データセットなどの中間ファイルを用いることが多く、ログおよび終了コードの管理も個別に設計する必要がある。

```
%* 外部 Python プログラムを実行;
systask command
  '"C:¥Python314¥python.exe" "C:¥Project¥program1.py"'
  taskname=python_task status=python_status waitfor;
```

第二の方法は、SAS 9.4M6 以降で利用可能となった PROC FCMP の Python object である。Python 関数を呼び出す SAS 関数を PROC FCMP で定義し、DATA ステップなどから利用できるため、レコード単位の値変換やモデルスコアリングに適している。一方、Python プログラム全体を実行する用途では、オブジェクトの定義、引数の受け渡し、結果取得などの記述が必要となる。これらに対して PROC PYTHON では、SAS プログラム内の SUBMIT ブロックに Python コードをまとまった形で記述できる。SAS から Python への値の受け渡し、SAS データセットの読み書き、外部 Python プログラムの呼び出しなどを一つの処理フローとして構成できるため、Excel 操作、図表作成、ファイル変換のような複数工程を含む処理に適している。

方法	実行形態	主な用途	特徴
SYSTASK	外部プロセス	既存.py の実行	幅広い環境で利用しやすいが、ファイル連携が中心
PROC FCMP	Python 関数	値変換・スコアリング	DATA ステップに組み込みやすい
PROC PYTHON	SAS 内で Python 実行	Excel・図表・データ変換	Python 処理をまとまって記述できる、ファイル連携中心であるものの外部プロセスで起動するより扱いやすい。

3. SAS と Python のデータ受け渡し

PROC PYTHON では、SAS.symget を用いて SAS マクロ変数を Python 側で取得できる。また、SAS データセットを Python の pandas DataFrame として読み込み、処理結果を SAS データセットへ書き戻すことができる。実務では、SAS 側で対象データを作成し、Python 側では Excel 出力や可視化など、得意な処理を担当させる

構成が分かりやすい。例えば SAS と通常の Python 実行環境で動作可能な Python ファイルを作成することにより、デバッグも容易になることが期待できる。

4. 実用例

4.1 Excel 連携

Python が提供する各種ライブラリを利用することで、SAS から Excel を柔軟に操作できる。従来、SAS から Excel を操作する手段としては、DDE（Dynamic Data Exchange）が広く利用されてきた。DDE は古くからある技術であるが、比較的単純な仕組みで Excel の操作とデータ転送を行うことができ、その軽量さと利便性から現在でも実務で有用な場面は多い。一方、Windows におけるアプリケーション連携技術は、その後 OLE（Object Linking and Embedding）、COM（Component Object Model）へと発展しており、DDE にはセキュリティや将来的な保守性の面で課題もある。DDE による Excel 連携では、`filename CMD DDE 'EXCEL|SYSTEM';` により Excel へコマンドを送り、`filename FF DDE 'EXCEL|sheet!range';` により指定したセル範囲へデータを書き込むことができる。このように、Excel アプリケーションの操作とセルへのデータ転送を簡潔に使い分けられることは、DDE の大きな利点の一つである。PROC PYTHON を利用すると、DDE が担ってきたこれらの機能を、目的に応じて Python のライブラリに置き換えることができる。Excel の起動、再計算、VBA マクロの実行、印刷設定など Excel アプリケーション自体の操作には、COM Automation を利用する`win32com`を使用できる。一方、セルへの値の書き込みなどのデータ操作には、Excel を起動せずに`.xlsx`ファイルを直接操作できる`openpyxl`を利用できる。このように、DDE における「Excel の操作」と「データ転送」という役割分担を Python でも維持できるため、従来の DDE プログラムを理解していれば、Python による Excel 連携にも比較的移行しやすいと考えられる。今回の症例フロー作成システム（1_Flow.sas）では、SAS で症例数を集計した後、PROC PYTHON から Python プログラムを実行し、レイアウト用 SAS データセットに定義したシート名とセル番地に従って、既存の Excel テンプレートへ集計結果を書き込む仕組みとした。`win32com`や`openpyxl`を用いることで、既存テンプレートの書式や数式を維持したまま値を書き込むことができ、集計ロジックは SAS、Excel 帳票への出力処理は Python という役割分担が可能となった。既存の DDE 資産を直ちに置き換える必要はないが、新規開発から Python による Excel 連携を取り入れることで、既存資産を活かしながら DDE への依存を段階的に減らし、より柔軟で保守しやすい構成へ移行できる。

4.2 Heatmap 作成

SAS は SGPLOT や SGPANEL などの統計グラフ作成機能を備えており、解析結果を迅速に確認するための図表を作成するうえで有用である。一方、Python には Matplotlib をはじめとする豊富な描画ライブラリがあり、レイアウト、注記、配色、画像サイズ、解像度、出力形式などを細かく制御できる。PROC PYTHON を利用すると、SAS による解析処理はそのまま維持しながら、図表作成部分だけを SGPLOT から Python へ切り替えるような使い方が可能となる。例えば、解析途中の迅速な確認には SGPLOT を使用し、最終的な Figure の作成や投稿用の清書には Matplotlib を使用するなど、目的に応じて描画手段を使い分けることができる。

今回の Heatmap 作成では、SAS で相関係数を算出し、相関行列と変数メタデータを作成した後、PROC PYTHON から Matplotlib を用いて描画した。SAS 側では解析対象変数、表示順、変数ラベルなど解析に関係する情報を管理し、Python 側では色範囲、カラーバー、軸ラベル、セル内の数値表示、画像サイズ、解像度など Figure としての表現を担当させた。Matplotlib では、`coolwarm`、`bwr`、`RdBu_r`など、Heatmap に適した多様なカ

ラーマップが標準で用意されており、Figure の目的に応じて容易に使い分けることができる。このように、PROC PYTHON による Figure 作成は SAS のグラフ機能を置き換えるものではなく、SAS の処理フローの中で描画モジュールを選択・切り替える手段を追加するものと位置付けられる。迅速な解析には SGPLOT、より細かな調整が必要な Figure 作成には Python と使い分けることで、SAS の解析資産を維持しながら図表作成の自由度を広げることができる。

4.3 CDISC Dataset-JSON の読み込み

Dataset-JSON は、臨床試験データを JSON 形式で交換するための CDISC 標準であり、従来の XPT 形式が抱える制約を補う、柔軟性・拡張性に優れた新たなデータ交換形式として、今後の普及と活用の拡大が期待されている。Python の json モジュールおよび pandas を利用すると、Dataset-JSON を読み込み、変数定義とレコードを DataFrame へ変換できる。今回の試行では、ADaM、SDTM、SEND の各フォルダにある JSON ファイルを一括処理し、SAS データセット、CSV および Excel へ出力した。ファイル名から SAS メンバー名を生成する際には、使用可能文字、先頭文字、32 文字制限など、SAS 命名規則への対応が必要であった。

5. 実行環境の構築

PROC PYTHON を利用するには、対応する SAS 9.4 環境に Hot Fix を適用し、制限オプション設定ファイルで ALLOWPROCPYTHON を有効化する必要がある。また、PROC_PYPATH に Python の実行ファイル、PROC_M2PATH に SAS と Python の連携モジュールを設定する。Python 側では、使用するライブラリとそのバージョンを requirements.txt で管理することで、別の PC においても同等の実行環境を再構築しやすくなる。今回の例では、pandas、matplotlib、pyreadstat、openpyxl および pywin32 を使用した。環境構築後は、まず簡単な Hello World プログラムを実行して PROC PYTHON が正常に動作することを確認する。その後、SAS と Python の基本的な連携方法を確認するため、いくつかの簡単なサンプルプログラムを紹介する。

1_HelloPython.sas

2_SASMACROVar.sas

3_SasDS.sas

4_ReadExcel_SJIS.sas

4_ReadExcel_UTF8.sas

6. メリットと課題

メリットは、既存の SAS 資産を維持しながら Python ライブラリを追加し、Excel 帳票、図表、JSON 変換を一つの SAS ジョブから実行できる点である。一方、Python 本体とライブラリのバージョン、文字コード、日本語ログ、例外処理などの管理が増える。開発環境と本番環境の Python 構成を一致させ、実行ログとエラーコードを保存する仕組みが重要となる。

7. 今後の展望

今後は、PDF・PowerPoint 出力、Web API、自然言語処理、機械学習への展開も考えられる。すべてを Python へ移行するのではなく、SAS のデータ処理・統計解析と、Python の外部形式連携・可視化を適切に分担することが重要である。PROC PYTHON は SAS を置き換えるのではなく、SAS を中心とした業務フローの選択肢を広げる。

8. まとめ

SAS から Python を利用する方法は、SYSTASK による外部実行、PROC FCMP Python object による関数連携を経て、PROC PYTHON による統合的な実行へ発展した。Excel 出力、Heatmap、Dataset-JSON 変換を SAS プログラムの一部として構成できることは、既存の SAS 資産を活かしながらデータ活用の可能性を広げる。

参考資料

1) Python Package Index (PyPI): https://pypi.org/

2) SAS Documentation, SYSTASK Statement: https://documentation.sas.com/

3) CDISC Dataset-JSON: https://www.cdisc.org/standards/data-exchange/dataset-json

見た目は「.」，中身は何か

― BASE SAS における欠測表現の入力挙動と情報損失の検証 ―

○小林　邦世
（イーピーエス株式会社）

What Lies Behind the Dot? — Input Behavior and Information Loss of Missing-Value Representations in BASE SAS —

Kuniyo Kobayashi
(EPS Corporation)

要旨

外部ファイルでは，欠測または値なしを示す表現として，空欄，「.」，「NA」，「N/A」，「NULL」，「-」等が混在し得る．本論文では，BASE SAS への外部データ入力時に生じる欠測表現の変換と情報損失について，固定位置テキストおよび CSV を対象に，データステップの入力形式，MISSING ステートメント，MISSING システムオプション，IMPORT プロシジャの型推測および GUESSINGROWS ステートメントの影響を検証した．その結果，入力形式によって原表記の保持状態が異なり，数値として解釈できない文字列が通常数値欠測「.」へ集約されること，同一の構成要素を持つ CSV でもレコード順序および GUESSINGROWS の設定により IMPORT プロシジャの推測型と格納結果が変化することを確認した．また，特殊欠測値は欠測として判定される一方で大小比較およびソートの対象となり，条件式によっては意図しない抽出につながることを確認した．さらに，NullFlavor コードは文字値として入力した場合には保持されたが，BEST32.による数値入力ではすべて通常数値欠測「.」へ集約された．以上より，欠測表現を取り扱う際には最終的な表示値のみでなく，原表記，入力時の解釈，内部値および入力エラーを区別して確認し，意味を持つ原表記を入力時点から追跡可能な形で保持することが重要であると考えられた．

キーワード：BASE SAS，欠測値，外部データ入力，IMPORT プロシジャ，NullFlavor

1. 背景・目的

臨床試験等で受領する外部データでは，欠測または値なしを示す表記が必ずしも統一されておらず，空欄，「.」，「NA」，「N/A」，「NULL」，「-」等が用いられることがある．SAS では通常数値欠測，文字欠測，特殊欠測が区別され，さらに入力方法や設定により，同一の外部表記であっても格納結果や表示が変化し得る．

Bost C (2014) [1]，Glanz H ら(2018)[2]によって，SAS の欠測値の種類，表示，比較および一般的な欠測コードの再符号化手法が整理されている．一方，外部ファイルの原表記が SAS 入力の各段階でどのように変換・集約され，完成した SAS データセットからどの程度復元可能かという観点は，実務上のデータ監査と密接に関係する．

本論文では「情報損失」を，異なる原表記または変換経路が入力後に識別不能となり，SAS データセットのみから元の表記・入力時の状態・意味的区別を復元できなくなることと定義する．外部ファイルの原表記，入力時の解釈，SAS 内部値，出力時の表示，比較・演算上の挙動を分離し，BASE SAS における欠測表現の入力挙動を検証することを目的とした． また，以降の検証を理解するために必要な外部データ入力と欠測値の基礎を先に整理する．

2. SAS における外部データ入力と欠測値の基礎

2.1 「$20.」,「$CHAR20.」および「BEST32.」の違い

「$20.」と「$CHAR20.」はいずれも入力値を文字値として解釈して読み込むための入力形式であり，入力値を文字型として扱う．「$20.」は標準的な文字の入力形式であるのに対し，「$CHAR20.」はフィールド内の文字をより直接的に読み込み，特に先頭空白を保持する点が異なる[3][4].

「BEST32.」は入力値を数値として解釈して読み込むための入力形式であり，入力文字列を数値として解釈する[5].

2.2 SAS における欠測値

SAS では，文字型の欠測は空白で表される．数値型では通常欠測「.」のほか，「._」および「.A」～「.Z」などの特殊欠測を使用できる．MISSING 関数は，指定した値が SAS 上で欠測値であるかを判定する関数であり，欠測値の場合は「1」，欠測でない場合は「0」を返す．通常欠測および特殊欠測はいずれも MISSING 関数では欠測として判定されるが，「value=.A」のように特定の欠測値と比較することで，通常欠測と各特殊欠測を区別できる[6].

2.3 MISSING システムオプションと MISSING ステートメント

SAS では，通常の数値欠測値は内部的には欠測値として保持されるが，画面上でどの文字を用いて表示するかは MISSING システムオプションによって変更できる．

例えば，「options missing='-';」を指定した場合，PRINT プロシジャにおいて，同じ通常欠測値「.」が「-」として表示される．MISSING システムオプションはあくまで表示方法を変更するオプションであり，SAS 内部に格納されている欠測値そのものを変更するものではない[7].

さらに，グローバルステートメントとして MISSING ステートメントがある．これは MISSING システムオプションとは役割が異なる．MISSING ステートメントは，特定の文字を数値欠測値として認識させるために用いる．例えば，データステップ内で「missing A R;」と指定すると，入力データ中の A および R を，それぞれ特殊欠測値「.A」および「.R」として扱うことができる(図 1 参照)[8].

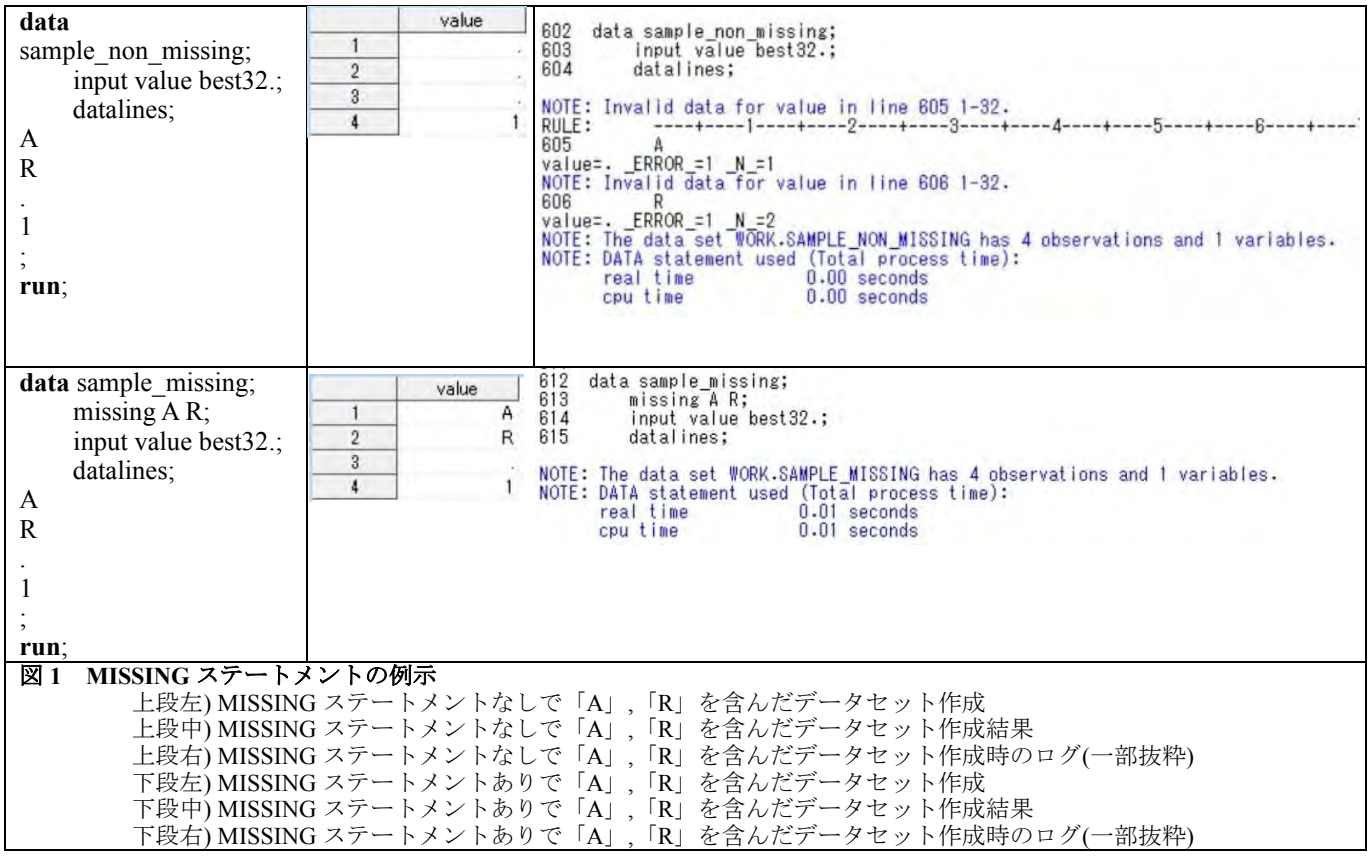


図 1　MISSING ステートメントの例示

上段左) MISSING ステートメントなしで「A」,「R」を含んだデータセット作成
上段中) MISSING ステートメントなしで「A」,「R」を含んだデータセット作成結果
上段右) MISSING ステートメントなしで「A」,「R」を含んだデータセット作成時のログ(一部抜粋)
下段左) MISSING ステートメントありで「A」,「R」を含んだデータセット作成
下段中) MISSING ステートメントありで「A」,「R」を含んだデータセット作成結果
下段右) MISSING ステートメントありで「A」,「R」を含んだデータセット作成時のログ(一部抜粋)

2.4　IMPORT プロシジャ

IMPORT プロシジャは，CSV 等の外部ファイルを SAS データセットとして読み込むためのプロシジャである．データステップによる入力では，変数型や informat 等の読み込み規則を利用者が明示的に指定できるのに対し，IMPORT プロシジャでは外部ファイルの内容をもとに，変数型，変数長等が自動的に推測される．

CSV ファイルを読み込む場合，GUESSINGROWS ステートメントによって，変数属性の推測に用いる行数を指定できる．推測対象となる行の内容によっては，同一列に数値と文字列が混在する場合などに，どの型として読み込まれるかが変化する可能性がある．IMPORT プロシジャ内で「GUESSINGROWS=MAX」を指定すると，データ内のすべての行を参照して型や変数長を推測するため，ファイル前半の値による偏った型推測を避けられる場合がある．GUESSINGROWS ステートメントでは 1～2,147,483,647 行まで指定可能であり，デフォルトの場合には最初の 20 行までが推測に用いられる[9].

2.5　NullFlavor

NullFlavor は，主たる値が null（欠測）である場合に，その理由や状態に関する追加情報を表現するための仕組みである．SDTMIG v3.3 では，Trial Summary（TS）ドメインにおいて，TSVAL が null の場合に TSVALNF を用いて NullFlavor を格納することが示されている．NullFlavor には「NI（No information）」,「INV（Invalid）」,「UNK（Unknown）」,「NASK（Not asked）」,「NA（Not applicable）」等があり，同じ欠測状態であっても，値が存在しない理由を区別して表現できる[10].

3. 方法

SAS 9.4 を用いた．各実験では SAS ログ，生成データセットおよび CONTENTS プロシジャによる変数属性

を保存した.

3.1 固定位置テキスト，CSV データ：データステップによる格納結果の差

固定位置テキストに 12 種類の値を作成した(test.txt)．この test.txt を SAS のデータステップにおいて，(1)標準的な文字入力形式「$20.」，(2)先頭空白を保持する文字入力形式「$CHAR20.」，(3)数値入力形式「BEST32.」を用いて読み込んだ．また，文字型として読み込んだ値については，画面上の表示だけでは空白や「.」の保持状態を判別できない場合があるため，MISSING 関数，LENGTHN 関数，NOTSPACE 関数および 16 進数表示(HEX 表示)を併用して格納状態を確認した(Appendix 1 参照)．MISSING 関数は SAS 上で値が欠測と判定されるかを確認するために用いた．LENGTHN 関数は末尾空白を除いた文字列長を，NOTSPACE 関数は最初の非空白文字の位置を確認するために用いた．HEX 表示は各文字を 16 進数で確認し，例えば半角空白は 20，ピリオドは 2E として識別する．これにより，画面上では同一に見える値についても，先頭空白や「.」が実際に保持されているかを確認した.

これらの検証を，固定位置テキストを CSV 形式にしたデータ(test.csv)でも同様に行った.

表 1　固定位置テキストに設定した値

Case	原表記	確認目的
1		基本的な欠測(空欄)
2	.	通常欠測表記(ピリオド)
3	.	空白の保持(前後空白付きピリオド)
4	.A	特殊欠測表記
5	NA	一般的な欠測表現
6	N/A	一般的な欠測表現
7	NULL	一般的な欠測表現
8	-	欠測記号候補
9	21.22	小数値
10	-1	負数
11	1.2E3	指数表記
12	A.B	ピリオドを含む文字列

3.2 MISSING システムオプションと MISSING ステートメントの挙動確認

MISSING システムオプションの挙動を確認するため，3.1 節と同一のデータ test.csv をデータステップを用いて「BEST32.」で読み込み，さらに読み込んだデータを，「options missing='-';」を指定したうえで，PRINT プロシジャを用いて表示した．内部値が変化していないことを確認するため，データステップ内で「MISSING(value)」，「value=.」，「value=.A」の各条件を用いて判定した(Appendix 2 参照).

また MISSING ステートメントの挙動を確認するため，「A」，「R」，「.」，「.A」，「.R」，「21.22」，「-1」で構成された固定位置テキスト(short_test.txt，表 2 参照)を用意し，データステップにおいて BEST32.で読み込む際の，MISSING ステートメントを指定しない場合と，指定した場合とで，入力結果および入力エラーの発生状況を比較した.

表 2　固定位置テキスト(short_test.txt)に設定した特殊欠測値と数値

Case	原表記	確認目的
1	A	MISSING ステートメントの影響確認
2	R	MISSING ステートメントの影響確認
3	.	基本的な欠測
4	.A	特殊欠測表記そのもの
5	.R	特殊欠測表記そのもの
6	21.22	小数値
7	-1	負数

3.3　IMPORT プロシジャの型推測

3.2 節でも使用した test.csv を IMPORT プロシジャの GUESSINGROWS ステートメントがデフォルト条件および「GUESSINGROWS=MAX」指定時の各条件でそれぞれ読み込み，読み込み結果を比較した．さらに，同一の value 変数(要素：1～20，「.」，「NA」，「NULL」，「.A」，「A.B」，「-」）を格納した numeric_first.csv(表 3 参照)と，value 変数内の順序を変更した text_first.csv(表 4 参照)を別途作成し，レコード順序と GUESSINGROWS ステートメントが型推測や格納結果に与える影響を検証した．

表 3　numeric_first.csv(一部抜粋)

Case	value
1	1
2	2
3	3
⋮	⋮
21	.
22	NA
23	NULL
24	.A
25	A.B
26	-

表 4　text _first.csv(一部抜粋)

Case	value
1	.
2	NA
3	NULL
4	.A
5	A.B
6	-
7	1
⋮	⋮
24	19
26	20

3.4　特殊欠測値の判定・比較

「._」，「.」，「.A」，「.R」，「.Z」，「-1」，「0」，「1」をデータステップで直接生成し(special_missing_test データ，図 2，Appendix 3 参照)，MISSING 関数で欠測判定を，特定の特殊欠測値との直接比較，SORT プロシジャで value 変数内の要素の大小関係を確認した．さらに，special_missing_test から「if value < 0;」で抽出される値と，「if not missing(value) and value < 0;」により欠測値を除外した場合に抽出される値を比較した．

	label	id	value
1	._	1	._
2	.	2	.
3	.A	3	A
4	.R	4	R
5	Z	5	Z
6	-1	6	-1
7	0	7	0
8	1	8	1

図 2　データステップで直接生成した特殊欠測値を含んだデータセット(special_missing_test データ)

SAS では，「.I」および「.M」 は，「.A」～「.Z」に含まれる特殊欠測値である．DIVIDE 関数は，除算結果が正の無限大または負の無限大に相当する場合に，それぞれ「.I」または 「.M」を返す[11]．なお，SAS 上では「.I」および「.M」も欠測値として扱われ，MISSING 関数では欠測と判定される．

本論文では，通常の式「1/0」と DIVIDE 関数（「DIVIDE(-1,0)」，「DIVIDE(1,0)」）の計算結果を比較し，「.I」および「.M」について欠測判定を行う．また，「.I」，「.M」，「.Z」，「-1」，「0」，「1」の 6 種の要素をもつデータセットを作成し，SORT プロシジャを用いてこの要素の並び順を確認した．

3.5　NullFlavor の読み込み

CDISC SDTMIG v3.3 では，主たる値が null である場合に，その理由に関する追加情報を表現する仕組みとして NullFlavor が示されている．

本論文では，SDTMIG v3.3 の NullFlavor Enumeration に掲載された 15 種類，すなわち「NI」，「INV」，「OTH」，「PINF」，「NINF」，「UNC」，「DER」，「UNK」，「ASKU」，「NAV」，「NASK」，「QS」，「TRC」，「MSK」，「NA」を対象とした．

15 種類の NullFlavor を value 変数に格納した CSV ファイル（null_flavor.csv，表 5 参照）を作成し，まずデータステップを用いて，(1)標準的な文字入力形式「$20.」，(2)先頭空白を保持する文字入力形式「$CHAR20.」，(3)数値入力形式「BEST32.」の 3 条件で読み込んだ．文字入力条件では，MISSING 関数，LENGTHN 関数，NOTSPACE 関数および HEX 表示を用いて原表記の保持状態を確認した．数値入力条件では，入力時エラーの発生状況および MISSING 関数による欠測判定を確認した．

さらに，同一の null_flavor.csv を IMPORT プロシジャで読み込み，NullFlavor の原表記が保持されるかを確認した．本データは 15 オブザベーションであり，GUESSINGROWS ステートメントのデフォルト条件における推測範囲内に全観測が含まれるため，追加条件として GUESSINGROWS=5 を指定し，型推測に用いる観測数を縮小した場合についても読み込み結果を比較した．

表 5　null_flavor.csv に設定した NullFlavor

Case	value
1	NI
2	INV
3	OTH
4	PINF
5	NINF
6	UNC

Case	value
7	DER
8	UNK
9	ASKU
10	NAV
11	NASK
12	QS
13	TRC
14	MSK
15	NA

4. 結果

4.1 固定位置テキスト，CSV データ：データステップによる格納結果の差

固定位置テキスト(test.txt)をデータステップで読み込んだ結果を図 3～6 に示す.

Case	原表記	$20.	SAS上で欠測か	最後の非空白文字までの長さ	最初の非空白文字が何文字目か	HEX表示
1			1	0	0	20202020202020202020202020202020202020
2	.		1	0	0	20202020202020202020202020202020202020
3	.		1	0	0	20202020202020202020202020202020202020
4	.A	.A	0	2	1	2E41202020202020202020202020202020202020
5	NA	NA	0	2	1	4E41202020202020202020202020202020202020
6	N/A	N/A	0	3	1	4E2F41202020202020202020202020202020202020
7	NULL	NULL	0	4	1	4E554C4C2020202020202020202020202020202020
8	-	-	0	1	1	2D2020202020202020202020202020202020202020
9	21.22	21.22	0	5	1	32312E32322020202020202020202020202020202020
10	-1	-1	0	2	1	2D31202020202020202020202020202020202020
11	1.2E3	1.2E3	0	5	1	312E32453320202020202020202020202020202020
12	A.B	A.B	0	3	1	412E42202020202020202020202020202020202020

図 3　$20.で test.txt を読み込んだ結果
1)SAS 上で欠測か:「0:いいえ」,「1:はい」

Case	原表記	$CHAR20.	SAS上で欠測か	最後の非空白文字までの長さ	最初の非空白文字が何文字目か	HEX表示
1			1	0	0	20202020202020202020202020202020202020
2	.	.	0	1	1	2E20202020202020202020202020202020202020
3	.	.	0	2	2	202E20202020202020202020202020202020202020
4	.A	.A	0	2	1	2E41202020202020202020202020202020202020
5	NA	NA	0	2	1	4E41202020202020202020202020202020202020
6	N/A	N/A	0	3	1	4E2F41202020202020202020202020202020202020
7	NULL	NULL	0	4	1	4E554C4C2020202020202020202020202020202020
8	-	-	0	1	1	2D2020202020202020202020202020202020202020
9	21.22	21.22	0	5	1	32312E32322020202020202020202020202020202020
10	-1	-1	0	2	1	2D31202020202020202020202020202020202020
11	1.2E3	1.2E3	0	5	1	312E32453320202020202020202020202020202020
12	A.B	A.B	0	3	1	412E42202020202020202020202020202020202020

図 4　$CHAR20.で test.txt を読み込んだ結果
1)SAS 上で欠測か:「0:いいえ」,「1:はい」

Case	原表記	BEST32.	入力時エラー	SAS上で欠測か	通常の数値欠測値「.」そのものか	特殊欠測か
1		.	0	1	1	0
2	.	.	0	1	1	0
3	.	.	0	1	1	0
4	.A	A	0	1	0	1
5	NA	.	1	1	1	0
6	N/A	.	1	1	1	0
7	NULL	.	1	1	1	0
8	-	.	1	1	1	0
9	21.22	21.22	0	0	0	0
10	-1	-1	0	0	0	0
11	1.2E3	1200	0	0	0	0
12	A.B	.	1	1	1	0

図 5　BEST32.で test.txt を読み込んだ結果

1)SAS 上で欠測か:「0:いいえ」,「1:はい」
2)入力時エラー:「0:なし」,「1:あり」
3)通常の数値欠測値「.」そのものか:「0:いいえ」,「1:はい」
4)特殊欠測「.A」か:「0:いいえ」,「1:はい」

```
NOTE: Invalid data for value in line 5 5-6.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0
5         005 NA 6
case_id=5 value=. method=BEST32. _ERROR_=1 _N_=5
NOTE: Invalid data for value in line 6 5-7.
6         006 N/A 7
case_id=6 value=. method=BEST32. _ERROR_=1 _N_=6
NOTE: Invalid data for value in line 7 5-8.
7         007 NULL 8
case_id=7 value=. method=BEST32. _ERROR_=1 _N_=7
NOTE: Invalid data for value in line 8 5-5.
8         008 - 5
case_id=8 value=. method=BEST32. _ERROR_=1 _N_=8
NOTE: Invalid data for value in line 12 5-7.
12        012 A.B 7
case_id=12 value=. method=BEST32. _ERROR_=1 _N_=12
NOTE: 12 records were read from the infile
```

図 6　BEST32.で固定位置テキストを読み込んだ際のログ(一部抜粋)

$20.では空欄,「.」,前後空白付き「.」がすべて欠測となった．一方,$CHAR20.では「.」が保持され,前後空白付き「.」では HEX の先頭が「202E」となり,先頭空白とピリオドが保持された.

BEST32.では「.」および前後空白付き「.」は正常な通常数値欠測,「.A」は特殊欠測として入力された．一方,「NA」,「N/A」,「NULL」,「-」,「A.B」は Invalid data となり,ログに NOTE が出力され,正常な欠測入力と変換エラーの結果が同一の「.」へ集約される結果となった.

次に,固定位置テキストを CSV 形式にしたデータ(test.csv)をデータステップで読み込んだ結果を図 7～10 に示す.

Case	原表記	$20.	SAS上で欠測か	最後の非空白文字までの長さ	最初の非空白文字が何文字目か	HEX表示
1			1	0	0	20202020202020202020202020202020202020
2	.		1	0	0	20202020202020202020202020202020202020
3	.		1	0	0	20202020202020202020202020202020202020
4	.A	.A	0	2	1	2E41202020202020202020202020202020202020
5	NA	NA	0	2	1	4E41202020202020202020202020202020202020
6	N/A	N/A	0	3	1	4E2F41202020202020202020202020202020202020
7	NULL	NULL	0	4	1	4E554C4C2020202020202020202020202020202020
8	-	-	0	1	1	2D2020202020202020202020202020202020202020
9	21.22	21.22	0	5	1	32312E3232202020202020202020202020202020
10	-1	-1	0	2	1	2D3120202020202020202020202020202020202020
11	1.2E3	1.2E3	0	5	1	312E3245332020202020202020202020202020
12	A.B	A.B	0	3	1	412E42202020202020202020202020202020202020

図 7　$20.で test.csv を読み込んだ結果

1)SAS 上で欠測か:「0:いいえ」,「1:はい」

Case	原表記	$CHAR20.	SAS上で欠測か	最後の非空白文字までの長さ	最初の非空白文字が何文字目か	HEX表示
1			1	0	0	20202020202020202020202020202020202020
2	.	.	0	1	1	2E20202020202020202020202020202020202020
3	.	.	0	2	2	202E20202020202020202020202020202020202020
4	.A	.A	0	2	1	2E41202020202020202020202020202020202020
5	NA	NA	0	2	1	4E41202020202020202020202020202020202020
6	N/A	N/A	0	3	1	4E2F41202020202020202020202020202020202020
7	NULL	NULL	0	4	1	4E554C4C2020202020202020202020202020202020
8	-	-	0	1	1	2D2020202020202020202020202020202020202020
9	21.22	21.22	0	5	1	32312E3232202020202020202020202020202020
10	-1	-1	0	2	1	2D3120202020202020202020202020202020202020
11	1.2E3	1.2E3	0	5	1	312E3245332020202020202020202020202020
12	A.B	A.B	0	3	1	412E42202020202020202020202020202020202020

図 8　$CHAR20.で test.csv を読み込んだ結果

1)SAS 上で欠測か:「0:いいえ」,「1:はい」

Case	原表記	BEST32.	入力時エラー	SAS上で欠測か	通常の数値欠測値「.」そのものか	特殊欠測か
1		.	0	1	1	0
2	.	.	0	1	1	0
3	.	.	0	1	1	0
4	.A	A	0	1	0	1
5	NA	.	1	1	1	0
6	N/A	.	1	1	1	0
7	NULL	.	1	1	1	0
8	-	.	1	1	1	0
9	21.22	21.22	0	0	0	0
10	-1	-1	0	0	0	0
11	1.2E3	1200	0	0	0	0
12	A.B	.	1	1	1	0

図 9　BEST32.で test.csv を読み込んだ結果

1)SAS 上で欠測か:「0:いいえ」,「1:はい」
2)入力時エラー:「0:なし」,「1:あり」
3)通常の数値欠測値「.」そのものか:「0:いいえ」,「1:はい」
4)特殊欠測「.A」か:「0:いいえ」,「1:はい」

```
NOTE: Invalid data for value in line 6 3-4.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0
6         5,NA 4
method=BEST32. case_id=5 value=. input_error=1 _ERROR_=1 _N_=5
NOTE: Invalid data for value in line 7 3-5.
7         6,N/A 5
method=BEST32. case_id=6 value=. input_error=1 _ERROR_=1 _N_=6
NOTE: Invalid data for value in line 8 3-6.
8         7,NULL 6
method=BEST32. case_id=7 value=. input_error=1 _ERROR_=1 _N_=7
NOTE: Invalid data for value in line 9 3-3.
9         8,- 3
method=BEST32. case_id=8 value=. input_error=1 _ERROR_=1 _N_=8
NOTE: Invalid data for value in line 13 4-6.
13        12,A.B 6
method=BEST32. case_id=12 value=. input_error=1 _ERROR_=1 _N_=12
NOTE: 12 records were read from the infile
```

図 10　BEST32.で test.csv を読み込んだ際のログ(一部抜粋)

ファイル形式が txt と csv で読み込み結果に差は見られなかった.

SAS の文字変数は固定長であり，格納時には変数長に満たない部分が空白で補われる．このため，原ファイル中の末尾空白と，SAS が格納時に補った空白とを入力後の値のみから区別することはできないと考えられる.

4.2　MISSING システムオプションと MISSING ステートメントの挙動確認

MISSING システムオプションの挙動を確認するため，test.csv をデータステップを用いて BEST32.で読み込み，MISSING システムオプションで「OPTIONS MISSING='-'」に変更した上で PRINT プロシジャを使用した結果を図 11 に示す.

Case	原表記	BEST32.	入力時エラー	SAS上で欠測か	通常の数値欠測値「.」そのものか	特殊欠測か
1		-	0	1	1	0
2	.	-	0	1	1	0
3	.	-	0	1	1	0
4	.A	A	0	1	0	1
5	NA	-	0	1	1	0
6	N/A	-	0	1	1	0
7	NULL	-	0	1	1	0
8	-	-	0	1	1	0
9	21.22	21.22	0	0	0	0
10	-1	-1	0	0	0	0
11	1.2E3	1200	0	0	0	0
12	A.B	-	0	1	1	0

図 11　「OPTIONS MISSING='-'」を用いて BEST32.で読み込んだ test.csv を PRINT プロシジャで表示した結果
1)入力時エラー:「0:なし」,「1:あり」
2)SAS 上で欠測か:「0:いいえ」,「1:はい」
3)通常の数値欠測値「.」そのものか:「0:いいえ」,「1:はい」
4)特殊欠測「.A」か:「0:いいえ」,「1:はい」

通常数値欠測「.」は「-」と表示されたが，特殊欠測「.A」や文字変数の欠測や文字列内の「.」「-」には同様の表示変更は認められなかった.

また，原表記の「-」について，BEST32.で読み込んだ際には，Invalid data として扱われ SAS 上での表示は「.」となっていたが，「OPTIONS MISSING='-'」を指定すると，表示は「-」となった．見た目だけでは原表記が保持されたか判別できなかったものの，MISSING 関数により欠測として判定されたままでいることが明らかになった．このことから MISSING システムオプションは表示のみを変更し，内部値には関与しないことが確認できた.

次に，MISSING ステートメントの挙動を確認するため，「A」,「R」,「.」,「.A」,「.R」,「21.22」,「-1」で構

成された固定位置テキスト(short_test.txt)をデータステップで読み込む際，MISSING ステートメントを指定しない場合と，指定した場合とで，入力結果および入力エラーの発生状況を図 12，13 に示す．

Case	原表記	BEST32.	入力時エラー	SAS上で欠測か	通常の数値欠測値「.」そのものか	特殊欠測か
1	A	.	1	1	1	0
2	R	.	1	1	1	0
3	.	.	0	1	1	0
4	.A	A	0	1	0	1
5	.R	R	0	1	0	1
6	21.22	21.22	0	0	0	0
7	-1	-1	0	0	0	0

図 12　MISSING ステートメント未指定時の short_test.txt を読み込んだ結果
1)入力時エラー:「0:なし」,「1:あり」
2)SAS 上で欠測か:「0:いいえ」,「1:はい」
3)通常の数値欠測値「.」そのものか:「0:いいえ」,「1:はい」
4)特殊欠測「.A」もしくは「.R」か:「0:いいえ」,「1:はい」

Case	原表記	BEST32.	入力時エラー	SAS上で欠測か	通常の数値欠測値「.」そのものか	特殊欠測か
1	A	A	0	1	0	1
2	R	R	0	1	0	1
3	.	.	0	1	1	0
4	.A	A	0	1	0	1
5	.R	R	0	1	0	1
6	21.22	21.22	0	0	0	0
7	-1	-1	0	0	0	0

図 13　MISSING ステートメント指定時の short_test.txt を読み込んだ結果
1)入力時エラー:「0:なし」,「1:あり」
2)SAS 上で欠測か:「0:いいえ」,「1:はい」
3)通常の数値欠測値「.」そのものか:「0:いいえ」,「1:はい」
4)特殊欠測「.A」もしくは「.R」か:「0:いいえ」,「1:はい」

MISSING ステートメント未指定時には，原表記 A および R は Invalid data となり，通常数値欠測「.」へ変換された．一方，原表記「.A」および「.R」は MISSING ステートメントを指定しなくても，それぞれ特殊欠測値「.A」および「.R」として正常に入力された．「MISSING A R;」を指定すると，原表記「A」および「R」もそれぞれ「.A」および「.R」として入力され，入力時エラーは発生しなかった．

4.3　IMPORT プロシジャの型推測

4.1 節でも利用した test.csv を，GUESSINGROWS ステートメントをデフォルト，「GUESSINGROWS=MAX」でそれぞれ IMPORT プロシジャを用いて読み込んだ結果を図 14，15 に示す．

Case	原表記	デフォルト	GUESSINGROWS=MAX
1			
2	.		
3	.		
4	.A	.A	.A
5	NA	NA	NA
6	N/A	N/A	N/A
7	NULL	NULL	NULL
8	-	-	-
9	21.22	21.22	21.22
10	-1	-1	-1
11	1.2E3	1.2E3	1.2E3
12	A.B	A.B	A.B

図 14　IMPORT プロシジャを用いて test.csv を読み込んだ結果

条件	TYPE	LENGTH
デフォルト	character	5
GUESSINGROWS=MAX	character	5

図 15　GUESSINGROWS ステートメントがデフォルトの場合と「MAX」を指定した場合の test.csv の IMPORT プロシジャを利用した読み込み結果

GUESSINGROWS ステートメントがデフォルト条件下での IMPORT プロシジャは VALUE を長さ 5 の文字型と推測した．デフォルト条件と「GUESSINGROWS=MAX」で読み込み結果，型(Type)，長さ(Length)に差は認められなかった．また，単独の「.」および空白を伴う「.」はいずれも文字欠測へ集約された．

ここで GUESSINGROWS ステートメントの詳細な挙動を検証した．

numeric_first.csv と text_first.csv は VALUE 変数の構成要素が同一で，そのレコード順のみを変更したファイルである．GUESSINGROWS ステートメントがデフォルト条件と「GUESSINGROWS=MAX」条件で IMPORT プロシジャを用いて読み込んだ結果を一部抜粋して図 16～19 に示す．

Case	原表記	デフォルト	GUESSINGROWS=MAX
20	20	20	20
21	.	.	
22	NA	.	NA
23	NULL	.	NULL
24	.A	A	.A
25	A.B	.	A.B
26	-	.	-

図 16
numeric_first.csv をデフォルトと「GUESSINGROWS=MAX」を用いて IMPORT プロシジャを用いてそれぞれ読み込んだ結果(case20～26 の一部抜粋)

Case	原表記	デフォルト	GUESSINGROWS=MAX
1	.		
2	NA	NA	NA
3	NULL	NULL	NULL
4	.A	.A	.A
5	A.B	A.B	A.B
6	-	-	-
7	1	1	1

図 17
text_first.csv をデフォルトと「GUESSINGROWS=MAX」を用いて IMPORT プロシジャを用いてそれぞれ読み込んだ結果(case1～7 の一部抜粋)

入力ファイル	GUESSINGROWS	Type	Length
Numeric first	デフォルト	Numeric	8
Numeric first	MAX	Character	4
Text first	デフォルト	Character	4
Text first	MAX	Character	4

図 18
numeric_first.csv と text_first.csv の読み込み結果まとめ

```
NOTE: Invalid data for VALUE in line 23 4-5.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+----8----+----9----+----0
23        22,NA 5
CASE_ID=22 VALUE=. _ERROR_=1 _N_=22
NOTE: Invalid data for VALUE in line 24 4-7.
24        23,NULL 7
CASE_ID=23 VALUE=. _ERROR_=1 _N_=23
NOTE: Invalid data for VALUE in line 26 4-6.
26        25,A.B 6
CASE_ID=25 VALUE=. _ERROR_=1 _N_=25
NOTE: Invalid data for VALUE in line 27 4-4.
27        26,- 4
CASE_ID=26 VALUE=.  ERROR_=1  N_=26
```

図 19
numeric_first.csv を GUESSINGROWS ステートメントがデフォルト条件で読み込んだ際のログ(一部抜粋)

GUESSINGROWS ステートメントがデフォルト条件での IMPORT プロシジャでは value 変数が numeric_first.csv で数値型，text_first.csv では文字型で読み込まれた．

value 変数内のレコード順の前半に数値が並べられた numeric_first.csv において，GUESSINGROWS ステートメントがデフォルト条件では，value 変数は数値型で読み込まれ，レコード順が後半の「NA」，「NULL」，「A.B」，「-」は Invalid data となり通常欠測「.」へ集約された．一方，「GUESSINGROWS=MAX」を利用した場合には value 変数が文字型で読み込まれ，「NA」，「NULL」，「A.B」，「-」の文字表記は保持された．

value 変数内のレコード順の前半に文字値が並べられた text_first.csv においては，GUESSINGROWS ステートメントがデフォルト条件でも「GUESSINGROWS=MAX」条件でも，value 変数は文字型で読み込まれ，「.」，「NA」，「NULL」，「A.B」，「-」の文字表記が保持された．

したがって，同一の構成要素でもレコード順と型推測に用いるレコード範囲によって，変数型，原表記の保持，入力エラーの発生状況が変化することが分かった．

4.4　特殊欠測値の判定・比較

「._」，「.」，「.A」，「.R」，「.Z」，「-1」，「0」，「1」を含む数値変数「value」をもつデータセット(special_missing_test)をデータステップで作成し，SAS 上での表記と通常欠測か特殊欠測かを判定した結果を図 20 に示す．

Case	原表記	value	欠測か	「.」	「.A」	「.R」	「.Z」	「._」
1	._	_	1	0	0	0	0	1
2	.	.	1	1	0	0	0	0
3	.A	A	1	0	1	0	0	0
4	.R	R	1	0	0	1	0	0
5	.Z	Z	1	0	0	0	1	0
6	-1	-1	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	1	1	0	0	0	0	0	0

図 20　special_missing_test の通常欠測と特殊欠測の判定結果
1)欠測か:「0:いいえ」，「1:はい」
2「.」(通常の数値欠測値「.」か):「0:いいえ」，「1:はい」
3)「.A」(特殊欠測「.A」か):「0:いいえ」，「1:はい」
4)「.R」(特殊欠測「.R」か):「0:いいえ」，「1:はい」
5)「.Z」(特殊欠測「.Z」か):「0:いいえ」，「1:はい」
6)「._」(特殊欠測「._」か):「0:いいえ」，「1:はい」

「._」,「.」,「.A」,「.R」,「.Z」はいずれも欠測と判定された一方，通常欠測と特殊欠測は「value=.」,「value=._」,

「value=.A」等の特定の欠測値と直接比較することで個別に識別可能であった.

また，同データセットの value 変数を SORT プロシジャを用いて並べ替えた結果を図 21 に示す.

Case	原表記	value
1	._	_
2	.	.
3	.A	A
4	.R	R
5	.Z	Z
6	-1	-1
7	0	0
8	1	1

図 21　special_missing_test の通常欠測の並べ替え結果

「._」，「.」，「.A」，「.R」，「.Z」，「-1」，「0」，「1」　の順に並んだ.

また，special_missing_test の value 変数の構成要素から「0 より小さい value (if value < **0)**」のみを抽出した場合と，「value が欠測値でなく，かつ，0 より小さい value (if not missing(value) and value < **0**)」のみを抽出した場合の結果を図 22，23 に示す.

Case	原表記	value
1	._	_
2	.	.
3	.A	A
4	.R	R
5	.Z	Z
6	-1	-1

図 22　special_missing_test 内の VALUE 変数から 0 より小さい値のみを出力した結果

Case	原表記	value
6	-1	-1

図 23　special_missing_test 内の VALUE 変数から，欠測値でなく，かつ，0 より小さい値のみを出力した結果

「0 より小さい value (if value < **0)**」のみの構成要素を抽出した場合，「 ._」，「.」，「.A」，「.R」，「.Z」，「-1」の 6 件が抽出された．一方，「value が欠測値でなく，かつ，0 より小さい value (if not missing(value) and value < **0**)」の構成要素を出力した場合には，「-1」のみが抽出された．これらの結果から，欠測値を明示的に除外しない大小条件は，実務上意図しないレコードを抽出する可能性があることが明らかになった.

4.5　DIVIDE 関数，特殊欠測値

通常の式「1/0」と「DIVIDE(1,0)」と「DIVIDE(-1,0)」の計算結果を図 24，25 に示す.

```
16326  data infinity_data;
16327       /* 通常の除算 */
16328       normal_div = 1 / 0;
NOTE: Division by zero detected during the compilation phase, detected at line 16328 column 20.
16329
16330       /* DIVIDE関数 */
16331       divide_pos = divide(1, 0);
16332       divide_neg = divide(-1, 0);
16333  run;

NOTE: The data set WORK.INFINITY_DATA has 1 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

図 24　通常の式「1/0」と「DIVIDE(1,0)」と「DIVIDE(-1,0)」の計算結果のログ

method	result	欠測か	特殊欠測との一致「.I」	特殊欠測との一致「.M」
1/0	.	1	0	0
divide(1, 0)	I	1	1	0
divide(-1, 0)	M	1	0	1

図 25　通常の式「1/0」と「DIVIDE(1,0)」と「DIVIDE(-1,0)」の計算結果
1)欠測か(MISSING 関数による欠測判定):「0:いいえ」,「1:はい」
2)特殊欠測との一致「.I」:「0:いいえ」,「1:はい」
3)特殊欠測との一致「.M」:「0:いいえ」,「1:はい」

通常の式「1/0」では通常欠測「.」が得られ，ログに「Division by zero」の NOTE が出力された.「DIVIDE(1,0)」は「.I」,「DIVIDE(-1,0)」は「.M」を返した．一方，データステップを用いた比較では，いずれも MISSING 関数では欠測と判定された.

また，特殊欠測値の大小関係を確認するために「.I」,「.M」,「.Z」,「-1」,「0」,「1」を SORT プロシジャを用いて並べ替えた結果を図 26 に示す.

原表記	value
.I	I
.M	M
.Z	Z
-1	-1
0	0
1	1

図 26　「.I」,「.M」,「.Z」,「-1」,「0」,「1」の並べ替え結果

「.I」,「.M」,「.Z」,「-1」,「0」,「1」の順に並んだ．本検証から，「.I」と「.M」は Infinity 概念と関係する特殊欠測値である一方，データステップ上では SAS の欠測値として欠測判定されることが確認された.

4.6　NullFlavor 入力時情報保持性

SDTMIG v3.3 の NullFlavor Enumeration に掲載された 15 種類のコード「NI」,「INV」,「OTH」,「PINF」,「NINF」,「UNC」,「DER」,「UNK」,「ASKU」,「NAV」,「NASK」,「QS」,「TRC」,「MSK」,「NA」を格納した null_flavor.csv を, データステップおよび IMPORT プロシジャを用いて読み込んだ結果を図 27～30 に示す.

Case	原表記	$20.	SAS上で欠測か	最後の非空白文字までの長さ	最初の非空白文字が何文字目か	HEX表示
1	NI	NI	0	2	1	4E492020202020202020202020202020202020
2	INV	INV	0	3	1	494E562020202020202020202020202020202020
3	OTH	OTH	0	3	1	4F544820202020202020202020202020202020
4	PINF	PINF	0	4	1	50494E46202020202020202020202020202020
5	NINF	NINF	0	4	1	4E494E4620202020202020202020202020202020
6	UNC	UNC	0	3	1	554E432020202020202020202020202020202020
7	DER	DER	0	3	1	44455220202020202020202020202020202020
8	UNK	UNK	0	3	1	554E4B2020202020202020202020202020202020
9	ASKU	ASKU	0	4	1	41534B5520202020202020202020202020202020
10	NAV	NAV	0	3	1	4E41562020202020202020202020202020202020
11	NASK	NASK	0	4	1	4E41534B202020202020202020202020202020
12	QS	QS	0	2	1	5153202020202020202020202020202020202020
13	TRC	TRC	0	3	1	5452432020202020202020202020202020202020
14	MSK	MSK	0	3	1	4D534B20202020202020202020202020202020

020

図 27　$20.で null_flavor.csv を読み込んだ結果
1)SAS 上で欠測か:「0:いいえ」,「1:はい」

Case	原表記	$CHAR20.	SAS上で欠測か	最後の非空白文字までの長さ	最初の非空白文字が何文字目か	HEX表示
1	NI	NI	0	2	1	4E4920202020202020202020202020202020202020
2	INV	INV	0	3	1	494E5620202020202020202020202020202020202020
3	OTH	OTH	0	3	1	4F5448202020202020202020202020202020202020
4	PINF	PINF	0	4	1	50494E4620202020202020202020202020202020
5	NINF	NINF	0	4	1	4E494E4620202020202020202020202020202020
6	UNC	UNC	0	3	1	554E4320202020202020202020202020202020202020
7	DER	DER	0	3	1	4445522020202020202020202020202020202020
8	UNK	UNK	0	3	1	554E4B2020202020202020202020202020202020
9	ASKU	ASKU	0	4	1	41534B5520202020202020202020202020202020
10	NAV	NAV	0	3	1	4E4156202020202020202020202020202020202020
11	NASK	NASK	0	4	1	4E41534B2020202020202020202020202020202020
12	QS	QS	0	2	1	5153202020202020202020202020202020202020
13	TRC	TRC	0	3	1	5452432020202020202020202020202020202020
14	MSK	MSK	0	3	1	4D534B2020202020202020202020202020202020
15	NA	NA	0	2	1	4E4120202020202020202020202020202020202020

図 28　$CHAR20.で null_flavor.csv を読み込んだ結果

1)SAS 上で欠測か:「0:いいえ」,「1:はい」

Case	原表記	BEST32.	入力時エラー	SAS上で欠測か	通常の数値欠測値「.」そのものか
1	NI	.	1	1	1
2	INV	.	1	1	1
3	OTH	.	1	1	1
4	PINF	.	1	1	1
5	NINF	.	1	1	1
6	UNC	.	1	1	1
7	DER	.	1	1	1
8	UNK	.	1	1	1
9	ASKU	.	1	1	1
10	NAV	.	1	1	1
11	NASK	.	1	1	1
12	QS	.	1	1	1
13	TRC	.	1	1	1
14	MSK	.	1	1	1
15	NA	.	1	1	1

図 29　BEST32.で null_flavor.csv を読み込んだ結果

1)SAS 上で欠測か:「0:いいえ」,「1:はい」
2)入力時エラー:「0:なし」,「1:あり」
3)通常の数値欠測値「.」そのものか:「0:いいえ」,「1:はい」

```
NOTE: Invalid data for value in line 2 3-4.
RULE:     ----+----1----+----2----+----3----+----4----+----5----+----6----+----7----+
2         1,NI 4
method=BEST32. case_id=1 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=1
NOTE: Invalid data for value in line 3 3-5.
3         2,INV 5
method=BEST32. case_id=2 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=2
NOTE: Invalid data for value in line 4 3-5.
4         3,OTH 5
method=BEST32. case_id=3 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=3
NOTE: Invalid data for value in line 5 3-6.
5         4,PINF 6
method=BEST32. case_id=4 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=4
NOTE: Invalid data for value in line 6 3-6.
6         5,NINF 6
method=BEST32. case_id=5 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=5
NOTE: Invalid data for value in line 7 3-5.
7         6,UNC 5
method=BEST32. case_id=6 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=6
NOTE: Invalid data for value in line 8 3-5.
8         7,DER 5
method=BEST32. case_id=7 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=7
NOTE: Invalid data for value in line 9 3-5.
9         8,UNK 5
method=BEST32. case_id=8 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=8
NOTE: Invalid data for value in line 10 3-6.
10        9,ASKU 6
method=BEST32. case_id=9 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=9
NOTE: Invalid data for value in line 11 4-6.
11        10,NAV 6
method=BEST32. case_id=10 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=10
NOTE: Invalid data for value in line 12 4-7.
12        11,NASK 7
method=BEST32. case_id=11 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=11
NOTE: Invalid data for value in line 13 4-5.
13        12,QS 5
method=BEST32. case_id=12 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=12
NOTE: Invalid data for value in line 14 4-6.
14        13,TRC 6
method=BEST32. case_id=13 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=13
NOTE: Invalid data for value in line 15 4-6.
15        14,MSK 6
method=BEST32. case_id=14 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=14
NOTE: Invalid data for value in line 16 4-5.
16        15,NA 5
method=BEST32. case_id=15 value=. input_error=1 missing_flag=1 _ERROR_=1 _N_=15
```

図 30　BEST32.で null_flavor.csv を読み込んだ際のログ(一部抜粋)

データステップにおいて， $20.および$CHAR20.を用いて読み込んだ場合，15 種類すべてについて原表記が保持された．いずれの条件でも MISSING 関数による欠測判定は 0 であり，文字列長および HEX 表示からも，入力された NullFlavor コードが文字列として保持されていることを確認した．今回使用した NullFlavor には単独の「.」や先頭空白を含む値が存在しないため，$20.と$CHAR20.の読み込み結果に差は認められなかった．

一方，数値入力形式 BEST32.を用いて同一ファイルを読み込んだ場合，15 種類すべての NullFlavor コードで Invalid data が発生し，VALUE は通常数値欠測「.」へ変換された．全観測で入力時エラーフラグは 1，MISSING 関数による欠測判定は 1，通常数値欠測「.」との比較結果も 1 となった．したがって，BEST32.による数値入力後の VALUE のみから，元の NullFlavor を識別することはできなかった．

次に，同一の null_flavor.csv を IMPORT プロシジャで読み込んだ結果を図 31～32 に示す．

Case	原表記	デフォルト	GUESSINGROWS=5
1	NI	NI	NI
2	INV	INV	INV
3	OTH	OTH	OTH
4	PINF	PINF	PINF
5	NINF	NINF	NINF
6	UNC	UNC	UNC
7	DER	DER	DER
8	UNK	UNK	UNK
9	ASKU	ASKU	ASKU
10	NAV	NAV	NAV
11	NASK	NASK	NASK
12	QS	QS	QS
13	TRC	TRC	TRC
14	MSK	MSK	MSK
15	NA	NA	NA

図 31　IMPORT プロシジャを用いて null_flavor.csv を読み込んだ結果

条件	TYPE	LENGTH
デフォルト	character	4
GUESSINGROWS=5	character	4

図 32　GUESSINGROWS ステートメントがデフォルトの場合と「5」を指定した場合の null_flavor.csv の IMPORT プロシジャを利用した読み込み結果

GUESSINGROWS ステートメントがデフォルトの条件および「GUESSINGROWS=5」を指定した条件のいずれにおいても，15 種類すべての NullFlavor について原表記が保持され，両条件間で格納結果に差は認められなかった．本データでは先頭 5 行の時点で文字列が含まれているため，推測に用いる観測数を 5 まで縮小した場合でも，NullFlavor を文字列として保持する結果となった．

以上より，本検証条件では，NullFlavor を文字値として入力した場合および IMPORT プロシジャが文字として取り込んだ場合には原表記が保持された一方，BEST32.を明示して数値として解釈した場合には，意味の異なる 15 種類のコードがすべて同一の通常数値欠測「.」へ集約された．

4.7　小括

各検証条件における情報の保持状況と，本検証条件で識別不能となった情報を表 6 にまとめる．

表 6　各検証条件における情報の保持と本検証条件で識別不能となった情報

検証条件	入力後に保持された情報	本検証条件で識別不能となった情報
$20.による文字入力	「NA」,「N/A」,「NULL」等の文字列表記	空欄，単独の「.」，空白を伴う「.」の区別
$CHAR20.による文字入力	単独の「.」，先頭空白を含む「 .」	原ファイルの末尾空白と SAS による補填空白の区別
BEST32.による数値入力	数値，通常数値欠測および特殊欠測	数値変換エラーとなった場合の各原表記の区別
MISSING システムオプション	SAS 内部に格納された欠測値	表示値のみから元の入力値や変換経路を判断すること
MISSING ステートメント	「A」,「R」に対応する特殊欠測値	原表記が「A」か「.A」か,「R」か「.R」かの区別
IMPORT プロシジャ	推測された型に適合する値	推測型と不整合な後続値の原表記
特殊欠測値	直接比較による特殊欠測値間の区別	MISSING 関数だけでは個々の欠測種類を区別できない
NullFlavor の文字入力	各 NullFlavor コード	—
NullFlavor の数値入力	数値変換エラーの発生	各 NullFlavor コードの原表記および意味的区別

注：—は，本検証条件で原表記の識別不能が確認されなかったことを示す.

以上の検証から，欠測表現に関する情報損失は，主に三つの段階で生じることが確認された．第一に，文字入力時の解釈により，異なる原表記が同一の文字欠測へ集約された．第二に，数値として解釈できない文字列が入力エラーを経て通常数値欠測へ変換され，原表記間の区別が失われた．第三に，IMPORT プロシジャによる型推測の結果，推測型と一致しない値が欠測へ変換される場合があった．一方，MISSING システムオプションは表示のみを変更し，SAS 内部値には影響しなかった．また，特殊欠測値は欠測として判定されながらも，個別の比較およびソートが可能であった.

5. 考察

5.1　欠測表現は最終的な値ではなく，入力から格納までの変換過程として確認する必要がある

本論文では，同一または類似した外部表記であっても，入力方法によって SAS 上の格納結果が異なることが確認された．特に，原表記「.」は BEST32.を用いた数値入力では正常な通常数値欠測「.」として入力された一方，NA，N/A，NULL，-，A.B 等の数値として解釈できない文字列も，入力エラーを経た後には同じ通常数値欠測「.」として格納された．このため，入力後の VALUE のみを確認した場合，正常に欠測として入力された値と，数値変換に失敗した結果として欠測となった値を区別することはできない.

また，文字型として入力する場合にも，必ずしも原表記が保持されるとは限らなかった．標準的な文字入力形式$20.では，空欄，単独の「.」，空白を伴う「.」がいずれも文字欠測へ集約された一方，$CHAR20.では単独の「.」および先頭空白を含む「.」を保持できた．さらに，この傾向は固定位置テキストと CSV の双方で同様に認められた．したがって，情報保持性は数値型か文字型かという型のみで決まるものではなく，入力方式および適用する入力形式にも依存すると考えられる.

さらに，MISSING システムオプションでは，通常数値欠測の内部値を変更することなく表示記号のみを変更できた．原表記「-」を BEST32.で入力した場合には数値変換エラーにより内部値は通常数値欠測「.」となったが，その後「OPTIONS MISSING='-';」を指定すると再び「-」と表示された．この例では，画面上

の表示だけを確認すると，原表記「-」がそのまま保持されたようにも見える．したがって，欠測値を確認する際には，「原表記」「入力時の解釈」「SAS 内部値」「表示」を分離して考える必要がある．

5.2　入力時の解釈規則は情報保持性を左右する

MISSING ステートメントの検証では，原表記「A」および「R」は，MISSING ステートメントを指定しない場合には数値として解釈できず，入力エラーを経て通常数値欠測「.」へ変換された．一方，「MISSING A R;」を指定した場合には，それぞれ特殊欠測「.A」および「.R」として正常に入力された．また，原表記「.A」および「.R」は MISSING ステートメントを指定しなくても特殊欠測として入力された．

この結果は，同一の文字列であっても，入力時にどのような解釈規則が設定されているかによって格納結果が変化することを示している．一方，「A」から「.A」への変換後は，入力後の特殊欠測値のみから，原表記が「A」であったのか「.A」であったのかを判別することはできない．すなわち，特殊欠測値を利用すること自体は欠測理由の区別に有用であるものの，変換前の原表記や変換経路を保持するものではない．

このことから，外部データの入力規則を設計する際には，最終的に得たい SAS 値だけではなく，どの原表記をどの値へ変換したかを追跡可能にしておくことが重要と考えられる．

5.3　IMPORT プロシジャの自動推測は利便性と再現性の両面から確認する必要がある

IMPORT プロシジャでは，変数型や変数長等を利用者が明示的に指定するのではなく，外部ファイルの内容から自動的に推測する．本検証では，同一の構成要素を持つ numeric_first.csv と text_first.csv について，レコード順序のみを変更したところ，GUESSINGROWS がデフォルトの条件では VALUE の推測型が異なった．numeric_first.csv では数値型と推測され，後半に現れた NA，NULL，A.B，-等が数値変換エラーとなって通常欠測「.」へ集約された．一方，GUESSINGROWS=MAX では VALUE が文字型と推測され，これらの文字列表記が保持された．

したがって，IMPORT プロシジャによる自動推測では，同一の値集合であっても，レコード順序や型推測に用いる範囲によって，変数型，入力エラーの発生および原表記の保持状態が変化する可能性がある．この性質は，ファイル作成元の都合等でレコード順が変化し得る実務データにおいて，再現性上の注意点となる．一方，GUESSINGROWS の範囲を広げればすべての情報損失を回避できるわけではない．test.csv では VALUE が文字型と推測された場合であっても，単独の「.」および空白を伴う「.」は文字欠測へ集約された．したがって，GUESSINGROWS=MAX は型推測範囲を拡大するための手段ではあるが，入力時の解釈そのものに由来する情報損失を防止するものではない．

IMPORT プロシジャを使用する場合には，生成されたデータセットの値だけでなく，推測された Type，Length，informat 等の変数属性および SAS ログを併せて確認することが望ましいと考える．

5.4　SAS の欠測値は「比較されない値」ではなく，SAS 固有の大小関係を持つ

特殊欠測値の検証では，「._」，「.」，「.A」，「.R」，「.Z」はいずれも MISSING 関数によって欠測と判定された一方，特定の欠測値との直接比較によって個別に識別できた．また，SORT プロシジャによる昇順並べ替えでは，「._」，「.」，「.A」，「.R」，「.Z」，「-1」，「0」，「1」の順となった．

この大小関係は条件抽出にも影響する．本検証では，「if value < 0;」とした場合，通常数値である-1 だけでなく，通常欠測および特殊欠測も抽出された．一方，「if not missing(value) and value < 0;」と欠測値を明示

的に除外した場合には-1 のみが抽出された．

この結果から，SAS における欠測値を「値がないため大小比較の対象にならない」と直感的に扱うと，条件式によって意図しないレコードが抽出される可能性がある．特に閾値を用いたデータ抽出やフラグ作成では，必要に応じて MISSING 関数等により欠測値を明示的に除外する必要がある．

また，通常の 1/0 では通常欠測「.」が得られたのに対し，DIVIDE 関数では正負の無限大に相当する結果として「.I」および「.M」が得られた．しかし，「.I」および「.M」も SAS 上では特殊欠測値として扱われ，MISSING 関数によって欠測と判定された．このことから，数学的な Infinity の概念と，SAS における特殊欠測値としての取扱いを区別する必要がある．

5.5　原表記の損失は欠測理由に関する意味情報の損失につながり得る

SDTMIG v3.3 では，主たる値が null の場合にその理由や状態を補足する仕組みとして NullFlavor が示されており，「NI」，「INV」，「UNK」，「NASK」，「NA」等，意味の異なる複数のコードが定義されている．これらはすべて主たる値が存在しない状態に関係するが，その理由は同一ではない．

本論文では，SDTMIG v3.3 の NullFlavor Enumeration に掲載された 15 種類のコードを対象とした．文字入力形式$20.および$CHAR20.では全 15 コードが原表記のまま保持され，IMPORT プロシジャでも今回の条件では文字列として保持された．一方，BEST32.を明示して数値として入力した場合には，全 15 コードが Invalid data となり，同一の通常数値欠測「.」へ集約された．

この場合，入力時エラーの発生状況を保存することで「数値変換時にエラーが発生した」という情報は保持できるが，変換後の値と入力時エラーのみから，元の表記が「NI」，「INV」，「UNK」，「NA」等のいずれであったかを復元することはできなかった．すなわち，原表記が業務上の意味を持つコードである場合，入力時の情報損失は単なる文字列表記の消失ではなく，後工程で必要となる意味的区別の消失につながる．

したがって，欠測理由等の意味を持つ原表記を含む外部データでは，入力時点で原表記を保持した上で，仕様や業務上の意味を確認した後に必要な変換を行うことが望ましいと考えられる．

6. まとめ

本論文は，BASE SAS における代表的な外部テキスト入力としてデータステップおよび IMPORT プロシジャを対象とし，限定したテストデータを用いて入力挙動を確認したものである．したがって，すべての外部データ形式，文字コード，SAS 環境および入力条件に同一の結果が生じることを示すものではない．

一方，異なる原表記が入力時に同一の SAS 欠測値へ集約される現象が，固定位置テキストおよび CSV，明示的な数値入力，IMPORT プロシジャによる型推測等，複数の条件で確認された．また，入力後に失われた原表記は，SAS 内部値や入力エラーの情報のみから完全には復元できない場合があった．

以上から，外部データの受領・変換工程では，最終的な解析用変数のみを保持するのではなく，少なくとも原表記，入力時のエラー状態，変換規則を追跡できる形で保持することが，情報損失の検出および原因究明に有用と考えられる．特に欠測理由等の意味を持つコードについては，自動的に特定の欠測値へ変換する前に，原表記を保持した状態で仕様との整合性を確認することが重要であることが示唆された．

参考文献

[1] Bost C. What You’re Missing About Missing Values. SAS Global Forum 2014, 1440-2014.

[2] Glanz H, Horstman J. The Missing Link: A Robust Macro for Recoding General Missing Data Values. Western Users of SAS Software Conference, 2018.

[3] SAS Institute Inc. SAS 9.4 Formats and Informats: Reference: $w. Informat. (https://documentation.sas.com/doc/da/pgmsascdc/v_073/leforinforref/n02aiion6v9r6an12db5exllrow3.htm , 最終アクセス: 2026/09/18)

[4] SAS Institute Inc. SAS 9.4 Formats and Informats: Reference: $CHARw. Informat. (https://documentation.sas.com/doc/da/pgmsascdc/v_073/leforinforref/n1oily9h2mwp5cn1dtb5qu8tcj7u.htm , 最終アクセス: 2026/09/18)

[5] SAS Institute Inc. SAS 9.4 Formats and Informats: Reference: BESTw. Informat. (https://documentation.sas.com/doc/da/pgmsascdc/v_073/leforinforref/p1fum54c93f8r0n1wrs5mrb05nzi.htm , 最終アクセス: 2026/09/18)

[6] SAS Institute Inc. SAS 9.4 Language Reference: Concepts. Missing Values / Order of Missing Values. (https://documentation.sas.com/doc/en/lrcon/9.4/p175x77t7k6kggn1io94yedqagl3.htm, 最終アクセス: 2026/09/18)

[7] SAS Institute Inc. SAS 9.4 System Options: Reference. MISSING= System Option. (https://documentation.sas.com/doc/ja/vdmmlcdc/8.1/lesysoptsref/n0qamf3yfjtwzhn1ran6xwblnlea.htm, 最終アクセス: 2026/09/18)

[8] SAS Institute Inc. SAS 9.4 Global Statements: Reference. MISSING Statement. (https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/lestmtsglobal/n0ewjmosjtyjbtn1t5zabfkrof4f.htm, 最終アクセス: 2026/09/18)

[9] SAS Institute Inc. Base SAS 9.4 Procedures Guide. IMPORT Procedure: GUESSINGROWS Statement. (https://documentation.sas.com/doc/da/pgmsascdc/v_073/proc/p13kvtl8ezj13in17i6m99jypcwi.htm, 最終アクセス: 2026/09/18)

[10] Clinical Data Interchange Standards Consortium (CDISC). Study Data Tabulation Model Implementation Guide: Human Clinical Trials Version 3.3. NullFlavor / TSVALNF.

[11] SAS Institute Inc. SAS 9.4 Functions and CALL Routines: Reference. DIVIDE Function. (https://documentation.sas.com/doc/da/pgmsascdc/v_073/lefunctionsref/p1xybc7e4naq2dn1shl7w9kbxqqz.htm, 最終アクセス: 2026/09/18)

Appendix 1

```
/*方法 A：$20 による文字入力*/
data read_dollar;
	infile "&path.¥test.txt" truncover;
	input
		@1 case_id 3.
		@5 value $20.
	;
	length method $20;
	method = "$20.";
run;

/*方法 B：$CHAR による文字入力*/
data read_char;
	infile "&path.¥test.txt"    truncover;
	input
		@1 case_id 3.
		@5 value $char20.
	;
	length method $20;
	method = "$CHAR20.";
run;

/*方法 C：  BEST32.による入力*/
data read_numeric;
	infile "&path.¥test.txt" truncover;
	input
		@1 case_id 3.
		@5 value best32.
	;
	length method $20;
	method = "BEST32.";
run;

data audit_dollar;
	set read_dollar;

	/* SAS 上で欠測か  */
	missing_flag = missing(value);

	/*  最後の非空白文字までの長さ  */
	lengthn_value = lengthn(value);

	/*  最初の非空白文字が何文字目か  */
	first_nonblank = notspace(value);

	/* HEX 表示  */
	length value_hex $40;
	value_hex = put(value, $hex40.);
run;
```

Appendix 2

```
data read_numeric_csv;
	infile "&path.¥test.csv"
		dsd
		dlm=","
		firstobs=2
		truncover;

	length method $20;
	informat	case_id best32.
			value	best32.;

	input case_id value;
	input_error = _error_;
	method = "BEST32.";
run;

data check_missing_option;
	set read_numeric_csv;

	missing_flag = missing(value);
	normal_missing = (value = .);
	special_a = (value = .A);
run;

options missing='-';
proc print data=check_missing_option;
	var case_id value missing_flag normal_missing special_a;
run;
```

Appendix 3

```
data special_missing_test;
	length label $10;

	id=1; value=._; label="._"; output;
	id=2; value=.;	label=".";	output;
	id=3; value=.A; label=".A"; output;
	id=4; value=.R; label=".R"; output;
	id=5; value=.Z; label=".Z"; output;
	id=6; value=-1; label="-1"; output;
	id=7; value=0;	label="0";	output;
	id=8; value=1;	label="1";	output;
run;
```

等張回帰入門 ― 理論、実装、物理アナロジー

田中 真史

（イーピーエス株式会社）

An Introduction to Isotonic Regression—Theory, Algorithms, and a Physical Analogy

Masashi Tanaka

EPS Corporation

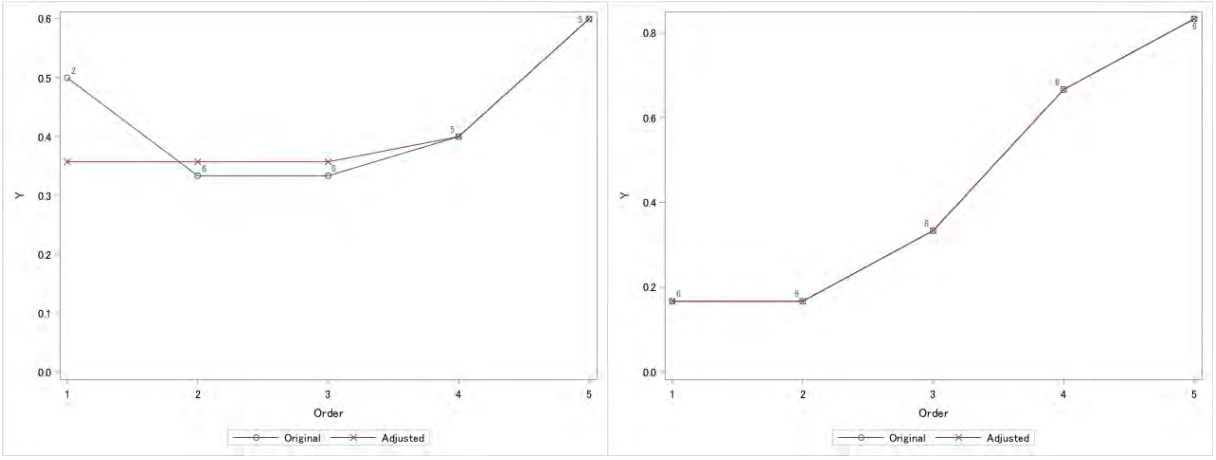
要旨

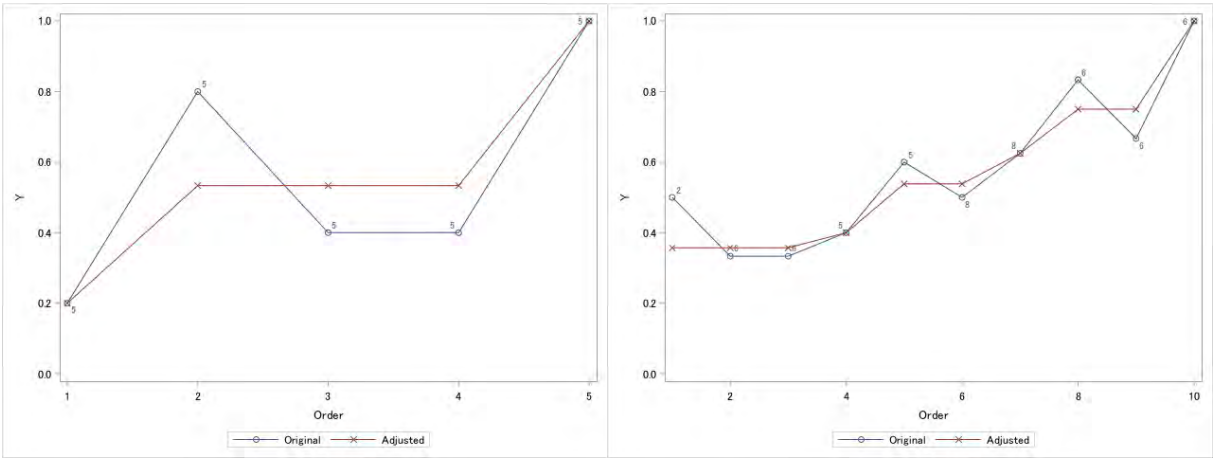
等張回帰はデータに単調増加性のみを仮定するノンパラメトリック回帰の手法で、臨床試験の用量漸増デザインなどで使用される。本稿では、制約付き最小二乗法、下側凸包、アルゴリズム、物理アナロジーの 4 つの観点から等張回帰を考察する。理論的背景は数理最適化の一般論には踏み込まずに展開し、PAVA とそれを簡略化したアルゴリズムを SAS で実装する際の手掛かりとした。等張回帰は棹秤（さおばかり）による物理アナロジーが可能で、実験してみると測定結果と SAS による計算結果は良く一致した。

キーワード：等張回帰、制約付き最小二乗法、下側凸包、PAVA、物理アナロジー

1. はじめに

日々の生活においては、単調増加する、すると思われる、あるいはしてほしい現象が存在する。等張回帰（等調回帰、単調回帰、アイソトニック回帰、Isotonic Regression）は、データに単調増加性のみを仮定してフィッティングを行うノンパラメトリック回帰の手法で、ときには減少してしまう現実のデータを「あるべき姿」に正すために用いる。下にいくつかの例をグラフで示した。





臨床試験の例を挙げると、第 I 相試験の用量漸増デザインのひとつである BOIN（Bayesian Optimal Interval）デザインで各用量の毒性発現率に単調増加性を仮定して推定して、最大耐用量（MTD: Maximum Tolerated Dose）を特定するのに使われる。重要な手法であるが、各種の解説では、結果のみを示している場合や理論を理解するのに数理最適化の予備知識を要求する場合が多いように思う。本稿では、数理最適化の予備知識をあまり仮定せずに等張回帰の理論を説明し、SAS による実装とひとつの物理アナロジーを示した。以下、制約付き最小二乗法、下側凸包、アルゴリズム、物理アナロジーの 4 つの観点から等張回帰について考察する。

2. 理論的背景

2.1. 制約付き最小二乗法と推定値の一意性

等張回帰は制約付き最小二乗法で定式化できる。群1から群nの実際のデータを$Y_1, \dots, Y_n$とする。データは非負（$Y_i \geq 0$）としておく。等張回帰によって得られる$Y_1, \dots, Y_n$の推定値$y_1, \dots, y_n$に

$$y_1 \leq y_2 \leq \cdots \leq y_{n-1} \leq y_n$$

の単調増加性（正確には単調非減少性）を仮定する。単調減少（単調非増加）を仮定しても良いが、本稿では単調増加の場合のみを扱う。評価関数Sは

$$S(\boldsymbol{y}, \boldsymbol{Y}) = \sum_{i=1}^{n} w_i (y_i - Y_i)^2, \qquad \boldsymbol{y} = (y_1, \dots, y_n), \qquad \boldsymbol{Y} = (Y_1, \dots, Y_n), \qquad w_i > 0$$

で定義する。w_iは群iの重みで、典型的には群iに含まれる治験参加者の人数である。$\boldsymbol{y}$を凸領域

$$\mathcal{D} = \{\boldsymbol{y} | y_1 \leq \cdots \leq y_n\}$$

の範囲で色々変えてSに最小値を与える$\boldsymbol{y}$を等張回帰による推定値$\boldsymbol{y}^*$と定義する。細かい点だが、推定値$\boldsymbol{y}^*$の存在は、$\mathcal{D}$を十分に大きな有界閉集合に制限することでSの連続性から保証されている。

もしデータが最初から$Y_1 \leq \cdots \leq Y_n$であれば$\boldsymbol{y}^* = \boldsymbol{Y}$が推定値で$S = 0$、逆に$S = 0$なら$\boldsymbol{y} = \boldsymbol{Y}$なので、この場合は推定値の一意性がいえる。データが単調増加とは限らない一般の場合には、Sが重み付きユークリッド距離の2乗で

$$S(\boldsymbol{y}_1, \boldsymbol{Y}) + S(\boldsymbol{y}_2, \boldsymbol{Y}) = 2\big(S(\boldsymbol{y}_3, \boldsymbol{Y}) + S(\boldsymbol{y}_3, \boldsymbol{y}_1)\big), \qquad \boldsymbol{y}_3 = (\boldsymbol{y}_1 + \boldsymbol{y}_2)/2$$

の中線定理を満たすことから推定値の一意性がいえる。ふたつの異なる$\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathcal{D}$で$s = S(\boldsymbol{y}_1, \boldsymbol{Y}) = S(\boldsymbol{y}_2, \boldsymbol{Y})$が最小値になったと仮定すると、$\mathcal{D}$は凸なので$\boldsymbol{y}_3 \in \mathcal{D}$で、

$$S(\boldsymbol{y}_3, \boldsymbol{Y}) = s - S(\boldsymbol{y}_3, \boldsymbol{y}_1) < s$$

だからsの最小性に反する。

推定値$\boldsymbol{y}^*$について$y_i^* \geq 0$であることを注意しておく。$y_i^* < 0$を仮定すると$y_1^* = \cdots = y_i^* = 0$と置きなおした点の方が$S$を小さくするからである。

2.2. 条件付き極値の必要条件

等張回帰の推定値、すなわち$\mathcal{D}$におけるSの最小点$\boldsymbol{y}^* = (y_1^*, \ldots, y_n^*)$を$\mathcal{D}$における$S$の条件付き極値点として扱う。数理最適化の精緻な一般論には触れず、次節の図形的解釈の準備を目的とする。$\boldsymbol{Y} \in \mathcal{D}$のときは$\boldsymbol{y}^* = \boldsymbol{Y}$である。$\boldsymbol{Y} \notin \mathcal{D}$のとき、$\boldsymbol{y}^*$は$\mathcal{D}$の境界上に存在する。評価関数の形状のために、$\boldsymbol{Y}$に近づこうすると$\mathcal{D}$の境界に到達するからである。$\mathcal{D}$の境界では順序制約が$y_i = y_{i+1}$の等式になる。

ラグランジュの未定乗数法で

$$L = S(\boldsymbol{y}, \boldsymbol{Y}) + \sum_{i=1}^{n-1} \lambda_i (y_i - y_{i+1})$$

をy_iで微分して、

$$\left.\frac{\partial L}{\partial y_i}\right|_{\boldsymbol{y}=\boldsymbol{y}^*} = 2w_i(y_i^* - Y_i) + \lambda_i - \lambda_{i-1} = 0, \qquad i = 1, \ldots, n \tag{1}$$

が得られる。$y_i^* = y_{i+1}^*$の制約が無いiに対しては$\lambda_i = 0$とし、端についても$\lambda_0 = \lambda_n = 0$とした。つまり

$$\lambda_i(y_i^* - y_{i+1}^*) = 0, \qquad i = 1, \ldots, n-1 \tag{2}$$

である。ここで$\lambda_i \geq 0$が分かる。評価関数を展開して式1を用いると

$$S(\boldsymbol{y}, \boldsymbol{Y}) = S(\boldsymbol{y}^*, \boldsymbol{Y}) + 2\sum_{i=1}^{n} w_i(y_i^* - Y_i)(y_i - y_i^*) + S(\boldsymbol{y}, \boldsymbol{y}^*) = S(\boldsymbol{y}^*, \boldsymbol{Y}) - \sum_{i=1}^{n} (\lambda_i - \lambda_{i-1})(y_i - y_i^*) + S(\boldsymbol{y}, \boldsymbol{y}^*)$$

になる。任意の正数εについて

$$\boldsymbol{z} = (y_1^*, \ldots, y_i^*, y_{i+1}^* + \varepsilon, \ldots, y_n^* + \varepsilon), \qquad \boldsymbol{z} \in \mathcal{D}$$

と置く。$\boldsymbol{y}^*$の最小性より$S(\boldsymbol{z}, \boldsymbol{Y}) > S(\boldsymbol{y}^*, \boldsymbol{Y})$だから

$$S(\boldsymbol{z}, \boldsymbol{Y}) = S(\boldsymbol{y}^*, \boldsymbol{Y}) - \varepsilon \sum_{j=i+1}^{n} \left(\lambda_j - \lambda_{j-1}\right) + \varepsilon^2 \sum_{j=i+1}^{n} w_j = S(\boldsymbol{y}^*, \boldsymbol{Y}) + \varepsilon\left(\lambda_i + \varepsilon \sum_{j=i+1}^{n} w_j\right) > S(\boldsymbol{y}^*, \boldsymbol{Y})$$

になる。$\lambda_i + \varepsilon \sum_{j=i+1}^{n} w_j > 0$で、$\varepsilon$が任意だから$\lambda_i \geq 0$である。

式1を足し合わせて$\lambda_i \geq 0$を用いると

$$\sum_{j=1}^{i} w_j Y_j = \sum_{j=1}^{i} w_j y_j^* + \frac{1}{2}\lambda_i \geq \sum_{j=1}^{i} w_j y_j^*, \qquad \sum_{j=1}^{n} w_j Y_j = \sum_{j=1}^{n} w_j y_j^* \tag{3}$$

になる。ここから等張回帰の図形的解釈が得られる。ふたつめの式は$\boldsymbol{Y}$全体と$\boldsymbol{y}^*$全体の力のモーメントが一致することを意味する。

2.3. 累積和の下側凸包による図形的解釈

平面上に

$$(x, y) = \left(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} w_j Y_j\right), \qquad (x, y) = \left(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} w_j y_j^*\right), \qquad i = 1, \ldots, n$$

をプロットし、$(0,0)$もグラフに含める。式3は$\boldsymbol{Y}$の累積和のグラフの下に$\boldsymbol{y}^*$の累積和のグラフが位置し、$i = n$の端が一致することを意味する。これらのグラフは$Y_i \geq 0, y_i^* \geq 0$より単調増加である。$\boldsymbol{y}^*$の累積和のグラフは傾きが増えていくので下に凸である（以下、「の累積和のグラフ」は適宜省略する）。さて、$\boldsymbol{Y}$の下側に存

在するふたつの下に凸なグラフを重ね合わせた上側の境界は、新たな下に凸なグラフである。同様に、$\boldsymbol{Y}$の下側のすべての下に凸なグラフを重ねた境界が$\boldsymbol{Y}$の下側に描ける最大の下に凸なグラフで、これを下側凸包（GCM: Greatest Convex Minorant）と呼ぶ。$\boldsymbol{y}^*$は実は$\boldsymbol{Y}$の下側凸包であることが次のようにして分かる。$\lambda_a = 0, \lambda_{a+1} \neq 0, \ldots, \lambda_{b-1} \neq 0, \lambda_b = 0$とする。式1を足し合わせると

$$\sum_{j=\mathrm{a}+1}^{b} w_j Y_j = \sum_{j=\mathrm{a}+1}^{b} w_j y_j^*$$

だから、式2より

$$y_{a+1}^* = \cdots = y_b^* = \frac{\sum_{j=\mathrm{a}+1}^{b} w_j Y_j}{\sum_{j=\mathrm{a}+1}^{b} w_j}$$

になる。$\lambda_a = \lambda_b = 0$なので式3より

$$\sum_{j=1}^{a} w_j Y_j = \sum_{j=1}^{a} w_j y_j^*, \qquad \sum_{j=1}^{b} w_j Y_j = \sum_{j=1}^{b} w_j y_j^*$$

になる。累積和のグラフでは群aと群b（にそれぞれ対応する点）は等張回帰の前後で動かず、群aと群bを直線で結ぶ操作が等張回帰になる。この操作は$\boldsymbol{Y}$の下側凸包を作る操作に一致し、$\boldsymbol{Y}$の下側にこのようにして描ける図形は下側凸包のみである。よって、$\boldsymbol{Y}$の下側凸包から$\boldsymbol{y}^*$が得られる。

等張回帰のアルゴリズムは、図形的には$\boldsymbol{Y}$の累積和の折れ線グラフを上手く変形して下側凸包を作ることに相当する。

3. SAS による実装

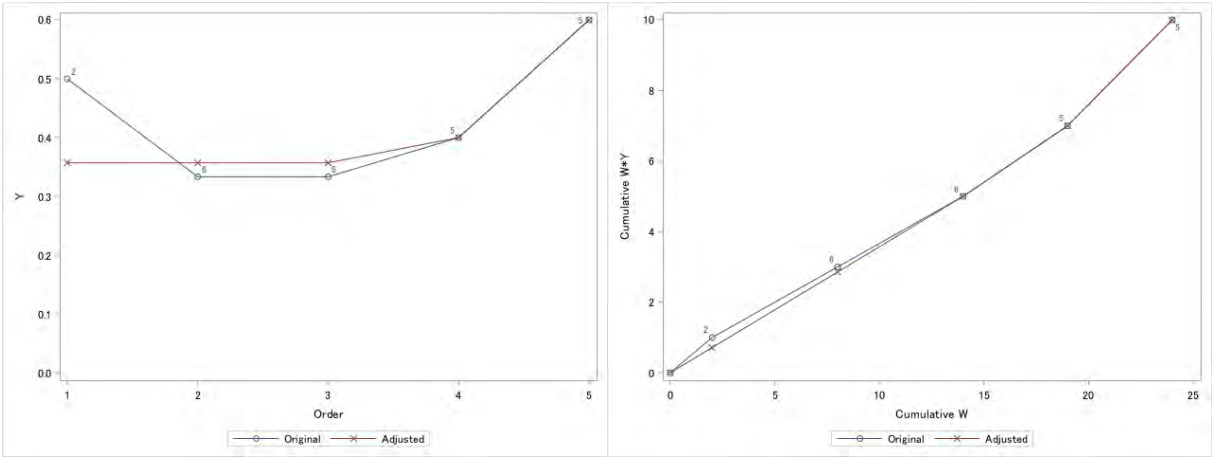
等張回帰によるフィッティングは、通常、Pool Adjacent Violators Algorithm（PAVA）と呼ばれる反復アルゴリズムで行う。初期値は$y_i = Y_i$とする。データ$\boldsymbol{Y}$が単調増加していればそこで終了し、そうでなければ隣接する群をクラスターにして重み付き平均でy_iを置き換える。それ以降の反復でも同様に、クラスターごとに与えられたy_iが単調増加していなければ、隣接（Adjacent）するクラスターを統合（Pool）して大きなクラスターを作り、y_iを重み付き平均で置き換えて、違反（Violation）を解消していく。

本章ではコードの簡単さを重視して PAVA（付録 2）のクラスター化の手順を省略したコードで説明する（付録 1）。PAVA に比べて計算量は増えるが、臨床試験で数個のデータに対して等張回帰をするような場合は問題にならない。コードの中心部分は DO WHILE ループで、説明のために記号 A、B、C を付与した。マクロ変数 LEN は群の個数である。

```
do i = 1 to &len.; * <-- A;
     _ay[i] = _y[i];
end;
do while(1); * <-- B;
     viol = 0;
     do i = 1 to &len. - 1;
          if round(_ay[i], 1e-10) > round(_ay[i + 1], 1e-10) then do; * <-- C;
               viol = 1;
               _ay[i]     = (_w[i]*_ay[i] + _w[i + 1]*_ay[i + 1])/(_w[i] + _w[i + 1]);
               _ay[i + 1] = _ay[i];
          end;
     end;
     iteration + 1;
     output;
     if viol = 0 then leave;
end;
```

A の DO ループで$y_i = Y_i$とする。B の DO WHILE ループで単調増加かどうかを判定して、$y_i > y_{i+1}$の違反があれば隣接するふたつの群の重み付き平均でy_i, y_{i+1}を置き換える。DO WHILE ループを打ち切るために C の ROUND 関数で端数処理を加えた。付録 1 のコードを実行すると出力データセットの最終オブザベーションに等張回帰の結果が得られ、等張回帰を図示したグラフ（左：$\boldsymbol{Y}$と$\boldsymbol{y}$、右：$\sum_i w_i Y_i$と$\sum_i w_i y_i$）も出力できる（Base SAS 9.4M8 を用いた）。

群	Y	w	y
1	0.5	2	0.35714
2	0.33333	6	0.35714
3	0.33333	6	0.35714
4	0.4	5	0.4
5	0.6	5	0.6



コードが等張回帰の推定値を与える理由は次の通りである。等張回帰の前後で累積和のグラフの特定の群aと群b（にそれぞれ対応する点、ただし上の例では原点と群 3）は動かず、それらの点を直線で結ぶ操作が等張回帰である。群aから群bまでをひとつのクラスターとすると、他のクラスターとの干渉は起こらないので独立に考えられる。本章で示したコードは、群aと群bを結ぶ折れ線を、それらを結ぶ直線に近づける操作に他ならない。十分に近づいたところで打ち切れば等張回帰の推定値が得られる。PAVA の実装ではコードをもう少し工夫して、群aと群bをより効率的に直線で結ぶ。

4. 棹秤によるアナロジー

等張回帰とそのアルゴリズムは棹秤（さおばかり）によって再現できる。筆者の自宅に棹秤が見当たらなかったので木の棒（900 mm×9 mm）と糸（太さ約 0.4 mm）で簡易的な装置を作成した。錘には 10 円玉を使用した（支点付近の 1 円玉は微調整用の錘）。データ$\boldsymbol{Y}$を cm に変換して、棹の右側の支点からその距離だけ離れた場所に、各群の重みの数だけ 10 円玉を吊り下げた。棹の左側には全体がつり合うように 10 円玉を吊り下げ、左側の錘の位置は固定した。右側の錘を群の小さい方から見ていき、支点から錘までの距離が減少していれば、ふたつの錘を近づけて群をまとめた。右側の錘をずらす過程では全体のつり合いを保ち、初期値のモーメントを維持した。3 章の表の数値を50倍して初期値（cm）とした場合の結果を下に示した。写真は棹秤の初期状態（左）と最終状態（右）である。表の一番右の列が結果で、大雑把な装置の割に値は良く一致した。

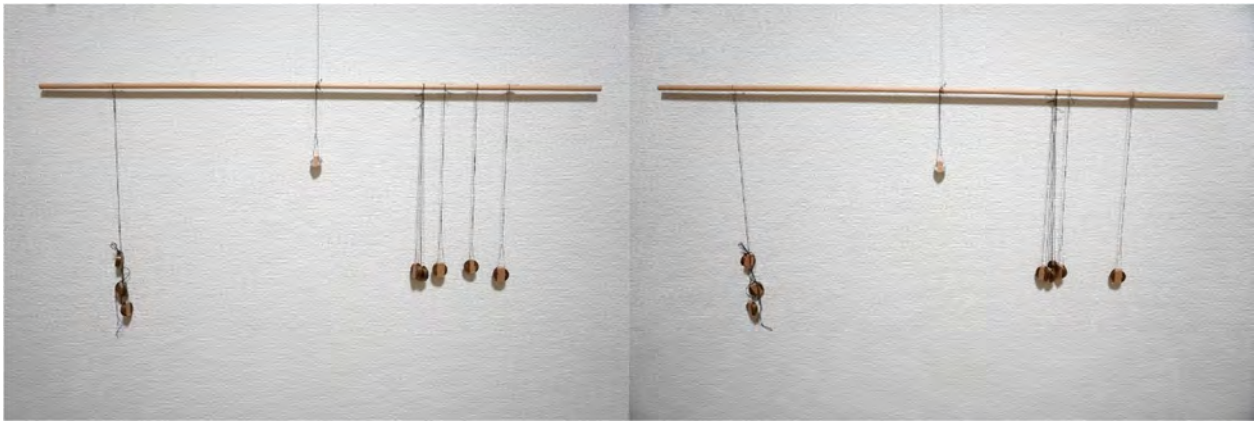
錘をまとめる手順は 3 章のコードと同様である。支点からの距離がL_i, L_{i+1}のふたつの錘（重さw_i, w_{i+1}）を距離L'の位置にまとめる際につり合いを保つことを力のモーメントで表すと

$$w_i L_i + w_{i+1} L_{i+1} = (w_i + w_{i+1}) L'$$

だから、錘をまとめた位置は

$$L' = \frac{w_i L_i + w_{i+1} L_{i+1}}{w_i + w_{i+1}}$$

になる。実際に作業すると、既にまとめたふたつ以上の錘をひとつのクラスターとして動かせることが分かる。これが PAVA の操作であった。



群	Y (cm)	w (枚)	y (cm) 途中	y (cm) 最終	$Y/50$	$y/50$
1	25	2	18.9	17.9	0.5	0.358
2	16.7	6	18.9	17.9	0.334	0.358
3	16.7	6	16.7	17.9	0.334	0.358
4	20	5	20	20	0.4	0.4
5	30	5	30	30	0.6	0.6
左側	33.4	15	33.4	33.4	-	-
右側モーメント (枚×cm)	500.4		501.4	500.6		
左側モーメント (枚×cm)	501		501	501		

5. まとめ

等張回帰について、筆者が分かる範囲で一貫した説明を試みた。個別の参考文献の列挙は省略したが、本稿は等張回帰に関する多種多様な解説や公開されているソースコードから学んだ知見に基づいている。より本格的な解説は Web で入手できるし、R の{BOIN}パッケージの関数を AI で SAS に変換すれば必要な SAS コードは容易に得られると思う。なお、{BOIN}パッケージの select.mtd 関数は測定値と各群の人数を補正して初期値と重みに使っている（実装は難しくない）。

等張回帰は実用的な手法であるとともに、制約付き最小二乗法、下側凸包、アルゴリズムといった様々な見方ができて理論的、プログラミング的に面白い手法だと思う。棹秤で再現した事例は見当たらなかったが、物理アナロジーとしては明らかである。単なる想像だが、PAVA の発想の端緒は案外このような物理的直観だったのかもしれない。

付録 1　等張回帰の簡易的な実装

```
data test;
     y1 = 1/2; w1 = 2;
     y2 = 2/6; w2 = 6;
     y3 = 2/6; w3 = 6;
     y4 = 2/5; w4 = 5;
     y5 = 3/5; w5 = 5;
run;
%macro isotonic_reg(ds, len, plots =, limit = 1000);
     data isotonic;
          set &ds.;
          array _y[*] y1-y&len.;
          array _w[*] w1-w&len.;
          array _ay[*] ay1-ay&len.;
          do i = 1 to &len.;
               _ay[i] = _y[i];
          end;
          do while(1);
               viol = 0;
               do i = 1 to &len. - 1;
                    if round(_ay[i], 1e-10) > round(_ay[i + 1], 1e-10) then do;
                         viol = 1;
                         _ay[i]     = (_w[i]*_ay[i] + _w[i + 1]*_ay[i + 1])/(_w[i] + _w[i + 1]);
                         _ay[i + 1] = _ay[i];
                    end;
               end;
               iteration + 1;
               output;
               if iteration >= &limit. then do;
                    put 'WARNING: Change LIMIT=.';
                    viol = 0;
               end;
               if viol = 0 then leave;
          end;
          keep y1-y&len. w1-w&len. ay1-ay&len. iteration;
     run;
     proc print data = isotonic;
     run;
     %if &plots. = on %then %do;
          data plots;
               length x1 y11 y12 x2 y21 y22 8;
               set isotonic end = eof;
               array _y[*] y1-y&len.;
               array _w[*] w1-w&len.;
               array _ay[*] ay1-ay&len.;
               if eof then do;
                    x2 = 0;
                    y21 = 0;
                    y22 = 0;
                    output;
                    do i = 1 to &len.;
                         x1 = i;
                         w = _w[i];
                         y11 = _y[i];
                         y12 = _ay[i];
                         x2 + _w[i];
                         y21 + _w[i]*_y[i];
```

```
                    y22 + _w[i]*_ay[i];
                    output;
               end;
          end;
          drop y1-y&len. w1-w&len. ay1-ay&len. iteration i;
     run;
     proc sgplot data = plots;
          label y11 = 'Original';
          label y12 = 'Adjusted';
          series x = x1 y = y11 / datalabel = w markers;
          series x = x1 y = y12 / markers markerattrs = (symbol = x);
          xaxis label = 'Order';
          yaxis label = 'Y' min = 0;
     run;
     proc sgplot data=plots;
          label y21 = 'Original';
          label y22 = 'Adjusted';
          series x = x2 y = y21 / datalabel = w markers;
          series x = x2 y = y22 / markers markerattrs = (symbol = x);
          xaxis label = 'Cumulative W';
          yaxis label = 'Cumulative W*Y';
     run;
%end;
%mend isotonic_reg;
%isotonic_reg(test, 5, plots = on);
```

付録 2　等張回帰の PAVA による実装

```
%macro pava(ds, len);
     data pava;
          set test;
          array _y[*] y1-y&len.;
          array _w[*] w1-w&len.;
          array _ay[*] ay1-ay&len.;
          do i = 1 to &len.;
               _ay[i] = _y[i];
          end;
          do while(1);
               viol = 0;
               do i = 1 to &len. - 1;
                    if _ay[i] > _ay[i + 1] then do;
                         viol = 1;
                         idx  = i;
                    end;
                    if viol = 1 then leave;
               end;
               if viol = 0 then leave;
               do left = idx to 2 by -1;
                    if _ay[left] ^= _ay[left - 1] then leave;
               end;
               do right = idx + 1 to &len. - 1;
                    if _ay[right] ^= _ay[right + 1] then leave;
               end;
               sumw  = 0;
               sumxw = 0;
               do i = left to right;
               sumw  = sumw  + _w[i];
                    sumxw = sumxw + _ay[i]*_w[i];
```

```
                end;
                pooled = sumxw/sumw;
                do i = left to right;
                     _ay[i] = pooled;
                end;
                iteration + 1;
           end;
           keep y1-y&len. w1-w&len. ay1-ay&len. iteration;
     run;
     proc print data = pava;
     run;
%mend pava;
%pava(test, 5);
```

MIMIC-IVを用いた教育機関におけるRWD解析の試み

○小林美保[1]、小野洋平[2]

（[1]聖路加国際大学大学院公衆衛生学研究科 修士課程,

[2]聖路加国際大学大学院公衆衛生学研究科）

A Practical Framework for Real-World Data Analysis in Educational Institutions Using MIMIC-IV

Miho Kobayashi[1], Yohei Ono[2]

[1]St. Luke's International University, Graduate School of Public Health Master's Program,

[2]St. Luke's International University, Graduate School of Public Health

要旨

近年、電子カルテ等の日常診療から得られるリアルワールドデータ（Real-World Data: RWD）の研究利用が進展しており、RWD を適切に扱う人材の育成が重要となっている。一方、医療機関の診療データへのアクセスには倫理的・制度的な手続きが必要であり、医学部や附属病院を持たない教育機関では、RWD を実際に扱う教育機会の確保が課題となっている。そこで本研究では、医療機関の RWD への直接的なアクセスが困難な教育環境において、RWD 研究の一連のプロセスを実践的に学ぶための教育スキームを構築し、その実行可能性を検討した。本スキームでは、研究仮説の設定、研究デザインの選択、データアクセス手続き、研究倫理教育、統計解析計画書の作成・事前登録、データ抽出・前処理、統計解析、結果の解釈および成果報告までを一連のプロセスとして実施した。RWD には MIMIC-IV v3.1 を用い、DuckDB・SQL によるデータ抽出・結合、Python によるデータ処理・統計解析を行った。実践例として、MIMIC-IV の全入院 546,028 件を対象に、主たる使用言語（英語／非英語）と院内死亡との関連を検討した。解析に必要な変数が揃った 523,039 件を対象に、単変量および共変量を投入した多変量ロジスティック回帰分析、逆確率重み付け法による分析を実施した結果、単変量モデルで認められた非英語群の死亡オッズ上昇は、多変量調整および逆確率重み付け法により消失・減衰した。本研究により、自施設の電子カルテデータや国内の RWD へのアクセスが困難な教育機関であっても、適切な申請・手続きを経て利用可能な MIMIC-IV を教育資源として用いることで、実際の RWD を対象とした研究プロセスを実践的に学習する機会を提供できる可能性が示された。これは、教育機関による RWD へのアクセス環境の違いを補完し、RWD を扱う技術と実践力を学習する機会の確保につながる可能性がある。

キーワード：リアルワールドデータサイエンス、医療データサイエンス教育、実践的教育、人材育成、
教育機会格差

1. はじめに

1.1 医療におけるリアルワールドデータの重要性

近年、医療のデジタル化に伴い、電子カルテ、レセプトデータ、患者レジストリ等、日常診療の過程で蓄積される医療情報を研究に活用する機会が拡大している[1]。Real-World Data（RWD）は、患者の健康状態や医療提供に関して日常的に収集されるデータとして整理され、RWD を解析して得られるエビデンスは Real-World Evidence（RWE）とされる。日本においても、RWD/RWE の活用は臨床開発や医療研究における重要な課題となっている[2]。従来、医療における効果の検証にはランダム化比較試験（Randomized Controlled Trials: RCTs）が重要な役割を果たしてきた。RCTs は介入とアウトカムの因果関係を検討するうえで高い内的妥当性を有する一方で、厳格な適格基準や管理された研究環境で実施されることから、実臨床における多様な患者集団や診療状況への適用に限界が生じる可能性がある。これに対して RWD は、実臨床を反映することから RCTs で得られた知見を実臨床の観点から補完する情報源となり得る。日本においても、実際の医療環境下で取得された RWD の利活用が進展している。厚生労働省は 2021 年に「承認申請等におけるレジストリの活用に関する基本的考え方」[3] を示し、レジストリデータを承認申請等に活用する際の基本的な考え方を示している。さらに、その後もレジストリデータの外部対照等への活用に関する検討が進められており、日本においても RWD を研究・開発に利用する環境は変化している。

1.2　RWD 研究を担う人材の必要性

しかし、RWD が存在することと、それを適切に利用できることは同義ではない。RWD は、研究目的で統一された手順により収集されたデータとは異なり、日常診療におけるデータ生成過程に由来する欠測、測定頻度の違い、施設間のデータ構造の違い、交絡および選択バイアスなどが存在する可能性がある[4]。そのため、RWD 研究を実施するためには、単に統計解析手法を習得するだけでなく、データがどのように生成されたかを理解し、研究仮説に応じて適切な研究デザインと解析計画を設定する能力が必要である[5]。すなわち、今後の RWD 活用においては、RWD から適切な RWE を生み出すことのできる人材の育成が重要となる。

1.3 日本における RWD 活用と教育上の課題

一方で、RWD を活用した研究を実施する環境には課題がある。医療機関が保有する診療データには個人情報等が含まれるため、研究目的で利用するためには適切な手続きが必要となる。また、電子カルテシステムやデータベース構造は医療機関によって異なり、施設内で蓄積されたデータを外部の教育機関が利用することは必ずしも容易ではない。日本の RWD/RWE の現状についても、RWD へのアクセスやデータリンケージが課題として指摘されている[2]。

この問題は、研究そのものだけでなく、RWD 研究を学習する機会の地域的な偏りにもつながる可能性がある。医学部、大学病院あるいは医療機関と一体となった教育環境では、医療データを用いた研究を実施する機会を得られる場合がある。一方、それ以外の教育機関では、実際の診療データを用いた研究経験を得ることが難しい場合がある。しかし、RWD を活用したデータサイエンス教育の目的は、必ずしも自施設の患者データを使用することではない。適切な手続きのもとで利用可能な RWD を教育資源として活用することで、所在地や所属機関にかかわらず、RWD 研究に必要な能力を段階的に学習できる可能性がある。

1.4 本研究の目的

そこで本研究では、医療機関の RWD への直接的なアクセスが困難な教育環境においても、RWD 研究の一連のプロセスを経験できる教育スキームを構築し、実際に RWD 研究を実践することで、その実行可能性を検討することを目的とした。本研究では、MIMIC-IV を教育用 RWD として利用し、データアクセスに必要な

手続きから研究計画の作成、統計解析計画書作成、事前登録、データ処理、統計解析および研究成果の報告までを一連のプロセスとして実施した。さらに、その実践の中で実施した RWD 研究を一つの具体例として示し、RWD 教育における本スキームの意義について検討した。本稿における全体の展開および各章の論述の流れを図 1 に示す。第 2 章で教育スキームの具体的設計を述べ、第 3 章において MIMIC-IV を用いた実証解析を展開し、第 4 章および第 5 章でその教育的意義について考察する。



図 1: 本研究における論文の論理展開と教育スキームの全体構成

2 本研究で構築した RWD 解析教育スキーム

2.1 教育スキーム

本研究では、RWD 教育を統計解析の方法を学ぶことに限定せず、一連の RWD を用いた研究を実施する能力を身につけることとした。具体的には、研究仮説の設定、研究デザインの選択、データベースの選択、データアクセスに必要な手続き、研究倫理教育の受講、統計解析計画書の作成、統計解析計画書の事前登録、データ抽出・前処理、統計解析、結果の解釈、成果の報告である。本スキームの導入段階として、日常診療における臨床的疑問を出発点としつつ、利用するデータベースに記録されている項目やデータ構造の制約を確認しながら、検証可能な研究課題へと具体化させる手順を採用した。本研究で構築した RWD 解析教育スキームを図 2 に示す。

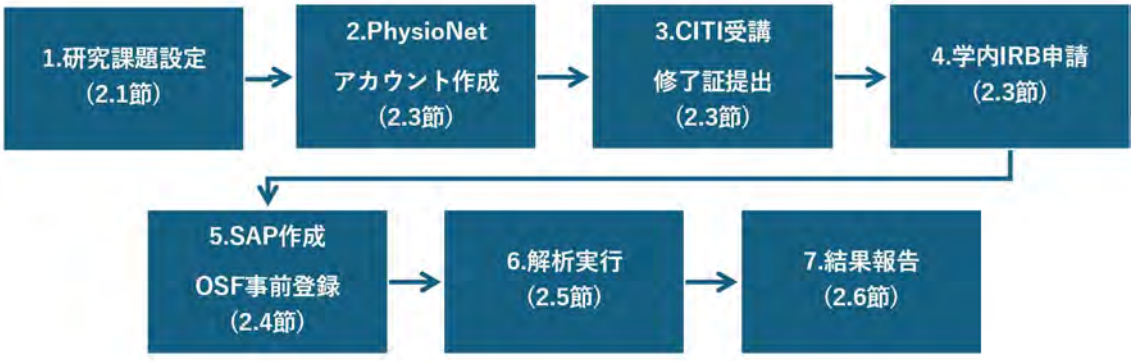


図 2: RWD 解析教育スキームの一連の研究プロセスフロー（2，3 については図 3 に詳述）

2.2 MIMIC-IV を用いた実践環境の構築

本研究では、RWD 解析を実施するためのデータベースとして MIMIC-IV を利用した。MIMIC-IV は、米国の Beth Israel Deaconess Medical Center の医療情報を基盤とした大規模な匿名化された電子健康記録データベースである [6,7]。MIMIC-IV v3.1 は 2024 年 10 月に公開され、hospital module および ICU module を中心に構成されている。hospital module には、入院情報、患者背景、検査、微生物検査、処方、診断、処置、請求情報な

どが含まれる。単純な表形式のデータだけではなく、複数のテーブルを結合し、患者単位、入院単位、検査単位などの異なる単位でデータが管理されている。そのため、必要な情報を複数のテーブルから抽出・結合し、研究目的に応じた解析用データセットを構築する必要がある。このようなデータ構造は、実際の RWD 研究におけるデータ抽出や前処理の過程を学習する教育環境として活用できると考えた。

2.3 データアクセスまでの手続き

MIMIC-IV の利用に際しては、所定のアクセス手続きが必要である。まず PhysioNet 上でユーザー登録を行い、credentialed アクセスに必要な条件を確認した [8]。MIMIC-IV v3.1 では、CITI の Data or Specimens Only Research training の修了および データ利用規約への同意が必要とされている [7]。これらの手続きを経てデータへのアクセスを取得した。また、本研究の実施にあたっては、所属機関における研究倫理手続きの一環として倫理審査委員会への申請を行い、承認を得た上で研究を推進した。これにより、所定の手続きにより利用可能な RWD を活用した解析であっても、研究倫理規範や所属機関のガバナンスを遵守する一連のプロセスを教育スキームに組み込んだ。本研究で実施したデータアクセス承認までの流れを図 3 に示す。

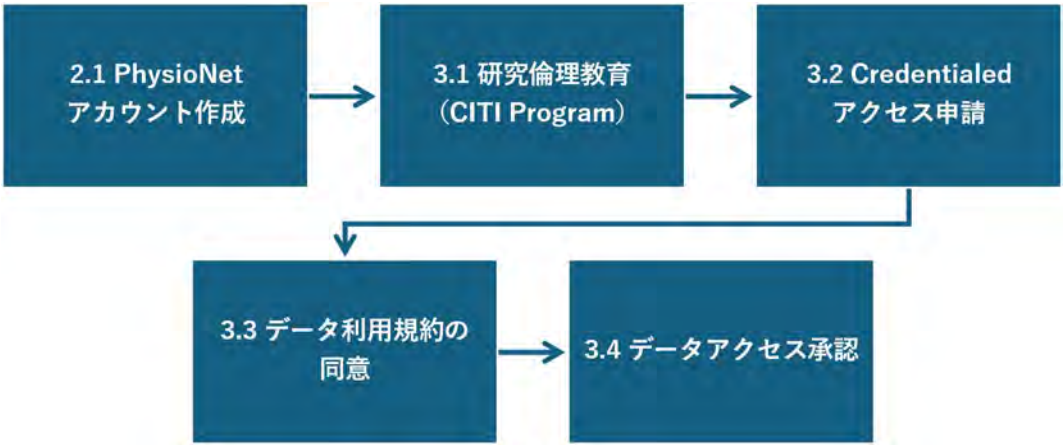


図 3: PhysioNet におけるデータ利用申請および承認フロー（図 2 のステップ 2，3 詳細）

2.4 研究計画書・統計解析計画書の作成と事前登録

RWD を用いた観察研究では、ランダム化比較試験と異なりデータを確認した後に研究対象や解析方法を変更することが容易である。そこで本研究では、データ解析を開始する前に研究計画を作成し、研究対象、曝露、主要評価項目、共変量、主要解析、欠損データの取り扱いについて事前に規定した。さらに、統計解析計画書を作成し、Open Science Framework（OSF）[9,10] に事前登録した。これにより、解析開始前に研究計画および解析方法を明確化し、解析上の自由度を抑えることで、研究の透明性および再現可能性を高めることを目指した。

2.5 データ処理・統計解析

MIMIC-IV は複数のテーブルから構成されており、大規模なデータを含んでいるため DuckDB および SQL を用いて必要なテーブルを結合し解析用データセットを作成した。その後、Python を用いてデータクリーニングおよび統計解析を実施した。

2.6 研究成果の報告と透明性の確保

結果の公表にあたっては、事後的な選択的報告を抑止し、研究の再現性と透明性を担保することが重要である。本スキームでは、事前に作成した統計解析計画書に基づき、観察研究の標準的な報告ガイドラインである STROBE 声明 [11] に準拠して対象者の抽出、欠損データの状況、未調整および多変量調整モデルの結果を体系的に整理・報告する作法を学習した。

3 RWD 解析教育の実践: MIMIC-IV を用いた言葉の壁と院内死亡に関する研究

3.1 研究の概要

本研究では、構築した RWD 解析教育スキームを研究者自身の RWD 研究に適用し、教育スキームの実践例として、MIMIC-IV を用いた主たる使用言語と院内死亡の関連に関する研究を実施した。多文化社会において、患者と医療従事者との間に言葉の壁が存在する場合、患者が自覚症状や既往歴などを十分に伝えることが困難となったり、医療従事者からの治療方針や治療上の指示の理解に影響したりする可能性がある。このような言葉の壁は、診療上のコミュニケーションに影響を及ぼし、適切な診断や治療の遅れに繋がり、患者の予後を悪化させる可能性が指摘されている。そこで本研究では、MIMIC-IV に記録された患者の使用言語に着目し、主たる使用言語と院内死亡との関連を検討することとし、リサーチクエスチョンとして「主たる使用言語が英語である患者と非英語である患者では、院内死亡に違いがあるか」を設定した。

3.2 研究対象

MIMIC-IV v3.1 の hospital module に含まれる全入院データ（計 546,028 件）を対象とした。主たる使用言語、院内死亡およびその他の解析に必要な変数を抽出し、対象変数に欠損のない症例 523,039 件を解析用データセットとした。

3.3 変数

本研究では、主たる使用言語を曝露、院内死亡を主要評価項目とした。主たる使用言語は英語と非英語の 2 値に分類した。そのほかの患者背景（年齢、性別、人種、社会経済的要因、入院経路）、臨床重症度としてチャールソン併存疾患指数 (CCI) を共変量として設定した。使用した変数について表 1 に示す。

表 1: 本解析で用いた変数の定義、参照テーブル

変数名	参照テーブル	役割・定義
使用言語	admissions	【主要曝露】自己申告による言語（英語 vs 非英語）
院内死亡	admissions	【主要評価項目】入院中の死亡
年齢（歳）	patients	【共変量】入院時の基準年齢
性別	patients	【共変量】男性/女性
人種	admissions	【共変量】白人、黒人、ヒスパニック、アジア系、その他/不明
保険種別	admissions	【共変量】メディケア、メディケイド、民間保険、費用免除、その他、不明
婚姻状況	admissions	【共変量】既婚、未婚、離婚、死別、不明
入院区分	admissions	【共変量】救急入院、救急観察、準緊急、予定入院
チャールソン併存疾患指数	diagnoses_icd	【共変量】CCIアルゴリズムに従い、ICDコードより算出

3.4 統計解析

MIMIC-IV から抽出したデータについて、まず対象集団の背景を記述統計により整理し、主たる使用言語別に比較した。群間の背景差は標準化平均差を用いて評価し、SMD≧0.1 以上を不均衡と評価した。その後、主たる使用言語と院内死亡との関連について以下のロジスティック回帰分析を実施した。

モデル 1（未調整モデル）：単変量ロジスティック回帰により粗オッズ比、95%CI の推定。

モデル 2（多変量調整モデル）：共変量を投入したモデルによる調整オッズ比、95%CI の推定。

モデル 3（逆確率重み付け法による感度分析）：傾向スコアを算出し、逆確率重み付け法による調整オッズ比、95%CI の推定。

3.5 結果

3.5.1 対象症例特性の比較

全入院 546,028 件のうち、英語群の入院は 492,316 件（90.2%）、非英語群の入院は 52,937 件（9.7%）、言語記録の欠損は 775 件（0.14%）であった。主要変数の欠損割合は、CCI 0.10%、保険種別 1.71%、婚姻状況 2.49%であり、共変量に欠損のない完全症例 523,039 件（95.8%）を後述の解析対象とした。

対象症例特性を主たる使用言語別に比較した結果を表 2 に示す。非英語群は英語群に比して平均年齢が高く（63.5 歳 vs 56.2 歳, SMD = 0.40）、臨床重症度を示す CCI も有意に高値であった（4.4 vs 3.4, SMD = 0.32）。人種構成では、非英語群においてヒスパニック（29.3% vs 3.4%, SMD = 0.75）およびアジア系（18.9% vs 2.0%, SMD = 0.58）の割合が顕著に高く、白人の割合は低かった（25.6% vs 70.4%, SMD = 1.00）。保険種別では、メディケイドの利用割合が非英語群で著しく高かった（35.9% vs 17.3%, SMD = 0.43）。また、婚姻状況においては非英語群で既婚者の割合が高かった（48.8% vs 41.2%, SMD = 0.15）。院内死亡は、英語群で 2.12%（10,446 件）、非英語群で 2.33%（1,234 件）であった。

表 2: 主たる使用言語別の対象症例特性

変数	全体 (N=546,028)	英語 (N=492,316)	非英語 (N=52,937)	言語欠損 (N=775)	英語vs非英語 SMD
人口統計学的特徴					
年齢（歳）、平均（標準偏差）	56.9 (19.0)	56.2 (19.0)	63.5 (17.7)	62.9 (19.8)	0.40
女性	284,097 (52.0%)	255,135 (51.8%)	28,611 (54.0%)	351 (45.3%)	0.04
人種・民族					
白人	360,519 (66.0%)	346,595 (70.4%)	13,537 (25.6%)	387 (49.9%)	1.00
黒人	89,057 (16.3%)	81,990 (16.7%)	7,032 (13.3%)	35 (4.5%)	0.09
ヒスパニック	32,210 (5.9%)	16,700 (3.4%)	15,492 (29.3%)	18 (2.3%)	0.75
アジア系	19,751 (3.6%)	9,716 (2.0%)	10,014 (18.9%)	21 (2.7%)	0.58
その他 / 不明	44,491 (8.1%)	37,315 (7.6%)	6,862 (13.0%)	314 (40.5%)	0.18
社会経済学的特徴					
保険種別					
メディケア	244,576 (44.8%)	220,002 (44.7%)	24,160 (45.6%)	414 (53.4%)	0.02
民間保険	173,399 (31.8%)	166,966 (33.9%)	6,241 (11.8%)	192 (24.8%)	0.55
メディケイド	104,229 (19.1%)	85,133 (17.3%)	18,997 (35.9%)	99 (12.8%)	0.43
その他	14,006 (2.6%)	11,770 (2.4%)	2,215 (4.2%)	21 (2.7%)	0.10
費用免除	463 (0.1%)	421 (0.1%)	41 (0.1%)	1 (0.1%)	0.00
└ 欠損 / 不明	9,355 (1.71%)	8,024 (1.63%)	1,283 (2.42%)	48 (6.19%)	-
婚姻状況					
既婚	229,134 (42.0%)	202,985 (41.2%)	25,842 (48.8%)	307 (39.6%)	0.15
未婚	206,232 (37.8%)	191,435 (38.9%)	14,634 (27.6%)	163 (21.0%)	0.24
死別	56,687 (10.4%)	48,360 (9.8%)	8,225 (15.5%)	102 (13.2%)	0.17
離婚	40,356 (7.4%)	37,501 (7.6%)	2,817 (5.3%)	38 (4.9%)	0.09
└ 欠損 / 不明	13,619 (2.49%)	12,035 (2.44%)	1,419 (2.68%)	165 (21.29%)	-
臨床及び入院時の特徴					
臨床重症度					
チャールソン併存疾患指数、平均（標準偏差）	3.5 (3.0)	3.4 (3.0)	4.4 (3.1)	4.2 (3.1)	0.32
└ 欠損	531 (0.10%)	483 (0.10%)	47 (0.09%)	1 (0.13%)	-
入院区分					
救急外来経由	177,459 (32.5%)	158,523 (32.2%)	18,691 (35.3%)	245 (31.6%)	0.07
救急観察	119,456 (21.9%)	107,832 (21.9%)	11,548 (21.8%)	76 (9.8%)	0.00
観察入院	84,437 (15.5%)	74,974 (15.2%)	9,427 (17.8%)	36 (4.6%)	0.07
準緊急	54,929 (10.1%)	50,440 (10.2%)	4,167 (7.9%)	322 (41.5%)	0.08
日帰り手術当日入院	42,898 (7.9%)	39,424 (8.0%)	3,425 (6.5%)	49 (6.3%)	0.06
直接観察	24,551 (4.5%)	22,235 (4.5%)	2,310 (4.4%)	6 (0.8%)	0.01
直接救急搬送	21,973 (4.0%)	20,120 (4.1%)	1,830 (3.5%)	23 (3.0%)	0.03
予定入院	13,130 (2.4%)	12,198 (2.5%)	921 (1.7%)	11 (1.4%)	0.05
外来観察	7,195 (1.3%)	6,570 (1.3%)	618 (1.2%)	7 (0.9%)	0.02
主要評価項目					
院内死亡	11,801 (2.16%)	10,446 (2.12%)	1,234 (2.33%)	121 (15.61%)	-

3.5.2 院内死亡に関する解析

完全症例 523,039 件を対象に実施したロジスティック回帰および感度分析の結果を表 3 に示す。モデル 1 においては、非英語群における院内死亡のオッズ比は 1.194（95%CI: 1.121-1.272, P < 0.001）と非英語群で院内死亡オッズが有意に高かった。しかしながら、共変量を投入したモデル 2 では、オッズ比 0.796（95%CI: 0.738 - 0.858, P < 0.001）と逆転し、非英語群において院内死亡オッズが有意に低いという結果が得られた。逆確率重み付け法による推定では、オッズ比 0.928（95%CI: 0.781-1.102, P = 0.393）となり、院内死亡オッズの統計学的有意差は認められなかった。

表 3: 主たる使用言語と院内死亡の関連における解析結果

モデル名	オッズ比 (95% CI)	P値
モデル1: 未調整	1.194 (1.121 - 1.272)	< 0.001
モデル2: 多変量調整*	0.796 (0.738 - 0.858)	< 0.001
モデル3: IPTW 感度分析**	0.928 (0.781 - 1.102)	0.393

*共変量（年齢、CCI、性別、人種、保険種別、婚姻状況、入院区分）を調整。
**傾向スコアに基づく逆確率重み付け法。

4 考察

4.1 一連の RWD 研究プロセスを経験する意義

本研究では、RWD 解析教育を個別の統計解析手法やプログラミング技術の習得に限定せず、研究課題の設定からデータアクセス、研究倫理に関する手続き、統計解析計画書の作成および事前登録、データ抽出・前処理、統計解析、結果の解釈、成果報告までを含む一連の研究プロセスとして位置づけた。このスキームに基づき、MIMIC-IV を用いた実際の観察研究を実施することで、RWD を用いた研究がどのような手順で計画され、実施されるのかを一連の流れとして経験した。RWD 研究では、研究開始時点で研究対象者、曝露、アウトカム、共変量等が必ずしも研究目的に合わせて整理されているわけではなく、研究者自身がデータの生成過程やデータベースの構造を理解した上で、研究仮説に適した研究デザインおよび解析方法を検討する必要がある。そのため、RWD 研究に必要な能力は、統計解析手法やプログラミング技術のみではなく、研究課題を具体的なリサーチクエスチョンに変換し、利用可能なデータから適切な変数を定義し、研究計画に基づいて解析を実施する一連の能力であると考えられる。

特に本研究では、解析開始前に統計解析計画書を作成し、研究対象、主要曝露、主要評価項目、共変量および統計解析方法等を事前に規定した上で、OSF に研究計画を登録した。RWD 研究では、データを確認した後に研究対象や解析方法を変更することが技術的には容易であるため、解析前に研究計画を明確化することは、研究の透明性や再現性を学習する上でも意義があると考えられる。したがって、本スキームの特徴は、単に RWD を解析することではなく、RWD 研究を実際の研究プロセスに沿って経験する機会を提供した点にある。

4.2 MIMIC-IV を教育資源として利用する意義

本研究では、MIMIC-IV を単なる解析対象となるデータセットではなく、RWD 研究を学習するための教育資源として利用した。MIMIC-IV は大規模な電子健康記録データベースであり、入院情報、患者背景、検査、投薬、診断、処置等の複数の情報を含んでいる。そのため、単一の表形式データを解析するだけではなく、複数のテーブルの構造を理解し、研究目的に応じて必要な情報を抽出・結合するという、実際の RWD 研究に近いデータ処理を経験することができる。また、MIMIC-IV の利用には、データアクセスに必要な手続き、研究倫理に関するトレーニング、利用規約への同意等が求められる。このような手続きそのものも、RWD を用いた研究を実施する上で必要となる研究者としての基礎的な能力を学習する機会となる。

すなわち、MIMIC-IV を教育資源として利用することにより、統計解析の実践だけではなく、RWD を研究目的で利用する際の倫理的・手続き的側面についても経験することが可能であった。さらに、本研究では DuckDB および SQL を用いたデータ抽出・結合と、Python を用いたデータ処理および統計解析を組み合わせた。MIMIC-IV のような大規模かつ複数テーブルから構成される RWD を扱うことで、データベースの構造を理解し、研究目的に応じた解析用データセットを構築する能力を実践的に学習できる可能性がある。

4.3 RWD 解析教育における教育機会格差解消への可能性

RWD を用いた研究を実践するためには、実際の医療機関が保有する診療データにアクセスできる環境が重要である。一方、医療機関が保有する診療データには個人情報等が含まれるため、研究目的で利用するためには適切な倫理的・制度的手続きが必要であり、教育機関が自由に利用できるとは限らない。また、電子カルテ等の医療情報システムは医療機関ごとに構成やデータ形式が異なるため、医療機関外の教育機関が診療

データを継続的に利用することには一定の障壁がある。このような環境では、医学部や大学病院等を有し、医療機関との連携基盤を持つ教育機関と、それらを持たない教育機関との間で、RWD を実際に扱う教育機会に差が生じる可能性がある。特に、自施設の医療機関や倫理審査委員会等の研究基盤を持たない教育機関では、実際の診療データを用いた研究を経験することが容易ではない。そのため、RWD 研究を担う人材を育成するためには、特定の医療機関との連携に依存しない教育環境を整備することが重要である。

本研究では、MIMIC-IV のように、所定の手続きをすることで教育・研究目的に利用可能な RWD を活用することで、医療機関を併設していない教育機関においても、RWD 研究の一連のプロセスを実践的に学習する環境を構築できる可能性を提示した。本研究で実施した手続きには、MIMIC-IV へのデータアクセスに必要なトレーニングおよび利用規約への同意に加え、研究倫理教育、統計解析計画書の作成、OSF への事前登録、データ抽出・前処理、統計解析および成果報告が含まれていた。これらを一連の教育プロセスとして経験することで、自施設内の RWD を直接利用することが困難な教育機関であっても、RWD 研究に必要となる研究計画の立案、倫理的配慮、データ管理および統計解析等を学習できる可能性がある。

また、本スキームは特定の地域や教育機関に限定されず、同様のデータアクセス環境を整備することで他の教育機関にも展開できる可能性がある。すなわち、医療機関との物理的な距離や、所属する教育機関が医学部・附属病院を有しているかに依存せず、RWD 研究を実践的に学習する機会を提供できる可能性がある。これは、RWD 解析教育における教育機関間および地域間の教育機会の偏りを縮小し、RWD を活用した研究を担う人材育成を広げるための一つの方法となり得る。

さらに、本スキームは、特定の医療職や専門領域に限定されず、多様な専門職を対象とした RWD 教育に応用できる可能性がある。RWD 研究では、多様な専門職がそれぞれの専門性を生かしてデータを活用することが期待される。MIMIC-IV には診療、検査、投薬、処置、さらには看護記録を含む多様な情報が含まれるため、特定の専門領域に限定されない RWD 教育の教材として利用できる可能性がある。今後、こうした教育環境を複数の教育機関へ展開し、教育効果を検証することで、地域や所属機関による RWD 教育機会の差を縮小し、RWD を活用することができる人材の育成につながることが期待される。

4.4 限界

本研究にはいくつかの限界がある。第一に、MIMIC-IV は米国の医療機関から得られたデータであり、日本の医療制度、診療体制、電子カルテのデータ構造等とは異なる。そのため、MIMIC-IV を用いて習得したデータ処理や解析の方法を、そのまま日本の RWD に適用できるとは限らない。一方で、RWD 研究に共通する研究計画の策定、データ生成過程の理解、変数定義、交絡への対応、解析結果の解釈等については、データベースが異なっても活用できる基礎的能力であると考えられる。第二に、本研究では、RWD 研究の一連のプロセスを実際に経験できる教育スキームを構築・実践することに重点を置いており、その教育効果を客観的・定量的に評価するまでには至っておらず、学習者の RWD に関する知識や研究プロセスへの理解がどの程度向上したかについては明らかではない。今後は、学習者の知識や技能、研究プロセスに対する理解度等を評価し、本教育スキームの教育効果を検証する必要がある。

5. 結論

本研究では、医療機関の RWD への直接的なアクセスが困難な教育環境においても、RWD 研究の一連のプロセスを経験できる教育スキームを構築し、MIMIC-IV を用いた RWD 研究を実践した。データアクセスに必

要な手続き、研究倫理教育、研究計画および統計解析計画書の作成、事前登録、データ抽出・前処理、統計解析、結果の解釈および報告までを一連のプロセスとして経験することで、RWD 研究を学習するための教育環境を構築することが可能であった。MIMIC-IV のように適切な手続きのもとで利用可能な RWD を教育資源として活用することは、自施設に医療データや医療機関との研究基盤を持たない教育環境においても、RWD 研究を学習する機会を提供する上で有用である可能性がある。今後、複数の教育機関や異なる専門領域において本スキームを展開し、その教育効果を検証することで、RWD を適切に利活用できる人材の育成につなげることが期待される。

参考文献

1. 土井麻理子. 日本におけるデータベース研究の現状と今後の展望. 保健医療科学. 2023;72(4):284–92.

2. Hiramatsu K, Barrett A, Miyata Y. Current Status, Challenges, and Future Perspectives of Real-World Data and Real-World Evidence in Japan. Drugs Real World Outcomes. 2021 Jun 19;8(4):459–80.

3. 厚生労働省. 「承認申請等におけるレジストリの活用に関する基本的考え方」について [Internet]. 2021 [cited 2026 Aug 28]. Available from: https://www.mhlw.go.jp/web/t_doc?dataId=00tc5763&dataType=1&pageNo=1

4. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-World Evidence — What Is It and What Can It Tell Us? N Engl J Med. 2016 Dec 8;375(23):2293–7.

5. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017 Sep;26(9):1033–9.

6. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data. 2023 Jan 3;10(1):1.

7. Johnson A, Bulgarelli L, Pollard T, et al. MIMIC-IV (version 3.1). PhysioNet; RRID:SCR_007345.

8. Pollard T, Moody BE, Lehman L, et al. PhysioNet as a global platform for biomedical research. Nat Health. 2026;1(8):792–5.

9. OSF [Internet]. [cited 2026 Sep 18]. Available from: https://osf.io/

10. Foster ED, Deardorff A. Open Science Framework (OSF). J Med Libr Assoc. 2017 Apr;105(2):203–6.

11. von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ. 2007;335(7624):806–8.

SASハッシュオブジェクトとそれを用いたデータ加工テクニック

○兵頭 照将

（イーピーエス株式会社）

SAS Hash Objects and Data Processing Techniques

Terumasa Hyodo
EPS Corporation

要旨

過去 SAS ユーザ総会で発表されてきたハッシュオブジェクトについて、今一度仕様を整理し、その特徴とそれを用いたデータ加工のテクニックを紹介する

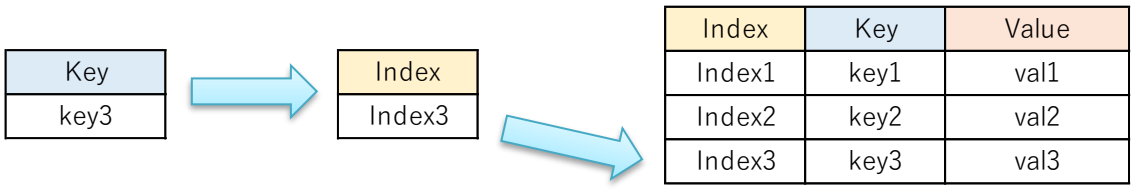
key ワード：ハッシュオブジェクト、データ加工、基準値範囲、アロワンス、SAS

はじめに

SAS のハッシュオブジェクトについては、弊社森岡裕氏の SAS ユーザー総会での紹介等もあり、日常のプログラム開発業務の中でも頻繁に利用されている。SAS ユーザーの中でもかなり浸透しているものとは思われるが、一方でマクロを介しての利用が中心で、ハッシュオブジェクトの操作を自分からコーディングする事はまだ浸透していないと思われる。
今一度、ハッシュオブジェクトについて整理し、実務にて利用可能なハッシュオブジェクトを使用しての加工ロジックについて紹介する。

ハッシュオブジェクトとは

単純に言うと「key」と「value」のペアをセットで管理する、データ構造のオブジェクトである。
一般的なプログラミング言語では、ハッシュテーブル、連想配列、辞書等と呼ばれる。
外部から key を与える事によって value を取り出せる。
その過程に key の値から限りなくユニークに近い値を生成する仕組み（ハッシュ関数）を用いて key から index の値を算出し、index から value を得る。



SAS では、複数変数を組み合わせた key、複数変数の value を格納出来る。
また、このハッシュオブジェクトはある種のテーブルとしても考えられ、データセットを丸ごと取り込む事が容易・データセットへの書き出しも容易である。
その為、柔軟な使い方が可能である。

ハッシュオブジェクト：補足

ハッシュオブジェクトは比較的高速に検索が可能であるが、それは、key の値からハッシュ関数を用いて index の値を直接計算出来る為である。
ただし、偶然にも、違う値の key から同じ index が算出される場合もある。この状況は衝突、コリジョンと呼ばれる。その場合には、AVL 木（平衡探索二分木の一種）を用い一つの index にぶら下がった複数の key-value のペアを元の key の値を用いて探索する。こちらについては図から省略している。
全体のパフォーマンスに影響するものとしては、SAS 特有となるが、ハッシュオブジェクトはデータセットを一気にメモリに展開してそのまま利用する形式になる為、ソートが必要な MERGE に比べて、ファイルに書き出す手間を減らせ、その分高速になる。

ハッシュオブジェクトを使ったデータ加工について

```
data DM ;
  attrib SUBJID length=$4 SEX length=$1 RFSTDTC length=$10 ;
  input SUBJID SEX RFSTDTC ;
cards ;
1001 M 2026-01-01
1002 M 2026-01-02
1003 F 2026-01-03
1004 M 2026-01-04
1005 F 2026-01-05
;run ;

data LB ;
  attrib SUBJID length=$4 LBSTDTC length=$10 LBTESTCD length=$4 VAL length=8  ;
  input SUBJID LBSTDTC LBTESTCD VAL ;
cards ;
1001 2025-11-10 WBC 1000
1001 2026-01-01 WBC 1000
1001 2026-01-07 WBC 1000
1001 2026-01-14 WBC 1002
1001 2026-01-21 WBC 1001
1001 2025-11-10 RBC  500
1001 2026-01-01 RBC  500
1001 2026-01-07 RBC  500
1001 2026-01-14 RBC  501
1001 2026-01-21 RBC  502
1001 2026-01-07 ALP  900
1009 2026-01-07 WBC 1000
;run ;

data TESTMST ;
  attrib LBTESTCD length=$4 LBTEST length=$20 ;
  input LBTESTCD LBTEST ;
cards ;
WBC Leukocyte
RBC Erythrocyte
;run ;

data AWMST ;
  attrib VISIT length=$10 AWTARGET AWLO AWHI length=8 ;
  input VISIT AWTARGET AWLO AWHI ;
cards ;
Screening -30   . -30
Week1       1   1   1
Week2       7   5   9
Week3      14  12  16
Week4      21  19  23
;run ;

data AE ;
  attrib SUBJID length=$4 AESEQ length=8 AETERM length=$10 ;
  input SUBJID AESEQ AETERM ;
cards ;
1001 1 発熱
1001 2 腹痛
1005 1 発熱
;run ;
```

```
data WK01 ;
  set LB ;
  attrib LBDY length=8 ;

  /* 1. 1症例1レコードのデータを結合 */
  if 0 then set DM(keep=SEX RFSTDTC) ;
  if _N_ = 1 then do ;
    declare hash h1(dataset: 'DM', duplicate: 'E');
      h1.definekey('SUBJID') ;
      h1.definedata('SEX', 'RFSTDTC') ;
      h1.definedone() ;
  end ;
  rc1=h1.find() ;
  if rc1 ne 0 then call missing(of SEX RFSTDTC) ;

  if not missing(RFSTDTC) and not missing(LBSTDTC) then do ;
    LBDY = input(LBSTDTC, yymmdd10.) - input(RFSTDTC, yymmdd10.) ;
    if LBDY >= 0 then LBDY = LBDY+1 ;
  end ;

  /* 2. 検査項目マスタの当てはめ */
  if 0 then set TESTMST(keep=LBTEST) ;
  if _N_ = 1 then do ;
    declare hash h2(dataset: 'TESTMST', duplicate: 'E');
      h2.definekey('LBTESTCD') ;
      h2.definedata('LBTEST') ;
      h2.definedone() ;
  end ;
  rc2=h2.find() ;
  if rc2 ne 0 then call missing(of LBTEST) ;

  /* 3. アロワンスのマスタ等keyが設定されない場合の結合 */
  if 0 then set AWMST ;
  if _N_ = 1 then do ;
    declare hash h3(dataset: 'AWMST');
      h3.definekey(All: 'YES') ; /* ダミー */
      h3.definedata(All: 'YES') ;
      h3.definedone() ;
    declare hiter hi3('h3') ;
  end ;

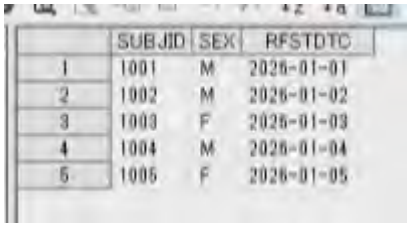
  call missing(of VISIT AWTARGET AWLO AWHI) ;
  if not missing(LBDY) then do ;
    rc3=hi3.first() ;
    do while(rc3=0) ;
      if AWLO = . then do ;
        if LBDY <= AWHI then leave ;
      end ;
      else if AWHI = . then do ;
        if AWLO < LBDY then leave ;
      end ;
      else if AWLO <= LBDY <= AWHI then leave ;
      rc3=hi3.next() ;
    end ;
    if rc3 ne 0 then call missing(of VISIT AWTARGET AWLO AWHI) ;
  end ;
run ;
```

```
data WK02 ;
  set DM ;
  length FEYN $1 ;
  /* 4. 特定条件のレコードの存在有無フラグの追加 */
  if _N_ = 1 then do ;
    rc1=dosubl("proc sql ; create view v1 as select SUBJID from AE where AETERM='発熱' ;
quit ;") ;
    declare hash h1(dataset: 'v1', multidata: 'Y');
      h1.definekey('SUBJID') ;
      h1.definedone() ;
    rc1=dosubl("proc sql ; drop view v1 ; quit ;") ;
  end ;
  rc1=h1.check() ;
  if rc1 ne 0 then FEYN='N' ;
  else FEYN = 'Y' ;
run ;
```

サンプルデータ

DM：

背景情報を格納しているデータセット。1 症例（SUBJID にて特定）に 1 レコードのデータを保持する。



	SUBJID	SEX	RFSTDTC
1	1001	M	2026-01-01
2	1002	M	2026-01-02
3	1003	F	2026-01-03
4	1004	M	2026-01-04
5	1005	F	2026-01-05

LB：

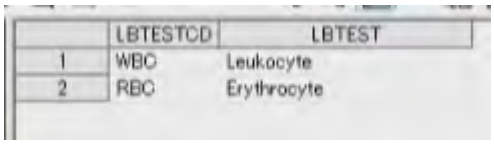
臨床検査情報を格納しているデータセット。1 症例につき複数の検査項目コード（LBTESTCD）および検査日（LBSTDTC）を保持する。



	SUBJID	LBSTDTC	LBTESTCD	VAL
1	1001	2025-11-10	WBC	1000
2	1001	2026-01-01	WBC	1000
3	1001	2026-01-07	WBC	1000
4	1001	2026-01-14	WBC	1002
5	1001	2026-01-21	WBC	1001
6	1001	2025-11-10	RBC	500
7	1001	2026-01-01	RBC	500
8	1001	2026-01-07	RBC	500
9	1001	2026-01-14	RBC	501
10	1001	2026-01-21	RBC	502
11	1001	2026-01-07	ALP	900
12	1009	2026-01-07	WBC	1000

TESTMST：

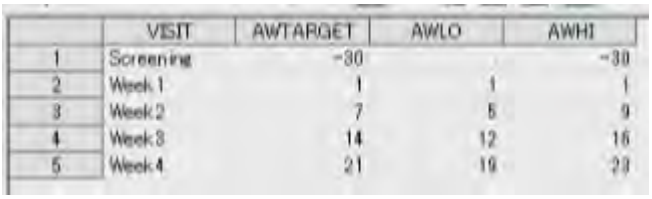
検査項目マスタを格納しているデータセット。検査項目コードと検査項目名（LBTEST）を保持する。



	LBTESTCD	LBTEST
1	WBC	Leukocyte
2	RBC	Erythrocyte

AWMST：

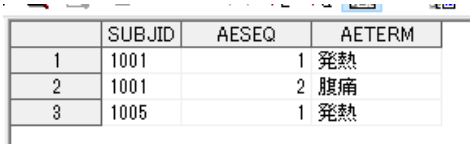
アロワンスのマスタを格納しているデータセット。各週（VISIT）の許容範囲（AW～）を保持する。



	VISIT	AWTARGET	AWLO	AWHI
1	Screening	-30		-30
2	Week 1	1	1	1
3	Week 2	7	5	9
4	Week 3	14	12	16
5	Week 4	21	19	23

AE：

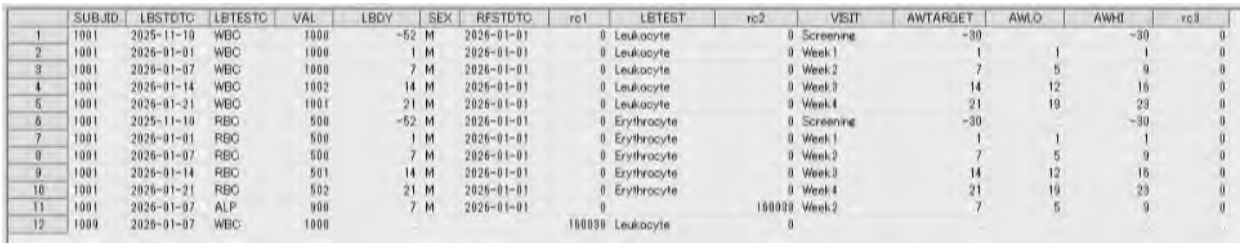
有害事象の情報を格納しているデータセット。症例毎の連番（AESEQ）と有害事象名（AETERM）を保持する。



	SUBJID	AESEQ	AETERM
1	1001	1	発熱
2	1001	2	腹痛
3	1005	1	発熱

WK01：

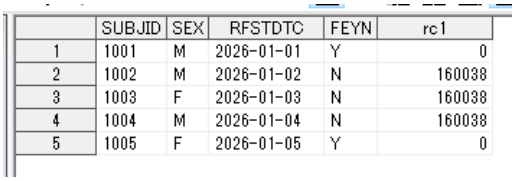
1.2.3.のデータ加工結果(後述)のデータセット。



	SUBJID	LBSTDTC	LBTESTC	VAL	LBDY	SEX	RFSTDTC	rc1	LBTEST	rc2	VISIT	AWTARGET	AWLO	AWHI	rc3
1	1001	2025-11-10	WBC	1000	-52	M	2026-01-01	0	Leukocyte	0	Screening	-30		-30	0
2	1001	2026-01-01	WBC	1000	1	M	2026-01-01	0	Leukocyte	0	Week 1	1	1	1	0
3	1001	2026-01-07	WBC	1000	7	M	2026-01-01	0	Leukocyte	0	Week 2	7	5	9	0
4	1001	2026-01-14	WBC	1002	14	M	2026-01-01	0	Leukocyte	0	Week 3	14	12	16	0
5	1001	2026-01-21	WBC	1001	21	M	2026-01-01	0	Leukocyte	0	Week 4	21	19	23	0
6	1001	2025-11-10	RBC	500	-52	M	2026-01-01	0	Erythrocyte	0	Screening	-30		-30	0
7	1001	2026-01-01	RBC	500	1	M	2026-01-01	0	Erythrocyte	0	Week 1	1	1	1	0
8	1001	2026-01-07	RBC	500	7	M	2026-01-01	0	Erythrocyte	0	Week 2	7	5	9	0
9	1001	2026-01-14	RBC	501	14	M	2026-01-01	0	Erythrocyte	0	Week 3	14	12	16	0
10	1001	2026-01-21	RBC	502	21	M	2026-01-01	0	Erythrocyte	0	Week 4	21	19	23	0
11	1001	2026-01-07	ALP	900	7	M	2026-01-01	0		160038	Week 2	7	5	9	0
12	1009	2026-01-07	WBC	1000				160038	Leukocyte	0					

WK02：

4.のデータ加工結果(後述)に該当するデータセット



	SUBJID	SEX	RFSTDTC	FEYN	rc1
1	1001	M	2026-01-01	Y	0
2	1002	M	2026-01-02	N	160038
3	1003	F	2026-01-03	N	160038
4	1004	M	2026-01-04	N	160038
5	1005	F	2026-01-05	Y	0

データ加工ロジックの説明

1.　一症例一レコードのデータの結合

治験に置ける各検査データ、有害事象等のイベントの情報に、背景情報を結合させるようなデータ加工はしばしば発生する。例として、SDTM を作成する場合での元データの LB や AE 等への DM の結合、ADaM を作成する場合での LB や AE への ADSL の結合等が考えられる。

通常の MERGE BY での結合と異なり、事前ソートを必要とせず、またサンプルデータ LB に結合したい DM 内の変数については definekey()メソッド、definedata()メソッド、if 0 then set~ にて定義しておけば、find()メソッドのみで簡単に取得される。

また、この際に、特にロジック制御されない場合、LB の各レコードについて保持され、LB に存在しない DM のデータ等は取得されないところも異なる。

使用される構文

・declare hash <ハッシュオブジェクト名> (<オプション>) ;

ハッシュオブジェクトの名称の宣言。カッコ内のオプションを指定する事で、先の宣言に加えハッシュオブジェクトのインスタンスが生成される。「duplicate: 'E'」は、definekey で定義される key がハッシュオブジェクト内で重複した際に、ERROR とする場合に設定する（指定しない場合には最初に格納された key,value が保持される）。

・<ハッシュオブジェクト名>.definekey('<key 変数名>') ;

ハッシュオブジェクトを検索する為の key 変数を定義する。変数名は引用符で囲み、複数ある場合にはカンマ区切りで指定する。

・<ハッシュオブジェクト名>.definedata('<value 変数名>') ;

ハッシュオブジェクトから value として取得する変数を定義する。

・<ハッシュオブジェクト名>.definedone() ;

定義を終了し、ハッシュオブジェクトの初期化を完了する。

・<リターンコードを格納する変数>=<ハッシュオブジェクト名>.find();

key 変数よりハッシュオブジェクトを介して検索が実行され、存在すれば value 変数に値が格納され、またリターンコードには 0 が格納される。存在しない場合には、リターンコードには 0 以外の値が格納される。

2.　検査項目マスタの当てはめ

検査項目コードに紐づいた形で検査項目名等を格納したいケースもしばしば発生する。この目的で別手法として、SAS のユーザ定義フォーマットを利用する場合もあるが、ユーザ定義フォーマットの場合、フォーマット名の長さ制限、また原則結合元の変数と値を格納する変数はそれぞれ 1 つしか指定出来ない為、柔軟性に欠ける。MERGE BY であれば同様のロジックを実装する場合に、事前に検査項目コードでソートしなければならない。

こちらも 1.の場合と同様の手法で実行出来る。これらのデータ加工については汎用マクロとして、PhamaForest にて%kvlookup()（sashash パッケージ内）が提供されており、スマートな実装が可能である。

3.　アロワンスのマスタ等 key が設定されない場合の結合

1.および 2.では、key の指定によって 1 レコードのみ特定であったが、日付が特定期間にあるかどうかを検索するような、マッチングが不等号を含む条件の場合、key によるマッチングは使えず、ハッシュオブジェクトを複数行にわたり検索していく必要がある。

例としてアロワンスにしたがって VISIT を設定していく場合を想定した。

当該手法は PROC SQL の LEFT JOIN、ON 句に指定する条件に不等号を含める事が出来る事を利用しての代替も可能であるが、時点によって LBDY だけでなく他変数の利用を行う場合等、条件が複雑化する場合には SQL では実装するのに苦労する事がある。他、ARRAY では文字変数と数値変数の混在するケースを同時に同時に処理しづらい。その場合に当該手法が向いていると考える。

使用される構文

・<ハッシュオブジェクト名>.definekey(All: 'Y') ;

ハッシュオブジェクトを検索する為の key 変数として全ての変数を設定する。実際には find()メソッドを用いないが、key 変数の指定は必要な為ダミーで設定している。

・<ハッシュオブジェクト名>.definedata(All: 'Y') ;

ハッシュオブジェクトから value として取得する変数として全ての変数を設定している。

・declare hiter <ハッシュ反復子オブジェクト名>('<ハッシュオブジェクト名>') ;

ハッシュオブジェクトを元にハッシュ反復子オブジェクトを設定する。

・<ハッシュ反復子オブジェクト名>.first()

ハッシュ反復子オブジェクトの先頭レコードに移動する

・<リターンコードを格納する変数>=<ハッシュ反復子オブジェクト名>.next();

ハッシュ反復子オブジェクトの現在の次のレコードに移動する。レコードが存在すれば value 変数に値が格納され、またリターンコードには 0 が格納される。存在しない場合には、リターンコードには 0 以外の値が格納される。

4.　特定条件のレコードの存在有無フラグの追加

ハッシュオブジェクトに読み込む側のデータが key に対して複数レコードを持つ場合には複数データの取り込みは通常不可であるがオプションを指定する事によってハッシュオブジェクトに取り込める。

また、dosubl 関数によって、DATA ステップ実行中に別途 PROC SQL を利用する事が可能。

サンプルでは AETERM が「発熱」となるレコードが症例毎に存在するかを check()メソッドにて確認しフラグ変数に格納する。

使用される構文

・<リターンコードを格納する変数>=dosubl('<実行したいコード>') ;

DATA ステップの実行中に別の SAS コードを実行する。実行が成功された場合には 0 を返す。

・declare hash <ハッシュオブジェクト名> (<オプション>,multidata: 'Y') ;

宣言自体は同様。他オプションに加え、「multidata: 'Y'」を指定すると同じ key の値を持つレコードが複数存在するのを可とする（通常は不可）。

・<リターンコードを格納する変数>=<ハッシュオブジェクト名>.check();

key 変数よりハッシュオブジェクトを介して検索が実行され、存在すればリターンコードには 0 が格納される。存在しない場合には、リターンコードには 0 以外の値が格納される。find()メソッドと異なり、value 変数が不要。
こちらも、PhamaForest にて%kvcheck()（sashash パッケージ内）が提供されている。

ハッシュオブジェクトを用いたデータ加工ロジックの利点

ハッシュオブジェクトは DATA ステップ内の実行フェーズにおいて実行され、動的に生成される。その為、既存の DATA ステップの場合のように、定義についてはどこに記載していても最初に一回だけ実行されるという訳ではない。また、構文等に違いがある為、通常の SAS の構文に慣れているプログラマーにとっては新たに覚える事も多い。
また、ハッシュオブジェクトはメモリ上に載せられるサイズに限定されるという制限はある。
しかし、例えば、異なる key を持つ複数のデータセットを結合する場合、MERGE BY による結合では其の都度ソート及び別々に MERGE する必要がある。PROC SQL による結合の場合にも副問合せを行い可読性が下がる。ハッシュオブジェクトの場合には、1 つの DATA ステップ内で利用する事が可能である。
ハッシュオブジェクトに格納するデータセットについてはサイズに多少注意が必要なものの、ベースとするデータセットのサイズが大規模なものであるほど、例えば臨床検査データ等は検査項目、検査時点の数の多さから長大になりやすいが、ソートに非常に時間を有する為、MERGE 結合に比べハッシュオブジェクトを用いた結合の方が圧倒的に高速である。
また、特に output ステートメント等で出力制御を行わない限り、ハッシュオブジェクトを用いる場合はベースとなるデータセットのレコード、並び順を保持している。
その他、投与後日数から解析週を求めるような場合についても、条件に使われる数値をデータセット化してハッシュオブジェクトとして取り込む事により、DATA ステップ内の複数の条件式で記述するのに比べて少量のコードで容易に利用出来る。こちらは、ARRAY でも同様の事が可能であるが、ARRAY が文字のみの配列か数値のみの配列に制限されるのに比べて柔軟に対応可能である。

結語

SAS ハッシュオブジェクトについて簡単にはなるが説明を行い、またデータ加工テクニックについて数パターンを解説した。
当該資料が、日々の開発業務の助けになれば幸いである。

参考文献：

【過去ユーザー総会でのハッシュオブジェクト系の演題】

2014 SAS ハッシュオブジェクトを利用して医薬品開発に使用するプログラムを効率化するー有害事象と併用薬、臨床検査値と途中変更のある施設基準値 のマッチングから SAS プログラムコードの分析まで 森岡裕，神田悟志(ナイフィックス株式会社)

2015 SAS ハッシュオブジェクトを利用したデータ集計ー ア イテムセットのカウントを例として ー 森岡裕(株式会社日経リサーチ)

2018 DOSUBL 関数内で生成したビューをハッシュオブジェクトに格納することにより PROC DS2 のハッシュパッケージの挙動を擬似的に再現する方法 森岡裕(イーピーエス株式会社)

2019 FCMP の STATIC statement, HASH object, DICTIONARY object それぞれによる LAG 関数機能の定義 森岡裕(イーピーエス株式会社)

2023 前処理大全 SAS バージョン 森岡裕(イーピーエス株式会社)

2024 SAS ハッシュオブジェクト基礎編 松永友貴(ノバルティス ファーマ株式会社)

【 SC19-2006 】

2006 How to Implement the SAS® DATA Step Hash Object Bill Parman, BlueCross-BlueShield of Tennessee, Chattanooga, TN

https://www8.sas.com/scholars/Proceedings/2006/SC/SC19_06.PDF

データステップとPDV入門

－ SASコードはどのような仕組みで動いているのか

○角　春佳

（イーピーエス株式会社）

Introduction to the DATA Step and PDV: Understanding the Mechanics Behind SAS Code

Haruka Sumi
(EPS Corporation)

要旨

SAS データステップのプログラムでは、変数の長さの宣言、出力する変数の選択、変数名の変更といった処理を頻繁に行う。このような処理は、上から順に書けば動くという理解でも利用できるため、内部の動作を十分に意識しないまま使われることも少なくない。発表者も同様であった。

本発表では、これら決まり文句と見なされがちなコードの挙動が、実は PDV(Program Data Vector)と、コンパイル・実行という仕組みによって決定されることを、具体的なコード例を通じて確認する。length、keep、Rename などを題材に、コードの実行結果を予想しながら、内部動作を一つずつ確かめていく。ばらばらの決まりごととして処理を覚えるのではなく、PDV という一つの仕組みから捉え直すことで、正確な挙動を理解し、堅牢なコードを記述できるようになることを目指す。

習慣的に記述しているコードの「なぜ」に関心のある方、また他言語から SAS に移行し、その挙動に戸惑っている方を主な対象とした入門的な内容である。

キーワード：データステップ、PDV、コンパイルフェーズ、実行フェーズ

1. はじめに

SAS 言語の特徴として、データを 1 行ずつ順に読み込んで処理をするデータステップが挙げられる。データステップ中では、代入文や IF 文等の一般的なステートメントに加え、変数の属性や出力そのものを制御する LENGTH、KEEP/DROP、RENAME 等の SAS 言語独自のステートメントを使用する。発表者はこれらのステートメントを日常的に用いてきたが、その内部動作を正確に説明できるかと問われると、明確には答えられなかった。

例えば、「LENGTH ステートメントは SET ステートメントより前に書く」という慣習に従ってコードを書いてきたが、なぜそう書かなければならないのか、また、この慣習を守らなかった場合に具体的にどのような不都合が生じるのか、これまで意識したことはなかった。

こうした反省を踏まえ、SAS のデータステップの内部処理について詳細に検証した。その結果、これらのステートメントの動作は、SAS のデータステップが持つ、(1)プログラムデータベクター（Program Data Vector, 以下 PDV）と呼ばれる仕組みと、(2)コンパイルフェーズ・実行フェーズという 2 段階の処理構造に着目することで、統一的に説明できることが分かった。本稿は、その検証の過程を整理したものである。

2 章でデータステップの基本構造である PDV およびコンパイル・実行フェーズについて整理する。続く 3 章では、LENGTH、KEEP/DROP、RENAME の 3 つのステートメントを具体的な事例として取り上げ、それぞれの動作をコンパイル・実行フェーズという枠組みと、PDV の状態から確認する。4 章では、これらの事例に共通する動作の規則を整理する。

2. PDV およびコンパイル・実行フェーズ

2.1. データステップとプログラムデータベクター（PDV）とは

SAS におけるデータステップとは、DATA ステートメントに始まり、RUN ステートメントに至るまでの一連の記述を指す。なお、RUN ステートメントを省略した場合は、次の DATA ステートメントや PROC ステートメントが現れた時点が区切りとなる。SAS のデータステップの処理の流れを図 1 に示す。データステップは、PDV と呼ばれる仕組みに基づき動作している[1]。PDV とは、データステップが扱うすべての変数について、その名前・型・長さといった属性と現在の値を保持する、メモリ上の作業領域である。なお、SAS データセットの 1 行をオブザベーションと呼ぶ。入力データセットの値について、PDV が保持するのは、1 オブザベーション分である。

SAS は、プログラム内の(1)SET ステートメントに達すると、(2)入力データセットの 1 行目のオブザベーションを PDV に読み込み、SAS コードを実行し、(3)PDV の内容を出力データセットに出力する（図 1）。その後、入力データセットの 2 行目のオブザベーションを PDV に読み込み、同様の処理を行う。これを入力データセットの最後のオブザベーションまで繰り返す。なお図 1 の例では「if Name = "Alfred";」という条件を記述しているため、実際に出力されるのは条件を満たしたオブザベーションのみであり、条件に合わないオブザベーションは出力されないまま次の読み込みに移る。このように、データステップはデータセット全体を一度にまとめて処理するのではなく、1 オブザベーションずつ処理を行う行指向の仕組みである。この「PDV が作業領域として繰り返し使われる」という性質が、データステップの大きな特徴である。

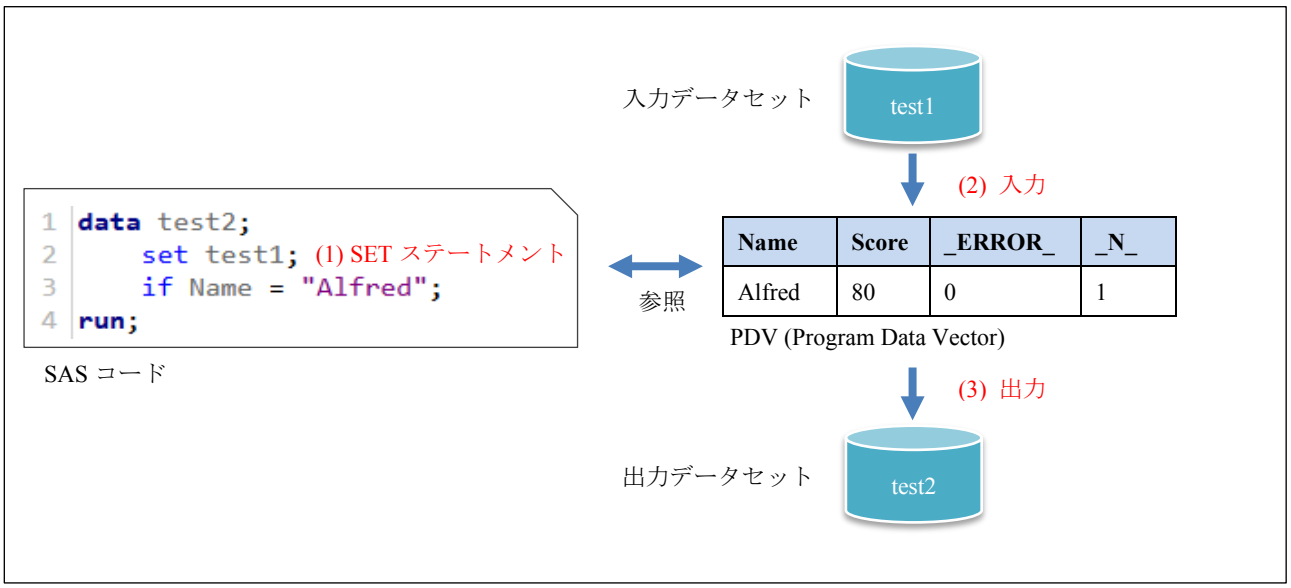


図 1　データステップの流れ

例として、図 2(a)のコードを考える。入力データセットとして、SAS に標準で用意されているサンプルデータセット sashelp.class を用いる（図 2(b)）。以降のコード例でも同じデータセットを用いる。

```
1 data test;
2     set sashelp.class;
3     Age_x2 = Age * 2;
4 run;
```

(a) SAS コード



入力データセット

	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69	112.5
2	Alice	F	13	56.5	84
3	Barbara	F	13	65.3	98
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

(b) 入力データセット sashelp.class の中身。5 行目までを表示

図 2　SAS コード例と入力データセット

上記のコードを実行すると、入力データセットから取得した変数（Name・Sex・Age・Height・Weight）と、新たに作成される変数（Age_x2）の型、長さといった属性が PDV に保持される。入力データセットの 1 行目のオブザベーションが PDV に読み込まれ、Age_x2 の計算が完了した後、PDV は図 3(a)の状態になる。コードの実行がデータステップ最終行まで進むと、PDV の内容が出力データセットへ出力され、SAS は次のオブザベーションの読み込みに移る。1 行目の処理と同様に入力データセットの 2 行目のオブザベーションが PDV に読み込まれ、計算が完了すると、PDV は図 3(b)の状態になる。この内容も出力データセットへ出力される。この処理は入力データセットの最後のオブザベーションまで同様に繰り返され、最終的に図 3(c)に示す出力データセット test が得られる。なお、以降の図で示す PDV の表現は、PDV の内部状態を分かりやすく示すための模式図であり、SAS 内部の実装を正確に反映したものではない。

PDV

Name	Sex	Age	Height	Weight	Age_x2	_ERROR_	_N_
Alfred	M	14	69	112.5	28	0	1

(a) 入力データセット 1 行目のオブザベーション処理後の PDV の状態

PDV

Name	Sex	Age	Height	Weight	Age_x2	_ERROR_	_N_
Alice	F	13	56.5	84	26	0	2

(b) 入力データセット 2 行目のオブザベーション処理後の PDV の状態



出力データセット

	Name	Sex	Age	Height	Weight	Age_x2
1	Alfred	M	14	69	112.5	28
2	Alice	F	13	56.5	84	26
3	Barbara	F	13	65.3	98	26
4	Carol	F	14	62.8	102.5	28
5	Henry	M	14	63.5	102.5	28

(c) 出力データセット test の中身。5 行目までを表示

図 3　1 行目、2 行目のオブザベーション処理後の PDV の状態および出力結果

本稿では、LENGTH、KEEP/DROP、RENAME 等のステートメントが、PDV に対してどの段階でどのような影響を与えるのかという観点から、それぞれの動作を確認する。

このうち「どの段階で」にあたるコンパイルフェーズと実行フェーズについては次節で説明し、「どのような影響を与えるか」の具体的な内容は、3 章で事例とともに確認する。

2.2. コンパイルフェーズと実行フェーズ

SAS コードを実行環境に提出（Submit）しても、SAS は Submit されたコードを即座に上から順に実行するわけではない。はじめに SAS はデータステップの構文をチェックし、実行可能な内部形式に変換する。この処理をコンパイルフェーズと呼ぶ[1]。コンパイルフェーズは 1 つのデータステップに対して一度だけ行われる。この過程において PDV が構築され、PDV を構成する変数とその属性が決定される。この段階では入力データセットの値は PDV に読み込まれていない。LENGTH、KEEP/DROP、RENAME 等のステートメントは、いずれもこのコンパイルフェーズで解釈され、PDV の構造や出力の設定に反映される。

コンパイルフェーズが終わると実行フェーズに移る。実行フェーズとは、SET ステートメントが入力データセットから次のオブザベーションを読み込めなくなるまで繰り返されるループ処理である[1]。各繰り返しの開始時には、データステップ内の代入文等で作成された変数の値が欠損値に初期化される。SET ステートメントで読み込まれる変数は初期化されない。その後、SET ステートメントによって入力データセットの 1 オブザベーションが PDV に読み込まれ、IF 文や代入文等の処理文が、コードに記述された順序どおりに実行されることで、PDV 上の値が更新される。

IF 文や代入文等、実行フェーズで動作する処理文は、記述された順序がそのまま結果に反映される。しかし、LENGTH、KEEP/DROP、RENAME 等、コンパイルフェーズで処理されるステートメントは、ステートメントが記述されている位置によって結果が変わるものと、変わらないものがある。この違いがなぜ生じるのかについては、3 章で事例とともに詳しく確認する。

なお、本稿では SAS データセットを読み込む SET ステートメントのみを扱う。外部ファイルを読み込む INPUT ステートメントは、入力バッファ（Input Buffer）という別の領域を経由する点に留意されたい。

3. 事例を用いて内部動作を確認

本章では、コンパイルフェーズで処理されるステートメントのうち、LENGTH、KEEP/DROP、RENAME の 3 つを取り上げる。これらに限定するのは、いずれも日常的に用いられる一方、その動作の理由が意識されにくいためである。

2 章で述べたとおり、これら 3 つはいずれもコンパイルフェーズで処理されるという共通点を持つ。しかし、ステートメントを使用した結果が記述する位置に依存するかどうかは、ステートメントによって異なる。本章では、まず LENGTH ステートメントを例に、書く位置によって結果が変わる事例を確認する（3.2）。続いて、KEEP/DROP および RENAME ステートメントを例に、書く位置によらず結果が一定となる事例を確認し、その違いがどこから生じるのかを考察する（3.3、3.4）。なお、本稿の SAS コードは Base SAS 9.4M8 で実行した結果に基づく。

3.1. PDV の確認方法: PUT _ALL_

PDV は通常、プログラマの目に触れることはない。しかし、PUT ステートメントに_ALL_を指定した PUT _ALL_を用いることで、その時点で PDV に存在するすべての変数と値をログに出力させることができる。_ERROR_や_N_のように、プログラマが明示的に宣言していない自動変数も、PDV に実在するためログに出力される。なお、PUT _ALL_が示すのは、それが記述された時点での PDV の状態である。したがって、コード中のどこに PUT _ALL_を配置するかによって出力結果が異なる点に注意されたい。

配置する位置によって PDV の出力結果が異なることを確認するため、図 4(a)に示すコード中の(1)～(3)の 3 か所に PUT _ALL_を配置し、それぞれの時点で PDV がどのような状態になっているかを比較する。

(1)の時点では、SAS データセットからの読み込みが行われていないため、_ERROR_と_N_を除くすべての変数が欠損値となる（図 4(b)）。(2)では、SET ステートメントによって Name, Sex, Age, Height, Weight の値が PDV に読み込まれている（図 4(c)）。Age_x2 はまだ計算されていないため欠損値のままである。(3)になって初めて、Age_x2 の計算が行われ、全ての変数に値が格納された状態となる（図 4(d)）。

このように、PDV には、実行フェーズの中でコードが上から実行されるにつれて段階的に値が格納されていく。PUT _ALL_を用いれば、この過程を任意の時点で確認できる。以降では、この方法を用いて PDV が保持する値の変化に着目し、各ステートメントの動作を確認していく。

```
1 data test;
2     put _ALL_;          <-- (1)
3     set sashelp.class;
4     put _ALL_;          <-- (2)
5     Age_x2 = Age * 2;
6     put _ALL_;          <-- (3)
7 run;
```

(a) SAS コード

PDV

Name	Sex	Age	Height	Weight	Age_x2	_ERROR_	_N_
		.	.	.	.	0	1

(b) (1)の時点での PDV の状態

PDV

Name	Sex	Age	Height	Weight	Age_x2	_ERROR_	_N_
Alfred	M	14	69	112.5	.	0	1

(c) (2)の時点での PDV の状態

PDV

Name	Sex	Age	Height	Weight	Age_x2	_ERROR_	_N_
Alfred	M	14	69	112.5	28	0	1

(d) (3)の時点での PDV の状態

図 4　データステップの SAS コード例と PDV の状態。表中、「.」は数値変数の欠損値を表す。空白は文字変数の欠損値を表す。

3.2. LENGTH ステートメント: 書く位置によって実行結果が異なる

LENGTH ステートメントは、変数に格納する値の長さを指定するためのステートメントである。文字変数の場合、この長さは格納できるバイト数を意味する。なお、数値変数に対する LENGTH ステートメントは、文字変数とは異なる規則に従うため、本稿では文字変数を対象とした場合について扱う[2]。LENGTH ステートメントの対象となる変数は、入力データセットから読み込まれる既存の変数と、データステップ内で新たに作成される変数に分けられる。以降、前者を既存変数、後者を新規変数と呼ぶ。

LENGTH ステートメントは SET ステートメントより前に記述するのが慣習とされる。しかし、なぜそうしなければならないのか、そして SET ステートメントの後に記述すると具体的にどのような不都合が生じるのかを、正確に説明できるだろうか。以下では、PDV の観点から、それらの動作を確認する。

(1) 既存変数の場合: SET ステートメントの後に LENGTH ステートメントを記述すると、何が起きるか

図 5(a)のコードを実行すると、SAS は WARNING を出力し（図 5(b)）、LENGTH ステートメントの指定は反映されない。

```
1 data test;
2     put _all_;          <-- (1)
3     set sashelp.class;
4     length Name $ 5;
5     put _all_;          <-- (2)
6 run;
```

(a) SET ステートメントの後に LENGTH ステートメントを記述したコード

```
WARNING: 文字変数Nameの長さはすでに設定されています。文字変数の長さを宣言するには、
         DATAステップの最初にLENGTHステートメントを使用してください。
```

(b) ログに出力される WARNING

出力データセット

	Name	Sex	Age	Height	Weight
1	Alfred	M	14	69	112.5
2	Alice	F	13	56.5	84
3	Barbara	F	13	65.3	98
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

(c) 出力データセット test の中身。5 行目までを表示

図 5　SET ステートメントの後に LENGTH ステートメントを記述した SAS コード例と出力結果

このとき、Name の属性がどのように扱われているのかを下記の PDV の状態から確認する。図 6(a)は、図 5(a)に示すコード中の(1)、すなわち SET ステートメントが実行される前の PDV の状態を示す。この時点で、Name の属性は長さ 8 として確定している。続く図 6(b)は、図 5(a)に示すコード中の(2)、すなわち LENGTH ステートメントの記述位置より後の PDV の状態であるが、Name の長さは 8 のままであり、LENGTH ステートメントで指定した 5 という長さは反映されていない。つまり、SET ステートメントによってすでに属性が確定した既存変数に対して、後から LENGTH ステートメントを記述してもその変数の属性や出力結果は変更されない。なぜこのようになるのか。理由はコンパイルフェーズにおいて、SAS がコードを上から順にスキャンし、変数が最初に参照された箇所の属性を採用する、という原則にある。上記の例では、SET ステートメントが LENGTH ステートメントより先に記述されているため、Name の長さは SET ステートメントの時点で入力データセット内の定義に基づいて確定してしまう。後続の LENGTH ステートメントは、既に確定した属性を変更することができない。

PDV

Name	Sex	Age	Height	Weight	_ERROR_	_N_
		.	.	.	0	1
Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8

(a) 図 5(a)の(1)の PDV の状態

PDV

Name	Sex	Age	Height	Weight	_ERROR_	_N_
Alfred	M	14	69	112.5	0	1
Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8

(b) 図 5(a)の(2)の PDV の状態

図 6　図 5(a)コード実行時の PDV の状態

(2) 既存変数の場合: LENGTH ステートメントを先に記述すると、指定した長さが反映される

LENGTH ステートメントを SET ステートメントより前に記述すれば、この問題は回避できる（図 7(a)）。図 7(b)はコンパイルフェーズから実行フェーズに移った後、SET ステートメントが実行されるより前の PDV の状態を示しており、Name の属性は長さ 5 となっている。また、図 7(c)は SET ステートメントによって 1 行目のオブザベーションが読み込まれた後の PDV の状態を示しているが、Name の値は"Alfred"ではなく、5 バイトに切り詰められた"Alfre"として格納されている。これは、LENGTH ステートメントが SET ステートメントより前にあることで、コンパイルフェーズの段階で Name の長さが 5 として確定し、その後、実行フェーズにおいて SET ステートメントで読み込まれる値が、長さ 5 に合わせて切り詰められたことを示している。これが「LENGTH ステートメントは SET ステートメントより前に書く」という慣習の根拠である。

```
1 data test;
2     put _all_;   <-- (1)
3     length Name $ 5;
4     set sashelp.class;
5     put _all_;   <-- (2)
6 run;
```

(a) SET ステートメントの前に LENGTH ステートメントを記述したコード

PDV

Name	Sex	Age	Height	Weight	_ERROR_	_N_
		.	.	.	0	1
Length: 5	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8

(b) (a)の(1)の PDV の状態

PDV

Name	Sex	Age	Height	Weight	_ERROR_	_N_
Alfre	M	14	69	112.5	0	1
Length: 5	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8

(c) (a)の(2)の PDV の状態

出力データセット

	Name	Sex	Age	Height	Weight
1	Alfre	M	14	69	112.5
2	Alice	F	13	56.5	84
3	Barba	F	13	65.3	98
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

(d) 出力結果

図 7　LENGTH ステートメントを SET ステートメントより前に記述した場合の SAS コード例と PDV の状態および出力結果

(3) 新規変数の場合: LENGTH ステートメントを指定しないと何が起きるか

データステップ内で新たに作成される変数についても、LENGTH ステートメントを明示的に指定しない限り、最初の代入に用いられた値の長さで、その変数の長さが確定する。以下のコード例（図 8(a)）では、Grade という変数に LENGTH の指定がない。

```
1 data test;
2     put _all_;          <-- (1)
3     set sashelp.class;
4     if Name = "Alfred" then Grade = "A";
5     else Grade = "Average";
6     put _all_;          <-- (2)
7 run;
```

(a) LENGTH ステートメントを明示的に指定しないコード

図 8　LENGTH ステートメントを明示的に指定しない SAS コード例

コンパイルフェーズにおいて、コードを上から構文チェックする過程で最初に現れる「Grade = "A"」によって、Grade の長さは 1 に確定する（図 9(a)、図 9(b)）。以降のオブザベーションで代入される"Average"はこの長さ 1 に切り詰められ、出力データセットでは Grade の値がすべて"A"となる（図 9(c)）。

PDV

Name	Sex	Age	Height	Weight	Grade	_ERROR_	_N_
		.	.	.		0	1
Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 1	Length: 8	Length: 8

(a) 図 8 の(1)の PDV の状態

PDV

Name	Sex	Age	Height	Weight	Grade	_ERROR_	_N_
Alfred	M	14	69	112.5	A	0	1
Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 1	Length: 8	Length: 8

(b) 図 8 の(2)の PDV の状態

出力データセット

	Name	Sex	Age	Height	Weight	Grade
1	Alfred	M	14	69	112.5	A
2	Alice	F	13	56.5	84	A
3	Barbara	F	13	65.3	98	A
4	Carol	F	14	62.8	102.5	A
5	Henry	M	14	63.5	102.5	A

(c) 出力結果

図 9　LENGTH ステートメントを明示的に指定しない場合の PDV の状態および出力結果

(4) 新規変数の場合: LENGTH ステートメントをあらかじめ指定することで、指定した長さが反映される

この問題も、代入文の前に LENGTH ステートメントで仕様に即した長さを確保しておくことで回避できる（図 10）。

```
1 data test;
2     length Grade $7;
3     put _all_;          <-- (1)
4     set sashelp.class;
5     if Name = "Alfred" then Grade = "A";
6     else Grade = "Average";
7     put _all_;          <-- (2)
8 run;
```

(a) LENGTH ステートメントを明示的に指定した場合のコード

PDV

Grade	Name	Sex	Age	Height	Weight	_ERROR_	_N_
			.	.	.	0	1
Length: 7	Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8

(b) (a)の(1)の PDV の状態

PDV

Grade	Name	Sex	Age	Height	Weight	_ERROR_	_N_
A	Alfred	M	14	69	112.5	0	1
Length: 7	Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8

(c) (a)の(2)の PDV の状態

出力データセット

	Grade	Name	Sex	Age	Height	Weight
1	A	Alfred	M	14	69	112.5
2	Average	Alice	F	13	56.5	84
3	Average	Barbara	F	13	65.3	98
4	Average	Carol	F	14	62.8	102.5
5	Average	Henry	M	14	63.5	102.5

(d) 出力結果

図 10　LENGTH ステートメントを明示的に指定した場合の SAS コード例と PDV の状態および出力結果

(5) 小括

以上 2 つの起こりやすい現象、すなわち、既存変数に対して LENGTH ステートメントを後から記述しても長さが変わらないことと、新規変数の長さが最初に現れた値によって確定してしまうことは、いずれも「文字変数の属性は、コンパイルフェーズにおいて、その変数が最初に参照された箇所で確定する」という共通のルールから説明できる。

LENGTH ステートメント自体はコンパイルフェーズで処理されるが、その処理が PDV の属性に反映されるのは、LENGTH ステートメントが対象変数への最初の参照となる場合に限られる。

「LENGTH ステートメントは SET ステートメントより前に書く」という慣習は、この既存変数の制約に合わせて、常に安全側で運用するための習慣であると理解できる。

3.3. KEEP/DROP ステートメント: 書く位置によらず出力結果が同じ

LENGTH ステートメントとは対照的に、KEEP/DROP ステートメントおよび RENAME ステートメントは、データステップのどこに記述しても結果が変わらない。本節と次節では、この 2 つのステートメントを例に、なぜ位置に依存しないのかを PDV の動作から確認する。

KEEP ステートメントは、出力データセットに含める変数を指定するステートメントであり、DROP ステートメントは、出力データセットから除外する変数を指定するステートメントである[3][4]。多くのプログラマは、これらをデータステップの最後にまとめて記述することが多いのではないだろうか。そのため、「KEEP ステートメントはデータステップの最後に書くものだ」、「対象の変数を計算するより前に書くと、正しく動作しないのではないか」と感じている人もいるかもしれない。

実際には、KEEP/DROP ステートメントはデータステップ内のどこに記述しても、結果は変わらない。以下の 2 つのコードは、KEEP ステートメントを書く位置が異なるにもかかわらず、全く同じ結果を出力する[3][4]。

```sas
data test1;
    set sashelp.class;
    Age_x2 = Age * 2;
    keep Name Age_x2;
run;
```

(a1) KEEP ステートメントを最後に書く場合のコード

出力データセット

	Name	Age_x2
1	Alfred	28
2	Alice	26
3	Barbara	26
4	Carol	28
5	Henry	28

(a2) 出力結果

```sas
data test2;
    keep Name Age_x2;
    set sashelp.class;
    Age_x2 = Age * 2;
run;
```

(b1) KEEP ステートメントを最初に書く場合のコード

出力データセット

	Name	Age_x2
1	Alfred	28
2	Alice	26
3	Barbara	26
4	Carol	28
5	Henry	28

(b2) 出力結果

図 11　KEEP ステートメントの書く位置を変更した場合の SAS コード例と出力結果

KEEP ステートメントを最初に書いたコード（図 11(b1)）では、Age_x2 という変数を定義する前に KEEP ステートメントで指定している。しかし、エラーにも WARNING にもならず、Age_x2 は正しく出力データセットに含まれる（図 11(b2)）。なぜ LENGTH ステートメントとは異なり、KEEP ステートメントは位置に影響されないのだろうか。

KEEP ステートメントは、「出力データセットに保持する変数のリスト」（図 12(c)）を作成する。そしてこのリストは、データステップのコード全体をコンパイルした結果として作られる。すなわち、KEEP ステートメントが実際にコードのどこに書かれているかにかかわらず、SAS はデータステップ内をコンパイルする際に、リストを確定する。

このことを確かめるため、実行フェーズの途中で PDV の状態を確認する。

```sas
data test3;
    keep Name Age_x2;
    set sashelp.class;
    Age_x2 = Age * 2;
    put _all_;        <-- (1)
run;
```

(a) PDV の状態を確認するためのコード

PDV

Name	Sex	Age	Height	Weight	Age_x2	_ERROR_	_N_
Alfred	M	14	69	112.5	28	0	1
Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8
Keep					Keep		

(b) (1)のときの PDV の状態

KEEP リスト
Name
Age_x2

(c) KEEP リスト

図 12　KEEP ステートメントの効果を確認する SAS コード例と PDV の状態

PUT _ALL_で PDV の状態を確認すると、Sex や Weight など、KEEP ステートメントで指定していない変数も、この時点では PDV 上に存在していることがわかる（図 12(b)）。つまり、KEEP ステートメントの効果が生じるのは、PDV から出力データセットへ出力される段階である。出力段階になって初めて、リストにない変数（Sex, Age, Height, Weight）が除外される。なお、除外したい変数を指定する DROP ステートメントについても、同様の性質が成り立つ。

LENGTH ステートメントは、変数の属性を指定するステートメントであり、変数の属性は対象変数への最初の参照が行われた時点で確定する。そのため、書く位置が SET ステートメントの前か後かによって、結果が変わった。

一方 KEEP ステートメントは、値や属性には影響を与えず、「最後に何を出力するか」という出力に関する判断だけを行う。この判断はコード全体を見渡したうえでコンパイル時に一度だけ確定し、実際に適用されるのは出力の瞬間である。位置によって「最初の参照」が変わる、という問題がそもそも生じないため、KEEP ステートメントは書く位置に依存しない。

なお、同一データステップ内で同じ変数を KEEP ステートメントと DROP ステートメントの両方に指定した場合、記述した順序にかかわらず DROP が優先される。ただし、両者を同一データステップ内で併用することは推奨されていない[3][4]。

3.4. RENAME ステートメント: 書く位置によらず出力結果が同じ

RENAME ステートメントは、出力データセットにおいて変数名を変更するためのステートメントである[5]。RENAME ステートメントも、KEEP/DROP ステートメントと同様に、データステップのどこに記述しても結果が変わらない。なぜ結果が変わらないのか、PDV の動作から確認する。

(1) RENAME ステートメントは PDV 上の変数名を変更しない

以下のコードは、Height 変数を Height_inch に変更する例である。

```
1 data test;
2     set sashelp.class;
3     rename Height = Height_inch; /*Height変数をHeight_inchに変更する*/
4     put _all_;                   <-- (1)
5 run;
```

(a) Height 変数を Height_inch に変更するコード

出力データセット

	Name	Sex	Age	Height_inch	Weight
1	Alfred	M	14	69	112.5
2	Alice	F	13	56.5	84
3	Barbara	F	13	65.3	98
4	Carol	F	14	62.8	102.5
5	Henry	M	14	63.5	102.5

(b) 出力結果

RENAME リスト	
変更前	変更後
Height	Height_inch

(c) RENAME リスト

図 13　Height 変数を Height_inch に変更する SAS コード例および出力結果

PDV

Name	Sex	Age	Height	Weight	_ERROR_	_N_
Alfred	M	14	69	112.5	0	1
Length: 8	Length: 1	Length: 8	Length: 8	Length: 8	Length: 8	Length: 8

図 14　図 13(a)の(1)の時点での PDV の状態

図 13(a)の(1)の時点での PDV の状態を図 14 に示す。PDV 上には Height_inch という名前の変数は存在せず、Height という元の変数名があるのみである。つまり、RENAME ステートメントを記述しても、PDV 上での変数名は変更されない。RENAME ステートメントがコンパイルフェーズで行っているのは、「出力時に、変数名を〇〇に変更する」という RENAME リストへの登録（図 13(c)）であり、PDV 上の変数名を書き換えることではない。したがって、実行フェーズの間、対象の変数は元の名前で維持される。

(2) 変更後の変数名をコード中で使用すると、変数名が競合する

RENAME ステートメントで変数名を変更しても、PDV 上は新しい変数名に変更されないため、新しい変数名をコード中で使用することはできない。以下のコードは、変更した変数名（Height_inch）を用いて、Height_cm を計算しようとする誤った例である。

```
1 data test;
2     set sashelp.class;
3     rename Height = Height_inch;
4     put _all_;                          (1)
5     Height_cm = Height_inch * 2.54;
6     put _all_;                          (2)
7 run;
```

(a) 変更した変数名を同一データステップ内で使用するコード

出力データセット

	Name	Sex	Age	Height_inch	Weight	Height_cm
1	Alfred	M	14	69	112.5	.
2	Alice	F	13	56.5	84	.
3	Barbara	F	13	65.3	98	.
4	Carol	F	14	62.8	102.5	.
5	Henry	M	14	63.5	102.5	.

(b) 出力結果

```
NOTE: 変数Height_inchは初期化されていません。
WARNING: 変数Height_inchはファイルWORK.TESTに既に存在します。
```

(c) ログに NOTE と WARNING が出力される

図 15　変更した変数名を同一データステップ内で使用する SAS コード例、ログおよび出力結果

SAS は新規変数 Height_inch が初期化されていない旨の NOTE と、Height_inch が WORK.TEST に既に存在する旨の WARNING を出力する（図 15(c)）。この WARNING の原因を、PUT _ALL_ で PDV の中身を直接確認することで確かめる。

PDV

Name	Sex	Age	Height	Weight	Height_cm	Height_inch	_ERROR_	_N_
Alfred	M	14	69	112.5	.	.	0	1

(a) 図 15(a)の(1)の PDV の状態

PDV

Name	Sex	Age	Height	Weight	Height_cm	Height_inch	_ERROR_	_N_
Alfred	M	14	69	112.5	.	.	0	1

(b) 図 15(a)の(2)の PDV の状態

図 16　図 15(a)を実行したときの PDV の状態

この出力を見ると、PDV 上には Height と Height_inch という変数が存在するが、Height_inch という変数に Height の値が格納されていない（図 16(a)）。Height_inch が欠損値であることから、Height_cm も計算されず欠損となる（図 16(b)）。これは、先ほど確認したとおり、RENAME ステートメントを記述しても、PDV 上での変更は発生しないためである。RENAME ステートメントは、RENAME リストに「Height を Height_inch に変更する」という規則を登録するだけである。

PDV 上に存在する Height_inch はコンパイル時に処理される 5 行目「Height_cm = Height_inch * 2.54」の Height_inch である。2 章で確認したとおり、コンパイルフェーズでは、コード中に登場する変数をもとに PDV が構築されるため、Height_inch も PDV 上に変数として作成される。RENAME ステートメントによって出力時に Height が Height_inch という名前に変更されるところへ、新規変数の Height_inch と変数名が競合してしまい WARNING となる（図 17）。なお、出力データセットには、RENAME ステートメントの効果が優先され、元の Height の値が Height_inch という名前で出力される。一方、実行フェーズ中に参照された新規変数の Height_inch は欠損値であったため、Height_cm の計算結果は欠損値のままとなる（図 15(b)）。

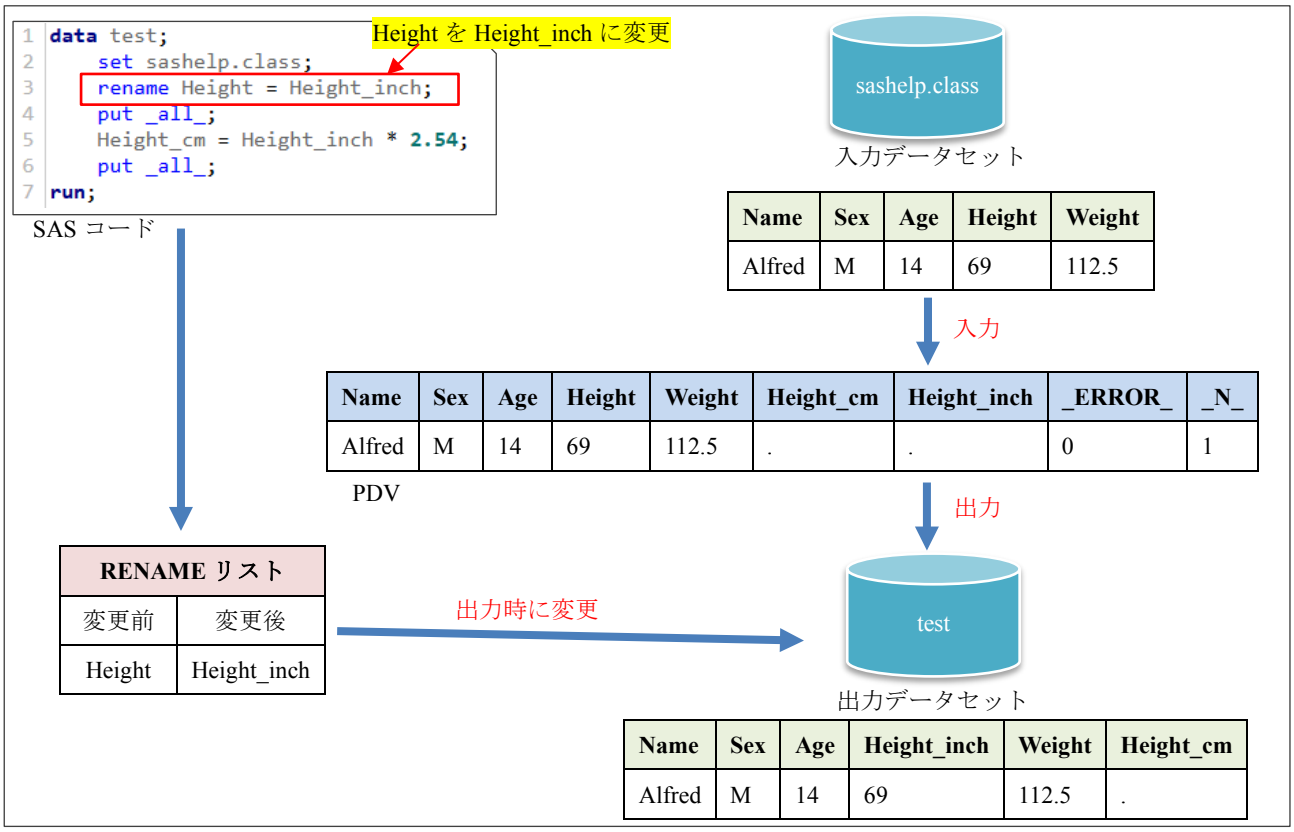


図 17　RENAME リストと PDV の間で変数名 Height_inch が競合する

KEEP/DROP ステートメントおよび RENAME ステートメントが、書く位置によらず同じ結果となるのは、これらのステートメントの動作がコンパイルフェーズで確定し、実行フェーズの出力のタイミングで適用されるからである。特に RENAME ステートメントについては、変更後の変数名を同一データステップ内で使用するとバグの原因となり得る。

4. まとめ

本稿では、LENGTH、KEEP/DROP、RENAME という 3 つのステートメントを題材に、それぞれの動作を PDV という仕組みの観点から確認してきた。

文字変数の属性は、コンパイルフェーズにおいてその変数が最初に参照された時点で確定する。LENGTH ステートメントの指定が反映されるのは、それが対象変数への最初の参照となる場合に限られるため、SET ステートメントより前に書くか後に書くかによって結果が異なる。

一方、KEEP/DROP ステートメントと RENAME ステートメントは、いずれもコンパイルフェーズで「何を」「どのような名前で」出力するかを決定するが、その効果が実際に現れるのは、PDV から出力データセットへ出力されるタイミングである。実行フェーズの間、対象の変数は PDV 上に通常どおり存在し（KEEP/DROP の場合）、あるいは元の名前のまま扱われる（RENAME の場合）。この違いにより、LENGTH とは異なり、KEEP/DROP、RENAME ステートメントの働きは書く位置に依存しない。

データステップの各種ステートメントを「慣習を守って書けば動くもの」として扱うのではなく、PDV の観点からどのような処理が行われているのかを理解することが重要と考える。こうした理解は、バグを防ぎ堅牢なコードを作成することにつながると期待できる。

なお、本稿で使用した SAS コードを別添に示す。別添には、PDV 上に存在するが出力データセットには含まれない変数の例として、BY ステートメントを使用したコード例も収録した。BY ステートメントは指定した変数の値でオブザベーションをグループ化するステートメントで、SET ステートメントと併用すると PDV 上に FIRST.変数名と LAST.変数名という変数が現れ、そのオブザベーションが BY 変数の値ごとのグループの先頭か末尾かを示す[6]。PUT _ALL_を用いて、これらの変数が PDV 上に実際に作られていることを確認されたい。本稿では SET ステートメントのみを扱ったが、MERGE ステートメントについては、PDV の初期化の観点から、データセットの結合時に想定と異なる結果が得られる事例が解説されている[7]。あわせて参照されたい。

参考文献

[1] SAS Institute Inc. (2023), "DATA Step Processing", SAS® 9.4 Language Reference: Concepts, Sixth Edition, <https://documentation.sas.com/doc/en/lrcon/9.4/p1topuaeb1ikf0n11f6ibw5ftral.htm>（参照日: 2026 年 8 月 1 日）

[2] SAS Institute Inc. (2026), "LENGTH Statement", SAS® 9.4 and SAS® Viya® 3.5 Programming Documentation, <https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/lestmtsref/p1hgqgmxm3dpqcn1d4w5za5qbz0d.htm>（参照日: 2026 年 8 月 1 日）

[3] SAS Institute Inc. (2026), "KEEP Statement", SAS® 9.4 and SAS® Viya® 3.5 Programming Documentation, <https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/lestmtsref/n1nnrzzsw6rzrjn1p2jfky6pdv23.htm>（参照日: 2026 年 8 月 1 日）

[4] SAS Institute Inc. (2026), "DROP Statement", SAS® 9.4 and SAS® Viya® 3.5 Programming Documentation, <https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/lestmtsref/n1capr0s7tilbvn1lypdshkgpaip.htm>（参照日: 2026 年 8 月 1 日）
[5] SAS Institute Inc. (2026), "RENAME Statement", SAS® 9.4 and SAS® Viya® 3.5 Programming Documentation, <https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.5/lestmtsref/n0x16kvqkxxdx5n1t04voifvo8wo.htm>（参照日: 2026 年 8 月 1 日）
[6] SAS Institute Inc. (2023), "FIRST. and LAST. DATA Step Variables", SAS® 9.4 Language Reference: Concepts, Sixth Edition, <https://documentation.sas.com/doc/en/lrcon/9.4/n01a08zkzy5igbn173zjz82zsi1s.htm>（参照日: 2026 年 9 月 12 日）
[7] 雨宮 祐輔 (2023), "SAS 初心者のための DATA ステップ処理・データセットマージ入門", SAS ユーザー総会 2023 論文集, pp.151-163.

生成 AI エンジニアリングの発展と SAS の統合

～ Prompt・Context・Harness・Loop・Graph Engineering の体系化 ～

○ 小野 潔

Independent Researcher　E-mail：ono2000j@gmail.com

Evolution of Generative AI Engineering and Integration with SAS

– From Prompt and Context to Harness Loop and Graph Engineering –

要旨

生成 AI を業務で安定して利用するには、指示文だけでなく、必要情報、外部ツール、検証、再実行、処理関係を一体として設計する必要がある。本稿では、これらをプロンプト、コンテキスト、ハーネス、ループ、グラフの 5 つのエンジニアリングとして整理し、SAS との統合方法を検討する。対象は、筆者が前職で経験した二つの事例であり、本稿はその概念を拡張したものである。事例 1 は SAS Enterprise Guide（SAS EG）から SAS Viya への移植であり、既存資産の解析、仕様書作成、コード再構成および新旧結果の比較を扱う。事例 2 は 2025 年に提案した SAS Intelligent Decisioning（以下、SAS ID）によるルールベース制御の与信モデルであり、業務ルール、Reason Code、MCP を介した Decision Service 呼出しおよび人間の承認を扱う。二つの事例を分けて説明した後、5 領域との対応を比較する。なお、統合システムの実装と実環境での総合検証は今後の課題である。

キーワード

生成 AI、SAS、SAS Intelligent Decisioning、MCP、プロンプトエンジニアリング、
コンテキストエンジニアリング、ハーネスエンジニアリング、ループエンジニアリング、グラフエンジニアリング

1. はじめに

1.1. 研究背景と目的

生成 AI は、文章作成だけでなく、プログラム生成、既存コードの説明、エラー調査、仕様書作成などにも利用されている。SAS 開発でも、生成 AI によって作業の一部を効率化できる可能性がある。一方、生成されたコードは、文法的に正しく見えても、元の業務要件やデータ構造を正確に反映しているとは限らない。

本研究の目的は、この仕組みを 5 つの生成 AI エンジニアリングとして整理し、SAS 開発における役割を明らかにすることである。対象事例は、事例 1 の SAS EG から SAS Viya への移植と、事例 2 の SAS ID によるルールベース制御の与信モデルである。前者では開発支援、後者では意思決定と統制を検討する。

2. 生成 AI エンジニアリングの全体像

2.1. 5 領域のエンジニアリング

生成 AI の業務利用は、指示の設計から始まり、必要な情報の選択、外部ツールとの接続、結果の反復的な改善、複雑な処理関係の構造化へと広がる。この関係を表 1 に示す。各領域は独立した技術ではなく、相互に補い合う

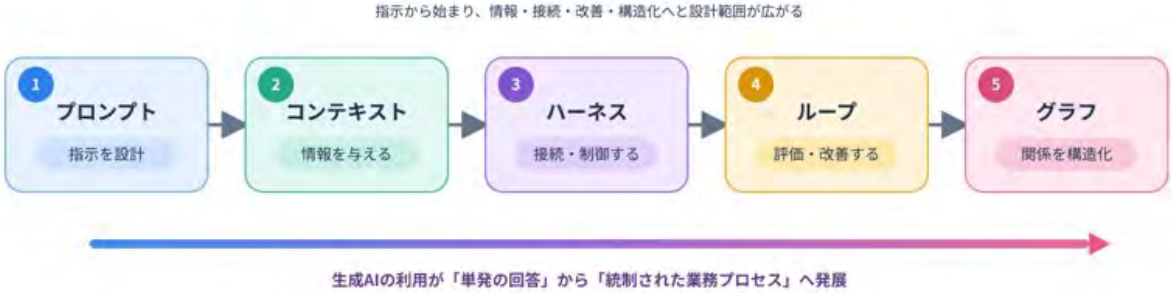
設計領域である。

下記の 5 領域のエンジニアリングは、同時に提唱された統一理論ではない。プロンプトエンジニアリングは 2021 年頃から広く利用され、コンテキストエンジニアリングは 2025 年 6 月頃、ハーネスエンジニアリングは 2026 年 2 月頃から注目されるようになった。一方、ループエンジニアリングとグラフエンジニアリングは、2026 年頃から実務者の間で使われ始めた新しい整理方法であり、提唱者や開始時期は明確に確定していない。本稿では、生成 AI と SAS の統合を説明するための設計領域として、この 5 領域を整理して説明する。

表1　5つの生成AIエンジニアリング

領域	注目された時期	主な設計対象	SAS開発での例
プロンプト	2021年頃	目的、指示、制約、出力形式	SASコードの解析手順を指定する
コンテキスト	2025年6月頃	データ定義、業務知識、既存資産	変数定義書や既存プログラムを与える
ハーネス	2026年2月頃	ツール接続、実行制御、ルール制御、記録、外部DB接続	生成AIからSASを呼び出して結果を制御する
ループ	2026年頃	評価、修正、再実行、終了条件	新旧プログラムの出力を比較して修正する
グラフ	2026年7月頃	データフロー、処理フロー、依存関係、分岐	SAS EGのノードと入出力関係を構造化する

図1　5つの領域の発展関係



2.2.　5 領域と SAS の関係

生成 AI は、自然言語の理解、情報の要約、コード候補の作成など、形式が固定されていない処理を得意とする。一方、SAS は、データ加工、統計解析、モデル実行、業務ルール、処理結果の保存など、再現性が必要な処理を担う。両者を組み合わせる際は、生成 AI が候補を作り、SAS が計算と検証を行い、人間が最終的に確認するという役割分担が基本となる。

生成 AI の 5 領域のエンジニアリングは、特定の製品名や厳密な年代順を示すものではない。生成 AI の利用範囲が、単発の回答から外部ツールを含む業務プロセスへ広がる関係を表している。

2.3.　本研究の位置づけ

本研究では、5 領域を同じ重要度を持つ設計領域として扱い、二つの事例へ適用する。事例 1 は SAS EG から SAS Viya への移植であり、プロンプト、コンテキスト、ループおよび処理関係のグラフ化を中心に扱う。事例 2 は SAS ID によるルールベース制御の与信モデルであり、ハーネス、業務ルール、Reason Code、MCP および説明可能性を中心に

扱う。第 3 章と第 4 章では事例を分け、第 6 章で両者と 5 領域の対応を比較する。

事例 1 では、生成 AI を用いて SAS コードを処理単位に分割し、コードから仕様書を作成した後、その仕様書を基にコードを再構成して元の処理と比較する方法を検討した。事例 2 は、2025 年に提案した三層型与信モデルをハーネスの観点から再整理したものである。SAS ID によるルールベース制御、MCP を介した外部接続、サロゲートモデルによる説明および人間承認の構成を扱う。両事例を通じて、5 領域が開発支援と意思決定の双方にどのように対応するかを示す。

3. 事例 1　SAS EG から SAS Viya への移植

3.1. 移植作業におけるプロンプト

事例 1 は、既存の SAS EG 資産を解析し、仕様を確認しながら SAS Viya へ移植する作業である。プロンプトでは、目的、対象範囲、処理順序、禁止事項、出力形式を明示する。例えば、「入力データ、出力データ、使用変数、処理条件の順に説明する」「プログラムに書かれていない内容は推測しない」と指定する。

明確なプロンプトは出力のばらつきを抑えるが、生成 AI が知らない企業固有の変数、処理順序、データ構造を補うことはできない。指示の改善だけで解決できない情報不足は、既存 SAS 資産をコンテキストとして与えることで補う。

3.2. 既存 SAS 資産のコンテキスト化

移植作業におけるコンテキストは、生成 AI が既存処理を理解するために参照する背景情報である。変数定義書、データレイアウト、既存プログラム、テスト条件、過去の設計書、SAS EG の処理フローなどが該当する。情報を大量に与えるのではなく、現在の処理に必要な部分を選び、意味と依存関係を保って提供することが重要となる。

表2　プロンプトとコンテキストの違い

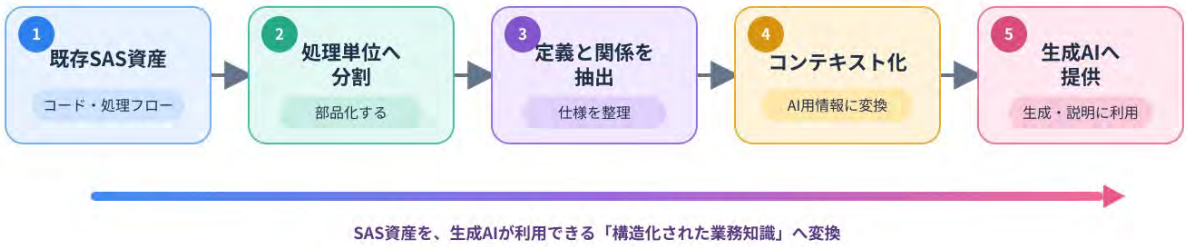
項目	プロンプト	コンテキスト
役割	何をどのように処理するかを指示する	処理に必要な知識と前提を与える
主な内容	目的、手順、制約、出力形式	変数定義、業務ルール、既存コード、データ
SASの例	処理単位ごとに仕様を説明するよう指示	処理フロー、入出力、変数情報を提供
限界	不足している業務知識は補えない	情報を与えても必ず守られるとは限らない

3.3. 仕様書を介した移植手順

長期間利用された SAS プログラムには、コードだけでなく、実行順序、入力ファイル、例外処理、担当者の運用知識が含まれている。特に SAS EG では、プログラム、クエリ、データセットが複数のノードとして接続されるため、一つのコードだけを読んでも処理全体を把握できない。

そこで、既存資産を（1）処理単位へ分割し、（2）入出力と使用変数を抽出し、（3）業務上の意味を設計書と照合し、（4）仕様書として整理し、（5）その仕様書を基に SAS Viya 向けコードを再構成する。生成 AI が作成した仕様書とコードは元の処理と照合する。これにより、コードから仕様書を作り、その仕様書からコードを再生成して比較する検証が可能になる。

図2　SAS資産からコンテキストを作る流れ



4.　事例 2　SAS ID によるルールベース与信モデル

4.1.　与信モデルにおけるハーネスの役割

事例 2 は、SAS ID によるルールベース制御を組み込んだ与信モデルである。ここでのハーネスは、生成 AI、予測モデル、業務ルール、外部接続および人間による確認を結び、判断の実行、検証、記録、例外処理を管理する。プロンプトとコンテキストが生成 AI への入力を設計するのに対し、ハーネスは入力後の処理全体を統制する。

重要な業務ルールは、コンテキストとして生成 AI へ与えるだけでは不十分である。生成 AI はルールを参照して判断案を作成できるが、常にルールどおりの結果を返すとは限らない。そこで、SAS ユーザ総会 2025 では、同じルールを SAS Intelligent Decisioning などに実装し、生成 AI の出力を機械的に再検証することを述べた。

4.2.　与信モデルのハーネス構成

SAS Intelligent Decisioning（ID）は、生成 AI の出力にルールベースの制御を組み込み、幻覚やバイアスを抑制すると同時に、スコアリングの根拠を Reason Code として提示できる。この仕組みにより「なぜそのスコアとなったのか」を透明かつ体系的に説明可能とし、規制当局に対する説明責任を果たすことができる。

表3　与信モデルのハーネスが管理する機能

機能	処理内容	管理する情報
入力管理	申込情報、モデルスコア、必要なコンテキストを確認	入力項目、欠損、データ版
実行管理	許可されたSAS IDのDecision Flowを実行	実行者、時刻、ルール版
結果管理	判定、Reason Code、警告を保存	判定結果、理由、例外
統制	権限、停止条件、人間承認を確認	承認者、変更履歴、監査ログ

図3　与信モデルのハーネス処理



4.3. 昨年度の三層モデルとハーネスの関係

2025 年研究で提案した三層モデルは、ChatGPT エージェント、Dify、SAS Intelligent Decisioning に役割を分けていた。Dify は、生成 AI と外部ツールを組み合わせた処理の流れを画面上で構築・管理する開発基盤であり、本稿では入力の受け取り、ツールの呼出し、実行順序およびデータ受渡しを担当する。当時は三つの製品・機能を連携させる構成として説明したが、現在の生成 AI エンジニアリングの観点では、生成 AI に実行環境と制御機構を装着するハーネスの初期的な構想と捉えられる。各層とハーネス機能の対応を表 4 に示す。

昨年度モデルでは、役割分担と接続の構想は示したが、生成 AI へ渡す情報の選択、実行結果の反復的な検証、複数処理の依存関係までは十分に扱っていなかった。本研究は、この三層モデルをハーネスとして位置づけ直し、コンテキスト、ループ、グラフを加えた 5 領域へ拡張する。これが昨年度研究からの主な発展点である。

表4　昨年度の三層モデルとハーネス機能の対応

	昨年度の構成	昨年度に示した役割	ハーネスとしての役割
1	ChatGPT エージェント	自然言語の理解、情報整理、処理候補の提示	利用者の要求を実行可能な指示へ変換する
2	Dify	ワークフローとAPIによる処理制御	ツールの選択、実行順序、データ受け渡しを管理する
3	SAS Intelligent Decisioning	モデル実行、業務ルール、Reason Code	再現可能な計算とルール検証を実行する
4	人間	結果の確認と最終判断	例外処理、承認、責任の所在を明確にする

昨年度の三層モデルを、実行・検証・承認を担うハーネスとして再整理

4.4. SAS Intelligent Decisioning によるルールベース制御

生成 AI の出力を業務へ渡す前に、SAS Intelligent Decisioning （SAS ID） の Decision Flow でモデル、Decision Table、ルールセットを順に実行する。生成 AI は候補を作るが、SAS ID が許容値、禁止条件、例外条件を決定論的に検査する。判定結果には承認可否だけでなく Reason Code を付け、入力、ルール版、実行時刻とともに保存する。これにより、同じ入力に同じルールを適用でき、生成 AI の文章だけに統制を依存しない。

実装上の示唆として、生成 AI にルール文を読ませる処理と、SAS ID で同じ制約を強制する処理は分ける。前者は説明と候補生成、後者は最終ゲートである。ルールの優先順位、適用開始日、版番号を管理し、変更前後のテストケースを保存することが重要となる。

表5　SAS IDに実装するSルール例

優先度	入力条件	SAS IDの処理	Reason Code
1	個人情報を含む、または必須値が欠落	自動実行を停止し、人間へ確認を戻す	REVIEW_REQUIRED
2	提案値が業務上限を超過	上限内へ補正するか却下する	LIMIT_EXCEEDED
3	モデル信頼度が基準未満	自動承認せず保留する	LOW_CONFIDENCE
4	すべての条件を満たす	次工程の実行を許可する	APPROVED

上位の停止・確認ルールから順に評価し、安全条件を満たした場合のみ次工程を許可

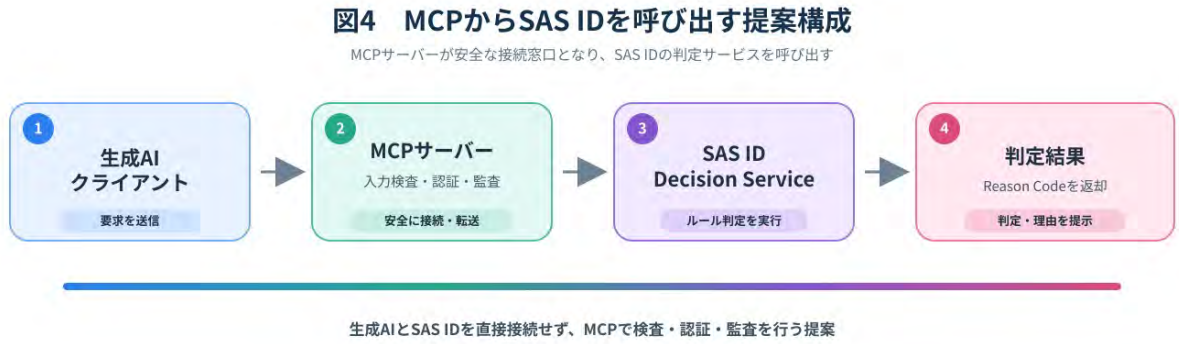
4.5. 予測モデルの説明と検証

本稿でサロゲートモデルを扱うのは、生成 AI そのものを説明するためではない。生成 AI が業務の中で利用する XGBoost や GNN などの予測モデルについて、その判断をハーネス内で検証・説明するためである。サロゲートモデル（代理モデル）は、複雑な本モデルの判断を、ロジスティック回帰や決定木などの単純なモデルで近似する。本モデルの代替ではなく、どの変数や条件が判断に影響したかを示す「通訳」として利用する。

与信では、予測精度に加えて判断理由を検証・説明できることが重要である。米国の FRB、OCC、FDIC および日本の金融庁も、重要な判断に用いるモデルの合理性を検証・説明できることを重視している［10、11］。両当局がサロゲートモデルを特定して推奨しているわけではないが、複雑なモデルの判断傾向を可視化する実務的な手段の一つとなる。元モデルの予測性能とサロゲートモデルの一致度は別の指標であるため、分けて評価する。具体的な SAS コードと評価方法は付録 A に示す。

4.6. MCP を介した外部からの SAS ID 呼出し

Model Context Protocol（MCP）は、生成 AI アプリケーションが外部のデータやツールを共通形式で利用するためのプロトコルである。SAS ID に対しては、SAS が公開する Decision Service の REST エンドポイントを MCP サーバーのツールとして包む構成が考えられる。これは SAS 標準の MCP 機能を意味するものではなく、本研究が提案する接続アダプターである。外部の AI は SAS の URL や認証情報を直接扱わず、入力スキーマを限定した evaluate_decision ツールだけを呼び出す。



リスト 2　MCP ツールから Decision Service を呼ぶ概念コード

```
from mcp.server.fastmcp import FastMCP
import os, requests

mcp = FastMCP('sas-id-gateway')

@mcp.tool()
def evaluate_decision(customer_id: str, score: float, action: str) -> dict:
    payload = {'inputs': [
        {'name': 'customer_id', 'value': customer_id},
        {'name': 'score',       'value': score},
        {'name': 'action',      'value': action}]}
    response = requests.post(
        os.environ['SAS_DECISION_URL'], json=payload,
        headers={'Authorization': 'Bearer ' + os.environ['SAS_TOKEN']},
        timeout=10)
    response.raise_for_status()
    return response.json()
```

@mcp.tool()の先頭にある@は、Python のデコレーターを表す。デコレーターは、直後の関数に役割を追加する仕組みである。ここでは、通常の Python 関数である evaluate_decision を MCP ツールとしてサーバーへ登録し、生成 AI から呼び出せるようにする。@mcp.tool()がなければ、この関数は同じ Python プログラム内から呼び出す通常の関数にと

どまる。mcp は FastMCP('sas-id-gateway')で作成した MCP サーバーであり、関数名、引数名および型情報を基に、生成 AI が利用するツールの入力形式を公開する。

URL と入出力形式は、SAS Viya で Decision を発行した際のサービス定義に合わせて設定する。概念実証では、（1）参照・判定系だけを公開、（2）ツール名と入力項目を許可リスト化、（3）アクセストークンを環境変数または秘密管理基盤で保管、（4）タイムアウトと再試行回数を固定、（5）要求と応答を監査ログへ保存、（6）ルール変更や本番発行は人間承認とする。SAS コードの任意実行機能を、そのまま MCP へ公開してはならない。任意のコードを受け付けると、生成 AI の誤操作や不正な指示によって、想定外のデータ参照、更新、削除、または過大な計算処理が実行される危険がある。そのため、MCP から呼び出せる処理は、事前に検証して登録した Decision Service などに限定し、入力項目、実行権限、処理時間、返却項目を制御する必要がある。

5. 二つの事例におけるループとグラフ

5.1. 移植作業におけるループ

事例 1 のループは、生成した SAS Viya 向けコードを評価し、問題があれば修正して再実行する仕組みである。単に処理を繰り返すのではなく、既存 SAS EG 処理との比較項目、修正対象、再実行条件および終了条件を明確にする。

既存プログラムと生成プログラムを同じテストデータで実行し、レコード件数、欠損値、集計値および出力データを比較する。不一致があれば、生成 AI が差分と関連コードを確認して修正案を作る。修正後に再実行し、基準を満たすまで検証する。ただし、移植完了の最終判断は人間が行う。

表6　移植ループで確認する主な項目

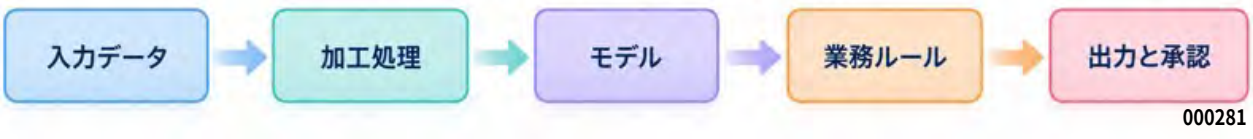
	確認対象	比較内容	問題がある場合の処理
1	データ加工	件数、型、欠損値、重複、結合結果	該当する処理単位を修正する
2	集計	合計、平均、分布、グループ別結果	条件式や対象期間を確認する
3	出力データ	変数名、型、並び順、値の一致	変数定義と生成コードを確認する
4	処理関係	実行順序、入出力、前後の依存関係	SAS EGフローと仕様書を照合する

5.2. 二つの事例におけるグラフ

グラフエンジニアリングは、データ、プログラム、モデル、業務ルール、人間の承認などの関係を、ノードと接続として構造化する考え方である。事例 1 では、SAS EG のデータ、プログラム、クエリおよび出力の依存関係を表す。事例 2 では、入力データ、予測モデル、SAS ID の業務ルール、Reason Code および人間承認の関係を表す。

事例 1 では、変更したプログラムが後続処理へ与える影響を追跡し、生成 AI へ渡すコンテキストを選ぶためにグラフを利用できる。事例 2 では、どのモデルとルールが判定に使われ、誰が承認したかを追跡するために利用できる。対象は異なるが、必要な関係だけを抽出し、処理の根拠と変更影響を確認する役割は共通する。

図6　二つの事例に共通する処理関係



6. 二つの事例の比較と今後の課題

6.1. 二つの事例と 5 領域の対応

事例 1 と事例 2 は対象業務が異なるが、5 つのエンジニアリングで整理できる。事例 1 では、プロンプトとコンテキストを用いて既存処理を理解し、ループとグラフによって移植結果を検証する。事例 2 では、ハーネスが予測モデル、SAS ID、MCP および人間承認を接続し、判断と理由を記録する。両者の対応を表 7 に示す。

表7 二つの事例と5領域の対応

領域	事例1 SAS EGからViyaへの移植	事例2 SAS ID与信モデル	共通する管理
プロンプト	仕様書とコード生成の指示	説明候補と処理要求の指示	目的、制約、出力形式
コンテキスト	既存コード、定義書、SAS EG フロー	申込情報、モデル情報、業務ルール	必要情報の選択と版管理
ハーネス	実行範囲、ログ、比較処理を管理	SAS ID、MCP、人間承認を管理	権限、停止、記録
ループ	生成、実行、比較、修正	予測、ルール検証、再判定	評価基準と終了条件
グラフ	データとプログラムの依存関係	モデル、ルール、判定、承認の関係	影響分析と追跡

6.2. 事例別の段階的な概念実証

概念実証は、二つの事例を混ぜずに別の経路で進める。事例 1 では、まず一つの SAS EG フローを対象に仕様書を作成し、次に SAS Viya 向けコードを生成して新旧結果を比較する。事例 2 では、SAS ID 単体、REST API、MCP、生成 AI の順に接続範囲を広げる。最後に、共通の監査ログと人間承認の方法を評価する。これにより、移植精度の問題と、与信モデルのルール・接続・統制の問題を切り分けて検証できる。

表8 事例別のPoCと合格条件

対象	段階	主な合格条件
事例1 移植	仕様書作成	入出力、使用変数、処理順序が元の SAS EG資産と一致
事例1 移植	コード生成・比較	SAS Viyaで実行でき、主要な新旧結果が一致
事例2 与信	SAS ID・REST API	判定とReason Codeが期待値に一致し、認証エラーを制御
事例2 与信	MCP・生成AI	入力制限、監査ログ、停止条件、人間承認を確認

6.3. 実用化に向けた課題

事例 1 では、SAS EG 固有機能の置換、テストデータの網羅性、移植前後の差分判定を検証する必要がある。事例 2 では、Dify と SAS の API 接続、認証と権限、処理時間、障害時の再実行、入力情報と判定結果の保存方法を検証する必要がある。共通課題として、生成 AI が作成したコードや説明の正確性、再現性および説明可能性を評価する基準が必要となる。

金融分野では、個人情報の保護、モデルリスク管理、判断理由の保存、人間による承認が重要である。生成 AI

へ最終判断を任せず、SAS による計算とルール検証を行い、その結果を人間が確認する構成を維持する必要がある。

6.4. 創薬への応用可能性

創薬では、化合物、遺伝子、タンパク質、標的、疾患、副作用などが相互に関係するため、これらをノード、関係をエッジとして表すグラフエンジニアリングが適用候補となる。化合物の合成ではグラフニューラルネットワーク（ GNN）が使われるが、GNN はグラフ構造を学習するニューラルネットワークの一種である。化合物―標的関係などから候補スコアを算出できる一方、スコアだけを見ても、どの分子特徴や関係が評価に影響したかを理解しにくい場合がある。

そこで、GNN の出力を目的変数として決定木や回帰モデルを学習するサロゲートモデルを併用し、GNN が候補を高く評価する条件を人間が確認できる形に近似する。これは説明・検証の有力な手段であるが、GNN の内部機構や生物学的作用機序を直接証明するものではない。FDA と EMA の創薬 AI 原則も、予測性能だけでなく解釈可能性、説明可能性、利用目的に応じた検証を重視している［12］。本モデルのスコア、サロゲートによる説明、SAS による統計的検証、専門家による生物学的評価を組み合わせて運用することが望ましい。SAS ID では毒性、適用条件、データ品質などを確認し、MCP は外部データベースや解析サービスとの接続、ループは予測、検証、条件修正、再評価を管理する。

7. 結論

本稿では、生成 AI の業務利用をプロンプト、コンテキスト、ハーネス、ループ、グラフの 5 領域として整理した。プロンプトは指示、コンテキストは必要情報、ハーネスは外部ツールとの接続と統制、ループは評価と改善、グラフは複雑な関係の構造化を担当する。

本稿では、事例 1 の SAS EG から SAS Viya への移植と、事例 2 の SAS ID によるルールベース与信モデルを分けて検討した。事例 1 では、既存資産のコンテキスト化と、生成、実行、比較、修正のループが中心となる。事例 2 では、SAS ID によるルール検証、MCP を介した Decision Service 呼出し、Reason Code、サロゲートモデルおよび人間承認が中心となる。二つの事例を 5 領域で比較することで、開発支援と意思決定統制に共通する設計要素を示した。今後は、各事例を別々に小規模試作し、段階的に有効性を検証する。

8. 参考文献

1）小野 潔：生成 AI と SAS Intelligent Decisioning を組み合わせた意思決定基盤，2025 年 SAS ユーザー総会発表資料.

2）SAS Institute Inc.：SAS Intelligent Decisioning Documentation.

3）Dify：Dify Documentation.

4）SAS Institute Inc.：SAS Intelligent Decisioning Support and Documentation，https://support.sas.com/en/software/intelligent-decisioning-support.html（2026 年 9 月 16 日参照）.

5）SAS Institute Inc.：SAS REST API Examples，https://github.com/sassoftware/devsascom-rest-api-samples（2026 年 9 月 16 日参照）.

6）Model Context Protocol：Specification and Server Tools，https://modelcontextprotocol.io/specification/2026-07-28（2026 年 9 月 16 日参照）.

7）Sahoo, P. et al.：A Systematic Survey of Prompt Engineering in Large Language Models，arXiv:2402.07927，2024.

8）Schmid, P.：The New Skill in AI is Not Prompting, It is Context Engineering，2025 年 6 月 30 日.
9）OpenAI：Harness engineering: leveraging Codex in an agent-first world，2026 年 2 月 11 日.
10）Board of Governors of the Federal Reserve System, FDIC, OCC：Revised Guidance on Model Risk Management，2026 年 4 月 17 日，https://www.federalreserve.gov/supervisionreg/srletters/SR2602.pdf（2026 年 9 月 16 日参照）.
11）金融庁：AI ディスカッションペーパー（第 1.1 版），2026 年 3 月 3 日，https://www.fsa.go.jp/news/r7/sonota/20260303/aidp_version1.1.pdf（2026 年 9 月 16 日参照）.
12）FDA and EMA：Guiding Principles of Good AI Practice in Drug Development，2026 年，https://www.fda.gov/about-fda/artificial-intelligence-drug-development/guiding-principles-good-ai-practice-drug-development（2026 年 9 月 16 日参照）.

付録 A　SAS によるサロゲートモデルの作成例

本モデル（XGBoost や GNN など）は予測精度を優先する一方、判断過程を説明しにくい場合がある。サロゲートモデルは、本モデルの判断を決定木や回帰モデルなどの単純な構造で近似するため、本モデルより予測精度や再現度が低くなることがある。しかし、どの変数や条件が判断に影響したかを人間が確認できるため、説明、検証、監査に利用できる。したがって、サロゲートモデルは本モデルの代替ではなく、本モデルを説明するために併用するモデルである。

サロゲートモデルには、元モデルへ入力した説明変数と、元モデルが出力した判定結果を使用する。1 行を 1 件の観測とし、x1～x3 を元モデルと共通の説明変数、p_blackbox を元モデルの予測確率、blackbox_flag をその確率から作成した 0 または 1 の判定とする。実際の目的変数を blackbox_flag として使用すると通常の予測モデルになるため、サロゲートモデルでは必ず元モデルの出力を目的変数とする。

リスト A1　PROC HPSPLIT による決定木サロゲートの作成

```
data work.surrogate_tree;
 set work.model_scored;
 blackbox_flag = (p_blackbox >= 0.5);
run;

proc hpsplit data=work.surrogate_tree seed=2026;
 class blackbox_flag;
 model blackbox_flag(event='1') = x1 x2 x3;
 partition fraction(validate=0.3);
 grow entropy;
 prune costcomplexity;
 code file='surrogate_tree_score.sas';
run;
```

PROC HPSPLIT は blackbox_flag を目的変数として決定木を作成する。PARTITION 文でデータの 30%を検証用に分け、GROW 文で分岐基準、PRUNE 文で枝刈り方法を指定する。CODE 文は、完成した決定木を別データへ適用するためのスコアコードを出力する。出力される予測確率の変数名は環境や設定により異なるため、実行後に確認し、次の例では p_surrogate へ統一する。

リスト A2　元モデルとの一致率の計算

```
data work.surrogate_eval;
 set work.surrogate_scored;
 surrogate_flag = (p_surrogate >= 0.5);
 agreement = (surrogate_flag = blackbox_flag);
run;

proc means data=work.surrogate_eval mean;
 var agreement;
run;
```

surrogate_scored は、検証用データへ surrogate_tree_score.sas を適用した結果である。surrogate_flag は決定木の判定、agreement は元モデルとの一致を表し、PROC MEANS で求めた agreement の平均が一致率になる。ただし、一致率は元モデルの再現度であり、実際の目的変数に対する予測精度ではない。元モデルの AUC 等とサロゲート

モデルの一致率を分けて評価する必要がある。

付録 B　主要用語集

本文で使用する主要な専門用語を、生成 AI や SAS に詳しくない読者向けに整理する。

表9　本稿で用いる主な用語

用語	本稿での意味
SAS ID	SAS Intelligent Decisioning の略称。予測モデルと業務ルールを組み合わせ、意思決定の流れを設計、実行、管理するSASの機能。
MCP	Model Context Protocol の略称。生成AIが外部のデータやツールを共通の形式で利用するための接続仕様。
Reason Code	判定結果に至った主な理由を示すコード。例として、審査で否決となった理由や、適用された業務ルールを記録する。
ハーネス	生成AI、外部ツール、業務ルール、人間による確認を結び、実行、検証、記録、例外処理を管理する制御の仕組み。
サロゲートモデル	複雑な予測モデルの判断を、決定木や回帰モデルなどの単純なモデルで近似し、人間が理解しやすくする説明用モデル。
Dify	生成AIと外部ツールを組み合わせた処理の流れを、画面上で構築・管理するための開発基盤。本稿では、入力の受け取り、ツールの呼び出し、実行順序およびデータ受渡しを担当する。

NDBオープンデータを用いた高血圧患者におけるアンメットメディカルニーズの探索

○芦沢 宏治[1]，渡邊 智昭[1]，齊藤 正樹[2]

（[1]株式会社タクミインフォメーションテクノロジー 医薬開発部，[2]戦略営業部）

Exploring Unmet Medical Needs Among Patients with Hypertension Using NDB Open Data

Koji Ashizawa, Tomoaki Watanabe, Masaki Saito

Takumi Information Technology Co., Ltd. Pharmaceutical Business Development Department, Strategic Sales Department

要旨

本研究では，NDB オープンデータ（2015～2023 年度のレセプトデータ及び特定健診データの簡易集計結果）を用いて，高血圧の有病状況と治療状況の乖離を全国，都道府県，性別・年齢層別に評価した．収縮期血圧 140mmHg 以上を高血圧患者とする場合を主解析とし，ポアソン回帰により年平均変化率（APC）を推定した．その結果，血圧降下薬の年間処方数量の APC は 1.6%，質問票における血圧降下薬使用割合の APC は 1.2%であった．一方で収縮期血圧 140mmHg 以上の割合と 160mmHg 以上の割合の APC はそれぞれ 1.2%および 2.1%であり，特に収縮期血圧 160mmHg 以上の割合の増加速度は，血圧降下薬処方数量の増加速度を上回っていた．また地域や性別・年齢層によって治療状況との乖離の程度は異なっていた．これらの結果から，地域や性別・年齢層によって異なる特徴的なパターンが観察され，未治療・未介入や治療継続に関するアンメットメディカルニーズ候補の探索につながる研究仮説を生成できることが示唆された．

キーワード：NDB オープンデータ，高血圧，ポアソン回帰

1 はじめに

高血圧は世界で約 14 億人が罹患する極めて頻度の高い疾患であり，重要な心血管リスク因子として知られている[1,2]．日本でも依然として高い有病割合を有し，50 歳以上の男性および 60 歳以上の女性では半数以上が高血圧に該当する[2]．高血圧は多くの場合無症状であるため，診断や治療の遅れにつながることがあり，さらに降圧薬治療を受けている患者においても血圧管理目標（収縮期血圧 140mmHg 未満，拡張期血圧 90mmHg 未満）の達成率は約 40～45%にとどまることが報告されている[1,3]．その結果，脳卒中や心筋梗塞・心不全・腎不全などに陥るケースが後を絶たず，国民の福祉や医療経済上も重大な問題になっている．また健康診断で高血圧が指摘されながらも医療機関受診や薬物治療に至っていない集団の存在が指摘されている[4]．そのため，高血圧の治療状況と血圧コントロールについて全国・都道府県レベルでのマクロな構造把握が必要とされる．

本研究では，厚生労働省ホームページで公開されているレセプト情報・特定健診情報データベース（NDB）オープンデータ[5]を利用して，日本における血圧高値患者の割合及び血圧降下薬処方状況の経年変化を調査し，両者の乖離や地域・性別・年齢層別の差異から特徴的なパターンを観察し，アンメットメディカルニーズ候補を探索することを目的とした．

2 方法

2.1 使用データ

NDB とは，厚生労働省保険局が保有するレセプト・特定健診等情報データベースであり，日本国民のほぼ全てのレセプト情報が集約されている．NDB オープンデータとは，NDB のレセプト・特定健診情報を簡易集計した結果を公表したものである．第 1 回では 2014 年度のレセプト情報及び 2013 年度の特定健診情報が公開されている．本研究では NDB オープンデータの第 2 回～第 11 回までを利用した．処方データは内服薬の「外来（院外）」「外来（院内）」「入院」を第 2 回～第 10 回（2015 年度～2023 年度）から取得した．特定健診データは「収縮期血圧」「拡張期血圧」「標準的な質問票（質問項目 1 ：血圧を下げる薬を現在使用しているか）」を第 3 回～第 11 回（2015 年度～2023 年度）から取得した．特定健診情報はレセプト情報より 1 年遅れて集計されているため，同一年度の解析が可能となるよう公開回次を調整して対応付けた．また公開品目数の変更による影響を避けるため，第 1 回の処方データは用いず，第 9 回以降の処方データからは上位 100 品目までに絞ったものを用いた. さらに, 2023 年度の都道府県別の年間処方数量と全国比を評価する際には，2023 年 10 月 1 日現在の総務省統計局人口推計のデータを用いて人口補正を行った[6].

表 1 特定健診及び標準的な質問票の項目とカテゴリー

項目	カテゴリー
収縮期血圧 (mmHg)	180 以上，160 以上 180 未満，140 以上 160 未満，130 以上 140 未満，120 以上 130 未満，120 未満
拡張期血圧 (mmHg)	110 以上，100 以上 110 未満，90 以上 100 未満，85 以上 90 未満，80 以上 85 未満，80 未満
質問項目 1 （現在，血圧を下げる薬を使用しているか）	はい/いいえ

2.2 定義

血圧降下薬は先行研究論文[7]に基づく一般名を対象とし, 当該研究と同様に内服薬を対象として定義した．全国及び性別・年齢層別の集計では，特定健診の受診者年齢層との整合性のため 40 歳～74 歳までとした．都道府県別の集計では，年齢の情報を持っていないため全年齢を対象とした．また特定健診の標準的な質問票における「血圧を下げる薬の使用の有無」に基づく副次解析を実施した．高血圧の指標は収縮期血圧が 140mmHg 以上と定義した．この閾値は高血圧の診断基準として広く用いられている収縮期血圧 140mmHg 以上または拡張期血圧 90mmHg 以上の基準に基づくものである[3]．副次解析として，収縮期血圧が 160mmHg 以上及び拡張期血圧が 90mmHg 以上も評価した．

2.3 統計手法

年次変化の評価にはポアソン回帰モデルを用いた．説明変数を年度，目的変数を年間処方数量または対象人数とし，回帰係数 β から以下の式により年平均変化率（APC）を算出した．

$$APC = (\exp(\beta) - 1) \times 100$$

高血圧割合および血圧降下薬使用割合のモデルでは，特定健診受診者数をオフセット項として組み込んだ．これにより，高血圧の有病状況と治療状況の経年変化を同一指標で比較可能とした．解析は全国，性別・年齢層別および都道府県別に実施した．また COVID-19 の影響を考慮し，2015 年度～2019 年度に限定した解析も実施した．地域差に関しては，2023 年度の各都道府県の人口で補正した年間処方数量と日本全国の年間処方数量の比についても算出し，2023 年度の各都道府県の処方状況を評価した．

すべての解析は SAS 9.4 を用いて実施した．

3 結果

3.1 年間推移

40 歳～74 歳における血圧降下薬の年間処方数量の年間推移を図 1 に，標準的な質問票の「血圧を下げる薬の使用」に「はい」と回答した割合の年間推移を図 2 に示す．年間処方数量は各年度で 50 億を超え，使用割合は各年度で 20%以上となった．年間処方数量と使用割合のいずれも増加傾向を示した．なお，全年齢を対象とした場合には各年度で 100 億を超えた．

収縮期血圧が 140mmHg 以上の割合の年間推移を図 3 に，160mmHg 以上の割合の年間推移を図 4 に，拡張期血圧が 90mmHg 以上の割合の年間推移を図 5 に示す．収縮期血圧が 140mmHg 以上は各年度で 16%以上となり，160mmHg 以上は 3%程度となり，拡張期血圧が 90mmHg 以上は 12%以上となった．いずれも増加傾向を示した．

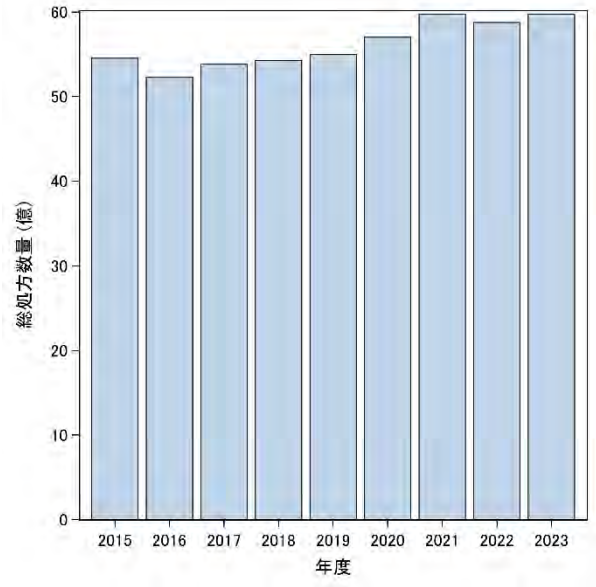


図 1 血圧降下薬の年間処方数量の推移

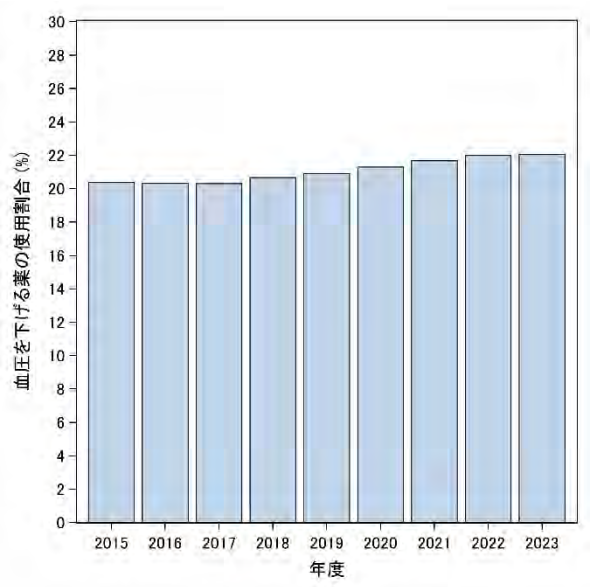


図 2 血圧を下げる薬の使用割合の推移

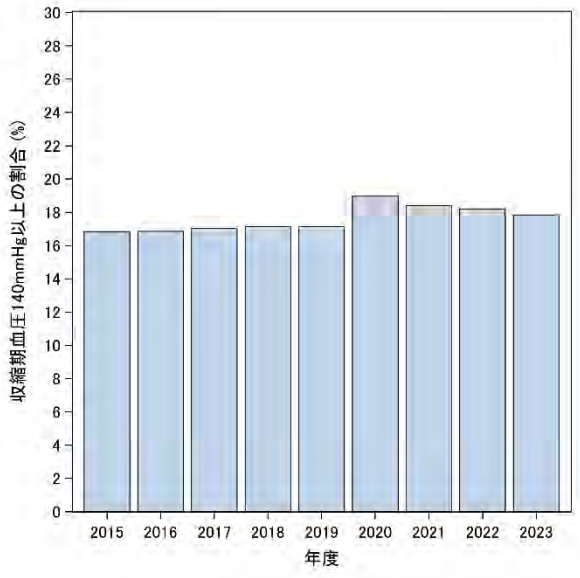


図 3 収縮期血圧が 140mmHg 以上の割合

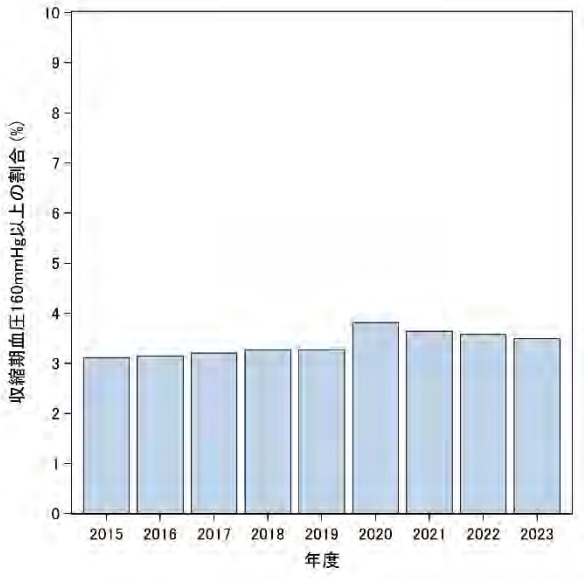


図 4 収縮期血圧が 160mmHg 以上の割合

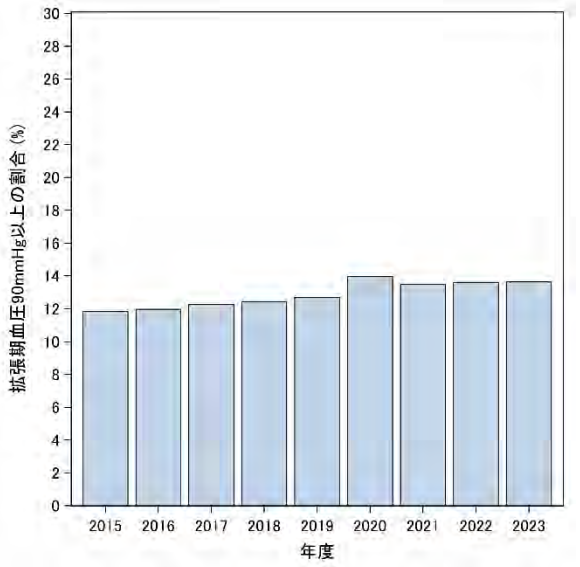


図 5 拡張期血圧が 90mmHg 以上の割合

3.2 全国の APC

全国では，40 歳～74 歳における血圧降下薬の年間処方数量の APC が 1.6%の増加を示し，質問票における使用割合の APC も 1.2%とやや緩やかだがどちらも増加傾向を示した．これに対し，高血圧患者については収縮期血圧が 140mmHg 以上の割合の APC が 1.2%の増加を示し，収縮期血圧が 160mmHg 以上の割合の APC では 2.1%と，より重症側で増加傾向が顕著であった．また拡張期血圧が 90mmHg 以上の割合の APC では 2.1%となり拡張期血圧でも同方向を示した．血圧高値の増加速度が，血圧降下薬使用の増加速度を上回っている事象が全国レベルで観察された．性別および年齢層を共変量として調整したモデルによる解析も実施したが，未調整の推定値とほぼ同様の結果となった．また COVID-19 の影響を考慮し 2015 年度～2019 年度を対象とした場合の APC を推定したが，いずれも同方向を示した．

3.3 サブグループ解析

都道府県別にみると，各 APC はほとんどの地域で正の値を示した．収縮期血圧が 140mmHg 以上の割合の APC については，山梨県は 3.7%，山形県は 3.1%と高血圧者割合の増加が顕著であった．一方，これらの地域においても血圧降下薬による治療の増加は限定的であった．収縮期血圧が 140mmHg 以上の割合の APC と血圧降下薬の年間処方数量の APC の散布図を図 6 に示す．2023 年度における各都道府県の人口当たり血圧降下薬処方数量の全国比には地域差が認められた．秋田県では 136%，青森県では 134%，山形県では 125%と全国平均を上回る一方，東京都では 85%，神奈川県では 88%，沖縄県では 89%と全国平均を下回っていた．2023 年度における人口当たり血圧降下薬処方数量の全国比と血圧降下薬年間処方数量の APC との関係を図 7 に示す．相関係数を見てみると，収縮期血圧が 140mmHg 以上の割合の APC と血圧降下薬年間処方数量の APC との間に相関は認められなかった（r＝-0.01）．また，2023 年度における人口当たり血圧降下薬処方数量の全国比と，収縮期血圧が 140mmHg 以上の割合の APC との間にも相関は認められなかった（r = 0.01）．一方，全国比と血圧降下薬年間処方数量の APC との間には中程度の負の相関が認められた（r＝-0.54）．感度解析として 2015 年度～2019 年度を対象として APC を推定すると，東京都や群馬県など一部の地域では COVID-19 流行期を含めた場合にのみ収縮期血圧が 140mmHg 以上の割合の APC が正の値となり増加を示した．ただし，血圧降下薬の年間処方数量については都道府県×年齢層の集計は存在しないため，40 歳～74 歳に絞っていない．

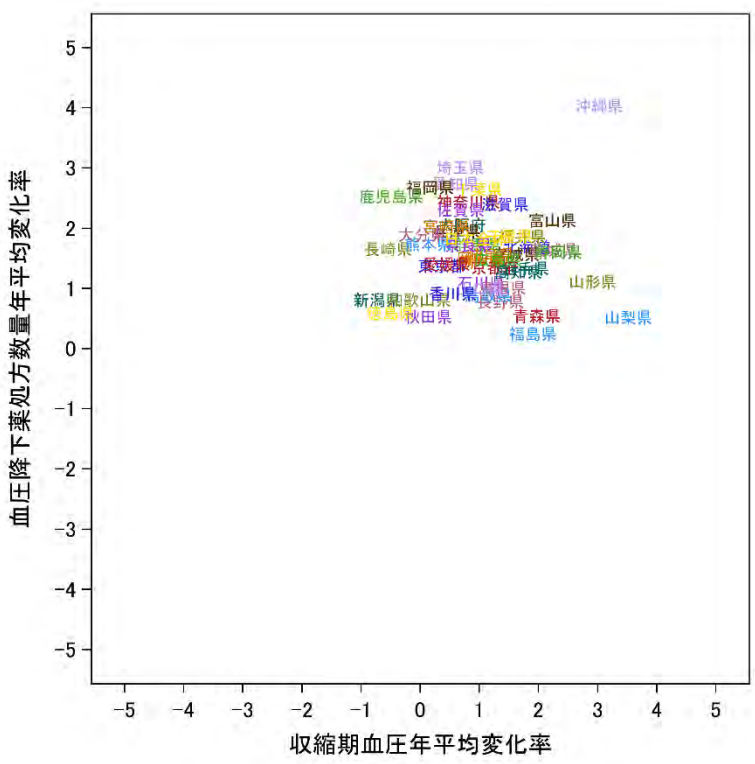


図 6 都道府県別の高血圧割合の APC と血圧降下薬処方数量の APC

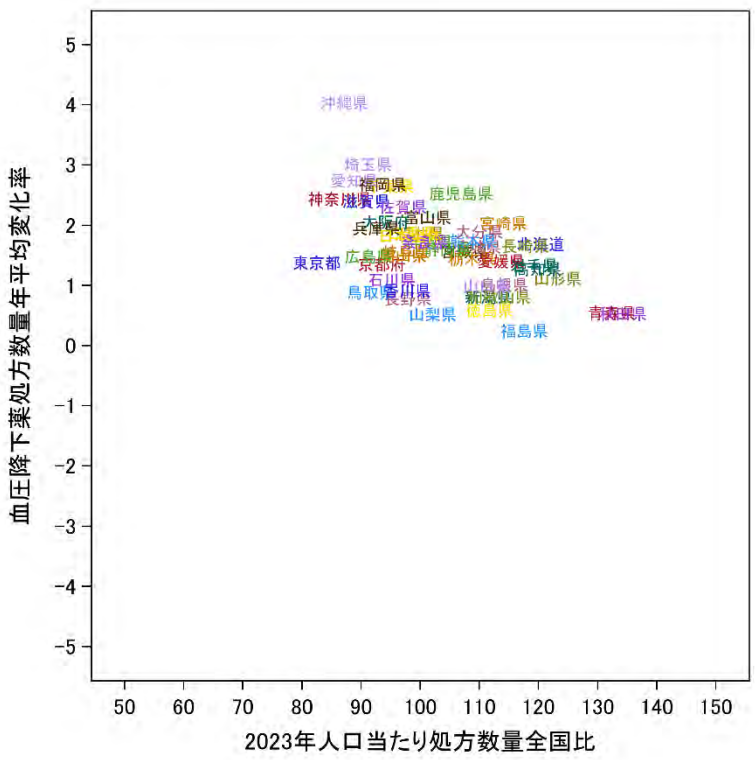


図 7 2023 年人口当たり処方数量全国比と血圧降下薬処方数量の APC

性別・年齢層ごとにみると，45～59 歳では血圧降下薬の年間処方数量の APC が男性も女性も増加し，60～64 歳では男性と女性で傾向が異なった．一方で 40～44 歳の男性及び 65～69 歳の男女では，血圧降下薬の年間処方数量の APC が減少する中で，収縮期血圧が 140mmHg 以上の割合の APC が増加していた．収縮期血圧が 140mmHg 以上の割合の APC と血圧降下薬の年間処方数量の APC の散布図を図 8 に示す.

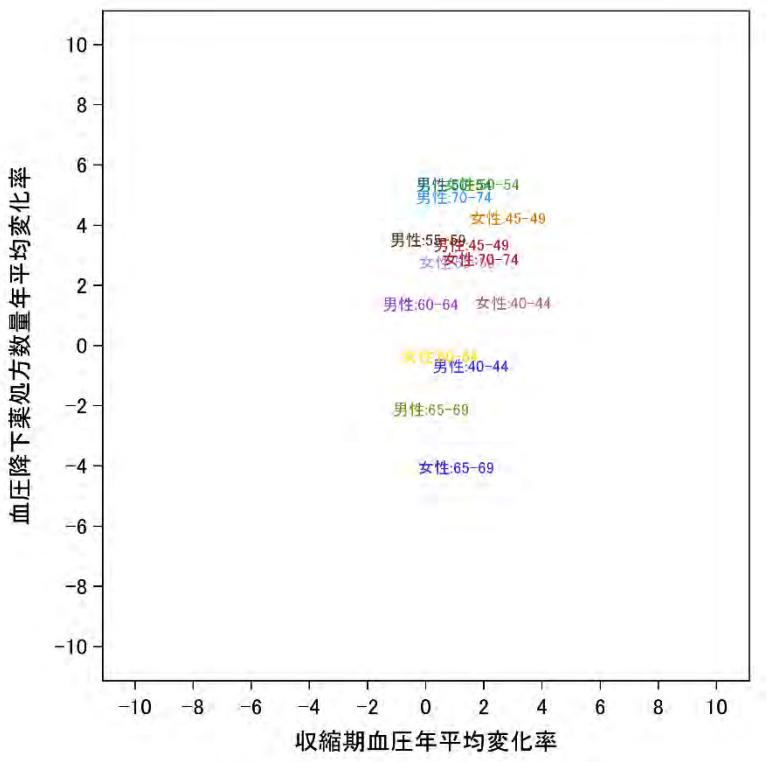


図 8 性別・年齢層別の高血圧割合の APC と血圧降下薬処方数量の APC

感度解析としてポアソン回帰ではなく対数線形回帰でも APC を推定したが，日本全体及び都道府県別や性別・年齢層別のサブグループ解析のいずれの結果もポアソン回帰とほぼ同じ結果となった．

4 考察

本解析により，NDB オープンデータから得られる血圧降下薬使用動向と特定健診血圧データを組み合わせることで，地域・性別・年齢層ごとの特徴的なパターンを観察し，アンメットメディカルニーズ候補となり得る課題を探索できることが示された．全国的には血圧降下薬の使用は増加しているものの，山梨県や山形県などでは高血圧者やハイリスク層（160mmHg 以上）の増加が上回っており，治療開始の遅れや治療中断，あるいは生活習慣介入のみで経過観察されているといった要因の存在が考えられた．一方で，2023 年時点の人口当たり年間処方数量には地域差が存在していた．高血圧割合の増加が顕著であった山形県では人口当たり処方数量も全国平均を上回っており，高血圧割合の増加は単純な治療不足だけでは説明できない可能性がある．また，処方数量水準が全国平均を下回る地域も存在し，未治療・未介入構造は地域ごとに異なる可能性がある．さらに，高血圧割合の地域差には治療状況だけでなく生活習慣要因の影響も考えられる．実際に，高血圧割合増加グループでは全国平均を上回る食塩摂取量が報告されており，食生活の地域差が関与している可能性も考えられる[8]．しかし東京都や群馬県など一部地域では COVID-19 流行期を含めた場合にのみ収縮期血圧が 140mmHg 以上の割合の APC の増加が観察されており，COVID-19 流行期に伴う特定健診受診状況の変化等の影響を受けたことが示唆された．人口当たり血圧降下薬処方数量の全国比と高血圧割合の増加との間に相関は認められず，高血圧割合の地域差は単純な治療量の差だけでは説明できない可能性が考えら

れる．一方で，人口当たり処方数量と年間処方数量の年平均変化率との間には負の相関が認められたことから，既に処方数量が多い地域では追加的な増加余地が小さい可能性が考えられる．

年齢層別には，特に 65 歳～69 歳の層で血圧降下薬の年間処方数量の APC が減少となっていた．定年退職後は被用者保険から国保へ移行することに伴い，健康管理や健診受診行動に変化が生じる可能性が指摘されている[9]．このことより，保険制度や受診行動の変化が影響している可能性が考えられる．

これらの結果から，個票 DB により検証が必要なアンメットメディカルニーズ候補としては以下の 3 つなどがあると考える．(1) 軽症～中等症高血圧段階での早期薬物治療導入，(2) 治療継続性を高める服薬アドヒアランス向上施策，(3) 生活習慣介入と薬物治療を組み合わせた包括的マネジメント，といった領域が挙げられる．特に 160mmHg 以上の重症高血圧者割合が増加している事実は，既存治療では十分にコントロールされていない患者層の存在や，未治療集団の存在や，重症高血圧患者増加につながる要因の存在が示唆された．

本研究の限界として，収縮期血圧や質問票は特定健診受診者を母集団とした傾向であり，一般人口における処方実態や有病率そのものを示すものではない．また NDB オープンデータは個票 DB ではなく簡易集計された結果の利用のため，特定健診受診者の背景情報の影響を完全に考慮することはできていない．

5 まとめ

本研究では，日本の高血圧患者の増加と薬物療法の増加の間に乖離の存在を示唆する結果を得た．NDB オープンデータは，個人の治療経過を追跡することはできない一方，全国規模で地域・性別・年齢層ごとの処方構造の異質性を俯瞰し特徴的なパターンを発見することで，次の研究仮説を生成するためのデータとして活用できる．今後，NDB などの個票レベルの医療情報 DB を活用することで，より詳細な検討が可能になると考えられる．第一に，特定健診で血圧高値と診断された後の治療開始や薬剤変更・中断に至るまでのペイシェントジャーニー分析である．これにより，治療がどの段階で途切れているかを可視化し，課題の所在を明確にすることが可能となる．第二に，リアルワールドデータを用いた治療効果や安全性の評価を通じて臨床および医療政策への応用が期待される．本解析で示されたマクロな乖離構造を起点として，医療情報 DB によるミクロ分析を重ねることで，高血圧領域におけるアンメットメディカルニーズに対応した戦略立案が可能になると考えられる．

参考文献・サイト

1. https://www.who.int/news-room/fact-sheets/detail/hypertension（最終アクセス日：2026-09-10）
2. Miura K, Nagai M, Ohkubo T. Epidemiology of Hypertension in Japan: Where Are We Now?. Circulation Journal. 2013;77(9):2226-2231. https://doi.org/10.1253/circj.CJ-13-0847
3. Miura K, Segawa H. わが国の高血圧の治療率とコントロール率の実態と考察. https://mhlw-grants.niph.go.jp/system/files/2018/182031/201809025A_upload/201809025A0013.pdf
4. Kitaoka K, et al. Blood pressure control and treatment status at 1 year after the first health check-up in individuals with observed referral-level blood pressure. Hypertension Research. 2025;48:2537–2547. https://doi.org/10.1038/s41440-025-02284-y
5. https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000177182.html（最終アクセス日：2026-09-10）
6. https://www.stat.go.jp/data/jinsui/2023np/index.html（最終アクセス日：2026-09-10）
7. Satoh M, et al. Differential coronary heart disease risk among antihypertensive and lipid-lowering medication users versus non-users: A real-world data analysis. Journal of Clinical Lipidology. 2025;19(2):294-302. https://doi.org/10.1016/j.jacl.2024.12.006
8. https://www.mhlw.go.jp/content/001675215.pdf（最終アクセス日：2026-09-10）
9. Koike S, et al. 定年退職等により新たに国民健康保険の被保険者になった者の特徴および国保連が行う保険者支援に関する実態調査.https://www.hws-kyokai.or.jp/images/ronbun/all/201907-01.pdf

逆ポーランド記法の実装

田中 真史

（イーピーエス株式会社）

Reverse Polish Notation via SAS

Masashi Tanaka

EPS Corporation

要旨

中置記法の計算規則を出発点にして、式の複雑化に伴う括弧の多用を避けるために考案されたポーランド記法及び逆ポーランド記法の仕組みと応用について解説した。ポーランド記法は、20 世紀初頭に括弧を一切使わずに論理式を表現する目的で導入され、この特性が計算機のスタック構造と相性が良いことから逆ポーランド記法として応用された。本稿では、SAS マクロやデータステップによる実装やテンパズルの解法における有用性を紹介する。最後に、臨床試験の解析業務への応用を考えた。

キーワード：中置記法、ポーランド記法、逆ポーランド記法、スタック、操車場アルゴリズム、テンパズル

1. 歴史と概略

1.1. 中置記法

3 世紀にアレクサンドリアのディオファントスが数式と呼べる表記を使って以来[1]、一貫して使われてきたのは、$1+2+3$のように演算子を数字の間に置いて接中辞（Infix、例：真っ黒）にする中置記法（ちゅうちきほう、Infix notation）であった。中置記法では演算の優先順位をあらかじめ決めておく必要があり、Base SAS も従う標準的なルールは次の通りである。

ルール 0　　括弧内から計算する。

ルール 1　　$*/$（乗除）、$+-$（加減）の順番で計算する。

ルール 2　　乗除と加減は左から計算する。　　$(2+3)/4*5=((2+3)/4)*5.$

冪乗が混ざった場合は、

ルール 1’　　^（冪乗）、$*/$（乗除）、$+-$（加減）の順番で計算する。

ルール 2’　　乗除と加減は左から、冪乗は右から計算する。

$$1/2/2=(1/2)/2=0.25,\qquad 2\text{^}3\text{^}2=2\text{^}(3\text{^}2)=512.$$

上のルールを気にしなくて良い場合も多い。加算（$+$）のみの場合は結合法則が成り立つのでどこから計算しても結果は変わらない。乗算（$*$）のみの場合も同様である。減算（$-$）も負数の加算と考えられる。

$$1+2+3=(1+2)+3=1+(2+3),$$
$$1*2*3=(1*2)*3=1*(2*3),$$
$$1-2-3=(1-2)-3=\big(1+(-2)\big)+(-3)=1+\big((-2)+(-3)\big)=1+(-2-3).$$

加減乗除が混ざると演算の順番によって結果が変わるのは周知のとおりである。ルール上括弧が不要であっても紛らわしい場合は括弧を付けた方が良い（例：$1/2/5 = 0.1$）。中置記法で括弧が増えることは避けられず、記号論理が発展した 20 世紀初頭の一部の論理学者にとって、この点が煩わしかった、あるいは本質的な問題だったようである。

1.2. ポーランド記法

括弧を使わずに演算の順番を一意的に表すために開発されたのが、演算子を前に置くポーランド記法（Polish notation）である。下にポーランド記法と中置記法を並べて書いた。ポーランド記法と後述の逆ポーランド記法ではトークン（数字や演算子）の間にスペースを置いて「$+123$」が「$1 + 23$」なのか「$12 + 3$」なのかを区別できるようにする。

ポーランド記法	ポーランド記法（括弧有り）	中置記法
$+\ +\ 1\,2\,3$	$+(+(1\ 2)\ 3)$	$(1 + 2) + 3$
$*\,/\,+\,2\,3\,4\,5$	$*\,(/(+(2\ 3)\ 4)\ 5)$	$((2 + 3)/4) * 5$
$*\ +1\,2\ -\ 3\,4$	$*\,\big(+(1\ 2) - (3\ 4)\big)$	$(1 + 2) * (3 - 4)$

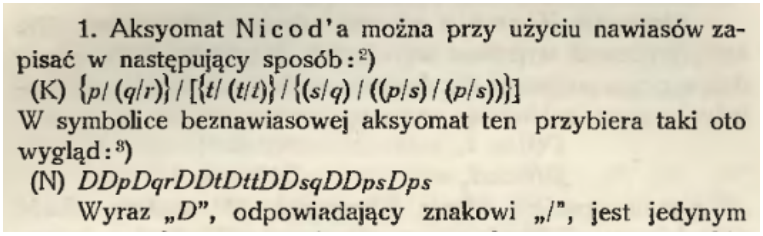
もともとポーランド記法は記号論理で括弧を使わず論理式を表すための記法で、ポーランドの論理学者 Jan Łukasiewicz（1878–1956）が着想した[2]。1931 年の論文[3]（下図参照）の冒頭で、ニコの公理を中置記法の

$$\{p|(q|r)\}|\big[\{t|(t|t)\}\big|\big\{(s|q)\big|\big((p|s)\big|(p|s)\big)\big\}\big]$$

から、ポーランド記法の

$$DDpDqrDDtDttDDsqDDpsDps$$

に書き換えて（|とDは NAND 演算子）、この記法を 1924 年に思いついた旨を脚注に記している（英訳 ”The idea for bracketless symbolism occurred to me in 1924. I first used this symbolism in print in the article by Łukasiewicz (1)”、(1)は 1929 年の論文）。



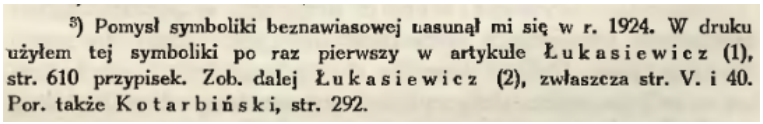
1. Aksyomat Nicod'a można przy użyciu nawiasów zapisać w następujący sposób:[2])

(K) $\{p/(q/r)\}/[\{t/(t/t)\}/\{(s/q)/((p/s)/(p/s))\}]$

W symbolice beznawiasowej aksyomat ten przybiera taki oto wygląd:[3])

(N) $DDpDqrDDtDttDDsqDDpsDps$

Wyraz „D", odpowiadający znakowi „/", jest jedynym



[3]) Pomysł symboliki beznawiasowej nasunął mi się w r. 1924. W druku użyłem tej symboliki po raz pierwszy w artykule Łukasiewicz (1), str. 610 przypisek. Zob. dalej Łukasiewicz (2), zwłaszcza str. V. i 40. Por. także Kotarbiński, str. 292.

結局のところポーランド記法は記号論理の標準的な記法にはならなかった。しかし、その後現れた計算機に逆ポーランド記法として組み込まれることになった。

1.3. 逆ポーランド記法

ポーランド記法と逆に数値を前に出して演算子を後に置くのが逆ポーランド記法（Reverse Polish notation）である。例として逆ポーランド記法、ポーランド記法、中置記法を並べた。

逆ポーランド記法	ポーランド記法	中置記法
1 2 + 3 +	+ + 1 2 3	(1 + 2) + 3
2 3 + 4 / 5 *	* / + 2 3 4 5	((2 + 3)/4) * 5
1 2 + 3 4 − *	* + 1 2 − 3 4	(1 + 2) * (3 − 4)

逆ポーランド記法の歴史を瞥見すると、計算機への応用は 1940 年代に遡り、Konrad Zuse による 1941 年に完成した計算機「Z3」への実装が最初である[4]。最初期の論文として、1954 年の Burks、Warren 及び Wright による研究[5]（下図参照）が知られている。同論文は、実際のところはポーランド記法を用いており、ポーランド記法の式を評価する際は右から左へ 1 回走査すれば十分であることを注意している。逆ポーランド記法とほぼ等価であるが、ずいぶんあっさりと記述している感はある。

(*Example 3: pKqN* is not a positive formula; *ApqNr* is positive but not well-formed; *KpNq* is well-formed.)

Note that the non-recursive definition 3F shows that a single scan of a formula from right to left is sufficient to determine whether it is well-formed.

逆ポーランド記法はスタック構造で計算できるので計算機と相性が良い。与えられた式を前から読んでトークンが数値であればスタックに積んで（プッシュ）、演算子であればスタックから数値をふたつ取り出して（ポップ）、それらを演算する単純なルールで計算できる。1 2 + 3 4 −（中値記法は(1 + 2) * (3 − 4)）を計算する例を示す。

トークン	スタック	演算
1	1	
2	1, 2	
+	3	= 1 + 2
3	3, 3	
4	3, 3, 4	
−	3, −1	= 3 − 4
*	−3	= 3 * (−1)

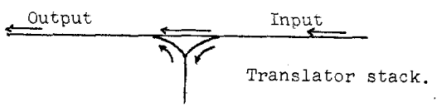
1960 年代に入って Dijkstra が文献[6]（下図参照）で中置記法の式を逆ポーランド記法に変換する操車場アルゴリズム（Shunting yard algorithm）を開発した。中置記法の演算子を演算子スタック（操車場の側線）に退避させて、適切にポップすることで逆ポーランド記法に変換できる（付録 3）。

computation of

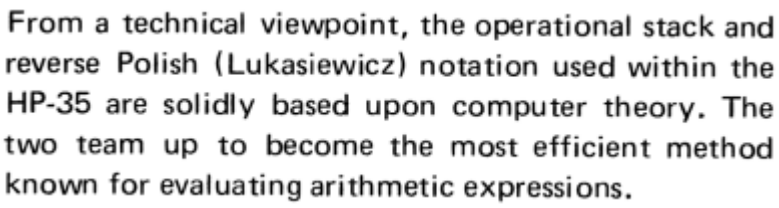
A +(B - C) * D + E

in the following form:

```
TAKE A          v0:= A
TAKE B          v1:= B
TAKE C          v2:= C
SUBTRACT        v1:= v1 - v2
TAKE D          v2:= D
MULTIPLY        v1:= v1 * v2
ADD             v0:= v0 + v1
TAKE E          v1:= E
ADD             v0:= v0 + v1
```



製品への応用も進み、良く知られたところでは、Hewlett-Packard 社の関数電卓（HP-35、1972 年）が逆ポーランド記法による入力を採用した[7][8]。下図は HP-35 のマニュアルで、逆ポーランド記法のことを、Polish と Łukasiewicz を併記して「reverse Polish (Lukasiewicz) notation」と呼んでいる。逆ポーランド記法の説明としては、商業と研究との違いは当然としても、率直に言ってこのマニュアルは研究論文よりも格段に読みやすかった。関数電卓が逆ポーランド記法の普及にどの程度寄与したかは分からなかったが、多くの技術者に逆ポーランド記法を広めたことは想像できる。


From a technical viewpoint, the operational stack and reverse Polish (Lukasiewicz) notation used within the HP-35 are solidly based upon computer theory. The two team up to become the most efficient method known for evaluating arithmetic expressions.

2. SAS による実装

2.1. データステップで逆ポーランド記法を使うためのマクロ

SAS のデータステップは逆ポーランド記法をサポートしていない。逆ポーランド記法の式を評価する場合は、逆ポーランド記法を中置記法に変換する処理を加える。データステップで「weight height 2 ** /」、すなわち BMI を計算する例を下に示した（本稿のコードはすべて Base SAS 9.4M8 で実行確認した）。マクロ RPN に直接逆ポーランド記法で式を書いて実行することができる。なお、SASHELP.CLASS の身長と体重の単位は（おそらく）インチとポンドなので計算結果の単位は$\mathrm{lb/in^2}$である。

```
%macro rpn(expression);
     %let top = 0;
     %do i = 1 %to %sysfunc(countw(&expression., %str( )));
          %let token = %scan(&expression., &i., %str( ));
          %if %sysfunc(whichc(&token., +, -, *, /, **)) %then %do;
               %let val2 = &&stack&top..;
               %let top  = %eval(&top. - 1);
               %let val1 = &&stack&top..;
               %let stack&top. = (&val1. &token. &val2.);
          %end;
          %else %do;
               %let top = %eval(&top. + 1);
               %let stack&top. = &token.;
          %end;
     %end;
     %do;
          &&stack&top.
     %end;
%mend rpn;
data bmi;
     set sashelp.class;
     bmi1 = weight/height**2;
     bmi2 = %rpn(weight height 2 ** /);
run;
```

OBS	Name	Sex	Age	Height	Weight	bmi1	bmi2
1	Alfred	M	14	69.0	112.5	0.023629	0.023629
2	Alice	F	13	56.5	84.0	0.026314	0.026314
3	Barbara	F	13	65.3	98.0	0.022983	0.022983
4	Carol	F	14	62.8	102.5	0.025990	0.025990

逆ポーランド記法を中値記法に変換するのは案外簡単だった。一方、中置記法を逆ポーランド記法に変換するのは手間がかかる（付録 3）。この差は、逆ポーランド記法が機械的な処理に向いていることの表れだと思う。

2.2. データセットで与えられた逆ポーランド記法による式の計算

逆ポーランド記法による式がデータセットで与えられたときに、その値を計算するコードを付録 1 に示した。コードの中の重要な箇所を下に抜き出した。データステップの配列をスタックに見立てて、トークンが数値だったらスタックに積んで、トークンが演算子だったらスタックから数値を取り出して計算することで結果が得られる。

```
if token in ('+', '-', '*', '/') then do;
     val2 = stack[top];
     top  = top - 1;
     val1 = stack[top];
     if     token = '+' then res = val1 + val2;
     else if token = '-' then res = val1 - val2;
     else if token = '*' then res = val1 * val2;
     else if token = '/' then res = val1 / val2;
     stack[top] = res;
end;
else do;
     top = top + 1;
     stack[top] = input(token, best.);
end;
```

OBS	expression	result	error
1	10 5 3 + *	80	
2	20 8 2 / -	16	
3	1 2 3 + 4 5 * + *	25	

2.3. テンパズルへの応用

1 桁の数字 4 つから加減乗除（など）を用いて 10 を作る遊びを「テンパズル」と呼ぶらしい。ここでは 9 を 4 つ使って四則演算のみで 10 を作ることを考える。例えば、

$$9 + 9 + 9 - 9 = 18 \neq 10,$$

$$(9 + 9)/9 + 9 = 11 \neq 10.$$

ひとつめは問題外でふたつめは惜しい。いじわるクイズは意図していないので$(99 - 9)/9 = 10$は無しとする。おそらく上手い解法は無くて、総当たりで調べるのが確実なようである。候補の式のすべてを作成する際に、括弧が不要な逆ポーランド記法が便利である。下のデータセットを用意して前節のコードで式の値を計算すれば正解が得られる（付録 2）。9 が 4 つの場合、10 を作る方法は実質的に一通りのみだった。答えは付録のコードで確かめてほしい。

OBS	expression	x1	x2	x3	x4	x5	x6	x7
1	＊＊＊9999	＊	＊	＊	9	9	9	9
2	＊＊＋9999	＊	＊	＋	9	9	9	9
3	＊＊－9999	＊	＊	－	9	9	9	9
4	＊＊／9999	＊	＊	／	9	9	9	9
5	＊＊9＊999	＊	＊	9	＊	9	9	9
6	＊＊9＋999	＊	＊	9	＋	9	9	9

3. まとめ — 臨床試験の解析業務への応用に向けて

逆ポーランド記法は、臨床試験の解析業務では通常使わない。しかしながら、括弧が多い数式を実装する際、逆ポーランド記法を使えば括弧の誤記は原理的になくなる。また、ダブルプログラミングの片側が逆ポーランド記法を使えばその部分の品質は上がるかもしれない。熟練すれば逆ポーランド記法の方が中置記法よりも式の入力が早くなるはずなので、練習次第で電卓による検算を効率化できるかもしれない。ポーランド記法の原点に戻って論理演算を考えると、間違えやすい(A and B) or CとA and (B or C)の表記が大きく変わる。この性質は、EDC（Electronic Data Capture）のロジカルチェックなどで複雑な抽出条件を扱う際に有用かもしれない。

中置記法	逆ポーランド記法
(A and B) or C	A B and C or
A and (B or C)	A B C or and

中置記法に慣れきってしまった私たちには容易ではないと思われるが、テンパズル以外に逆ポーランド記法の応用があれば、筆者としては幸いである。

参考文献

1. Boyer, C. B., & Merzbach, U. C. (2011). *A history of mathematics (3rd ed.)*. John Wiley & Sons.
2. Jan Łukasiewicz. (2026, August 4). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Jan_%C5%81ukasiewicz&oldid=1367663303
3. Łukasiewicz, J. (1931). Uwagi o aksjomacie Nicoda i "dedukcji uogólniającej" [Comments on Nicod's Axiom and on "Generalizing Deduction"]. *In Księga pamiątkowa Polskiego Towarzystwa Filozoficznego We Lwowie, 12. II.* 1904-12. II. 1929 (pp. 366–383). Wydawnictwo Polskie Towarzystwo Filozoficzne. https://sbc.org.pl/dlibra/publication/19517/edition/18864 [リンク切れ]
4. Reverse Polish notation. (2026, August 18). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Reverse_Polish_notation&oldid=1370001662
5. Burks, A. W., Warren, D. W., & Wright, J. B. (1954). An analysis of a logical machine using parenthesis-free notation. *Mathematical Tables and Other Aids to Computation, 8(46)*, 53–57. https://scispace.com/pdf/an-analysis-of-a-logical-machine-using-parenthesis-free-4pbw5u8xbx.pdf
6. Dijkstra, E. W. (1961). Making a translator for ALGOL 60. *APIC-Bulletin, (7)*. https://www.cs.utexas.edu/~EWD/MCReps/MR35.PDF
7. HP-35. (2025, December 13). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=HP-35&oldid=1327204043
8. Hewlett-Packard. (1972). *HP-35 operating manual*. https://literature.hpcalc.org/community/hp35-om-en-reddot.pdf

付録 1　逆ポーランド記法の計算

```
data rpn;
     length expression $200;
     expression = '10 5 3 + *'; output;
     expression = '20 8 2 / -'; output;
     expression = '1 2 3 + 4 5 * + *'; output;
run;
data rpn_calc;
     length expression $200 result 8 error token $20;
     array stack[20] _temporary_;
     set rpn;
     top = 0;
     call missing(of stack[*]);
     do i = 1 to countw(expression, ' ');
          token = scan(expression, i, ' ');
          if token in ('+', '-', '*', '/') then do;
               if top = 0 then do;
                    error = 'Invalid Expression';
                    goto label;
               end;
               val2 = stack[top];
               top  = top - 1;
               if top = 0 then do;
                    error = 'Invalid Expression';
                    goto label;
               end;
               val1 = stack[top];
               if     token = '+' then res = val1 + val2;
               else if token = '-' then res = val1 - val2;
               else if token = '*' then res = val1 * val2;
               else if token = '/' then do;
                    if val2 = 0 then do;
                         error = 'Division by Zero';
                         goto label;
                    end;
                    else do;
                         res = val1 / val2;
                    end;
               end;
               stack[top] = res;
          end;
          else do;
               top = top + 1;
               stack[top] = input(token, best.);
          end;
     end;
     if top ^= 1 then do;
          error = 'Invalid Expression';
          goto label;
     end;
     result = stack[top];
     label:;
     keep expression result error;
run;
```

付録 2　テンパズル用のデータセット

```
%let n = 9;
data tokens;
     length x $20;
     x = "&n."; output;
     x = '+'; output;
     x = '-'; output;
     x = '*'; output;
     x = '/'; output;
run;
proc sql;
     create table wk01 as
     (
     select t1.x as x1
          , t2.x as x2
          , t3.x as x3
          , t4.x as x4
          , t5.x as x5
          , t6.x as x6
          , t7.x as x7
     from tokens as t1
          , tokens as t2
          , tokens as t3
          , tokens as t4
          , tokens as t5
          , tokens as t6
          , tokens as t7
     )
     ;
quit;
data rpn;
     length expression $200;
     set wk01;
     expression = catx(' ', of x1-x7);
     if countc(expression, "&n.") = 4;
run;
proc sort data = rpn;
     by expression;
run;
```

付録 3　操車場アルゴリズム

中置記法を逆ポーランド記法に変換する操車場アルゴリズムの実装例を示す。「tokenization」のステップでトークンの切れ目を解析した後、「shunting_yard」のステップで中置記法を前から読んで、演算子と数値に分類しながら逆ポーランド記法を組み立てた。

OBS	expression	expr	rpn
1	4*-2	4 * -2	4 -2 *
2	3+(5)	3 + (5)	3 5 +
3	(3+(-5))	(3 + (-5))	3 -5 +
4	2**3**4	2 ^ 3 ^ 4	2 3 4 ** **
5	(3+5)*2**4-6	(3 + 5) * 2 ^ 4 - 6	3 5 + 2 4 ** * 6 -
6	((7-2)**3+4)/(5*(-6-1))	((7 - 2) ^ 3 + 4) / (5 * (-6 - 1))	7 2 - 3 ** 4 + 5 -6 1 - * /
7	3+4*-2/(1-5)**2	3 + 4 * -2 / (1 - 5) ^ 2	3 4 -2 * 1 5 - 2 ** / +
8	((2+3**2**2)/(4-1))**2+5*(6-2**3)	((2 + 3 ^ 2 ^ 2) / (4 - 1)) ^ 2 + 5 * (6 - 2 ^ 3)	2 3 2 2 ** ** + 4 1 - / 2 ** 5 6 2 3 ** - * +
9	((1+2)*3-4)/(5-6/3)	((1 + 2) ^ 3 - 4) / (5 - 6 / 3)	1 2 + 3 ** 4 - 5 6 3 / - /
10	8+((3+1)*2*(5-3)-9)/5	8 + ((3 + 1) * 2 * (5 - 3) - 9) / 5	8 3 1 + 2 * 5 3 - * 9 - 5 / +

```
data infix_notation;
     length expression $200;
     expression = '4*-2'; output;
     expression = '3+(5)'; output;
     expression = '(3+(-5))'; output;
     expression = '2**3**4'; output;
     expression = '(3+5)*2**4-6'; output;
     expression = '((7-2)**3+4)/(5*(-6-1))'; output;
     expression = '3+4*-2/(1-5)^2'; output;
     expression = '((2+3**2**2)/(4-1))**2+5*(6-2**3)'; output;
     expression = '((1+2)**3-4)/(5-6/3)'; output;
     expression = '8+((3+1)*2*(5-3)-9)/5'; output;
run;
proc format;
     invalue order
          '^'      = 3
          '*', '/' = 2
          '+', '-' = 1
          ;
run;
data tokenization;
     length expression _expr expr $200;
     set infix_notation;
     _expr = compress(tranwrd(expression, '**', '^'));
     call missing(expr);
     do i = 1 to length(_expr);
          char = char(_expr, i);
          if char = '.' then do;
               expr = cats(expr, char);
          end;
          else if indexc(char, '1234567890') then do;
               if indexc(char(_expr, i - 1), '1234567890.') then do;
                    expr = cats(expr, char);
               end;
               else if char(_expr, i - 1) in ('+', '-')
                         and char(_expr, i - 2) in ('*', '/', '^', '(', '') then do;
                    expr = cats(expr, char);
               end;
               else do;
                    expr = catx(' ', expr, char);
               end;
          end;
          else do;
               expr = catx(' ', expr, char);
          end;
     end;
     keep expression expr;
run;
data shunting_yard;
     length expression expr token operator_stack rpn $200;
     set tokenization;
     do i = 1 to countw(expr, ' ');
          token = scan(expr, i, ' ');
          fstopr = first(operator_stack);
          if indexc(token, '1234567890') then do;
               rpn = catx(' ', rpn, token);
          end;
          else do;
```

```
            if missing(operator_stack) then do;
                operator_stack = token;
            end;
            else if token = '(' then do;
                operator_stack = cats('(', operator_stack);
            end;
            else if token = ')' then do;
                if first(operator_stack) = '(' then do;
                    operator_stack = substr(operator_stack, 2);
                end;
                else do;
                    rpn = catx(' ', rpn,
                        prxchange('s/([¥+¥-¥*¥/¥^])/ $1 /', -1, scan(operator_stack, 1, '(')));
                    operator_stack = substr(operator_stack, index(operator_stack, '(') + 1);
                end;
            end;
            else do;
                if fstopr = '(' or (input(token, order.) > input(fstopr, ??order.))
                    or (token = fstopr = '^') then do;
                    operator_stack = cats(token, operator_stack);
                end;
                else do;
                    do while(not missing(operator_stack)
                        and input(token, order.) <= input(first(operator_stack), ??order.));
                        rpn = catx(' ', rpn, first(operator_stack));
                        operator_stack = substr(operator_stack, 2);
                    end;
                    operator_stack = cats(token, operator_stack);
                end;
            end;
        end;
        if i = countw(expr, ' ') then do;
            rpn = catx(' ', rpn, prxchange('s/([¥+¥-¥*¥/¥^])/ $1 /', -1, operator_stack));
            call missing(operator_stack);
        end;
    end;
    rpn = compbl(rpn);
    rpn = tranwrd(rpn, '^', '**');
    keep expression expr rpn;
run;
```

抗がん剤第I相試験における3+3，BOIN及びBOIN12の用量選択特性：SASを用いたシミュレーションによる評価

○牛尾 英孝

（株式会社ピープロ・ジャパン）

Dose Selection of the 3+3, BOIN, and BOIN12 Designs in Phase I Oncology Trials
: A Simulation Study Using SAS

Hidetaka Ushio
Ppro-Japan Co.,Ltd.

要旨

3+3，BOIN 及び BOIN12 について，毒性・有効性の用量反応関係が用量選択に及ぼす影響を SAS によるシミュレーションにより評価し，Target Dose の正選択率，選択用量の方向・程度及び最大耐用量と最適生物学的用量の関係を比較する．

キーワード：3+3 デザイン，BOIN（Bayesian Optimal Interval），BOIN12

1　はじめに

抗がん剤第 I 相試験の用量探索では，従来から 3+3 デザインが広く用いられている．一方，近年では，目標毒性率に基づき，観察された DLT 発現割合と目標毒性率との関係に応じて用量増減を決定する BOIN や，毒性及び有効性を utility として統合的に評価し，毒性が許容範囲内で，かつ有効性を考慮すると最も望ましい用量である OBD を選択する BOIN12 が提案されている．本演題では，SAS による共通のシミュレーション環境下で 3+3，BOIN 及び BOIN12 を比較し，毒性及び有効性の用量反応が各デザインの用量探索及び用量選択特性に及ぼす影響を評価する．毒性及び有効性の用量反応が異なる複数の Scenario を設定し，各デザインについて，Target Dose を正しく選択する割合，Target Dose 未満の用量を選択する割合及び Target Dose を超える用量を選択する割合を評価する．さらに，最終選択用量と Target Dose との差を評価し，誤選択の方向及び程度を比較する．また，BOIN12 については，これらの指標に加えて，OBD を選択する割合及び MTD と OBD の一致率を評価し，毒性及び有効性の用量反応が異なる状況における各デザインの用量選択特性を多面的に検討する．

2　Scenario 設定

用量探索における真の DLT 発現割合及び真の有効性割合の用量反応関係が各デザインの用量選択特性に及ぼす影響を評価するため，6 段階の用量（Dose 1～Dose 6）を設定し，目標 DLT 発現割合を 30%とした．全ての Scenario において，目標 DLT 発現割合に最も近い真の DLT 発現割合を Target Dose と定義し，本 Scenario では Dose 4 を Target Dose とした．Scenario 1～3 では DLT 発現割合の用量反応関係を変化させ，Scenario 1 では用量増加に伴って DLT 発現割合が緩やかに上昇する標準的な毒性反応曲線，Scenario 2 では Dose 4 を超えた用量で DLT 発現割合が急峻に上昇する毒性反応曲線，Scenario 3 では Dose 4 周辺の用量で DLT 発現割合が目標 DLT 発現割合に近く，DLT 発現割合が緩やかに変化する毒性反応曲線を設定した．Scenario 4～5 では DLT 発現割合の用量反応関係を共通とし，有効性の plateau の程度を変化させた．これらの Scenario（表 2.1）を 3+3，BOIN 及び BOIN12 に共通して適用し，DLT 発現割合及び有効性の用量反応関係の違いが各デザインの用量選択特性に及ぼす影響を比較した．

[表 2.1：Scenario 1～Scenario 5 における真の DLT 発現割合及び真の有効性割合]

Scenario	DLT probability (%)	Efficacy probability (%)
Scenario 1	5 / 10 / 20 / 30 / 40 / 50	10 / 20 / 30 / 40 / 50 / 60
Scenario 2	5 / 10 / 20 / 30 / 50 / 70	10 / 20 / 30 / 40 / 50 / 60
Scenario 3	5 / 15 / 25 / 30 / 35 / 40	10 / 20 / 30 / 40 / 50 / 60
Scenario 4	5 / 10 / 25 / 30 / 50 / 70	10 / 25 / 40 / 50 / 65 / 68
Scenario 5	5 / 10 / 25 / 30 / 50 / 70	10 / 25 / 40 / 55 / 55 / 55

DLT 及び Efficacy probability は Dose 1，Dose 2，Dose 3，Dose 4，Dose 5，Dose 6 の順に示した．

3　評価項目

各 Scenario 1～Scenario 5 及び各用量探索デザイン（3+3，BOIN 及び BOIN12）について，最終的に選択された用量を Dose 4（Target Dose）と比較し，以下の指標を算出する．

- 最終選択用量が Dose 4 である割合を算出し，Target Dose を正しく選択できる確率（Correct selection）
- 最終選択用量が Dose 4 未満（D1～D3）である割合，Target Dose より低い用量を選択する確率（Under-dose）
- 最終選択用量が Dose 4 を超える（D5～D6）割合，Target Dose より高い用量を選択する確率（Over-dose）
- Target Dose を超える用量に割り付けられた割合（Over-dose allocation）
- 最終選択用量と Target dose（Dose 4）との差，選択用量が Target Dose からどの程度離れているか（*）

（*）この値が 0 の場合は Target dose を正しく選択したことを示し，負の値は under-dose、正の値は over-dose を示す。さらに，各値（−3～+2）の発生割合を評価することで，誤選択の方向及び程度を比較する．

なお，BOIN12 については，これらに加えて，下記も評価する．

- 最終選択用量が OBD である割合を算出し，OBD を正しく選択できる確率（OBD selection probability）．
- 毒性の観点から定義した MTD と，毒性及び有効性を考慮して定義した OBD が一致する割合．

4　各デザインの概要及び SAS によるシミュレーション

4.1　[3+3]

第 I 相試験で広く用いられてきた Rule-based の用量探索デザインであり，3 例を 1 コホートとして，DLT が 0 例の場合は増量，1 例の場合は同一用量に 3 例を追加，2 例以上の場合は増量を中止する．6 例投与した場合は，DLT が 1 例以下であれば増量し，2 例以上であれば増量を中止する．一般的には，6 例中 1 例以下で DLT が認められた最高用量を MTD として選択する．シミュレーションに用いた主要な SAS プログラムのアルゴリズムを下記へ示す．なお，最大症例数は 36 例として，シミュレーション回数は 10,000 回とした．BOIN，BOIN12 も同様で 1 コホートあたり 3 例とした．

```
/**********************/
/* 3+3 Simulation          */
/**********************/
/*array tox[&ndose] _temporary_(&tox);にて各用量の真のDLT発現割合を設定*/
/*各被験者3例ずつを1コホートとしてDLT発現をBernoulli分布により発生*/
        dlt = 0;
        do i = 1 to 3; /*[level] :  各 Scenario における DLT probability tox[1]～tox[6]が dose1～dose6 に対応*/
                dlt = dlt + rand("Bernoulli",tox[level]);
        end
/*DLT発現例数に基づく用量エスカレーション*/
/* 0/3 DLT →  増量*/
        if dlt = 0 then do;
                if level < &ndose then level = level + 1;     /*&ndose=6*/
                else do;
                        mtd = level;
                        stop_trial = 1;
                end;
        end;
/* 1/3 DLT →  同用量に 3 例追加*/
        else if dlt = 1 then do;
                dlt2 = 0;
                do i = 1 to 3;
                        dlt2 = dlt2 + rand("Bernoulli",tox[level]);
                end;
/*現在の用量に3例追加し，DLT数と総症例数を更新  n_dose[1]であればdose1に割付られた例数*/
                n_dose[level] = n_dose[level] + 3;
                n_dlt[level]    = n_dlt[level] + dlt2;     /*n_dlt[1]であれば dose1 で発生した DLT 数*/
                n_pts           = n_pts + 3;
```

```
/* 1/6 以下 → 増量 */
                    if n_dlt[level] <= 1 then do;
                              if level < &ndose then level = level + 1;     /*&ndose=6*/
                              else do;
                                        mtd = level;
                                        stop_trial = 1;
                              end;
                    end;
/* 2/6 以上 → 1 つ下の level を MTD */
                    else do;
                              mtd = ifn(level > 1,level - 1,.);
                              stop_trial = 1;
                    end;
          end;
/* 2/3 以上 → 1 つ下の用量を MTD*
          else if dlt >= 2 then do;
                    mtd = ifn(level > 1,level - 1,.);
                    stop_trial = 1;
          end;
```

4.2 [BOIN]

Model-assisted 型の用量探索デザインであり，DLT 発現率と目標毒性率に基づき増量・維持・減量を決定し，isotonic regression により DLT 発現率を推定し，目標毒性率に最も近い用量を MTD として選択する．シミュレーションに用いた主要な SAS アルゴリズムを下記へ示す．

```
/***********************/
/* BOIN                */
/***********************/
data _null_;   /*用量移行境界値を算用*/
          target = &target;     /*&target=0.30*/
          p_saf = 0.6 * target;
          p_tox = 1.4 * target;
          lambda_e = log((1-p_saf)/(1-target))
                    / log((target*(1-p_saf))/(p_saf*(1-target)));
          lambda_d = log((1-target)/(1-p_tox))
                    / log((p_tox*(1-target))/(target*(1-p_tox)));
          call symputx("lambda_e",lambda_e);
          call symputx("lambda_d",lambda_d);
run;
```

```
/* Safety elimination */
        if n_dose[level] >= 3 then do;
                alpha = n_dlt[level] + 1;
                beta = n_dose[level] - n_dlt[level] + 1;
                prob_over = 1 - cdf("BETA",&target,alpha,beta);
                if prob_over > &cutoff_eli then do;     /*&cutoff=0.95*/
                        do k = level to &ndose;
                                elimi[k] = 1;
                        end;
                end;
```

当該用量 n_dose[level]に 3 例以上投与された場合について，DLT 発現率の事後分布 Beta([alpha],[beta])=Beta(n_dlt[level] + 1, n_dose[level] - n_dlt[level] + 1)のパラメータを設定し，目標 DLT 発現率を超える事後確率[prob_over]を算出する．この確率があらかじめ設定した基準[cutoff_eli=0.95]を超えた場合には，当該用量およびそれ以上の用量を以後の投与対象から除外した．

```
/* 推定 DLT 発現率に基づく用量移行 */
p_hat = n_dlt[level] / n_dose[level];
/* 増量境界値（lambda_e=0.236）以下の場合は増量 */
        if p_hat <= &lambda_e then do;
                if level < &ndose and elimi[level+1] = 0 then level = level + 1;
        end;
/* 減量境界値（lambda_d=0.358）より大きい場合は減量 */
        else if p_hat > &lambda_d then do;
                if level > 1 then level = level - 1;
                else do;
                        stop_trial = 1;  /*  試験中止フラグを設定  */
                        leave;
                end;
        end;
```

当該用量 n_dose[level]における DLT 発現率の推定値 p_hat を算出し，あらかじめ設定した増量境界値 lambda_e 及び減量境界値 lambda_d と比較した．増量先の用量が除外されている場合は増量せず，最低用量で減量が必要となった場合には試験を中止した．

また，下記に単調回帰のアルゴリズムも示す．投与例数が 1 例以上の各用量について，DLT 発現率に Beta(0.05, 0.05)を事前分布として設定し，その事後平均を DLT 発現率の推定値とした．さらに，事後分布の分散の逆数を重みとして weighted PAVA（weighted pool-adjacent-violators algorithm）を適用した．隣接する用量間で毒性発現率の単調性が満たされない場合には，隣接する両ブロックを併合し，重み付き平均により推定値を更新する処理を繰り返した．得られた単調推定値のうち，安全性基準により除外されていない用量を MTD 選択の対象とした．

```
/* Pool Adjacent Violators Algorithm*/
nblock = 0;
do j = 1 to &ndose;
        if n_dose[j] > 0 and elimi[j] = 0 then do;
                nblock = nblock + 1;
                lo[nblock] = j;
                hi[nblock] = j;
                /* 補正推定値・分散・重み*/
                p_dlt_blk[nblock] = (n_dlt[j] + 0.05) / (n_dose[j] + 0.10);
                var_p = ((n_dlt[j] + 0.05) * (n_dose[j] - n_dlt[j] + 0.05))
                        / ((n_dose[j] + 0.10)**2 * (n_dose[j] + 1.10));
                wt_pava[nblock] = 1 / var_p;
        end;
end;

/* Weighted PAVA */
changed = 1;
do while(changed = 1 and nblock > 1);
        changed = 0;
        do j = 1 to nblock-1;
                if p_dlt_blk[j] > p_dlt_blk[j+1] then do;
                        wt_pool = wt_pava[j] + wt_pava[j+1];
                        p_pool = (wt_pava[j]*p_dlt_blk[j] + wt_pava[j+1]*p_dlt_blk[j+1]) / wt_pool;
                        hi[j] = hi[j+1];
                        p_dlt_blk[j] = p_pool;
                        wt_pava[j] = wt_pool;
                        do k = j+1 to nblock-1;
                                lo[k] = lo[k+1];
                                hi[k] = hi[k+1];
                                p_dlt_blk[k] = p_dlt_blk[k+1];
                                wt_pava[k] = wt_pava[k+1];
                        end;
                        nblock = nblock - 1;
                        changed = 1;
                        leave;
                end;
        end;
end
```

```
/* MTD selection */
        mtd = .;
        min_diff = constant("BIG");

        do j = 1 to &ndose;
            if n_dose[j] > 0 and elimi[j] = 0 then do;
                diff = abs(p_iso[j] - &target);
                if diff < min_diff then min_diff = diff;
            end;
        end;

        low_dose = .;
        high_dose = .;

            do j = 1 to &ndose;
                if n_dose[j] > 0 and elimi[j] = 0 then do;
                    diff = abs(p_iso[j] - &target);
                    if abs(diff - min_diff) < 1E-12 then do;
                        if p_iso[j] < &target then do;
                            if missing(low_dose) or j > low_dose then low_dose = j;
                        end;
                        else do;
                        if missing(high_dose) or j < high_dose then high_dose = j;
                        end;
                    end;
                end;
            end;
            if not missing(high_dose) then mtd = high_dose;
            else if not missing(low_dose) then mtd = low_dose;
```

MTD は PAVA（Pool Adjacent Violators Algorithm）により得られた単調 DLT 推定値と目標 DLT 発現率の差の絶対値が最小となる用量として選択した．差の絶対値が同一の場合，単調 DLT 推定値が目標値未満では高用量側，目標値より大きな場合は低用量側を選択する実装とした．

4.3 BOIN12

Model-assisted 型の用量探索デザインであり，DLT 発現率と有効性を統合した utility に基づき増量・維持・減量を決定し，isotonic regression により DLT 発現率を推定し，MTD 以下で utility が最大となる用量を OBD として選択する．

```
/**********************/
/* BOIN 12    */
/**********************/
dlt = 0;
efficacy = 0;
te = 0;
        do i = 1 to &cohort;
                z1 = rand("NORMAL");
                z2 = &rho*z1 + sqrt(1-&rho**2)*rand("NORMAL");    /*&rho=0.25*/
                dlt_i = (z1 <= quantile("NORMAL",tox[level]));
                eff_i = (z2 <= quantile("NORMAL",eff[level]));
                dlt = dlt + dlt_i;
                efficacy = efficacy + eff_i;
                if dlt_i = 1 and eff_i = 1 then
                        te = te + 1;
        end;
```

各コホートの被験者について，設定した真の DLT 発現率及び有効性割合に基づいて DLT・有効性を生成し，コホート内の DLT 数，有効性発現数及び DLT と有効性が同時に発現した例数を集計する．DLT 及び有効性は相関を有する潜在正規変数を用いて生成し，潜在変数間の相関係数を ρ=0.25 とした．

```
        do j = 1 to &ndose;
                if n_dose[j] > 0 then do;
                        prob_tox = 1-cdf("BETA",&phi_t,n_dlt[j]+1,n_dose[j]-n_dlt[j]+1);
                        if prob_tox >= &cutoff_t then do;      /*&cutoff_t=0.95*/
                                do k = j to &ndose;
                                        elimi[k] = 1;
                                end;
                        end;

                        prob_futile = cdf("BETA",&phi_e,n_eff[j]+1,n_dose[j]-n_eff[j]+1);
                        if prob_futile >= &cutoff_e then      /*&cutoff_e=0.90*/
                                elimi[j] = 1;
                        end;
        end;
```

各用量について，DLT 発現率が毒性の許容上限を超える事後確率および有効割合が許容下限を下回る事後確率を算出した．DLT 発現率が許容上限を超える事後確率が 0.95 以上となった場合には，当該用量及びそれ以上の用量を除外し，有効割合が許容下限を下回る事後確率が 0.90 以上となった場合には，当該用量を除外した．除外後，BOIN12 の用量移行ルールに基づき，次コホートで投与する用量を決定した．

000313

```
/* Case B (steps 2b)*/
else if p_hat > &lambda_e    /*&lambda_e=0.276*/
        and p_hat <= &lambda_d    /*&lambda_d=0.418*/
        and n_dose[level] >= &nstar then do;      /*&nstar=6*/

        best_dose = .;
        best_desir = -1;

        do k = max(1,level-1) to level;
                if elimi[k] = 0 then do;
                        n_u1 = n_eff[k] - n_te[k];
                        n_u2 = n_dose[k] - n_dlt[k] - n_eff[k] + n_te[k];
                        n_u3 = n_te[k];
                        n_u4 = n_dlt[k] - n_te[k];
                /*&u1=100, &u2=40, &u3=60, &u4=0*/
                        x = (&u1*n_u1+&u2*n_u2+&u3*n_u3+&u4*n_u4)/100;
                        alpha_u = 1 + x;
                        beta_u = 1 + n_dose[k] - x;
                /*&phi_t=0.35, &phi_e=0.25*/
                        ubar_benchmark
                           = &u1*(1-&phi_t)*&phi_e+&u2*(1-&phi_t)*(1-
                              &phi_e)+&u3*&phi_t*&phi_e+&u4*&phi_t*(1-&phi_e);
                           ub = ubar_benchmark+(100-ubar_benchmark)/2;

                        p_desir = 1-cdf("BETA",ub/100,alpha_u,beta_u);

                        if p_desir > best_desir then do;
                                best_desir = p_desir;
                                best_dose = k;
                        end;
                end;
end;
```

現在用量の観察 DLT 発現割合 p_hat が増量境界 lambda_e より大きく減量境界 lambda_d 以下であり，かつ現在用量の投与例数が 6 例以上の場合には，BOIN12 による steps 2b を適用した．この場合，現在用量及び 1 段階低い用量を候補とし，安全性または有効性の基準により除外された用量は候補から除外した．各候補用量について，被験者を DLT 及び有効性の発現状況に基づく 4 つのアウトカムカテゴリーに分類し，それぞれに設定した効用値を用いて観察効用を算出した．本シミュレーションでは，「DLT なし・有効性あり」，「DLT なし・有効性なし」，「DLT あり・有効性あり」，「DLT あり・有効性なし」に対して，それぞれ 100,

40，60，0 の効用値を設定した．これらの観察効用に基づいて効用の事後分布を設定し，各候補用量について benchmark 効用を超える事後確率を望ましさとして算出した．次コホートの投与用量は，現在用量および 1 段階低い用量を候補として，望ましさが高い用量を選択することとした．原著 BOIN12 では，候補用量の望ましさを Rank-based Desirability Score（RDS）に変換し，その順位に基づいて用量移行を決定する．一方，本シミュレーションでは，候補用量間の望ましさを直接比較することを目的として，RDS への変換は行わず，各候補用量の benchmark 効用を超える事後確率を直接比較し，その値が最大となる用量を次コホートの投与用量として選択した．

```
/* OBD */

obd = .;
best_ubar = -1;

        do j = 1 to &ndose;
                if n_dose[j] > 0
                        and elimi[j] = 0
                        and j <= mtd then do;

                        x = (&u1*(n_eff[j]-n_te[j])
                            +&u2*(n_dose[j]-n_dlt[j]-n_eff[j]+n_te[j])
                            +&u3*n_te[j]+&u4*(n_dlt[j]-n_te[j]))/100;

                                ubar = (x+1)/(n_dose[j]+2);
                                if ubar > best_ubar then do;
                                        best_ubar = ubar;
                                        obd = j;
                                end;

                end;

        end;
```

最終的な OBD は，投与例数が 1 例以上で，安全性または有効性の基準により除外されておらず，かつ MTD 以下の用量を候補とした．各候補用量について，設定した効用の事後分布の事後平均を平均効用とし，平均効用が最大となる用量を OBD として選択した．

000315

5　SAS によるシミュレーション結果

シミュレーションの結果を下記の[表 5-1], [表 5-1], [表 5-3]に示す.

[表 5-1：最終選択用量と Target dose（Dose 4）との差（ / は MTD 算出不可）]

Scenario	デザイン	/	-3	-2	-1	0	+1	+2
Scenario 1	3+3	2.95 %	8.95 %	25.39 %	31.83 %	21.18 %	8.18 %	1.52 %
	BOIN	0 %	1.15 %	4.42 %	29.67 %	40.87 %	19.89 %	4.00 %
	BOIN12	0 %	3.79 %	12.38 %	32.08 %	32.24 %	15.75 %	3.76 %
Scenario 2	3+3	2.85 %	8.90 %	25.55 %	31.86 %	25.73 %	4.94 %	0.17 %
	BOIN	0 %	1.15 %	4.43 %	30.31 %	52.23 %	11.60 %	0.28 %
	BOIN12	0 %	3.96 %	12.85 %	34.82 %	35.78 %	12.07 %	0.52 %
Scenario 3	3+3	2.87 %	18.00 %	31.42 %	23.97 %	14.48 %	6.55 %	2.71 %
	BOIN	0 %	1.95 %	13.18 %	32.54 %	27.44 %	16.63 %	8.26 %
	BOIN12	0 %	4.47 %	21.02 %	31.03 %	20.82 %	15.30 %	7.36 %
Scenario 4*	3+3	2.80 %	9.01 %	34.53 %	27.21 %	21.92 %	4.38 %	0.15 %
	BOIN	0 %	1.15 %	9.50 %	35.72 %	43.37 %	10.00 %	0.26 %
	BOIN12	0 %	4.39 %	13.24 %	42.41 %	28.84 %	10.78 %	0.34 %
Scenario 5*	3+3	2.80 %	9.01 %	34.53 %	27.21 %	21.92 %	4.38 %	0.15 %
	BOIN	0 %	1.15 %	9.50 %	35.72 %	43.37 %	10.00 %	0.26 %
	BOIN12	0 %	4.24 %	12.88 %	41.82 %	29.99 %	10.74 %	0.33 %

[表 5-2：各種評価項目]

Scenario	デザイン	Correct selection	Under-dose	Over-dose	Over-dose allocation
Scenario 1	3+3	21.18 %	66.17 %	9.7 %	8.32 %
	BOIN	40.87 %	35.24 %	23.89 %	16.23 %
	BOIN12	32.24 %	48.25 %	19.51 %	17.94 %
Scenario 2	3+3	25.73 %	66.31 %	5.11 %	7.33 %
	BOIN	52.23 %	35.89 %	11.88 %	12.50 %
	BOIN12	35.78 %	51.63 %	12.59 %	13.25 %
Scenario 3	3+3	14.48 %	73.39 %	9.26 %	6.68 %
	BOIN	27.44 %	47.67 %	24.89 %	14.51 %
	BOIN12	20.82 %	56.52 %	22.66 %	6.63 %
Scenario 4*	3+3	21.92 %	70.75 %	4.53 %	6.32 %
	BOIN	43.37 %	46.37 %	10.26 %	10.68 %
	BOIN12	28.84 %	60.04 %	11.12 %	11.53 %
Scenario 5*	3+3	21.92 %	70.75 %	4.53 %	6.32 %
	BOIN	43.37 %	46.37 %	10.26 %	10.68 %
	BOIN12	29.99 %	58.94 %	11.07 %	9.97 %

[表 5-3: BOIN12 OBD 選択割合 / OBD MTD 一致率（ / は OBD 算出不可）]

Scenario	/	1	2	3	4	5	6	MTD = OBD
Scenario 1	0.84 %	19.28 %	29.81 %	26.01 %	15.98 %	6.74 %	1.34 %	38.57 %
Scenario 2	0.79 %	20.19 %	31.53 %	27.35 %	16.60 %	3.42 %	0.12 %	38.35 %
Scenario 3	1.67 %	22.54 %	30.20 %	20.21 %	13.31 %	8.94 %	3.13 %	42.70 %
Scenario 4*	0.87 %	15.24 %	32.02 %	29.74 %	17.24 %	4.84 %	0.05 %	48.01 %
Scenario 5*	0.81 %	14.44 %	30.65 %	28.48 %	23.19 %	2.41 %	0.02 %	49.14 %

5.1 3+3

BOIN 及び BOIN12 と比較して Target dose 選択割合が低く，また，Target dose より高い用量への割付割合も低かった．3+3 では，あらかじめ定められた DLT 発現数に基づいて用量移行を判断するため，DLT が一定数観察された場合には高用量への移行が抑制される一方，Target dose より低い用量に留まる場合が多くなり，Target dose 選択割合が BOIN 及び BOIN12 より低くなったと考えられる．

5.2 BOIN

3+3 及び BOIN12 と比較して Target dose 選択割合が高かった．これは，BOIN が事前に設定した目標 DLT 発現率に基づく用量増量・減量境界を用いて用量移行を判断するため，観察された DLT 発現率に応じて MTD 付近の用量を探索しやすいことによるものと考えられる．一方，BOIN は主として DLT 発現率に基づいて用量移行及び MTD の選択を行うため，有効性の用量反応関係を直接考慮した用量選択には限界がある．

5.3 BOIN12

いずれの Scenario においても，Target dose より 2 用量低い用量が OBD として選択される割合が最も高かった．一方 MTD に関して，Scenario 1 及び 2 では Target dose 選択割合が最も高く，Scenario 3～5 では Target dose より 1 用量低い選択割合が最も高かった．また，有効性の plateau の程度を変化させた Scenario 4 及び 5 では各用量の選択割合，MTD と OBD の一致割合はいずれも概ね同様の傾向を示した．

6 まとめと考察

3+3 では，高用量への移行が抑制される一方，Target dose 付近の用量を選択する割合は BOIN 及び BOIN12 より低かった．BOIN では，目標 DLT 発現率に基づく明確な用量移行規則により，観察された DLT 発現率に応じて MTD 付近を探索しやすく，3+3 と比較して毒性に基づく用量探索において異なる特性が示された．一方，BOIN12 では，毒性に加えて有効性を考慮して OBD を選択するため，MTD とは異なる用量が OBD として選択される場合が認められた．なお BOIN12 において，本シミュレーションでは遅発性毒性を考慮しておらず，観察された DLT を各用量での用量移行判断に利用した．遅発性毒性が用量探索及び用量選択に及ぼす影響については，今後の検討が必要である．

000317

7　参考文献

[1]　Ruitao Lin, Yanhong Zhou, Fangrong Yan, Daniel Li, Ying Yuan

BOIN12:BayesianOptimal Interval Phase I/II Trial Design for Utility-Based Dose Finding in Immunotherapy and Targeted Therapies

https://pmc.ncbi.nlm.nih.gov/articles/PMC7713525/pdf/PO.20.00257.pdf

（最終アクセス 2026/09/18）

[2]　Fangrong Yan, Liangcai Zhang, Yanhong Zhou, Haitao Pan, Suyu Liu, Ying Yuan

BOIN: An R Package for Designing Single-Agent and Drug-Combination Dose-Finding Trials Using Bayesian Optimal Interval Designs

https://www.jstatsoft.org/article/view/v094i13

（最終アクセス 2026/09/18）

[3]　R/boin12_rds.R

escalation source: R/boin12_rds.R

（最終アクセス 2026/09/18）

[4]　橋詰公一　武田健太朗　佐藤宏征　平川晃弘　寒水孝司

がん第 1 相用量探索試験におけるモデル支援型デザインの最近の展開

https://www.jstage.jst.go.jp/article/jjb/43/1/43_3/_pdf/-char/ja

（最終アクセス 2026/09/18）

[5]　日本製薬工業協会　医薬品評価委員会　データサイエンス部会　2022　年度　タスクフォース 3 Ver 1.0 2023 年 6 月

近年のがん第 I 相試験デザインとその選択

https://www.jpma.or.jp/information/evaluation/results/allotment/g75una0000002m1p-att/DS_202306_oncoP1DE.pdf

（最終アクセス 2026/09/18）

報告するまでがEstimand

～ SAS を用いた実装検討 2026 ～

○森田　祐介，太田　裕二，内藤　寿輝

（ノーベルファーマ株式会社　データサイエンス部）

Well-Specified Estimands Still Require Appropriate Analysis and Reporting

Yusuke Morita，Yuji Ohta，Toshiki Naito

Data Science Department, Nobelpharma Co., Ltd.

要旨

治験実施計画書に規定された Estimand は，試験デザイン，データ収集，統計解析及び報告を通じて一貫して実装される必要がある．しかし，Estimand の設計に関する議論に比べ，データ収集，統計解析及び報告への具体的な反映方法を示した事例は限られている．そこで本検討では，PhUSE ホワイトペーパーを参考に，北海道大学病院が公開する CDISC 標準の臨床試験サンプルデータセットを用いて，中間事象の特定から ADaM への反映, 統計解析及び解析結果の提示までを SAS で実装した. 中間事象を解析用データセット ADICE に集約し，有効性評価レコードとの関係及び解析ごとの取扱いを有効性解析用データセット ADEFF に反映することで，SDTM から解析帳票に至るトレーサビリティを確保する一つの実装例を示した．計画された Estimand に基づく解析結果を, 関心のある臨床的疑問への回答として適切に報告するためには, データ標準, 統計解析及び報告を一貫して結び付ける実装が重要であり，その実現において SAS ユーザーの貢献が期待される．

キーワード：ICH E9(R1), Estimand，中間事象 (Intercurrent Event), SDTM, ADaM, ADICE, SAS, Traceability

1. 背景と目的

2019 年 11 月に ICH E9(R1)「臨床試験における Estimand と感度分析」が国際的に合意され[1]，2024 年 6 月には本邦においても「臨床試験のための統計的原則」の補遺として通知された[2]．ICH E9(R1)では，推定対象となる治療効果を明確にする枠組みとして Estimand が導入され，試験目的，試験デザイン，データ収集，統計解析及び結果の解釈を相互に整合させるフレームワークが提案された．

ICH E9(R1)の公表以降，Estimand の概念，設計方法，適用事例・提案に関する文献は蓄積されつつある[3-12]．また，Estimand を規定した試験が完了し，統計解析及び報告の段階に移行するとともに，国内外の規制当局による審査報告書においても，Estimand や中間事象の取扱い，それらに対応する解析及び報告が具体的に議論される事例がみられるようになった[13-16]．すなわち，実務上の関心は，Estimand を「どのように設定するか」だけでなく，規定した Estimand を「どのように解析し，報告するか」にも広がっている[17]．

当社においても，Estimand を規定した臨床試験が報告段階に移行するなかで，その実装に関する課題に直面した[18]．具体的には，実際に試験で発生した中間事象の特定，その影響を受けるデータの識別，各 Estimand に対応する解析用データセットの作成，並びに中間事象の発現状況及び解析結果の提示方法を検討する必要があった．しかし，Estimand を解析用データセット，解析プログラム及び解析帳票に反映し，治験総括報告書までのトレーサビリティを確保するための具体的な実装方法を示した公表文献は限られており[19]，その対応に苦慮した．

そこで本稿では，PhUSE のホワイトペーパー「Implementation of ICH E9(R1) Estimands Framework using Data Standards」を参考に[20]，北海道大学病院 医療・ヘルスサイエンス研究開発機構 データサイエンスセンターが公開する CDISC 標準の臨床試験サンプルデータセット[21]を用いて，Estimand の統計解析及び報告への実装方法を検討した．具体的には，中間事象の特定，中間事象と主要評価データとの関係の ADaM への反映，中間事象の発現状況の要約，並びに主解析，感度分析及び補足的解析の結果提示について，SAS を用いた一連の実装例を示す．

2. Estimand の報告とデータ標準

2.1. ICH E9(R1)ガイドラインにおける報告上の要請

ICH E9(R1)では，主とする解析，感度分析及び補足的解析の結果を，各解析が事前規定，盲検性が維持されている間の追加又は事後的な解析のいずれであるかを明示した上で，治験総括報告書において体系的に報告することが求められている．また，感度分析により主とする推定量に基づく推測の安定性を評価するとともに，各治療群における中間事象の発現数及び発現時期を要約することが示されている．さらに，試験計画時に予見されなかった中間事象が明らかになった場合には，Estimand 及び試験結果の解釈への影響を考察し，Estimand を変更する場合は通常，治験実施計画書の改訂に反映すべきとされている[2]．

2.2. PhUSE ホワイトペーパーによるデータ標準への実装提案

ICH E9(R1)は，Estimand を実装するための具体的なデータ構造及び方法までは示していない．このため，PhUSE のホワイトペーパー「Implementation of ICH E9(R1) Estimands Framework using Data Standards」では，データ収集，SDTM への格納，ADaM への反映及び関連文書への記載について，CDISC 標準を用いた実装方法が提案されている[20]．データ収集については，中間事象の理由及び発現時期を適切に収集することの重要性が示され，CRF 及び SDTM へのマッピング例が提示されている．既存の CRF で必要な中間事象を十分に収集できない場合には，試験に応じた CRF を作成することも提案されている．また，中間事象の収集方法及び SDTM ドメイン・変数へのマッピングを cSDRG に記載することが推奨されている．
ADaM については，複数のデータソースから中間事象に関する情報を集約する解析用データセットとして ADICE を作成し，中間事象の種類，発現時期，情報源及び適用するストラテジー等を記録するとともに，他の解析用データセットにおいて，中間事象の影響を受けるデータポイントと ADICE との関係を示す方法が提案されている．さらに，Define-XML に ADICE 及び関連変数のメタデータやトレーサビリティに関する情報を記載し，ADRG では Estimand と推定量（Estimator），解析用データセット，解析対象レコード及び解析方法等の関係を説明することが提案されている．

3. 使用データ及び解析設定

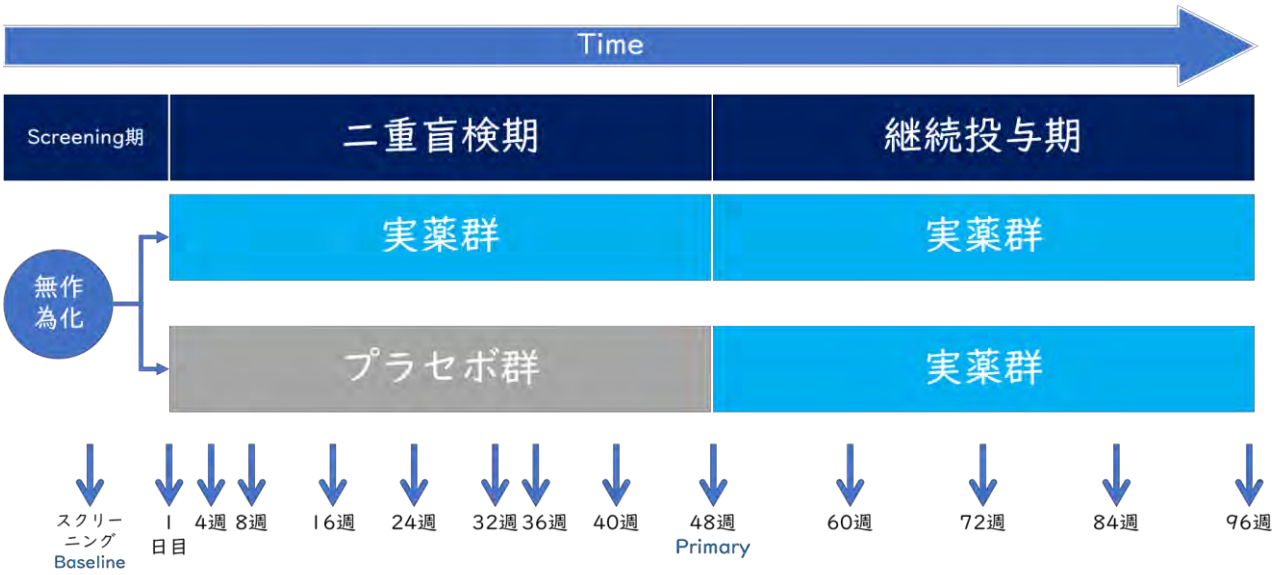
3.1. 使用するデータ

本稿では，北海道大学病院 医療・ヘルスサイエンス研究開発機構 データサイエンスセンターが公開する CDISC 標準の臨床試験サンプルデータセットを使用した．本データは，医薬品等規制調和・評価研究事業「アカデミアにおける CDISC 標準利用促進に関する研究開発」（課題管理番号：24mk0101242）により作成されたものである[21].

3.2. 試験デザイン

対象とした試験は，ウルソデオキシコール酸錠 100 mg を評価する第 III 相試験であり，二重盲検期及び継続投与期から構成された．試験デザインの概要を下表及び下図に示す．なお，本稿では，「Week 48 における ALT のベースラインからの変化率」を主要評価項目と設定して，Estimand の解析及び報告への実装方法を検討した．

項目	二重盲検期（48 週間）	継続投与期（48 週間）
対象被験者数	186 例	164 例
投与群	実薬群：93 例 プラセボ群：93 例	実薬群：164 例
投与方法	二重盲検，無作為割付	実薬投与
主な有効性評価項目	48 週時の ALT 及び AST のベースラインからの変化率，正常化割合等	―



3.3. 中間事象

公開データに基づいて本稿で設定した中間事象を下表に示す．有害事象による割付治療中止後にも主要評価項目が評価された事例を設定するため，死亡に至った事象を除く重篤な有害事象を発現した被験者については，当該有害事象の発現日を割付治療の中止日とみなし，発現日の翌日以降は割付治療が投与されなかっ

たものとして取り扱った．なお，これは Estimand に関連する実装方法を例示するための仮想的な設定であり，原データにおける実際の治療経過を示すものではない．また，同意撤回による治験の中止は，中間事象とせず，欠測の問題として取り扱った．

本稿で設定した中間事象	関連する SDTM ドメイン
有害事象による割付治療中止	EX，AE，DS
被験者の死亡	DS，AE

3.4. 主要 Estimand 及び解析計画

本稿で対象とする主要 Estimand を下表に示す．

治療	被験薬 vs プラセボ	
対象集団	選択・除外基準にて規定された患者集団	
変数	48 週時の ALT のベースラインからの変化率	
中間事象とその取扱い	有害事象による割付治療中止	仮想ストラテジー：割付治療中止日の翌日以降のデータは欠測値として取り扱う
	被験者の死亡	仮想ストラテジー：死亡後の予定評価時点には評価値を設定せず，解析上は欠測と同様に取り扱う*
要約指標	変数の平均値の群間差	

*：ICH E9(R1)ガイドラインでは，被験者の死亡後には測定値自体が存在せず，欠測データとは区別されるが，本稿では解析実装上の便宜から欠測と同様に取り扱う．

主解析には Mixed Model for Repeated Measures（MMRM）を用いた．目的変数を ALT のベースラインからの変化率とし，説明変数として割付群，評価時点（カテゴリカル），割付群と評価時点の交互作用項及び ALT のベースライン値をモデルに含めた．被験者内誤差の分散共分散構造には無構造（unstructured：UN）を用い，固定効果の推定量の分散共分散行列の補正及び検定の分母自由度の近似には Kenward–Roger 法を用いた．また，感度分析として，主解析の欠測メカニズムの仮定（Missing at Random）に対する推測の安定性を検討するため，欠測値に対して Jump to Reference 法による多重補完を実施した[22]．さらに，補足的 Estimand（Estimand 02）は，主要 Estimand（Estimand 01）のうち，有害事象による割付治療中止に対するストラテジーのみを治療方針ストラテジーに変更し，その他の構成要素は主要 Estimand と同一とした．この補足的 Estimand に対する補足的解析では，割付治療中止後に収集された評価値を含めて主解析と同様の解析を実施した．

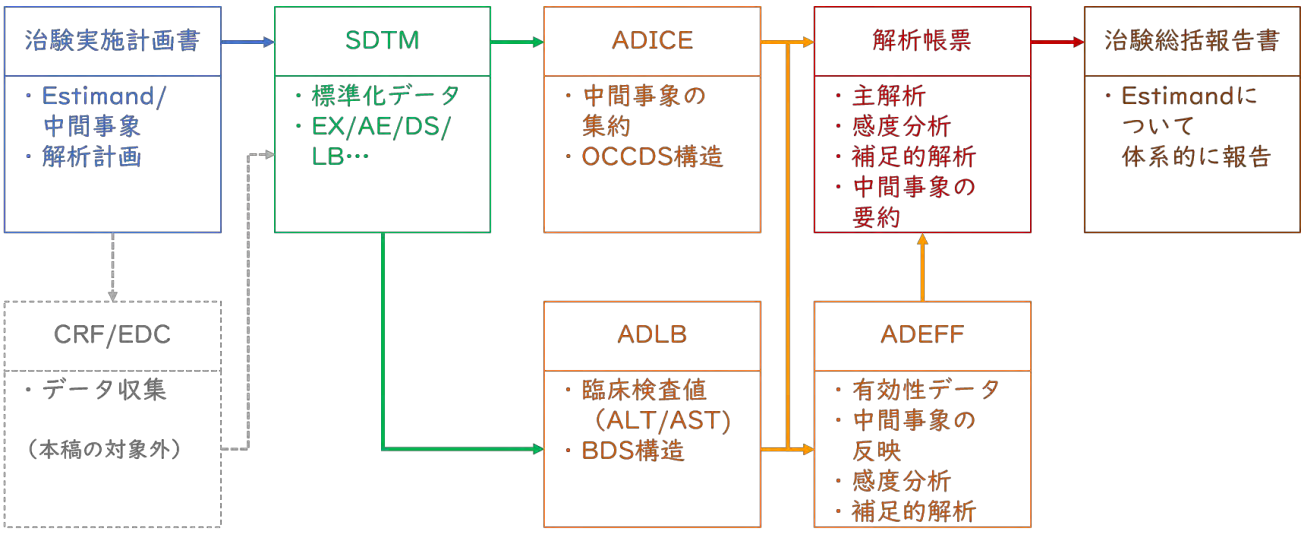
なお，本稿における各種設定は，Estimand に関連する SAS での実装方法を検討することを主目的としたものであり，被験薬の臨床的意義のある治療効果を推定することを意図したものではない．

4. Estimand の実装方法

4.1. 実装の全体像

第 3 章に示した中間事象，Estimand 及び解析計画に基づき，SDTM から中間事象に関する情報を抽出して ADICE を作成するとともに，ADaM の ADLB を基に ADEFF を作成し，ADICE に格納した中間事象との関係及び各解析に用いる解析対象フラグを ADEFF に設定した．ADICE を用いて中間事象の発現状況を要約す

るとともに，ADEFF に設定した解析対象フラグに基づいて主解析，感度分析及び補足的解析を実施し，解析帳票を作成した．SDTM，ADICE，ADLB，ADEFF 及び解析帳票の関係を下図に示す．



4.2. 解析用データセット：ADICE の作成

第 3.3 節に示した判定基準に基づき，SDTM データセットから有害事象による割付治療中止に関する情報及び死亡に関する情報を抽出し，被験者ごとの中間事象を ADICE に集約した．ADICE には ADaM の Occurrence Data Structure（OCCDS）を参考に，1 件の中間事象につき 1 レコードとし，被験者識別コード，中間事象の識別番号，種類，発現日，情報源となる SDTM ドメイン及び元レコードの識別情報を格納した．また，第 3.4 節に示した Estimand との対応を明確にするため，中間事象ごとに適用するストラテジーを格納した．

4.3. 解析用データセット：ADEFF への中間事象の反映

ADEFF は，ADaM の ADLB データセットに格納された AST 及び ALT データを基に作成した．ADEFF は ADaM の Basic Data Structure（BDS）に準じた構造とし，各評価レコードについて，評価日と ADICE に格納された中間事象の発現日を比較し，評価値が中間事象の発現日以前又は発現後のいずれに測定されたものであるかを判定した．有害事象による割付治療中止日の翌日以降の評価レコードには，対応する ADICE の中間事象識別番号を設定し，評価値と中間事象との関係を追跡可能とした．一方，死亡後の予定評価時点に対応する評価レコードは ADEFF に作成しなかった．

また，各評価レコードについて，Estimand ごとの中間事象の影響及び各解析に使用するか否かを示す変数を設定した．主要 Estimand では仮想ストラテジーを適用するため，有害事象による割付治療中止日の翌日以降の評価値を主解析及び感度分析の解析対象から除外した．一方，有害事象による割付治療中止に治療方針ストラテジーを適用する補足的解析では，割付治療中止後に評価された評価値を解析対象に含めた．

4.4. 統計解析及び解析帳票の作成

4.4.1. 中間事象の発現状況の要約

各期間における中間事象の発現状況について，中間事象の有無及び種類別の被験者数と割合を二重盲検期の割付群別に要約した．割合の分母には，各期間に参加した各割付群の被験者数を用いた．

また，中間事象の発現状況を経時的に把握するため，各評価時点における中間事象の種類別の累積発現割合を算出し，割付群別の積み上げ棒グラフとして示した．累積のため，中間事象を発現した被験者は，発現時点以降，当該中間事象のカテゴリーに分類した．中間事象を発現していない被験者は「No IE」，同意撤回等により以降の追跡が行われなかった被験者は「Lost/Withdrawal」に分類した．各カテゴリーの割合は，各割付群の解析対象被験者数を分母として算出した．

さらに，中間事象の発現時期を連続的な時間軸上で把握するため，治療開始から最初の中間事象発現までの期間について，種類別の累積発現率を累積発生関数により推定し，積み上げ面グラフとして示した．種類別の推定では，対象とする中間事象以外の中間事象を競合事象として扱い，同意撤回はその時点で打ち切りとして扱った．例えば，有害事象による割付治療中止の累積発現率を推定する際には，被験者の死亡を競合事象として扱った．

4.4.2. 主解析，感度分析及び補足的解析

主解析では，主要 Estimand に対応する解析対象フラグを用いて，割付治療中止日の翌日以降に収集された評価値を解析の対象外とし，解析上欠測として取り扱った．これにより，実際に観測された評価値をデータセット内に保持しつつ，仮想ストラテジーに対応する解析データを作成した．

感度分析では，主解析と同じ Estimand を対象として，主解析用データに Jump to Reference 法による多重補完を適用した．プラセボ群の欠測値は自群の条件付き分布に基づいて補完し，被験薬群の欠測値はプラセボ群の条件付き分布を参照して補完した．補完回数は 100 回とし，補完データセットごとに第 3.4 節に規定した解析を実施し，得られた推定値及び標準誤差を Rubin のルールに基づいて統合した．

補足的解析では，補足的 Estimand に対応する解析対象フラグを用い，有害事象による割付治療中止の発現にかかわらず，ベースライン後の二重盲検期に収集されたすべての評価値を解析対象とした．

5. Estimand の実装結果

5.1. ADICE の実装結果

中間事象を集約した ADICE の出力イメージ（抜粋）を下表に示す．なお，本稿では複数の中間事象を発現した被験者はいなかったため，被験者内の中間事象の発現順序を示す ASEQ は全レコードで 1 となった．

USUBJID	ASEQ	ACAT1	ACAT1N	ATERM	ASTDT	SRC DOM	SRC SEQ	EST01 STP	EST02 STP
01-10	1	Treatment Discontinuation due to AE	1	Cholangitis	2021-04-13	AE	2	HYPOTHETICAL	TREATMENT POLICY
01-14	1	Treatment Discontinuation due to AE	1	Urticaria	2021-04-30	AE	6	HYPOTHETICAL	TREATMENT POLICY

USUBJID	ASEQ	ACAT1	ACAT1N	ATERM	ASTDT	SRC DOM	SRC SEQ	EST01 STP	EST02 STP
01-15	1	Death	2	DEATH	2021-08-11	DS	4	HYPOTHETICAL	HYPOTHETICAL
01-18	1	Treatment Discontinuation due to AE	1	Constipation	2021-10-31	AE	1	HYPOTHETICAL	TREATMENT POLICY
01-27	1	Death	2	DEATH	2022-02-11	DS	5	HYPOTHETICAL	HYPOTHETICAL
01-32	1	Treatment Discontinuation due to AE	1	Urticaria	2022-06-06	AE	4	HYPOTHETICAL	TREATMENT POLICY

ADICE の各変数の説明を下表に示す．なお，紙面の都合上，主要な変数のみを抽出して提示した．

変数名	ラベル	説明
USUBJID	Unique Subject Identifier	被験者識別子（ここでは試験番号を割愛して表示）
ASEQ	Analysis Sequence Number	各被験者内の中間事象の連番（発現順）
ACAT1	Analysis Category 1	中間事象の分類
ACAT1N	Analysis Category 1 (N)	中間事象の分類コード
ATERM	Analysis Term	中間事象を記述する解析用語
ASTDT	Analysis Start Date	中間事象の発現（開始）日
SRCDOM	Source Data	中間事象の元ドメイン
SRCSEQ	Source Sequence Number	中間事象の元ドメインにおける連番（XXSEQ）
EST01STP	Estimand 01 Planned Handling Strategy	Estimand 01（主要 Estimand）に関連する中間事象の影響を受けるデータポイントの取扱いについて，事前に計画されたストラテジー名
EST02STP	Estimand 02 Planned Handling Strategy	Estimand 02（補足的 Estimand）に関連する中間事象の影響を受けるデータポイントの取扱いについて，事前に計画されたストラテジー名

5.2. ADEFF の実装結果

有効性解析に用いる ADEFF の出力イメージ（抜粋）を下表に示す．ADICE に基づく中間事象に関する情報を示す変数，レコード採否を示す変数をそれぞれ追加した．

USUBJID	TRTP	AVISIT	AVAL	ICC1S	ICC2S	ICC1E01F	ICE01S	ANL01FL	ICC1E02F	ICE02S	ANL02FL
01-14	Placebo	Baseline	0.00	.	.		.			.	
01-14	Placebo	Day 1	-3.26	.	.		.	Y		.	Y
01-14	Placebo	Week 4	-2.72	.	.		.	Y		.	Y
01-14	Placebo	Week 8	-7.61	.	.		.	Y		.	Y
01-14	Placebo	Week 16	-9.78	.	.		.	Y		.	Y
01-14	Placebo	Week 24	-13.04	.	.		.	Y		.	Y
01-14	Placebo	Week 32	-14.67	.	.		.	Y		.	Y
01-14	Placebo	Week 36	-14.13	1	.	Y	1			.	Y
01-14	Placebo	Week 40	-17.39	1	.	Y	1			.	Y
01-14	Placebo	Week 48	-18.48	1	.	Y	1			.	Y

ADEFF の各変数の説明を下表に示す．なお，紙面の都合上，主要な変数のみを抽出して提示した．

変数名	ラベル	説明
USUBJID	Unique Subject Identifier	被験者識別子（ここでは試験番号を割愛して表示）
TRTP	Planned Treatment	割り付けられた投与群
AVISIT	Analysis Visit	解析用 VISIT
AVAL	Analysis Value	解析に用いる値．本稿では ALT のベースラインからの変化率
ICC1S	Impacting ICE Cat. Val. 1 Seq. Num	評価レコードに影響し得る中間事象分類 1（有害事象による割付治療中止）に対応する ADICE の ASEQ
ICC2S	Impacting ICE Cat. Val. 2 Seq. Num	評価レコードに影響し得る中間事象分類 2（死亡）に対応する ADICE の ASEQ（ただし，本実装では変数体系を例示するため ICC2S も設定したが，死亡後レコードを作成しないため，全レコードで欠測となる.）
ICC1E01F	Imp. ICE Cat. Val. 1 for Est. 01 Flag	中間事象分類 1 が Estimand 01（主要 Estimand）における当該評価レコードの取扱いに影響する場合に「Y」
ICE01S	Impacting ICE for Est. 01 Seq. Num	優先順位を考慮した結果，Estimand 01 における当該評価レコードの取扱いを決定する中間事象に対応する ADICE の ASEQ
ANL01FL	Analysis Flag 01	主解析及び感度分析で評価値として採用するレコードに「Y」
ICC1E02F	Imp. ICE Cat. Val. 1 for Est. 02 Flag	中間事象分類 1 が Estimand 02（補足的 Estimand）における当該評価レコードの取扱いに影響する場合に「Y」
ICE02S	Impacting ICE for Est. 02 Seq. Num	優先順位を考慮した結果，Estimand 02 における当該評価レコードの取扱いを決定する中間事象に対応する ADICE の ASEQ
ANL02FL	Analysis Flag 02	補足的解析に採用するレコードに「Y」

本実装では，PhUSE のホワイトペーパーの変数体系を参考に，評価レコードと中間事象との関係を示す ICCyS，Estimand ごとの中間事象の実際の影響を示す ICCyEzzF，優先順位を考慮して最終的に取扱いを決定する中間事象を示す ICEzzS，及び解析ごとのレコード採否を示す ANLxxFL を設定した．これらの変数名において，y は中間事象の分類，zz は Estimand を識別する番号を表す．本実装では，ADICE の ACAT1N に対応して，中間事象分類 1 を「有害事象による割付治療中止」，分類 2 を「死亡」とし，主要 Estimand を 01，補足的 Estimand を 02 とした．

ICCyS には，各中間事象分類に対応する ADICE の ASEQ を格納した．したがって，ICC1S は有害事象による割付治療中止，ICC2S は死亡との関係を示す．なお，y は中間事象の発現順を示すものではない．本データでは，死亡後の予定評価時点に対応する評価レコードを ADEFF に作成しなかったため，ICC2S は全レコードで欠測となり，実際に値が設定された中間事象種類別のトレーサビリティ変数は ICC1S のみであった．

Estimand ごとの変数では，例えば ICC1E01F は，有害事象による割付治療中止が主要 Estimand における評価レコードの取扱いに実際に影響するかを示し，ICE01S は，主要 Estimand において優先順位を考慮して最終的に取扱いを決定する中間事象の ASEQ を示す．zz は解析番号ではなく Estimand 番号であるため，同一の Estimand に複数の解析が対応する場合も同じ zz を用い，解析ごとのレコード採否は ANLxxFL で区別した．

5.3. 解析帳票による結果提示

5.3.1. 中間事象の発現状況の要約

各期間における中間事象の発現状況を下表に示す.

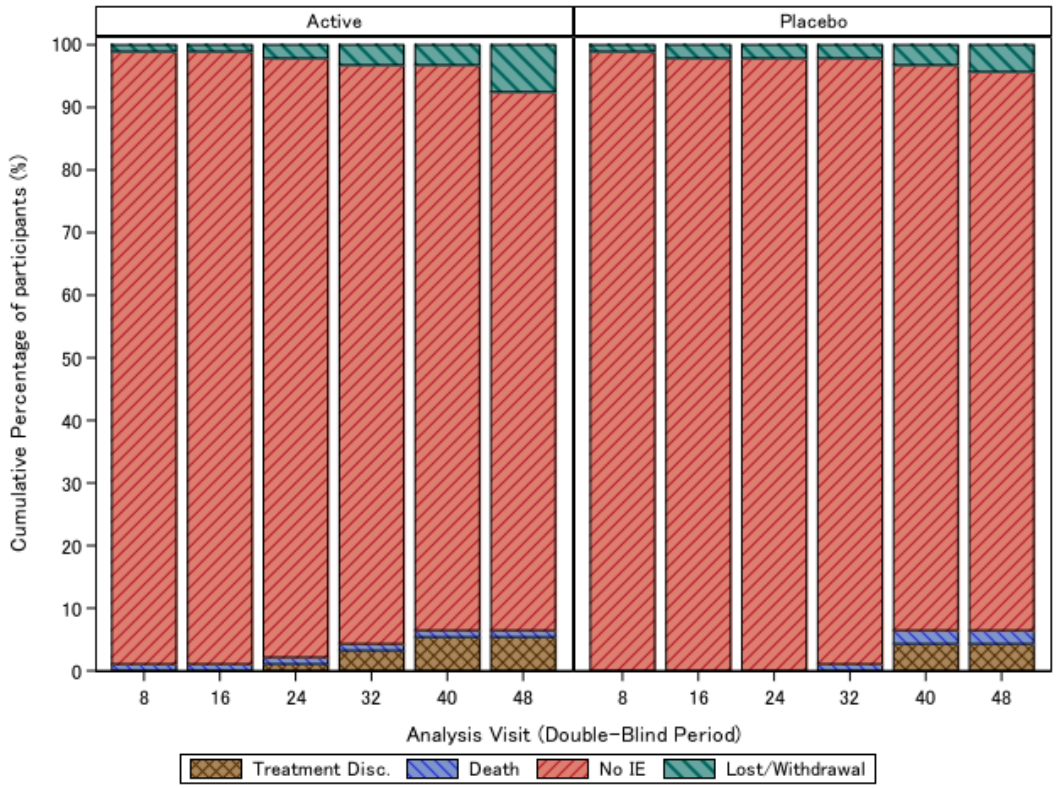
Table X.X　Summary of Intercurrent Events

Epoch	Category	Active n (%)	Placebo n (%)
Double-Blind Period	Number of Participants	93	93
	Participants without any intercurrent event	87 (93.5)	86 (92.5)
	Participants with at least one intercurrent event	6 (6.5)	7 (7.5)
	Treatment Discontinuation due to AE	5 (5.4)	4 (4.3)
	Death	1 (1.1)	3 (3.2)
Extension Period	Number of Participants	81	83
	Participants without any intercurrent event	77 (95.1)	80 (96.4)
	Participants with at least one intercurrent event	4 (4.9)	3 (3.6)
	Treatment Discontinuation due to AE	2 (2.5)	2 (2.4)
	Death	2 (2.5)	1 (1.2)

Active and Placebo indicate the randomized treatment groups during the double-blind period.

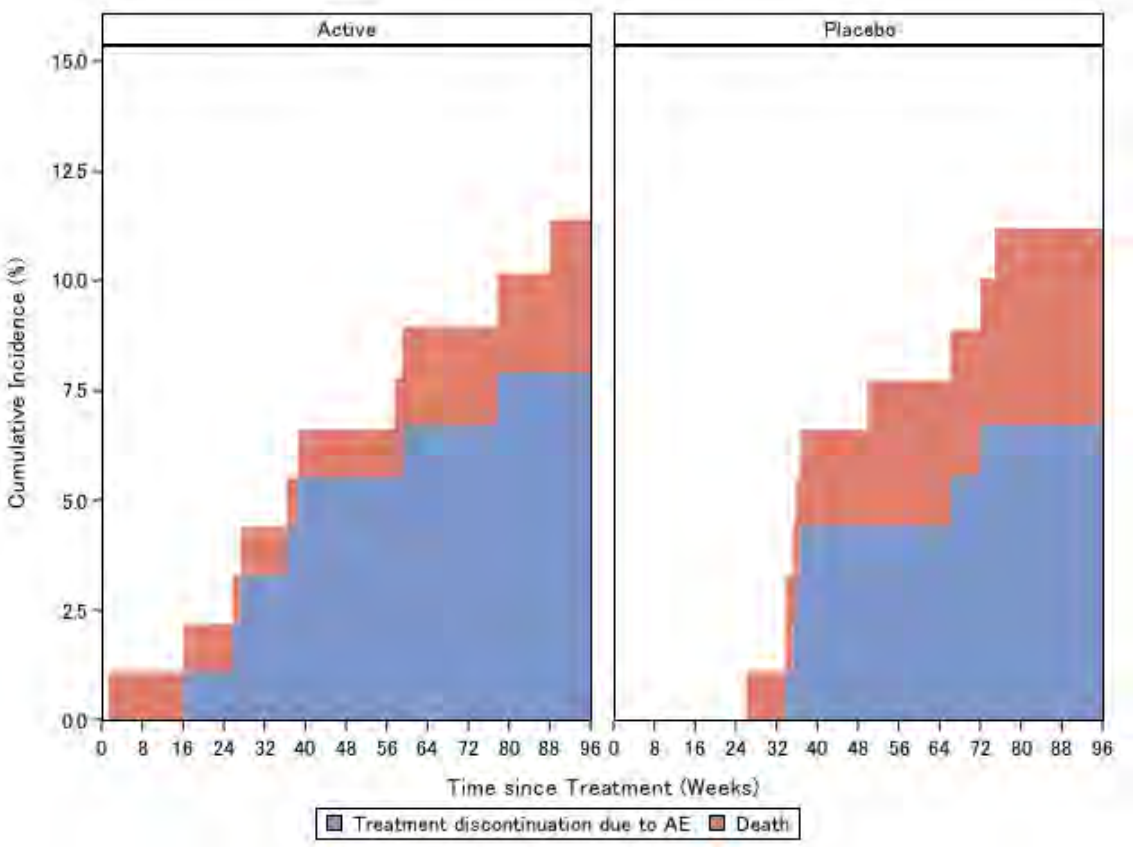
二重盲検期の評価時点における中間事象の種類別の累積割合を割付群別の積み上げ棒グラフに示す.

Figure X.X　Intercurrent Events and Study Withdrawal Over Time (Double-Blind Period)



治療開始から最初の中間事象発現までの期間について，種類別の累積発現率を積み上げ面グラフとして示す.

Figure X.X　Cumulative Incidence of Intercurrent Events Over Time



5.3.2. 主解析，感度分析及び補足的解析

主要評価項目に対する各解析の結果を下表に示す．群間差は主解析−10.3%，感度分析（Jump to Reference 法）−9.9%，補足的解析−10.3%であり，解析間で推定値は概ね同様であった.

Table X.X　Treatment Effect Estimates for Percent Change from Baseline in ALT at Week 48 by Estimand and Analysis

	LS Mean at Week 48 (95% CI)		Difference (Active − Placebo)	
	Active	Placebo	Estimate (95% CI)	p-value
Estimand 01. Primary: Hypothetical Strategy with MAR assumption				
	−33.0 (−34.8 to −31.2) %	−22.7 (−24.5 to −20.9) %	−10.3 (−12.8 to −7.7) %	<.0001
Estimand 01. Sensitivity: Hypothetical Strategy with MNAR assumption (Jump to Reference)				
	−32.6 (−34.4 to −30.8) %	−22.7 (−24.5 to −20.9) %	−9.9 (−12.4 to −7.4) %	<.0001
Estimand 02. Supplementary: Treatment-Policy Strategy for Treatment Discontinuation; Otherwise as Primary (MAR)				
	−33.0 (−34.8 to −31.2) %	−22.7 (−24.5 to −20.9) %	−10.3 (−12.9 to −7.8) %	<.0001

6. 考察

本検討では，PhUSE のホワイトペーパーで提案された実装方法を公開 CDISC 標準データセットに適用し，中間事象の特定から ADaM への反映，統計解析及び解析帳票による結果提示までを一連の実装として具体化した．中間事象を解析用データセット ADICE に集約し，有効性評価データとの関係を有効性解析用データセット ADEFF に反映することで，SDTM から ADaM 及び解析帳票に至るトレーサビリティ（追跡可能性）を確保する一つの実装例を示した．さらに，実務では，cSDRG，ADRG，Define-XML 及び Analysis Results Metadata（ARM）も活用し，データの導出方法並びに各 Estimand，解析方法及び解析結果の対応を明示することで，解析から報告までの透明性向上にもつながると考えられる．

ICH E6(R3)の国際合意を受け，国内では GCP 省令等の改正作業が進められ，2026 年 7 月には改正省令案が公示された[23]．ICH E6(R3) Annex 1 では，「Statistical Programming and Data Analysis（統計プログラミング及びデータ解析）」として独立した項が設けられ，統計プログラミング及びデータ解析に対する適切かつ文書化された品質管理，データ処理及び解析におけるデータ変換・導出のトレーサビリティ，解析集団への組入れ・除外基準の事前規定及び参加者又は特定のデータポイントを除外する理由の文書化，並びに計画された統計解析からの逸脱等の文書化及び正当化が求められている[24]．これらを踏まえると，設計された Estimand に基づくデータ収集，統計解析及び報告のプロセス・手順を明確にし，適切な文書及び記録によって説明可能に保つことは，今後さらに重要になってくる．

中間事象を集約した ADICE は，他の解析用データセットへの中間事象情報の反映や中間事象の発現状況の要約に用いられることから，関連情報を一元的に管理するために有用である．本実装では，中間事象の分類を示す ACAT1 に加え，分類と ADEFF の変数との対応を明確にするため ACAT1N を設定し，その値を ICCyS 等の添字 y と対応させた．このため，ACAT1 の分類粒度が ADEFF の変数構成に直接影響したことから，その粒度は Estimand ごとのストラテジー及び解析上の取扱いを踏まえて慎重に定める必要がある．また，ADICE を中間事象の発現状況の要約にも利用する場合には，発現日など，要約や時間解析に必要な変数も併せて検討する必要がある．

有効性解析用データセット ADEFF では，中間事象と評価レコードとの関係を示す変数と，各解析に用いるレコードを識別する変数の役割を区別した点が重要である．ICCyS，ICCyEzzF 及び ICEzzS は，それぞれ評価レコードに影響し得る中間事象，Estimand ごとのデータの取扱いへの影響及び最終的に優先される中間事象を示し，ANLxxFL には各解析におけるレコードの採否を格納した．これらを別の変数で管理することで，中間事象の影響と解析上のレコード採否を区別しつつ，トレーサビリティを確保できることを確認した．一方，各 Estimand の解析に採用するレコードを直接識別するフラグとしては，ホワイトペーパーのドラフト版に存在した ESTzzRFL の方が，ANLxxFL より直感的に理解しやすいと考えられた[25].

本稿では，中間事象の発現状況について 3 種類の要約方法を例示したが，適切な報告方法は全試験で一様ではなく，対象とする Estimand，中間事象の種類，採用するストラテジー等を踏まえて試験ごとに検討する必要がある[19]．本検討では作成していないが，個々の中間事象の発現状況及び取扱いを確認できるよう，被

験者ごとの中間事象の種類，発現日及び Estimand ごとの取扱い等を示す一覧も実務上有用と考えられる．また，発現時期が重要な場合には，被験者数及び割合に加えて初発の中間事象までの時間を Failure Plot 等により示すことが考えられる．中間事象の種類別の初回発現までの時間を評価する場合には，他の種類の中間事象が競合事象となり得るため，競合リスクを考慮するか否かを試験ごとに検討する必要がある．本検討では，種類別の累積発現率を累積発生関数により推定した．なお，中間事象の要約方法によっては，ADICE から派生した解析用データセット，例えば time-to-event 型の解析に対応した ADICETTE が有用と考えられた．

ICH E9(R1)では，主解析，感度分析及び補足的解析の結果を治験総括報告書において体系的に報告することが求められている[2]．本検討では，関連する解析結果を比較しやすくするため，先行文献を参考に，主解析，感度分析及び補足的解析を同一の解析帳票に提示した[19]．実務上，主解析，感度分析及び補足的解析は別個の図表として作成される場合も多い．しかし，感度分析を含めて一つの図表に関連する解析結果を併記することで，各解析が対象とする Estimand，ストラテジー及び解析上の仮定の違いが明確となり，仮定に対する安定性の評価及び臨床的解釈がしやすくなる．

本検討の限界として，単一の公開サンプル試験を対象とし，主要評価項目及び中間事象の種類を単純化したため，実際の臨床試験の多様な状況を網羅していない．また，中間事象を正確かつ十分な粒度で収集することは Estimand を適切に実装する上で重要であるが，本検討では既存の公開サンプルデータを用いたため，CRF 設計を含むデータ収集段階については検討していない．さらに，同一被験者に複数種類の中間事象が発現して優先順位が評価レコードの取扱いに影響する状況や，多数の Estimand を並行して実装・管理する際のデータセットの作成及び解析上の課題は十分に検討していない．加えて，Estimand に関する情報及び解析結果を治験総括報告書のどの章・節に配置し，相互に関連付けて報告するかも，重要な検討課題である[19]．一方で，本検討で示した実装例は，より複雑な試験への適用方法を検討する上での出発点になり得ると考える．

7. まとめ

適切に報告するまでが Estimand である．臨床的疑問を反映した Estimand に基づき，試験デザイン，データ収集，統計解析及び報告までを一貫して実施し，その結果を適切に解釈できて初めて意義をもつ．本稿では，その実装例として，中間事象の特定から解析結果の提示までを検討した．Estimand を報告まで確実につなぐうえで，SAS ユーザーの貢献が期待される．

謝辞：本稿の検討に際し，CDISC 標準の臨床試験サンプルデータセットを公開いただいた北海道大学病院 医療・ヘルスサイエンス研究開発機構 データサイエンスセンターの関係者の皆様に，深く感謝申し上げます．なお，本稿では当該データに仮想的な設定を加えており，本稿の内容及び解釈は著者らの責任によるものである．

連絡先：morita.yusuke@nobelpharma.co.jp

Appendix：本稿に添付する SAS プログラム

プログラム名	内容
adice.sas	中間事象に関する情報を集約した ADICE を作成するプログラム
adeff.sas	Estimand に対応した解析用データセット ADEFF を作成するプログラム
summary_ice.sas	ADICE に基づき中間事象の発現状況を要約するためのプログラム
primary_endpoint.sas	ADEFF に基づき各 Estimand に対応した解析を行うプログラム

なお，プログラムの動作確認は SAS 9.4 TS1M7 で実施した．

引用文献・サイト

[1] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials E9(R1) [Internet]. ICH; 2019 [cited 2026 Sep 18]. Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf

[2] 厚生労働省医薬局医薬品審査管理課. 「臨床試験のための統計的原則」の補遺について. 医薬薬審発 0620 第 1 号 [Internet]. 2024 Jun 20 [cited 2026 Sep 18]. Available from: https://www.pmda.go.jp/files/000269154.pdf

[3] Kahan BC, Hindley J, Edwards M, Cro S, Morris TP. The estimands framework: a primer on the ICH E9(R1) addendum. BMJ. 2024;384:e076316. doi:10.1136/bmj-2023-076316.

[4] Lynggaard H, Keene ON, Mütze T, Rehal S. Applying the Estimand Framework to Non-Inferiority Trials. Pharm Stat. 2024;23(6):1156-1165. doi:10.1002/pst.2433.

[5] Manitz J, Kan-Dobrosky N, Buchner H, Casadebaig ML, Degtyarev E, Dey J, et al. Estimands for overall survival in clinical trials with treatment switching in oncology. Pharm Stat. 2022;21(1):150-162. doi:10.1002/pst.2158.

[6] Fleming TR, Carroll KJ, Wittes JT, Emerson SS, Rothmann MD, Collins S, Levin G. A Perspective on the Appropriate Implementation of ICH E9(R1) Addendum Strategies for Handling Intercurrent Events. Stat Med. 2025;44(10-12):e70104. doi:10.1002/sim.70104.

[7] Kahan BC, Li F, Harhay MO, Cro S. What Makes an Estimand Useful? Guidance on the Choice of Intercurrent Event Strategies. Stat Med. 2026;45(6-7):e70452. doi:10.1002/sim.70452.

[8] Keene O, Wright D, Fletcher C. Commentary on Fleming et al. “A Perspective on the Appropriate Implementation of ICH E9(R1) Addendum Strategies for Handling Intercurrent Events”. Stat Med. 2026;45(6-7):e70453. doi:10.1002/sim.70453.

[9] Morris TP, Ocampo A, Madsen J, Michiels H, Roels S. A Causal Perspective on “Appropriate Implementation of ICH E9(R1) Addendum Strategies” (Comment on Fleming et al.). Stat Med. 2026;45(6-7):e70455. doi:10.1002/sim.70455.

[10] Fleming TR, Carroll KJ, Wittes JT, Emerson SS, Rothmann MD, Collins S, Levin G. Rejoinder to Commentaries on “A Perspective on the Appropriate Implementation of ICH E9(R1) Addendum Strategies for Handling Intercurrent Events”. Stat Med. 2026;45(15-17):e70622. doi:10.1002/sim.70622.

[11] Troxel AB, Gatsonis CA, Hogan JW, Hubbard RA, Hunter DJ, Normand SL. Statistics in Medicine—What’s in an Estimand? N Engl J Med. 2026;394(1):6-9. doi:10.1056/NEJMp2513633.

[12] Keene O, Lynggaard H, Fletcher C, Wright D, Terrill P. Different Clinical Questions Need Different Estimands. Ther Innov Regul Sci. 2026. doi:10.1007/s43441-026-01012-z.

[13] 独立行政法人医薬品医療機器総合機構. エキシデンサー皮下注 100 mg シリンジ、同皮下注 100 mg ペン（デペモキマブ（遺伝子組換え））審査報告書 [Internet]. 2025 Nov 17 [cited 2026 Sep 18]. Available from: https://www.pmda.go.jp/drugs/2026/P20260116002/340278000_30700AMX00273_A100_1.pdf

[14] 独立行政法人医薬品医療機器総合機構. デノスマブ BS 皮下注 120 mg RM「F」（デノスマブ（遺伝子組換え）［デノスマブ後続 1］）審査報告書 [Internet]. 2025 Aug 4 [cited 2026 Sep 18]. Available from: https://www.pmda.go.jp/drugs/2025/P20251027002/670109000_30700AMX00220_A100_1.pdf

[15] U.S. Food and Drug Administration, Center for Drug Evaluation and Research. Statistical Review(s): Application Number 210598Orig1s000 (revefenacin) [Internet]. Silver Spring (MD): U.S. Food and Drug Administration; 2018 [cited 2026 Sep 18]. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210598Orig1s000StatR.pdf

[16] European Medicines Agency. Wainzua (eplontersen): EPAR – Public Assessment Report. EMA/CHMP/515905/2024 [Internet]. Amsterdam: European Medicines Agency; 2025 [cited 2026 Sep 18]. Available from: https://www.ema.europa.eu/en/documents/assessment-report/wainzua-epar-public-assessment-report_en.pdf

[17] Cro S, Kahan BC, Rehal S, Chis Ster A, Carpenter JR, White IR, et al. Evaluating how clear the questions being investigated in randomised trials are: systematic review of estimands. BMJ. 2022;378:e070146. doi:10.1136/bmj-2022-070146.

[18] 森田 祐介. 小さな組織の Estimand 実践報告 2026 ～グローバル・日本・茅場町～. 臨床評価研究会 2026 年度総会; 2026 Jun 26; 東京.

[19] Lanius V, Glocker B, Lösch C, Bratton DJ, Callegari F, Wright M, et al. Realizing the benefits of the estimand framework when reporting and communicating clinical trial results—some recommendations. Trials. 2025;26:241. doi:10.1186/s13063-025-08915-6.

[20] PhUSE. Implementation of ICH E9(R1) Estimands Framework using Data Standards [Internet]. PhUSE; 2025 [cited 2026 Sep 18]. Available from: https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/WP-092.pdf

[21] 北海道大学病院 医療・ヘルスサイエンス研究開発機構 データサイエンスセンター. CDISC 標準の臨床試験サンプルデータセット [Internet]. [cited 2026 Sep 18]. Available from: https://helios.huhp.hokudai.ac.jp/dsc/tool/cdisc/

[22] 藤原 正和. 欠測のあるデータの解析のための SAS プログラムの紹介 ～データ発生・DIA マクロとプロシジャの進展～ (5) Pattern Mixture Model と Multiple Imputation に基づく解析 2 (Reference-based imputation). SAS ユーザー総会 2016 [Internet]. 2016 [cited 2026 Sep 18]. Available from: https://www.sas.com/content/dam/SAS/ja_jp/doc/event/sas-user-groups/usergroups2016-a-02-05.pdf

[23] 厚生労働省. 医薬品の臨床試験の実施の基準に関する省令等の一部を改正する省令（案）に関する御意見の募集について [Internet]. 2026 Jul 31 [cited 2026 Sep 18]. Available from: https://public-comment.e-gov.go.jp/pcm/detail?CLASSNAME=PCMMSTDETAIL&Mode=0&id=495260127

[24] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. ICH Harmonised Guideline: Guideline for Good Clinical Practice E6(R3). Final version [Internet]. ICH; 2025 [cited 2026 Sep 18]. Available from: https://database.ich.org/sites/default/files/ICH_E6%28R3%29_Step4_FinalGuideline_2025_0106.pdf

[25] PhUSE. Implementation of ICH E9(R1) Estimands Framework using Data Standards. Public Review Draft [Internet]. Updated 2023 Nov 15 [cited 2026 Aug 22]. Available from: https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Under+Review/Implementation+of+Estimands+Using+Data+Standards+-+White+Paper.pdf

MMRMにおける分散共分散構造及び
ロバスト分散推定が推定結果に与える影響の検討

○牛尾 英孝

（株式会社ピープロ・ジャパン）

Effects of Covariance Structure and Robust Variance Estimation
in a Mixed Model for Repeated Measures

Hidetaka Ushio
Ppro-Japan Co.,Ltd.

要旨

MMRM における分散共分散構造とロバスト分散推定の組み合わせに着目し，それらが推定結果に与える影響をロバスト分散による推定分散の頑健性，標本サイズ及び欠測割合等の条件も踏まえて統合的に評価する．

キーワード：MMRM（Mixed Model for Repeated Measures），分散共分散構造，ロバスト分散推定（sandwich 推定量）

1　はじめに

近年，臨床試験における反復測定データの解析では，MMRM が標準的な解析手法として広く用いられている．一方，MMRM では，分散共分散構造の誤指定により，欠測データの存在下で回帰パラメータの推定値に bias が生じる可能性があるほか，標準誤差が過小または過大に評価される可能性が報告されている．分散共分散構造の誤指定に対する推定の頑健性を高める手法として，ロバスト分散推定（sandwich 推定量）が提案されており，その性能について，分散共分散構造，標本サイズ，欠測割合などの異なる条件下で評価されている．ただし，ロバスト分散推定は分散共分散構造が誤指定された場合に対しても分散の推定値を頑健にする一方，推定値の bias 自体を解消するものではない．これまで，分散共分散構造の誤指定による影響やロバスト分散推定の性能について検討されてきたが，本演題では真の分散共分散構造と解析時に仮定する分散共分散構造を明示的に組み合わせ，さらにロバスト分散推定の有無，標本サイズ及び欠測割合などの条件を体系的に比較することで，分散共分散構造の誤指定が推定結果に与える影響を統合的に評価する．

2　シミュレーションに用いるパラメータ条件設定

本シミュレーションでは，Baseline 及び治療後 4 時点（V1，V2，V3，V4）において，連続変数を繰り返し測定する 2 群間比較試験（連続変数の値が増加するほど改善傾向であるものとする）を想定した．投与群として Control 群及び Treatment 群の 2 群を設定し，各群の症例数はシミュレーション条件により 40 例（欠測がない場合の総症例数は 80 例）または 80 例（欠測がない場合の総症例数は 160 例）とした．各症例について，V1，V2，V3，V4 の 4 時点の測定値は多変量正規分布に従うものとし，平均構造及び分散共分散構造をそれぞれ設定した．平均構造は，分散共分散構造の設定とは独立に規定し，すべてのシミュレーション条件で共通とした．時点の順序を Baseline，V1，V2，V3，V4 とし，Control 群及び Treatment 群の真の平均ベクトルをそれぞれ

$$\mu_c = (0,0.5,1.0,1.5,2.0)^t,\quad \mu_t = (0,0.8,1.6,2.3,3.0)^t$$

とした．なお，t はベクトルの転置を表す．両群とも Baseline における平均値は 0 とし，治療後は時間の経過に伴って平均値が増加する構造とした．Treatment 群では，Control 群と比較して V1 以降の平均値が一貫して高くなるように設定し，Treatment 群において経時的に大きな改善を示す平均構造を仮定した．各時点における Treatment 群と Control 群の真の平均差は V1，V2，V3 及び V4 でそれぞれ 0.3，0.6，0.8 及び 1.0，主要な評価時点として V4 を想定し，V4 における真の Treatment 群と Control 群の平均差は 1.0 である．以降のシミュレーションでは，真の分散共分散構造，解析時に指定する分散共分散構造，ロバスト分散推定の有無，症例数及び欠測割合を設定し，それらの組み合わせが推定結果に及ぼす影響を評価した．

3　真の分散共分散構造

3.1　[Unstructured：UN]

UN では，各時点の分散及び時点間の共分散を個別に設定し，測定値のばらつきと時点間の関連性を柔軟に表現した．行列の対角成分は各測定時点での分散を示し，本シミュレーションでは V1～V4 の各時点での分散をすべて 1.00 とした．一方，非対角成分は 2 時点間の共分散を示し，例えば V1 と V2 の共分散は 0.45，V1 と V4 の共分散は 0.20 とした．UN では，測定時点間の距離に応じて共分散が一定の規則に従って低下するという制約を設けず，各時点間の関連性を個別に表現できる．本シミュレーション条件で設定した UN の分散共分散行列を以下に示す．

$$\Sigma = \begin{pmatrix} 1.00 & 0.45 & 0.30 & 0.20 \\ 0.45 & 1.00 & 0.50 & 0.35 \\ 0.30 & 0.50 & 1.00 & 0.40 \\ 0.20 & 0.35 & 0.40 & 1.00 \end{pmatrix}$$

3.2　[Autoregressive of order 1：AR(1)]

AR(1)では，各測定時点での分散を共通とし，測定時点間の相関が時点間の距離に応じて減衰することを仮定した．具体的には，時点間の距離が 1，2，3 と増加するにつれて，相関はそれぞれ 0.5，0.5^2，0.5^3

で表される．本シミュレーションでは，V1～V4 の各時点での分散を 1.00，隣接する測定時点間の相関係数を 0.5 と設定し，これにより，測定時点間の距離が大きくなるほど相関が小さくなる構造とした．なお，本シミュレーションで設定した AR(1)の分散共分散行列を以下に示す．

$$\Sigma=\begin{pmatrix} 1.00 & 0.50 & 0.25 & 0.125 \\ 0.50 & 1.00 & 0.50 & 0.25 \\ 0.25 & 0.50 & 1.00 & 0.50 \\ 0.125 & 0.25 & 0.50 & 1.00 \end{pmatrix}$$

4　Scenario 設定及び評価項目

分散共分散構造について，真の分散共分散構造が UN，解析時に仮定する分散共分散構造として AR(1)，真の分散共分散構造が AR(1)，解析時に仮定する分散共分散構造として UN の 2 通り，ロバスト分散推定の有/無，症例数の 2 水準，欠測割合の 2 水準を組み合わせ，計 16 条件（表 4.1）でシミュレーションを各々1000 回実施し，表 4.2 に示した評価項目を算出する．なお，欠測データは MAR（Missing At Random）を仮定して発生させた．Baseline，V1，V2 及び V3 は全症例で観測されるものとし，V4 のみを欠測対象とした．V4 の欠測は，それまでに観測された Baseline，V1，V2，V3 の値に基づいて発生させ，欠測率は 10%または 20%とした．

[表 4.1：Scenario（シミュレーション条件）]

分散共分散構造	UN / AR(1)
ロバスト分散推定の有無	有 / 無
1 群あたりの症例数	40 例 / 80 例
全例に対する欠測割合	10 % / 20 %

[表 4.2：評価項目]

1	bias	群間差の平均推定値と真値（V4 時点における群間差=1.0）との差
2	SE ratio	推定された SE / 推定値の経験的 SD
3	95% CI coverage	95% CI が真値（V4 時点における群間差=1.0）を含む割合

5　解析方法

Treatment 群及び Control 群における Baseline からの変化量を応答変数とし，固定効果として投与群，評価時点，投与群と評価時点の交互作用及び Baseline 値をモデルに含めた MMRM を用いて解析した．症例内の反復測定値間の分散共分散構造には，解析時に仮定する分散共分散構造として[UN / AR(1)]を指定した．パラメータの推定には制限付き最尤法（REML）を用い，ロバスト分散推定を用いない場合は自由度の調整に Kenward-Roger 法を用いた．V4 時点における Treatment 群と Control 群の Baseline からの変化量の差を，最小二乗平均値の差として推定した．

6　解析 SAS プログラムコード

シミュレーションに用いた SAS プログラムのアルゴリズムを示す.

```
/************************************************/
/* MMRM Simulation: True covariance = UN        */
/************************************************/
%let n_per_group=40;
%let sigma=1;
%let miss_rate=0.10;
%let nsim=1000;

ods _all_ close;
%macro simulation;
%do sim=1 %to &nsim;
/* Data generation */
data simdata_&sim;
    call streaminit(123456+&sim);
    do trt=0,1;
        do i=1 to &n_per_group;
            usubjid=trt*&n_per_group+i;
            baseline=rand("NORMAL",0,1);

            if trt=0 then do;
                mu1=0.5; mu2=1.0; mu3=1.5; mu4=2.0;
            end;
            else do;
                mu1=0.8; mu2=1.6; mu3=2.3; mu4=3.0;
            end;
```

Treatment 群と Control 群の各群の症例数を 40 例に設定（[n_per_group = 40 例/ 80 例]）し，V1～V4 の誤差項の標準偏差[sigma]を 1，V4 の欠測率を 10%（[miss_rate= 0.10 / 0.20]）とした．これらの条件で 1,000 回[nsim]のシミュレーションを実施した.

各症例について，V1，V2，V3，V4 の 4 時点の測定値は，多変量正規分布に従うものとし，各症例の Baseline 値は，平均 0，標準偏差 1 の正規分布から生成した．V1～V4 の平均値は，Control 群でそれぞれ 0.5，1.0，1.5，2.0，Treatment 群でそれぞれ 0.8，1.6，2.3，3.0 とし，V1～V4 における真の Treatment 群と Control 群の群間差は，それぞれ 0.3，0.6，0.8，1.0 となるように設定した．V1～V4 の分散共分散構造は UN（[UN（unstructured）/ AR(1)（Autoregressive order 1）]）とし，これに基づいて測定値を生成した．乱数の初期値にはシミュレーションごとに異なる値を設定し，再現可能な乱数系列を用いた.

000336

```
            z1=rand("NORMAL"); z2=rand("NORMAL");
            z3=rand("NORMAL"); z4=rand("NORMAL");

            /* True UN covariance: Cholesky decomposition */
            e1=z1;
            e2=0.4500000*z1+0.8930286*z2;
            e3=0.3000000*z1+0.3527323*z2+0.8863295*z3;
            e4=0.2000000*z1+0.2351549*z2
              +0.2900200*z3+0.9058645*z4;
        /* % let sigma=1 */
            v1=mu1+&sigma*e1;
            v2=mu2+&sigma*e2;
            v3=mu3+&sigma*e3;
            v4=mu4+&sigma*e4;
            output;
            end;
        end;
    drop i mu1-mu4 z1-z4 e1-e4;
run;
```

RAND("NORMAL")を用いて，平均 0，標準偏差 1 の互いに独立な標準正規乱数(z1～z4)を生成した．これらの乱数を Cholesky 分解に基づく係数で線形結合することにより，4 時点間に相関を持つ誤差項(e1～e4)を生成した．具体的には，e1 は z1 のみを用いて生成し，e2 は z1 と z2，e3 は z1～z3，e4 は(z1～z4)を用いて生成した．これらの係数は Cholesky 分解により得られる下三角行列の要素である．その後，各時点の平均値（mu1～mu4）に，誤差項（e1～e4）に標準偏差（sigma）を乗じた値を加えることで，V1～V4 の測定値を生成した．すなわち，V1～V4 はそれぞれ「各時点の平均値+相関を持つ誤差項」として生成される．

```
/* True AR(1) covariance:
/* % let rho = 0.5 */
e1=z1;
e2=&rho*e1 + sqrt(1-&rho**2)*z2;
e3=&rho*e2 + sqrt(1-&rho**2)*z3;
e4=&rho*e3 + sqrt(1-&rho**2)*z4;
```

AR(1)の場合，UN と同様に平均 0，標準偏差 1 の互いに独立な標準正規乱数(z1～z4)を生成し，これらを用いて 4 時点間に相関を持つ誤差項 e1～e4 を逐次的に生成した．具体的には，e1=z1 とし，e2 以降の誤差項は，直前の時点の誤差項に(&rho=0.5)を乗じた項と，新たに生成した標準正規乱数に(sqrt{1-&rho ^2})を乗じた項の和として生成した．

```
/* MAR missingness at V4 */
data simdata_&sim;
      set simdata_&sim;
      score=baseline+v1+v2+v3;
run;

proc sort data=simdata_&sim; by score; run;

data simdata_&sim;
      set simdata_&sim;
      retain nmiss 0;
      if nmiss<round(2*&n_per_group*&miss_rate) then do;
         v4=.; nmiss+1;
      end;
      drop score nmiss;
run;
```

V4 の欠測は，欠測メカニズムとして Baseline 及び V1～V3 の観測値に依存する MAR を仮定し，Baseline，V1，V2，V3 を用いて欠測判定用のスコア[score]を算出した．具体的には，score = baseline + v1 + v2 + v3 とした．このスコアは，V4 の欠測判定に用いる指標であり，V4 の値そのものは含めていない．V4 が欠測となるかどうかは，欠測が生じる時点で既に観測されている Baseline 及び V1～V3 の情報に基づいて決定される構造とした．次に，算出したスコアの昇順に症例を並べ替え，スコアの低い症例から V4 を欠測値とした．具体的には，proc sort により score の昇順にデータを並べ替えた後，nmiss により V4 を欠測とした症例数を累積的にカウントした．round(2*&n_per_group*&miss_rate)により欠測とする症例数を設定し，設定した欠測数に達するまで V4 を欠測値（.）に置き換えた．上記では，n_per_group = 40，miss_rate=0.10 としたため，V4 の欠測数は 2*40*0.10 = 8 例となる．

```
/* Wide to Long */
proc sort data=simdata_&sim;
            by usubjid trt baseline;
run;

proc transpose data=simdata_&sim
            out=simdata_long&sim(rename=(col1=aval));
            by usubjid trt baseline;
      var v1-v4;
run;
```

生成したV1～V4の測定値について，USUBJID，投与群（TRT）及びBaseline値をキーとしてデータを並べ替えた後，V1～V4の各測定値を縦持ち形式に変換した．具体的には，Proc sortを用いてUSUBJID，TRT及び

000338

Baselineの順にデータを並べ替え，Proc Transposeを用いてV1～V4を1症例1時点1レコードとなる形式に変換した．変換後の測定値は解析用変数AVALとして格納した．

```
/* MMRM data */
data mmrm_data&sim;
	set simdata_long&sim;
	avisitn=input(substr(_name_,2),8.);
	avisit=_name_;
	if not missing(aval) then chg=aval-baseline;
	keep usubjid trt baseline avisit avisitn aval chg;
run;
```

simdata_longを用いて解析用データセットを作成した．測定時点（V1～V4）を数値化したAVISITN及び時点名を示すAVISITを作成し，ベースライン値との差として変化量CHGを算出した．

```
/* MMRM */
ods output Diffs=Diffs&sim;

/*  ロバスト分散推定あり  */
proc mixed data=mmrm_data&sim method=reml empirical;
	class usubjid trt avisitn;
	model chg=baseline trt avisitn trt*avisitn;
	repeated avisitn / subject=usubjid type=AR(1);
	lsmeans trt*avisitn / diff cl;
run;

/*  ロバスト分散推定なし  */
proc mixed data=mmrm_data&sim method=reml;
	class usubjid trt avisitn;
	model chg=baseline trt avisitn trt*avisitn / ddfm=kr;
	repeated avisitn / subject=usubjid type=AR(1);
	lsmeans trt*avisitn / diff cl;
run;
%end;
%mend simulation;

%simulation;
```

各シミュレーションデータセットについて，Proc mixedを用いてMMRMによる解析を実施し，解析結果をシミュレーションごとに保存した．

7　SAS によるシミュレーション結果

シミュレーションの結果を下記の[表 7-1]に示す.

[表 7-1：各シミュレーション条件における評価項目]

Scenario	真の分散共分散構造/解析時に使用する分散共分散構造 *1	ロバスト分散推定の有無	両群の症例数	欠測割合	bias*2	SE ratio*3	95% CI coverage*4
Scenario 1	UN / AR(1)	有	80	10 %	-0.025	0.950	93.9 %
Scenario 2	UN / AR(1)	有	80	20 %	-0.032	0.948	93.3 %
Scenario 3	UN / AR(1)	有	160	10 %	-0.016	0.991	93.8 %
Scenario 4	UN / AR(1)	有	160	20 %	-0.022	0.991	94.5 %
Scenario 5	UN / AR(1)	無	80	10 %	-0.025	0.971	94.7 %
Scenario 6	UN / AR(1)	無	80	20 %	-0.032	0.971	94.3 %
Scenario 7	UN / AR(1)	無	160	10 %	-0.016	1.003	94.1 %
Scenario 8	UN / AR(1)	無	160	20 %	-0.022	1.004	94.7 %
Scenario 9	AR(1) / UN	有	80	10 %	-0.012	0.950	94.1 %
Scenario 10	AR(1) / UN	有	80	20 %	-0.014	0.941	93.4 %
Scenario 11	AR(1) / UN	有	160	10 %	-0.008	0.991	93.8 %
Scenario 12	AR(1) / UN	有	160	20 %	-0.008	0.983	94.8 %
Scenario 13	AR(1) / UN	無	80	10 %	-0.012	0.969	94.7 %
Scenario 14	AR(1) / UN	無	80	20 %	-0.014	0.968	94.4 %
Scenario 15	AR(1) / UN	無	160	10 %	-0.008	1.001	94.0 %
Scenario 16	AR(1) / UN	無	160	20 %	-0.008	0.997	95.2 %

[*1]　UN / AR(1)であれば真の分散共分散構造がUN，解析時に指定する分散共分散構造がAR(1)

[*2]　群間差の推定値と真値（V4における群間差1.0）の差の平均値

[*3]　標準誤差の平均 / 群間差の推定値の標準偏差（推定値の経験的SD）

[*4]　95% CIが真値（V4における群間差1.0）を含む割合

*2～*4については1000回のシミュレーションに対して算出

[真の分散共分散構造：UN]

[解析時に指定する分散共分散構造：AR(1)]

ロバスト分散推定を用いた場合においても，群間差の点推定値における bias は変化しなかった．ロバスト分散推定は，分散共分散構造の誤指定に対して分散の推定値を頑健にする手法であり，点推定値の bias を直接低減する手法ではないことから，本シミュレーション結果はこの特性と整合するものと考えられた.

SE ratio は実際の推定値のばらつき（経験的 SD）に対して，解析で推定された SE がどの程度の大きさであったかを示す指標であり，症例数，欠測割合の条件が同じである場合，ロバスト分散推定により SE ratio は低下傾向であった．また，95%信頼区間が真値（V4 における群間差 1.0）を含む割合はロバスト分散を指定しない方が大きくなる傾向であった．

[真の分散共分散構造：AR(1)]
[解析時に指定する分散共分散構造：UN]

真の分散共分散構造を UN，解析時に指定する分散共分散構造として AR(1)を指定した場合と同様の傾向が見られた．

8　まとめと考察

群間差の bias は，真の分散共分散構造を AR(1)，解析時に UN を指定した条件の方が，真の構造を UN，解析時に AR(1)を指定した条件より小さかった．この結果から，群間差の bias は真の分散共分散構造と解析時に指定した分散共分散構造の組合せによっても異なる可能性を示している．また，ロバスト分散推定を用いた場合には，SE ratio が小さくなる傾向とともに，95%信頼区間の coverage も低下する傾向であったものの，大きな差は認められなかったことから，本シミュレーション条件においては，分散共分散構造の誤指定が及ぼす影響は限定的であると考えられる．

9　参考文献

[1] 五所 正彦，丸尾 和司
経時測定データ解析における mixed-effects models for repeated measures（MMRM）の利用
https://www.jstage.jst.go.jp/article/jappstat/46/2/46_53/_article/-char/ja
（最終アクセス 2026/09/18）

[2] tomokichi
mmrm を SAS で実装｜PROC MIXED で反復測定混合モデル｜生物統計家への道
（最終アクセス 2026/09/18）

[3] 飯田 絢子 MMRM 入門
https://sas-user2023.ywstat.jp/download.html?n=9&key=paper
（最終アクセス 2026/09/18）

[4] 阿部 貴行『欠測データの統計解析』MMRM 入門 朝倉書店 (2016)

『What if』の視点で読み解くPROC CAUSALTRT：IPWと標準化の理解と注意点

○峯脇 颯汰[1]、浦狩 保則[1]

([1]シミック株式会社　データサイエンス事業本部)

Understanding PROC CAUSALTRT Through the “What If” Lens: IPW, Standardization, and Considerations

Sota Minewaki[1], Yasunori Uragari[1]

[1]Data Science Division, CMIC Co., Ltd.

要旨

観察研究では研究者は治療割付を制御できないため、治療群間の背景因子の不均衡が交絡を生じ、平均治療効果(ATE)の推定にバイアスをもたらす。交絡を調整しATEを推定する手法として逆確率重み付け(IPW)と標準化がある。これらの手法は因果推論の名著である What if で中心的に扱われており SAS の PROC CAUSALTRT で簡単に実装できる。CAUSALTRT では IPW を実装する際、method = IPW/IPWR/IPWS の3つのオプションがある。IPW/IPWR はそれぞれ What if の Horvitz-Thompson 推定量、Hajek 推定量に対応する。これら 3 つのオプションの違いが ATE の推定精度に与える影響と実務における使い分けの指針について解説する。一方、CAUSALTRT で標準化を実装する際、治療群ごとにアウトカムモデルを想定するのに対し、What if は単一のアウトカムモデルを想定する。しかし、単一モデルと群別モデルの使い分けについて整理されていない。これらのモデルの違いや単一モデルにおける積項が ATE の推定精度に与える影響について解説し、実務における指針を示す。

キーワード：因果推論、IPW、標準化、CAUSALTRT

1. 序論

観察研究は治療や介入の「効果」を推定することを主な目的として行われる。しかし、治療割り付けを制御できないため、治療群間で背景因子の不均衡により交絡が生じる。そのため、群間のアウトカムを単純比較するだけでは治療の効果を正しく推定できない。Hernán and Robins の Causal inference: What if では「もし集団全体が治療群ならどのような結果になるか」「もし集団全体が対照群だったらどのような結果になるか」という反事実的な視点のもと、観察データから治療の因果効果を識別・推定する方法を体系的に説明している[1]。集団全体の治療の因果効果を推定する方法として IPW と標準化がある。これらの手法は What if で中心的に扱われており、SAS9.4 SAS/STAT14.2 から PROC CAUSALTRT で簡単に実装できるようになった[1,2]。本稿の目的は CAUSALTRT で扱える IPW と標準化を What if の視点から読み解くことで理論

と実装のつながりを明確にすることにある。また、CAUSALTRT で扱える IPW の手法の推定精度の違いや What if で説明されている標準化と CAUSALTRT の標準化（CAUSALTRT 上では正しくは回帰調整（Regression adjustment）と表現される）の推定精度の違いについても数値実験で検討し、これらの手法を扱う上での注意点を解説する。

2. 諸前提

本稿では、ベースライン時点で定義される二値治療A 、ベースライン共変量L 、およびアウトカムYからなる 1 時点の治療を考える。ここで$A=1$は治療あり、$A=0$は治療なしを表す。観察データでは、治療群と対照群で共変量分布が異なることが多く、観測されたアウトカムの差をそのまま治療の効果として解釈できない。因果効果を議論するため、潜在アウトカムを用いて治療の因果効果を定義し、そのうえで観測データからその量を識別するための条件を整理する必要がある。

2.1. 潜在アウトカムと ATE

因果推論では、各症例について「治療を受けた場合のアウトカム」と「治療を受けなかった場合のアウトカム」を考える。治療を受けた場合の潜在アウトカムをY^1、治療を受けなかった場合の潜在アウトカムをY^0とすると個体レベルの因果効果はY^1-Y^0として定義される[1]。しかし、同一症例のY^1とY^0は同時に観測できないため症例ごとの因果効果は求められない。そのため、関心の対象は母集団の平均的な因果効果となることが多い。母集団の平均的な因果効果である ATE は$E[Y^1]-E[Y^0]$と定義される。ATE は「対象集団全体が治療を受けた場合の平均アウトカム」と「対象集団全体が治療を受けなかった場合の平均アウトカム」の差として解釈できる。

2.2. 識別条件

実際はY^1とY^0は観測できないため、観測データから ATE を推定するために以下の 3 つの識別条件を仮定する必要がある[1]。これらの識別条件をすべて満たすことで観察研究を条件付きランダム化比較試験とみなせ、観測データYから ATE を推定することができる。

1. （条件付き）交換可能性

   観察研究では治療割り付けが無作為でないため、一般的に交換可能性$Y^a \perp A$は成立しない。そのため、共変量Lで条件づけた条件付き交換可能性$Y^a \perp A \mid L$を仮定する必要がある。

2. 一致性

   観察研究では、研究者が想定する治療の内容が曖昧だと対応する潜在アウトカムY^aも不明瞭になるため比較する治療aの定義を明確にする必要がある。そのうえで、実際に治療$A=a$を受けた症例は観測アウトカムYが潜在アウトカムY^aに一致するという一致性の仮定が必要である。一致性は$Y=AY^1+(1-A)Y^0$と表される。

3. 正値性

   条件付ける共変量Lの各水準において各治療水準が観測されることである。正値性は$P(A=a \mid L=l)>0$と表される。

2.3. IPW

2.3.1. What if で説明される IPW の役割

IPW の役割は観測集団を重みづけることで、あたかも治療がランダムに割り付けられ交換可能性が成り立つと考えられる疑似集団を作り出すことにある[1]。図 1 の右にある疑似集団を比較することで因果効果を推定できるようになる。つまり、疑似集団内では関連を因果とみなせる。

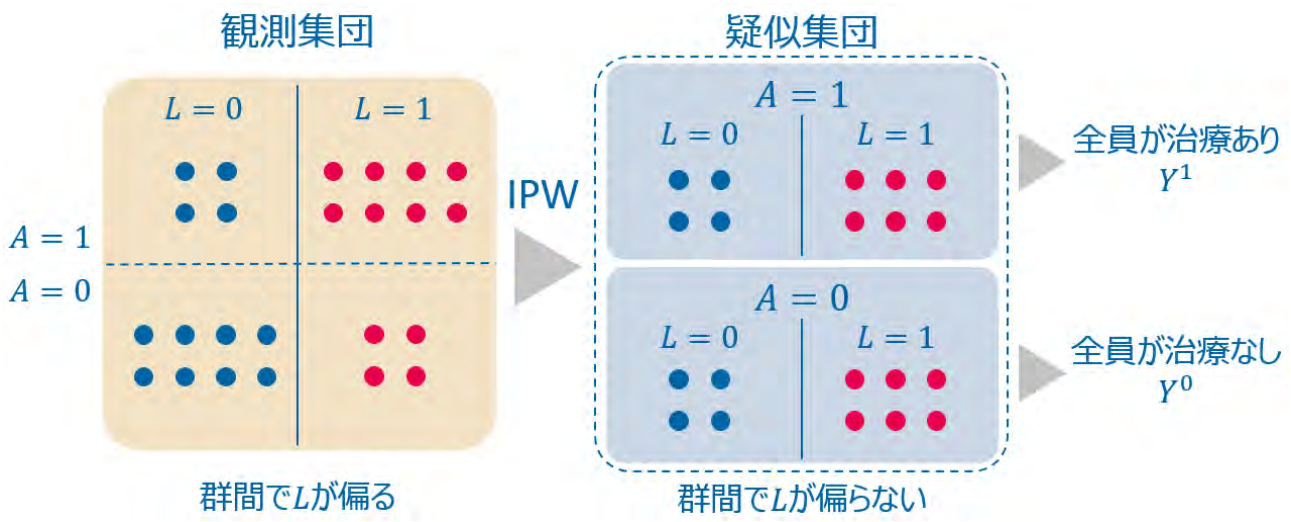


図 1：IPW の役割

各症例の重みは割付確率（傾向スコア）の逆数である$W^A = 1/P(A \mid L)$で定義され、傾向スコアを$e(L) = P(A = 1 \mid L)$とすると治療群では$1/e(L)$、対照群では$1/(1 - e(L))$が重みとなる。この重みを用いることで ATE は以下の式で求められる。

$$ATE = E[Y^1] - E[Y^0] = E\left[\frac{AY}{e(L)}\right] - E\left[\frac{(1-A)Y}{1-e(L)}\right]$$

What if ではW^Aにくわえて安定化重み$SW^A = P(A)/P(A \mid L)$も紹介されている。安定化重みを用いても交換可能性が成り立つと仮定できる疑似集団を作り出すという役割は変わらない。安定化重みを用いると極端な重みをもつことを防ぐため、有限標本における推定の不安定性を軽減できる。

2.3.2. Horvits-Thompson 型推定量、Hájek 型推定量

What if では ATE の推定において Horvits-Thompson（HT）型推定量と Hájek（H）型推定量を紹介しており、治療が二値の場合において$f(A|L)$をW^AまたはSW^Aとするとそれぞれ以下の式で表せる[1,3,4]。

$$\text{HT 型推定量：} \hat{E}[Y^a] = \hat{E}\left[\frac{\mathbb{I}(A=a)Y}{f(A|L)}\right] = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbb{I}(A_i=a)y_i}{\hat{f}(A|L_i)}$$

$$\text{H 型推定量：} \hat{E}[Y^a] = \left(\hat{E}\left[\frac{\mathbb{I}(A=a)}{f(A|L)}\right]\right)^{-1}\hat{E}\left[\frac{\mathbb{I}(A=a)Y}{f(A|L)}\right] = \left(\sum_{i=1}^{n}\frac{\mathbb{I}(A_i=a)}{\hat{f}(A|L_i)}\right)^{-1}\sum_{i=1}^{n}\frac{\mathbb{I}(A_i=a)y_i}{\hat{f}(A|L_i)}$$

HT 型推定量は単純な重みづけ平均であるのに対し、H 型推定量は重みの総和で正規化した推定量であることが分かる。実務上は、H 型推定量のほうが有限標本で安定しやすく推定値が自然な範囲に収まりやすいという利点がある。HT 型推定量を用いる際はSW^Aを用いるとバイアスが生じるので注意が必要である。

2.3.3. CAUSALTRT における IPW 系の method と What if との対応

CAUSALTRT では、ATE を推定するための IPW 法として method=IPW / IPWR / IPWS が実装されている。重みはW^Aのみを想定しており、以下のように psmodel ステートメントで傾向スコアモデルを指定する。

```
proc causaltrt data = simdata method = IPW/IPWR/IPWS;
   psmodel A(event='1') = L;
   model Y / dist=normal;
   output out=result_IPW
run;
```

What if と CAUSALTRT の各 method は次のように対応付けることができる。

- method=IPW：What if の HT 型推定量に対応
- method=IPWR：What if の H 型推定量に対応
- method=IPWS：What if では紹介されていないが、ATE 推定量の大標本分散を最小化するために HT / H 型推定量を拡張したものであり、交換可能性が成り立つと仮定できる疑似母集団を作り出すという役割は変わらない[5]。また、H 型推定量と同様に重みの総和で正規化した推定量である。

HT 型推定量と H 型推定量の正規化の有無と使用できる重みの対応を表 1 に示す。

表 1：HT 型推定量と H 型推定量の正規化の有無と使用できる重みの対応

	$W^A = 1/P(A \mid L)$	$SW^A = P(A)/P(A \mid L)$
Horvits-Thompson 型（正規化なし）	What if のW^A CAUSALTRT の method=IPW	
Hájek 型（正規化あり）	CAUSALTRT の method=IPWR	What if のSW^A

CAUSALTRT で実装できる IPW 、IPWR、IPWS の ATE 推定精度を数値実験にて様々な条件下で比較する。

2.4. 標準化

2.4.1. What if で説明される標準化の役割

標準化の役割は各共変量組合せの層ごとの条件付き平均を集団全体に拡張して ATE を求めることにある[1]。離散的な共変量Lを考えると潜在アウトカムの平均は以下のように表せる。

$$E[Y^a] = \sum_l E[Y \mid A = a, L = l]\, P(L = l)$$

$$= E_L[E[Y \mid A = a, L]]$$

つまり、各層$L = l$における条件付き平均$E[Y \mid A = a, L = l]$を求め、各層の出現率$P(L = l)$で重みづけ平均することで集団全体が治療aを受けた場合の潜在アウトカムの平均を得る。

共変量が高次元または連続変数を含む場合、各層における条件付き平均をノンパラメトリックに推定するのは非現実的である。そのため、条件付き平均に対して回帰モデルを仮定し、その推定値を用いて標準化する。これが parametric g-formula である[1]。手順としては、まず観測データセットを用いてアウトカム回帰モデル$E[Y \mid A = a, L = l]$を推定する。次に観測データセットを共変量はそのままにして、「すべての症例が治療群」、「すべての症例が対照群」であるデータセットをそれぞれ用意する。これらのデータセットに対してアウトカム回帰モデルの推定式$\hat{E}[Y \mid A = a, L = l]$を用いて、各症例の潜在アウトカムの予測値$\hat{Y}^1$、$\hat{Y}^0$を得る。最後に$\hat{\mu}_1 = (1/n)\sum_{i=1}^n \hat{Y}_i^1$、$\hat{\mu}_0 = (1/n)\sum_{i=1}^n \hat{Y}_i^0$を算出し、$\widehat{ATE} = \hat{\mu}_1 - \hat{\mu}_0$を得る。これは条件付き平均アウトカムを plug-in して g-formula を評価する方法であるため、plug-in g-formula とよばれる[1]。

2.4.2. CAUSALTRT における REGADJ と What if との対応

CAUSALTRT では ATE を推定するための標準化に対応する方法として method=REGADJ が実装されている。以下のように model ステートメントで推定するアウトカム回帰モデルを指定する。

```
proc causaltrt data = simdata method = REGADJ;
  psmodel A(event='1');
  model Y = L1-L10 / dist=normal;
  output out=result_IPW
run;
```

CAUSALTRT では群ごとにアウトカム回帰モデル$\hat{Y}^1 = g^{-1}(\boldsymbol{L}\hat{\boldsymbol{\beta}}_{\boldsymbol{t}})$、$\hat{Y}^0 = g^{-1}(\boldsymbol{L}\hat{\boldsymbol{\beta}}_{\boldsymbol{c}})$ を推定する[2]。ここでgはリンク関数、$\hat{\beta}_t$と$\hat{\beta}_c$は治療群と対照群で推定されたパラメータである。観測データからアウトカム回帰モデルを推定し、推定したモデルから「全員が対照群」「全員が治療群」という 2 つの反事実アウトカムを予測し、平均するという点で、What if の標準化と同じ手順を踏む。ところがアウトカム回帰モデルの推定方法が少し異なる。What if では$E[Y \mid A = a, L = l] = A\beta_A + \boldsymbol{L\beta} + A\boldsymbol{L\gamma}$というかたちで 1 つのモデルでアウトカム回帰モデルを推定するのに対し、CAUSALTRT では群ごとにアウトカム回帰モデルを推定する。この違いが ATE 推定にどのような影響を与えるのか数値実験にて様々な条件下で検討する。

3. 方法

3.1. IPW のシミュレーションの設定

3.1.1. シミュレーション手順

以下の手順でシミュレーションデータを生成する。

1. 共変量$L_1, \dots, L_{10}$の生成

   平均$\boldsymbol{0}$、分散共分散行列が$\boldsymbol{\Sigma}$である 10 変量正規分布から中間変数$X_1, \dots, X_{10}$をそれぞれn個生成する。nはサンプルサイズであり、nと$\boldsymbol{\Sigma}$の設定は 3.1.2 で述べる。そのあと、以下の式で表すように$L_1, \dots, L_5$は中央化して二値化し、$L_6, \dots, L_{10}$はそのまま連続量として扱う。

$$L_i = \begin{cases} \mathbb{I}\{X_i > 0\} - 0.5 & (i = 1, \dots, 5) \\ X_i & (i = 6, \dots, 10) \end{cases}$$

2. 治療Aの生成

   以下の式で表される処理生成モデルによって治療Aを生成する。

$$P(A = 1) = \exp\{\alpha_0 + \eta_A(\boldsymbol{L})\}/[1 + \exp\{\alpha_0 + \eta_A(\boldsymbol{L})\}]$$
$$A \sim \mathrm{Ber}\{P(A = 1)\}$$

   ここで、α_0は$P(A = 1)$の割合を調整する定数であり、$\eta_A(\boldsymbol{L})$は治療割付に対する共変量$\boldsymbol{L}$の影響を表す線形予測子である。$E[L_i] = 0$より$E[\eta_A(\boldsymbol{L})] = 0$であるので、$\alpha_0$を調整することで治療割合$P(A = 1)$の期待値を調整することができる。一方、共変量の影響$\eta_A(\boldsymbol{L})$により個々の症例割付に偏りが生じる。α_0と$\eta_A(\boldsymbol{L})$の設定は 3.1.2 で述べる。

3. アウトカムYの生成

   以下の式で表されるアウトカム生成モデルによってアウトカムYを生成する。

$$Y = \tau A + \boldsymbol{L\beta} + \varepsilon, \qquad \varepsilon \sim N(0,1),$$
$$\boldsymbol{\beta} = (0.6,\ -0.5,\ 0,\ 0.4,\ -0.3,\ 0.4,\ 0,\ 0.2,\ 0.1,\ 0)^{\mathrm{T}},$$

   Aの回帰係数τの設定は 3.1.2 で述べる。

4. 各 IPW 法の実行

   手順 1.~3.で生成したデータセットに対して CAUSALTRT を用いて IPW、IPWR、IPWS を実行

しATEの推定値$\hat{\theta}_r$ $(r = 1, \ldots, R,\ R$はシミュレーション回数)を得る。各methodに対してpsmodelステートメントはpsmodel A(event='1') = L1-L10;と指定する。

3.1.2. シナリオの設定

本シミュレーションでは、IPW、IPWR、IPWS の ATE の推定精度がサンプルサイズ、共変量間の共分散、治療割り付けの期待値、共変量による治療割り付けの偏りの強さ、アウトカム生成モデルにおける A の係数τによってどう変わるかを系統的に比較するため、以下の表 2 に示した項目を組み合わせた 120 通りのシナリオを設定する。各シナリオに対してシミュレーション回数$R = 3000$とする。

表 2：IPW のシミュレーションのシナリオ設定

項目	設定
サンプルサイズn（2 通り）	$n = 250, 1000$
共変量の分散共分散行列$\boldsymbol{\Sigma}$（2 通り）	$V[L_i] = 1, Cov[L_i, L_j] = 0$ $V[L_i] = 1, Cov[L_i, L_j] = 0.25$
$P(A = 1)$の期待値を調整する定数α_0（5 通り）	$P(A = 1) = 0.1:\ \alpha_0 = -2.197$ $P(A = 1) = 0.3:\ \alpha_0 = -0.847$ $P(A = 1) = 0.5:\ \alpha_0 = 0$ $P(A = 1) = 0.7:\ \alpha_0 = 0.847$ $P(A = 1) = 0.9:\ \alpha_0 = 2.197$ $\eta_A(\boldsymbol{L}) = 0$のときに上記の$P(A = 1)$を満たす$\alpha_0$を設定する
処理生成モデルにおける治療割付の偏りの強さ$\eta_A(\boldsymbol{L})$（3 通り）	偏りなし：$\eta_A(\boldsymbol{L}) = 0$ 偏りあり：$\eta_A(\boldsymbol{L}) = 0.5L_1 - 0.5L_2 + 0.4L_3 + 0.3L_6 - 0.3L_7$ 偏り強い：$\eta_A(\boldsymbol{L}) = 1.5L_1 - 1.5L_2 + 1.2L_3 + 1.0L_6 - 1.0L_7$
Aの回帰係数τ（2 通り）	$\tau = 0, 0.5$

3.1.1 節の手順 2.3.における処理生成モデル、アウトカム生成モデルに含まれる共変量の対応を表 3 に示す。

表 3：処理生成モデル、アウトカム生成モデルに含まれる共変量の対応

処理生成モデル	アウトカム生成モデル	共変量
含む	含む	L_1, L_2, L_6
含む	含まない	L_3, L_7
含まない	含む	L_4, L_5, L_8, L_9
含まない	含まない	L_{10}

表 3 において処理生成モデルとアウトカム生成モデルの両方に影響を与えるL_1, L_2, L_6が交絡因子である。

3.1.3. 評価指標

各シナリオに対して、3.1.1 節の手順 3.のアウトカム生成モデルにおける群間差τ を真の ATE とする。各 IPW 手法の ATE の推定精度を比較するため表 4 に示す評価指標を用いる。

表 4：各 IPW 手法の ATE の推定精度を比較するための評価指標一覧

評価指標	式
バイアス	$\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_r-\tau\right)$
RMSE	$\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}_r-\tau\right)^2}$
95% CI Coverage	$\frac{1}{R}\sum_{r=1}^{R}\mathbb{I}\{L_r\le\tau\le U_r\}$ L_rは$\hat{\theta}_r$の 95%信頼下限、U_rは$\hat{\theta}_r$の 95%信頼上限
Empirical Power Rate	$\frac{1}{R}\sum_{r=1}^{R}\mathbb{I}\{L_r>0 \text{ or } U_r<0\}$

Empirical Power Rate は$\tau=0$のとき第一種の過誤確率、$\tau=0.5$のとき検出力として解釈できる。

3.2. 標準化のシミュレーションの設定

3.2.1. シミュレーション手順

基本的に 3.1.1 と同じ手順で以下のようにシミュレーションデータを生成する。

1. 共変量$L_1,\dots,L_{10}$の生成

   3.1.1 の手順 1.と同様の方法で共変量$L_1,\dots,L_{10}$を生成する。nと$\boldsymbol{\Sigma}$の設定は 3.2.2 で述べる。

2. 治療Aの生成

   以下の式で表される処理生成モデルによって治療Aを生成する。

$$P(A=1)=\exp\{\alpha_0+\eta_A(\boldsymbol{L})\}/[1+\exp\{\alpha_0+\eta_A(\boldsymbol{L})\}]$$
$$A\sim\mathrm{Ber}\{P(A=1)\}$$

   ここで、$\eta_A(\boldsymbol{L})=0.5L_1-0.5L_2+0.4L_3+0.3L_6-0.3L_7$である。$\alpha_0$の設定は 3.2.2 で述べる。

3. アウトカムYの生成

   以下の式で表されるアウトカム生成モデルによってアウトカムYを生成する。

$$Y=\tau A+\boldsymbol{L\beta}+A\boldsymbol{L\gamma}+\varepsilon,\qquad \varepsilon\sim N(0,1),$$
$$\boldsymbol{\beta}=(0.6,\ -0.5,\ 0,\ 0.4,\ -0.3,\ 0.4,\ 0,\ 0.2,\ 0.1,\ 0)^{\mathrm{T}}$$

   τと$\boldsymbol{\gamma}$の設定は 3.2.2 で述べる

4. 標準化の実行

   手順 1.~3.で生成したデータセットに対して以下に示す解析モデル M1~M3 を用いて標準化を実行し ATE の推定値$\hat{\theta}_r$を得る。

$$\mathrm{M1}: Y=\beta_0+A\beta_A+\boldsymbol{L\beta}$$
$$\mathrm{M2}: Y=\beta_0+A\beta_A+\boldsymbol{L\beta}+A\boldsymbol{L\gamma}$$
$$\mathrm{M3}: Y=\beta_{a0}+\boldsymbol{L\beta}_a$$

   M1 と M2 は PROC GENMOD を用いて標準化し、ATE の推定値は ESTIMATE ステートメントにおいて、M1 では estimate “ATE” A 1、M2 では estimate “ATE” A 1 A*L_1 $\bar{L}_{r1}$ … A * L_{10} $\bar{L}_{r10}$ とすることで得る。一方、M3 は CAUSALTRT の model ステートメントを model Y = L1-L10 / dist=normal;として標準化する。M1 と M2 は 1 つのモデルで推定するのに対し M3 は CAUSALTRT の仕様上、群ごとにモデルを推定する。

3.2.2. シナリオの設定

本シミュレーションでは、各解析モデルの ATE の推定精度がサンプルサイズ、共変量間の共分散、治療割り付けの期待値、アウトカム生成モデルにおけるAと$\boldsymbol{L}$の積項の係数$\boldsymbol{\gamma}$ と A の係数τによってどう変わるかを系統的に比較するため、以下の表 5 に示した項目を組み合わせた 120 通りのシナリオを設定する。各シナリオに対してシミュレーション回数$R = 3000$とする。

表 5：IPW のシミュレーションのシナリオ設定

項目	設定
サンプルサイズn（2 通り）	$n = 250, 1000$
共変量の分散共分散行列$\boldsymbol{\Sigma}$（2 通り）	$V[L_i] = 1, Cov[L_i, L_j] = 0$ $V[L_i] = 1, Cov[L_i, L_j] = 0.25$
$P(A = 1)$の期待値を調整する定数α_0（5 通り）	$P(A = 1) = 0.1：\ \alpha_0 = -2.197$ $P(A = 1) = 0.3：\ \alpha_0 = -0.847$ $P(A = 1) = 0.5：\alpha_0 = 0$ $P(A = 1) = 0.7：\ \alpha_0 = 0.847$ $P(A = 1) = 0.9：\ \alpha_0 = 2.197$ $\eta_A(\boldsymbol{L}) = 0$のときに上記の$P(A = 1)$を満たす$\alpha_0$を設定する
アウトカム生成モデルにおける Aと$\boldsymbol{L}$の積項の係数$\boldsymbol{\gamma}$（3 通り）	積項なし：$\boldsymbol{\gamma} = \boldsymbol{0}$ 積項あり：$\boldsymbol{\gamma} = (0.4,\ 0,\ 0,\ 0,\ 0,\ -0.5,\ 0,\ 0.3,\ 0,\ 0)^T$ 積項強い：$\boldsymbol{\gamma} = (2.0,\ 0,\ 0,\ 0,\ 0,\ -2.5,\ 0,\ 1.5,\ 0,\ 0)^T$
Aの回帰係数τ（2 通り）	$\tau = 0, 0.5$

また、3.2.1 節の手順 2.3.における処理生成モデル、アウトカム生成モデルに含まれる共変量の対応は 3.1.2 節の表 3 と同じである。

3.2.3. 評価指標

各シナリオに対して手順 4.で実行した各方法の ATE の推定精度を比較するため表 4 の評価指標を用いる。Empirical Power Rate は$\tau = 0$のときは第一種の過誤確率、$\tau = 0.5$のときは検出力として解釈できる。

4. 結果

4.1. IPW のシミュレーションの結果

紙面の都合上、方法間の相違が比較的大きかった結果のみ図示し、残りの結果は補足資料に記載した。

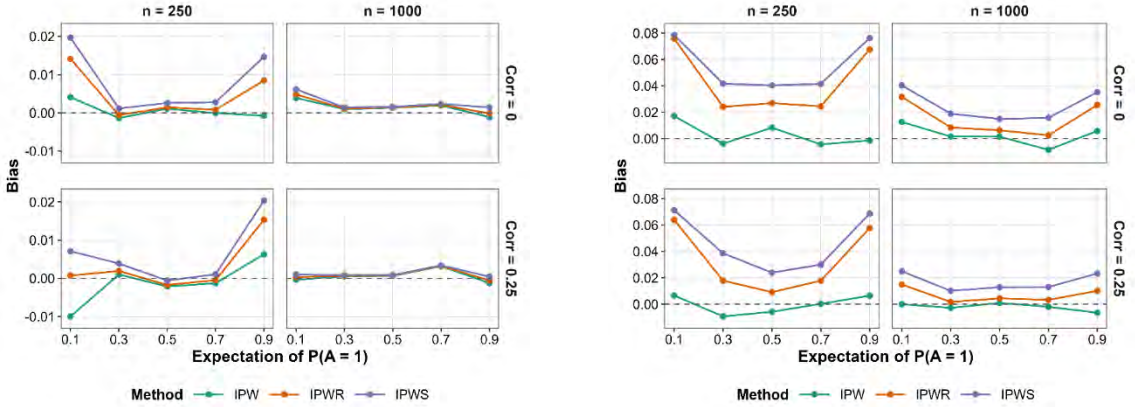


図 2：$\tau = 0$で処理生成モデルにおける治療割付の偏りがある（左）と偏りが強い（右）ときのバイアス

図 2 に示すように処理生成モデルにおける治療割付の偏りが強くなるほど、IPWR、IPWS のバイアスが大きくなった。このバイアスはサンプルサイズが大きくなるほど、治療割付の期待値が均等になるほど小さくなった。RMSE は処理生成モデルにおける治療割付の偏りが強いときのみ IPW の RMSE が大きくなったが、そうでないときは方法間で差はなかった。95% CI Coverage は処理生成モデルにおける治療割付の偏りが強いときのみ IPWR、IPWS の Coverage が 95%を大きく下回った。そうでないときは治療割付の期待値が 0.1, 0.9 のとき以外は方法間で差はなかった。

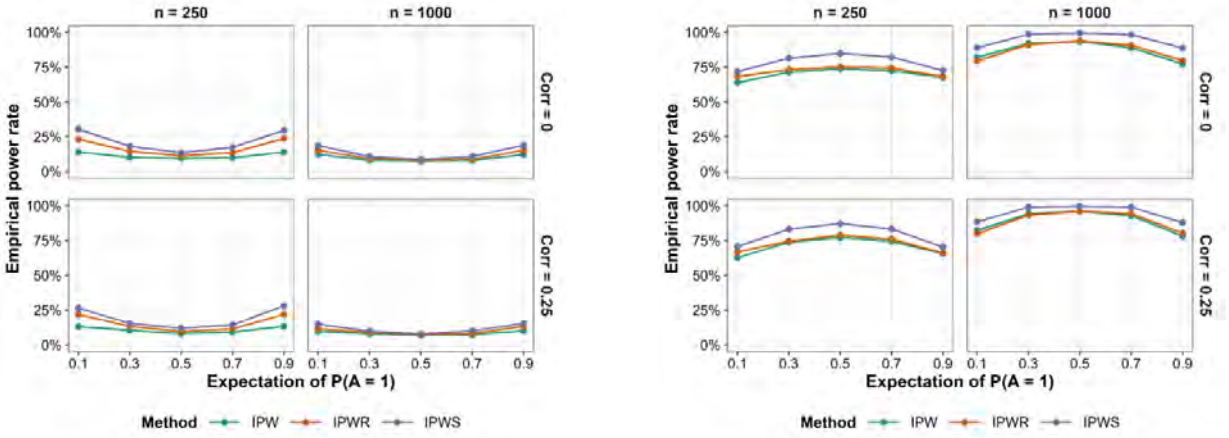


図 3：処理生成モデルにおける治療割付の偏りが強いときにおける$\tau = 0$（左）と$\tau = 0.5$（右）の Empirical Power Rate

図 3 に示すように治療割付の偏りが強いとき、$\tau = 0$（図 3 の左図）から分かるように IPWR、IPWS の第一種の過誤確率が大きくなった。サンプルサイズが大きいと IPW の方法間の差は小さくなった。$\tau = 0.5$（図 3 の右図）から分かるように IPWS の検出力が大きくなった。治療割付の偏りがない場合とある場合で方法間の差はほとんどなかった。

4.2. 標準化のシミュレーションの結果

紙面の都合上、解析モデル間の相違が比較的大きかった結果のみ図示し、残りの結果は補足資料に記載した。アウトカム生成モデルに$\boldsymbol{A}$と$\boldsymbol{L}$の積項がある場合、解析モデル M1 のバイアスは大きくなった。また、治療割合の期待値が 0.5 から乖離するほど M1 のバイアスが大きくなった。M2 と M3 はいずれの設定においてもバイアスがほとんどなかった。積項がない場合はどの解析モデルもバイアスはほとんどなかった。RMSE はいずれの設定でも M2 と M3 の値がほとんど同値だった。

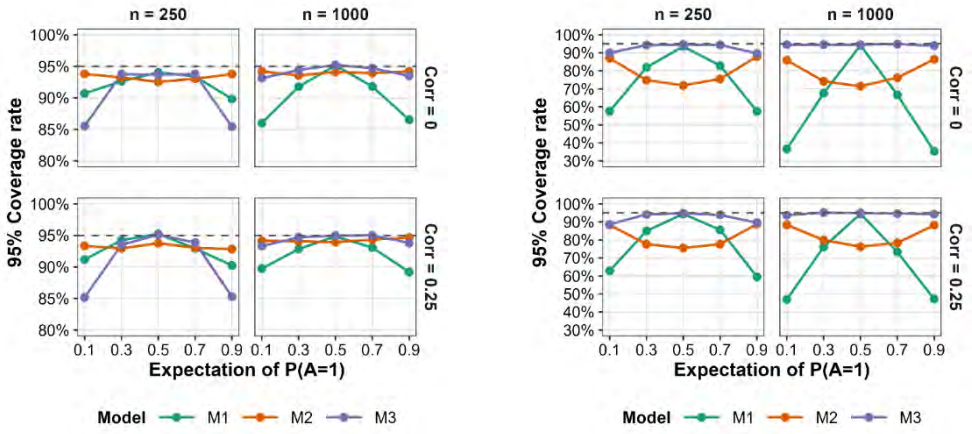


図 4：$\tau = 0$でアウトカム生成モデルにおける積項がある（左）と積項が強い（右）ときの 95% CI Coverage

95% CI Coverage は図 4 に示したように M2 は積項の影響が強い場合は名目水準から乖離した。M3 は治療割合の期待値が 0.1, 0.9 のときは名目水準から乖離するものの、積項の影響が強い場合でも名目水準に近い値を示した。M1 は治療割合の期待値が 0.5 から乖離するほど名目水準から乖離した。

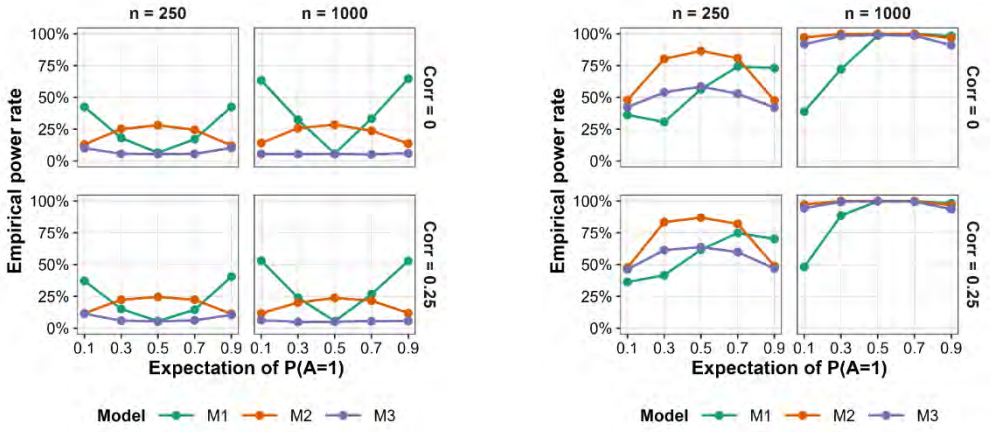


図 5：アウトカム生成モデルにおけるの積項が強いときにおける$\tau = 0$（左）と$\tau = 0.5$（右）の Empirical Power Rate

図 5 に示すようにアウトカム生成モデルの積項が強いとき、M2 の第一種の過誤確率と検出力が大きくなった。$\tau = 0.5$（図 5 の右図）でサンプルサイズが大きいとき、M2、M3 の検出力はほとんど同じだった。アウトカム生成モデルの積項が強くないとき、解析モデル間で第一種の過誤確率と検出力に差はほとんどなかった。

5. 考察

5.1. IPW のシミュレーション

処理生成モデルにおける治療割付の偏りが強い場合、IPWR と IPWS は正規化しているので RMSE は小さくなったがバイアスが生じた。これは IPWR と IPWS が比率推定量であることに起因すると考えられる。2.3.2 節において$A = \sum \frac{\mathbb{I}(A_i=a)y_i}{\hat{f}(A|L_i)}$、$B = \sum \frac{\mathbb{I}(A_i=a)}{\hat{f}(A|L_i)}$とすると H 推定量は$E\left[\frac{A}{B}\right] \approx \frac{E[A]}{E[B]} - \frac{Cov[A,B]}{E[B]^2} + \frac{E[A]Var[B]}{E[B]^3}$と一次近似できる。右辺第 2 項以降が有限標本バイアスとして現れるため、正規化を用いる IPWR と IPWS でバイアスが生じたと考えた。

処理生成モデルにおける治療割付の偏りが強い場合、IPWR と IPWS の 95% CI Coverage は名目水準である 95%より小さくなり、第一種の過誤確率と検出力が大きくなった。これは上記と同様の理由で標本バイアスが生じたことと、重みづけにより標準誤差が小さくなることに起因すると考えた。

5.2. 標準化のシミュレーション

標準化のシミュレーションでは、M1 と M2 が What if に由来する 2 群を併合した単一の解析モデル、M3 が CAUSALTRT の群ごとに推定する解析モデルであった。アウトカム生成モデルに積項が含まれる場合、積項を考慮していない解析モデル M1 は治療割合の期待値が 0.5 から乖離するほどバイアスが大きくなった。治療割合の期待値が小さいあるいは大きい条件では、片方の群のデータが少なくなるため、モデル誤特定に基づく予測誤差がより強く現れたと考えた。M2 は積項を考慮しており、M3 は群別にモデルを推定しているので、いずれの設定でもバイアスが小さくなったと考えた。それに伴って M2 と M3 の RMSE はほとんど同じ値を示したと考えた。

95% CI Coverage と Empirical Power Rate について、アウトカム生成モデルの積項の影響が強い場合、95% CI Coverage と Empirical Power Rate のトレードオフの関係が顕著であることが分かる。M3 の 95% CI Coverage は比較的、名目水準に近い値を示した。これは CAUSALTRT で用いられるロバスト・サンドイッチ分散に起因すると考えられる。CAUSALTRT の REGADJ では、群別の解析モデルのスコア関数と、潜在アウトカムの平均の推定方程式を積み上げたロバスト・サンドイッチ分散が用いられる。ロバスト分散では回帰係数の不確実性だけでなく、母集団から抽出された標本における共分散の不確実性も考慮する。つまり、母集団から抽出された各標本の患者背景の違いによる ATE の不確実性も考慮できる。そのため、母集団の ATE 推定の不確実性をより正確に評価でき、M3 では名目水準に近い 95% CI Coverage が得られたと考えた。一方、M2 の 95% CI Coverage が名目水準から乖離したのは ATE 推定の際に GENMOD の ESTIMATE ステートメントを用いたことに起因すると考えられる。ESTIMATE ステートメントで ATE を推定する際、3.2.1 節の手順 4.で示したように各共変量の標本平均を用いる必要がある。これにより母集団から抽出された標本における共変量の不確実性を考慮できなくなる。そのため、GENMOD で推定された ATE 推定値の分散は過小評価され、95% CI Coverage が名目水準より小さくなったと考えた。M2 の検出力は大きいが、ATE 推定値の分散は過小評価されているため解釈には注意が必要である。M1 はモデル誤特定による影響で 95% CI Coverage が名目水準より小さくなり、第一種の過誤が大きくなったと考えた。

6. 総括

本稿では、What if の視点で IPW および標準化の考え方を読み解くとともに、CAUSALTRT で扱うことのできる各手法とそれらの What if との対応を整理し、シミュレーションにより各手法の ATE 推定精度を比較した。

IPW のシミュレーションの結果から、共変量による治療割付の偏りが弱いと仮定できる状況では、IPW、IPWR および IPWS の間で結果が大きく異なる可能性は低いと考えられる。一方、共変量による治療割付の偏りが強く、かつサンプルサイズが小さい場合、正規化を伴う IPWR および IPWS で有限標本バイアスが大きくなった。このような場合、IPWR、IPWS を用いた ATE の推定はバイアスの観点から妥当でない可能性が高く、推定値の扱いは慎重に行うべきである。

What if で説明される標準化では、2 群のデータを併合した単一の解析モデルを用いる。共変量の値によって治療のアウトカムへの効果が異なるという効果修飾が想定される共変量について、治療とその共変量の積項を含めるといったモデル化が可能である。一方、CAUSALTRT は群ごとに解析モデルを推定する。治療と共変量の積項を事前に特定しなくてもよい利点があるが、群ごとにモデル化するため各解析モデルを推定するための群ごとのサンプルサイズが小さくなり、推定が不安定になる場合がある。標準化のシミュレーションの結果から、単一の解析モデル、群ごとの解析モデルのどちらを用いてもバイアスなく推定できることが分かった。一方、ATE 推定値の信頼区間や検出力を算出する際は ATE 推定値の分散推定の方法に注意が必要である。CAUSALTRT では共分散の不確実性も考慮して母集団の ATE を推定できるため妥当な分散を推定できるが、GENMOD の場合は共分散の不確実性を考慮できないため ATE 推定値の分散が過小評価される恐れがある。単一のモデルで ATE 推定とその信頼区間を求めたい場合は積み上げ式ロバスト分散を考慮すべきである。

本シミュレーションでは、実務において複数の交絡因子候補を同時に考慮しながら因果効果を推定する状況を模擬するため 10 個の共変量が存在する状況を設定した。解析に含める共変量について What if では、交絡調整に必要な共変量がすべて測定され解析に含まれていること、すなわち未測定交絡がないことを仮定す

る。ところが、この仮定が成立するかを観察データのみから確認することは困難である。したがって、実務では、測定された共変量のうち、治療選択およびアウトカムに共通して関係する可能性のある因子を当該領域の専門知識や因果ダイアグラム等を用いて検討し適切に解析へ含めることが重要となる。

本稿では、IPW と標準化それぞれについて整理検討したが、両手法の使い分けについては議論しなかった。モデル誤特定への対策として、What if は IPW と標準化がどちらとも可能である場合は両方実行し、さらに可能なら二重ロバスト推定量を用いるという立場をとっている。モデル誤特定という観点から IPW、標準化、二重ロバスト推定量の推定精度を比較することを今後の検討課題としたい。

参考文献

[1] Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC. 2020.

[2] SAS Institute Inc. *SAS/STAT® 14.2 User's Guide*. SAS Institute Inc. 2016.

[3] Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*. 1952; **47**: 663-685. doi:10.1080/01621459.1952.10483446.

[4] Hájek J. Comment on a paper by D. Basu. In: Godambe VP, Sprott DA (eds). *Foundations of Statistical Inference*. 1971: 236.

[5] Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*. 2004; **23**: 2937-2960. doi:10.1002/sim.1903.

ベイズ統計の漸近理論と SAS による実装

○玉乃井 崚太

（シミック株式会社　データサイエンス事業本部）

Asymptotic Theory of Bayesian Posteriors and Its Implementation in SAS

Ryota Tamanoi

Data Science Division, CMIC Co., Ltd.

要旨

近年、医療統計においてもベイズ流の解析が用いられる場面が増えている。頻度論的な統計学では、標本平均のような推定量の妥当性が大数の法則や中心極限定理によって保証されている。ベイズ流の解析にも同じ役割を果たす理論が必要であり、それは「観測データが十分に存在するとき、事後分布はどのようにふるまうか」という問いになる。素直な答えは「事後分布は正規分布に近づく」であり、最適なパラメータが一点に定まるような性質のよい正則なモデルでは実際にそのとおりになる。しかし説明変数に完全な共線性がある回帰のような非正則なモデルにおいてはこの結論は破綻する。本稿では、正則なモデルの漸近正規性とその破綻を PROC MCMC で観察したうえで、破綻した場合にも成立する漸近理論、渡辺澄夫ベイズ理論を紹介する。パラメータそのものを追うことをやめ、真の分布との KL 情報量の値だけを追うという見方をとると、正則・非正則を問わず同じ形の漸近分布が得られる。応用として、正則なモデルにおけるパラメータ数に代わってモデルの実質的な自由度を表す実対数閾値と、正則性を仮定しない情報量規準である WAIC を導入し、SAS による実装を与える。

キーワード：ベイズ統計、漸近理論、実対数閾値、WAIC、PROC MCMC、非正則モデル

1. はじめに

1.1. 背景

近年、医療統計分野においてベイズ統計の応用が活発に進展している。ベイズ統計では観測データに加え外部情報を「事前分布」として組み込むことができるほか、複雑な試験デザインにも柔軟に対応できるなど多くの利点がある。ベイズ流でない頻度論的な統計学においては、標本平均という推定量の正当性は大数の法則と中心極限定理によって保証されている。ベイズ推定にも同じ役割を果たす理論はあるだろうか。すなわち、「観測データが十分に存在するとき、事後分布はどのようにふるまうか」という問いである。

素直な答えは「事後分布は正規分布に近づく」であり、実際に多くのモデルでそのとおりになる。信用区間などのベイズ流でよく用いられる道具は、すべてこの事実の上に立っている。ところが、モデルの書き方をわずかに変えるだけでこの答えは完全に破綻する。破綻するのは「近似の精度が落ちる」という意味ではなく、近似の枠組みそのものが使えなくなるという意味である。

破綻するモデルは決して病的 (人為的に作成した反例) ではない。説明変数に完全な共線性がある回帰、混合分布、ニューラルネットワーク、行列分解はいずれもこちら側に属する[1][2]。

渡辺澄夫によって整備された特異学習理論（本稿では渡辺澄夫ベイズ理論と呼ぶ[2]）は、この破綻した世界でも成り立つ漸近論を与える。本稿の目的は、その骨格を SAS で実際に事後分布を描きながら追うことである。

1.2. 本稿の構成

2 章でベイズ推定の基礎と PROC MCMC による実装を確認する。3 章で正則なモデルの漸近正規性を定式化し、SAS で確認する。4 章で共線性のある回帰を用いてそれが破綻する様子を観察する。5 章が本稿の中心で、パラメータを追うのをやめて$K(\theta)$の値だけを追うという発想を導入し、実対数閾値λを定義して SAS で測定する。6 章では応用として WAIC を導き、事後標本から計算する実装を与える。

2. ベイズ統計の基礎と SAS による実装

2.1. 事後分布

観測データを$y_1,\cdots,y_n$とし、統計モデルを$f(Y|\theta)$、事前分布を$\varphi(\theta)$とする。ベイズ推定の出発点は事後分布である。

$$\varphi(\theta|y_{1:n}) = \frac{1}{Z_n}\varphi(\theta)\prod_{i=1}^{n} f(y_i|\theta)$$

$$Z_n = \int \varphi(\theta)\prod_{i=1}^{n} f(y_i|\theta)\,d\theta$$

例えば身長のデータ$(y_1,y_2,y_3,y_4,y_5) = (165,167,170,173,180)$を用いて平均 μ 及び分散 σ^2 を推定する場合モデルの確率密度関数は$f(Y|\mu,\sigma^2) = N(Y|\mu,\sigma^2)$、事前分布を$\varphi(\mu,\sigma^2) = N(\mu|170,100^2)\cdot HalfCauchy(\sigma|0,25)$などのようにすることができる。ここで事前分布の正規分布の平均や分散、半コーシー分布の scale はハイパーパラメータと呼ばれる。この例ではそれぞれ 170、10000、25 としており、観測データとは無関係に事前に決定される。この例では 170 が「成人の身長はこのくらい」という事前の知識を、標準偏差 100 cm が「その知識をほとんど信用しない (解析に組み入れない)」という意思を表している。

このとき事後分布はμについて事後平均 171.0、95%信用区間 [163.2, 179.7]、σについて事後平均 8.40 (右にゆがんだ分布、事後中央値は 7.25)、95% 信用区間 [3.67, 20.96] となる。

従来の解析が観測データから点推定量や信頼区間を求めるのに対し、ベイズ流の統計は観測データと事前分布を合わせて事後分布を求め、そこから点推定量や信用区間を得る。

2.2. PROC MCMC による実装

SAS では PROC MCMC で事後分布からのサンプルが得られる[3]。2.1 章の例では以下のコードになる。

```
data height;
   input subjid $ y;
   datalines;
001 165
002 167
003 170
004 173
005 180
```

```
;
run;
proc mcmc data=height seed=12345
         nbi=1000  /* MCMCサンプルの最初はバイアスが大きいため捨てる(Buin in) その回数 */
         nmc=10000  /* MCMCの回数 */
         thin=20    /* 隣接するMCMCサンプルは独立でないためthin回ごとに取得する。*/
                    /* nmc / thin 個のMCMCサンプルが作成される。 */
         outpost=post_out
         monitor=(mu sigma)
         statistics(alpha=0.05 percent=(2.5 25 50 75 97.5))=(summary interval)
         diagnostics=none plots=none;
   parms mu 170 sigma 6;                          /* MCMCサンプルの初期値 */
   beginnodata;
      lp_sigma = -log(1 + (sigma/25)**2);         /* コーシー分布 */
   endnodata;
   prior mu    ~ normal(170, var=10000);          /* muの事前分布 */
   prior sigma ~ general(lp_sigma, lower = 0); /* lower = 0によって正に限定 */
   model y ~ normal(mu, sd=sigma);                /* モデルの分布 */
   ods output   PostSummaries = PostSummaries;
run;
```

注意として、PROC MCMC が返すのは事後分布そのものではなく事後分布に従うサンプルである。OUTPOST= で指定したデータセットにそのサンプルが格納される。ヒストグラムを描けば事後分布の概形が分かり、昇順に並べて下から 2.5% から 97.5% までの範囲が 95% 信用区間になる。ODS OUTPUT PostSummaries = で数値としても取り出せる。

2.3. 信用区間と信頼区間

上記のサンプルコードでは分散の大きい事前分布を指定しているが、無情報事前分布を指定することも可能である。平均と分散をともに未知とし、Jeffreys 事前分布 (無情報事前分布としてよく用いられる) の場合、平均に対する 95% 信用区間は t 検定による 95% 信頼区間と一致する[1]。

3. 正則なモデルの漸近理論

3.1. KL 情報量と最適なパラメータ

真の分布を$f_{TRUE}(Y)$とする。これは頻度論的な解析と同様、設定したモデルなどに全く依らず決定されているものであり、我々が推定したい分布の正体である。モデル$f(Y|\theta)$が真の分布からどれだけずれているかを表す量として、Kullback-Leibler 情報量を次で定義する[4]。

$$K(\theta) = E_{Y_{TRUE}}\left[\log\frac{f_{TRUE}(Y_{TRUE})}{f(Y_{TRUE}|\theta)}\right] = \int f_{TRUE}(Y)\log\frac{f_{TRUE}(Y)}{f(Y|\theta)}dY$$

ここで$E_{Y_{TRUE}}$は真の分布に従う確率変数Y_{TRUE}による期待値を表す。この$K(\theta)$が小さいほど真の分布に近い「良い」θであるといえる。$K(\theta)$を最小にするθの全体をΘ_{BEST}とおく。このようなθを推定することがベイズ

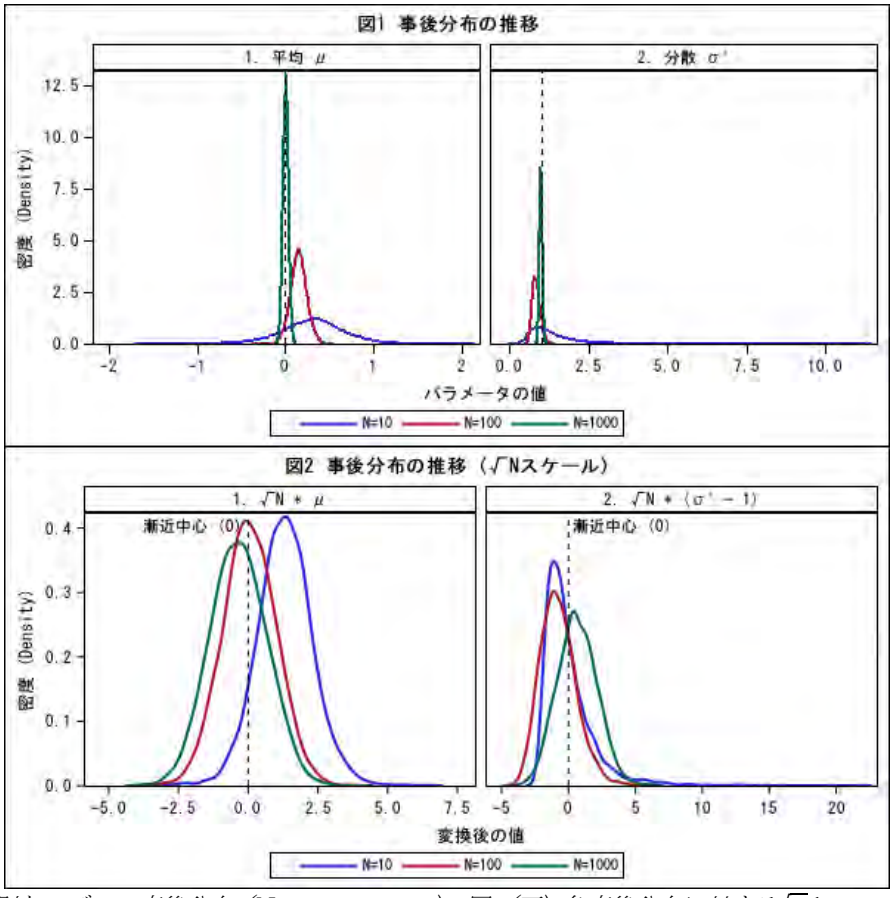


図 1(上) 正則なモデルの事後分布 (N=10, 100, 1000)。図 2(下) 各事後分布に対する$\sqrt{N}(\theta-\theta_{BEST})$の分布。

流の目標といえる。真の分布がモデルに含まれる場合、Θ_{BEST}上で$K(\theta)$は 0 になる。以下 3,4,5,6 章では簡単のため真の分布がモデルに含まれることを仮定する (実現可能性の仮定)[1]。

$$\Theta_{BEST} = \{\theta | K(\theta) \text{が最小値}\}$$
$$= \{\theta | K(\theta) = 0\}$$

多くの議論は実現可能性の仮定なしに成立するが、AIC など一部でこの仮定が必要になる。

3.2. 正則なモデルの定義

本稿では、次の 2 条件を満たすモデルを***正則なモデル***と呼ぶ。①Θ_{BEST}が 1 点$\{\theta_{BEST}\}$であり、θ_{BEST}はパラメータ空間の内点である(すなわち端の点などではない)こと。②その点で$K(\theta)$のヘッセ行列$J = \nabla^2 K(\theta)$が正定値であること。②は一見難しい数式のように見えるかもしれないが、これは単に関数$K(\theta)$が谷底θ_{BEST}のまわりで純粋な 2 次形式になり、いわば「きれいなお椀」のような形になることを意味する。すなわち次の近似がθ_{BEST}の周りで成立する。

$$K(\theta) \simeq \frac{1}{2}(\theta-\theta_{BEST})' J (\theta-\theta_{BEST})$$

この 2 条件のどちらが欠けても、3 章の以降の議論はすべて崩れる。4 章では①が破れる例、①が成り立っているが②が破れる例をそれぞれ紹介する。

3.3. 正則なモデルの例と漸近正規性

例として平均分散未知の正規分布を扱う$\theta = (\mu, \sigma^2)$, $f(Y|\mu,\sigma^2) = N(Y|\mu,\sigma^2)$。事前分布は Jeffreys 事前分布とする。また真の分布を平均 0 分散 1 の正規分布とする $f_{TRUE}(Y) = N(Y|0,1^2)$。最適なパラメータは$(\mu_{BEST}, \sigma^2_{BEST}) = (0,1^2)$であり、

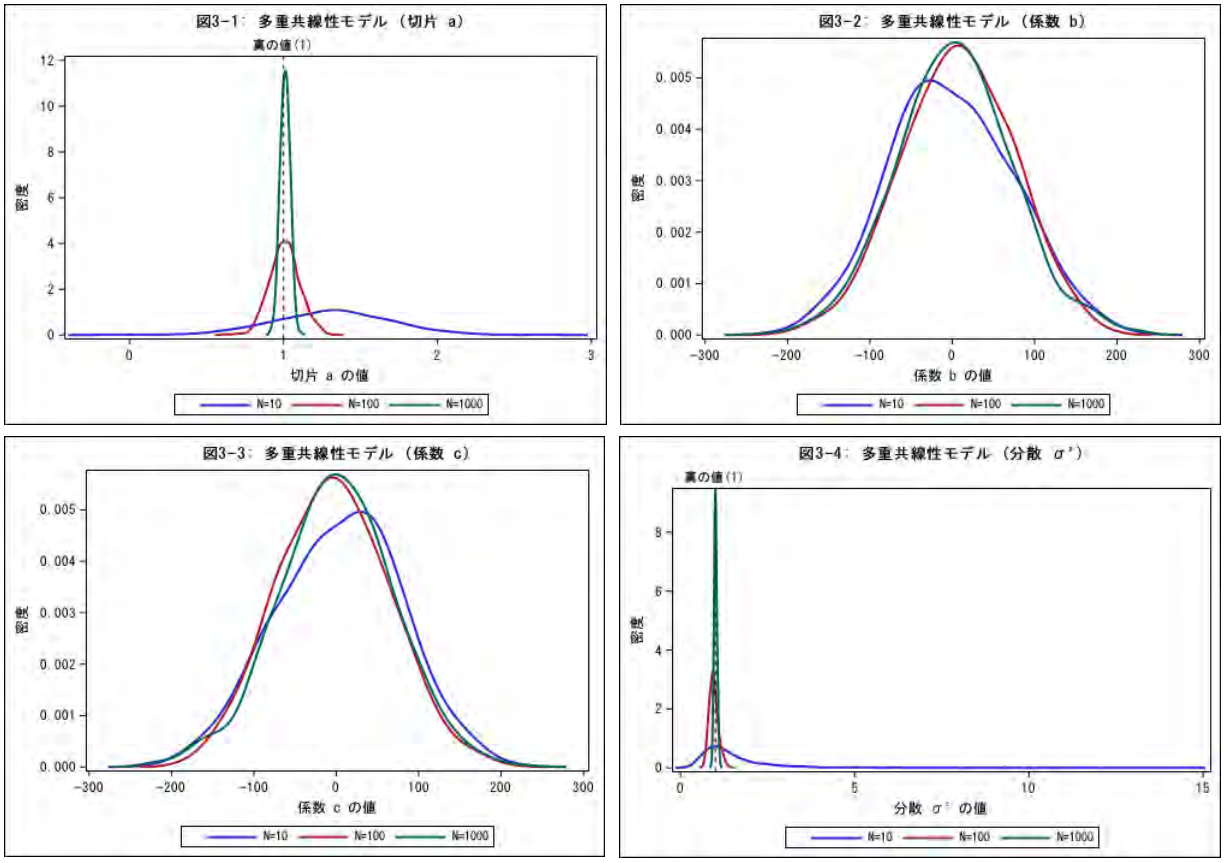


図 3-1, 図 3-2, 図 3-3, 図 3-4: 共線性を持つモデルの事後分布。a, σ^2 は観測データを増やすごとに分散が小さくなることがわかるが、b, c は観測データの個数の影響がない。

$$K(\mu, \sigma^2) = \frac{1}{2}\left(\log\sigma^2 + \frac{(1+\mu^2)}{\sigma^2} - 1\right)$$

$$J = \begin{pmatrix} 1/\sigma^2 & -\mu/\sigma^4 \\ -\mu/\sigma^4 & -\frac{1}{2\sigma^4} + \frac{1+\mu^2}{\sigma^6} \end{pmatrix}$$

になる。このとき$n = 10,100,1000$で事後分布は図 1 のようになる。例数が大きくなるにつれて真の平均、分散に収束していることがわかる。さらに収束の仕方の詳細を追うために、事後分布のθ_{BEST}からの離れ方を$\sqrt{n}$倍した確率変数を考える。

$$\sqrt{n}(\theta - \theta_{BEST})$$

$N = 10,100,1000$に対してこの確率変数は図 2 のようになる。$\sqrt{n}$倍することで一点に収束せず、正規分布に分布収束する（Bernstein–von Mises の定理)。

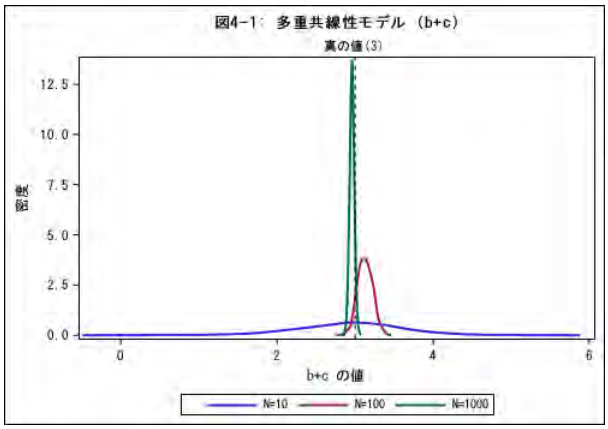
3.4. 事後分布の漸近正規性

3.3 章で見たように、正則なモデルでは、次の意味で事後分布は正規分布に近づく。

$$\sqrt{n}(\theta - \theta_{BEST} - \Delta_n) \to N(\theta|0, J^{-1})$$

ここでΔ_nはデータのゆらぎを表すベクトルで、経験過程η_nを用いて次で与えられるが本稿では深く立ち入らない。

$$\Delta_n = \frac{J^{-1}\nabla\eta_n(\theta_{BEST})}{\sqrt{n}}$$



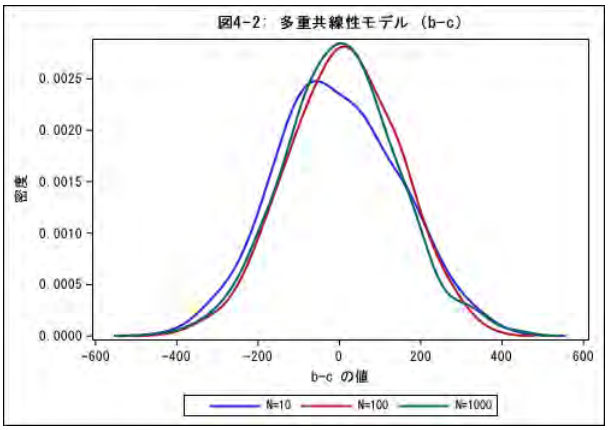


図 4-1,図 4-2: $(b+c)$の事後分布及び$(b-c)$の事後分布。$b+c$は観測データを増やすことで真の値 3 に収束する。

$$\eta_n(\theta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{E_{Y_{TRUE}}\left[\log\frac{f(Y_{TRUE}|\theta_{BEST})}{f(Y_{TRUE}|\theta)}\right]-\log\frac{f(y_i|\theta_{BEST})}{f(y_i|\theta)}\right\}$$

重要な点は、Δ_n自体は 0 に収束するが、$\sqrt{n}\Delta_n$は 0 に収束しないことである。すなわち$\sqrt{n}(\theta-\theta_{BEST})$のバイアスである$\sqrt{n}\Delta_n$は消えず、$N(0,J^{-1})$に分布収束しない。

4. 非正則なモデルの例

3.2 章で挙げた 2 条件のうち、4.1 章では①が破れる例を、4.2 章では①は成り立つが②が破れる例を見る。

4.1. 共線性のある回帰

説明変数に完全な共線性がある回帰を考える。

$$Y=a+bx_1+cx_2+\epsilon$$
$$\epsilon\sim N(0,\sigma^2)$$

とし、真の分布を$x_1=x_2\sim N(0,1^2)$かつ$Y=1+3x_1+\epsilon$、$\epsilon\sim N(0,1)$とする。この設定は極端に見えるが、共線性の強い共変量を扱う場面の単純化である。このとき KL 情報量は次のようになる。

$$K(a,b,c,\sigma^2)=\frac{1}{2}\left\{\log\sigma^2-1+\frac{1+(a-1)^2+(b+c-3)^2}{\sigma^2}\right\}$$

よって$\Theta_{BEST}=\{(a,b,c,\sigma^2)|a=1,b+c=3,\sigma^2=1\}$となり一点ではなく直線となる。3 章の条件①が破れている。このとき$N=10,100,1000$の事後分布は図 3 のようになる。b,cについて一点に収束しないことがわかる。原因は明らかで、このモデルではb,c個別の情報はなく、$b+c$について収束するのみである(図 4)。特に、情報の全くない$b-c$の分布は観測データによって変化しない。

4.2. 二乗平均のモデル

3 章の条件②が破れる例として次のモデルを考える

$$f(Y|\theta)=N(Y|\theta^2,1^2)$$

真の分布を$f_{TRUE}(Y)=N(Y|0,1^2)$とする。このとき

$$K(\theta)=\theta^4/2$$

であり、最適なパラメータは$\theta_{BEST}=0$の一点のみである。この一点におけるヘッセ行列は

$$J=6\theta_{BEST}^2=0$$

であり、これは②を満たさない。このモデルにおいて事後分布は図 5 のようになる。N を増やしても正規分

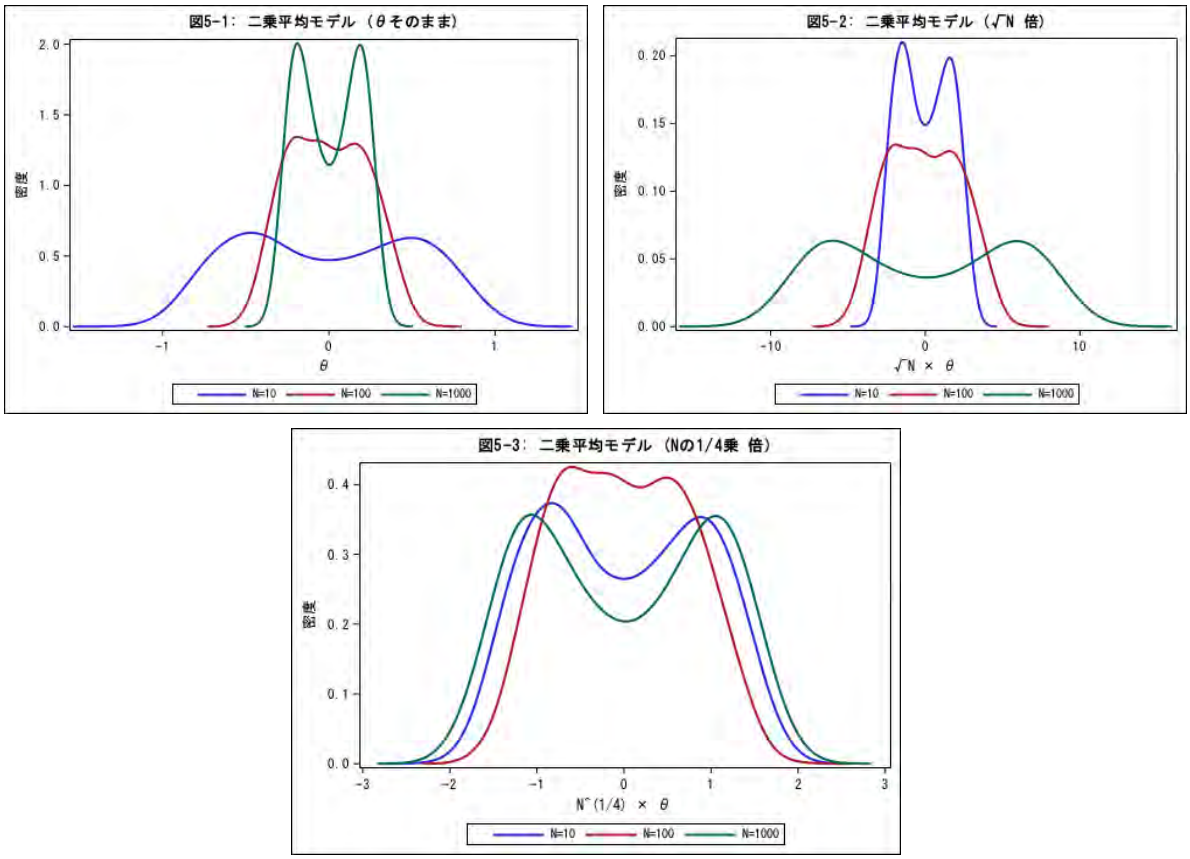


図 5-1, 図 5-2, 図 5-3:スケールを調整した二乗平均のモデルの事後分布。元の分布では観測データを増やすごとに分散が小さくなり、$\sqrt{N}$倍のスケールでは逆に大きくなる。$N^{1/4}$倍のスケールでは分散も 0 に収束もしない

布にはならず、また 3 章の例とは異なり$\sqrt{n}$倍すると発散していく。これは最適な点への収束が正則なモデルより遅いことを意味する（$\sqrt{n}$に速度で負けてしまう）。実際のスケールである$n^{1/4}$倍をかけると発散も一点への収束もしていないことが観察できる。しかし正規分布にはならない。

さらに図 5 では事後分布が二峰になっているがこれはデータに依存し、50%の確率で二峰に、50%の確率で単峰の分布となる。実際$u = n^{1/4}\theta$と置きなおすとuの事後分布は

$$\varphi(u|y_1,\cdots,y_n) \propto \exp(G \cdot u^2 - u^4/2)$$
$$G = \sqrt{n}\bar{y}$$

となり、観測データの平均$\bar{y}$が 0 より大きい場合は二峰、小さい場合は単峰である。正則なモデルでは観測データのゆらぎΔ_nによって分布の平均が変化していたが、このモデルでは分布の形も観測データによるゆらぎの影響をうけることがわかる。

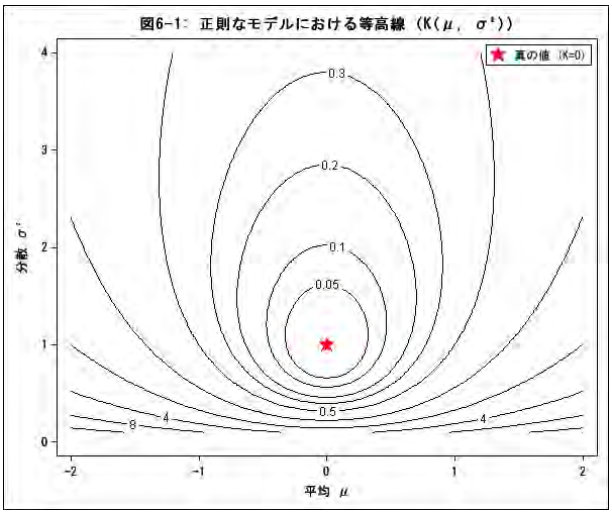
5. 非正則なモデルの漸近理論

5.1. 発想

4 章で見たように、非正則なモデルでは「Θ_{BEST}を中心にして$\theta - \theta_{BEST}$を測る」という測り方そのものが成立しない。Θ_{BEST}が 1 点でなければ中心が定まらず、Kのヘッセ行列が退化していれば$\sqrt{n}$という尺度も合わない。そこで測り方を変える。

θ がどこにあるかを追うのをやめて、K(θ) の値を追う。

θは d 次元のベクトルだが、$K(\theta)$は非負の実数である。d 次元の情報を 1 次元に潰す乱暴な操作に見えるが、



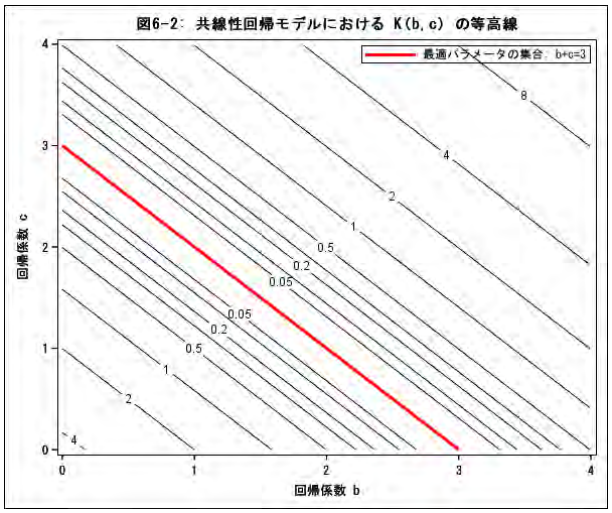
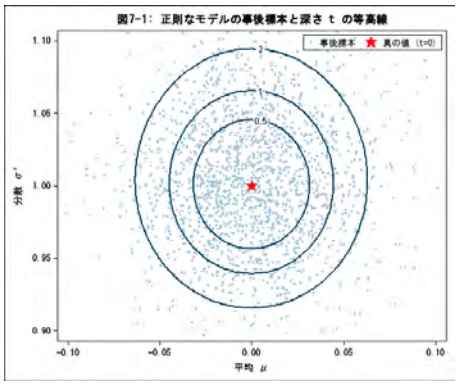
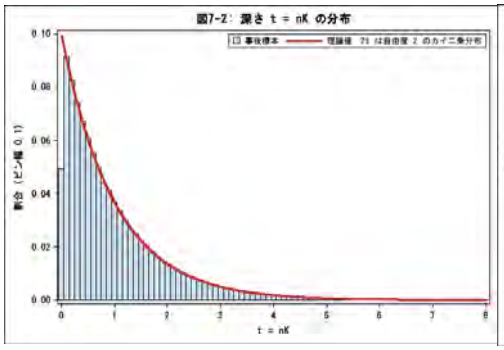


図 6-1:正則なモデルにおける$K(\mu,\sigma^2)$の値。図 6-2:共線性のある回帰モデルにおける$K(a,b,c,\sigma^2)$。ただし$a=1,\sigma^2=1$としてb,cに対してのみ図示している。





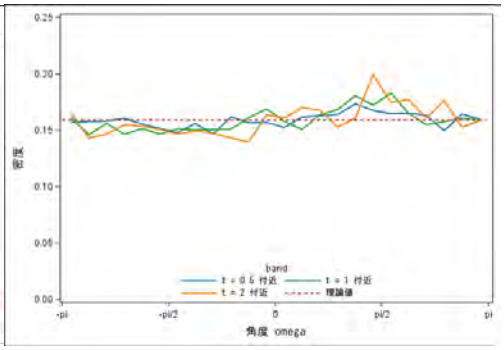


図 7-1: 事後分布の MCMC サンプルに t=0.5,1,2 の線をを重ねた図。(n=1000) 図 7-2: t で周辺化した分布。ヒストグラムは MCMC サンプルから作成され、赤線が理論値。図 7-3: t=0.5, 1, 2 の等高線の付近のωの分布。

$K(\theta)$はΘ_{BEST}が 1 点でなくとも定義でき、かつ真の分布との近さという、いちばん知りたい情報を持っている。

$\boldsymbol{t=nK(\theta)}$と置く。この$t$でパラメータ空間にいわば「等高線」を引き、各等高線の間と、各等高線内にわけて考える。図 6 は 3.3 章の正則モデルと 4.1 章のモデルである。ただし 4.1 章のモデルは簡単のため b,c についてのみ図示している。形はまったく違うが、「高さtでラベルを付ける」という操作自体はどちらでも同じようにできる。

一般論に入る前に、例として正則なモデルの場合のtの分布と各等高線内の分布を考える (図 7)。$K(\theta)\simeq$

$\frac{1}{2}(\theta-\theta_{BEST})^{\mathsf{T}}J(\theta-\theta_{BEST})$であったから

$$2t=\left(\sqrt{n}(\theta-\theta_{BEST})\right)^{\mathsf{T}}J\left(\sqrt{n}(\theta-\theta_{BEST})\right)$$

となる。3 章でみたように$\sqrt{n}(\theta-\theta_{BEST})$は観測データのブレを除いて正規分布であったから、**$2t$はおよそ自由度 d のカイ二乗分布になる。**($\varphi(t)\propto t^{\frac{d}{2}-1}\exp(-t)$)。また、

$$x=\mu,\ y=\frac{\sigma^2-1}{\sqrt{2}},\ x=r\cos\omega,\ y=r\sin\omega$$

とすると、ωの分布は観測データのブレを除いて一様分布となる。

5.2. 結論の形

前節で見たものは、実は正則なモデルに限った話ではない。同じ形が非正則なモデルでもそのまま成り立つ。先に答えを書く。θを「どの等高線か」を表すtと「その等高線の中のどこか」を表すuに分けて書き直すと、事後分布は次の形になる[1]。

$$\varphi(t,u|y_1,\cdots,y_n)\propto t^{\lambda-1}\exp\left(-t+\sqrt{t}\xi_n(u)\right)\cdot\varphi(u|t)$$

ここで各記号は次のとおりである。

$$t=nK(\theta)$$

$$\xi_n(u)=\frac{t-\sum_{i=1}^{n}\log\frac{f_{TRUE}(y_i)}{f(y_i|\theta)}}{\sqrt{t}}$$

$\xi_n(u)$はデータのゆらぎを表す量であり、$\varphi(u|t)$は各等高線に制限した時のuの事前分布である。

本稿ではこの式の導出は行なわず、この式を認めたうえでλなどについて考える。

5.3 章でこの漸近モデルを前提に、3.3 章、4.1 章、4.2 章の例に対して実際にλを天下り的に計算する。このときλの計算に当たっては$t\to 0$における挙動を考えればよい。$t\to 0$では$\exp\left(-t+\sqrt{t}\xi_n(u)\right)$は 1 に収束するため、原点における特異性はだけが担っており、λはこの因子から読み取れる。

5.4 章では一般のモデルにおいてλを事前分布から計算する方法を考える。

5.3. 計算例

前節の形が実際にどのように出てくるかを、3.3 章、4.1 章、4.2 章の例で確かめる。

5.3.1. 正則なモデル（3.3 章の例）

5.1 章で図 7 として見た「2t が自由度 d のカイ二乗分布になる」「各等高線の中で角度ω が一様になる」ことを、ここで導出する。真の分布を$f_{TRUE}(Y)=N(Y|0,1)$、モデルを$f(Y|\mu,\sigma^2)$とすると最適なパラメータは$(\mu_{BEST},\sigma^2_{BEST})=(0,1)$であった。3.3 章で記載したヘッセ行列 J を用いると、最適なパラメータの周りで$K(\mu,\sigma^2)$は次の二次形式で近似できる。

$$K(\mu,\sigma^2)\simeq\frac{1}{2}\mu^2+\frac{1}{4}(\sigma^2-1)^2$$

ここで高さsの等高線$K(\mu,\sigma^2)=s$を考えると、楕円

$$\mu^2+\frac{1}{2}(\sigma^2-1)^2=2s$$

で近似できることがわかる。この楕円を円にするために縦軸をスケーリングし、$x=\mu$、$y=\frac{1}{\sqrt{2}}(\sigma^2-1)$と座標変換する。すると$K(x,y)\simeq\frac{1}{2}(x^2+y^2)$となるため、中心からの距離$r=\sqrt{2s}$を用いて極座標

$$x=r\cos\omega,\ \ y=r\sin\omega$$

とできる。ここでsが「どの等高線か」を、角度ωが「その等高線の中のどこか」を表しており、これがまさに 5.2 章の(t,u)にほかならない。正則なモデルにおいて(t,u)は極座標表示によって実現できる。

このx,y,r,ωを用いてλを導出する。上述の変換によって(x,y)は独立な正規分布になる。

$$x,y\sim N(\Delta_n',1/n)$$

Δ_n'は観測データの揺らぎΔ_nを変換したものである。このときよく知られているように、$n(x^2+y^2)=nr^2=2ns=2nK(\theta)=2t$の分布は自由度 2 の非心カイ二乗分布となる。$t\to 0$においてこれは自由度 2 のカイ二乗分布で近似することができ、事後分布$\varphi(\theta|y_1,\cdots,y_n)$を$t$で周辺化すると

$$\varphi(t|y_1,\cdots,y_n)\propto t^{\frac{2}{2}-1}\exp(-t)$$

となる。したがって$\lambda=1$である。

以上の議論はパラメータが d 個ある正則なモデルにおいてほとんどそのまま成立する。θはd個の独立な正規分布$N(0,1/n)$に分解され、$2t$は自由度dのカイ二乗分布となる。したがって

$$\boldsymbol{\lambda=d/2}$$

となる。次の例から観察できるように、非正則なモデルにおいてλは$d/2$にはならない。実質的にモデルに関与しないパラメータが存在するためである。**2λは「実質的なモデルの自由度」**と考えることもできる。

5.3.2. 非正則なモデル (4.1 章の例)

説明変数$x_1=x_2$に対して真のモデルを$f_{TRUE}(Y)=N(Y|1+3x_1,1^2)$、モデルを$f(Y|a,b,c,\sigma^2)=N(Y|a+bx_1+cx_2,\sigma^2)$とすると最適なパラメータは直線$\Theta_{BEST}=\{(a,b,c,\sigma^2)|a=1,b+c=3,\sigma^2=1\}$であった。

$\alpha=a-1$、$\beta=b+c-3$、$\epsilon=\sigma^2-1$とおくとΘ_{BEST}の周りで$K(a,b,c,\sigma^2)$は次の二次形式で近似できる。

$$K(a,b,c,\sigma^2)\simeq\frac{1}{2}\left(\alpha^2+\beta^2+\frac{\epsilon^2}{2}\right)$$

このとき、$\gamma=b-c$はこの式に現れない。つまりこのモデルは、パラメータを 4 個持ちながら実質的には 3 個しか使っていない。

言い換えるとヘッセ行列 J はγの方向に 0 の固有値をもち、$\mathrm{rank}\,J=3$である。5.4 章で見るように$K(\theta)$が二次形式で近似できる場合$\lambda=\mathrm{rank}(J/2)$であるから、$\lambda=3/2$を得る。$d/2=4/2$にはならない。

5.3.3. 非正則なモデル (4.2 章の例)

最後に 4.2 章の$f(Y|\theta)=N(Y|\theta^2,1^2)$の例を見る。$K(\theta)=\theta^4/2$であった。

このモデルでは 5.3.1 章, 5.3.2 章と異なり$K(\theta)$は二次形式にならない。直接計算することでλが計算できる。

4.2 で紹介したように、$u=n^{1/4}\theta$と置きなおすとuの事後分布は

$$\varphi(u|y_1,\cdots,y_n)\propto\exp(G\cdot u^2-u^4/2)$$

である。$t=nK(\theta)=u^4/2$であるから$u=(2t)^{1/4}$であり、

$$\varphi(t)\propto t^{-\frac{3}{4}}\exp\left(-t+\sqrt{2}G\sqrt{t}\right)$$

となる。$t^{\lambda-1}\exp\left(-t+\sqrt{t}\xi_n(u)\right)$と比較すると、

$$\lambda=\frac{1}{4}$$

である。特筆すべき点として、このモデルは$2\lambda=1/2$であり整数値にはならない。λはパラメータの個数を数

えるだけでは得られず、一般の定義が必要になる。

5.4. 実対数閾値λ

本節で一般のλの定義を述べる。5.3 章の 3 つの例で共通して使ったのは、$K(\theta)$の形そのものではなく「高さsの等高線」に事前分布の重みがどれだけ乗っているかだけであった。これを**状態密度**と呼び、$v(s)$と書く。

$$v(s) = \int \delta\big(s - K(\theta)\big)\varphi(\theta)d\theta = \int_{s=K(\theta)} \frac{\varphi(\theta)}{|\nabla K(\theta)|} d\theta$$

ここでδはディラックのδ関数である。平たく言えば「$s = K(\theta)$のθにおいて事前分布$\varphi(\theta)$を積分する」ことを意味する。この$v(s)$の原点における振る舞い（極）によってλが定義できる

$$v(s) = v_0 s^{\lambda-1} \qquad (s \to 0)$$

このλを**実対数閾値**（real log canonical threshold）と呼ぶ。なお一般には、この式に重複度と呼ばれる自然数で定まる対数の因子が付く。本稿で扱う 3 つのモデルはいずれも重複度が 1 であり、その場合は上の形のままでよい。定義から、$K(\theta)$が二次形式で近似できる場合、

$$\lambda = \frac{\operatorname{rank} J}{2}$$

特に正則なモデルにおいては$\lambda = d/2$であり、一般に$\lambda \le d/2$である[5]。ただし事前分布がΘ_{BEST}において 0 になる場合$\lambda > d/2$となりうる。

5.5. 実対数閾値λの推定量と実装

以上の議論ではすべて$K(\theta)$を用いており、これは真の分布$f_{TRUE}(Y)$がわかっている状況でしか厳密なλは導出できない。現実的には与えられた観測データからλに収束する何らかの統計量を作ることでλを推定することになる。詳細はここでは触れないが、以下の理論によってλの推定量を作ることができる[6]。

まず「逆温度βの事後分布」を以下で定義する

$$\varphi_\beta(\theta|y_1,\cdots,y_n) \propto \varphi(\theta) \times \prod_{i=1}^{n} f(y_i|\theta)^\beta$$

各観測データの尤度$f(y_i|\theta)$をβ乗した形であり、これは「各観測データがβ個得られた」と考えていることと等しい。$\beta = 2$であればすべての観測データを二回ずつかけており、$\beta = 0.5$であれば各観測データを 0.5 個分で計算している。特に$\beta = 0$の場合明らかに$\varphi_\beta(\theta|y_1,\cdots,y_n) = \varphi(\theta)$である。

このとき、次が成り立つ

$$E^\beta\left[\sum_{i=1}^{n} \log f(y_i|\theta)\right] \simeq C_n - \frac{\lambda}{\beta}$$

すなわち、$\varphi_\beta(\theta|y_1,\cdots,y_n)$で対数尤度の期待値をとり、それを$1/\beta$に対してプロットすると、傾きが$-\lambda$になる。

実際に本稿の 3 つのモデルに対して計算すると以下の値が得られた。(SAS コードは別添)

モデル	理論値	N=10	N=100	N=1000
正則$N(\mu,\sigma^2)$	1	1.35	1.10	0.998
共線性モデル	3/2	2.71	1.66	1.56
$N(\theta^2,1)$	1/4	0.188	0.249	0.288

ただしこの計算には逆温度βの事後分布を繰り返し計算する必要があり、特に観測データが多い場合時間がかかる。また非正則モデルでは特に観測データによる誤差が大きく、N=1000 程度では正しいλの算出は難しいモデルが存在する。

6. 応用：WAIC

5 章で導入した漸近近似の応用例として**渡辺・赤池情報量規準（WAIC）**を導出する。

6.1. 汎化損失と経験損失

与えられた観測データ$(y_1,\cdots,y_n)$から真の分布を予測するうえで、より予測精度の良いモデルを探索する手法を考える。例えば回帰分析において説明変数を選択するような状況である。

事後分布$\varphi(\theta|y_1,\cdots,y_n)$に関する期待値を$E_{post}[\cdot]$と書くことにする。このとき、**事後予測分布**が次で定義される。

$$r(Y|y_1,\cdots,y_n)=E_{post}[f(Y|\theta)]$$

すなわち、観測データ$y_1,\cdots,y_n$の下で新しいデータYの予測をする分布である。事後予測分布をもちいて、真の分布に対する予測の悪さである**汎化損失**、手元のデータの当てはまりの悪さである**経験損失**を次で定義する。

$$G_n=E_{Y_{TRUE}}[-\log r(Y_{TRUE}|y_1,\cdots,y_n)]$$

$$T_n=-\frac{1}{n}\sum_{i=1}^{n}\log r(y_i|y_1,\cdots,y_n)$$

汎化損失G_nは事後予測分布が真の分布$f_{TRUE}(Y)$に当てはまるほど小さくなる関数であり、G_nが小さいモデルは「良いモデル」であるといえる。一方でG_nは$f_{TRUE}(Y)$が既知でないと計算できないため、観測データ$(y_1,\cdots,y_n)$からG_nを推定することが必要になる。

経験損失T_nは一見するとG_nを推定する量になっているように見える。しかし、T_nはいわゆる過適合（over fitting）を起こすことが知られる。すなわち、事後分布の予測とG_nの予測両方に同じ観測データを用いているため、バイアスが発生する（$E[G_n]\neq E[T_n]$）。このずれは以下になることが知られる[1][7]。

$$G_n-T_n=(S-S_n)+\frac{E_{post}\left[\sqrt{t}\xi_n(u)\right]}{n}+o(1/n)$$

ここで、Sは真の分布のエントロピー、S_nは経験エントロピー：

$$S=-E_{Y_{TRUE}}[\log f_{TRUE}(Y_{TRUE})]$$

$$S_n=-\frac{1}{n}\sum_{i=1}^{n}\log f_{TRUE}(y_i)$$

$E_{Y_{TRUE}}[S-S_n]=0$であるから、バイアスの正体は$E_{post}\left[\sqrt{t}\xi_n(u)\right]/n$であることがわかる。

$$2\nu=E_{Y_{TRUE}}\left[E_{post}\left[\sqrt{t}\xi_n(u)\right]\right]$$

とおくと、

$$E_{Y_{TRUE}}[G_n]=E_{Y_{TRUE}}[T_n]+\frac{2\nu}{n}+o(1/n)$$

となる。もちろん2νの値は真の分布がわからないと計算できない。よってこの2νを観測データから推定することが必要になる。推定量を作ることができれば、それを罰則項として汎化損失のバイアスのない推定が可能になる。これが WAIC である。

正則かつ実現可能なモデルでは$2\nu=d$となる[1]。したがって次の**正則の情報量規準**が与えられる。

$$T_n+\frac{d}{n}$$

これは赤池情報量規準 AIC[8]のベイズ版に当たる。

一方で非正則なモデルでは$2v = d$とは限らず、正則の情報量規準はバイアスを含む。また、3.1 章で仮定した実現可能性（真のモデルが$f(Y|\theta)$に含まれている）を仮定しない場合も同様である。WAIC はいずれの場合にも成立する点で有用である。

6.2. WAIC

$2v$は観測データ$(y_1, \cdots, y_n)$から推測することができる。各観測の対数尤度の事後分布に関する分散を足し合わせた量を**汎関数分散**と呼ぶ。

$$V_n = \sum_{i=1}^{n} Var_{post}[\log f(y_i|\theta)]$$

これが$2v$の推定量になる[1]。WAIC は次で定義される。

$$WAIC = T_n + \frac{V_n}{n}$$

このとき

$$E_{Y_{TRUE}}[G_n] = E_{Y_{TRUE}}[WAIC] + O(1/n^2)$$

が成り立つことが知られる[1]。

6.3. SAS による実装

2.2 章のサンプルコードのデータセット height, post_out から WAIC を算出できる。

```
proc sql noprint;
   select count(*) into :M trimmed from post_out;
quit;
%put &=M : the number of posterior MCMC samples;
data loglik;
   set height;                              /* 被験者 subjid */
   do j = 1 to &M.;
      set post_out(keep=mu sigma) point=j;   /* 事後標本 j */
      ll  = logpdf('NORMAL', y, mu, sigma);
      lik = exp(ll);
      output;
   end;
run;
proc means data=loglik noprint;
   by subjid notsorted;
   var lik ll;
   output out=stat mean(lik)=Ebar var(ll)=v_i;
run;
data waic;
   set stat end=last;
   sumT + (-log(Ebar));           /* SUM -log E[f]  */
   Vn   + v_i;                   /* SUM Var[log f] */
```

```
  if last then do;
    Tn   = sumT / _n_;
    WAIC = Tn + Vn / _n_;
    output;
  end;
  keep Tn Vn WAIC;
run;
```

このコードでは$\log r(y_i|y_1,\cdots,y_n)$、$Var_{post}[\log f(y_i|\theta)]$を全ての$y_i$について計算するため「観測データの個数$n$ × 事後分布の MCMC サンプルの個数」行のデータセットを作成している。観測データを 5 個としているため問題ないが、観測データを増やすと過剰に重くなる。

Yが 1 変数で説明変数を持たない場合、Yの代表値をいくつか選び、その点を用いて事後予測分布及び$Var_{post}[\log f(Y|\theta)]$を推定することで軽量化できる。例えば 50 点の$y$に対して$r(y|y_1,\cdots,y_n)$、$Var_{post}[\log f(y|\theta)]$を求め、ラグランジュ 5 点内挿 (各$y_i$に近い 5 点を用いた 4 次近似) を用いて$r(y_i|y_1,\cdots,y_n)$、$Var_{post}[\log f(y_i|\theta)]$ $(i=1,\cdots,n)$ を求めることができる。

3.3 章、4.1 章、4.2 章の例に対してV_n、T_n、WAIC を計算すると以下になる。(SAS コードは別添) (4.2 章の例の2νの理論値は数値積分による。R コードを別添。)

モデル	n	2ν（理論値）	V_n	T_n	WAIC	G_n
正則$N(\mu,\sigma^2)$	10	2	1.97	1.35	1.54	1.47
	100	2	2.04	1.36	1.38	1.42
	1000	2	1.98	1.41	1.41	1.42
共線性モデル	10	3	2.81	1.28	1.56	1.52
	100	3	3.13	1.46	1.49	1.42
	1000	3	2.92	1.40	1.41	1.42
$N(\theta^2,1)$	10	0.345…	0.471	1.33	1.38	1.45
	100	0.345…	0.293	1.36	1.36	1.42
	1000	0.345…	0.697	1.41	1.41	1.42

7. まとめ

本稿では「観測データが十分に存在するとき、事後分布はどのようにふるまうか」という問いを追ってきた。頻度論において標本平均の妥当性を中心極限定理が支えているのと、同じ役割を果たす定理である。答えはモデルが正則であるかどうかで分かれる。Θ_{BEST}が 1 点であり、その点で$K(\theta)$のヘッセ行列Jが正定値であれば、事後分布は$\sqrt{n}(\theta-\theta_{BEST})$の意味で正規分布に近づく。事後分布から作った信用区間が頻度論的にも妥当な被覆をもつことも、パラメータの個数を罰則とする AIC が使えることも、この事実の上に立っている。

ところがこの 2 条件は簡単に壊れる。そのときはθの位置を追うことをやめ、真の分布との KL 情報量の値$t=nK(\theta)$を追えばよい。事後分布は正則・非正則を通じて$t^{\lambda-1}\exp\left(-t+\sqrt{t}\xi_n(u)\right)\cdot\varphi(u|t)$という同じ形になり、モデルの違いは実対数閾値$\lambda$ただ一つに集約される。正則なモデルでは$2\lambda=d$となるが、非正則ではそうならない。応用として、汎化損失と経験損失のずれ2νを事後標本から推定する WAIC を導き、SAS による実装を与えた。

ここで強調しておきたいのは非正則なモデルは決して特殊なものではないということである。4.2 章の$N(\theta^2,1)$は、ごく普通の正規分布$N(\mu,1)$を変数変換しただけのモデルにすぎない。つまり非正則性は、モデルそのものというよりパラメータの付け方から生じている。パラメータを冗長に取ることや、意味のある量を非線形に書き直すことは、実務では自然に起こる。

応用として WAIC を導いた。汎化損失と経験損失のずれ2νは、正則かつ実現可能なモデルではdに一致するが、非正則ではそうならない。$N(\theta^2,1)$では$2\nu \simeq 0.35$であり、$2\lambda = 0.5$とも$d = 1$とも異なる値になる。それでも2νは各観測の対数尤度の事後分散を足し合わせるだけで推定できるため、モデルが正則かどうかを判定することなく汎化損失を見積もることができる。

本稿で扱わなかったことを 3 点挙げる。第一に、λを解析的に求めるには特異点解消と呼ばれる代数幾何の手続きが必要になる。実は正則なモデルにおいて用いた極座標は正に特異点解消の例である。一般の非正則なモデルでは「極座標の一般化」のような概念を導入する必要がある。第二に、理論が成り立つための条件である。本稿では 5.2 章の形を認めて話を進めたが、これは無条件に成り立つものではない。まずモデル$f(Y|\theta)$がθについて解析的であること。単に滑らかであるだけではなく、解析性が必要となる。次に、相対的に有限な分散を持つことが必要である。これは真の分布から離れたパラメータにおいても対数尤度比が$K(\theta)$に比べて暴れないことを要求している。このほかパラメータ空間が有界であることなどが必要である。第三に、WBIC である。5.5 章で用いた逆温度によるλの推定は、周辺尤度Z_nの対数を推定する WBIC と呼ばれる枠組みから得られたものだが、その導出には立ち入らなかった。WAIC が予測の良さを測るのに対し、WBIC はモデルと事前分布の適切さを測る規準であり、目的が異なる。

θの位置を追うことをやめて$K(\theta)$の値を追えば、正則・非正則を通じた漸近論が得られ、そこから正則性を仮定しない情報量規準 WAIC が、PROC MCMC の事後標本だけで計算できる形で導かれる。

8. 参考文献

[1] 渡辺澄夫. ベイズ統計の理論と方法. コロナ社, 2012

[2] 鈴木譲. 渡辺澄夫ベイズ理論 100 問 with R/Stan. 共立出版, 2023

[3] SAS Institute Inc. SAS/STAT® User's Guide. Chapter 80, "The MCMC Procedure". Cary, NC: SAS Institute Inc., 2026 https://documentation.sas.com/api/collections/pgmsascdc/v_077/docsets/statug/content/mcmc.pdf

[4] Kullback, S., Leibler, R. A. On information and sufficiency. Annals of Mathematical Statistics, 1951, 22(1): 79-86. https://doi.org/10.1214/aoms/1177729694

[5] Watanabe, S. Algebraic Analysis for Nonidentifiable Learning Machines. Neural Computation, 2001, 13(4): 899-933. https://doi.org/10.1162/089976601300014402

[6] Watanabe, S. A widely applicable Bayesian information criterion. J. Mach. Learn. Res., 2013, 14: 867-897 https://www.jmlr.org/papers/v14/watanabe13a.html

[7] Watanabe, S. Equations of States in Singular Statistical Estimation. Neural Networks, 2010, 23(1): 20-34. https://doi.org/10.1016/j.neunet.2009.08.002

[8] Akaike, H. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, 1974, 19(6): 716-723. https://doi.org/10.1109/TAC.1974.1100705

PROC IMLなしで実装するPattern-Mixture Multiple Imputation

○ビービ 彩加[1]、室永 遼太郎[1]

([1]シミック（株）データサイエンス事業本部）

Implementation of Pattern-Mixture Multiple Imputation Without PROC IML

Ayaka Beebe, Ryotaro Muronaga
Data Science Div., CMIC Co., Ltd.

要旨

本発表では、反復測定データの欠測値に対し、欠測が初めて生じた時点とそれ以降に欠測が生じた時点で異なる仮定を置く NFMV＋CCMV 法による多重代入法を、SAS の標準プロシジャのみを用いて実装する手順を紹介する。
キーワード: 反復測定データ、欠測データ、単調脱落、多重代入法、Pattern-Mixture Model、NFMV＋CCMV

1　はじめに

反復測定データの解析では、観測期間中に欠測データが生じることは少なくない。特に臨床試験のように、被験者を複数の時点で繰り返し観測するデータでは、試験途中での脱落や被験者が来院しないなどにより、計画された時点の観測値が得られない場合がある。このような欠測データの取り扱いは解析結果に影響を及ぼす可能性があり、適切な解析手法を事前に検討しておくことが重要である。

Estimand の考え方では、まず何を推定したいのかを明確にしたうえで、欠測データをどのように扱うかについての仮定を整理することが重要である。主解析で置いた仮定が結果にどの程度影響するかを確認するためには、欠測データについて異なる仮定を置いた感度分析を行い、結果の頑健性を評価する必要がある。したがって、欠測データの扱いに配慮した解析手法を、実務にて使える形へ整備しておくことには大きな意義がある。

特に欠測データへの対応方法の一つとして、多重代入法は広く用いられている。中でも Pattern-Mixture Model に基づく多重代入は、脱落後の未観測データに対して具体的な仮定を置くことができるため、感度分析の枠組みとして有用である。しかし、Pattern-Mixture Multiple Imputation の実装は複雑になりやすく、特に初回の欠測とそれ以降の欠測で異なる仮定を適用する場合には、データの整形、脱落パターンの識別、参照する被験者集団の設定、欠測補完後データの再構成など、複数の処理を組み合わせる必要がある。そのため、実務上では、実装の難しさが大きな負担となることがある。

SAS は臨床試験データに対する解析を含む多くの実務領域で広く利用されているが、Pattern-Mixture Multiple Imputation の実装例としては、PROC IML を前提としたものも少なくない。PROC IML は柔軟な行列演算が可能である一方、SAS/STAT には含まれておらず、SAS/IML が追加導入されている環境が必要となる。よっ

て、SAS/IML が無くとも SAS の標準プロシジャのみを用いて、再利用しやすく見通しのよい形で実装する方法を示すことは、実務上有用と考えられる。
そこで本論文では、2 群 4 時点の単調な欠測データを例に、初回の欠測には NFMV（Non Future Missing Values）、それ以降の欠測には CCMV（Complete Case Missing Values）を適用する Pattern-Mixture Multiple Imputation を、PROC IML を用いずに PROC MI、PROC MIXED、PROC MIANALYZE のみで実装する方法を示す。NFMV＋CCMV を取り上げたのは、初回欠測時点とその後の欠測について仮定を変えるためである。NFMV＋CCMV を用いることで、欠測後データに対する考え方を段階的に反映でき、感度分析としても扱いやすい。また、今回のような単調脱落データであれば、SAS の標準プロシジャのみで実装できる点も実務上の利点である。あわせて、データの並べ替え、脱落時点ごとの処理、多重補代入の解析用データの作成、MMRM の実行、および結果統合までの一連の流れを整理し、SAS の標準機能を組み合わせて、分かりやすく再利用しやすい実装とするための工夫について述べる。

2 方法

本論文では、2 群比較の反復測定データを対象として、単調な欠測がある場合の Pattern-Mixture Multiple Imputation の実装方法を示す。ここで単調な欠測とは、ある時点以降の観測値が連続して欠測となることを意味する。このような欠測パターンは、試験の途中中止や来院の中断が生じるようなデータでみられることが多い。
今回の実装では、欠測データに対する仮定として、初回の欠測には NFMV（Non Future Missing Values）、それ以降の欠測には CCMV（Complete Case Missing Values）を適用する。すなわち、脱落直後の未観測値については対象被験者より 1 時点先まで観測されている被験者を参照し、残りの欠測部分については試験を最後まで完了した被験者を参照する構成とした。上記の仮定に基づき、欠測後のデータの扱いを段階的に設定した Pattern-Mixture 型の多重代入を実装する。
なお、本手法は Estimand 自体を定義するものではなく、脱落後の未観測データにどのような仮定を置くかを具体化する方法である。本論文では、その仮定を SAS の標準プロシジャのみで実装し、代入後に MMRM を適用して結果を統合するまでの一連の手順を示す。

2.1 想定するデータの構造

本論文で扱うデータは、各被験者について、ベースライン値、投与群および複数の評価時点における反応変数が収集された反復測定データである。被験者を i=1, …, N(=200)、投与群を $g \in \{0,1\}$、予定された評価時点を j=1, …, K(=4)とする。i 番目の被験者について、時点 j における反応変数を Y_{ij} とし、観測されている最後の評価時点を脱落時点（dropvisit）として定義する。たとえば dropvisit が $j=s$ の被験者は、時点 1〜s までの値は観測され、時点 s+1 以降の値は欠測である。一方、dropvisit が $j=K$ である被験者は、全評価時点で観測を完了した完遂例（completer）である。表 1 にて具体的なデータ構造を示す。なお、sim_long は実装例においてシミュレーションにより作成したデータである。

表 1. 解析開始時の反復測定データ（sim_long）の変数一覧

変数名	型	内容	格納値

id	numeric	被験者 ID	1〜200
trt	numeric	投与群	0: 対照群, 1: 投与群
base	numeric	ベースライン値	正規分布から生成
visit	numeric	評価時点	1〜4
y_obs	numeric	各時点の観測値	脱落後は欠測値

多重代入の実装にあたっては、long 形式データを wide 形式へ変換して用いる。実装全体の流れは 3 章および図 2 に示す。

2.2 NFMV＋CCMV に基づく多重代入アルゴリズム

Pattern-Mixture Model は、欠測パターンごとにデータを捉え、脱落後の未観測データに対して仮定を置く方法である。本論文で扱う NFMV＋CCMV による多重代入は、そのような Pattern-Mixture Model に基づく実装の一例である。

2.1 で定義した脱落時点（dropvisit）は、各被験者で観測された最後の評価時点を表す。したがって、dropvisit が s の被験者は、評価時点 1〜s までは観測され、評価時点 t=s+1 以降が欠測である。ここで s は脱落時点を表す一般の添字であり、1, ..., K−1 の値をとる。また、最初の欠測時点を t=s+1 と定義する。本論文では、各投与群にて欠測パターンごとに以下の 2 段階の代入を行う。具体的には、初回の欠測時点に対して NFMV、それ以降の欠測に対して CCMV の仮定をおく。適用関係を図 1 に示す。

図 1. 単調な欠測データに対する NFMV+CCMV の適用

例: dropvisit=2 の被験者

	Visit 1	Visit 2	Visit 3	Visit 4
対象被験者（dropvisit=2）	観測	観測	欠測	欠測
NFMV donor（dropvisit=3）	観測	観測	観測	欠測
CCMV donor（dropvisit=4）	観測	観測	観測	観測

NFMV

CCMV

Step 1: 初回の欠測時点に対する NFMV

最初の段階では、2.1 で定義した dropvisit を用い、dropvisit が s の被験者を target、dropvisit が t の被験者を donor とする。ここで t=s+1 である。すなわち、target と donor は同一投与群内にあり、donor は target より 1 時点先まで観測されている被験者である。この donor の情報を用いて、target の最初の欠測時点 t に値を代入する。

実装では、target と donor を同じデータセットにまとめ、ベースライン値を説明変数、時点 1 から t までの反応変数を目的変数として PROC MI の monotone regression（単調回帰法）法を適用する。ここで、変数リストを Y_1 から Y_t までに限定することにより、初回欠測時点の値のみを補完し、それ以降の欠測部分は未補完のまま残す。このとき、donor の時点 tにおける観測値は、target の時点 tの欠測値を推定するための情報として利用されるのであり、donor の観測値をそのまま target に代入するわけではない。したがって、この処理により、初回欠測時点について、より近い脱落パターンを持つ被験者の情報を反映した代入値を得ることができる。

Step 2: 残りの欠測部分に対する CCMV

Step 1 で初回欠測時点の値を補完した後、残りの欠測部分、すなわち時点 t+1 から K までの未観測値に対して、CCMV を適用する。この段階では、同一投与群内の完遂例、すなわち dropvisit が K の被験者を donor とする。

具体的には、Step 1 により初回欠測時点が補完された target と、同一群内の完遂例を結合し、ベースライン値および全時点の反応変数 $Y_1, \ldots, Y_K$ を用いて PROC MI の monotone regression 法を再度適用する。このとき、Step 1 で得られた代入データごとに Step 2 を実施することで、初回欠測に対する NFMV の結果を保持したまま、残りの欠測部分に対する CCMV を実現する。

なお、脱落時点が K−1 の被験者については、初回欠測時点が最終時点 K に一致するため、Step 1 の時点で欠測部分がすべて補完される。この場合、Step 2 は不要であり、省略する。

群別・脱落パターン別の繰り返し

上記の Step 1 および Step 2 は、各投与群に対して独立に実施する。また、dropvisit が $1, \ldots, K-1$ の各脱落パターンについて同様の処理を繰り返す。最終的に、各群・各脱落パターン・各代入回数に対して得られたデータを結合し、全被験者を含む多重代入後データセットを作成する。

このように、本手法は「同一群内で donor を選択すること」「初回欠測とその後の欠測で異なる donor 集団を用いること」「脱落パターンごとに処理を分けること」を特徴とする。これにより、欠測後データに対する仮定を反映しつつ、SAS の標準機能で実装可能な多重代入手順を構成できる。

2.3 代入後データの解析

多重代入後に得られた wide 形式データは、各時点の反応変数を縦持ちに戻し、long 形式データとして解析に用いる。各代入データに対して、ベースライン値、投与群、時点、および投与群と時点の交互作用を固定効果とする MMRM を適用する。被験者内相関については、時点を反復要因として非構造型共分散構造を仮定する。

各代入データで得られた推定値および標準誤差は、PROC MIANALYZE を用いて Rubin の方法に基づき統合する。本論文の実行例では、最終時点における群間差を対象として、各代入データで得られた推定結果を統合することで、Pattern-Mixture Multiple Imputation の下での最終的な推定値を得る。

3 実装

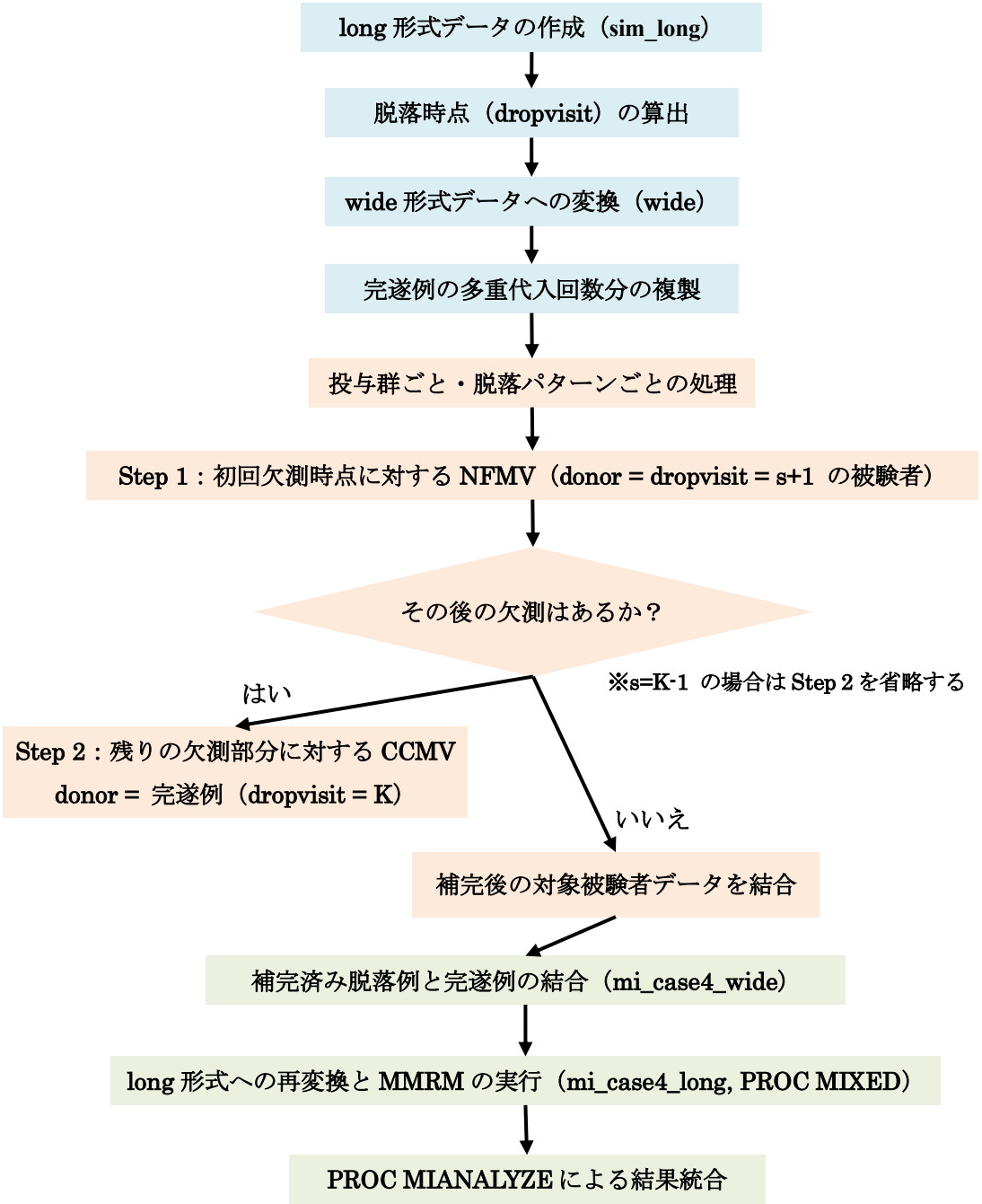
本論文では、前節で述べた NFMV＋CCMV に基づく Pattern-Mixture Multiple Imputation を、PROC IML を用いずに SAS の標準プロシジャおよびデータステップのみで実装した。実装の全体像は、(1) long 形式データの準備、(2) 脱落時点の算出と wide 形式への変換、(3) 群別・脱落パターン別の多重代入、(4) 代入後データの再構成、(5) MMRM の実行と結果統合、の 5 段階から構成される。

本実装では、初回欠測時点に対する NFMV と、その後の欠測部分に対する CCMV を明確に分離して処理するため、投与群ごと、欠測脱落パターンごと、さらに代入回数ごとにデータを整理する構成とした。これにより、欠測後の仮定を段階的に反映した代入データを、SAS の標準機能のみで構築できるようにしている。

000372

本実装の全体的な流れを図 2 に示す。本実装に用いた SAS コード全体を付録に示す。SAS での実装は、SAS Version 9.4 を使用した。以下では、各処理の内容を順に説明する。

図 2. PROC IML を使用しない Pattern-Mixture Multiple Imputation の SAS 実装フロー
long 形式データから dropvisit を算出し、wide 形式への変換、完遂例の複製、投与群別・脱落パターン別の NFMV および CCMV、多重代入データの結合、long 形式への再構成、MMRM、および MIANALYZE による結果統合までの流れを示す。



3.1 データ準備と脱落時点の算出

実装の起点となるのは、被験者 ID、投与群、ベースライン値、評価時点、観測値を含む long 形式データである。まず各被験者について、観測されている最終時点を dropvisit として算出する。この処理には PROC SQL を用い、各被験者について欠測でない時点の最大値を取得することで脱落時点を定義した。

次に、この dropvisit 情報を元の long 形式データに結合し、PROC TRANSPOSE により各評価時点の観測値を wide 形式に変換した。wide 形式では、各被験者につき 1 行のレコードに Y_1 から Y_K までの変数が並ぶため、後続の多重代入処理において、脱落パターン別の抽出や対象時点の制御が容易になる。また、dropvisit が K に一致する被験者については、完遂例であることを示す補助変数を付与し、CCMV における donor 抽出に利用した。

```
/* 各被験者の最終観測時点を dropvisit として定義 */
proc sql;
   create table dropinfo as
   select id, max(case when y_obs is not null then visit end) as dropvisit
   from sim_long
   group by id;
quit;
```

3.2 NFMV の実装

NFMV の実装では、脱落時点が s の被験者を target、脱落時点が t=s+1 の被験者を donor とし、両者を同一投与群内で結合したデータセットを作成する。ここで重要なのは、初回欠測時点 t の値のみを補完対象とし、それ以降の欠測はこの段階では補完しないことである。

この目的のため、PROC MI の monotone regression 法を用いる際の変数リストを、ベースライン値および Y_1 から Y_t までに限定した。すなわち、PROC MI が参照する変数範囲を初回欠測時点までに制御することで、初回欠測までを補完し、その後の時点は欠測のまま保持する構成とした。これにより、NFMV の仮定に従って、target の初回欠測値に対し、1 時点先まで観測されている target の情報を反映させることができる。

得られた出力データからは、Imputation が 0 より大きいレコード、かつ dropvisit が s の target 被験者のみを抽出した。さらに、後続の CCMV 処理で wide 形式の構造を保つため、未作成の時点変数については欠測値を設定し、各代入データ間で変数構成が一致するように調整した。

```
/* 初回欠測時点までを補完 */
data nfmv_in_arm&arm._s&s.;
   set wide(where=(trt=&arm and dropvisit in (&s., &t.)));
   keep id trt base dropvisit y1-y&t.;
run;
proc mi data=nfmv_in_arm&arm._s&s.
            out=nfmv_out_arm&arm._s&s.
            nimpute=&nimp.
            seed=%eval(&seed. + 1000*&arm. + 10*&s.);
   monotone reg;
   var base y1-y&t.;
run;
```

3.3 CCMV の実装

CCMV の実装では、Step 1 で初回欠測時点が補完された target と、同一投与群内の完遂例を結合して、新たな代入用データセットを作成する。ここでは完遂例が donor となり、初回欠測以降に残る未観測値を補完する。この処理は、Step 1 で生成された各代入データに個別に実施した。すなわち、Imputation ごとに target を取り出し、それに完遂例を加えた上で PROC MI の monotone regression 法を再度適用した。変数リストにはベースライン値および Y_1 から Y_K までの全時点を含めることで、Step 1 後に残された欠測部分を補完した。これにより、初回欠測値には NFMV の仮定を保持しつつ、それ以降の欠測部分には CCMV の仮定を反映した多重代入データが得られる。

なお、dropvisit が $K-1$ の被験者では、初回欠測時点が最終時点 K に相当するため、Step 1 で欠測部分がすべて補完される。この場合、CCMV による追加の補完は不要であり、PROC MI による再代入は行わず、そのまま最終的な代入データとして採用した。このように処理を分けることで、不要な処理を避けつつ、欠測構造に応じた効率的な実装としている。

000375

```
/* Step 1 の各 _Imputation_ ごとに Step 2 を実施 */
data work_ccmv_arm&arm._s&s._m&m;
   set wide(where=(trt=&arm. and dropvisit=&K))
         targets_step1_arm&arm._s&s.(where=(_Imputation_=&m.));
   drop _Imputation_ completer;
run;
proc mi data=work_ccmv_arm&arm._s&s._m&m
            out=ccmv_out_arm&arm._s&s._m&m
            nimpute=1
            seed=%eval(&seed + 2000*&arm + 100*&s + &m);
   monotone reg;
   var base y1-y&K.;
run;
```

3.4 代入データの結合と解析用データの作成

群別、脱落パターン別、および代入回数別に得られた補完済み target データは、PROC APPEND を用いて順次結合した。また、完遂例については各 Imputation に共通して存在する必要があるため、あらかじめ代入回数分だけ複製し、補完済み dropout 被験者と合わせて最終的な wide 形式の多重代入データセットを構成した。その後、この wide 形式データを再び long 形式に変換し、各被験者について各時点の観測値を 1 行ずつ持つ解析用データを作成した。このデータセットは、PROC MIXED による MMRM に直接入力可能な形式となっている。実装は、各時点の変数を順に取り出し、評価時点と観測値を対応付けながら出力することで、簡潔に long 形式へ戻した。

3.5 MMRM の実行と結果統合

作成した長形式の多重代入データに対して、Imputation ごとに PROC MIXED を実行した。モデルには、ベースライン値、投与群、評価時点および投与群と評価時点の交互作用を固定効果として含め、被験者内の繰り返し測定に対しては非構造型共分散構造を仮定した。これにより、各代入データについて MMRM に基づく推定結果を得た。

本論文の実行例では、最終時点における投与群間差を関心対象とし、PROC MIXED の LS-means 差の出力から当該推定値と標準誤差を抽出した。最後に、PROC MIANALYZE を用いて各代入データの推定結果を Rubin の規則に基づいて統合し、最終的な推定値とその不確実性を評価した。これにより、多重代入から MMRM、推定結果の統合までの一連の解析フローを、PROC IML を用いずに完結させることができた。Imputation ごとに MMRM を実行し、その結果を PROC MIANALYZE で統合する SAS コードを以下に示す。

000376

```
/* 各 Imputation ごとに MMRM を実施 */
proc mixed data=mi_case4_long;
   by _Imputation_;
   class id trt visit;
   model y_obs = base trt visit trt*visit / solution;
   repeated visit / subject=id type=un;
   lsmeans trt*visit / diff;
   ods output diffs=diffs;
run;
/* Rubin の規則に基づいて統合 */
proc mianalyze data=diffs_final;
   modeleffects Estimate;
   stderr StdErr;
run;
```

4 実行例

本章では、前章で示した実装方法に基づき、2 群 4 時点の単調脱落データに対する実行例を示す。ここでは、群別・脱落パターン別に NFMV＋CCMV を適用した Pattern-Mixture Multiple Imputation を行い、その後に MMRM および結果統合を実施する。

4.1 シミュレーションデータの作成

実行例では、被験者数 *N*=200、予定評価時点数 *K*=4、代入回数 *NIMP*=20 とした。各被験者には投与群、ベースライン値、および被験者ごとのランダム効果を与え、各時点の連続反応変数を正規分布から生成した。さらに、観測値に依存する形で途中脱落を発生させることにより、単調脱落構造を持つデータを作成した。これにより、実務で想定される反復測定データに近い状況を再現した。

生成された long 形式データについて、各被験者の最終観測時点を dropvisit として算出し、wide 形式データへ変換した。wide 形式では、完遂例と脱落例が明確に区別され、後続の群別・脱落パターン別処理に利用できる形となった。

4.2 多重代入の実施

各投与群について、dropvisit が 1 から *K*−1 までの各脱落パターンに対し、まず NFMV により初回欠測時点を補完した。続いて、初回欠測後に残る欠測部分に対して、完遂例を donor とする CCMV を適用した。これらの処理を代入回数 *NIMP* 回分繰り返し、最終的に Imputation ごとに全被験者を含む wide 形式の多重代入データセットを得た。

代入後のデータについて、各 Imputation に含まれる被験者数を確認したところ、すべての代入データで元の被験者数と一致しており、完遂例と補完済み脱落例が適切に結合されていることが確認できた（表 2）。この

000377

点は、実装の妥当性を確認する上で基本的かつ重要な確認項目である。

表 2. 各多重代入データにおける被験者数

多重代入番号（Imputation）	被験者数
1	200
2	200
3	200
4	200
5	200
6	200
7	200
8	200
9	200
10	200
11	200
12	200
13	200
14	200
15	200
16	200
17	200
18	200
19	200
20	200

4.3 MMRM による解析と結果統合

多重代入後の wide 形式データを long 形式へ変換した後、Imputation ごとに PROC MIXED を用いて MMRM を実行した。関心対象は最終時点における投与群間差とし、LS-means 差から推定値および標準誤差を抽出した。これらを PROC MIANALYZE に入力することで、代入間のばらつきを考慮した最終的な推定結果を得た。統合後の推定値、標準誤差および 95%信頼区間を表 3 に示す。

表 3. 最終時点における投与群間差の統合推定結果

評価時点	推定値	標準誤差	95%信頼区間下限	95%信頼区間上限
Visit 4	4.5001	1.8589	0.7537	8.2465

この実行例により、NFMV＋CCMV に基づく Pattern-Mixture Multiple Imputation を、SAS の標準プロシジャのみを用いて実行可能であることが確認できた。また、データ整形から多重代入、MMRM、結果統合までの一連の処理を一貫したフローとして構成できることから、実務における再利用しやすい実装例となることが示

された。
なお、参考のため、観測値のみを用いた MMRM による結果との比較を表 4 に示す。本論文の主目的は実装方法の提示であるため、ここでは結果の差異の有無を確認するにとどめる。

表 4. 観測値のみを用いた MMRM と Pattern-Mixture Multiple Imputation 後の MMRM の比較

解析方法	推定値	標準誤差	95%信頼区間下限	95%信頼区間上限	p 値
Pattern-Mixture Multiple Imputation 後の MMRM	4.5001	1.8589	0.7537	8.2465	0.0197
観測値のみを用いた MMRM	3.9481	1.5525	0.8865	7.0097	0.0118

5 考察

本論文では、単調脱落がある 2 群 4 時点の反復測定データを対象として、初回欠測時点に NFMV（Non Future Missing Values）、それ以降の欠測部分に CCMV（Complete Case Missing Values）を適用する Pattern-Mixture Multiple Imputation を、PROC IML を使用せずに SAS の標準プロシジャのみで実装する方法を示した。実装例を通じて、データの整形、欠測パターンの識別、多重代入、MMRM、結果統合までを一連の処理として構築できることを確認した。
本論文の主な貢献は、NFMV と CCMV を組み合わせた Pattern-Mixture Multiple Imputation を、PROC IML に依存せず、SAS の標準プロシジャのみで実装する方法を具体的に示した点にある。さらに、多重代入そのものにとどまらず、データの整形、解析用データセットの再構成、MMRM、結果統合までを含めた一連の解析フローとして整理したことで、実務で再利用しやすい実装例を提示した点にも意義がある。
本実装の主な利点の一つは、PROC MI、PROC MIXED、PROC MIANALYZE といった比較的利用頻度の高い標準プロシジャにより構成できる点にある。Pattern-Mixture Model に基づく多重代入は、欠測後データに対する仮定を明確に扱える一方で、実務上は実装の複雑さが導入の妨げとなることが少なくない。特に、初回欠測時点とそれ以降の欠測部分で異なる仮定を適用する場合、対象データの選択、donor 群の設定、欠測補完後データの再構成など、多段階の処理が必要となる。本論文で示した方法は、こうした処理を SAS の標準機能の範囲で整理し、比較的見通しのよい手順として実装できる点で有用と考えられる。
また、本手法では、同一投与群内で donor を選択し、脱落パターンごとに処理を分けることで、脱落後の未観測データに対する仮定を段階的に反映している。初回欠測時点では、対象被験者より 1 時点先まで観測されている被験者の情報を参照することで、脱落直後の挙動に対して比較的限定的な仮定を置いている。一方、それ以降の欠測部分では完遂例を参照することで、残りの未観測部分に対する補完を行っている。このように、NFMV と CCMV を組み合わせることで、脱落後の未観測データに対する仮定を単一の形に固定せず、時点に応じて切り分けて扱える点は、本手法の特徴である。
一方で、本論文の実装にはいくつかの前提と制約がある。第一に、対象とする欠測構造は単調脱落を前提としているため、間欠的欠測を多く含むデータに対しては、そのままでは適用しにくい。第二に、実行例は 2 群 4 時点の設定に基づいており、より多くの評価時点や複雑な群構成に拡張する場合には、マクロ処理や中間

000379

データの管理を含めた追加の工夫が必要となる。第三に、NFMV および CCMV は欠測後データに対する仮定を具体化する一例であり、その妥当性は解析目的、対象疾患、脱落理由、観測スケジュールなどを踏まえて検討されるべきである。したがって、本手法はあくまで一つの感度分析の実装例として位置づけるのが適切である。

さらに、Estimand の観点からは、本論文で扱った手法は推定対象そのものを定義するものではなく、脱落後の未観測データをどのように扱うかという解析上の仮定を具体化する方法と理解する必要がある。すなわち、まず何を推定したいのかという Estimand を明確にし、その上で、その Estimand に対応した欠測後データの仮定として NFMV＋CCMV 型の補完を採用する、という順序で整理することが重要である。この点を明示することで、本手法を主解析を補完する感度分析として適切に位置づけることができる。

今後の課題としては、評価時点数や群数の増加に対する一般化、より柔軟な欠測構造への対応、解析目的に応じた donor 設定の検討、ならびに実務で再利用しやすいテンプレート化やモジュール化などが挙げられる。また、主解析との比較や、異なる欠測後仮定を置いた複数の Pattern-Mixture Multiple Imputation 手法との比較検討を行うことで、本手法の位置づけをさらに明確にできると考えられる。

6 まとめ

本論文では、単調脱落がある反復測定データに対して、初回欠測時点に NFMV、以降の欠測部分に CCMV を適用する Pattern-Mixture Multiple Imputation を、PROC IML を用いずに SAS の標準プロシジャのみで実装する方法を示した。具体的には、PROC MI、PROC MIXED、PROC MIANALYZE、ならびに PROC SQL、PROC TRANSPOSE、データステップ、マクロ処理を組み合わせることで、データ整形から多重代入、MMRM、結果統合までの一連の解析フローを構築した。

本実装により、脱落後の未観測データに対する仮定を反映した Pattern-Mixture 型の多重代入を、比較的見通しよく再利用しやすい形で SAS 上に構成できることを示した。特に、PROC IML を前提としない点は、実務上の導入可能性を高める上で意義があると考えられる。

反復測定データにおける欠測の扱いは解析結果に影響する重要な要素であり、Estimand の観点からも、欠測後データに対する仮定を明確にした上で感度分析を行うことが求められる。本論文で示した実装は、そのような感度分析を SAS の標準機能で実施するための一つの具体例として、実務において有用であると考えられる。

000380

置換ブロック法による被験者割付におけるブロックサイズ混合手法と固定手法の性能比較

○上原 歩夢[1]、室永 遼太郎[1]、菊池 宏和[1]

([1]シミック株式会社)

Performance Comparison of Mixed and Fixed Block Size Methods in Permuted Block Randomization: A SAS Simulation Study

Uehara Ayumu[1], Muronaga Ryotaro[1], Kikuchi Hirokazu[1]
[1]Data Science Division, CMIC Co., Ltd.

要旨

置換ブロック法による被験者割付において、予見可能性排除に推奨されるブロックサイズ混合手法と、ブロックサイズを固定した場合の性能をシミュレーションにより比較する。

キーワード：割付表、置換ブロック法、ブロックサイズ、無作為化、層別化、モンテカルロシミュレーション、PROC PLAN

1. はじめに

1.1 背景、目的

無作為化比較試験（Randomized Controlled Trial: RCT）では、治療群間の比較可能性（背景因子の等質性）を確保するため、ランダム化による被験者割付が実施される。特に被験者数が限られる治験や多施設共同試験では、群間の被験者数不均衡を抑制する目的で置換ブロック法（Permuted Block Randomization）が広く利用されている。置換ブロック法では、一定数の被験者が登録されるごとに治療群間の均衡が維持されるという利点がある一方で、ブロックサイズが固定されている場合には割付の予見可能性が高まり、選択バイアスの発生につながる可能性が指摘されている。この予見可能性の問題に対し、ICH E9 ガイドライン（2.3.2 節）では複数のブロックサイズをランダムに組み合わせる手法（ブロックサイズ混合手法）が紹介されている。一方で、ブロックサイズ混合手法は割付表の作成に手間を要するため、ブロックサイズを固定した場合と比べて、群間均衡性および検出力の観点でどの程度の性能差が生じるのかを把握しておくことは、実務上の手法選択において重要である。

本発表では、割付因子を設定した場合の割付手法である層別置換ブロック法を用いて層別無作為化を行う状

況を想定し、ブロックサイズが固定の場合および混合の場合について、割付比が計画から乖離する「可能性の大きさ」を定量化することを目的とする。

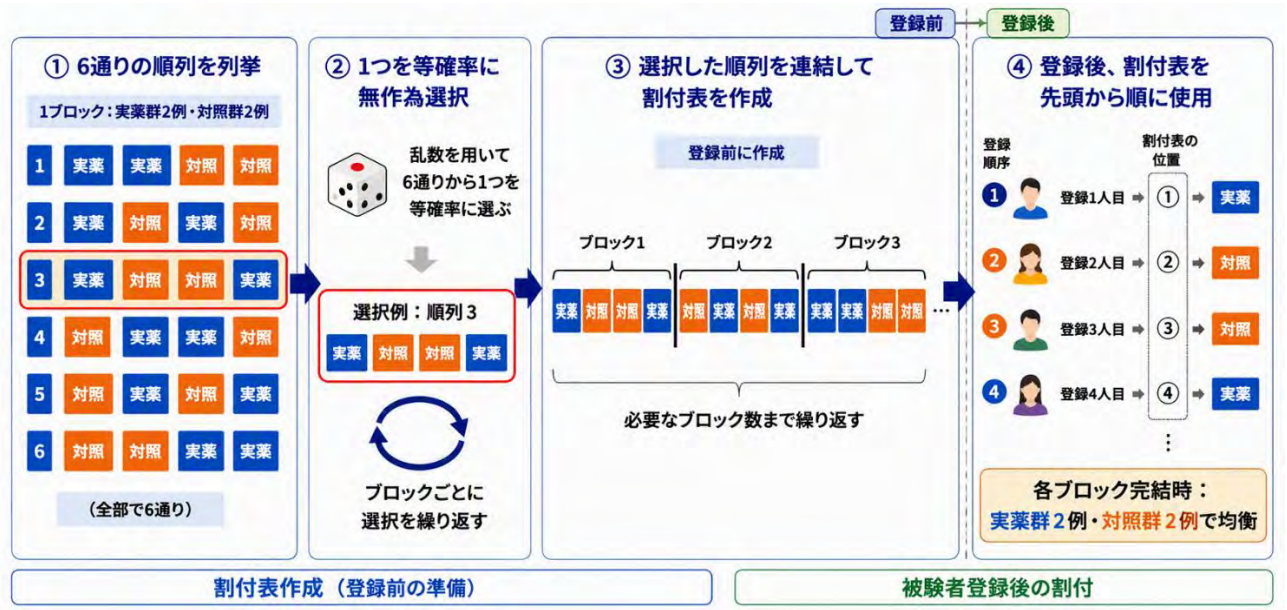
1.2 置換ブロック法の基礎

置換ブロック法はブロック内の割付順序をあらかじめランダムに並べ替えることで群間の被験者数偏りを一定範囲に抑える手法であり、ブロックサイズを固定すると割付の予見可能性が高まるという課題がある。また、割付因子（層別因子）を設定する場合には、層ごとに独立した割付表を作成する層別置換ブロック法が用いられる。層ごとに割付表を作成することにより、各層の内部で治療群間の被験者数の均衡が保たれるため、割付因子の分布が治療群間で偏ることを防ぐことができる。以下、ブロックサイズを固定・混合それぞれについて具体的な数値例で割付表作成の手順を示す。

1.2.1 ブロックサイズが固定の割付表

ブロックサイズ 4（実薬群 2 例・対照群 2 例）を例として、割付表作成および被験者割付の手順を示す。
まず、1 つのブロック内で実薬群 2 例および対照群 2 例を配置する順列をすべて列挙する。ブロックサイズ 4 では、実薬群 2 例・対照群 2 例の並び方は全部で 6 通り存在する。
次に、6 通りの順列の中から乱数を用いて等確率に 1 つを選択し、割付表に追加する。
この「順列の選択」と「割付表への追加」を、必要なブロック数に達するまで繰り返すことにより、あらかじめ治療群の並びが定められた 1 つの割付表が作成される。
被験者の登録後は、作成済みの割付表を先頭から順に使用する。すなわち、1 人目の登録被験者には割付表の 1 番目に記載された治療群を、2 人目には 2 番目の治療群を割り付ける。この操作を登録順に繰り返すことで、各ブロックが完結した時点で実薬群・対照群の被験者数が等しくなるように割り付けられる。

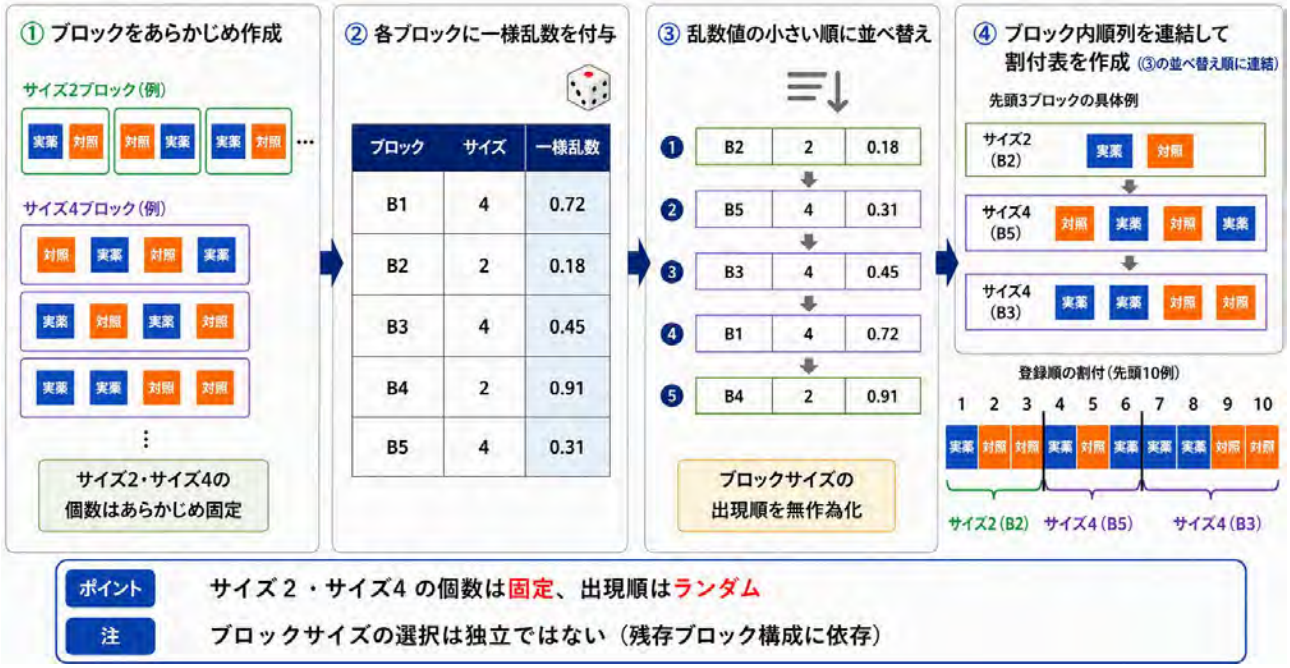
図 1. ブロックサイズが固定の場合の割付表作成



1.2.2 ブロックサイズが混合の割付表

本発表で用いるブロックサイズ混合手法では、ブロックサイズ 2 およびブロックサイズ 4 のブロックをあらかじめ所定数作成し、それらのブロックをランダムな順序で連結する。すなわち、あらかじめ各ブロックサイズの個数を固定したうえで、その出現順を無作為化する方式である。具体的には、作成した全ブロックに対して独立な一様乱数を付与し、乱数値の小さい順にブロックを並べ替える。並べ替え後の順に各ブロックを連結することにより、ブロックサイズ 2 および 4 がランダムな順序で出現する割付表を作成する。この方法を採用する場合のメリットとして、割付表の大きさをコントロールしやすいことがあげられる。
本方式では、サイズ 2 ブロックおよびサイズ 4 ブロックの個数をあらかじめ固定している一方、各ブロックの出現順はランダムである。したがって、ブロックサイズの出現順は無作為化されるが、各ブロックにおけるサイズ選択は独立ではない。例えば、すでにサイズ 2 ブロックが多く出現している場合には、残りの割付表におけるサイズ 4 ブロックの割合が相対的に高くなる。

図 2. ブロックサイズが混合の場合の割付表作成



2. 方法

2.1 割付表の作成

被験者割付には層別置換ブロック法を用いた。SAS の PROC PLAN を用いて、以下の 3 条件について割付表を作成した。SAS での実装は、SAS Version 9.4 を使用した。

- 固定ブロックサイズ 2（すべてのブロックのサイズを 2 とする）
- 固定ブロックサイズ 4（すべてのブロックのサイズを 4 とする）
- ブロックサイズ混合（サイズ 2 とサイズ 4 のブロックを混在させる）

PROC PLAN は実験計画および無作為割付表の作成に用いられるプロシジャであり、factors 文で割付の構造（割付因子の層数、ブロック数、ブロック内順列数など）を指定し、treatments 文で 1 ブロック内に並べる割付枠の数（key 値の水準数。ブロックサイズ 4 であれば key=4）を指定することで、ブロック内の並び順を無作為に入れ替えた割付表を生成する。被験者数 100 例に対応する割付表として、固定ブロックサイズ 2 では 50 ブロック、固定ブロックサイズ 4 では 25 ブロック、ブロックサイズ混合ではサイズ 2 を 20 ブロックおよびサイズ 4 を 15 ブロック生成した。ブロックサイズ混合では、作成した全ブロックに独立な一様乱数を付与し、乱数値の小さい順にブロックを並べ替えたうえで、その順に各ブロックを連結することにより、サイズ 2 および 4 のブロックがランダムな順序で出現する割付表を作成した。なお、PROC PLAN が出力する key 値はブロック内での通し番号であり、本発表では key 値 1 および 3 を実薬群、2 および 4 を対照群に対応させた（ブロックサイズ 2 のブロックでは、key 値 1 が実薬群、2 が対照群となる）。

SAS PROC PLAN のサンプルコード

例：ブロックサイズ 4 に固定した層別置換ブロック法の割付表

```
proc plan seed=20260918;
    factors strata=10 ordered block=25 ordered ban=4;
    treatments key=4;
    output out= plan4;
run;
```

各ステートメントの意味は以下である。

factors 文：割付表の構造を指定する。

- strata=10 ... 割付因子の層を 10 水準作成する。
- block =25 ... 1 層あたり 25 ブロックを配置する。
- ban = 4 ... 1 ブロックの割付枠(順列)を 4 つにする。
- ordered ... 層番号(strata)・ブロック番号(block)は昇順にする。

treatments 文：ブロック内の key 値 1〜4 を無作為な順序で割り当てる。

出力されたデータセット plan4 (図 3)を見ると、treatments ステートメントに指定した key という変数がランダムに並ぶ。この値に対して、群を設定する。

例： key = 1 or 3 → 実薬群、key = 2 or 4 → 対照群

図 3.データセット plan4 の一部抜粋

	strata	block	ban	key
1	1	1	2	4
2	1	1	3	1
3	1	1	1	2
4	1	1	4	3
5	1	2	3	2
6	1	2	1	4
7	1	2	2	3
8	1	2	4	1
9	1	3	1	3
10	1	3	4	2
11	1	3	3	4
12	1	3	2	1

2.2 シミュレーション条件

被験者数は 100 例とした。
割付の層数については 10・25・50・100・200 層の 5 条件を設定した。
層別置換ブロック法を用いるため、割付表は層ごとに 100 例分作成した。
各条件について 10,000 回のモンテカルロシミュレーションを実施し、各試行において割付表を独立に再生成した。
各被験者の割付は、(i)どの層に登録されるか、および(ii) 登録された層内でどの割付順序を引くか、という 2 種類の乱数によって決まる。本発表では(i)を各層へ等確率で登録されるものとして設定し、(ii)は前述の PROC PLAN およびブロックの並べ替えに用いた乱数により独立に生成されている。

2.3 評価指標

割付性能の評価指標として、全体割付差、割付比 1:1 からの乖離率(%)、および割付の偏りが実務上問題となる水準に達する確率（割付比が 2:1 以上となる確率、ならびに割付比 1:1 から 10%以上および 20%以上乖離する確率）を用いた。

・全体割付差

全体割付差は、

$$|N_{treat}\text{-}N_{control}|$$

により定義した。
ここで

- N_{treat}：実薬群の被験者数
- $N_{control}$：対照群の被験者数

を表す。

・割付比 1:1 からの乖離率および割付の偏りに関する確率指標

割付比 1:1 からの乖離率(%)は、全体割付差を総被験者数で除して百分率化した指標であり、（全体割付差／総被験者数×100）により定義した。また、実務上の意思決定に資する指標として、割付比が 2:1 以上に達する確率（いずれかの群の被験者数がもう一方の群の 2 倍以上となる試行の割合)、および割付比 1:1（各群 50 例）から 10%以上・20%以上乖離する確率に着目した。ここでいう「10%以上の乖離」とは、上記の乖離率が 10%以上となること、すなわち被験者数 100 例においては両群の被験者数の差が 10 例以上（55 例対 45 例以上の偏り）となることを指す。同様に「20%以上の乖離」は、両群の差が 20 例以上（60 例対 40 例以上の偏り）となることを指す。

・層ごとの平均割付差（層内不均衡の指標）

層別無作為化を行った試験では、割付で用いた層別因子を解析でも調整することが原則とされている（ICH E9, 5.7 節)。これは層別因子が真の治療効果の推定に影響し得る交絡因子となり得るためであり、層内で群間の比較可能性（均衡性）が保たれて初めて、この交絡の影響が適切に制御される。したがって、全体の被験者数比が 1:1 に近い場合であっても、個々の層内で実薬群・対照群の被験者数が大きく偏っていれば、割付因子が重要な予後因子である場合には交絡を十分に抑制できていない可能性がある。そこで本発表では、全体割付差とは別に、層ごとの割付差の絶対値を、実際に 1 例以上の被験者が登録された層についてのみ合計した「層別割付差の総和」を試行ごとに算出したうえで、これを登録があった層数（S^*）で除した「1 層あたりの平均割付差」を算出した。すなわち、1 層あたりの平均割付差は次式により定義される。

$$\frac{1}{S^*}\sum_{s:N_{treat,s}+N_{control,s}>0}\left|N_{treat,s}-N_{control,s}\right|$$

ここでS^*は実際に 1 例以上の被験者が登録された層の数（登録層数)、$N_{treat,s}$および$N_{control,s}$は第s層における実薬群および対照群の被験者数を表す。登録が 0 例の層は、分子・分母のいずれからも除外される。

2.4 全体割付差が検出力に与える影響の評価

割付比の偏り（全体割付差）が検定の検出力にどの程度影響するかを評価するため、各試行で得られた群別の被験者数から理論的な検出力を算出した。ここでの検出力は、2 群の平均値の差を対応のない t 検定で比較する状況を想定し、非心 t 分布を用いて算出したものである。

まず、被験者数 100 例が両群に均等（各群 50 例）に割り付けられた理想的な場合に検出力がちょうど 80%となる標準化効果量として、

$$d=0.5659$$

を設定した（有意水準は両側 5%とした)。

そのうえで、各試行で実際に得られた群別の被験者数を用いて非心パラメータを算出し、それに対応する検出力を求めた。これにより、割付が均等からずれた場合に検出力がどの程度低下するかを定量化することができる。

具体的には、2 群の t 検定において、両群の被験者数をそれぞれn_1、n_2、標準化効果量をdとすると、検定統計量は自由度n_1+n_2-2、非心パラメータλの非心 t 分布に従う。ここで非心パラメータは次式で与えられる。

$$\lambda = \frac{d}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

各群 50 例（全体割付差 0）の場合は$\lambda = 2.830$ となり、有意水準両側 5%における棄却限界値と比較することで、検出力 80.0%が得られる。一方、効果量 d の値を一定にしたうえで、割付が偏って全体割付差が 10（55 例対 45 例）となった場合など、群ごとの被験者数に差が生まれると検出力が低下する。割付差が大きくなるにつれて検出力がどのような推移になるかを確認するため、3.2 節で検出力曲線を作成した。

3. 結果

3.1 割付差の比較

全体割付差および 1 層あたりの平均割付差の平均値、標準偏差、最大値を表 1 に示す。1:1 からの平均乖離率は被験者数を 100 例に固定しているため、割付比 1:1 からの乖離率(%)は全体割付差を 100 で除して 100 倍した値、すなわち全体割付差の平均値と同じ数値になることに注意する。すべての指標について、固定ブロックサイズ 2、ブロックサイズ混合、固定ブロックサイズ 4 の順に大きくなる単調な傾向が全ての層数条件で一貫して観察された。また、全体割付差の平均値・標準偏差はいずれも層数が増えるほど増加した。1 層あたりの平均割付差についても、平均値は層数が増えるほど増加したが、標準偏差は逆に層数が増えるほど減少した。

表 1. 全体割付差および 1 層あたりの平均割付差の平均値・標準偏差・最大値

層数	条件	全体割付差の平均（=1:1 からの平均乖離率）(標準偏差)	全体割付差の最大	1 層あたりの平均割付差の平均 (標準偏差)	1 層あたりの平均割付差の最大
10	固定ブロックサイズ 2	1.69 (1.49)	8	0.50 (0.16)	1.00
10	ブロックサイズ混合	2.01 (1.71)	10	0.60 (0.19)	1.40
10	固定ブロックサイズ 4	2.17 (1.86)	12	0.66 (0.20)	1.40
25	固定ブロックサイズ 2	2.74 (2.20)	14	0.51 (0.10)	0.88
25	ブロックサイズ混合	3.27 (2.62)	16	0.61 (0.12)	1.04
25	固定ブロックサイズ 4	3.61 (2.84)	18	0.68 (0.13)	1.12
50	固定ブロックサイズ 2	3.98 (3.10)	18	0.57 (0.07)	0.83
50	ブロックサイズ混合	4.67 (3.63)	22	0.68 (0.09)	1.00
50	固定ブロックサイズ 4	5.27 (4.07)	26	0.79 (0.09)	1.13
100	固定ブロックサイズ 2	5.18 (3.98)	32	0.68 (0.06)	0.88
100	ブロックサイズ混合	5.80 (4.43)	30	0.77 (0.06)	1.02
100	固定ブロックサイズ 4	6.53 (4.94)	30	0.88 (0.07)	1.18
200	固定ブロックサイズ 2	6.34 (4.80)	30	0.80 (0.04)	0.96
200	ブロックサイズ混合	6.69 (5.18)	36	0.86 (0.05)	1.01

200	固定ブロックサイズ 4	7.30 (5.55)	40	0.93 (0.05)	1.13

続いて、1:1 の割付比率から 10%以上・20%以上乖離する確率および割付比が 2:1 以上に達する確率を表 2 に示す。2:1 以上の極端な不均衡は層数 200 の固定ブロックサイズ 4 条件でも最大 0.07%と、実務上ほぼ生じない水準にとどまった。一方、10%以上の乖離は層数が多い条件（100・200）ではいずれの条件でも 16〜33%程度の試行で発生しており、層数の設定が割付比の乖離抑制に関わる実務上の検討事項であることが示唆された。

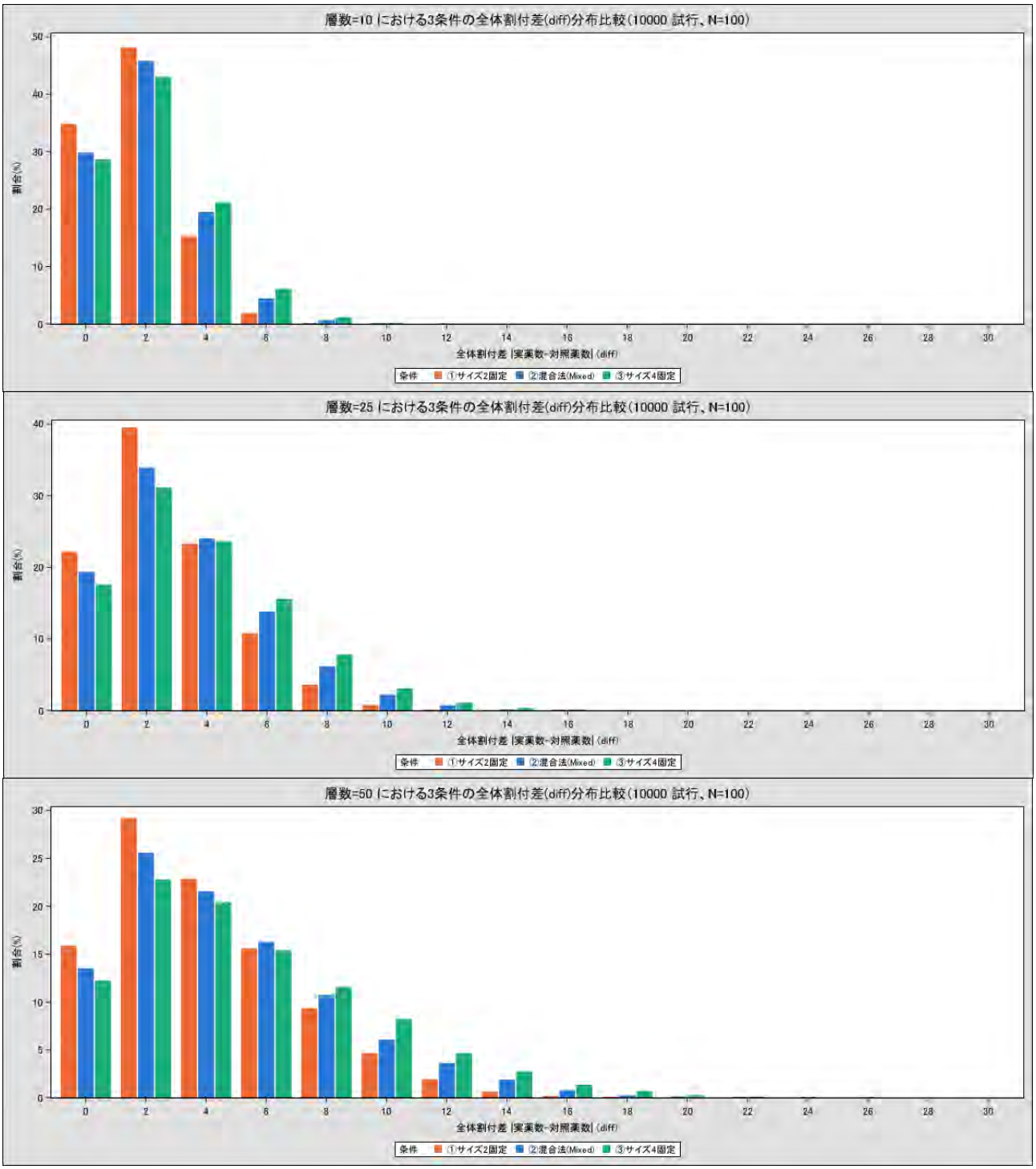
表 2. 割付比 1:1 から 10%以上・20%以上乖離する確率および割付比が 2:1 以上に達する確率（%）

層数	条件	10%以上乖離確率(%)	20%以上乖離確率(%)	2:1 以上到達確率(%)
10	固定ブロックサイズ 2	0.00	0.00	0.00
10	ブロックサイズ混合	0.05	0.00	0.00
10	固定ブロックサイズ 4	0.17	0.00	0.00
25	固定ブロックサイズ 2	0.83	0.00	0.00
25	ブロックサイズ混合	3.04	0.00	0.00
25	固定ブロックサイズ 4	4.57	0.00	0.00
50	固定ブロックサイズ 2	7.32	0.00	0.00
50	ブロックサイズ混合	12.52	0.12	0.00
50	固定ブロックサイズ 4	17.83	0.35	0.00
100	固定ブロックサイズ 2	16.27	0.36	0.00
100	ブロックサイズ混合	21.48	0.93	0.00
100	固定ブロックサイズ 4	27.45	1.89	0.00
200	固定ブロックサイズ 2	26.06	1.65	0.00
200	ブロックサイズ混合	28.01	2.69	0.02
200	固定ブロックサイズ 4	32.59	3.70	0.07

注：確率は 10,000 回のモンテカルロシミュレーションにおける観測割合である。0.00%は 10,000 回中 0 回を表し、真の発生確率が 0 であることを意味しない

次に、5 つの層数条件それぞれについて、3 条件の全体割付差の分布を図 4 に示す。いずれの層数条件でも、固定ブロックサイズ 2 の分布は小さい割付差の領域に集中し、固定ブロックサイズ 4 は右方（大きな割付差側）に分布が広がり、ブロックサイズ混合は両者の中間的な分布を示した。
また、層数条件別・3 条件別の 1:1 の割付比率からの乖離率(%)の分布の箱ひげ図を作成した（図 5）。中央値・四分位範囲ともに固定ブロックサイズ 2 で最小、固定ブロックサイズ 4 で最大となり、層数が増えるほど外れ値の頻度も増加する傾向が認められた。

図 4. 層数別（10・25・50・100・200）× 3 条件の全体割付差（図中の diff）の分布比較



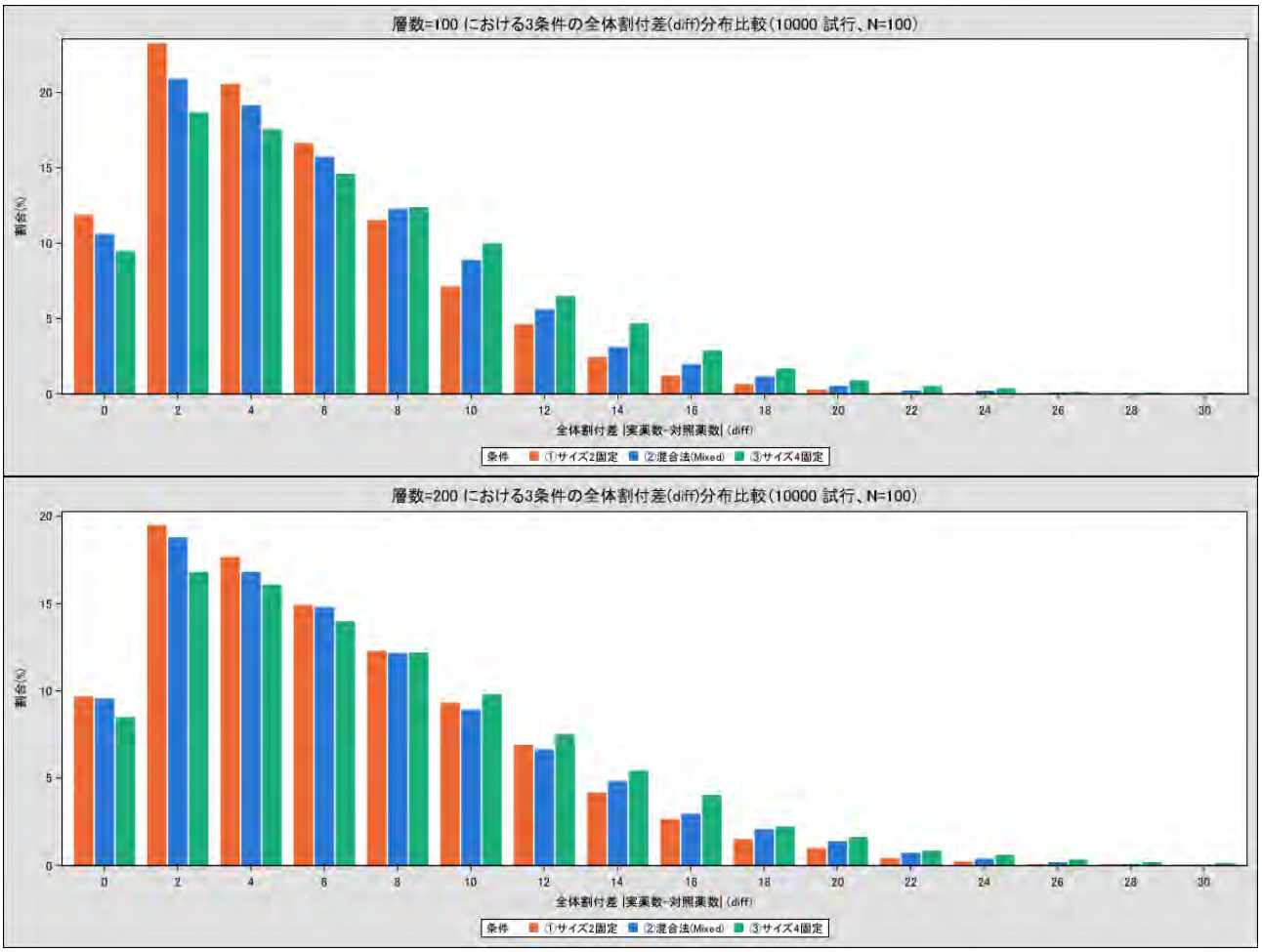
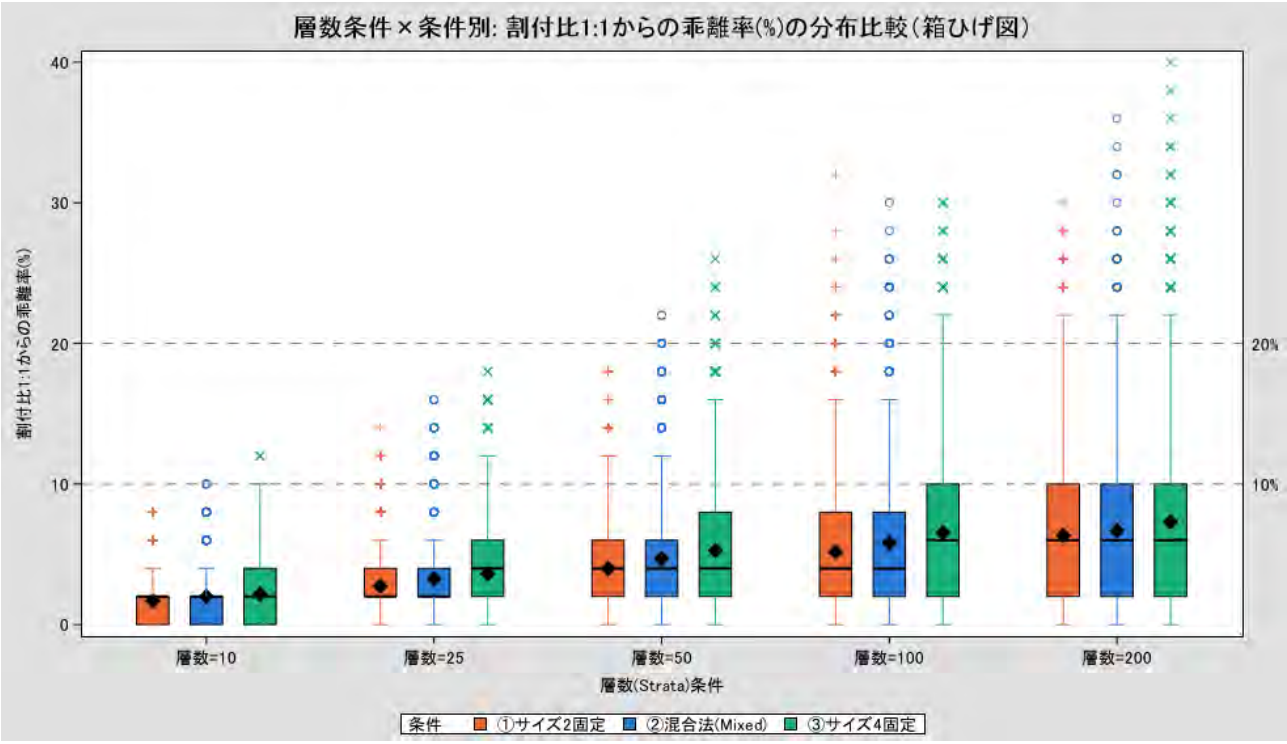


図 5. 層数別・3 条件ごとの割付比 1:1 からの乖離率(%)の箱ひげ図



3.2 全体割付差が検出力へ与える影響

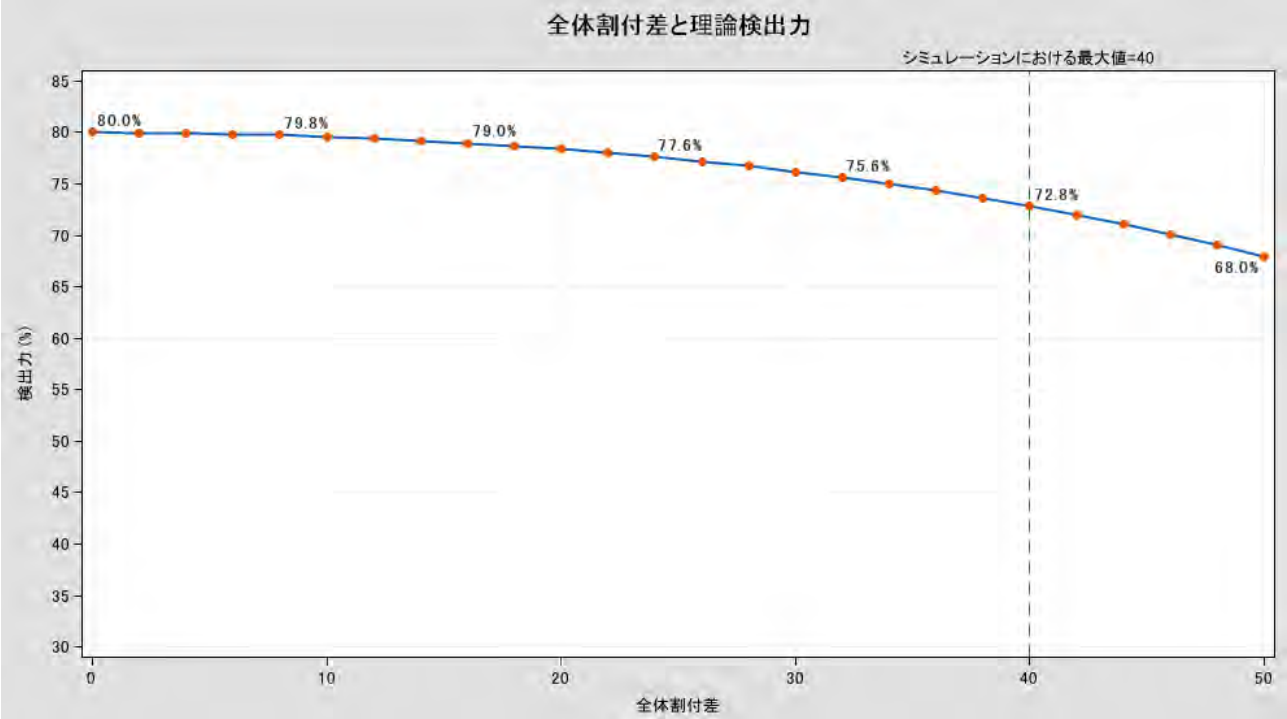
図 6 に、全体割付差と理論検出力の関係を示す（理論検出力は 2.4 節で述べた非心 t 分布に基づき算出した）。

全体割付差 10 で理論検出力 79.6%、全体割付差 20 で 78.4%であった。

今回のシミュレーションで観測された最大の全体割付差（層数 200・固定ブロックサイズ 4、全体割付差 40）でも理論検出力は 72.8%であり、全体割付差 0（80.0%）からの低下幅は 7.2 ポイント程度に留まった。

一方、各条件の平均的な割付差（表 1 の全体割付差の平均、いずれも 10 未満）に基づく理論検出力は 79.8〜80.0%の範囲であり、条件間の差はほぼ認められなかった。

図 6. 全体割付差と理論検出力の関係（理論検出力曲線）



4. 考察

4.1 割付差の比較

3.1 節で示したとおり、全体割付差の平均値は固定ブロックサイズ 2、ブロックサイズ混合、固定ブロックサイズ 4 の順に単調に増加した。これは置換ブロック法の定義上、ブロックサイズが小さいほど群間均衡がより早期に達成されることから理論的に予想される結果と整合的である。ただし、最大値（10,000 試行中の極端な実現値）は必ずしも平均値と同じ順位になるとは限らない点に留意が必要である。全体割付差の標準偏差が層数の増加とともに拡大したことは、層数が多い試験ほど、割付を実施したときにどの程度の不均衡が生じるかが試験ごとにばらつきやすい、つまり割付差を事前に見通しにくいことを意味する。

また、層数が被験者数に対して相対的に多い条件（100・200）では 1 層あたりの平均登録被験者数が極めて少なくなり（層数 100 で 1 例、層数 200 で 0.5 例）、ブロックが完成しない層の割合が高くなる。この未完成ブロックの偏りによる影響が、1 層あたりの平均割付差、全体割付差および乖離率のばらつき（図 5 の外れ値増加）が層数の増加とともに大きくなったことの要因と考えられる。
一方、1 層あたりの平均割付差の標準偏差は層数が増えるほど縮小した。これは、平均を取る対象となる登録層数自体が層数条件の増加とともに大きくなるためである。

4.2 全体割付差と検出力

3.2 節の結果、全体割付差の平均値に基づく理論検出力は、いずれの層数条件においても 3 条件とも 79.8〜80.0%の範囲にとどまり、条件間の実質的な差はほとんど認められなかった。
一方、各条件で観測された最大割付差に基づく理論検出力は 72.8〜79.8%の範囲となり、層数が多い条件ほど低下する傾向が認められたが、これらは 10,000 試行中の極端な実現値である。この 80%からの低下幅（7.2 ポイント）は割付差が理想値（0）から最悪の実現値まで変動した場合の低下幅であり、全体割付差の平均同士の比較（79.8〜80.0%の範囲）では条件間の差はほぼ認められない。今回のシミュレーションの結果からは、割付差が検出力に与える影響は小さいと考えられる。
ただし、検出力への影響が小さいことをもって割付差を無視してよいわけではない。割付差が大きくなった場合には、実薬群に割り付けられる被験者数が計画より少なくなる可能性がある。実薬群の被験者数が治験実施計画書で予定していた数を大きく下回ると、有害事象の発現割合の推定精度が低下するなど、安全性プロファイルの確認に影響を及ぼす可能性がある。したがって、割付差のリスクは検出力の観点のみで評価するのではなく、安全性評価に必要な実薬群の被験者数が確保されるかという観点も含めたうえで、ブロックサイズや層数の設定、割付状況のモニタリングなど、適切な対応を検討すべきである。

4.3 層別不均衡の実務的留意点と本発表の限界

本発表では、層を解析で調整しない場合の層別不均衡が検出力へ及ぼす影響は検証していないが、層内の割付不均衡が大きい場合には層別因子による交絡を十分に抑制できていない可能性がある点に留意が必要である。また、表 2 のとおり層数が被験者数に対し相対的に多い条件では割付比 1:1 から 10%以上乖離する試行が 16〜33%発生しており、層数の設定（層あたりの登録数の確保）がブロックサイズの選択以上に割付比の乖離抑制に重要と考えられる。
本発表の主な限界として、被験者数 100 例・割付比 1:1 に固定しており他の条件での性能は未検証である点、およびブロックサイズ混合の本来の意義である割付の予見可能性の低減を直接測定しておらず、その定量化には選択バイアス指標（例：Blackwell-Hodges の予測可能性指標）などを用いた追加検討が必要である点が挙げられる。検討したブロックサイズの組み合わせや層への登録確率の仮定も本発表固有の設定であり、異なる設定下では結果が変わり得る。さらに、本発表のブロックサイズ混合手法は各ブロックサイズの個数をあらかじめ固定する方式であり、ブロックごとに独立にサイズを無作為選択する方式とは異なる点にも留意が必要である。

5. 結論

本発表では、置換ブロック法における固定ブロックサイズ 2、固定ブロックサイズ 4 およびブロックサイズ混合（サイズ 2・4）について、試行ごとに割付表を独立に再生成する 10,000 回のモンテカルロシミュレーションにより、被験者数 100 例・層数 10〜200 の 5 水準の条件下で比較した。

その結果、全体割付差および割付比 1:1 からの乖離率の平均値は、いずれの条件でも固定ブロックサイズ 2 で最小、固定ブロックサイズ 4 で最大となり、ブロックサイズ混合は両者の中間的な性能を示した。1 層あたりの平均割付差においても同様の順位が全条件で一貫して確認された。理論検出力に換算すると、平均ケース同士で比較した条件間の差はほぼ認められず（79.8〜80.0%の範囲）、割付差が最悪の実現値まで変動した場合の低下幅（同一条件内でのばらつき）も 7.2 ポイント程度に留まった。割付比が 2:1 以上に達する確率は全条件で 1%未満と低かった一方、割付比 1:1 から 10%以上乖離する確率は層数が多い条件で 16〜33%に達し、層数の設定が割付比の乖離抑制において重要であることが示された。

以上より、本発表で検討した条件下では均衡性・検出力の観点からブロックサイズ手法間の実質的な差は小さく、手法選択の判断基準は割付の予見可能性（選択バイアスのリスク）に置かれるべきである。ブロックサイズ混合手法は、この予見可能性を低減する手段として ICH E9（2.3.2 節）で紹介されており、本発表の結果は、均衡性を大きく損なうことなく予見可能性の低減を図れるという従来の理解と矛盾しないものであった。ただし、割付差が大きい試行では実薬群の被験者数が予定を下回り得ることから、安全性評価に必要な被験者数が確保されるかという観点も併せて検討することが望ましい。

6. 参考文献

1. ICH Harmonised Tripartite Guideline E9. Statistical Principles for Clinical Trials. 1998. Section 2.3.2 (Randomisation), Section 5.7 (Subgroups, Interactions and Covariates).

2. Blackwell D, Hodges JL Jr. Design for the control of selection bias. Annals of Mathematical Statistics. 1957;28(2):449-460.

3. SAS Institute Inc. SAS/STAT(R) User's Guide: The PLAN Procedure.

MMRM における重みの取り扱いの検討
— SAS PROC MIXED / GLIMMIX の重み指定の違いと実務上の注意点 —

○江口 凌平[1]、中西 豊支[1]

([1]シミック株式会社 データサイエンス事業本部)

An Examination of Weight Handling in Weighted Mixed Models
— Specification of Weights in SAS PROC MIXED / GLIMMIX and Practical Considerations —

○Ryohei Eguchi[1], Yushi Nakanishi[1]

([1]Data Science Div., CMIC Co., Ltd.)

要旨

反復測定混合モデル（MMRM）は、臨床試験の経時データ解析において標準的な手法として広く用いられている。近年、欠測や選択に起因する偏りへの対応として、観測データの代表性を考慮した重みを組み合わせる重み付き MMRM (IPW-MMRM 等)が注目されている。しかし、SAS の PROC MIXED で指定する WEIGHT ステートメントと、尤度に直接重みを掛ける「尤度重み付け」には数理的・意味合い的に重大な相違がある。本論文では、SAS における重み付けの挙動についてシミュレーションを用いた検証を行い、その違いについて考察する。PROC MIXED の WEIGHT は、観測値または被験者の尤度寄与を直接重み付けするものではなく、残差側（R-side）の分散共分散構造を重み（精度重み（Precision Weight））に応じて変換する。一方、PROC GLIMMIX の OBSWEIGHT オプション等は、個々の観測値の代表性を調整する「尤度重み（Likelihood Weight）」として機能する。さらに、PROC GLIMMIX において観測値固有の重み（尤度重み）を適用する場合には、繰り返し測定の残差分散共分散構造（UN 等）を R-side でモデル化できないという実装上の技術的制約が存在することを解説する。本論文は、実務において重み付き混合モデルを適切に適用するための実践的な指針を提示する。

キーワード：反復測定混合モデル（MMRM）、精度重み付け、尤度重み付け、PROC MIXED、PROC GLIMMIX

1. MMRM 解析において仮定される分散共分散構造

反復測定データ解析（MMRM; Mixed Model for Repeated Measures）では、同一被験者から得られた複数時点の測定値の間に相関があることを前提とする。標準的な MMRM においては、同じ測定スケジュールを持つすべての被験者について、共分散行列は等しいと仮定される。例えば、時点 1、時点 2、時点 3 の 3 時点の評価スケジュールにおいて、被験者 A と被験者 B のそれぞれの残差共分散行列 R_A および R_B は共通であり、非構造化（UN; Unstructured）共分散構造を仮定した場合は以下のように定式化される。

$$R_A = R_B = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix}$$

ここで、σ_t^2 は各時点 t における分散、σ_{ts} は時点 t と s の間の共分散を表す。しかし、臨床試験の実データにおいては、被験者ごとに測定の精度やばらつきが異なる場合が想定される。例えば、被験者 B の測定データの分散が被験者 A の 2 倍大きいと仮定する。このとき、測定精度を考慮して解析を補正するために、被験者 A には $W_A = 2$、被験者 B には $W_B = 1$ という重みを与えて調整を行うことができる。これが「精度重み（Precision Weight）」の基本的な考え方であり、測定精度の高い被験者のデータを重視して分散を補正するアプローチとなる。

2. 精度重みづけと尤度重みづけの考え方の違い

2.1 精度重みづけ（Precision Weighting）の数理と特徴

PROC MIXED において WEIGHT ステートメントを指定した場合、指定された重みは、観測値の尤度寄与を直接重み付けするものではなく、残差側（R-side）の分散共分散構造を変換するために用いられる。以下では、この PROC MIXED における WEIGHT ステートメントの数理的な作用を説明する。

観測値数を n、観測値 i (i = 1, …, n)に指定した重みを w_i とする。このとき、重み行列 W は、各観測値の重みを対角成分に配置し、それ以外の成分を 0 とした次の対角行列である。

$$W = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_n \end{pmatrix}$$

また、W の逆平方根行列を L とすると、L は次のように表される。

$$L = \begin{pmatrix} \frac{1}{\sqrt{w_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{w_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{w_n}} \end{pmatrix}$$

すなわち、L の各対角成分は対応する重み w_i の逆平方根となる。通常の線形混合モデルでは、観測値ベクトル Y の分散共分散行列 V は、以下のように表される。

$$V = ZGZ' + R$$

ここで、G はランダム効果 (G-side)の分散共分散行列、R は残差側 (R-side)の分散共分散行列である。REPEATED ステートメントを指定した場合、PROC MIXED の WEIGHT ステートメントは、残差分散共分散行列 R を LRL と変換する。したがって、WEIGHT ステートメントを指定した場合の分散共分散行列 V は、以下のようになる。

$$V = ZGZ' + LRL$$

例えば、R の i 行 i 列の対角成分 R_{ii} は、観測値 i の残差分散を表す。重みを適用した後の観測値 i の残差分散は、以下のようになる。

$$Var(\varepsilon_i, \varepsilon_j) = R_{ij} \,/\, w_i$$

したがって、重み w_i が大きい観測値ほど残差分散は小さくなる。すなわち、重みが大きい観測値は、より測定精度が高い観測値として扱われる。反対に、重みが小さい観測値では残差分散が大きくなり、相対的に測定精度が低い観測値として扱われる。つまり、PROC MIXED の WEIGHT ステートメントは、残差分散共分散行列を重みに応じて変換する精度重み付けとして実装されている。

2.2 尤度重みづけ（Likelihood Weighting）の数理と特徴

疫学や因果推論の解析では、欠測、脱落、治療選択、あるいは標本抽出などにより生じ得る対象集団との構成の違いを考慮するため、各被験者または各観測値に重みを設定することがある。このような重みは、対象集団における個々のデータの頻度、重要度、または代表性を調整することを目的とする。被験者別の尤度重みを考慮した経時データ解析について、PROC GLIMMIX を用いた実装が考えられる。PROC GLIMMIX では、データが階層構造になっている際のモデル(Multilevel model)を実装することが可能である。例えば、地域、施設、世帯、個人など、複数の階層において抽出確率が異なる場合があり、各階層の単位が対象集団において代表する程度を推定に反映するために、階層ごとの重みを用いることがある。一方、通常の臨床試験における経時データ解析では、複雑な標本抽出設計を用いる調査のように、施設の抽出確率や上位階層の抽出確率を反映した重みを用いることは一般的ではない。例えば、施設を第 3 レベル、被験者を第 2 レベル、被験者内の反復測定値を第 1 レベルとみなすことができるが、通常、施設が母集団から抽出される確率に基づいて施設別の重みを設定することはない。臨床試験において重みを用いる場合には、被験者の観測されやすさ、解析対象集団に対する代表性、あるいは欠測・脱落等に関連する要因を考慮し、被験者単位の重みを設定する場面が中心となる。以下では、各被験者に一つの重みを設定し、その重みを当該被験者の全測定時点に共通して適用する場合を考える。

PROC GLIMMIX において尤度重みを適用する場合、目的関数である対数尤度の各被験者または観測値の寄与に対して、直接重み w を乗じる数理的アプローチをとる。被験者 i の反復測定値を y_i とし、被験者 i のデータが対象集団を代表する程度に関係する確率をπ_iとする。このとき、当該単位に与える重み w_i は、観測確率の逆数として次のように定義される。

$$w_i = 1/\pi_i$$

ここで、観測確率π_iが小さい、すなわち対象集団に存在するにもかかわらずデータとして観測されにくい単位ほど、大きな重みが与えられる。この重みは、その単位が対象集団において何単位分を代表するかを表す量として解釈できる。全体の対数擬似尤度は、各単位の対数尤度への寄与を重み付きで加算したものとして、次のように表される。

$$\log\big(p(y)\big) = \sum w_i \log(p(y_i))$$

ここでの$p(y_i)$は i 番目のデータが得られる確率(尤度)、$\log(p(y_i))$は対数尤度への寄与を表す。この式は、各単位の対数尤度に対して重みを直接乗じることを意味するため、重みが大きい単位は、パラメータ推定においてより大きな寄与を持つ。例えば、観測確率が 0.25 であれば、その重みは 4 となる。この単位は、対象集

団において概念的には 4 単位分を代表するものとして、対数尤度への寄与が 4 倍される。このように、重みが大きい単位ほど推定への寄与が大きくなり、観測されにくい単位が対象集団において代表する割合を推定に反映することができる。

以上の数式から分かるように、重み w_i は被験者 i の残差分散のみを変化させるものではない。被験者 i の平均構造、分散共分散構造、および観測値と予測値との乖離を含む、被験者 i の対数尤度への寄与全体に対して乗じられる。この点は、PROC MIXED の WEIGHT ステートメントによる精度重み付けと本質的に異なる。PROC MIXED の WEIGHT は、残差分散共分散行列を重みに応じて変換することにより、観測値の測定精度または異分散性をモデル化する。一方、尤度重みでは、被験者が対象集団において代表する程度を反映するため、被験者全体の尤度寄与を直接調整する。ここで、精度重みづけと尤度重みづけを整理する意味で、両者の違いを表 1 にまとめる。

表 1. 精度重みづけと尤度重みづけの違い

項目	精度重みづけ (Precision Weighting)	尤度重みづけ (Likelihood Weighting)
考え方	誤差分散の調整	頻度・代表性(寄与度)の調整
意味づけ	測定精度が高いデータほど分散が小さくなる。	重みが大きいほど、その観測値が対象集団においてより多くの頻度・重要度(代表性)を持つ。

2.3 SAS での実装コードの比較

PROC MIXED における精度重み付き MMRM の実装は、以下のコードのように WEIGHT ステートメントを単純に追加する形で行う。本コードでは、解析データセット targ を入力データとし、CHG を目的変数としている。CHG は、ベースラインからの変化量を表す連続変数とする。モデルには、ベースライン値 BASE、投与群 TRT、評価時点 AVISIT、および投与群と評価時点の交互作用 TRT*AVISIT を含めている。

```
proc mixed data=targ;
  class USUBJID TRT AVISIT;
  model CHG = BASE TRT AVISIT TRT*AVISIT / ddfm=kr solution;
  repeated AVISIT / subject=USUBJID type=un;
  weight WT; /* 精度重みとして WT を適用 */
  lsmeans TRT*AVISIT / diff cl;
run;
```

一方、PROC GLIMMIX では、重みを指定するオプションとして、MODEL ステートメントの OBSWEIGHT オプション（観測レベル重み）と、RANDOM ステートメントの WEIGHT オプション（被験者クラスターレベル重み）が存在する。RANDOM ステートメントの WEIGHT オプションは、地域、施設、世帯、個人など、複数の階層において抽出確率が異なる場合を想定している。臨床試験の経時データ解析において、個々の被験者の代表性を補正する目的では、通常は前者の OBSWEIGHT を使用する。本コードでは、dist=normal を指定することで、連続アウトカムに対する正規分布モデルとしての解析を想定している。次に、プロシジャ名の後に指定した empirical オプションの指定によって、サンドイッチ型分散共分散行列を用いている。この指定は主として、固定効果の標準誤差、信頼区間および検定統計量の計算に影響するものであり、

固定効果の点推定値そのものを変更するものではない。また、RANDOM _RESIDUAL_ は、残差側（R-side）の分散共分散構造を指定するための記述である。subject=USUBJID により、同一被験者内の反復測定値間に相関があることをモデル化する。type=un は非構造化分散共分散構造を表す。

```
proc glimmix data=targ empirical;
 class USUBJID TRT AVISIT;
 model CHG = BASE TRT AVISIT TRT*AVISIT / dist=normal obsweight=WT;
 random _residual_ / subject=USUBJID type=un;
run;
```

ただし、PROC GLIMMIX では、観測値または被験者に固有の重みを OBSWEIGHT= 等で指定した場合、これらの重みを用いた解析では尤度ベースの推定法の指定が必要となるため、Gauss-Hermite adaptive quadrature を用いる METHOD=QUAD (Pfeffermann et al. (1998), Rabe-Hesketh and Skrondal (2006))等を採用し、モデルパラメータを最尤推定する必要がある。一方、METHOD=QUAD では、上記の RANDOM _RESIDUAL_ による R-side の残差分散共分散構造をサポートしない。したがって、本コードは、観測値レベルの尤度重みと R-side の非構造化共分散構造を同時に指定することを意図したコード例ではあるが、PROC GLIMMIX 内ではこの組合せを適合できない点に注意を要する。詳細は 3 章にて説明する。

3. 実務上の注意点

3.1 SAS PROC GLIMMIX を用いた実装上での注意点

個体別の尤度重みを考慮できる PROC GLIMMIX の使用を検討したが、2.3 章にて述べた通り、実務上の制約が生じる。経時データの繰り返し測定における時間相関構造を適切に考慮するため、R-side（残差側）の共分散モデル（random _residual_ / subject=USUBJID type=un; など）を指定し、かつ個体別重みを考慮するために MODEL ステートメントで obsweight=WT を指定すると、以下のエラーが発生する。

> ERROR: Subject or observation specific weights are only supported for nested models that are estimated using METHOD=QUAD.

エラーを回避するために推定手法を変更（例えば METHOD=QUAD 等に指定）すると、次に以下のエラーが発生する。

> ERROR: R-side random effects are not supported for METHOD=QUAD.

この挙動の背景には以下の数理的・システム上の制約があることが判明した。

1. GLIMMIX プロシジャにおいて、観測値または被験者個別の重み（尤度重み）を適用したモデルの適合は、推定手法として METHOD=QUAD を使用する場合にのみサポートされる。
2. 一方、METHOD=QUAD では、繰り返し測定の相関構造を表現する R-side（残差側）のランダム効果（UN、AR(1)、TOEP などの共分散構造）をモデリングすることができない。
3. したがって、OBSWEIGHT=による観測値・被験者固有の重みと、METHOD=QUAD 下での R-side 残差共分散構造を同時に適合することはできなかった。

この制約は、尤度重みを用いた経時データ解析の実装を検討する際に重要な留意点となる。

3.2 実務におけるその他の注意事項

上記の制約に対しては、単に利用可能なプロシジャを選択するのではなく、推定したい対象集団、重みの統計学的意味、および被験者内相関の表現を踏まえて、解析モデルと推測方法を一体として検討する必要がある。以下に、実務上の主な選択肢と解析計画上の留意点を示す。

- **1. G-side のランダム効果による代替（ランダム切片モデルの適用）:** R-side のランダム効果でモデル化することができないため、代わりに G-side のランダム効果を用いて被験者内相関をモデル化する。具体的には、以下のようにランダム切片モデルを適合させる。

```
proc glimmix method=quad empirical=classical;
  class USUBJID TRT AVISIT;
  model CHG = TRT AVISIT TRT*AVISIT / dist=normal obsweight=WT;
  random int / subject=USUBJID;
run;
```

  上記のモデルは、同一被験者内の異なる時点間の相関が一様であると仮定する。時間間隔に依存する相関（測定時点が近いほど相関が高く、離れるほど低い）はモデル化できないが、時点が異なっても相関がほぼ一定に近い状況であれば有効なモデルとなる。また、連続変数としての時点（AVISITN）を用いて random int AVISITN / subject=USUBJID type=un; としてランダム切片＋ランダム傾きモデルを当てることで、近い時点ほど相関が高く離れるほど低いという経時データの挙動をある程度近似して表現できる場合がある。ただし、誘導される相関構造はランダム効果の分散共分散構造により制約され、UN や AR(1)等の R-side 共分散構造を一般に再現するものではない。そのため、対象データに対する近似の妥当性は、適合度および感度分析等を通じて確認する必要がある。
- **2. PROC NLMIXED による多変量正規尤度の直接実装:** R-side の UN（非構造化）共分散行列等を数学的に維持しつつ、各被験者に対して厳密な尤度重みを掛けたい場合は、PROC GLIMMIX ではなく PROC NLMIXED を用いることも可能である。NLMIXED 内で経時データの多変量正規尤度を直接数式として定義し、その対数尤度に重み WT を乗じた上で最尤法によるパラメータ探索を行う。プログラミングの難易度は高いが、数理的整合性を最優先とする場合の解決策として挙げられる。
- **3. 臨床試験の意思決定・当局相談における留意点:** 臨床試験の統計解析計画書（SAP）を作成する際には、使用する SAS のプロシジャ制限を正確に把握し、解析モデルの選択（R-side の共分散構造を諦め G-side ランダム切片とするか、あるいは NLMIXED による直接プログラミングとするか）について、あらかじめ規制当局や臨床統計の専門家と十分な合意形成を図ることが必須である。

4. シミュレーション

PROC MIXED の WEIGHT 指定と NLMIXED による尤度重み付けとの間に、どれくらいの治療効果の点推定値に関する実質的な差が見られるのかシミュレーションを実施した。本シミュレーションでは、対象集団への選択確率が共変量 X に依存する状況を設定し、被験者の代表性を補正する重みを用いた場合の推定性能を評価した。具体的には、以下の 3 手法を比較した。

・重み無し MMRM（PROC MIXED）

・PROC MIXED の WEIGHT ステートメントに指定した MMRM

・被験者単位で対数尤度に重み(IPW)を乗じた尤度重み付き多変量正規モデル（PROC NLMIXED）

各反復における対象集団の被験者数は 1,500 例とし、二値共変量 X は確率 0.5 で生成した。解析対象への選択確率は、X=0 の被験者では 0.80、X=1 の被験者では 0.20 とした。このため、選択後の解析集団では X=1 の被験者が過少となる設定とした。アウトカムは 4 時点の反復測定値とし、被験者内相関を有する多変量正規分布に従う誤差を発生させた。最終時点における治療群差は、X=0 で−2.0、X=1 で−3.0 とした。対象集団における X=1 の割合は 0.5 であることから、対象集団における最終時点の平均治療群差の真値は、条件付き治療群差の期待値として、(−2.0)*0.5 + (−3.0)*0.5 ＝−2.5 とした。各手法について 500 反復のシミュレーションを実施し、最終時点における治療群差の平均推定値、バイアス、モンテカルロ標準偏差（Monte Carlo standard deviation; MC SD）、平均モデル標準誤差を評価した。ここでのバイアスは、平均推定値−真値として算出を行った。それぞれの算出結果を表 2 に示す。

表 2: 最終時点の治療群差に関するシミュレーション結果

解析方法	平均推定値	バイアス	MC SD	平均モデル標準誤差
PROC MIXED WEIGHT=IPW	−2.501104	-0.00110	0.090996	0.076112
重み無し MMRM	−2.199045	0.30095	0.072194	0.074783
尤度重み付きモデル(NLMIXED)	−2.501096	-0.00109	0.090998	0.075569

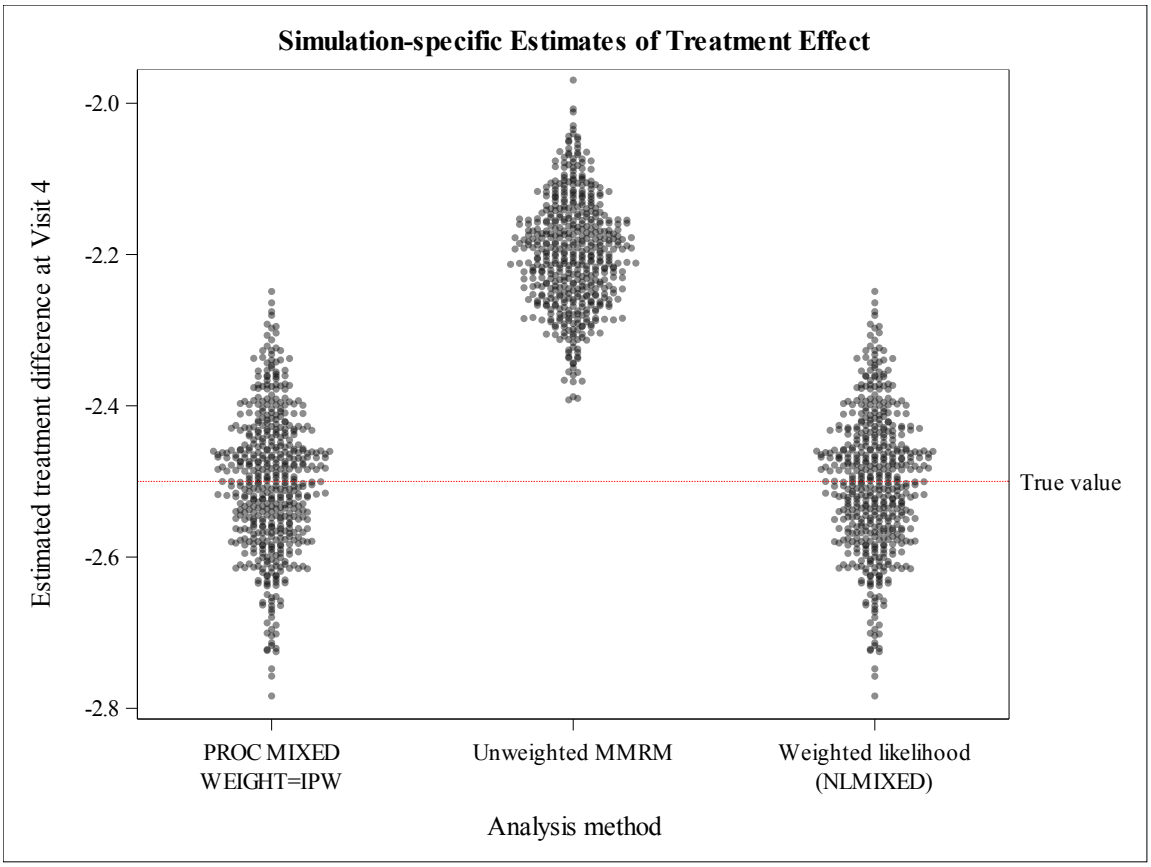


図 1. 各反復における最終時点の治療群差推定値

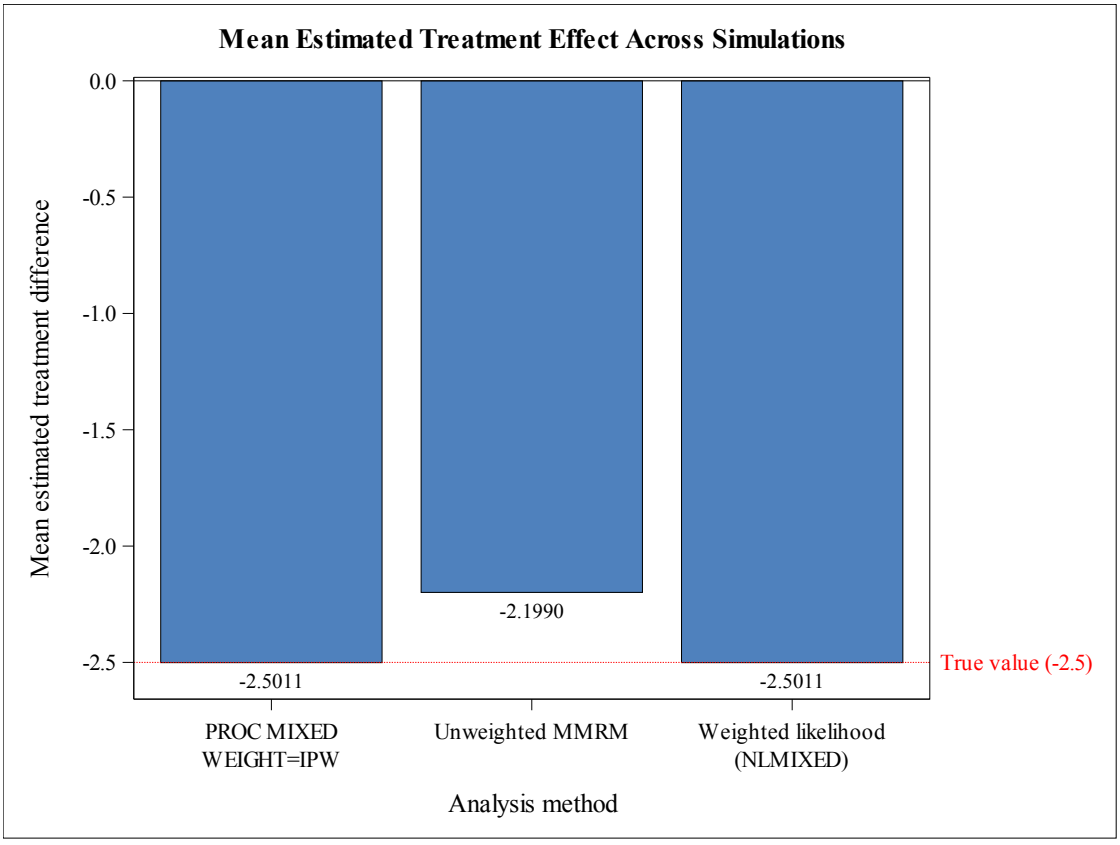


図 2. 最終時点の治療群差推定値の分布

図 1 に、各反復における最終時点の治療群差推定値を示す。重み無し MMRM による推定値は、真値である−2.5 より一貫して大きい値、すなわち治療効果を過小評価する方向に分布した。一方、PROC MIXED の WEIGHT ステートメントを使用した方法および NLMIXED による尤度重み付きモデルでは、推定値の分布の中心はともに真値付近に位置した。図 2 に、治療群差推定値の分布を示す。重み無し MMRM の平均推定値は−2.1990 であり、真値−2.5 に対するバイアスは約 0.3010 であった。これは、選択後の解析集団では X=1 の被験者が過少となり、当該集団における平均治療効果が対象集団における平均治療効果より小さくなるように設定されていたことと整合的である。これに対し、PROC MIXED の WEIGHT ステートメントに IPW を指定した方法および NLMIXED による尤度重み付きモデルの平均推定値はいずれも−2.5011 であり、いずれも真値−2.5 との差は約−0.0011 であり、実質的なバイアスは認められなかった。

また、PROC MIXED の WEIGHT=IPW および尤度重み付き NLMIXED の結果は、平均推定値だけでなく、バイアス、MC SD および平均モデル SE についても近い値を示した。具体的には、MC SD はそれぞれ 0.0910、平均モデル SE はそれぞれ 0.0761 および 0.0756 であった。ただし、PROC MIXED の WEIGHT ステートメントによる重み付けと、NLMIXED で実装した尤度重み付けは、数理的解釈としては異なる。本シミュレーションにおいて両者の平均治療効果推定値が一致したことは、PROC MIXED の WEIGHT ステートメントが尤度重みを一般に再現することを示すものではない。今回の設定では、以下の特徴があった。

・重みは被験者単位で定義され、同一被験者の全評価時点に共通の値であった。

・全被験者について 4 時点のアウトカムが観測されていた。

このような比較的単純で均衡的な条件下では、PROC MIXED における精度重み付けであっても、平均構造の推定に関しては被験者単位の重み付き推定と同様の挙動を示し得る。その結果、最終時点の治療群差に関する点推定値が NLMIXED による尤度重み付き推定値と一致したものと考えられる。欠測パターンが被験者間で異なる場合、重みが時点ごとに異なる場合、平均構造に時点に関する制約を置く場合、あるいはより複雑なランダム効果構造を用いる場合には、点推定値にも差が現れる可能性がある。したがって、PROC MIXED の WEIGHT ステートメントに重みを指定して本シミュレーションと同様の結果が得られたとしても、これを、被験者代表性を補正する尤度重み付き MMRM として一般化して解釈することには注意を要する。実務では、重みの作成目的、推定したい対象集団、モデルの平均・共分散構造、および標準誤差の推定方法を明確にした上で、使用するプロシジャおよび実装方法を選択する必要がある。

5. まとめ

本稿では、反復測定混合モデルにおける重みの取り扱いについて、SAS PROC MIXED の WEIGHT ステートメントによる重み付けと、被験者または観測値の対数尤度への寄与に直接重みを乗じる尤度重み付けとの数理的・実装上の相違を整理した。

PROC MIXED の WEIGHT ステートメントは、観測値ごとの重みを測定精度、すなわち残差分散の逆数として扱う精度重み付けである。したがって、重みの大きい観測値は残差分散が小さい、より精度の高い観測値としてモデル化される。一方、逆確率重みなど、被験者が対象集団において代表する程度を反映するため

の重みは、各被験者または観測値の尤度寄与を調整する尤度重みとして解釈される。このように、両者は重みが作用する対象および統計学的な意味が本質的に異なる。

また、PROC GLIMMIX では、OBSWEIGHT=等を用いて観測値または被験者に固有の尤度重みを考慮する実装が考えられる。しかし、個別重みを用いる場合には METHOD=QUAD が必要となる一方、METHOD=QUAD では RANDOM _RESIDUAL_による R-side の残差分散共分散構造を指定できない。このため、非構造化共分散構造（UN）等を用いた通常の MMRM と、個別の尤度重みを PROC GLIMMIX 内で同時に実装することには制約があることを示した。

シミュレーションでは、選択確率が共変量 X に依存し、選択後の解析集団において X=1 の被験者が過少となる状況を設定した。重み無し MMRM では、対象集団における最終時点の真の平均治療群差である−2.5 に対して、平均推定値は−2.1990 であった。一方、IPW を PROC MIXED の WEIGHT ステートメントに指定した解析および NLMIXED で被験者単位の尤度重みを実装した解析では、平均推定値はいずれも−2.5011 であり、真値に極めて近かった。ただし、本シミュレーションにおいて PROC MIXED の WEIGHT 指定と尤度重み付き NLMIXED の点推定値が一致したことは、両者の重み付けが同一の意味を持つこと、あるいは PROC MIXED の WEIGHT が一般に尤度重みを代替できることを意味するものではない。今回の設定では、重みが被験者単位で定義されて同一被験者の全時点に共通であったこと、全被験者について 4 時点の測定値が得られていたこと、および比較的均衡的な平均構造を用いたことが、両手法の点推定値が近接した要因と考えられる。実データでは、脱落や間欠欠測により被験者ごとに観測時点が異なる場合、時点ごとに異なる重みを用いる場合、重みがアウトカムや欠測パターンと複雑に関連する場合、あるいは平均構造や共分散構造に制約を置く場合がある。このような状況では、精度重みと尤度重みの相違が、治療効果の点推定値のみならず、標準誤差、信頼区間および検定結果にも影響する可能性がある。したがって、PROC MIXED の WEIGHT ステートメントを指定する場合には、単に「重み付き解析を実施した」と解釈するのではなく、その重みが測定精度を表すものなのか、対象集団に対する代表性を補正するものなのかを明確にする必要がある。

対象集団への一般化、選択バイアスの補正または欠測・脱落に対する重みづけ解析を目的とする場合には、推定対象（estimand）、重みの定義、平均構造、被験者内相関の表現、および標準誤差の算出方法を一体として検討することが重要である。R-side の柔軟な共分散構造を維持しつつ尤度重みを厳密に適用する必要がある場合には、PROC NLMIXED による多変量正規尤度の直接実装が選択肢となる。ただし、この方法ではプログラムの実装および推測の妥当性確認がより重要となる。実務においては、解析目的と重みの統計学的意味に適合した方法を選択し、必要に応じてシミュレーション、ロバスト分散推定またはリサンプリング法等によって推定および推測の性能を確認することが望まれる。

6. 参考文献

1. SAS Institute Inc. (2025). "The MIXED Procedure: WEIGHT Statement". SAS/STAT User's Guide. https://documentation.sas.com/doc/zh-TW/pgmsascdc/v_041/statug/statug_mixed_syntax17.htm
2. SAS Institute Inc. (2025). "The GLIMMIX Procedure: Weighted Multilevel Model for Survey Data". SAS/STAT User's Guide. https://documentation.sas.com/doc/en/pgmsascdc/v_074/statug/statug_glimmix_details25.htm
3. SAS Institute Inc. (2025). "The GLIMMIX Procedure: Pseudo-likelihood Estimation for Weighted Multilevel Models". SAS/STAT User's Guide. https://documentation.sas.com/doc/en/pgmsascdc/v_074/statug/statug_glimmix_details02.htm

4. Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., and Rasbash, J. (1998). “Weighting for Unequal Selection Probabilities in Multilevel Models.” Journal of the Royal Statistical Society, Series B 60:23–40.

5. Rabe-Hesketh, S., and Skrondal, A. (2006). “Multilevel Modelling of Complex Survey Data.” Journal of the Royal Statistical Society, Series A 169:805–827.

McNemar 検定のサンプルサイズ設計に関する性能比較

○尾崎 龍成，団野 友也，魚住 龍史

東京科学大学 工学院 経営工学系

Performance Comparison of Sample Size Methods for McNemar's Test

Ryusei Ozaki, Tomoya Danno, Ryuji Uozumi
Department of Industrial Engineering and Economics, Institute of Science Tokyo

要旨

McNemar 検定は，対応のある 2 値データの解析に用いられ，FREQ プロシジャで実行可能である．POWER プロシジャでは McNemar 検定のサンプルサイズ設計が実行可能であり，Miettinen 法，Exact 法，Connor 法が採用されている．同じ目標検出力を指定しても，計算法により求められるサンプルサイズは異なる．本稿では，3 種類の方法で求めたサンプルサイズにおける McNemar 検定の検出力と第 1 種の過誤確率を計算し，その比較結果を報告する．

キーワード：対応のある 2 値データ，McNemar 検定，サンプルサイズ設計，PROC FREQ，PROC POWER

1　はじめに

臨床研究や疫学調査では，同一対象者の介入前後における症状の有無の比較，年齢や性別などをそろえて対象者を 1 対 1 に組み合わせたペアの比較，同じ対象者に適用した二つの診断法の感度・特異度の比較などに，McNemar 検定が用いられる．例えば，抗がん剤治療を受ける乳がん患者を対象とした臨床研究では，利き手にフローズングローブを装着し，反対側の手を対照として，触覚低下などの末梢神経障害の発現割合を左右の手で比較するために McNemar 検定が用いられた (Hanai et al., 2018).

このような左右比較では，片方の手だけに障害が生じた患者が「不一致ペア」に対応する．McNemar 検定の統計量は，左右で結果が異なった不一致ペアの度数から計算される．

McNemar 検定で必要となるペア数は，二つの応答の周辺確率の差に加え，応答が一致しない確率にも依存する (Lachin, 1992). このため, 計算法の特徴を評価する際には, 確率設定による違いを検討する必要がある．本稿では，評価する確率の範囲を定める資料として，McNemar 検定の適用例をまとめた先行研究を参照する (Pembury Smith and Ruxton, 2020).

SAS では，FREQ プロシジャで McNemar 検定を実行でき，POWER プロシジャの `PAIREDFREQ` 文でサンプルサイズが設計できる．McNemar 検定の場合，サンプルサイズはペア数に相当する．連続性補正なしの McNemar 検定に対応する `DIST=NORMAL` では，Miettinen 法，Exact 法，Connor 法の 3 種類の計算法が利用できる．同じ目標検出力を指定しても，計算法によって算出されるペア数は異なる．各方法で得られたペア数において目標検出力が達成されるか，また目標達成に必要なペア数にどの程度の差が生じるかを明らかにすることが，設計時の方法選択に役立つ．

そこで本稿では, 実際の適用例を参考にした複数の確率設定のもとで, SAS の Miettinen 法, Exact 法, Connor 法による必要ペア数を比較する. 各方法で得られたペア数における補正なし McNemar 検定の実際の検出力と第 1 種の過誤確率を二項分布から計算し，各計算法の特徴を評価する．これらの結果を通じて，SAS を用いたサンプルサイズ設計と検出力確認の手順を示すことを目的とする.

2　McNemar 検定

2.1　対応のある 2 値データと統計的仮説検定

ペア数を N とし，ペア間は独立であると仮定する．各ペアの二つの応答を A，B とし，いずれも 0 または 1 の値をとるものとする．表 1 の各セルの度数を n_{ij}，対応する確率を $p_{ij} = \Pr(A = i, B = j)$ と定義する．同じ対象者を 2 回測定する場合，N は対象者数に等しい.

表 1　対応のある 2 値データの度数表

	$B = 0$	$B = 1$	合計
$A = 0$	n_{00}	n_{01}	$n_{00} + n_{01}$
$A = 1$	n_{10}	n_{11}	$n_{10} + n_{11}$
合計	$n_{00} + n_{10}$	$n_{01} + n_{11}$	N

A と B の周辺確率は，それぞれ $\Pr(A = 1) = p_{10} + p_{11}$，$\Pr(B = 1) = p_{01} + p_{11}$ である．これらが等しいという帰無仮説は $H_0 : p_{01} = p_{10}$ と同値であり，両側対立仮説は $H_1 : p_{01} \neq p_{10}$ となる．二つの応答の一方が 0，他方が 1 である不一致ペアは，表 1 の非対角セルに対応する．その総数を $D = n_{01} + n_{10}$ とすると，連続性補正を行わない McNemar 検定の統計量は次式で表される：

$$Q = \frac{(n_{01} - n_{10})^2}{D}, \quad D > 0. \tag{1}$$

帰無仮説 H_0 のもとで，統計量 Q の分布を自由度 1 のカイ二乗分布で近似し，その上側確率から両側 p 値を求める.

本稿では，不一致ペア数によらず，補正なしの統計量 Q を用いる．両側有意水準 $\alpha = 0.05$ に対して，Q が自由度 1 のカイ二乗分布の 95%分位点（約 3.84）以上であれば帰無仮説を棄却する．$D = 0$ の場合は帰無仮説を棄却しない.

参考として，連続性補正を行う場合の統計量を T とすると，次式で表される：

$$T = \frac{(|n_{01} - n_{10}| - 1)^2}{D}, \quad D > 0. \tag{2}$$

一方，カイ二乗近似を用いずに検定する方法もある．不一致ペア数 D を固定すると，帰無仮説のもとで n_{01} は二項分布 $\mathrm{Bin}(D, 1/2)$ に従う．この二項分布の確率から両側 p 値を求める方法を，以下では「正確検定」と呼ぶ.

2.2　FREQ プロシジャによる実行例

以下は，二つの応答の組合せごとに度数を集計したデータを用いる例である．AGREE を指定すると，McNemar 検定の統計量とカイ二乗近似による p 値が出力される (浜田・魚住, 2021)．EXACT 文で MCNEM を指定すると，正確検定の p 値も出力される．

```
data paired;
  input a b count;
  datalines;
0 0 44
0 1  1
1 0 11
1 1 44
;
run;
proc freq data=paired;
  tables a*b / agree;
  weight count;
  exact mcnem;
run;
```

この例では $Q = 8.33$ となり，カイ二乗近似による両側 p 値は 0.004，正確検定の両側 p 値は 0.006 となる．個票データを用いる場合は，1 行を 1 ペアとし，WEIGHT 文を削除する．

FREQ プロシジャの McNemar 検定には連続性補正のオプションはない．連続性補正ありの統計量 T は，二つの不一致セルの度数を式 (2) に代入して求める．

一方，R の mcnemar.test はカイ二乗近似を利用しており，correct=TRUE（既定値）で連続性補正を行い，correct=FALSE で補正を行わない．正確検定のオプションはないため，正確検定の両側 p 値は binom.test で求める．

mcnemar.r

```
tab <- matrix(c(44, 1, 11, 44), nrow = 2, byrow = TRUE)
n01 <- tab[1, 2]; n10 <- tab[2, 1]
mcnemar.test(tab, correct = FALSE)  # 補正なし
mcnemar.test(tab, correct = TRUE)   # 補正あり
binom.test(n01, n01+n10, p = 0.5)   # 正確検定
```

3　サンプルサイズ設計

McNemar 検定のサンプルサイズ設計では，二つの不一致セルの確率を仮定し，目標検出力を満たすペア数を求める．検出力は，想定した効果がある場合に帰無仮説を棄却する確率である．POWER プロシジャの PAIREDFREQ 文では，同じ確率設定に対して複数の計算法を利用できる．

3.1　検定と計算法の指定

PAIREDFREQ 文では，DIST=で解析に用いる検定を，METHOD=で検出力の計算法を指定する．表 2 に対応を示す．本稿では DIST=NORMAL を指定し，式 (1) の補正なし McNemar 検定に対応する 3 種類の計算法を比較する．既定値は DIST=EXACT_COND，METHOD=EXACT である．以後，DIST=NORMAL METHOD=EXACT による検出力の計算法を「Exact 法」と呼ぶ．

表 2　検定と検出力計算の対応

DIST	METHOD	計算方法
NORMAL	**MIETTINEN**	式 (1) のカイ二乗近似による検定に対し，検出力の近似式を用いる
NORMAL	**EXACT**	式 (1) のカイ二乗近似による棄却域の確率を二項分布から計算する
NORMAL	**CONNOR**	式 (1) のカイ二乗近似による検定に対し，検出力の近似式を用いる
EXACT_COND	EXACT	二項分布で棄却域を定めた正確検定の検出力を計算する

3.2　Miettinen 法と Connor 法の近似式

二つの応答が一致しない確率を $S = p_{01} + p_{10}$，周辺確率の差の絶対値を $\delta = |p_{10} - p_{01}|$ とする．$z_{1-\alpha/2}$ は標準正規分布の $(1-\alpha/2)$ 分位点，Φ はその分布関数を表す．以下の式の添字 j は計算法を表し，Miettinen 法では $j = \text{Miettinen}$，Connor 法では $j = \text{Connor}$ とする．式 (4) の各方法に対応する v_j を式 (3) に代入すると，両側検定の検出力の近似値が得られる (Miettinen, 1968; Connor, 1987)：

$$\text{Power}_j(N) = \Phi\left(\frac{\delta\sqrt{N} - z_{1-\alpha/2}\sqrt{S}}{v_j}\right) + \Phi\left(\frac{-\delta\sqrt{N} - z_{1-\alpha/2}\sqrt{S}}{v_j}\right), \tag{3}$$

$$v_{\text{Miettinen}}{}^2 = S - \frac{\delta^2(3+S)}{4S}, \quad v_{\text{Connor}}{}^2 = S - \delta^2. \tag{4}$$

Miettinen 法と Connor 法では，片側検定で目標検出力 $1-\beta$ を得るために必要なペア数の近似式が知られている．これらの式の有意水準を $\alpha/2$ に置き換えると，両側検定の必要ペア数に対する次の近似式が得られる：

$$N_{Miettinen} \approx \frac{\left(z_{1-\alpha/2}S + z_{1-\beta}\sqrt{S^2 - \frac{\delta^2(3+S)}{4}}\right)^2}{S\delta^2}, \tag{5}$$

$$N_{Connor} \approx \frac{\left(z_{1-\alpha/2}\sqrt{S} + z_{1-\beta}\sqrt{S-\delta^2}\right)^2}{\delta^2}. \tag{6}$$

ここで $\delta > 0$，$S > 0$ とし，得られた値を整数に切り上げる．式 (3) の第 1 項は，観測された差が想定した方向に現れて帰無仮説を棄却する確率，第 2 項は，差が逆方向に現れて帰無仮説を棄却する確率を表す．式 (5), (6) は，第 2 項を十分小さいものとして省略し，第 1 項を目標検出力に等しいとおいて，必要なペア数 N について解いた近似式である．SAS では，第 1 項と第 2 項の両方を含む式 (3) を用いて，必要なペア数を数値的に求める．本稿では，SAS が算出したペア数を比較に用いた．

3.3　Exact 法による検出力の計算

Exact 法では，二つの不一致セルの度数について，補正なし McNemar 検定で帰無仮説を棄却する組合せの確率を合計し，検出力を求める (Duffy, 1984)．総ペア数 N が同じでも，不一致ペア数 D は標本ごとに変わるため，まず D を固定した場合を考える．

不一致となる確率は S であるため，D は二項分布 $\text{Bin}(N, S)$ に従う．また，不一致ペアがセル $(0,1)$ に入る確率を $q = p_{01}/S$ とすると，$D = d$ のもとで，そのセルの度数 $X = n_{01}$ は二項分布 $\text{Bin}(d, q)$ に従う．もう一方の

セルの度数は残りのペア数で決まる．この分布を使って，式 (1) の検定で帰無仮説を棄却する確率を求め，$g(d,q)$ と表す．$D=0$ の場合は帰無仮説を棄却しないため，$g(0,q)=0$ とする．

不一致ペア数が d となる確率と，そのときに帰無仮説を棄却する確率 $g(d,q)$ を掛け，$d=0$ から N まで合計すると，全体の検出力が得られる：

$$\mathrm{Power}(N) \;=\; \sum_{d=0}^{N} \Pr(D \;=\; d)\; g(d,q)\;. \tag{7}$$

SAS の `DIST=NORMAL METHOD=EXACT` は，式 (7) によって補正なし McNemar 検定の検出力を計算する．第 1 種の過誤確率も同じ手順で計算でき，その際には S を固定したまま，帰無仮説に対応する $q=1/2$ を用いる．

4　SAS によるサンプルサイズ設計と検出力の計算

$p_{01}=0.01$，$p_{10}=0.11$，両側有意水準 $\alpha=0.05$，目標とする検出力を 0.80 とする．介入前の応答を A，介入後の応答を B とし，1 を症状ありとすれば，悪化する確率を 0.01，改善する確率を 0.11 と仮定した例に相当する．

`PAIREDFREQ` 文で求められる量とその指定方法を表 3 に示す．Miettinen 法と Connor 法では，目標検出力を与え，`NPAIRS=.` で必要なペア数を求められる．Exact 法では `NPAIRS=.` を指定できないため，候補となるペア数を与え，`POWER=.` で検出力を計算する．

表 3　`PAIREDFREQ` 文で求められる量と指定方法

METHOD	検出力の計算	必要ペア数の算出
MIETTINEN	POWER=.	NPAIRS=.
EXACT	POWER=.	指定不可
CONNOR	POWER=.	NPAIRS=.

4.1　Miettinen 法と Connor 法によるペア数の算出

`DISCPROPORTIONS` には，$p_{10}\,|\,p_{01}$ の順に確率を指定する．

```
proc power;
  pairedfreq test=mcnemar dist=normal method=miettinen
    discproportions=0.11 | 0.01
    sides=2 alpha=0.05 power=0.80 npairs=.;
  pairedfreq test=mcnemar dist=normal method=connor
    discproportions=0.11 | 0.01
    sides=2 alpha=0.05 power=0.80 npairs=.;
run;
```

算出されるペア数は，Miettinen 法で 77，Connor 法で 92 となる．`NPAIRS` は，応答が一致したペアも含めた総ペア数を表す．

4.2　Exact 法による検出力の計算

以下のコードでは，`NPAIRS=1 TO 150 BY 1` と `POWER=.` を指定し，1〜150 ペアの各整数について検出力を計算する．出力から，検出力が 0.80 以上となる最小のペア数を求める．

```
proc power;
  pairedfreq test=mcnemar dist=normal method=exact
    discproportions=0.11 | 0.01
    sides=2 alpha=0.05
    npairs=1 to 150 by 1 power=.;
run;
```

$N = 78$ では検出力が 0.797，$N = 79$ では 0.802 となる．1〜150 ペアの範囲で目標検出力 0.80 を満たす最小のペア数は 79 であった．

4.3　設計時の検出力と数値評価の比較

表 4 では，各方法が設計時に算出した検出力と，得られたペア数を式 (7) に代入して求めた検出力を比較する．

表 4　$p_{01} = 0.01$，$p_{10} = 0.11$ におけるペア数と補正なし検定の検出力

計算法	N	設計時の検出力	実際の検出力
Miettinen 法	77	0.801	0.791
Exact 法	79	0.802	0.802
Connor 法	92	0.801	0.862

Miettinen 法の設計時の出力は 0.801 であったが，式 (7) による検出力は 0.791 となり，目標値を下回った．Exact 法と Connor 法で求めたペア数では目標値を満たした．

4.4　正確検定を用いる場合

解析に `EXACT` 文の `MCNEM` で求めた正確検定の p 値を用いる場合は，`DIST=EXACT_COND METHOD=EXACT` を指定してサンプルサイズを設計する．以下のコードでは，70〜110 ペアについて正確検定の検出力を計算する．

```
proc power;
  pairedfreq test=mcnemar dist=exact_cond method=exact
    discproportions=0.11 | 0.01
    sides=2 alpha=0.05
    npairs=70 to 110 by 1 power=.;
run;
```

$p_{01} = 0.01$，$p_{10} = 0.11$ として，二項分布から正確検定の棄却確率を計算すると，79 ペアでの検出力は 0.689 となる．1〜200 ペアについて計算した場合，検出力が 0.80 以上となる最小のペア数は 95 であり，その検出力は 0.802 である．同じ確率を仮定しても，式 (1) のカイ二乗近似による検定と正確検定では棄却域が異なるため，必要なペア数も異なる．

5　数値評価の方法

5.1　パラメータの設定

各計算法で求めたペア数と，そのペア数で得られる検出力を，二つの不一致確率 p_{01}，p_{10} を変えて比較した．確率の範囲を定めるにあたり，実際の McNemar 検定の適用例を参考にした．Pembury Smith and Ruxton (2020) は，行動学・生態学・進化学分野の 50 論文における McNemar 検定の利用状況を調査している．本稿では，同論文に掲載された分割表から各不一致セルの割合を算出し，その分布を参考に評価する確率の範囲を定めた．

二つの割合のうち小さい方について，50 件での中央値は 0.029，四分位範囲は 0〜0.061 であった．また，割合の差の絶対値の中央値は 0.344，四分位範囲は 0.205〜0.453 であった．これらをもとに，小さい方の確率 p_{01} を 0.01〜0.10，差 δ を 0.01〜0.60 とし，いずれも 0.01 刻みで設定した．

$p_{10} = p_{01} + \delta$ とし，計 600 条件を設定した．両側有意水準を $\alpha = 0.05$，目標とする検出力を 0.80 とした．SAS の `DIST=NORMAL` を用い，方法ごとにペア数 N を求めた．Miettinen 法と Connor 法で求めたペア数を，それぞれ N_M，N_C とする．Exact 法では，$L = \max\{1, \min(N_M, N_C) - 50\}$ から $U = \max(N_M, N_C) + 100$ までの各整数 N について補正なしの検出力を計算し，0.80 以上となる最小の N を求めた．

5.2　検出力と第 1 種の過誤確率の評価

各方法で求めたペア数を式 (7) に代入して検出力を算出し，目標値 0.80 以上となった条件数と，方法間のペア数の差を比較した．第 1 種の過誤確率は，各条件で求めたペア数と不一致確率の和 S を固定し，帰無仮説に対応する $p_{01} = p_{10} = S/2$ を設定して算出した．計算には SAS/STAT 15.3 を使用した．

6　数値評価の結果

6.1　各方法で求めたペア数と検出力

表 5 は 600 条件の結果をまとめたものである．「目標達成」は検出力が 0.80 以上となった条件数（割合）を表す．ペア数の差は，各方法のペア数から Exact 法のペア数を引いた値とし，その最小値と最大値を示した．そして，各計算法について，600 条件における検出力の分布を図 1 に示す．なお，図 1 は SGPLOT プロシジャで作成しており，ダイヤマークは平均値を表している (魚住, 2026).

Miettinen 法では，224 条件 (37.3%) で実際の検出力が目標値 0.80 を下回った．検出力が最も低かったのは $p_{01} = 0.10$，$p_{10} = 0.68$ の条件であり，Miettinen 法による必要ペア数は 15 と算出されたものの，そのペア数における実際の検出力は 0.783 にとどまった．

一方，Connor 法では，全 600 条件 (100.0%) で目標検出力を満たしたものの，559 条件 (93.2%) で Exact 法より多いペア数が算出された．Miettinen 法で求めたペア数を Exact 法と比較すると，少ない条件が 224，多い条件が 159，同じ条件が 217 であった．

図 2 に，式 (7) で求めた検出力を，$p_{01} = 0.01, 0.04, 0.07, 0.10$ の 4 水準について示す．横軸は不一致確率の差 δ，縦軸は各方法で求めたペア数における検出力である．線は各条件の計算値を不一致確率の差の順に結んだ．灰色の破線は目標検出力 0.80 を示す．Exact 法の検出力はおおむね 0.80 付近で推移した一方，Connor

法では目標値を大きく上回る条件がみられた．Miettinen 法の検出力は不一致確率の差に応じて上下し，0.80 を下回る条件もみられた．

表 5　600 条件における補正なし検定の検出力とペア数の差

計算法	目標達成	平均検出力	検出力の範囲	ペア数の差	
				最小値	最大値
Miettinen	376 (62.7%)	0.806	0.783–0.889	−7	11
Exact	600 (100.0%)	0.808	0.800–0.849	0	0
Connor	600 (100.0%)	0.841	0.800–0.934	0	59

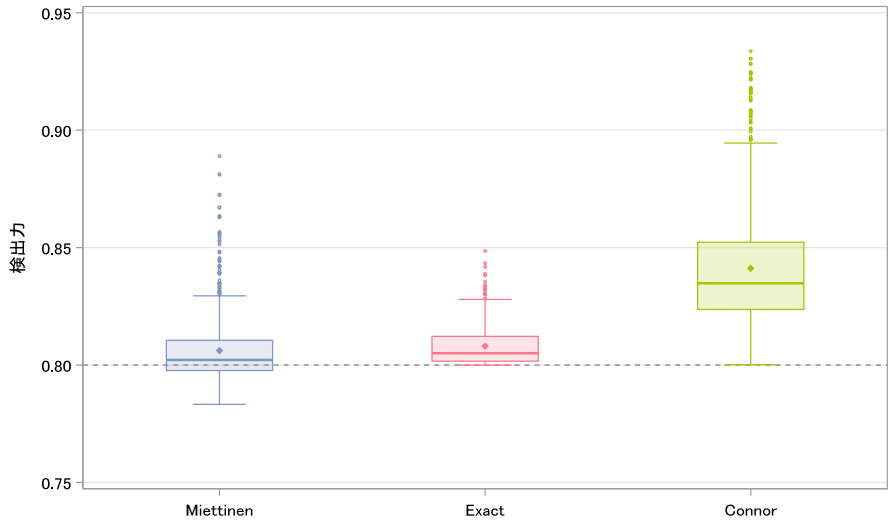


図 1　600 条件における各計算法の検出力の分布

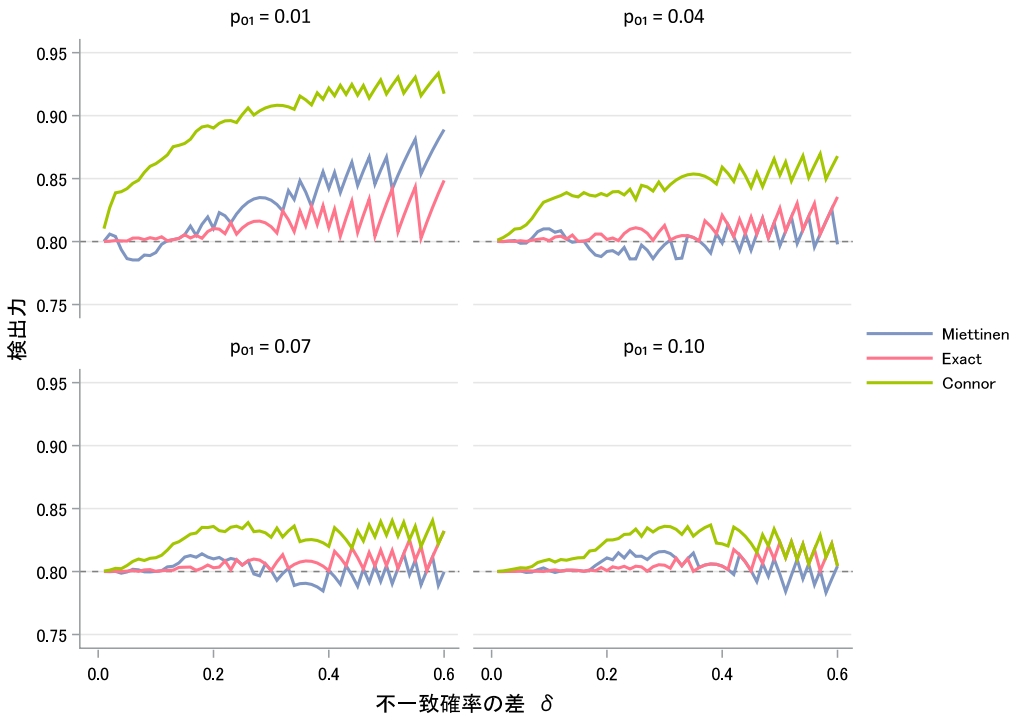


図 2　不一致確率の差に対する検出力の変化

6.2　各方法における第 1 種の過誤確率

Miettinen 法では，p_{01} の全 10 水準で検出力が 0.80 を下回る条件が認められた．各水準 60 条件のうち，その条件数は 8〜34 条件であり，最多は $p_{01} = 0.06$ の 34 条件であった．図 3 では，各方法で求めたペア数における第 1 種の過誤確率を，同一条件ごとに比較した．各点は 1 条件を表し，横軸は Exact 法，縦軸は Miettinen 法または Connor 法である．対角線は両方法の値が一致する位置，破線は有意水準 0.05 を示す．

第 1 種の過誤確率は，3 種類の計算法全体で 0.0409〜0.0536 となり，平均値は Miettinen 法 0.0464，Exact 法 0.0467，Connor 法 0.0465 であった．図 3（左）では多くの点が対角線付近に位置した．図 3（右）では点が対角線の上下に分布し，Connor 法と Exact 法の大小関係は条件によって異なった．同一条件における Exact 法との差の絶対値は，Miettinen 法で最大 0.0037，Connor 法で最大 0.0082 であった．いずれの方法にも 0.05 を上回る条件が認められた．

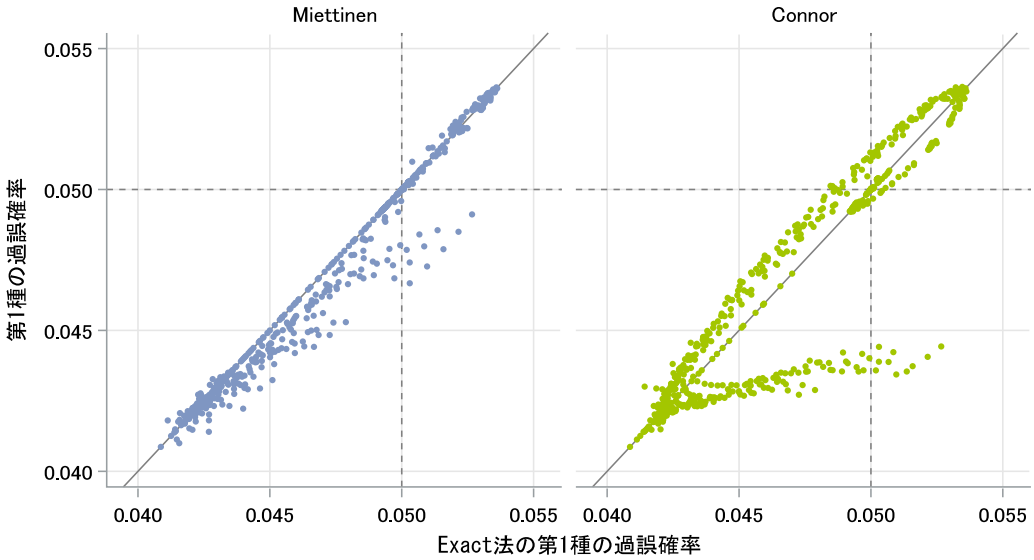


図 3　同一条件における第 1 種の過誤確率の比較

7　考察

既報では，Miettinen 法は必要ペア数を小さめに，Connor 法は大きめに見積もる傾向が指摘されている (Lachin, 1992)．両方法で用いる式 (4) の分散項の違いは，設計時の検出力と必要ペア数に差をもたらす．本稿における Miettinen 法の検出力不足と，Connor 法で Exact 法より多いペア数が算出された結果は，これらの近似法の特徴と整合していた．

図 3 に示した第 1 種の過誤確率の差は，共通の補正なし McNemar 検定に対して，各方法が異なるペア数を選択したことによる．不一致セルの度数は整数であり，棄却域に含まれる度数の組合せとその確率はペア数に応じて変わる．実際に，Connor 法で Exact 法よりペア数が増えた条件でも，第 1 種の過誤確率が高くなる場合があった．Exact 法は，仮定した確率のもとで補正なし検定の検出力を二項分布から計算する．表 2 に示したように，本稿で評価した Exact 法は，解析に用いる棄却域は二項分布に基づき計算するわけではなく，カイ二乗近似に基づくため，目標検出力を満たすペア数についても，第 1 種の過誤確率を別途確認する必要がある．

本稿の評価範囲は，両側有意水準$\alpha = 0.05$，目標検出力 0.80，p_{01}の範囲 0.01～0.10，δの範囲 0.01～0.60 における補正なし McNemar 検定に限られる．目標達成割合などの集計値は，設定した 600 条件に等しい重みを与えた結果であり，この検討範囲内での比較として解釈すべきである．

8　おわりに

SAS の `DIST=NORMAL` で利用できる 3 種類の計算法を 600 条件で比較した．Miettinen 法では実際の検出力が目標値を下回る条件があり，Connor 法では Exact 法より多いペア数が算出される傾向を確認した．

補正なし McNemar 検定の設計では，近似法で得たペア数を起点として，Exact 法で候補となるペア数ごとの検出力を再計算し，目標値を満たすペア数を選ぶ手順を提案する．選択したペア数について，第 1 種の過誤確率も併せて評価することが求められる．

参考文献

[1] Connor, R. J. Sample size for testing differences in proportions for the paired-sample design. Biometrics. 43:207–211, 1987.

[2] Duffy, S. W. Asymptotic and exact power for the McNemar test and its analogue with R controls per case. Biometrics. 40:1005–1015, 1984.

[3] Hanai, A., Ishiguro, H., Sozu, T., Tsuda, M., Yano, I., Nakagawa, T., Imai, S., Hamabe, Y., Toi, M., Arai, H. and Tsuboyama, T. Effects of cryotherapy on objective and subjective symptoms of paclitaxel-induced neuropathy: Prospective self-controlled trial. Journal of the National Cancer Institute. 110:141–148, 2018.

[4] Lachin, J. M. Power and sample size evaluation for the McNemar test with application to matched case-control studies. Statistics in Medicine. 11:1239–1251, 1992.

[5] Miettinen, O. S. The matched pairs design in the case of all-or-none responses. Biometrics. 24:339–352, 1968.

[6] Pembury Smith, M. Q. R. and Ruxton, G. D. Effective use of the McNemar test. Behavioral Ecology and Sociobiology. 74:133, 2020.

[7] 魚住龍史 監修．臨床評価研究会基礎解析分科会 著．実用 SAS 生物統計グラフハンドブック．東京：サイエンティスト社，2026.

[8] 浜田知久馬，魚住龍史 監修．臨床評価研究会基礎解析分科会 著．新版 実用 SAS 生物統計ハンドブック．初版第 2 刷．東京：サイエンティスト社，2021.

傾向スコアマッチングによる群間バランスの視覚的評価

○魚住 龍史
東京科学大学工学院経営工学系

Visual assessment of between-group balance before and after propensity score matching

Ryuji Uozumi [1, 2]
[1] *Department of Industrial Engineering and Economics, Institute of Science Tokyo*
[2] *Total Health Design Visionary Initiative, Institute of Science Tokyo*

要旨

観察研究では，ランダム化が行われていないことから，群間比較としてはバイアスが混在した結果となってしまう場合がある．比較可能性を担保する方法のひとつとして，傾向スコアによるマッチングが挙げられる．SAS/STAT では，PSMATCH プロシジャが利用可能であり，傾向スコアによる分析を行うためのさまざまなツールが提供されている．本稿では，傾向スコアマッチングによって，群間バランスが改善されたかどうかの評価に焦点を当てる．群間バランスの評価として，定量的な方法には標準化平均差が挙げられる．その一方，視覚的な方法として，PSMATCH プロシジャではいくつかのプロットを出力できる．しかし，群間バランスの視覚的な評価方法としてよく用いられているミラーヒストグラムに関して，PSMATCH プロシジャでは出力できず，SGPLOT プロシジャなどを用いても容易には作成できない．2026 年 3 月にサイエンティスト社より出版された『実用 SAS 生物統計グラフハンドブック』では，ミラーヒストグラムを作成するためのプログラムが解説されている．そこで本稿では，ミラーヒストグラムを用いる動機付けを論じたうえで，ミラーヒストグラムの作成方法について詳説する．

キーワード：SAS グラフィックス　傾向スコア　テンプレート　マッチング　ミラーヒストグラム
ASSESS　ODS GRAPHICS　PSMATCH　SGPLOT　SGRENDER

1　はじめに

医薬品開発では，ランダム化比較試験によりデザインされ，統計解析が行われることが多い．しかし，ランダム化ができない医学研究では，観察研究 (observational study) によりデザインされる．観察研究では，ランダム化比較試験と異なり比較可能性が担保されず，群間比較としてはバイアスが混在した結果となってしまう．比較可能性を担保する方法として，マッチング，層別解析，および回帰モデルによる解析などが挙げられる．

本稿では，傾向スコアを用いたマッチングに焦点を当てる．観察研究において，試験治療と対照治療の比較を行うと仮定すると，背景因子で条件付けたもと，試験治療を受ける確率が傾向スコアである．

SAS/STAT では，PSMATCH プロシジャが利用可能であり，傾向スコアによる分析を行うためのさまざまなツールが提供されている．PSMATCH プロシジャでは，大きく分けると図 1 のような役割を持つステートメントが利用可能である (魚住ら, 2017)．これらのステートメントのうち，ASSESS ステートメントに焦点を当てると，指定した変数のマッチング前後の群間差を評価できる．ASSESS ステートメントで出力されるアウトプットとしては，標準化群間差 (standardized mean difference, SMD) であり，単純な群間の平均値の差を分子，各群の標準偏差をプールした標準偏差を分母とした統計量である．ODS GRAPHICS による出力結果では，標準化群間差を横軸とするグラフも出力される．さらに，ODS GRAPHICS による機能では，ASSESS ステートメントで指定した変数のクラウドプロット，箱ひげ図，度数分布も出力される．

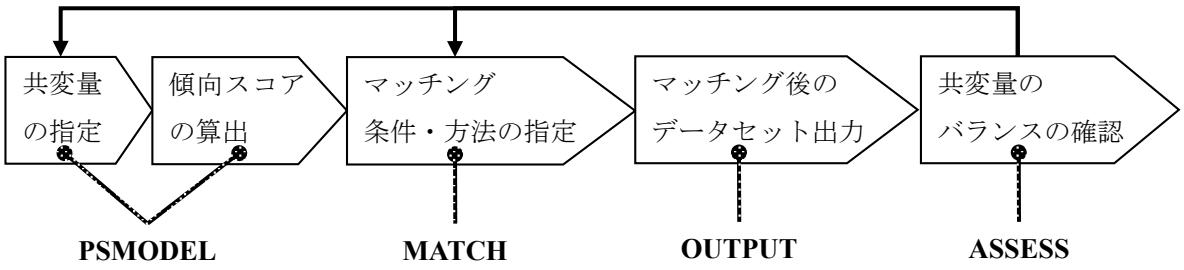


図 1： PSMATCH プロシジャの主なステートメントの役割 (魚住ら, 2017)

傾向スコアマッチング前後の群間バランスとして，標準化群間差のような定量的評価ではなく，クラウドプロットのような視覚的評価に着目すると，実際の観察研究ではどの程度適用されているだろうか．著者が調べたところ，残念ながらクラウドプロットなどで群間バランスの視覚的評価が行われている観察研究を見つけられなかった．群間バランスの視覚的評価としては，マッチング前後の傾向スコアのヒストグラムが示されていることが多いと感じている．このようなヒストグラムはミラーヒストグラム (mirror histogram) と呼ばれ，群間バランスが保たれている場合，上下対称あるいは左右対称のヒストグラムになるため，マッチングによって群間バランスが改善されているか視覚的に評価できる．しかし，残念ながら，PSMATCH プロシジャでミラーヒストグラムを出力することはできない．SAS でミラーヒストグラムを作成するために，PSMATCH プロシジャの結果を出力データに保存し，SGPLOT プロシジャなどの統計グラフを用いてみるとどうだろうか．何年も前の話になるが，著者自身，ミラーヒストグラムを出力する必要がある場面があったが，SGPLOT プロシジャや SGPANEL プロシジャでミラーヒストグラムを作成することは容易ではなかった．

さて，2026 年 3 月にサイエンティスト社より『実用 SAS 生物統計グラフハンドブック』が出版され，本稿の著者も監修の立場で参画した．書籍の第 2 章 第 6 節はヒストグラムとして構成されており，2 群のヒストグラムとしてミラーヒストグラムの作成プログラムが掲載されている．そこで本稿では，ミラーヒストグラムによる傾向スコアマッチング前後の群間バランスが行われた事例を議論したうえで，SAS によるミラーヒストグラムの作成方法について詳説する．

本稿の構成は次の通りである．2 節では，傾向スコアマッチング前後の群間バランスを評価した観察研究の事例を導入する．3 節では，傾向スコアマッチングの概説を踏まえたうえで，SAS で実施可能なマッチング前後の群間バランスの評価方法についての現状を議論する．4 節では，SAS によるミラーヒストグラム作成方法として，3 種類のアプローチについて議論する．最後に，5 節では，SAS 社への要望を示し，本稿のまとめを述べる．

2　傾向スコアマッチング前後の群間バランス評価事例

2.1　ミラーヒストグラムによる群間バランスの評価

新型コロナウイルス感染症入院患者に対する観察研究では，気管挿管または死亡が発生するまでの時間を主要評価項目として，ヒドロキシクロロキン投与群と非投与群の群間比較が行われた (Geleris et al., 2020). 観察研究のため，傾向スコアによる統計解析が行われ，Cox の比例ハザードモデルによるハザード比が推定された．主たる解析は傾向スコアによる重み付き解析であったものの，傾向スコアマッチングも実施されていた．傾向スコアマッチング後の対象集団に対するハザード比の出力前に，マッチングによる群間バランスの評価として，ミラーヒストグラムが出力されていた．ミラーヒストグラムは，マッチング前後の 2 つがパネル別に示され，SAS でたとえるならば，SGPANEL プロシジャの出力結果のようになっていた．ミラーヒストグラムにもとづき，マッチング前は上下非対称であったものの，マッチング後は上下対称のヒストグラムになっていることが確認された．ただし，本研究では，標準化群間差のような定量的なアプローチでマッチング前後の群間バランスの評価は行われていなかった．

2.2　ミラーヒストグラムと標準化群間差による群間バランスの評価

悪性胸膜中皮腫患者に対する観察研究では，死亡までの時間を主要評価項目として，手術ベースの集学的治療と手術単独群の群間比較が行われた (Nelson et al., 2017)．傾向スコアマッチング後の対象集団に対して，Cox の比例ハザードモデルによるハザード比が推定された．傾向スコアマッチングによる群間バランスの評価として，ミラーヒストグラムが出力されていた．ミラーヒストグラムはマッチング前後で別々の図で示されていた．ミラーヒストグラムにもとづき，マッチング前は上下非対称であったものの，マッチング後は上下対称のヒストグラムになっていることが確認された．さらに，標準化群間差による定量的な評価も行われ，縦軸に共変量，横軸に標準化群間差をとる散布図として示された．各共変量の群間バランスとして，マッチング後の標準化群間差が $\pm 10\%$ 以内であれば，その共変量の群間バランスは保たれていると判断された．結果，14 個の共変量について，マッチング後の標準化群間差はすべて $\pm 10\%$ 以内であることが確認された．加えて，P 値も出力されていたが，上野・魚住 (2024) で議論しているように，P 値よりも標準化群間差を使用すべきといえる．

3　傾向スコアマッチングによる群間バランス評価の現状

3.1　傾向スコアマッチングの概説

はじめに，X_i を被験者 $i \in \{1,2,\cdots,N\}$ で観測される共変量ベクトル，Z_i を被験者 i の群を表す指示変数 ($Z_i = 1$：治療群，$Z_i = 0$：対照群) とする．このとき，X_i で条件付けたもとで治療群である確率 $e_i = \Pr(Z_i = 1 \mid X_i)$ が傾向スコアと定義される (Rosenbaum and Rubin, 1983)．傾向スコア e_i で条件付けたもとで，各群で観察される共変量の分布が似通っていることが望まれる ($Z_i \perp X_i \mid e_i$)．PSMATCH プロシジャでは，PSMODEL ステートメントにより傾向スコアの推定が行われる．

傾向スコアによるマッチングを行う場合，治療群と対照群における被験者のうち，傾向スコアが似通った被験者をペアとして，新たな解析対象集団を構成する．傾向スコアモデルに連続変数を含めた場合，ある被験者と全く同一の傾向スコアを持つ被験者を見つけることは困難であるためであるため，被験者の傾向スコア間の差異を測るための尺度が必要となる．被験者 i と被験者 j の傾向スコア間の差異を測る尺度とし

て，傾向スコア自体の差 $|e_i - e_j|$ や傾向スコアのロジット変換した値の差 $|\mathrm{logit}(e_i) - \mathrm{logit}(e_j)|$ が用いられる．

3.2 SAS による傾向スコアマッチング

PSMATCH プロシジャを用いて，傾向スコアによるマッチングをどのように実施し，群間バランスの評価として，どのような結果が出力されるか示す．対象データとして，観察研究のデータ (データセット名：Drugs) を用いる ($n = 486$)．データセット Drugs は，新規治療を行った「治療群」(変数：Drug = 'Drug_X') と標準治療を行った「対照群」(変数：Drug = 'Drug_A') を比較するために収集されたデータであるが，被験者あるいは主治医がどちらの治療を行うか選択できたため，観察研究のデータとして扱う．ベースラインの被験者背景として，年齢 (変数名：Age)，性別 (変数名: Gender; Gender = 'Male'「男性」，'Female'「女性」)，BMI (変数名 Bmi) が挙げられる．結果の解釈を述べる．

PSMATCH プロシジャによる各ステートメントの役割は図 1 の通りであるが，それぞれのステートメントの詳細は，魚住ら (2017) を参照されたい．本稿で焦点を当てているのは，傾向スコアマッチングによる群間バランスの評価である． PSMATCH プロシジャの ASSESS ステートメントで，すべての解析対象データセット，あるいはマッチされた対象データセットに対して，指定した変数の群間差を評価できる．プログラム 1 では，MATCH ステートメントで傾向スコアマッチングを実施している．ASSESS ステートメントで LPS を指定して，傾向スコアをロジット変換した値の群間差を評価する．もし PS も指定すれば，傾向スコアの群間差も評価可能である．さらに，VAR = で指定した変数についても，群間差の評価を行う．また，VARINFO オプションを指定することで，各群に対する変数情報が出力される (表 2)．プログラム 1 では，最後に，PLOTS = で ODS GRAPHICS による出力結果を得ている．

プログラム 1：PSMATCH プロシジャにおける ASSESS ステートメントの実行例

```
ods graphics on;
proc psmatch data=Drugs region=allobs;
   class Drug Gender;
   psmodel Drug(treated='Drug_X')= Gender Age Bmi;
   match method=greedy(k=1 order=descending)
         stat=lps caliper(mult=stddev)=0.20;
   assess lps var=(Gender Age Bmi)
         / weight=none varinfo plots=(all);
   output out(obs=all)=psmatch00 ps=ps;
run;
```

表 1 では，傾向スコアの情報を一部のみ示しており，Observation = All と Observation = Matched の結果を比較すれば，データセットのうち，どの程度のオブザベーションでペアが構成されたか確認できる．データセット Drugs の場合，治療群：113 名，対照群：373 名のうち，治療群：113 名，対照群：113 名がマッチされたことがわかる．表 2 は VARINFO オプションを指定して出力される内容で，ASSESS ステートメントで指定した変数情報の要約統計量が出力される．性別のようなカテゴリカルデータの場合，Mean の列には，

平均の代わりに，最初の順番のカテゴリにおける割合が出力される．表 2 における性別 Gender では，Female「女性」の割合が出力されている．

表 1：傾向スコアの情報

Propensity Score Information						
Observations	Treated (Drug = Drug_X)			Control (Drug = Drug_A)		
	N	Mean	Std Dev	N	Mean	Std Dev
All	113	0.310773	0.132467	373	0.208801	0.131969
Region	113	0.310773	0.132467	373	0.208801	0.131969
Matched	113	0.310773	0.132467	113	0.308755	0.130649

表 2：変数情報

Variable Information							
Variable	Observations	Treated (Drug = Drug_X)			Control (Drug = Drug_A)		
		N	Mean	Std Dev	N	Mean	Std Dev
LPS	All	113	-0.880615	0.681761	373	-1.520586	0.844486
	Region	113	-0.880615	0.681761	373	-1.520586	0.844486
	Matched	113	-0.880615	0.681761	113	-0.889200	0.674687
Age	All	113	36.309735	5.534114	373	40.404826	6.579103
	Region	113	36.309735	5.534114	373	40.404826	6.579103
	Matched	113	36.309735	5.534114	113	35.884956	5.468664
Bmi	All	113	24.492566	1.863797	373	23.753271	1.980778
	Region	113	24.492566	1.863797	373	23.753271	1.980778
	Matched	113	24.492566	1.863797	113	24.309027	1.882123
Gender	All	113	0.433628	0.495575	373	0.458445	0.498270
	Region	113	0.433628	0.495575	373	0.458445	0.498270
	Matched	113	0.433628	0.495575	113	0.495575	0.499980

表 3 では，ASSESS ステートメントで指定した変数に対して，標準化群間差が出力される．Mean Difference は単純な群間の平均値の差であり，Standardized Mean Difference は Mean Difference / Divisor の計算結果である．Divisor 「除数」は各群の標準偏差をプールした標準偏差で構成される．例えば，年齢を表す変数 Age の All Obs「すべてのオブザベーション」を対象とした出力の場合，表 2 で求めた各群の標準偏差から，

$$\sqrt{\frac{(5.534114^2 + 6.579103^2)}{2}} = 6.079104$$

が Divisor となり，表 3 に出力された Mean Difference より，

$$\frac{\text{Mean Difference}}{\text{Divisor}} = \frac{-4.095091}{6.079104} = -0.673634$$

が Standardized Mean Difference として出力されている．ODS GRAPHICS ON の状態で実行しているので，表 3 で示した各変数に対する標準化群間差に対応するグラフも出力される (図 2)．図 2 では，± 0.25 がカットオフ値として参照線が出力され，$[-0.25, 0.25]$ であればバランスが取れているという目安として表している．カットオフ値として ± 0.10 を推奨する報告も行われている (Normand et al., 2001)．2.2 節の事例では，カットオフ値として ± 0.10 のもと，縦軸に 14 個の変数，横軸に標準化群間を示したグラフを用いて，マッチング前後の群間バランスを評価している．

表 3：各変数に対する標準化群間差

Standardized Variable Differences (Treated - Control)							
Variable	Mean Difference			Standardized Mean Difference			
				Divisor	Mean Difference		
	All Obs	Region Obs	Matched Obs		All Obs	Region Obs	Matched Obs
LPS	0.639971	0.639971	0.008584	0.767448	0.833894	0.833894	0.011185
Age	-4.095091	-4.095091	0.424779	6.079104	-0.673634	-0.673634	0.069875
Bmi	0.739296	0.739296	0.18354	1.923178	0.384414	0.384414	0.095436
Gender	-0.024817	-0.024817	-0.061947	0.496925	-0.049941	-0.049941	-0.124661

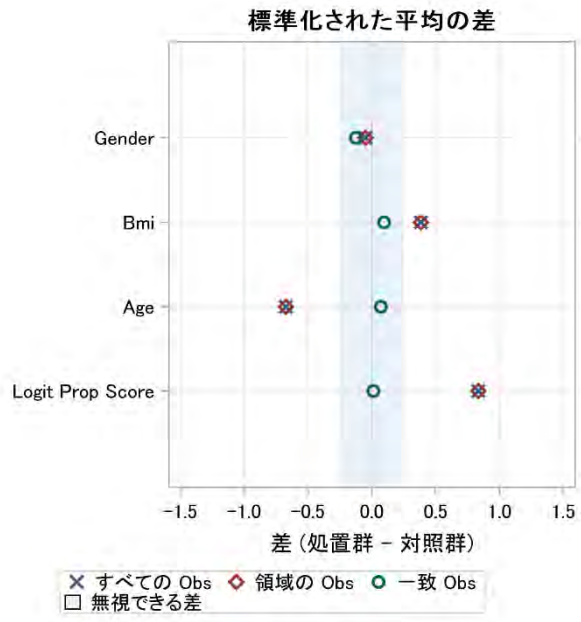


図 2：各変数に対する標準化群間差 (ODS GRAPHICS)

ASSESS ステートメントで PLOTS = ALL を指定することで，標準化群間差のプロット以外にも，ODS GRAPHICS による出力結果として，連続変数のクラウドプロット，箱ひげ図，累積分布，カテゴリカル変数の度数分布を得ることができる (魚住ら, 2017)．しかし，現在の PSMATCH プロシジャでは，ミラーヒストグラムが出力されないことが問題点として挙げられる．

4　SAS によるミラーヒストグラムの作成

本稿では，ミラーヒストグラムの作成方法として，3 種類のアプローチを議論する．いずれも PSMATCH プロシジャの出力結果を活用する．プログラム 1 を実行させると，出力データセット PSMATCH00 が出力される．OUTPUT ステートメントにおいて，OBS＝ALL を指定しているので，データセット PSMATCH00 にはマッチされなかったデータセットも含まれている．その後の手順として，プログラム 2 を実行させることで，マッチング前後のデータセットをどちらも含んだデータセット HIST00 を準備する．データセット HIST00 では，マッチング後のデータセット HIST01 とマッチング前のデータセット HIST02 を準備し，これらのデータセットをひとつにまとめている．

プログラム 2：PSMATCH プロシジャの出力データセットをデータハンドリング

```
data hist01;set psmatch00; where _MATCHWGT_^=.; match=1;
data hist02;set psmatch00; match=0;
data hist00; set hist01 hist02; run;
```

SAS プログラム

```
proc sgpanel data=hist00 noautolegend;
  panelby match Drug
    / novarname onepanel layout=lattice;
  histogram ps /  name='hist' group=Drug
            binwidth=0.02 transparency=0.5 ;
  rowaxis label='Proportion (%)' ;
  colaxis values=(0 to 1 by 0.2) label='Propensity score';
format match matchf. ;
run;
```

出力結果

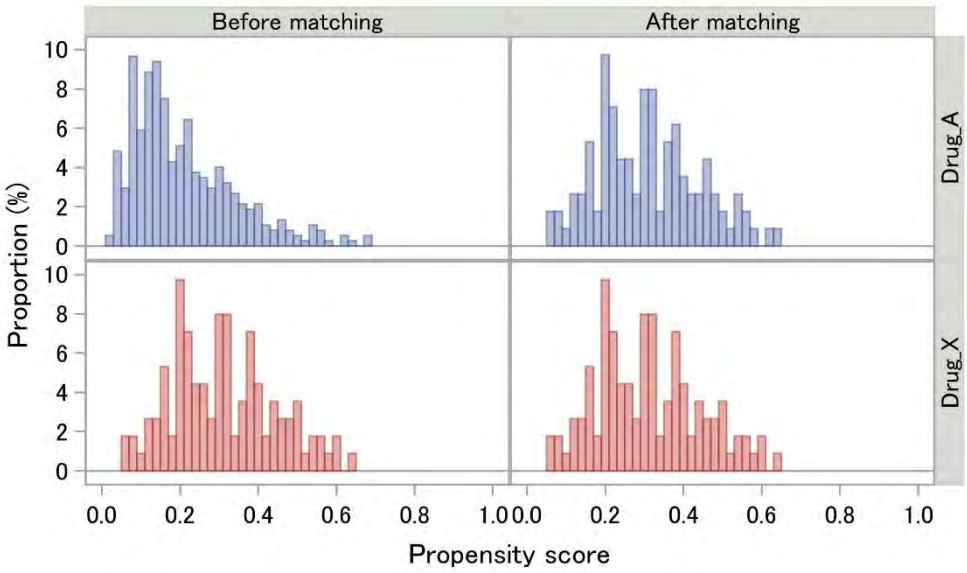


図 3：SGPANEL プロシジャによるミラーヒストグラムの作成 (方法 1)

4.1　SGPANEL プロシジャによるミラーヒストグラムの作成

2.1 節の事例のように，マッチング前後の 2 つミラーヒストグラムをパネル別に示すためには，SGPANEL プロシジャが有用と考えられる．SGPANEL プロシジャによるプログラムと出力結果を図 3 にまとめている．図 3 の出力結果は，縦軸に割合，横軸に傾向スコアをとったヒストグラムを表している．図 3 の出力結果にもとづき，マッチング前は分布が大きく異なっている一方，マッチング後は分布が似通ったヒストグラムになっていることを確認できる．しかし，上下対称のミラーヒストグラムには作成できていないことがわかる．ただし，マッチング前後の群間バランスの評価に必要なヒストグラムとしては出力できているので，ミラーにこだわらなければ，図 3 のアプローチで問題ないといえる．

プログラム 3 : テンプレート MYSTYLE1 の作成プログラム

```
proc template;
  define statgraph mystyle1;
    begingraph;
      layout lattice / order=rowmajor columndatarange=union rowgutter=0;
        columnaxes; columnaxis / label="Propensity Score"; endcolumnaxes;
        layout overlay / yaxisopts=(label="Count");
          histogram eval(ifn(TRTN=1, AVAL,.))
              / binstart=0 binwidth=0.05 binaxis=false
               scale=count fillattrs=(color=gray);
          entry halign=center "Drug_X" / valign=top;
        endlayout;
        layout overlay / yaxisopts=(label="Count" reverse=true);
          histogram eval(ifn(TRTN=2, AVAL,.))
              / binstart=0 binwidth=0.05 binaxis=false
                scale=count fillattrs=(color=white);
          entry halign=center "Drug_A" / valign=bottom;
        endlayout;
      endlayout;
    endgraph;
  end;
run;
```

4.2　テンプレートを用いたミラーヒストグラムの作成

魚住・浜田 (2011) では，テンプレートを定義したうえで，SGRENDER プロシジャにテンプレート情報を参照させて，Kalan-Meier プロットを作成する方法が示されている．同様の方法で，テンプレートを定義したうえで，ミラーヒストグラムの作成することが可能である．『実用 SAS 生物統計グラフハンドブック』の第 2 章 第 6 節では，テンプレートを定義したうえで，SGRENDER プロシジャでミラーヒストグラムを作成するプログラム例が示されている．プログラム 3 は，『実用 SAS 生物統計グラフハンドブック』のプログラ

ムを改変して，ミラーヒストグラムを作成するためのテンプレート MYSTYLE1 を作成するプログラムである．

テンプレート MYSTYLE1 を参照して，SGRENDER プロシジャでミラーヒストグラムを作成するプログラムと出力結果を図 4 に示す．図 4 のミラーヒストグラムでは，マッチング前のデータセットのみを用いているため，上下非対称のヒストグラムになる．群を表す変数 TRTN とヒストグラム作成の対象となる変数 AVAL については，書籍『実用 SAS 生物統計グラフハンドブック』のプログラムに合わせたデータセット SGRENDER00 を準備したうえで，SGRENDER プロシジャを実行している．テンプレート MYSTYLE1 では，LAYOUT LATTICE を指定することで，格子の形を定義して，上下に 2 つのヒストグラムが出力されるようにしている．LAYOUT OVERLAY が 2 回記述されているが，2 回目の LAYOUT OVERLAY 内の YAXISOPTS で REVERSE = TRUE を指定することで，2 回目に指定したヒストグラムは上下逆に出力するように設定している．

なお，図 3 の出力結果では，縦軸を割合とした．これに対して，図 4 の出力結果では，書籍『実用 SAS 生物統計グラフハンドブック』で公開されているプログラムに合わせて，縦軸を頻度としている．図 4 の出力結果より，マッチング前のヒストグラムが作成できたものの，縦軸の出力範囲がそれぞれの群で異なっていることがわかる．そこで，縦軸の出力範囲を群に依らずに共通になるように設定したうえで，ミラーヒストグラムを作成するために，新たにテンプレート MYSTYLE2 を作成する．

SAS プログラム

```
data sgrender00;set psmatch00;
  if Drug="Drug_X" then TRTN=1;
  else if Drug="Drug_A" then TRTN=2;
  AVAL=ps;
run;
proc sgrender data=sgrender00 template=mystyle1; run;
```

出力結果

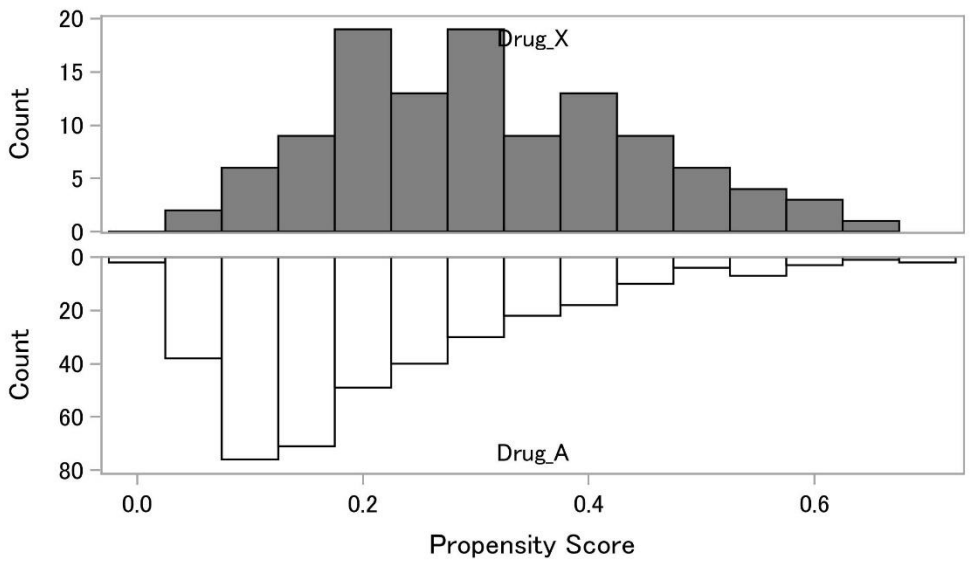


図 4：SGRENDER プロシジャによるミラーヒストグラムの作成 (方法 2)

プログラム 4 : テンプレート MYSTYLE2 の作成プログラム

```
proc template;
  define statgraph mystyle2;
    begingraph;
      layout lattice / order=rowmajor columndatarange=union rowgutter=0;
        columnaxes;
          columnaxis
            / linearopts=(viewmin=0 viewmax=1
              tickvaluesequence=(start=0 end=1 increment=0.2))
              display=(tickvalues);
        endcolumnaxes;
        layout overlay /
          yaxisopts=(linearopts=(viewmin=0 viewmax=80
                     tickvaluesequence=(start=0 end=80 increment=10) )
                     display=(tickvalues label) label="Count");
          histogram eval(ifn(TRTN=1, AVAL,.))
               / binstart=0 binwidth=0.05 binaxis=false
                 scale=count fillattrs=(color=gray);
          entry halign=center "Drug_X" / valign=top;
        endlayout;
        layout overlay /
          yaxisopts=(linearopts=(viewmin=0 viewmax=80
                     tickvaluesequence=(start=0 end=80 increment=10) )
                     display=(tickvalues label) label="Count" reverse=true);
          histogram eval(ifn(TRTN=2, AVAL,.))
                / binstart=0 binwidth=0.05 binaxis=false scale=count
                  fillattrs=(color=white);
          entry halign=center "Drug_A" / valign=bottom;
        endlayout;
      endlayout;
    endgraph;
  end;
run;
```

プログラム 4 は，テンプレート MYSTYLE2 を作成するためのプログラムである．テンプレート MYSTYLE2 では，COLUMNAXIS 内で横軸，YAXISOPTS 内で縦軸の設定を行っている．図 4 のプログラムにおいて，参照するテンプレートを MYSTYLE2 に変更すれば，同様のプログラムでミラーヒストグラムを作成できる．プログラム 4 のテンプレートを参照して作成したミラーヒストグラムを図 5 に示す．図 4 のミ

ラーヒストグラムと異なり，図 5 のミラーヒストグラムでは，マッチング前後のヒストグラムで縦軸の出力範囲が同じになっていることがわかる．

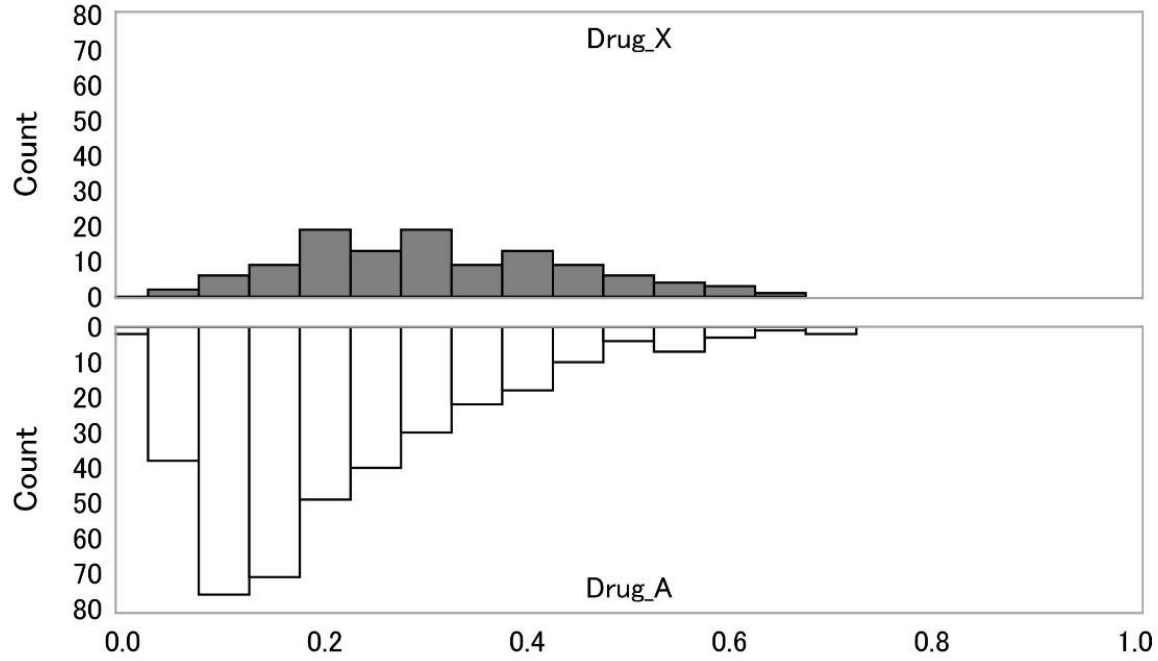


図 5：SGRENDER プロシジャによる軸の設定を加えたミラーヒストグラム

4.3　SGPANEL プロシジャによるミラーヒストグラムの作成を再考

4.2 節では，テンプレートを定義したうえで，SGRENDER プロシジャを実行することで，ミラーヒストグラムを作成することができた．ただ，テンプレートを定義するためのプログラムが複雑であり，容易に作成できるものではない．Kaplan-Meier プロット作成プログラムを例にあげても，魚住・浜田 (2011) ではテンプレートを定義した SGRENDER プロシジャによる方法が述べられているものの，著書 (大橋ら, 2016, 2022) にはテンプレートを定義した作成方法は深く言及されていなく，手軽に実施できない方法だといえる．その一方，4.1 節のミラーヒストグラムでは，テンプレートを定義する手間もなく，マッチング前後の群別のヒストグラムを作成することができたので，マッチング前後の群間バランスの評価は可能であるが，群別のヒストグラムがミラーのように上下対称になっていないことが課題であった．そこで，一旦 SGPANEL プロシジャでマッチング前後の群別のヒストグラムを作成するためのデータセットを出力し，データハンドリングしたうえで，改めて SGPANEL プロシジャを実行すれば，ミラーヒストグラムを作成することが可能である．

図 6 は，SGPANEL プロシジャによる実行結果をデータセットに出力したうえで，片方の群におけるヒストグラムの縦軸の値のみマイナスになるよう -1 倍している．SGPANEL プロシジャでヒストグラムを作成した際に出力したデータセット SGPANEL00 では，頻度あるいは割合を表すヒストグラムの縦軸は BIN_PS_BINSTART__BINSTART__B___Y，傾向スコアを表すヒストグラムの横軸は BIN_PS_BINSTART__BINSTART__B___X という変数名で格納されている．図 6 では，これらの複雑に自動生成された変数名を分かりやすく定義し直したうえで，再び SGPANEL プロシジャを実行している．テンプレートを定義して作成した図 5 のグラフでは，マッチング前のミラーヒストグラムのみ出力しており，マッ

チング後のミラーヒストグラムも作成する必要がある．しかし，図 6 のミラーヒストグラムでは，マッチング前後の 2 つのグラフをパネルに分けて出力しており，テンプレートを定義する必要もない．

SAS プログラム

```
data sgpanel00;set sgpanel00;
  ps=BIN_PS_BINSTART__BINSTART__B___X;
  if Drug="Drug_X" then
       percent=-BIN_PS_BINSTART__BINSTART__B___Y;
  else if Drug="Drug_A" then
       percent=BIN_PS_BINSTART__BINSTART__B___Y;
run;
proc sgpanel data=sgpanel00 noautolegend;
  panelby match  / novarname onepanel  layout=columnlattice;
  vbar ps /  name='hist' group=Drug response=percent
             transparency=0.5  ;
  rowaxis label='Proportion (%)' offsetmax=0.03
          values=(-12 to 12 by 2);
  colaxis display=(novalues noticks) valueshint
          label='Propensity score';
  keylegend 'hist' / noborder position=right;
  format match matchf.  ;
run;
```

出力結果

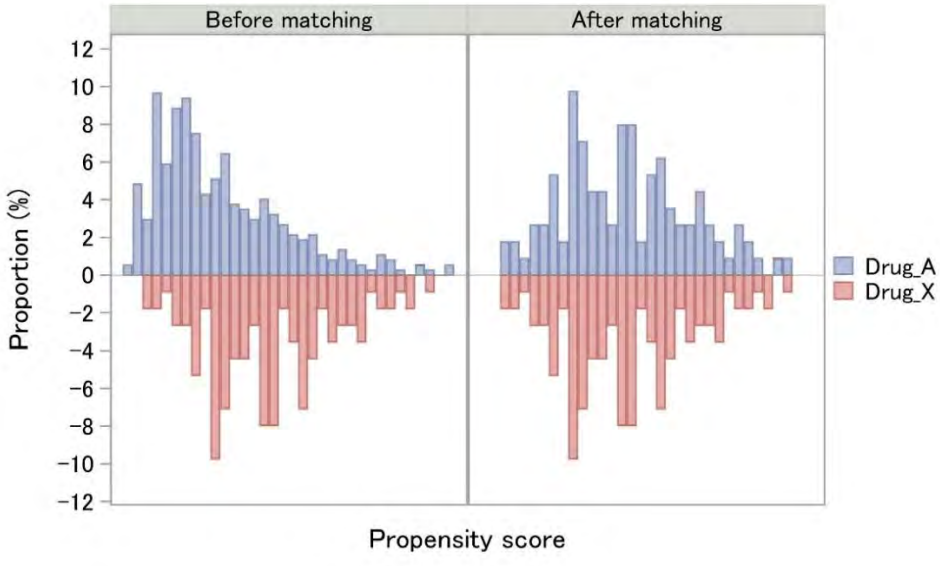


図 6：SGPANEL プロシジャによるミラーヒストグラムの作成 (方法 3)

5　まとめ

本稿では，傾向スコアマッチングを用いる際によく実施される群間バランスの評価方法として，視覚的な評価に焦点をあて，ミラーヒストグラムの作成方法について議論した．PSMATCH プロシジャでは，クラウドプロット，箱ひげ図，累積分布が出力されるが，傾向スコアの分布に関するミラーヒストグラムは出力されない．そこで，SAS によるミラーヒストグラム作成方法として，SGPANEL プロシジャによる方法，テンプレートを定義したうえで SGRENDER プロシジャによる方法，SGPANEL プロシジャを 2 回実行する方法の 3 つのアプローチを取りあげた．傾向スコアマッチング解析において，どのような場面でマッチング前後の群間バランスの評価が求められるかによって，これらの使い分けが変わってくるであろう．

本稿では，PSMATCH プロシジャによる出力データセットを用いて，ミラーヒストグラムを作成する方法について議論したが，本来であれば PSMATCH プロシジャによる ODS GRAPHICS の機能として，ミラーヒストグラムも出力されるべきである．SAS Institute Japan の皆様には，本稿での議論内容を是非 North Carolina の米国本社に共有いただきたい．観察研究のデザイン及び統計解析に従事する多くの実務家にとって，本稿の内容がお役に立てれば幸いである．

参考文献

[1] Geleris J, Sun Y, Platt J, et al. Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19. *New England Journal of Medicine*. **382**:2411–2418, 2020.

[2] Nelson DB, Rice DC, Niu J, et al. Long-Term Survival Outcomes of Cancer-Directed Surgery for Malignant Pleural Mesothelioma: Propensity Score Matching Analysis. *Journal of Clinical Oncology*. **35**:3354–3362, 2017.

[3] Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *Journal of Clinical Epidemiology*. **54**:387–398, 2001.

[4] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. **70**:41–55, 1983.

[5] 上野航太郎・魚住龍史. 傾向スコアマッチング前後の群間バランスの 定量的な評価指標に対する考察. SAS ユーザー総会 論文集 2024, 395.

[6] 魚住龍史・矢田真城・山本倫生・川口淳. SAS による傾向スコアマッチング. SAS ユーザー総会 論文集 2017, 263–281.

[7] 魚住龍史 (監修)・臨床評価研究会 基礎解析分科会 (執筆). 実用 SAS 生物統計グラフハンドブック. サイエンティスト社, 2026.

[8] 大橋靖雄・浜田知久馬・魚住龍史. 生存時間解析 応用編 –SAS による生物統計. 東京大学出版会, 2016.

[9] 大橋靖雄・浜田知久馬・魚住龍史. 生存時間解析 第 2 版 –SAS による生物統計. 東京大学出版会, 2022.

連絡先

E-mail : uozumi.r.ab@m.titech.ac.jp

ベイズ最適区間デザインによるがん第I相試験

○栗田凪，魚住龍史

東京科学大学工学院経営工学系

Bayesian Optimal Interval Designs for Phase I Oncology Trials

Nagi Kurita and Ryuji Uozumi
Department of Industrial Engineering and Economics, Institute of Science Tokyo

要旨

がん第 I 相試験は抗悪性腫瘍薬の安全性および忍容性を評価し，最大耐用量を特定することを主な目的として実施される. 近年注目されている用量探索法のカテゴリーとして, 統計的な根拠に基づく用量選択に加え，用量調整規則を事前に定めることができる Model-assisted デザインがある．Liu and Yuan（2015）は Model-assisted デザインの一つとしてベイズ最適区間デザインを提案している．ベイズ最適区間デザインは，用量調整規則の構成方法が異なる局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの 2 つに大別される．本稿ではこれら 2 つの用量探索法を整理し，SAS を用いたシミュレーションによってその動作特性を評価する．シミュレーションの結果，局所的ベイズ最適区間デザインの方が最大耐用量を正しく特定できる確率が高い一方で，大域的ベイズ最適区間デザインの方が保守的であり，平均被験者数が少ないことが示唆された.

キーワード：がん第 I 相試験, Model-assisted デザイン, ベイズ最適区間デザイン, BOIN デザイン, local BOIN, global BOIN，用量探索法

1. はじめに

がん第 I 相試験の目的の一つは，抗悪性腫瘍薬の安全性および忍容性を評価し，毒性が許容できる範囲内の最大の用量と定義される最大耐用量（Maximum Tolerated Dose，MTD）を特定することである．忍容性を検討する主たる評価項目として用量制限毒性（Dose Limiting Toxicity，DLT）を事前に治験実施計画書で定義し，定められた期間中に発現する DLT の頻度に基づいて MTD を決定する.

がん第 I 相試験において提案されている用量探索法は，Rule-based デザイン，Model-based デザイン，Model-assisted デザインという 3 つのカテゴリーに分類することができる．3+3 デザインに代表される Rule-based デザインは，試験の計画や実施が簡便である一方で，Model-based デザインや Model-assisted デザインと比較して用量選択の効率性に課題がある．連続再評価法に代表される Model-based デザインは，Rule-based デザインと比較して効率的な用量選択が可能である一方で，試験の計画や実施が煩雑になるという課

題がある．そこで近年，統計的な根拠に基づく用量選択に加え，用量調整規則を事前に定めることができる Model-assisted デザインが提案された．

Model-assisted デザインの一つとして，Liu and Yuan（2015）はベイズ最適区間デザイン（Bayesian Optimal Interval Design，BOIN）を提案している．ベイズ最適区間デザインは，検討中の用量における DLT 発現確率が，維持区間を下回るか，維持区間に含まれるか，維持区間を上回るかによって，それぞれ漸増，維持，漸減を行うデザインである．検討中の用量における被験者数と DLT 発現例数によってのみ用量調整を行うため，用量調整規則を事前に規定することができる．

Liu and Yuan（2015）はベイズ最適区間デザインとして，局所的ベイズ最適区間デザイン（local BOIN）と大域的ベイズ最適区間デザイン（global BOIN）という 2 つのデザインを提案している．現状ほとんどの場合，ベイズ最適区間デザインあるいは BOIN デザインという言葉は，局所的ベイズ最適区間デザインを指す．これは，局所的ベイズ最適区間デザインの方が大域的ベイズ最適区間デザインよりも簡便であることに加え動作特性も優れていることを理由に，Liu and Yuan（2015）が局所的ベイズ最適区間デザインを推奨しているためだと考えられる．実際，R の BOIN パッケージで実装されている動作特性を確認するためのシミュレーションプログラムは，局所的ベイズ最適区間デザインを用いている（Yan et al., 2020）．また無情報事前分布下の局所的ベイズ最適区間デザインは，米国食品医薬品局（U. S. Food and Drug Administration, FDA）から Fit-for-Purpose に指定されている（FDA, 2021）．

そこで本稿では，局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの共通点および相違点を整理する．その後，SAS でこれら 2 つのベイズ最適区間デザインを実装し，シミュレーションによってその動作特性を評価する．

2. ベイズ最適区間デザイン

はじめに局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインに共通する用量調整規則を説明する．ベイズ最適区間デザインは，検討中の用量d_jにおける真の DLT 発現確率p_jを閾値$\lambda_{ej}, \lambda_{dj}$と比較することによって用量調整を行う．次の図 1 のように，$p_j \leq \lambda_{ej}$ならば漸増（E），$\lambda_{ej} < p_j \leq \lambda_{dj}$ならば維持（$S$），$\lambda_{dj} < p_j$ならば漸減（$D$）という判断が下される．

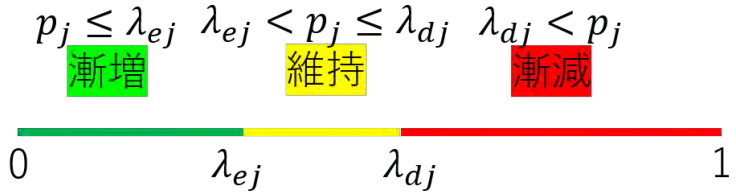


図 1　ベイズ最適区間デザインの用量調整規則

実際には真の DLT 発現確率p_jを観測することはできないため，被験者数n_jと DLT 発現例数y_jから$\hat{p}_j = y_j/n_j$として推定する．

大域的ベイズ最適区間デザインと局所的ベイズ最適区間デザインの違いは，$\lambda_{ej}, \lambda_{dj}$の構成方法の違いに由来する．したがって次節以降，その構成方法および用量調整表の例を示す．

2.1. 局所的ベイズ最適区間デザイン

局所的ベイズ最適区間デザインでは，次のような点仮説を考える．

$$H_{0j}: p_j = \phi, H_{1j}: p_j = \phi_1, H_{2j}: p_j = \phi_2$$

またそれぞれの点仮説が真である確率を次のように表す.

$$\Pr(H_{0j}) = \pi_{0j}, \Pr(H_{1j}) = \pi_{1j}, \Pr(H_{2j}) = \pi_{2j}$$

ここで，ϕは目標とする DLT 発現確率，ϕ_1は漸増と判断されるϕより低い毒性確率，ϕ_2は漸減と判断されるϕより高い毒性確率を表す．このとき閾値が与えられた下で誤った用量調整を行う確率$\alpha(\lambda_{ej}, \lambda_{dj})$は次のように表される．

$$\alpha(\lambda_{ej}, \lambda_{dj}) = \Pr(H_{0j}) \Pr(\bar{S}|H_{0j}) + \Pr(H_{1j}) \Pr(\bar{E}|H_{1j}) + \Pr(H_{2j}) \Pr(\bar{D}|H_{2j})$$

$\alpha(\lambda_{ej}, \lambda_{dj})$はそれぞれの点仮説の下で，誤った用量調整を行う確率の和である．例えばH_{0j}が真であるとき，正しい用量調整は維持，誤った用量調整は漸増あるいは漸減であるため，誤った用量調整を行う確率は$\Pr(H_{0j}) \Pr(\bar{S}|H_{0j})$となる．二項分布の累積分布関数$Bin(b; n, \phi)$を用いると，$\alpha(\lambda_{ej}, \lambda_{dj})$は次のように変形できる．

$$\begin{aligned}\alpha(\lambda_{ej}, \lambda_{dj}) = {} & \pi_{0j}\{Bin(n_j\lambda_{ej}; n_j, \phi) + 1 - Bin(n_j\lambda_{dj} - 1; n_j, \phi)\} \\ & + \pi_{1j}\{1 - Bin(n_j\lambda_{ej}; n_j, \phi_1)\} \\ & + \pi_{2j} Bin(n_j\lambda_{dj} - 1; n_j, \phi_2)\end{aligned}$$

局所的ベイズ最適区間デザインにおける閾値は，$\alpha(\lambda_{ej}, \lambda_{dj})$を最小化する$\lambda_{ej}, \lambda_{dj}$として次のように得られる．

$$\lambda_{ej} = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right) + n_j^{-1}\log\left(\frac{\pi_{1j}}{\pi_{0j}}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)}, \lambda_{dj} = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right) + n_j^{-1}\log\left(\frac{\pi_{0j}}{\pi_{2j}}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)}$$

検討中の用量d_jについて，毒性確率がϕ, ϕ_1, ϕ_2となる確率がそれぞれ等しい，すなわち$\pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$というベイズの事前分布を与えると，対数関数の真数の部分が 1 になることから，$\alpha(\lambda_{ej}, \lambda_{dj})$を最小化する$\lambda_{ej}, \lambda_{dj}$は次のように変形できる．

$$\lambda_e = \frac{\log\left(\frac{1-\phi_1}{1-\phi}\right)}{\log\left(\frac{\phi(1-\phi_1)}{\phi_1(1-\phi)}\right)}, \lambda_d = \frac{\log\left(\frac{1-\phi}{1-\phi_2}\right)}{\log\left(\frac{\phi_2(1-\phi)}{\phi(1-\phi_2)}\right)} \quad (1)$$

ゆえに毒性確率がϕ, ϕ_1, ϕ_2となる確率がそれぞれ等しいというベイズの事前分布を与えると，n_jを含む項が消えるため，被験者数によらず，すべての用量レベルで共通の閾値が得られる．これには実用上簡便であるという利点がある．したがって局所的ベイズ最適区間デザインを用いた用量調整は，一般に(1)式の閾値によって実施される．またϕ_1およびϕ_2は，$\phi_1 = 0.6\phi, \phi_2 = 1.4\phi$とすることが推奨されている．よって本稿では以降，この推奨値を用いた(1)式の閾値による用量調整を行うデザインを局所的ベイズ最適区間デザインと呼ぶ．

例として，$\phi = 0.30$の場合における局所的ベイズ最適区間デザインを用いた用量調整表を示す．(1)式より$\lambda_e = 0.236, \lambda_d = 0.358$と算出されるため，用量調整表は次の表 1 のようになる．

表 1　$\phi = 0.30$の場合における局所的ベイズ最適区間デザインを用いた用量調整表

	検討中の用量における被験者数					
	3	6	9	12	15	18
漸増 if # of DLT ≤	0	1	2	2	3	4
維持 if # of DLT =	1	2	3	3-4	4-5	5-6
漸減 if # of DLT ≥	2	3	4	5	6	7

2.2. 大域的ベイズ最適区間デザイン

大域的ベイズ最適区間デザインでは，次のような区間仮説を考える．

$$H_{0j}: \phi_1 < p_j < \phi_2, H_{1j}: 0 \le p_j \le \phi_1, H_{2j}: \phi_2 \le p_j \le 1$$

H_{0j}は検討中の用量d_jの毒性確率が目標に近く，用量を維持すべきであること，H_{1j}はd_jが過小用量であり漸増すべきであること，H_{2j}はd_jが過大用量であり漸減すべきであることを表す．局所的ベイズ最適区間デザインでは真の DLT 発現確率がϕ, ϕ_1, ϕ_2の中からいずれかの値をとる状況のみを考えたのに対し，大域的ベイズ最適区間デザインでは，$p_j \in [0,1]$のすべての値を場合分けして考える点が異なる．それぞれの区間仮説の下で，p_jに対する事前分布として次のような一様分布を与える．

$$f(p_j|H_{0j}) = Unif(\phi_1, \phi_2), f(p_j|H_{1j}) = Unif(0, \phi_1), f(p_j|H_{2j}) = Unif(\phi_2, 1)$$

このとき閾値が与えられた下で誤った用量調整を行う確率$\alpha_g(\lambda_{ej}, \lambda_{dj})$は，各仮説の下での誤りの確率を$p_j$について周辺化することで次のように表される．

$$\alpha_g(\lambda_{ej}, \lambda_{dj}) = \Pr(H_{0j}) \int f(p_j|H_{0j}) \Pr(\bar{S}|p_j, H_{0j})\, dp_j + \Pr(H_{1j}) \int f(p_j|H_{1j}) \Pr(\bar{E}|p_j, H_{1j})\, dp_j$$

$$+ \Pr(H_{2j}) \int f(p_j|H_{2j}) \Pr(\bar{D}|p_j, H_{2j})\, dp_j$$

ここで二項分布の確率関数を$[0, c]$で積分すると$\int_0^c {}_nC_y p^y (1-p)^{n-y} dp = Beta(c; y+1, n-y+1)/(n+1)$となることから，ベータ分布の累積分布関数$Beta(c; a, b)$を用いて$\alpha_g(\lambda_{ej}, \lambda_{dj})$は次のように変形できる．

$$\alpha_g(\lambda_{ej}, \lambda_{dj}) = \pi_{0j} + \pi_{1j}$$

$$+ \sum_{y=0}^{b_{ej}} \left[\frac{\pi_{0j}\{Beta(\phi_2; y+1, n_j - y + 1) - Beta(\phi_1; y+1, n_j - y + 1)\}}{(\phi_2 - \phi_1)(n_j + 1)} \right.$$

$$\left. - \frac{\pi_{1j} Beta(\phi_1; y+1, n_j - y + 1)}{\phi_1 (n_j + 1)} \right]$$

$$+ \sum_{y=0}^{b_{dj}-1} \left[\frac{\pi_{2j}\{1 - Beta(\phi_2; y+1, n_j - y + 1)\}}{(1 - \phi_2)(n_j + 1)} \right.$$

$$\left. - \frac{\pi_{0j}\{Beta(\phi_2; y+1, n_j - y + 1) - Beta(\phi_1; y+1, n_j - y + 1)\}}{(\phi_2 - \phi_1)(n_j + 1)} \right]$$

ただし$b_{ej} = \lfloor n_j \lambda_{ej} \rfloor, b_{dj} = \lceil n_j \lambda_{dj} \rceil$である．すなわち$b_{ej}$は漸増と判断される DLT 発現例数の最大値，$b_{dj}$は漸減と判断される DLT 発現例数の最小値に相当する．局所的ベイズ最適区間デザインとは異なり，$\alpha_g(\lambda_{ej}, \lambda_{dj})$を最小化する$\lambda_{ej}, \lambda_{dj}$は閉じた形では得られない．したがって，$0 \le b_{ej} \le b_{dj} \le n_j$を満たす組み合わせをすべて数え上げ，$\alpha_g(\lambda_{ej}, \lambda_{dj})$を最小にする組を選ぶという数値探索によって閾値を求めることができる．$\alpha_g(\lambda_{ej}, \lambda_{dj})$は$b_{ej}$に関する項と$b_{dj}$に関する項に分離できるため，それぞれ独立に累積和を最小化すればよい．また$\pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$としても，$\alpha_g(\lambda_{ej}, \lambda_{dj})$から$n_j$を消去することはできない．したがって大域的ベイズ最適区間デザインの閾値は，事前確率を等しく与えた場合であっても検討中の用量における被験者数n_jに依存する．この点が被験者数によらず一貫した閾値が得られる局所的ベイズ最適区間デザインとの実用上の大きな違いである．

例として，$\phi = 0.30, \phi_1 = 0.6\phi = 0.18, \phi_2 = 1.4\phi = 0.42, \pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$とした場合の大域的ベイズ最適区間デザインを用いた用量調整表を表 2 に示す．

表 2　$\phi = 0.30$の場合における大域的ベイズ最適区間デザインを用いた用量調整表

	検討中の用量における被験者数					
	3	6	9	12	15	18
漸増 if # of DLT ≤	0	0	1	2	2	3
維持 if # of DLT ＝	1	1-3	2-4	3-5	3-7	4-8
漸減 if # of DLT ≥	2	4	5	6	8	9

2.3. 局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの整理

前節までに整理した局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインによる用量調整規則を踏まえて，その特徴をまとめる．局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの共通点は，毒性確率と閾値の大小関係によって用量調整を行うこと，用量調整に用いる情報が検討中の用量における被験者数と DLT 発現例数のみであること，誤った用量調整を行う確率の最小化によって閾値を決定すること，用量調整表の事前規定が可能であることが挙げられる．局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの相違点としては，次の図 2 に示す要素が挙げられる．また$\phi = 0.30$の場合について，検討中の用量における被験者数ごとの維持区間を図 3 に示す．

局所的ベイズ最適区間デザイン		大域的ベイズ最適区間デザイン
点仮説	毒性確率の仮定	区間仮説
閉形式から得られる	閾値の求値	数値探索が必要
依存しない	閾値の被験者数への依存	依存する
一定	維持区間の幅	被験者数によって変化

図 2　局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの相違点

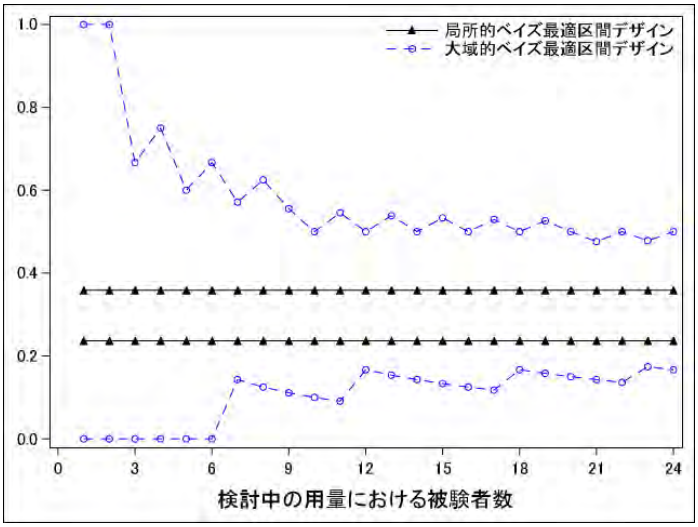


図 3　$\phi = 0.30$の場合における維持区間

3. シミュレーション

本節では，2 節で述べた局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインを SAS で実装し，シミュレーションによって両者の動作特性を比較する．

3.1. シミュレーション設定

目標 DLT 発現確率は$\phi = 0.30$とし，推奨値にしたがって$\phi_1 = 0.6\phi = 0.18, \phi_2 = 1.4\phi = 0.42$とする．また事前確率は$\pi_{0j} = \pi_{1j} = \pi_{2j} = 1/3$とする．このとき局所的ベイズ最適区間デザインによる用量調整表は 2.1 節の表 1，大域的ベイズ最適区間デザインによる用量調整表は 2.2 節の表 2 のようになる．

最大被験者数は 30 例，コホートサイズは 3 例とし，最大 10 コホートを登録する．ただし検討中の用量における被験者数が 12 例に達し，当該用量における用量調整が維持の場合は，その時点で試験を終了する．検討する用量はd_1, d_2, d_3, d_4, d_5の全 5 用量とし，開始用量は最低用量であるd_1とする．また用量除外規則として，ベータ二項モデルに基づき$n_j \geq 3$かつ$\Pr(p_j > \phi | y_j, n_j) > \xi$となった場合に$d_j$以上の用量を以降の試験から除外する．$p_j$の事前分布には無情報事前分布$Beta(1,1)$を仮定する．$\xi$は$d_2$以上の用量については0.95，最低用量である$d_1$については0.90とする．最低用量が除外された場合には試験を中止し，MTD は選択しないものとする．MTD は試験終了時に得られた各用量の観測 DLT 発現割合$\hat{p}_j = y_j/n_j$に対して等張回帰を適用し，単調性を満たすように変換した推定値$\tilde{p}_j$を求め，$\tilde{p}_j$がϕに最も近い用量を選択する．ただし局所的ベイズ最適区間デザインの漸減境界である$\lambda_d = 0.358$を上回る毒性確率をもつ用量，除外された用量，1 例も被験者が割り付けられなかった用量は MTD の候補から除くものとする．条件を揃えて比較するため，大域的ベイズ最適区間デザインでも局所的ベイズ最適区間デザインの漸減境界を用いて MTD 選択を行うことに注意されたい．

次の表 3 のような MTD の位置が異なる 6 つのシナリオを設定する．シナリオ 6 はすべての用量における毒性確率が目標 DLT 発現確率を超える，真の MTD が存在しない過毒性シナリオである．また図 4 にシナリオ別の DLT 発現確率の推移を示す．

表 3　各用量における真の DLT 発現確率

シナリオ	真の DLT 発現確率				
	用量d_1	用量d_2	用量d_3	用量d_4	用量d_5
1	**0.30**	0.45	0.55	0.65	0.75
2	0.15	**0.30**	0.45	0.55	0.65
3	0.10	0.20	**0.30**	0.40	0.50
4	0.05	0.15	0.24	**0.30**	0.36
5	0.03	0.05	0.08	0.15	**0.30**
6	0.35	0.45	0.50	0.55	0.60

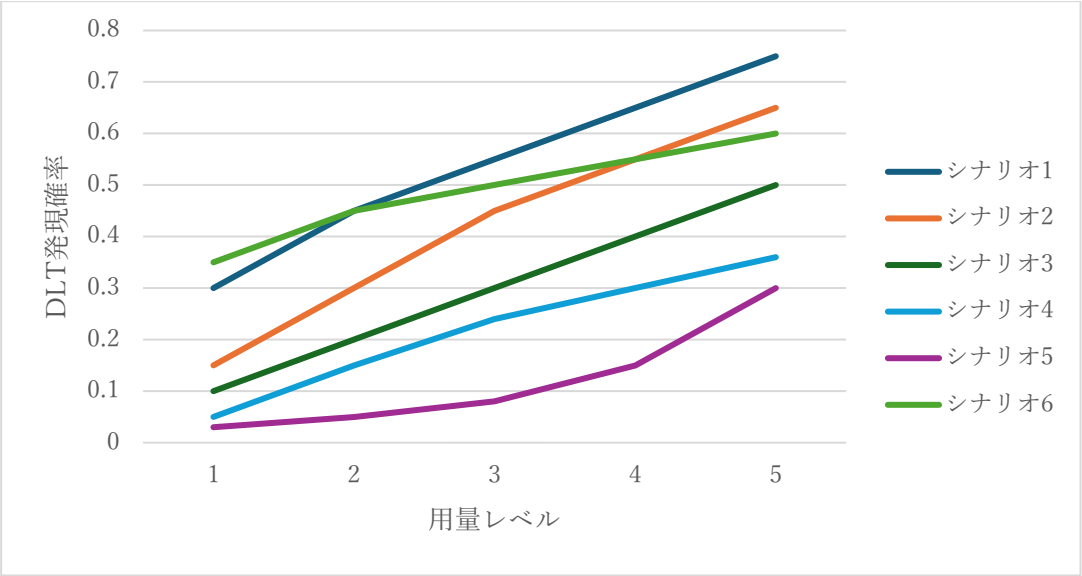


図 4　シナリオ別の DLT 発現確率の推移

それぞれのシナリオに対して 10000 回のシミュレーションを実施する．局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの動作特性は，MTD の選択割合，平均被験者数，平均 DLT 発現例数，中止割合によって評価する．

3.2. SAS によるシミュレーションの実装

前節で説明した設定を踏まえて，局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインのシミュレーションプログラムをそれぞれ SAS 上で実装する．作成した局所的ベイズ最適区間デザインのシミュレーションプログラムの概要を次に示す．

```
/*局所的ベイズ最適区間デザイン*/
proc iml;
/*シミュレーション設定*/
scenario = {
  0.30 0.45 0.55 0.65 0.75,
  0.15 0.30 0.45 0.55 0.65,
  0.10 0.20 0.30 0.40 0.50,
  0.05 0.15 0.24 0.30 0.36,
  0.03 0.05 0.08 0.15 0.30,
  0.35 0.45 0.50 0.55 0.60
};target = 0.30; p_saf = 0.6; p_tox = 1.4; target_eli = 0.30; n_max = 30;
n_max_dose = 12; cohort_size = 3; stop_boundary = {0.90 0.95 0.95 0.95 0.95};
start_dose = 1; n_simulations = 10000; mtd_target = 0.30; mtd_cut = 0.358; seed = 1234;
display_result = 1;
/*閾値の計算*/
start lambda_e_cal(phi, psaf);
   phi1 = phi * psaf;
   return( log( (1 - phi1) / (1 - phi) ) /
           log( (phi * (1 - phi1)) / (phi1 * (1 - phi)) ) );
finish lambda_e_cal;
start lambda_d_cal(phi, ptox);
   phi2 = phi * ptox;
   return( log( (1 - phi) / (1 - phi2) ) /
           log( (phi2 * (1 - phi)) / (phi * (1 - phi2)) ) );
finish lambda_d_cal;
```

```
/*局所的ベイズ最適区間デザインの実装*/
start BOIN(dlt_probs, mtd_sim, n_patients, n_dlts)
      global(b_esc, b_dees, b_elim, n_max, n_max_dose,
             cohort_size, start_dose, mtd_target, mtd_cut);
/* … (中略) … */
   do while (stopflag = 0);
      m = min( cohort_size, n_max - sum(n_patients) );
      if m > 0 then do;
         u = j(1, m, .);
         call randgen(u, "Uniform");
         new_dlts = sum( u < dlt_probs[current_dose] );
      end;
      else new_dlts = 0;
      n_patients[current_dose] = n_patients[current_dose] + m;
      n_dlts[current_dose]     = n_dlts[current_dose] + new_dlts;
      /*用量調整*/
      if nd <= b_esc[np] then do;
         /*漸増*/
         if (current_dose ^= n_doses) & (current_dose < cap_dose) then
            current_dose = current_dose + 1;
         else if np >= n_max_dose then stopflag = 1;
      end;
      else if (np >= 3) & (nd >= b_elim[current_dose, np]) then do;
         /*除外*/
         if current_dose = 1 then do;
            cap_dose     = 0; current_dose = 0; stopflag     = 1;
         end;
         else do;
            cap_dose = current_dose - 1; current_dose = current_dose - 1;
         end;
      end;
      else if nd >= b_dees[np] then do;
         /*漸減*/
         if current_dose ^= 1 then current_dose = current_dose - 1;
         else if np >= n_max_dose then stopflag = 1;
      end;
      else do;
         /*維持*/
         if np >= n_max_dose then stopflag = 1;
      end;
      if sum(n_patients) >= n_max then stopflag = 1;
   end;
   /*MTD選択*/
finish BOIN;
```

大域的ベイズ最適区間デザインのシミュレーションプログラムは閾値の計算を除き，局所的ベイズ最適区間デザインのシミュレーションプログラムと共通した枠組みを用いて実装することができる．したがってここでは大域的ベイズ最適区間デザインを用いた閾値を求めるプログラムを次に示す．

```
/*大域的ベイズ最適区間デザインによる閾値の計算*/
start global_boundary(n, phi1, phi2, pi0, pi1, pi2, b_e, b_d);
   eps = 1e-12;
   g0 = j(1, n+1, .);  g1 = j(1, n+1, .);  g2 = j(1, n+1, .);
   do k = 0 to n;
      aa = k + 1; bb = n - k + 1;
      I1 = cdf('BETA', phi1, aa, bb); I2 = cdf('BETA', phi2, aa, bb);
      g0[k+1] = (I2 - I1) / (phi2 - phi1);
```

```
      g1[k+1] = I1 / phi1;
      g2[k+1] = (1 - I2) / (1 - phi2);
   end;
   Aterm = pi0 # g0 - pi1 # g1;
   Bterm = pi2 # g2 - pi0 # g0;
   best = 0;  b_e = -1;  s = 0;
   do k = 0 to n;
      s = s + Aterm[k+1];
      if s < best - eps then do;
         best = s; b_e  = k;
      end;
   end;
   best = 0;  b_d = 0;  s = 0;
   do k = 1 to n+1;
      s = s + Bterm[k];
      if s <= best + eps then do;
         if s < best then best = s;
         b_d = k;
      end;
   end;
finish global_boundary;
```

局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの全シミュレーションプログラムは付録を参照されたい.

3.3. 結果

シミュレーションによって局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインの動作特性を評価した結果を次に示す．それぞれのシナリオにおいて，真の MTD の用量に該当するセル内の数値を太字で示している．なお表 4 における中止割合は，最低用量が除外され試験が中止された場合と MTD が選択されなかった場合の合計を表す．また図 5 から図 13 に関しては，真の MTD が存在しないシナリオ 6 を除いた結果を示している．

表 4　MTD 選択割合（%）と中止割合（%）

シナリオ	MTD として選択される割合（local BOIN / global BOIN）					中止割合 （local BOIN / global BOIN）
	用量d_1	用量d_2	用量d_3	用量d_4	用量d_5	
1	**47.3/51.4**	15.1/11.7	1.2/0.8	0.1/0.0	0.0/0.0	36.4/36.0
2	27.0/34.4	**50.4/47.2**	13.9/10.6	1.5/0.7	0.1/0.0	7.2/7.1
3	7.4/11.8	34.1/38.2	**36.0/34.0**	16.0/11.4	3.3/1.3	3.2/3.2
4	2.1/3.5	20.1/25.7	32.0/36.3	**26.6/23.6**	18.6/10.1	0.7/0.8
5	0.1/0.4	0.6/1.5	4.4/6.6	35.0/38.2	**59.6/53.2**	0.2/0.2
6	38.9/41.0	10.6/9.1	1.8/0.9	0.2/0.1	0.0/0.0	48.5/48.9

表 5　平均被験者数

シナリオ	平均被験者数（local BOIN / global BOIN）					合計
	用量d_1	用量d_2	用量d_3	用量d_4	用量d_5	（local BOIN / global BOIN）
1	**8.9/8.7**	5.0/3.8	1.0/0.5	0.1/0.0	0.0/0.0	15.0/13.0
2	7.0/6.9	**9.1/8.1**	4.4/3.1	0.9/0.5	0.1/0.0	21.5/18.7
3	5.0/5.3	7.5/7.4	**7.0/5.9**	3.7/2.3	1.1/0.4	24.2/21.4
4	4.0/4.2	6.3/6.7	7.0/6.8	**5.2/3.9**	3.2/1.6	25.8/23.3
5	3.3/3.6	3.6/4.1	4.3/4.8	6.7/6.7	**9.3/7.4**	27.2/26.6
6	8.4/8.2	3.8/2.9	0.9/0.5	0.1/0.1	0.0/0.0	13.3/11.6

表 6　平均 DLT 発現例数

シナリオ	平均 DLT 発現例数（local BOIN / global BOIN）					合計
	用量d_1	用量d_2	用量d_3	用量d_4	用量d_5	（local BOIN / global BOIN）
1	**2.7/2.6**	2.2/1.7	0.5/0.3	0.1/0.0	0.0/0.0	5.5/4.6
2	1.1/1.0	**2.7/2.4**	2.0/1.4	0.5/0.3	0.1/0.0	6.4/5.1
3	0.5/0.5	1.5/1.5	**2.1/1.8**	1.4/0.9	0.6/0.2	6.1/4.9
4	0.2/0.2	1.0/1.0	1.7/1.6	**1.6/1.2**	1.2/0.6	5.7/4.6
5	0.1/0.1	0.2/0.2	0.3/0.4	1.0/1.0	**2.8/2.2**	4.4/3.9
6	3.0/2.9	1.7/1.3	0.4/0.3	0.1/0.0	0.0/0.0	5.2/4.5

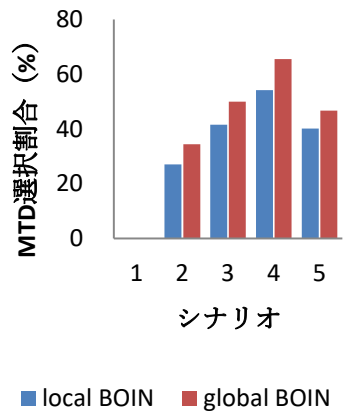


図 5　MTD 未満の用量を MTD として選択した割合（%）

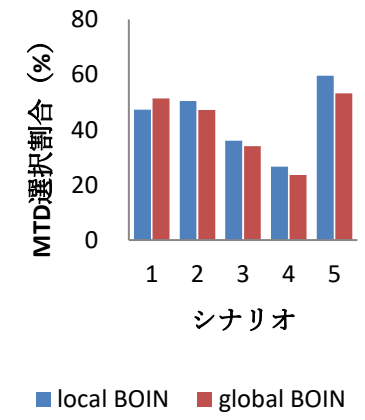


図 6　真の MTD を正しく MTD として選択した割合（%）

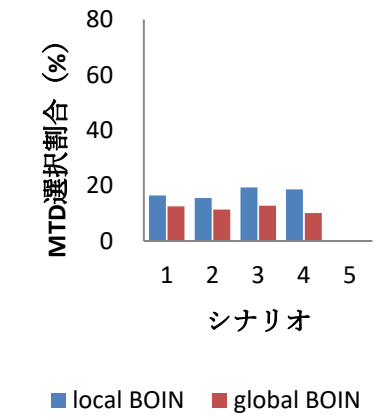


図 7　MTD より高い用量を MTD として選択した割合（%）

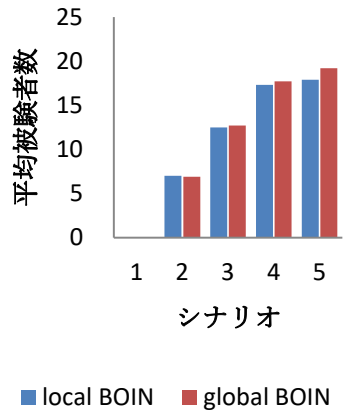


図 8　MTD 未満の用量に割り付けた被験者数の平均

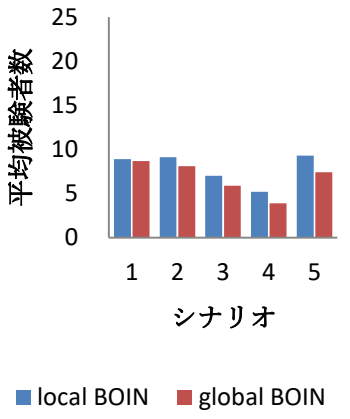


図 9　真の MTD に割り付けた被験者数の平均

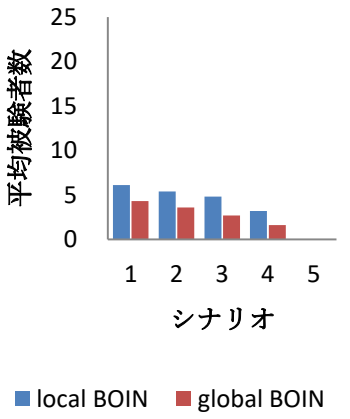


図 10　MTD より高い用量に割り付けた被験者数の平均

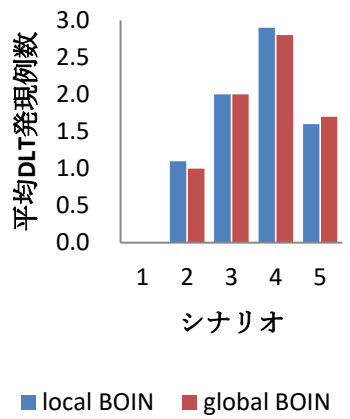


図 11　MTD 未満の用量における DLT 発現例数の平均

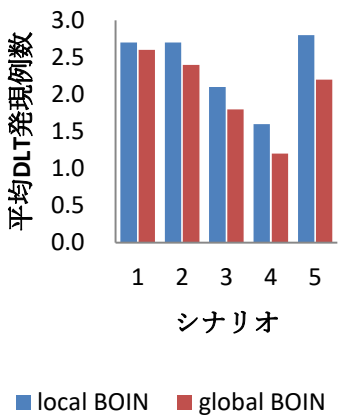


図 12　真の MTD における DLT 発現例数の平均

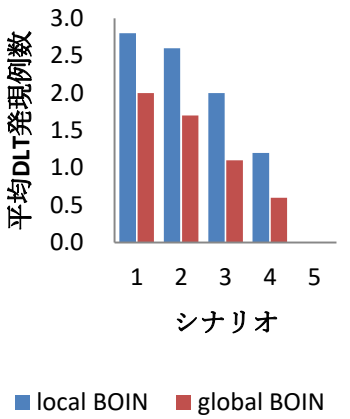


図 13　MTD より高い用量における DLT 発現例数の平均

まず図 6 に着目すると，シナリオ 1 を除き局所的ベイズ最適区間デザインの方が大域的ベイズ最適区間デザインよりも真の MTD を正しく MTD として選択した割合が高いことが確認される．シナリオ 1 については，大域的ベイズ最適区間デザインは局所的ベイズ最適区間デザインよりも維持区間が広く漸増しにくいため，開始用量でありかつ真の MTD である用量d_1より高い用量へ移りにくい．したがって，大域的ベイズ最適区間デザインにとって有利な状況であったと推測される．次に図 7 に着目すると，大域的ベイズ最適区間デザインの方が局所的ベイズ最適区間デザインよりも MTD より高い用量を MTD として選択した割合が低かった．また図 10 および図 13 においても同様に，大域的ベイズ最適区間デザインの方が局所的ベイズ最適区間デザインよりも MTD より高い用量への割り付けと DLT 発現例数が少なかった．したがって安全性の観点からは，大域的ベイズ最適区間デザインの方が局所的ベイズ最適区間デザインよりも優れていることが示唆される．逆に図 5，図 8，図 11 に着目すると多くのシナリオで，大域的ベイズ最適区間デザインの方が局所的ベイズ最適区間デザインよりも MTD 未満の用量を MTD として選択した割合が高く，MTD 未満の用量に割り付けた被験者数および MTD 未満の用量における DLT 発現例数が多かった．これらの結果から，大域

的ベイズ最適区間デザインは局所的ベイズ最適区間デザインよりも保守的なデザインであることが示唆される．次に表 5 に着目すると，大域的ベイズ最適区間デザインは局所的ベイズ最適区間デザインよりも試験全体での被験者数が少なかった．特にシナリオ 2 とシナリオ 3 では 2.8 例少なかった．これは，大域的ベイズ最適区間デザインは維持区間が広いため，検討中の用量における被験者数が 12 例に達し当該用量における観測 DLT 発現割合が維持区間に含まれる場合はその時点で試験を終了するという試験終了条件が発動しやすいからだと考えられる．同様に表 6 から，平均 DLT 発現例数に関しても大域的ベイズ最適区間デザインの方が少ないことが確認された．

4. まとめ

本稿では，がん第 I 相試験における用量探索法の一つであるベイズ最適区間デザインに注目した．ベイズ最適区間デザインとして，局所的ベイズ最適区間デザインと大域的ベイズ最適区間デザインという用量調整規則が異なる 2 つのデザインを紹介した後，シミュレーションによってその動作特性を評価した．シミュレーションの結果，局所的ベイズ最適区間デザインの方が MTD を正しく特定できる確率が高い一方で，大域的ベイズ最適区間デザインの方が保守的であり，平均被験者数が少ないことが示唆された．実際にがん第 I 相試験における用量探索法を選択する際には，動作特性だけでなく，運用上の難易度や臨床的な観点などと合わせて判断することが必要である．抗がん剤開発に従事する多くの研究者がベイズ最適区間デザインを検討する際，本稿がその一助となれば幸いである．

参考文献

Liu S, Yuan Y. Bayesian optimal interval designs for phase I clinical trials. Journal of the Royal Statistical Society: Series C. 2015;64(3):507-523.

O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. Biometrics. 1990;46(1):33-48.

Storer BE. Design and analysis of phase I clinical trials. Biometrics. 1989;45(3):925-937.

U.S. Food and Drug Administration. Determination Letter: Fit-for-Purpose Initiative — Bayesian Optimal Interval (BOIN) Design. Silver Spring (MD): FDA; 2021 Dec 10. [cited 2026 Sep 16]. Available from: https://www.fda.gov/media/155363/download

Yan F, Zhang L, Zhou Y, et al. BOIN: An R package for designing single-agent and drug-combination dose-finding trials using Bayesian optimal interval designs. Journal of Statistical Software. 2020;94(13):1-32.

独立行政法人医薬品医療機器総合機構．抗悪性腫瘍剤の第 I 相試験計画時の統計学的留意事項 ― 安全性の観点から（Early Consideration），令和 6 年 12 月．https://www.pmda.go.jp/review-services/symposia/0177.html（2026 年 9 月 15 日最終閲覧）

日本製薬工業協会 医薬品評価委員会 データサイエンス部会（2023）．近年のがん第 I 相試験デザインとその選択．https://www.jpma.or.jp/information/evaluation/results/allotment/DS_202306_oncoP1DE.html（2026 年 9 月 15 日最終閲覧）

付録

A. 局所的ベイズ最適区間デザインのシミュレーションプログラム

```
/*局所的ベイズ最適区間デザイン*/
proc iml;
/*シミュレーション設定*/
scenario = {
  0.30 0.45 0.55 0.65 0.75,
  0.15 0.30 0.45 0.55 0.65,
  0.10 0.20 0.30 0.40 0.50,
  0.05 0.15 0.24 0.30 0.36,
  0.03 0.05 0.08 0.15 0.30,
  0.35 0.45 0.50 0.55 0.60
};target = 0.30; p_saf = 0.6; p_tox = 1.4; target_eli = 0.30; n_max = 30;
n_max_dose = 12; cohort_size = 3; stop_boundary = {0.90 0.95 0.95 0.95 0.95};
start_dose = 1; n_simulations = 10000; mtd_target = 0.30; mtd_cut = 0.358; seed = 1234;
display_result = 1;
/*閾値の計算*/
start lambda_e_cal(phi, psaf);
   phi1 = phi + psaf;
   return( log( (1 - phi1) / (1 - phi) ) /
           log( (phi + (1 - phi1)) / (phi1 + (1 - phi)) ) );
finish lambda_e_cal;
start lambda_d_cal(phi, ptox);
   phi2 = phi + ptox;
   return( log( (1 - phi) / (1 - phi2) ) /
           log( (phi2 + (1 - phi)) / (phi + (1 - phi2)) ) );
finish lambda_d_cal;
/*等張回帰*/
start isoreg_weight(xin, wt);
   x = xin;
   n = ncol(x);
   if n <= 1 then return(x);
   lvlsets = 1:n;
   done = 0;
   do while (done = 0);
      viol = ( (x[1, 2:n] - x[1, 1:(n-1)]) < 0 );
      if any(viol) then do;
         i    = min(loc(viol)); lvl1 = lvlsets[i]; lvl2 = lvlsets[i+1];
         ilvl = loc( (lvlsets = lvl1) | (lvlsets = lvl2) );
         x[1, ilvl]       = sum( x[1,ilvl] # wt[1,ilvl] ) / sum( wt[1,ilvl] );
         lvlsets[1, ilvl] = lvl1;
      end;
      else done = 1;
   end;
   return(x);
finish isoreg_weight;
/*局所的ベイズ最適区間デザインの実装*/
start BOIN(dlt_probs, mtd_sim, n_patients, n_dlts)
      global(b_esc, b_dees, b_elim, n_max, n_max_dose,
             cohort_size, start_dose, mtd_target, mtd_cut);
   n_doses      = ncol(dlt_probs);
   n_patients   = j(1, n_doses, 0);
   n_dlts       = j(1, n_doses, 0);
   cap_dose     = n_doses;
   mtd_sim      = j(1, n_doses + 2, 0);
   current_dose = start_dose;
   stopflag = 0;
   do while (stopflag = 0);
      m = min( cohort_size, n_max - sum(n_patients) );
      if m > 0 then do;
         u = j(1, m, .);
         call randgen(u, "Uniform");
         new_dlts = sum( u < dlt_probs[current_dose] );
      end;
      else new_dlts = 0;
      n_patients[current_dose] = n_patients[current_dose] + m;
      n_dlts[current_dose]     = n_dlts[current_dose] + new_dlts;
      np = n_patients[current_dose];
      nd = n_dlts[current_dose];
      /*用量調整*/
      if nd <= b_esc[np] then do;
```

```
         /*漸増*/
         if (current_dose ^= n_doses) & (current_dose < cap_dose) then
            current_dose = current_dose + 1;
         else if np >= n_max_dose then stopflag = 1;
      end;
      else if (np >= 3) & (nd >= b_elim[current_dose, np]) then do;
         /*除外*/
         if current_dose = 1 then do;
            cap_dose     = 0; current_dose = 0; stopflag     = 1;
         end;
         else do;
            cap_dose     = current_dose - 1; current_dose = current_dose - 1;
         end;
      end;
      else if nd >= b_dees[np] then do;     /*漸減*/
         if current_dose ^= 1 then current_dose = current_dose - 1;
         else if np >= n_max_dose then stopflag = 1;
      end;
      else do;   /*維持*/
         if np >= n_max_dose then stopflag = 1;
      end;
      if sum(n_patients) >= n_max then stopflag = 1;
   end;
   /*MTD選択*/
   if cap_dose = n_doses then cap_dose_index = j(1, n_doses, 1);
   else do;
      cap_dose_index = j(1, n_doses, 0);
      if cap_dose >= 1 then cap_dose_index[1, 1:cap_dose] = 1;
   end;
   dose_eli  = n_patients # cap_dose_index;
   dose_plus = 0; check     = 0;
   if all(dose_eli = 0) then do;
      mtd_sim[1] = 1; check = 1;
   end;
   else do;
      if any(dose_eli = 0) then do;
         zpd   = loc(dose_eli = 0); zpd_n = loc(n_patients = 0);
         if ncol(zpd_n) > 0 then do;
            keep_n = loc(zpd_n ^= n_doses);
            if ncol(keep_n) > 0 then do;
               zpd_n_eli = zpd_n[keep_n];
               do jj = 1 to ncol(zpd_n_eli);
                  d = zpd_n_eli[jj];
                  if n_patients[d+1] ^= 0 then dose_plus = d;
               end;
            end;
         end;
         keep_idx       = setdif(1:n_doses, zpd);
         n_dlts_mtd     = n_dlts[1, keep_idx];
         n_patients_mtd = n_patients[1, keep_idx];
      end;
      else do;
         n_dlts_mtd     = n_dlts; n_patients_mtd = n_patients;
      end;
      phat_mean = (n_dlts_mtd + 0.05) / (n_patients_mtd + 0.1);
      phat_var  = ( (n_dlts_mtd + 0.05) # (n_patients_mtd - n_dlts_mtd + 0.05) ) /
                  ( ((n_patients_mtd + 0.1)##2) # (n_patients_mtd + 0.1 + 1) );
      phat = isoreg_weight(phat_mean, 1 / phat_var);
   end;
   if check = 0 then do;
      if all(phat > mtd_cut) then mtd_sim[2] = 1;
      else do;
         ix = loc(phat > mtd_cut);
         if ncol(ix) > 0 then phat[1, ix] = -1;
         dif       = abs(phat - mtd_target); mtd_index = loc( dif = min(dif) );
         mtd       = dose_plus + max(mtd_index); mtd_sim[mtd+2] = 1;
      end;
   end;
finish BOIN;
/*用量調整表*/
n_dose = ncol(scenario);nvec   = 1:n_max;
phi1 = target * p_saf;phi2 = target * p_tox;
lambda_e = lambda_e_cal(target, p_saf);lambda_d = lambda_d_cal(target, p_tox);
b_esc    = j(1, n_max, 0); b_dees   = j(1, n_max, 0);
do n = 1 to n_max;
   b_esc[n]    = floor( n * lambda_e ); b_dees[n]   = ceil(  n * lambda_d );
end;
```

```
b_elim = j(n_dose, n_max, n_max + 1);
do dose = 1 to n_dose;
   do n = 3 to n_max;
      prob = 0; k    = 0;
      do while (prob <= stop_boundary[dose] & k < n);
         k    = k + 1; prob = 1 - cdf('BETA', target_eli, 1 + k, 1 + n - k);
      end;
      if prob <= stop_boundary[dose] then k = n + 1;
      b_elim[dose, n] = k;
   end;
end;
lam = lambda_e || lambda_d;
print "local BOIN", lam[colname={"lambda_e" "lambda_d"} format=8.4];
/*シミュレーション*/
call randseed(seed);
n_scen   = nrow(scenario);out_dose = j(n_scen * n_dose, 6, .);
out_sum  = j(n_scen, 5, .);cn = j(1, n_dose, "          ");
do i = 1 to n_dose; cn[i] = concat("dose_", strip(char(i)));end;
rn = {"Average number of patients",
      "Average number of DLTs",
      "MTD selection rate (%)",
      "Stop rate (%)",
      "MTD not selected (%)",
      "Stop + not selected (%)",
      "Average total sample size"};
do s = 1 to n_scen;
   mtd_results     = j(1, n_dose + 2, 0); number_patients = j(1, n_dose, 0);
   number_dlt      = j(1, n_dose, 0); total_patients  = 0;
   do sim = 1 to n_simulations;
      run BOIN(scenario[s,], mtd_sim, np_sim, nd_sim);
      mtd_results     = mtd_results + mtd_sim; total_patients  = total_patients + sum(np_sim);
      number_patients = number_patients + np_sim; number_dlt      = number_dlt + nd_sim;
   end;
   mtd_rate           = 100 * mtd_results / n_simulations;
   average_patients   = number_patients / n_simulations;
   average_dlt_number = number_dlt / n_simulations;
   result_data = j(7, n_dose, .);
   result_data[1, ]  = round(average_patients,   0.01);
   result_data[2, ]  = round(average_dlt_number, 0.01);
   result_data[3, ]  = round(mtd_rate[1, 3:(n_dose+2)], 0.01);
   result_data[4, 1] = round(mtd_rate[1], 0.01);
   result_data[5, 1] = round(mtd_rate[2], 0.01);
   result_data[6, 1] = round(mtd_rate[1] + mtd_rate[2], 0.01);
   result_data[7, 1] = round(sum(average_patients), 0.01);
   if display_result = 1 then do;
      ttl = concat("=== local BOIN : Scenario ", strip(char(s)),
                   "  true DLT rate = ", rowcatc(char(scenario[s,], 6, 2)), " ===");
      print ttl;
      print result_data[rowname=rn colname=cn format=10.2];
   end;
   do d = 1 to n_dose;
      r = (s-1)*n_dose + d;
      out_dose[r, 1] = s; out_dose[r, 2] = d;  out_dose[r, 3] = scenario[s, d];
      out_dose[r, 4] = average_patients[d]; out_dose[r, 5] = average_dlt_number[d];
      out_dose[r, 6] = mtd_rate[d + 2];
   end;
   out_sum[s, 1] = s; out_sum[s, 2] = mtd_rate[1]; out_sum[s, 3] = mtd_rate[2];
   out_sum[s, 4] = mtd_rate[1] + mtd_rate[2]; out_sum[s, 5] = total_patients / n_simulations;
end;
create LOCAL_DOSE from out_dose[colname={"scenario" "dose" "true_dlt"
                                         "avg_n_patients" "avg_n_dlt" "mtd_sel_rate"}];
append from out_dose;
close LOCAL_DOSE;
create LOCAL_SUM from out_sum[colname={"scenario" "stop_rate" "not_selected_rate"
                                       "stop_plus_notsel" "avg_total_n"}];
append from out_sum;
close LOCAL_SUM;
quit;
```

B. 大域的ベイズ最適区間デザインのシミュレーションプログラム

```
/*大域的ベイズ最適区間デザイン*/
proc iml;
/*シミュレーション設定*/
scenario = {
  0.30 0.45 0.55 0.65 0.75,
  0.15 0.30 0.45 0.55 0.65,
  0.10 0.20 0.30 0.40 0.50,
  0.05 0.15 0.24 0.30 0.36,
  0.03 0.05 0.08 0.15 0.30,
  0.35 0.45 0.50 0.55 0.60
}; target = 0.30; p_saf = 0.6; p_tox = 1.4; target_eli = 0.30; n_max = 30;
n_max_dose = 12; cohort_size = 3; stop_boundary = {0.90 0.95 0.95 0.95 0.95};
start_dose = 1; n_simulations = 10000; mtd_target = 0.30; mtd_cut = 0.358; seed = 1234;
pi0 = 1/3;   pi1 = 1/3;   pi2 = 1/3; display_result = 1;
/*大域的ベイズ最適区間デザインによる閾値の計算*/
start global_boundary(n, phi1, phi2, pi0, pi1, pi2, b_e, b_d);
   eps = 1e-12;
   g0 = j(1, n+1, .);  g1 = j(1, n+1, .);  g2 = j(1, n+1, .);
   do k = 0 to n;
      aa = k + 1; bb = n - k + 1;
      I1 = cdf('BETA', phi1, aa, bb);I2 = cdf('BETA', phi2, aa, bb);
      g0[k+1] = (I2 - I1) / (phi2 - phi1); g1[k+1] = I1 / phi1;
      g2[k+1] = (1 - I2) / (1 - phi2);
   end;
   Aterm = pi0 # g0 - pi1 # g1; Bterm = pi2 # g2 - pi0 # g0; best = 0;  b_e = -1;  s = 0;
   do k = 0 to n;
      s = s + Aterm[k+1];
      if s < best - eps then do;
         best = s; b_e  = k;
      end;
   end;
   best = 0;  b_d = 0;  s = 0;
   do k = 1 to n+1;
      s = s + Bterm[k];
      if s <= best + eps then do;
         if s < best then best = s;
         b_d = k;
      end;
   end;
finish global_boundary;
/*等張回帰*/
start isoreg_weight(xin, wt);
   x = xin; n = ncol(x);
   if n <= 1 then return(x);
   lvlsets = 1:n;
   done = 0;
   do while (done = 0);
      viol = ( (x[1, 2:n] - x[1, 1:(n-1)]) < 0 );
      if any(viol) then do;
         i    = min(loc(viol)); lvl1 = lvlsets[i]; lvl2 = lvlsets[i+1];
         ilvl = loc( (lvlsets = lvl1) | (lvlsets = lvl2) );
         x[1, ilvl]       = sum( x[1,ilvl] # wt[1,ilvl] ) / sum( wt[1,ilvl] );
         lvlsets[1, ilvl] = lvl1;
      end;
      else done = 1;
   end;
   return(x);
finish isoreg_weight;
/*大域的ベイズ最適区間デザインの実装*/
start BOIN(dlt_probs, mtd_sim, n_patients, n_dlts)
      global(b_esc, b_dees, b_elim, n_max, n_max_dose,
             cohort_size, start_dose, mtd_target, mtd_cut);
   n_doses      = ncol(dlt_probs);  n_patients   = j(1, n_doses, 0);
   n_dlts       = j(1, n_doses, 0); cap_dose     = n_doses;
   mtd_sim      = j(1, n_doses + 2, 0); current_dose = start_dose; stopflag = 0;
   do while (stopflag = 0);
      m = min( cohort_size, n_max - sum(n_patients) );
      if m > 0 then do;
         u = j(1, m, .);
         call randgen(u, "Uniform");
         new_dlts = sum( u < dlt_probs[current_dose] );
      end;
      else new_dlts = 0;
```

```
      n_patients[current_dose] = n_patients[current_dose] + m;
      n_dlts[current_dose]     = n_dlts[current_dose] + new_dlts;
      np = n_patients[current_dose];
      nd = n_dlts[current_dose];  /*用量調整*/
      if nd <= b_esc[np] then do;   /*漸増*/
         if (current_dose ^= n_doses) & (current_dose < cap_dose) then
            current_dose = current_dose + 1;
         else if np >= n_max_dose then stopflag = 1;
      end;
      else if (np >= 3) & (nd >= b_elim[current_dose, np]) then do; /*除外*/
         if current_dose = 1 then do;
            cap_dose     = 0; current_dose = 0; stopflag     = 1;
         end;
         else do;
            cap_dose     = current_dose - 1; current_dose = current_dose - 1;
         end;
      end;
      else if nd >= b_dees[np] then do; /*漸減*/
         if current_dose ^= 1 then current_dose = current_dose - 1;
         else if np >= n_max_dose then stopflag = 1;
      end;
      else do; /*維持*/
         if np >= n_max_dose then stopflag = 1;
      end;
      if sum(n_patients) >= n_max then stopflag = 1;
   end;
   /*MTD選択*/
   if cap_dose = n_doses then cap_dose_index = j(1, n_doses, 1);
   else do;
      cap_dose_index = j(1, n_doses, 0);
      if cap_dose >= 1 then cap_dose_index[1, 1:cap_dose] = 1;
   end;
   dose_eli  = n_patients # cap_dose_index;
   dose_plus = 0; check      = 0;
   if all(dose_eli = 0) then do; mtd_sim[1] = 1; check = 1; end;
   else do;
      if any(dose_eli = 0) then do;
         zpd   = loc(dose_eli = 0); zpd_n = loc(n_patients = 0);
         if ncol(zpd_n) > 0 then do;
            keep_n = loc(zpd_n ^= n_doses);
            if ncol(keep_n) > 0 then do;
               zpd_n_eli = zpd_n[keep_n];
               do jj = 1 to ncol(zpd_n_eli);
                  d = zpd_n_eli[jj];
                  if n_patients[d+1] ^= 0 then dose_plus = d;
               end;
            end;
         end;
         keep_idx       = setdif(1:n_doses, zpd);
         n_dlts_mtd     = n_dlts[1, keep_idx]; n_patients_mtd = n_patients[1, keep_idx];
      end;
      else do;
         n_dlts_mtd     = n_dlts; n_patients_mtd = n_patients;
      end;
      phat_mean = (n_dlts_mtd + 0.05) / (n_patients_mtd + 0.1);
      phat_var  = ( (n_dlts_mtd + 0.05) # (n_patients_mtd - n_dlts_mtd + 0.05) ) /
                  ( ((n_patients_mtd + 0.1)##2) # (n_patients_mtd + 0.1 + 1) );
      phat = isoreg_weight(phat_mean, 1 / phat_var);
   end;
   if check = 0 then do;
      if all(phat > mtd_cut) then mtd_sim[2] = 1;
      else do;
         ix = loc(phat > mtd_cut);
         if ncol(ix) > 0 then phat[1, ix] = -1;
         dif       = abs(phat - mtd_target); mtd_index = loc( dif = min(dif) );
         mtd       = dose_plus + max(mtd_index); mtd_sim[mtd+2] = 1;
      end;
   end;
finish BOIN;
/*用量調整表*/
n_dose = ncol(scenario);nvec   = 1:n_max;phi1 = target * p_saf;phi2 = target * p_tox;
b_esc    = j(1, n_max, 0);b_dees   = j(1, n_max, 0);
do n = 1 to n_max;
   run global_boundary(n, phi1, phi2, pi0, pi1, pi2, b_e, b_d);
   b_esc[n]    = b_e; b_dees[n]   = b_d;
end;
b_elim = j(n_dose, n_max, n_max + 1);
```

```
do dose = 1 to n_dose;
   do n = 3 to n_max;
      prob = 0; k    = 0;
      do while (prob <= stop_boundary[dose] & k < n);
         k    = k + 1; prob = 1 - cdf('BETA', target_eli, 1 + k, 1 + n - k);
      end;
      if prob <= stop_boundary[dose] then k = n + 1;
      b_elim[dose, n] = k;
   end;
end;
/*シミュレーション*/
print "global BOIN";
call randseed(seed);
n_scen   = nrow(scenario);out_dose = j(n_scen * n_dose, 6, .);
out_sum  = j(n_scen, 5, .);cn = j(1, n_dose, "          ");
do i = 1 to n_dose; cn[i] = concat("dose_", strip(char(i)));end;
rn = {"Average number of patients",
      "Average number of DLTs",
      "MTD selection rate (%)",
      "Stop rate (%)",
      "MTD not selected (%)",
      "Stop + not selected (%)",
      "Average total sample size"};
do s = 1 to n_scen;
   mtd_results     = j(1, n_dose + 2, 0); number_patients = j(1, n_dose, 0);
   number_dlt      = j(1, n_dose, 0);  total_patients  = 0;
   do sim = 1 to n_simulations;
      run BOIN(scenario[s,], mtd_sim, np_sim, nd_sim);
      mtd_results     = mtd_results + mtd_sim; total_patients  = total_patients + sum(np_sim);
      number_patients = number_patients + np_sim; number_dlt      = number_dlt + nd_sim;
   end;
   mtd_rate           = 100 * mtd_results / n_simulations;
   average_patients   = number_patients / n_simulations;
   average_dlt_number = number_dlt / n_simulations;
   result_data = j(7, n_dose, .);
   result_data[1, ]  = round(average_patients,   0.01);
   result_data[2, ]  = round(average_dlt_number, 0.01);
   result_data[3, ]  = round(mtd_rate[1, 3:(n_dose+2)], 0.01);
   result_data[4, 1] = round(mtd_rate[1], 0.01); result_data[5, 1] = round(mtd_rate[2], 0.01);
   result_data[6, 1] = round(mtd_rate[1] + mtd_rate[2], 0.01);
   result_data[7, 1] = round(sum(average_patients), 0.01);
   if display_result = 1 then do;
      ttl = concat("=== global BOIN : Scenario ", strip(char(s)),
                   "  true DLT rate = ", rowcatc(char(scenario[s,], 6, 2)), " ===");
      print ttl; print result_data[rowname=rn colname=cn format=10.2];
   end;
   do d = 1 to n_dose;
      r = (s-1)*n_dose + d;
      out_dose[r, 1] = s; out_dose[r, 2] = d; out_dose[r, 3] = scenario[s, d];
      out_dose[r, 4] = average_patients[d]; out_dose[r, 5] = average_dlt_number[d];
      out_dose[r, 6] = mtd_rate[d + 2];
   end;
   out_sum[s, 1] = s; out_sum[s, 2] = mtd_rate[1]; out_sum[s, 3] = mtd_rate[2];
   out_sum[s, 4] = mtd_rate[1] + mtd_rate[2]; out_sum[s, 5] = total_patients / n_simulations;
end;
create GLOBAL_DOSE from out_dose[colname={"scenario" "dose" "true_dlt"
                                          "avg_n_patients" "avg_n_dlt" "mtd_sel_rate"}];
append from out_dose;
close GLOBAL_DOSE;
create GLOBAL_SUM from out_sum[colname={"scenario" "stop_rate" "not_selected_rate"
                                        "stop_plus_notsel" "avg_total_n"}];
append from out_sum;
close GLOBAL_SUM;
quit;
```

Rich Kaplan-Meier Plotの提案

森岡　裕

（イーピーエス株式会社）

The Rich Kaplan–Meier Plot: A New Visualization for Time-to-Event Data

Yutaka Morioka

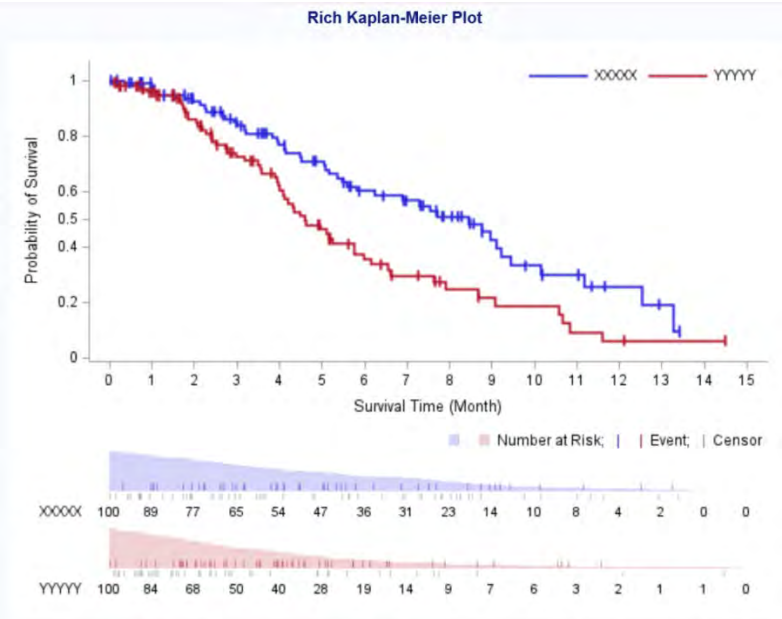
(EPS Corporation)

１.要旨

Kaplan-Meier（KM）plot は、臨床試験における time-to-event endpoint の結果報告に広く用いられている。一般的な KM plot では、各時点のリスク集合の大きさを示す Number at Risk Table がグラフ下部に併記される。特に追跡後期では、イベントや打ち切りによってリスク集合が減少し、1 イベントが KM 推定値に与える影響や推定の不確実性が大きくなるため、Number at Risk を併せて確認することが重要である。しかし、KM 曲線と表を行き来して情報を統合することは、認知的負荷を増加させる可能性がある。

また、一般的な KM plot では、打ち切りを KM 曲線上のマーカーで表現するため、打ち切り時点が Y 軸方向にも分散し、時間方向における打ち切りの集中や疎密を直感的に把握しにくい可能性がある。

そこで本研究では、Number at Risk と打ち切りの情報を時間軸との対応を保ちながら、より直接的かつ統合的に提示する Rich Kaplan-Meier Plot を提案する（下図）。本可視化により、追跡後期における情報量の減少と打ち切りの時間的分布を、従来の KM plot より直感的に把握できることを目指す.



Keywords：Kaplan–Meier Plot; Data Visualization; Number at Risk; Censoring; Clinical Trials

2. KM plot の表現の変遷

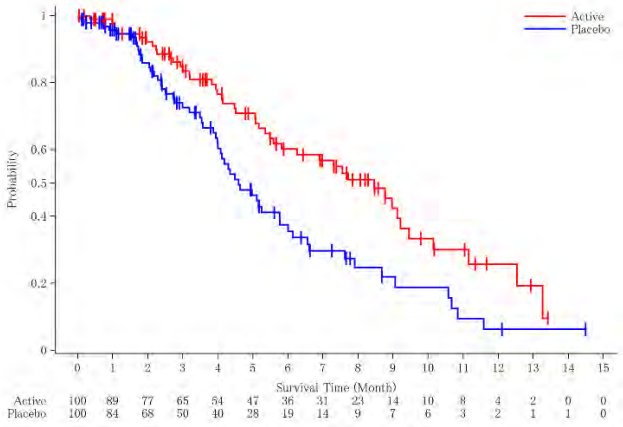
Kaplan-Meier（KM）推定量が初めて提案された Kaplan と Meier の論文、 Nonparametric Estimation from Incomplete Observations（1958）[1] では、現在一般的に用いられているような KM plot は示されておらず、主として推定方法そのものの説明が行われていた。

KM plot の表現がその後どのように変化してきたかを歴史的に正確にたどることは容易ではないが、少なくとも 1960 年代のランダム化比較試験の報告には、KM 曲線を用いた生存時間解析の図を確認することができる [2]。これらの初期の KM plot では、現在一般的な打ち切りマーカーや Number at Risk Table が示されていない例もみられる。

1970 年代に入ると、KM 曲線上に打ち切り時点をマーカーで示す表現が徐々にみられるようになる [3]。また、Number at Risk についても、現在のような表形式ではなく、図の脚注や本文中に各時点の人数を列記する形で補足する報告がみられるようになった。

その後、1980 年代から 1990 年代にかけて KM plot の表示形式は次第に現在の形に近づき、1990 年代以降は、多少のデザイン上の違いはあるものの、KM 曲線上に打ち切りマーカーを配置し、グラフ下部に Number at Risk Table を併記する形式が広く用いられるようになった。

このように、KM plot は当初、生存曲線そのものを示す比較的単純な表現から出発し、打ち切り情報や Number at Risk といった補助情報を段階的に加えながら、現在一般的にみられる以下のような形式へと発展してきたと考えられる。



3. Kaplan–Meier (KM) plot の付加情報表現の課題

現在一般的に用いられている Kaplan–Meier（KM）plot では、生存確率を示す KM 曲線に加えて、Number at Risk Table や打ち切りマーカーなどの付加情報が表示される。これらは KM 曲線を適切に解釈するために重要な情報である一方、その表示方法にはいくつかの課題がある。

3.1. Number at Risk Table の課題

Number at Risk は、各時点においてイベントを経験しておらず、かつ観察下にある被験者数を示す。特に追跡後期では、イベントや打ち切りによってリスク集合が小さくなるため、1 イベントによる KM 推定値の低下幅が相対的に大きくなり、推定の不確実性も増大する。そのため、KM 曲線の後半を解釈する際には、曲線だけでなく、その時点の Number at Risk を併せて確認することが重要である。

しかし、一般的な KM plot では、Number at Risk は曲線とは別にグラフ下部の表として配置される。このため、読者は KM 曲線上の特定の時点を確認した後、視線を Number at Risk Table へ移動し、対応する時点

および治療群の値を探し、再び曲線に戻って両者を統合して解釈する必要がある。すなわち、視覚的に分離された複数の情報を対応付ける操作が必要となり、曲線のみを読む場合と比較して認知的負荷が増加する可能性がある。

さらに、Number at Risk Table は通常、3 か月、6 か月、12 か月など一定の時間間隔でのみ値を表示する。そのため、表に示された時点間においてリスク集合がどのように減少したかを直接把握することはできない。KM 曲線は連続した時間軸上に表示される一方、Number at Risk は離散的な時点で提示されるため、両者の情報密度にも違いがある。

3.2.打ち切り情報の課題

一般的な KM plot では、打ち切りは KM 曲線上に「｜」「＋」「〇」などのマーカーを配置することで表現される。この方法は、個々の打ち切りが発生した時点を確認できるという利点を持つ。一方、打ち切りは本来、主として「いつ観察が終了したか」という時間軸上の情報であるにもかかわらず、マーカーはその時点の KM 推定値に対応する Y 座標上に配置される。その結果、打ち切り情報は X 軸方向だけでなく Y 軸方向にも分散して表示される。

このような二次元的な配置では、特定の時間帯に打ち切りが集中しているか、あるいはどの時点から打ち切りが増加しているかといった、時間方向の分布や疎密を直感的に把握しにくい可能性がある。特に複数の治療群を同時に表示する場合や打ち切り数が多い場合には、多数のマーカーが異なる Y 座標に配置されるため、この問題はより顕著になる。

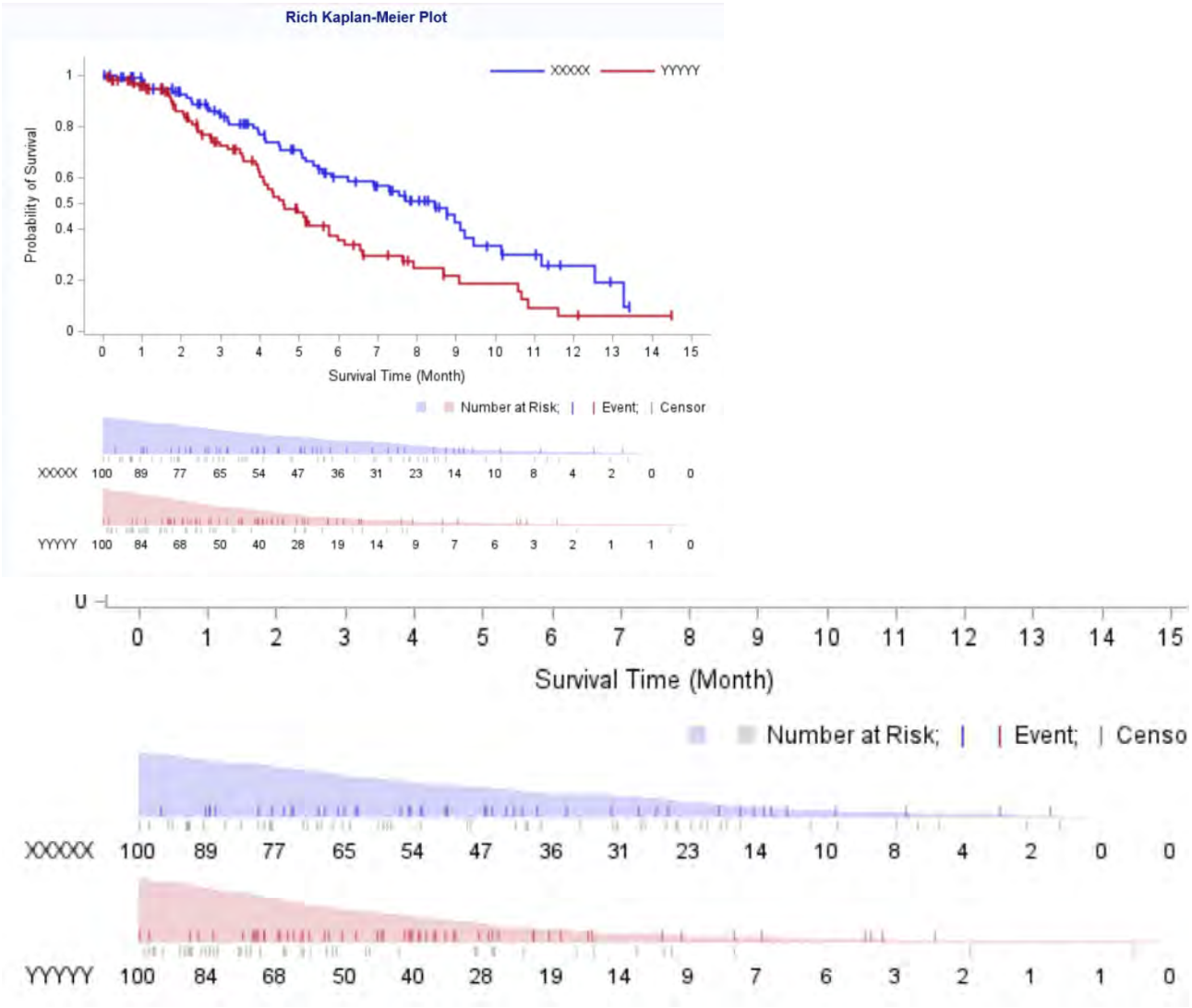
また、打ち切りマーカーは KM 曲線そのものに重ねて表示されるため、読者の注意は KM 曲線の形状に向きやすく、マーカーが独立した情報系列として認識されにくい可能性がある。個々の打ち切り時点を示すという従来の表現目的には適しているものの、「どの時間帯で観察対象が多く失われているか」という分布としての情報を読み取ることには必ずしも適していない。

3.3.付加情報を統合して解釈する際の課題

Number at Risk と打ち切りは互いに独立した情報ではない。時間の経過に伴うリスク集合の減少は、イベントの発生だけでなく打ち切りによっても生じる。そのため、KM 曲線の信頼性を時間軸に沿って理解するためには、KM 曲線の変化、Number at Risk の減少、打ち切りの発生を関連付けて解釈することが望ましい。しかし、従来の KM plot では、KM 推定値は曲線、打ち切りは曲線上のマーカー、Number at Risk はグラフ下部の表という異なる視覚表現に分散している。読者はこれらを頭の中で対応付けながら、「どこまで十分な被験者が観察されているのか」「どの時間帯で情報量が減少しているのか」を判断する必要がある。したがって、現在の KM plot が持つ統計情報を維持しながら、Number at Risk と打ち切りを時間軸上でより連続的かつ統合的に把握できる表現には、改善の余地があると考えられる。

4. Rich Kaplan–Meier Plot（RKM）による課題の解決

前章で述べた従来の Kaplan-Meier（KM）plot の課題に対し、本研究では Number at Risk および打ち切り情報の視覚表現を拡張した Rich Kaplan-Meier Plot（RKM）を提案する。RKM では、従来の KM 曲線そのものを大きく変更するのではなく、その下部に表示される付加情報を時間軸に沿った視覚表現として再構成することで、KM 曲線とリスク集合、イベントおよび打ち切りの関係をより直接的に把握できるようにする。



4.1. Number at Risk の連続的な視覚化

従来の Number at Risk Table では、Number at Risk はあらかじめ設定された時点の数値として離散的に表示される。そのため、表に示された時点間でリスク集合がどのように減少しているかを直接把握することは難しい。

RKM では、従来の Number at Risk の数値表示を保持しながら、その推移を KM 曲線と共通の時間軸上に面積として表示する。これにより、読者は個々の数値を参照するだけでなく、時間経過に伴ってリスク集合がどの程度残っているかを連続的な視覚情報として捉えることができる。

特に追跡後期では、Number at Risk の減少に伴って表示される領域も縮小する。このため、KM 曲線の後半を解釈する際に、Number at Risk Table の数値を逐次確認しなくても、「この時点では既にリスク集合が小さい」という情報を視覚的に認識しやすくなる。これは、少数の被験者に基づく追跡後期の KM 推定値を過度に解釈することを避けるための補助となることが期待される。

一方、正確な Number at Risk を確認する必要がある場合には、従来と同様に各時点の数値を参照できる。すなわち RKM では、概略的な変化を面積によって把握する視覚表現と、正確な値を確認する数値表現を併存させている。

4.2. イベントおよび打ち切りの一次元的な配置

従来の KM plot では、打ち切りマーカーは KM 曲線上に配置される。そのため、打ち切り時点という時間情報が、各時点の KM 推定値に応じて Y 軸方向にも分散して表示される。この表現は個々の打ち切り時点を確認するには適している一方、時間方向における打ち切りの集中や疎密を把握する際には、二次元に分散したマーカーを時間軸方向へ読み替える必要がある。

RKM では、打ち切りマーカーを KM 曲線上から分離し、各治療群の付加情報領域において共通の時間軸上にラグプロットとして配置する。これにより、打ち切り情報を主として時間方向の一次元的な分布として捉えることが可能となり、どの時間帯に打ち切りが集中しているかを視覚的に把握しやすくする。

さらに RKM では、イベントについても同じ時間軸上にマーカーとして表示する。イベントと打ち切りを異なるマーカーとして並置することで、リスク集合の減少が生じている時間帯と、その背景にあるイベントおよび打ち切りの発生状況を対応付けて観察できる。

4.3. KM 曲線と付加情報の時間軸による統合

従来の KM plot では、KM 推定値は曲線、打ち切りは曲線上のマーカー、Number at Risk はグラフ下部の表という異なる形式で表現される。そのため、読者はこれらの情報を視線移動しながら頭の中で対応付ける必要がある。

RKM では、KM 曲線、Number at Risk の推移、イベントおよび打ち切りを、すべて共通の X 軸、すなわち時間軸によって対応付ける。たとえば KM 曲線の特定の時点に着目した際、その直下を見ることで、その時点付近のリスク集合の大きさとイベント・打ち切りの発生状況を確認できる。

この構成の目的は、すべての情報を KM 曲線そのものに重ねることではなく、異なる統計情報を時間という共通の位置情報によって整列させることにある。これにより、曲線と表を独立した情報として読み取った後に統合するのではなく、時間軸に沿って関連する情報を近接した位置から取得できる。

4.4. RKM が目指す情報量の直感的な把握

RKM におけるこれらの表現は、KM 推定値そのものを変更するものではない。従来の KM plot で示される生存確率に加えて、「その時点までどの程度の被験者が観察対象として残っているか」「どの時点でイベントが発生しているか」「どの時点で打ち切りが生じているか」という、KM 曲線を解釈するための背景情報を視覚的に統合するものである。

特に RKM では、時間の経過に伴う Number at Risk の減少を面積として、イベントと打ち切りの発生を時間軸上のマーカーとして表現する。これらを組み合わせることで、追跡初期には多くの被験者から情報が得られている一方、追跡後期では徐々に情報が限られていくという time-to-event 解析特有のデータ構造を、一つの図の中で把握しやすくすることを目的としている。

したがって RKM は、従来の KM plot を置き換えるものではなく、KM 曲線が持つ情報を保持しながら、その解釈に必要な Number at Risk、イベントおよび打ち切りの情報への視覚的アクセスを改善する拡張表現として位置付けられる。

5. RKM 作成 SAS プログラムの説明

Appendix 1.に 2 群の RKM のサンプルコードを掲載したが、長大なため、特徴的な部分のみ本稿で説明する.

概略としては以下の流れでコード開発した.

ODS OUTPUT による PROC LIFETEST 結果の取得→ PROC STDIZE による Number at Risk の視覚量への変換→ Event/Censor の時間変数への変換 → GTL の LATTICE レイアウトで KM 曲線＋BANDPLOT（Num at risk のボリューム）＋SCATTERPLOT（ラグプロット）による描画.

[1]　PROC LIFETEST の ODS 出力を二次利用する

```
01  ods output Survivalplot=SurvivalPlotData;
02  ods output ProductLimitEstimates=ProductLimitEstimates;
03
04  proc lifetest data=dummy_adtte
05       plots=survival(atrisk=0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 cl);
06    time AVAL * CNSR(1);
07    strata TRTPN;
08    id USUBJID;
09  run;
```

SAS における RKM 作成では「PROC LIFETEST のグラフを加工する」のではなく、「PROC LIFETEST が計算した結果を取得し、GTL(Graph Template Language)でグラフを再構築する」アプローチが適している.

[2] PROC STDIZE による Number at Risk の視覚量への変換

```
01  proc stdize data= _SurvivalPlotData_1
02       out = atrisk_volum/*(keep=Stratum atrisk time)*/
03       method=range
04       add = 0
05       mult = 1
06  ;
07  where ^missing(Stratum);
08  by Stratum;
09  var atrisk;
10  run;
```

PROC STDIZE は様々な形の標準化を実装可能なプロシジャである．METHOD=RANGE を使用することで、治療群ごとに Number at Risk を 0～1 の範囲に標準化する。この前の段階で、最終的に 0 の Number at Risk をデータに作っているため、一番最初の Number at Risk 数が最大値 1、最後の 0 例数が 0 とし標準化される．その標準化された値を後ほど BANDPLOT の高さとして使用している

[3] GTL (Graph Template Language)の LATTICE＋BANDPLOT＋SCATTERPLOT による描画

```
01  layout lattice /
02       rows=5
03       rowweights=(0.68 0.09 0.05 0.09 0.05)
04       rowgutter=0.01
```

GTL (Graph Template Language)の LATTICE layout を利用し、KM 曲線、Number at Risk の視覚表現、Event/Censor および従来型の Number at Risk Table を、共通の時間軸を持つ複数の領域として配置した。配置比率の例は以下になる

1	Kaplan-Meier Curve	68%
2	グループ 1 の Num at Risk のプロットと打ち切り/イベントのラグプロット	9%
3	グループ 1 の Number at Risk Table	5%
4	グループ 2 の Num at Risk のプロットと打ち切り/イベントのラグプロット	9%
5	グループ 2 の Number at Risk Table	5%

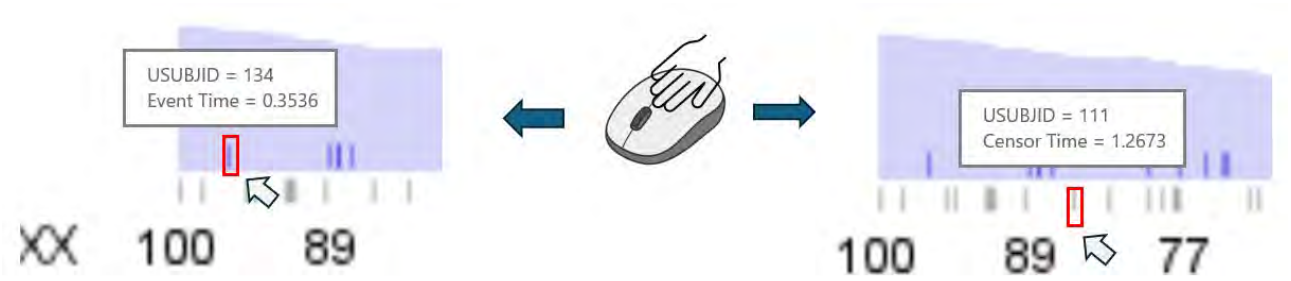
[4] インタラクティブグラフの実装

```
01  rolename=(tip1=USUBJID tip2=Cens_time)
02  tip=(tip1 tip2)
03  tiplabel=(tip1="USUBJID" tip2="Censor Time")
```

```
01  rolename=(tip1=USUBJID tip2=Event_time)
02  tip=(tip1 tip2)
03  tiplabel=(tip1="USUBJID" tip2="Event Time")
```

```
01  ods graphics on /imagemap=on tipmax=5000;)
```

IMAGEMAP=ON と HTML 出力を使用しているため、静的なグラフだけでなくマーカーにマウスを重ねることで被験者 ID や Event/Censor time を確認できるインタラクティブな RKM を HTML 形式で作成することが可能である。

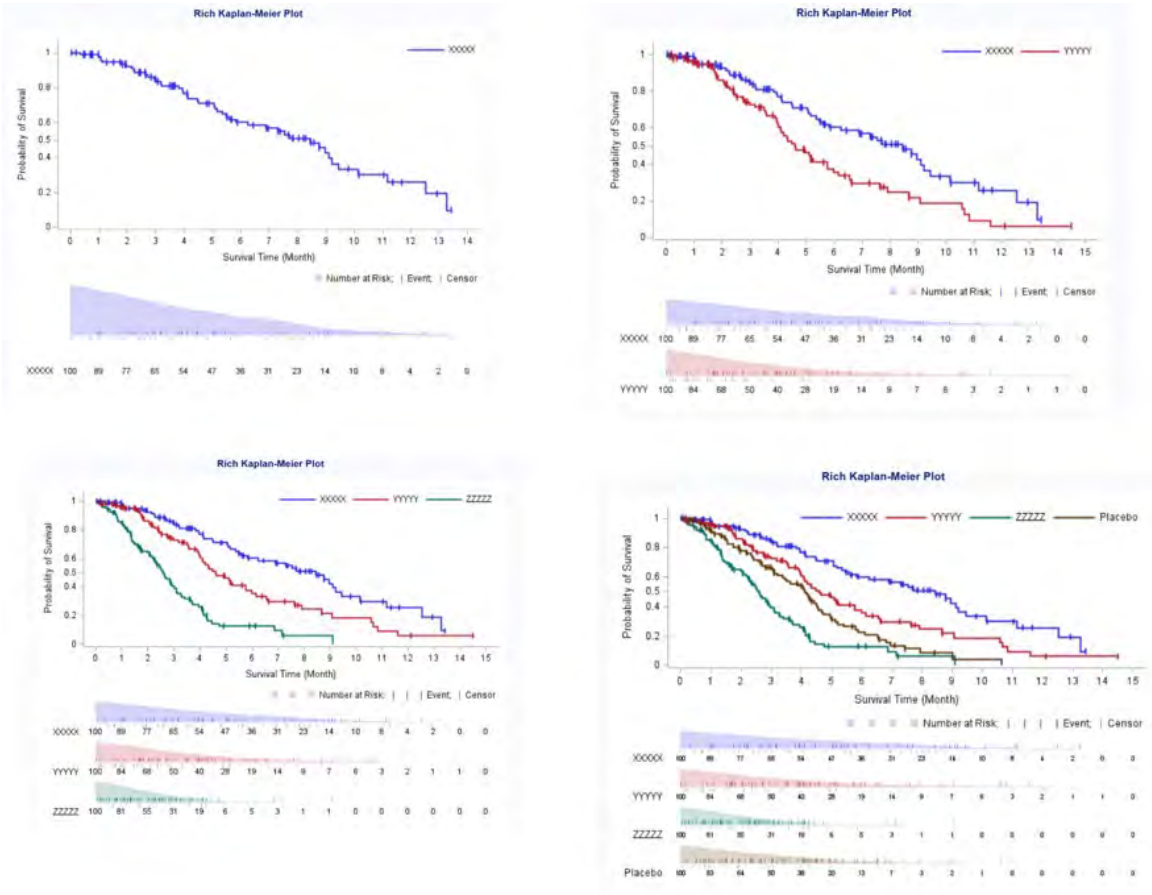


6. SAS Package [rich_kaplan_meier_plot]の紹介

Rich Kaplan-Meier Plot（RKM）の作図には Graph Template Language（GTL）を用いるため、利用者が独自に実装する場合には GTL に関する一定の知識と比較的複雑なプログラミングが必要となる。そこで、GTL の詳細な知識を必要とせず、必要な引数を指定することで RKM を簡便に作図できるよう、SAS Packages Framework（SPF）を用いた SAS Package：rich_kaplan_meier_plot を開発し、GitHub 上で公開している。

本パッケージでは、解析対象データや time-to-event 解析に必要な変数等を引数として指定することで、前節で示したデータ加工および GTL による描画処理をパッケージ内部で実行し、RKM を作成することができる。これにより、利用者は RKM の描画に必要となる一連の GTL コードを直接記述することなく、本可視化を利用することが可能である。

現時点では、1 群から 4 群までの RKM の作図に対応している。パッケージの導入方法、各引数の指定方法および実行例については、GitHub リポジトリ（PharmaForest/rich_kaplan_meier_plot）の README に掲載している。

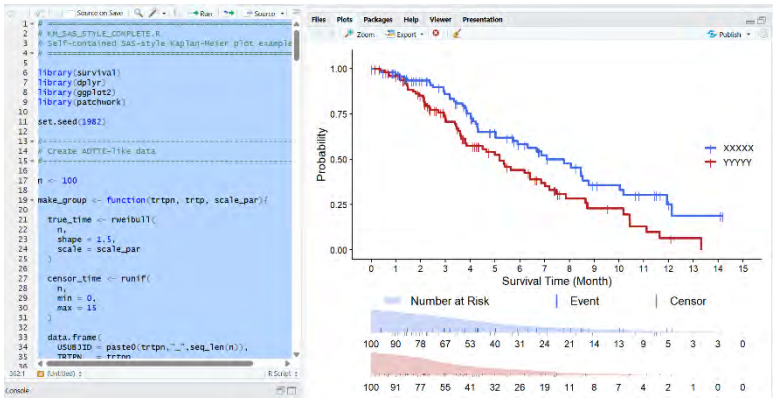


7.R 言語での実装例

Rich Kaplan-Meier Plot（RKM）のような新たな可視化手法を広く利用可能なものとするためには、特定のソフトウェアやプログラミング言語に依存せず、複数の環境で実装できることが望ましい。本研究では主として SAS による RKM の実装を行ったが、R 言語においても同様の可視化が実現可能であることを確認した。R では、生存時間解析およびグラフィックスに関する既存の機能を組み合わせることで、Number at Risk の推移、イベントおよび打ち切りの時間的分布を、KM 曲線と共通の時間軸上に配置した RKM を比較的容易に実装することができた。

このことは、RKM が SAS の Graph Template Language（GTL）に依存した表現ではなく、Number at Risk、イベントおよび打ち切りを時間軸上で統合的に提示するという可視化の考え方自体が、異なる解析環境にも展開可能であることを示している。

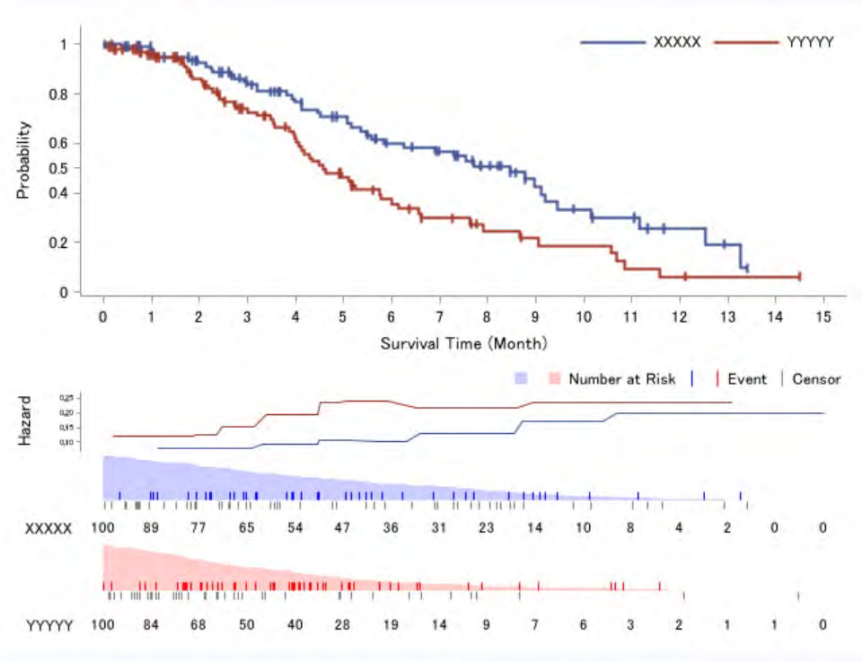
R 言語による実装例については、SAS ユーザー総会の Web サイトに参考プログラムとして公開する予定である。現時点では参考実装の位置付けであるが、今後はより簡便に RKM を利用できるよう、R パッケージとしての提供についても検討している。



8.将来的な改良

RKM は、新たな生存時間データのビジュアライゼーション手法として一定の有用性を有すると考えられる一方、現時点でその方法論が完成されたものとは考えておらず、今後さらに改良・拡張する余地がある。

その一つとして、時間依存性を有するハザードの表現方法が挙げられる。現状の表示では、局所的な変動が視覚的な解釈を難しくする場合があるため、下図に示すようにハザードを平滑化して表示することが考えられる。平滑化によって短期的な変動の影響を抑え、時間経過に伴うハザードの変化傾向をより直感的に把握できる可能性がある。一方で、平滑化の程度によっては重要な局所的変化が失われる可能性もあることから、平滑化手法やパラメータの選択については、今後検討する必要がある。



また、ラグプロットにおける打ち切り情報の表現についても改良の余地がある。特に、打ち切りに複数の種類や理由が存在する場合には、それらを同一の記号として扱うのではなく、シンボルや色を変更して区別することが有効と考えられる。これにより、打ち切りの発生時点だけでなく、その種類や分布についても同一の図上で把握することが可能となり、打ち切り機構とイベント発生との関係を探索的に検討する際にも有用であると考えられる。

さらに、RKM の視認性および解釈可能性を高めるためには、表示する情報量をデータの特性に応じて調整できる仕組みも重要である。例えば、症例数が多い場合には点や線が重なり、個々の観測値の把握が困難になる可能性がある。このような場合には、点の透過度や大きさの調整、重なりを考慮した表示、あるいは表示対象となる情報の選択などを導入することで、大規模なデータに対しても視認性を維持できると考えられる。

加えて、群間比較への拡張も今後の課題である。治療群や患者背景などによって層別した RKM を比較可能な形で提示することができれば、Kaplan-Meier 曲線のみでは把握しにくい、イベント発生時期や打ち切りの分布、さらには時間依存的なパターンの群間差を視覚的に探索することが可能になると考えられる。ただし、群数の増加に伴って図が複雑化するため、色、シンボル、配置などを含めた適切な表示方法について検討が必要である。

また、RKM から得られる視覚的な特徴と統計的指標との対応関係を整理することも重要である。例

えば、RKM 上で特徴的に観察されるパターンが、比例ハザード性からの逸脱や時間依存的な治療効果などとどのように関連するかを明らかにすることで、単なる可視化にとどまらず、後続する統計解析やモデル選択を支援する探索的ツールとしての役割をより明確にできる可能性がある。

今後は、これらの改良を進めるとともに、シミュレーションデータや実データを用いて、RKM によってどのような特徴が把握しやすくなるのかを体系的に評価する必要がある。特に、従来の Kaplan-Meier 曲線等との比較を通じて、RKM が付加的に提供する情報とその適用場面を明確にすることが、本手法の実用性を確立する上で重要な課題である。

9.Raincloud plot によせて

本稿で提案する RKM は、Mika Allen 氏らによる Raincloud plot の提案と、その後の普及に向けた一連の活動から強い影響を受けている。氏らは、Raincloud plot という新たなデータビジュアライゼーション手法をオープンな形で発表[4]するとともに、特定のソフトウェアやプログラミング言語に限定することなく、多様な環境での実装を試み、その成果を広く共有してきた。こうした取り組みは、単に新しいグラフを提案するだけでなく、誰もが実際に利用し、検証し、さらに改良できる環境を整えることの重要性を示している。

RKM の開発にあたっては、この姿勢から多くの示唆を得た。新しいビジュアライゼーション手法を提案する際には、その方法を論文として提示するだけでは十分ではない。手法そのものをオープンに議論できる形で公開し、図を作成するためのコードや実装例、そこから得られた成果についても可能な限り共有することで、第三者による利用、検証、改良が可能となる。そして、そのような過程を通じて初めて、一つのアイデアが実際に利用されるビジュアライゼーション手法へと発展していく可能性があると考える。

本稿において RKM を提案し、その実装や成果を積極的に公開しようとする背景には、Raincloud plot が示したこのようなオープンな研究・開発のあり方がある。RKM についても、完成された手法として提示するのではなく、多くの研究者や実務家が利用し、議論し、必要に応じて改良することのできる手法として発展していくことを期待している。

最後に、Raincloud plot という優れたビジュアライゼーション手法を提案したことに加え、新しい可視化手法をどのように共有し、育てていくかという一つのあり方を示した Mika Allen 氏ならびに共同研究者の方々に、深い敬意と感謝を表したい。

まとめ

本稿では、臨床試験における time-to-event endpoint の結果提示に広く用いられている Kaplan-Meier（KM）plot について、その付加情報の提示方法に着目し、新たな可視化手法として Rich Kaplan-Meier Plot（RKM）を提案した。従来の KM plot では、KM 曲線、打ち切りマーカー、Number at Risk Table がそれぞれ異なる視覚表現として配置されており、読者はこれらの情報を視線移動しながら対応付けて解釈する必要がある。また、Number at Risk は限られた時点の数値として提示され、打ち切りは KM 曲線上に二次元的に分散して配置されるため、時間経過に伴う情報量の減少や、イベントおよび打ち切りの時間的な分布を直感的に把握しにくいという課題がある。

RKM では、こうした課題に対し、Number at Risk の推移を面積として連続的に可視化するとともに、イベントおよび打ち切りを KM 曲線とは分離した領域にラグプロットとして配置する。さらに、こ

れらを KM 曲線と共通の時間軸上に整列させることで、ある時点の KM 推定値を確認した際に、その時点でどの程度の被験者が観察下にあり、どのようなイベントや打ち切りが生じているかを、近接した位置から同時に把握できる構成とした。RKM は KM 推定量そのものを変更するものではなく、従来の KM plot が持つ統計情報を保持しつつ、その解釈に必要な背景情報への視覚的アクセスを改善する拡張表現として位置付けられる。

また、本稿では RKM の概念を提示するだけでなく、SAS における具体的な実装方法を示した。PROC LIFETEST の ODS 出力を二次利用し、Number at Risk を視覚量に変換した上で、GTL の LATTICE、BANDPLOT、SCATTERPLOT 等を組み合わせることで RKM を構成した。さらに、利用者が GTL の詳細な知識を必要とせず RKM を作図できるよう、SAS Packages Framework を用いた rich_kaplan_meier_plot パッケージを開発し、公開した。加えて、R 言語でも同様の可視化が実装可能であることを確認しており、RKM の考え方が特定の解析環境に依存しないことも示した。

一方、RKM は現時点で完成された可視化手法ではなく、今後さらに検討すべき点も多い。時間依存性を有するハザードの平滑化表示、打ち切り理由に応じたシンボルや色の使い分け、大規模データにおける視認性の確保、群間比較への拡張など、表示方法そのものにも改良の余地がある。また、RKM 上で認識される視覚的な特徴と、比例ハザード性からの逸脱や時間依存的治療効果などの統計的特徴との対応関係を明らかにすることも重要である。今後は、シミュレーションデータおよび実データを用いた評価を通じて、従来の KM plot と比較した際に RKM がどのような情報を付加的に提供し、どのような場面で特に有用となるのかを検証する必要がある。

本研究の目的は、従来の KM plot を置き換えることではなく、長年にわたり発展してきた KM plot の基本的な枠組みを維持しながら、その周辺に存在する情報をより自然に読み取れる形へ再構成することにある。とりわけ、time-to-event 解析では、時間の経過とともに観察対象が減少し、図の後半ほど限られた情報に基づいて推定が行われる。RKM は、この time-to-event データ特有の構造を、KM 曲線、Number at Risk、イベント、打ち切りという複数の情報を時間軸上で統合することによって、より直感的に提示することを目指したものである。

さらに、本稿では Raincloud plot の普及過程から得た示唆を踏まえ、RKM についても手法の提案にとどまらず、実装やコードをオープンに共有し、第三者による利用、検証、改良を可能とすることを重視した。新しい可視化手法は、一度の提案によって完成するものではなく、多くの利用者による実践と議論を通じて改善され、次第に有用な表現として定着していくものと考える。

RKM についても、本稿を一つの出発点として、今後さまざまな実装や改良が加えられ、生存時間解析をより適切に理解し伝えるための可視化手法の一つとして発展していくことを期待する。

参考文献

[1] Kaplan EL、 Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958;53(282):457–481. doi:10.1080/01621459.1958.10501452.

[2] Freireich EJ、 Gehan EA、 Frei E III、 et al. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: a model for evaluation of other potentially useful therapy. *Blood.* 1963;21(6):699–716. doi:10.1182/blood.V21.6.699.699.

[3] Peto R、 Pike MC、 Armitage P、 Breslow NE、 Cox DR、 Howard SV、 Mantel N、 McPherson K、 Peto J、 Smith PG. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer.* 1977;35(1):1–39. doi:10.1038/bjc.1977.1.

[4] Allen M、 Poggiali D、 Whitaker K、 Marshall TR、 van Langen J、 Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*. 2021;4:63. doi:10.12688/wellcomeopenres.15191.2.

Appendix 1.

```
01  data dummy_adtte;
02  attrib
03  USUBJID label="Unique Subject Identifier" length=$20.
04  TRTP    label="Planned Treatment" length=$20.
05  TRTPN              label="Planned Treatment (N)" length=8.
06  PARAM    label="Parameter" length=$50.
07  PARAMCD label="Parameter Code" length=$20.
08  PARAMN    label="Parameter (N)" length=8.
09  AVAL    label="Analysis Value" length=8.
10  CNSR    label="Censor" length=8.
11  ;
12  call streaminit(1982);
13  do TRTPN = 1 to 2;
14  do _USUBJID = 1 to 100;
15  do PARAMN = 1 to 1;
16  if TRTPN =1 then time =rand('WEIBULL'、 1.5、 10);
17  else if TRTPN =2 then time =rand('WEIBULL'、 1.5、 7);
18  else if TRTPN =3 then time =rand('WEIBULL'、 1.5、 3);
19  else time =rand('WEIBULL'、 1.5、 5);
20  USUBJID = cats(TRTPN、_USUBJID);
21  censor_limit = rand('UNIFORM') * 15;
22  CNSR = ^(time <= censor_limit);
23  AVAL = min(time、 censor_limit);
24  TRTP = choosec(TRTPN、"XXXXX"、"YYYYY"、"ZZZZZ"、"Placebo");
25  PARAMCD = choosec(PARAMN、"PFS");
26  PARAM = choosec(PARAMN、"Progression Free Survival (Months)");
27  output;
28  end;
29  end;
30  end;
31  keep USUBJID -- CNSR;
32  run;
33  proc sort data=dummy_adtte(keep=TRTP TRTPN) out=group_fmt nodupkey;
34  by TRTP TRTPN;
35  run;
36  data group_fmt;
37  set group_fmt;
38  FMTNAME = "$KM_GR";
39  START = cats(TRTPN);
40  LABEL = TRTP;
41  call symputx(cats("label"、TRTPN)、LABEL);
42  run;
43  proc format cntlin=group_fmt;
44  run;
45  ods graphics on;
46  ods noresults;
47  ods select none;
48  ods output Survivalplot=SurvivalPlotData;
49  ods output ProductLimitEstimates =ProductLimitEstimates;
50  proc lifetest data=dummy_adtte plots=survival(atrisk=0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 cl);
51    time AVAL * CNSR(1);
52    strata TRTPN ;
53    id USUBJID;
54  run;
55  ods results;
56  ods select all;
57
58
59  proc sort data=SurvivalPlotData(keep = Stratum) out=Stratum nodupkey;
60    by Stratum;
61  run;
```

```
62  proc sort data=SurvivalPlotData(keep = tAtRisk) out=tAtRisk nodupkey;
63    where ^missing(tAtRisk);
64    by tAtRisk;
65  run;
66  data atrisk;
67  set Stratum;
68  if _N_=1 then do;
69     declare hash h1(dataset:"SurvivalPlotData(keep=Stratum tAtRisk)");
70     h1.definekey("Stratum"、 "tAtRisk");
71     h1.definedone();
72  end;
73  do i=1 to obs;
74     set tAtRisk nobs=obs point=i;
75     AtRisk=0;
76     if h1.check() ne 0 then output;
77  end;
78  run;
79  data SurvivalPlotData_1;
80  set SurvivalPlotData atrisk;
81  if ^missing(Censored) then do;
82     tick_marks_upper = Censored + 0.02;
83     tick_marks_lower = Censored - 0.02;
84  end;
85  run;
86  proc sort data=SurvivalPlotData_1 out=_SurvivalPlotData_1;
87  by Stratum;
88  run;
89  proc stdize data= _SurvivalPlotData_1
90         out = atrisk_volum/*(keep=Stratum atrisk time)*/
91         method=range
92         add = 0
93         mult = 1
94  ;
95  where ^missing(Stratum);
96  by Stratum;
97  var atrisk;
98  run;
99
100 data SurvivalPlotData_2;
101 set SurvivalPlotData_1 atrisk_volum(in=ina rename=(atrisk=atrisk_band));
102 if ina then do;
103   NaR_time=time;
104   call missing(of time);
105 end;
106 else do;
107    call missing(of atrisk_band);
108 end;
109
110 format atrisk_band best.;
111 run;
112
113 data Productlimitestimates_1;
114 set Productlimitestimates(rename=(STRATUM=_STRATUM));
115 where ^missing(USUBJID);
116 if Censor = 0 then Event_time=AVAL;
117 if Censor ne 0 then Cens_time=AVAL;
118 STRATUM=cats(_STRATUM);
119 keep STRATUM USUBJID Event_time Cens_time;
120 run;
121
122 data SurvivalPlotData_3;
123    set SurvivalPlotData_2(in=ina) Productlimitestimates_1(in=inb );
124
```

```
125     length TRTNAME $20;
126     TRTNAME=put(Stratum、$KM_GR.);
127
128     dumm_y1=0;
129
130   if ina then do;
131     if Stratum="1" then do;
132           band1_upper = atrisk_band;
133           band1_lower = dumm_y1;
134           risk1             = atrisk;
135           label risk1="&label1";
136     end;
137     /* YYYYY */
138     if Stratum="2" then do;
139           band2_upper = atrisk_band;
140           band2_lower = dumm_y1;
141           risk2             = atrisk;
142           label risk2="&label2";
143     end;
144   end;
145
146    if inb then do;
147
148     if Stratum="1" then do;
149           censor1          = dumm_y1;
150             event1          = dumm_y1;
151     end;
152     /* YYYYY */
153     if Stratum="2" then do;
154           censor2          = dumm_y1;
155           event2          = dumm_y1;
156     end;
157   end;
158
159    bar="|";
160   run;
161
162
163
164   proc template;
165
166   define statgraph KM_PLOT_SPLIT;
167
168   dynamic _title;
169
170   begingraph / border=false;
171
172   entrytitle _title;
173
174   discreteattrmap name='trtmap';
175
176       value '1' /
177             lineattrs=(color=blue thickness=2)
178             markerattrs=(color=blue)
179             fillattrs=(color=blue);
180
181       value '2' /
182             lineattrs=(color=red thickness=2)
183             markerattrs=(color=red)
184             fillattrs=(color=red);
185
186   enddiscreteattrmap;
187
```

```
discreteattrvar attrvar=TRTATTR
                        var=Stratum
                        attrmap='trtmap';

legendItem type=FILL name="red1" /
    fillattrs=(color=red transparency =0.8) filldisplay=(fill) label="Number at Risk";
legendItem type=FILL name="blue1" /
    fillattrs=(color=blue transparency =0.8) filldisplay=(fill) ;
legendItem type=text name="red2" /
    text="|" textattrs=(color=red )    label="Event";
legendItem type=text name="blue2" /
    text="|" textattrs=(color=blue ) ;
legendItem type=text name="gray" /
    text="|" textattrs=(color=gray )    label="Censor";




layout lattice / outerpad=(top=0px bottom=0px)    pad=(top=0px bottom=0px left=0px right=0px)
    rows=5
    rowweights=(0.68 0.09 0.05 0.09 0.05)
    rowgutter=0.01;

    /*****************************************/
    /* Row1 : KM Curve                       */
    /*****************************************/

    layout overlay /
        walldisplay=none

        xaxisopts=(
            label='Survival Time (Month)'
            linearopts=(
                viewmin=0
                viewmax=15
                tickvaluelist=(0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15)
            )
        )

        yaxisopts=(
            label='Probability'
            linearopts=(
                viewmin=0
                viewmax=1.05
                tickvaluelist=(0 .2 .4 .5 .6 .8 1.0)
            )
        );

        stepplot
            x=time
            y=survival
            /
            group=TRTATTR
            lineattrs=(thickness=2)
            name='step';

        scatterplot
            x=time
            y=censored/
            group=TRTATTR
            yerrorupper=tick_marks_upper
            yerrorlower=tick_marks_lower
            markerattrs=(size=0)
```

```
            errorbarattrs=(thickness=2)
            errorbarcapshape=none;

        discretelegend 'step' /
            location=inside
            halign=right
            valign=top
            border=false;
    discretelegend "blue1" "red1" "blue2" "red2" "gray"/location=outside halign=right valign=bottom across=5 border=false;


    endlayout;


    /****************************************/
    /* Row2 : XXXXX Band                    */
    /****************************************/

    layout overlay /outerpad=(top=0px bottom=0px)
        walldisplay=none

        xaxisopts=(display=none)

        yaxisopts=(
            display=none
            linearopts=(viewmin=-0.15 viewmax=1)
            offsetmax=0
            offsetmin=0
        );

        bandplot
            x=NaR_time
            limitupper=band1_upper
            limitlower=band1_lower /
            fillattrs=(color=blue)
            datatransparency=0.8
;

        scatterplot
            x=Cens_time
            y=censor1 /
            markercharacter=bar
            markercharacterposition=bottom
            markercharacterattrs=(
                color=gray
                size=3
            )
            rolename=(tip1=USUBJID tip2=Cens_time)
            tip=(tip1 tip2)
            tiplabel=(tip1="USUBJID" tip2="Censor Time")
;
        scatterplot
            x=Event_time
            y=event1 /
            markercharacter=bar
            markercharacterposition=top
            markercharacterattrs=(
                color=blue
                size=3
            )
            rolename=(tip1=USUBJID tip2=Event_time)
            tip=(tip1 tip2)
            tiplabel=(tip1="USUBJID" tip2="Event Time")
```

```
;

    endlayout;


    /*****************************************/
    /* Row3 : XXXXX At Risk                  */
    /*****************************************/

    layout overlay / pad=(top=0px bottom=0px left=0px right=0px) outerpad=(top=0px bottom=5px)
        walldisplay=none

        xaxisopts=(display=none)
        yaxisopts=(display=none);

        axistable
            x=tAtRisk
            value=risk1 /
            valueattrs=(size=8);

    endlayout;


    /*****************************************/
    /* Row4 : YYYYY Band                     */
    /*****************************************/

    layout overlay / pad=(top=0px bottom=0px left=0px right=0px) outerpad=(top=0px bottom=0px)
        walldisplay=none

        xaxisopts=(display=none)

        yaxisopts=(
            display=none
            linearopts=(viewmin=-0.15 viewmax=1)
           offsetmax=0
           offsetmin=0

        );

        bandplot
            x=NaR_time
            limitupper=band2_upper
            limitlower=band2_lower /
            fillattrs=(color=red)
            datatransparency=.8;

        scatterplot
            x=Cens_time
            y=censor2 /
            markercharacter=bar
            markercharacterposition=bottom
            markercharacterattrs=(
                color=gray
                size=3
            )
            rolename=(tip1=USUBJID tip2=Cens_time)
            tip=(tip1 tip2)
            tiplabel=(tip1="USUBJID" tip2="Censor Time")
;

        scatterplot
            x=Event_time
```

```
            y=event2 /
            markercharacter=bar
            markercharacterposition=top
            markercharacterattrs=(
                color=red
                size=3
            )
            rolename=(tip1=USUBJID tip2=Event_time)
            tip=(tip1 tip2)
            tiplabel=(tip1="USUBJID" tip2="Event Time")
;

    endlayout;


    /****************************************/
    /* Row5 : YYYYY At Risk                 */
    /****************************************/

    layout overlay / pad=(top=0px bottom=0px left=0px right=0px) outerpad=(top=0px bottom=0px)
        walldisplay=none

        xaxisopts=(
            display=none
            linearopts=(
                viewmin=-0
                viewmax=15
                tickvaluelist=(0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15)
            )
        )

        yaxisopts=(display=none);

        axistable
            x=tAtRisk
            value=risk2 /
            valueattrs=(size=8)
;

    endlayout;

endlayout;

endgraph;
end;
run;

ods select all;
ods graphics on /imagemap=on tipmax=5000;

ods html path=".¥RichKM" file="RichKM.html";

proc sgrender data=SurvivalPlotData_3 template=KM_PLOT_SPLIT;
   dynamic _title="";
   format Stratum $KM_GR.;
run;

ods html close;
```

SAS/IMLで実装する推定方程式：

修正ポアソン回帰, IPW, G-computationへの応用

○西尾　和恭

（イーピーエス株式会社）

Implementing Estimating Equations in SAS/IML:
Applications to Modified Poisson Regression, IPW, and G-Computation

Kazutaka Nishio
EPS Corporation

要旨

生物統計学で用いられる多くの推定量は，推定方程式の解として捉えることができる．例えば，回帰モデルのパラメータ推定や，因果推論における逆確率重み付け（Inverse Probability Weighting: IPW）推定量はその代表例である．そのため，推定方程式を柔軟に記述し，その解を数値的に求めることができれば，多様な推定量を SAS 上で実装できる可能性が広がる．本稿では，行列演算や数値計算を柔軟に実行できる IML プロシジャに着目し，推定方程式に基づく推定量の実装方法を解説する．推定方程式の記述から数値的な求解までの基本的な流れを紹介するとともに，実装例を通じて，異なる解析手法が共通の推定方程式の枠組みで理解できることを示す．推定方程式の視点から各手法を捉えることで，推定量の構造や仮定への理解が深まるだけでなく，標準的な SAS プロシジャでは実装が困難な推定量であっても，推定方程式を定式化できれば柔軟に実装できることを示す．

キーワード：IML プロシジャ，　推定方程式

1. はじめに

生物統計学で用いられる多くの推定量は推定方程式の解として捉えることができる．線形回帰やロジスティック回帰における回帰係数の推定をはじめ，修正ポアソン回帰や逆確率重み付け（Inverse Probability Weighting: IPW）推定量などの手法も適切な推定方程式を定義することで統一的に表現することができる．この視点は，推定量の導出や仮定を体系的に理解する上で重要であり，異なる解析手法の共通点や相違点を整理するうえでも有用である．

一方，SASによる統計解析では，GENMODプロシジャやLOGISTICプロシジャなどの既存プロシジャを利用することで多くの解析を実施できるが，推定方程式そのものを記述し数値的に解く機会は必ずしも多くない．

そのため, 新たな推定量を実装したい場合や, 既存プロシジャでは対応できない解析手法を適用したい場合などには, 実装上の制約が生じることがある. SAS/IMLプロシジャは, 行列演算や数値最適化を柔軟に実行できる環境を提供しており, 推定方程式を直接記述して解くことが可能である. これにより, 標準的な統計手法だけでなく, 既存のプロシジャでは実装されていない手法や, 研究者が独自に考案した推定量についても実装できる可能性が広がる. また, 推定方程式に基づいて推定量を構成することで, 傾向スコアやアウトカムモデルなどの複数のモデルを用いる解析においても, それぞれの推定過程を含めた推定量全体を一つの積み上げ推定方程式として表現することができる. このような表現を用いることで, 第1段階で推定されたモデルパラメータの不確実性を考慮した分散推定へ自然に拡張できるほか, 独自の尤度, 補正項, 推定対象などを含む推定量についても, 同一の枠組みで定式化・実装することが可能となる.

本稿では, 推定方程式に基づく推定量の実装という観点から, IMLプロシジャを用いて修正ポアソン回帰, IPW, G-computationを実装し, さらに二値アウトカムの誤分類を考慮するG-computationを提案し, 推定方程式を直接解くことによる実装方法を示す. これらの実装例を通じて, 異なる解析手法を推定方程式という統一的な枠組みで理解することの有用性を示すとともに, 複数の推定過程を含む推定量の構築や, 既存の解析手法を拡張した独自の推定量を実装する上でのIMLプロシジャの有用性について共有する.

2. 推定方程式の概要

生物統計で用いられる多くの推定量は, 未知パラメータに関する方程式の解として表現することができる. このような方程式は一般に推定方程式と呼ばれ, パラメータベクトルを θ とすると,

$$\sum_{i=1}^{n} \psi(O_i, \theta) = 0$$

の形で表すことができる(Stefanski and Boos, 2002 や Ross et al., 2024). ここで, $i = 1,..,n$ は観測単位, O_i は第 i 番目の観測データ, $\psi(\cdot)$ は推定関数と呼ばれ, 観測データとパラメータから構成される関数である. なお, 本稿ではベクトルとスカラーを区別せずに表記するため, 適宜読み替えていただけると幸いである. このとき推定量 $\hat{\theta}$ は, この推定方程式を満たす解として得られる. 例えば, 線形回帰における最小二乗推定量は, $X^T(Y - X\beta) = 0$ という正規方程式の解として与えられる. ここで, 上付きの T は転置を表す. また, 一般化線形モデルにおける最尤推定量やIPW 推定量などもそれぞれ対応する推定方程式の解として表現することができる.

推定方程式に基づく推定量の重要な性質として, 真のパラメータ θ_0 のもとで

$$E_{\theta_0}[\psi(O, \theta_0)] = 0$$

となるならば, いくつかの正則条件のもとで, 推定方程式の解 $\hat{\theta}$ は一致性と漸近正規性をもつことがあげられる. 最尤推定量や IPW 推定量などの一致性は, それぞれの手法に必要な仮定や識別条件, 正則条件のもとで, 上記の結果から導くことができる.

このように, 推定方程式は推定量の定義だけでなく, その理論的性質を統一的に記述する枠組みを提供する. さらに, 異なる統計手法の違いは主として $\psi(\cdot)$ の定義に反映されるため, 推定方程式を柔軟に記述し, その解を数値的に求めることができれば, 多様な推定量を統一的な枠組みで実装することが可能となる.

3. 既存のプロシジャの再現

本節では，推定方程式に基づく実装の例として，修正ポアソン回帰および IPW 推定量を取り上げる．いずれも SAS では GENMOD プロシジャおよび CAUSALTRT プロシジャにより実装可能な手法であるが，ここでは推定方程式を IML プロシジャ上で直接記述し，その解を数値的に求めることで同一の推定結果が得られることを確認する．

3.1 修正ポアソン回帰

修正ポアソン回帰は，二値アウトカムに対してポアソン回帰モデルを適用し，サンドイッチ分散により分散を推定することで，対数リスク比とその標準誤差を推定する方法である(Zou, 2004)．観測データを(Y_i, X_i)とし，$E[Y_i|X_i] = \exp(X_i^T\beta)$ を仮定する．ここで，Y_i はアウトカム，X_i は各個人の共変量ベクトルである．二値アウトカム Y_i に対してポアソン分布を仮定した場合，β に依存しない定数項を除く対数尤度は，

$$l(\beta) = \sum_{i=1}^{n}\{Y_i X_i^T\beta - \exp(X_i^T\beta)\}$$

となる．この対数尤度を β について微分すると，

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} X_i(Y_i - \exp(X_i^T\beta))$$

となるため，ポアソン尤度に基づく推定方程式は以下の式で与えられる．

$$\sum_{i=1}^{n} X_i(Y_i - \exp(X_i^T\beta)) = 0$$

IML プロシジャでは，パラメータ β に対して上式を定義し，その解を数値的に求めることで回帰係数を推定することができる．

また，アウトカムが二値変数であるため，ポアソン分布が仮定する条件付き分散は真の分散構造とは一致しない．そのため，分散推定にはサンドイッチ分散を用いる．一般に推定方程式 $\sum_{i=1}^{n}\psi_i(\hat{\theta}) = 0$ に対して，推定量の漸近分散は次式で推定することができる．

$$Var(\hat{\theta}) = \frac{1}{n}\widehat{A^{-1}}\hat{B}\widehat{A^{-T}}$$

ここでA, B はそれぞれ以下である．

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} -\frac{\partial\psi_i(\hat{\theta})}{\partial\theta^T}, \qquad \hat{B} = \frac{1}{n}\sum_{i=1}^{n}\psi_i(\hat{\theta})\psi_i(\hat{\theta})^T$$

以下のコードは，SAS help の bweight データセットを用いて，乳児の体重に対する母親の喫煙(MomSmoke)の対数リスク比を GENMOD プロシジャおよび IML プロシジャで推定するコードである．ここで，便宜的に乳児の体重を 3000 未満と 3000 以上に二値化し，母親の年齢(MomAge)，婚姻状況(Married)を共変量としてモデルに含めた．

```
data dummy;  /*データの作成*/
 set sashelp.BWEIGHT;
 outcome = Weight < 3000; /*Weight 3000未満/以上で二値アウトカムを定義*/
```

```
 id = _N_;
run;

proc genmod data=dummy descending; /*GENMODプロシジャによる対数リスク比の推定*/
 class MomSmoke(ref='0') Married(ref='0') id;
 model outcome = MomSmoke MomAge Married      / dist=poisson link=log;
 repeated subject=id / type=ind;
run;

proc iml; /*IMLプロシジャによる対数リスク比の推定*/
 use dummy;
  read all var {outcome MomSmoke MomAge Married} into data;
 close dummy;
 n = nrow(data); /*レコード数の保持*/
 p = 4; /*推定するパラメータ：切片項, MomSmoke, MomAge, Married*/
  /*推定方程式の定義*/
 start ef(par) global(data,p,n);
 f1 = j(1,1,0); f2 = j(1,1,0); f3 = j(1,1,0); f4 = j(1,1,0); f = j(p,1,0);
 /*推定関数の総和の計算*/
 do i = 1 to n;
  f1 = f1 + (data[i,1] - exp(par[1] + par[2]*data[i,2] + par[3] * data[i,3] + par[4] * data[i,4])); /*切片項*/
  f2 = f2 + data[i,2]*(data[i,1] - exp(par[1]    + par[2]*data[i,2] + par[3]* data[i,3] + par[4]*data[i,4])); /*処置変数
MomSmoke(基準:0)*/
  f3 = f3 + data[i,3]*(data[i,1] - exp(par[1]    + par[2]*data[i,2] + par[3]*data[i,3] + par[4]*data[i,4])); /*MomAge*/
  f4 = f4 + data[i,4]*(data[i,1] - exp(par[1]    + par[2]*data[i,2] + par[3]*data[i,3] + par[4]*data[i,4])); /*Married(基
準:0)*/
 end;
 f[1] = f1; f[2] = f2; f[3] = f3; f[4] = f4;
 return (f);
 finish ef;
 x0 = {0.5 0.5 0.5 0.5}; /*初期値の設定*/
 optn = p || 1;
 call nlplm(rc,xr,"ef",x0,optn);
 /*サンドイッチ分散の計算*/
 par = j(1,3,.);
 par[1] = p;
 /*Aの部分*/
 call nlpfdd(func, bread, hess, "ef", xr, par);
```

```
  bread = -bread / n;
  /*Bの部分*/
  meat = j(p,p,0);
  e1 = j(1,1,0); e2 = j(1,1,0); e3 = j(1,1,0); e4 = j(1,1,0); e = j(p,1,0);
  do i = 1 to n;
    e1 = (data[i,1] - exp(xr[1]    + xr[2]*data[i,2] + xr[3]*data[i,3] + xr[4]*data[i,4]));
    e2 = data[i,2]*(data[i,1] - exp(xr[1] + xr[2]*data[i,2] + xr[3]*data[i,3] + xr[4]*data[i,4]));
    e3 = data[i,3]*(data[i,1] - exp(xr[1] + xr[2]*data[i,2] + xr[3]*data[i,3] + xr[4]*data[i,4]));
    e4 = data[i,4]*(data[i,1] - exp(xr[1] + xr[2]*data[i,2] + xr[3]*data[i,3] + xr[4]*data[i,4]));
    e[1] = e1; e[2] = e2; e[3] = e3; e[4] = e4; meat = meat + (e*e`);
  end;
  meat = meat / n;
  /*サンドイッチ分散*/
  sandwich = (inv(bread) * meat * inv(bread)`) / n;
  stderr = sqrt(vecdiag(sandwich));
  bhat = xr`; lower = bhat - 1.96 # stderr; upper = bhat + 1.96 # stderr;
  print n,, bhat[format=8.4] stderr[format=8.4] lower[format=8.4] upper[format=8.4];
quit;
```

上記コードの start ~ finish の部分が推定方程式を定義している部分となり，定義した推定方程式に対して，call nlplm により Levenberg-Marquardt 法を用いて目的関数を最小化している．また，call nlpfdd により定義した推定関数の勾配を求めることで $\hat{A}$ を求めている．堀本(2025)では call nlpnra により Newton–Raphson 法を用いて目的関数を最小化しているが，引数などはほとんど同じであるため，IML プロシジャの詳細は堀本(2025)を参照されたい．

上記コードを実行した結果が表 1 となり，パラメータ推定値，サンドイッチ標準誤差，および Wald 型の 95%信頼区間が GENMOD プロシジャによる結果と一致することが確認できる．

表 1：GENMOD プロシジャおよび IML プロシジャによる対数リスク比の推定結果

プロシジャ	対数リスク比の推定値	サンドイッチ標準誤差	95%信頼区間
GENMOD	0.4915	0.0207	0.4510, 0.5320
IML	0.4915	0.0207	0.4510, 0.5320

3.2 逆確率重み付け法

次に，Ross et al. (2024)を参考に，因果推論で広く用いられる IPW 推定量を IML プロシジャを用いて実装する．Y^1, Y^0 をそれぞれ処置を受けた場合，受けなかった場合の二値の潜在アウトカムとし，推定したい効果を以下の平均処置効果(ATE)とする．

$$ATE = E[Y^1] - E[Y^0] = \mu_1 - \mu_0 = \mu_{dif}$$

ここで，$E[Y^1] = \mu_1$, $E[Y^0] = \mu_0$ とした．本節では詳細な導出は省略し，$E[Y^1] = \mu_1$および $E[Y^0] = \mu_0$を以

下の IPW 推定量により推定し，その差として ATE を推定することを考える．

$$\widehat{\mu_1} = \frac{1}{n}\sum_{i=1}^{n}\frac{A_i Y_i}{\hat{e}(X_i)}, \qquad \widehat{\mu_0} = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-A_i)Y_i}{1-\hat{e}(X_i)}$$

$$\widehat{\mu_{dif}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{A_i Y_i}{\hat{e}(X_i)} - \frac{(1-A_i)Y_i}{1-\hat{e}(X_i)}\right)$$

ここで，Y_i は発生していれば 1，発生していなければ 0 をとる実際に観測されるアウトカム，X_i は各個人の共変量ベクトル，A_i は処置ありならば 1，処置なしならば 0 をとる二値の処置変数とし，$\hat{e}(X_i)$ は以下のロジスティックモデルから推定される傾向スコアとする．

$$e_i = P(A_i = 1|X_i) = \frac{1}{1+\exp(-X_i^T\beta)} = expit(X_i^T\beta)$$

このロジスティック回帰の推定方程式は以下で与えられる．

$$\sum_{i=1}^{n} X_i(A_i - expit(X_i^T\beta)) = 0$$

これらの推定方程式をそれぞれ積み上げて 1 つの方程式とみなすと以下となる．

$$\sum_{i=1}^{n}\psi(\theta) = \sum_{i=1}^{n}\begin{pmatrix} X_i\left(A_i - expit(X_i^T\beta)\right) \\ \frac{A_i Y_i}{\hat{e}(X_i)} - \mu_1 \\ \frac{(1-A_i)Y_i}{1-\hat{e}(X_i)} - \mu_0 \\ (\mu_1 - \mu_0) - \mu_{dif} \end{pmatrix} = 0$$

この推定方程式に対して，前節と同様に以下のサンドイッチ分散推定量を用いることで，ATE の推定誤差に加え，傾向スコア推定の不確実性も考慮した標準誤差を算出することができる．

$$Var(\hat{\theta}) = \frac{1}{n}\widehat{A^{-1}}\hat{B}\widehat{A^{-T}}$$

以下のコードは修正ポアソン回帰と同様のデータを用いて，IML プロシジャにより ATE およびサンドイッチ標準誤差を推定するプログラムである．ただし，推定方程式の定義以外は 3.1 節のコードと基本的に同様であるため，紙面の都合上省略した．

```
proc causaltrt data=dummy method = ipw; /*CAUSALTRTプロシジャによるATEの推定*/
  class MomSmoke(ref='0') Married(ref='0') /desc;
  psmodel MomSmoke = MomAge Married ;
  model outcome;
run;


start ef(par) global(data,p,n); /*IMLプロシジャの推定方程式の定義部分*/
  f1 = j(1,1,0); f2 = j(1,1,0); f3 = j(1,1,0); f4 = j(1,1,0); f5 = j(1,1,0);
  f6 = j(1,1,0); f = j(p,1,0);
  do i = 1 to n;
  /*傾向スコアの推定*/
```

```
  pi = 1 / (1 + exp(-(par[1]    + par[2]*data[i,3] + par[3] * data[i,4] )));
  f1 =    f1 + (data[i,2] - pi); /*切片項の推定方程式*/
  f2 = f2 + data[i,3]*(data[i,2] - pi); /*MomAgeの推定方程式*/
  f3 = f3 + data[i,4]*(data[i,2] - pi); /*Marriedの推定方程式*/
  /*重み付き平均*/
  f4 = f4 + ((data[i,2]*data[i,1]) / pi) - par[4]; /*  処置=1の推定方程式  */
  f5 = f5 + (((1-data[i,2])*data[i,1]) / (1-pi)) - par[5];/*  処置=0の推定方程式  */    f6 = f6 + ((par[4]    - par[5] ) -
  par[6]); /*リスク差の推定方程式  */
end;
f[1] = f1; f[2] = f2; f[3] = f3; f[4] = f4; f[5] = f5; f[6] = f6;
return (f);
finish ef;
```

上記コードを実行した結果が表 2 となり，パラメータ推定値，サンドイッチ標準誤差，および Wald 型の 95% 信頼区間が CAUSALTRT プロシジャによる結果と一致することが確認できる．

表 2：CAUSALTRT プロシジャおよび IML プロシジャによる ATE の推定結果

プロシジャ	ATE の推定値	サンドイッチ標準誤差	95%信頼区間
CAUSALTRT	0.1399	0.00652	0.1271, 0.1527
IML	0.1399	0.00652	0.1271, 0.1527

4. G-computation

本節では IML プロシジャを用いて時間依存性交絡に対する G-computation を実装する．G-computation を直接実行する専用のプロシジャはないものの，ロジスティック回帰や線形回帰などの既存のプロシジャを組み合わせることで実装することが可能である．本節では，各ステップに対応する推定方程式を IML プロシジャ上で直接記述し，それらを逐次的に解くことで G-computation の推定手順を明示的に示す．

時間依存性交絡とは，治療開始後の時間経過に伴って変化する共変量が，その後の治療選択とアウトカムの双方に影響を与える一方で，過去の治療の影響も受ける状況を指す．すなわち，ある時点における治療曝露が，その後の交絡因子の値に影響を与え，さらにその交絡因子が次時点以降の治療選択およびアウトカムに影響を及ぼす場合に生じる．時間依存性交絡では，時間依存共変量は交絡因子であると同時に「治療の影響を受ける変数」となる．このような変数を通常の回帰モデルなどで調整すると，過去の治療による影響を受けた変数を調整することになるため，過剰調整により因果効果の推定にバイアスが生じる可能性がある．一方で，調整しない場合には交絡が残存するため，従来の回帰分析では適切な因果効果の推定が困難となる．複数時点で治療が行われる状況ではいくつかの推定対象を定義することができるが，本稿では主に，全ての対象者がどの時点でも治療を受けなかった場合の潜在アウトカムのリスクを推定することを考える．なお以降の議論では，因果効果の識別に必要な因果一致性，逐次的な条件付き交換可能性，逐次的な条件付き正値性の仮定は成り立っているものとする．

以降, Shinozaki and Suzuki (2020)の表 2 から一部抜粋した仮想的なコホートデータ(表 3)を用いて具体的な説明を行う. このコホートは時点 1 の治療 A_1, 時点 2 の共変量 L_2, 時点 2 の治療 A_2 の組み合わせによって 8 つの層に分類されている. 各層について, 対象者数 N , 治療レジメン (A_1, A_2) を全員に割り付けた場合の潜在アウトカム $(Y^{0,0}=1, Y^{1,1}=1)$, 実際に観察されたアウトカム $(Y=1)$ が示されている. 例えば, $A_1=1$, $L_2=1$, $A_2=1$ の層には 720 例が存在し, それらの症例が時点 1 および時点 2 でともに治療を受けない場合 $(A_1=A_2=0)$ は 648 例でイベントが発生するが $(Y^{0,0}=1$の列), この層では実際には $A_1=1$, $L_2=1$, $A_2=1$ が観察されており, 観察されるデータでのイベント発生例($Y=1$ の列) は 576 例となる.

表 3：Shinozaki and Suzuki (2020)の表 2 から一部抜粋した仮想的なコホートのデータ

Stratum			N	Potential outcome				Observed outcome
A_1	L_2	A_2		$Y^{0,0}=1$	Risk	$Y^{1,1}=1$	Risk	$Y=1$
1	1	1	720	648	0.9	576	0.8	576
1	1	0	180	162	0.9	144	0.8	108
1	0	1	1,800	1,080	0.6	720	0.4	720
1	0	0	1,800	1,080	0.6	720	0.4	900
0	1	1	5,670	5,103	0.9	3,402	0.6	4,536
0	1	0	630	567	0.9	378	0.6	567
0	0	1	840	252	0.3	252	0.3	294
0	0	0	3,360	1,008	0.3	1,008	0.3	1,008
Total			15,000	9,900	0.66	7,200	0.48	

表 3 から, 全ての対象者が治療を受けなかった場合のリスクは 0.66 となる. ここで, Shinozaki and Suzuki (2020) に記載されている通り, 時間依存性交絡を考慮せず, 時点 2 の共変量 L_2 を通常の交絡因子として扱う標準化を適用すると,

$$\sum_{l_2} E[Y|A_1=0, A_2=0, L_2=l_2]P(L_2=l_2)$$

より 0.59 と計算され, この値は表 3 における潜在アウトカム $(Y^{0,0}=1)$ のリスク 0.66 と一致しない. これは L_2 が時点 1 の治療の影響を受ける一方で, 時点 2 の治療とアウトカムに影響を与えるためである.

上記のような時間依存性交絡が生じている状況において, 本稿ではパラメトリックモデルを用いて潜在アウトカムの期待値を推定する. 具体的には, Iterative Conditional Expectation (ICE) 法に基づく G-computation を用いる(篠崎ら, 2026). なお, 表 3 のデータではパラメトリックモデルを使用する必要はなく, Shinozaki and Suzuki (2020)のようにノンパラメトリックに算出することが可能であるが, 本稿ではより一般的な状況での適用も考え, パラメトリックモデルにより実装する.

ICE 法は, G-computation を逐次的な条件付き期待値の計算によって推定する手法であり, 共変量過程をシミュレーションするモンテカルロ法に基づく G-computation と比較して, 時間依存共変量の条件付き分布をモデル化する必要がないという利点がある. 表 3 の状況の場合, 識別仮定のもとで逐次的な条件付き期待値は

$$E[Y^{a_1,a_2}] = E[E[E[Y|L_2, A_1 = a_1, A_2 = a_2]|A_1 = a_1]]$$

と表すことができる．この条件付き期待値の部分をそれぞれパラメトリックモデルを用いて計算していく．ここで，本稿では条件付き期待値のモデル化にロジスティック回帰と線形回帰を用いる．なお，本稿のデータにはベースライン共変量が存在しないため，最外層におけるベースライン共変量の分布に関する標準化は，単純な周辺平均の計算となる．ただし，より一般的な状況を想定し，推定手順には最終的な平均化のステップを含める．

具体的な手順は以下の通りである．

ステップ 1：

観測データを用いて，Y を目的変数，A_1, L_2, A_2 とそれら全ての交互作用項を説明変数とするロジスティック回帰モデル(飽和モデル)を当てはめパラメータを推定する．

ステップ 2：

$A_1 = 0$ および $A_2 = 0$ と固定し，ステップ 1 で推定されたパラメータと各対象者の観測された L_2 の値に基づき，治療レジメン $A_1 = A_2 = 0$ の下でのイベント発生確率を算出する．

ステップ 3：

得られたイベント発生確率を目的変数とし，A_1を説明変数とする線形回帰モデルを当てはめパラメータを推定する．

ステップ 4：

ステップ 3 で推定されたパラメータを用いて $A_1 = 0$ としたときの予測値を算出し，これらの値の平均を全ての対象者がどの時点でも治療を受けなかった場合のリスク推定値とする．

上記手順に必要な推定方程式は以下で表すことができる．

$$\sum_{i=1}^{n} \psi(\theta) = \sum_{i=1}^{n} \begin{pmatrix} X_i\left(Y_i - expit(X_i^T\beta)\right) \\ Z_i\left(expit\left(X_i^{0,0,T}\beta\right) - Z_i^T\gamma\right) \\ Z_i^{0,T}\gamma - \mu \end{pmatrix} = 0$$

1 つ目の式はステップ 1 に対応する，全ての交互作用項を含むロジスティック回帰に対する推定方程式である．なお，本稿では表 3 における潜在アウトカムのリスクを再現するために飽和モデルを用いて推定を行っている．2 つ目の式はステップ 3 に対応する線形回帰の推定方程式であり，$X_i^{0,0,T}$ は $A_1 = 0$ および $A_2 = 0$ としたデザイン行列を表す．ステップ 2 で算出した，時点 1 および時点 2 ともに治療を受けなかった場合のイベント発生確率の予測値を目的変数とし，これを時点 1 の治療A_1 で回帰する．3 つ目の式はステップ 4 に対応しており，$Z_i^{0,T}$ は $A_1 = 0$ としたデザイン行列を表し，全ての対象者が時点 1 および時点 2 で治療を受けなかった場合のイベント発生リスクを表すパラメータ μ に対する推定方程式である．

以下のコードは IML プロシジャにより $Y^{0,0} = 1$ のリスクを推定するプログラムである．ただし，推定方程式の定義以外の部分は 3 節のコードとほとんど同じであるため紙面の都合上部分的に割愛した．また，本節の手順はロジスティック回帰と線形回帰の組み合わせであるため，既存のプロシジャでも実装可能である．

```
/*Shinozaki and Suzuki(2020)からデータ作成部分を引用*/
data dummy;
 input A1 L2 A2 N N1;
 cumA = A1 + A2;
```

```
 do i = 1 to N1; Y = 1; ID + 1; output; end;
 do i = N1 + 1 to N; Y = 0; ID + 1; output; end;
 cards;
 1 1 1 720 576
 1 1 0 180 108
 1 0 1 1800 720
 1 0 0 1800 900
 0 1 1 5670 4536
 0 1 0 630 567
 0 0 1 840 294
 0 0 0 3360 1008
;
run;

proc iml; /*時点1, 時点2でともに治療を受けなかった場合のリスクの推定*/
 use dummy; /*dummyデータから使用する変数の抽出*/
  read all var {Y A1 L2 A2} into data; /*記載順にdata[i,k]のk列目にデータを格納*/
 close dummy;
 n = nrow(data); /*レコード数の保持*/
 p = 11; /*パラメータ数の指定*/

 /*推定方程式の定義*/
 start ef(par) global(data,p,n);
 f1 = j(1,1,0); ～ f11 = j(1,1,0); f = j(p,1,0); /*f1と同様にf11まで作成する*/
 /*推定関数の総和の計算*/
 do i = 1 to n;
 /*ロジスティック回帰モデルの定義*/
pi = 1 / (1 + exp(-(par[1] + par[2]*data[i,2] + par[3]*data[i,3] + par[4]*data[i,4] + par[5]*data[i,2]*data[i,3] +
par[6]*data[i,2]*data[i,4] + par[7]*data[i,3]*data[i,4] + par[8]*data[i,2]*data[i,3]*data[i,4])));
 /*線形回帰の定義*/
   linear = par[9] + par[10] * data[i,2];
   /*時点1, 2の治療を0に固定した場合のロジスティック回帰の定義*/
   pi_0_0 = 1 / (1 + exp(-(par[1] + par[2]*0 + par[3]*data[i,3] + par[4]*0 + par[5]*0*data[i,3] + par[6]*0*0 +
par[7]*data[i,3]*0 + par[8]*0*data[i,3]*0)));
   /*時点1の治療を0に固定した場合の線形回帰の定義*/
   pi_0 = par[9] + par[10] * 0;

   f1 = f1 + (data[i,1] - pi); /*f1~f8はロジスティック回帰の各パラメータ*/
```

```
    f2 = f2 + data[i,2]*(data[i,1] - pi);
    f3 = f3 + data[i,3]*(data[i,1] - pi);
    f4 = f4 + data[i,4]*(data[i,1] - pi);
    f5 = f5 + data[i,2]*data[i,3]*(data[i,1] - pi);
    f6 = f6 + data[i,2]*data[i,4]*(data[i,1] - pi);
    f7 = f7 + data[i,3]*data[i,4]*(data[i,1] - pi);
    f8 = f8 + data[i,2]*data[i,3]*data[i,4]*(data[i,1] - pi);
    f9 = f9 + (pi_0_0 - linear); /*f9,f10は線形回帰のパラメータ*/
    f10 = f10 + data[i,2]*(pi_0_0 - linear);
    f11 = f11 + (pi_0 - par[11]); /*リスクのパラメータ  */
  end;
  f[1] = f1; ~~ f[11] = f11; /*f[1]と同様にf[11]まで作成する*/
  return (f);
  finish ef;
  x0 = {0 0 0 0 0 0 0 0 0 0 0}; /*初期値の設定*/
/*以降 3.1 節のコードと同様のため割愛*/
quit;
```

上記コードによる推定結果を表 4 に示す．推定されたリスクは，表 3 における潜在アウトカムのリスクと一致しており，G-computation により表 3 のリスクが再現されることが確認できる．また，ステップ 2 およびステップ 4 で$A_1 = 1$ および $A_2 = 1$ とすることで，$Y^{1,1} = 1$ のリスクも推定することが可能であり，こちらも表 3 のリスクが再現されていることが確認できる．

表 4：G-computation によるリスクの推定値

推定対象	推定値
$Y^{0,0} = 1$ のリスク	0.66
$Y^{1,1} = 1$ のリスク	0.48

5. アウトカムに誤分類がある場合の G-computation

最後に，本節では ICE 法に基づく，二値アウトカムの誤分類を考慮するための G-computation を提案する．Edwards et al. (2018) は，死亡原因の感度および特異度を仮定し，動的治療レジメン下の原因別累積死亡リスクに対する誤分類の影響を評価する感度分析を，共変量をシミュレーションする G-computation に組み込んでいる．これに対して本稿では，静的治療レジメンの下で，追跡終了時点における二値アウトカムを対象とし，誤分類を考慮したアウトカムモデルから得られる条件付き期待値を ICE 法により逐次的に回帰することで推定する．

誤分類とは，本来の状態とは異なるカテゴリーに対象者を分類してしまうことを指し，疫学研究における代表的な情報バイアスの一つである．誤分類は，曝露，アウトカム，交絡因子などの測定において生じ，推定

される関連の大きさや方向に影響を及ぼす可能性がある. 例えば, 不完全なアウトカム定義や, レセプトデータにおいて疑い病名が登録されることなどによって生じる. 二値変数の場合, 誤分類の程度は一般に感度および特異度によって表される. 感度は真にイベントが発生している対象者を正しくイベントありと分類する確率, 特異度は真にイベントが発生していない対象者を正しくイベントなしと分類する確率を意味する.

以降, 4 節で用いたデータに対して, (現実的かはさておき)人工的に誤分類を発生させたデータを用いて議論を進める. 表 5 は表 3 のデータに対して, 治療レジメンや共変量の値によらず, 感度および特異度をそれぞれ 0.7 と仮定し, 真のアウトカムに対して誤分類を生成したデータである. 表 3 との違いは Misclassified outcome の列であり, 真のアウトカムが1 である対象の30% , および真のアウトカムが0である対象の30% がそれぞれ誤分類されるため, 実際に解析に用いられる観測データは Correct outcome 列ではなく, Misclassified outcome 列となる. なお, 誤分類の有無を明示的に示すため, $Y = 0$ および $Y^* = 0$ も追記している. 例えば, $A_1 = 1$, $L_2 = 1$, $A_2 = 1$ の層における $Y^* = 1$ の値は $576 \times 0.7 + 144 * 0.3 = 446.4$ となり, 小数点を四捨五入することで算出した.

表 5：表 3 のデータから感度および特異度ともに 0.7 で発生させた仮想的なコホートのデータ

Stratum			N	Potential outcome				Correct outcome		Misclassified outcome	
A_1	L_2	A_2		$Y^{0,0} = 1$	Risk	$Y^{1,1} = 1$	Risk	$Y = 1$	Y =0	$Y^* = 1$	$Y^* = 0$
1	1	1	720	576	0.9	576	0.8	576	144	446	274
1	1	0	180	144	0.9	144	0.8	108	72	98	82
1	0	1	1,800	720	0.6	720	0.4	720	1,080	828	972
1	0	0	1,800	720	0.6	720	0.4	900	900	900	900
0	1	1	5,670	3,402	0.9	3,402	0.6	4,536	1,134	3,515	2,155
0	1	0	630	378	0.9	378	0.6	567	63	416	214
0	0	1	840	252	0.3	252	0.3	294	546	370	470
0	0	0	3,360	1,008	0.3	1,008	0.3	1,008	2,352	1,412	1948
Total			15,000	9,900	0.66	7,200	0.48	8,709	6,291	7,985	7,015

ここで, 誤分類の有無に注目するため, 縦に真のアウトカムの有無, 横に誤分類が生じているアウトカムの有無としたクロス表を作成すると(表 6), $Y = 1$ の対象者 8709 例のうち $Y^* = 1$ の症例数は 6097 例, 同様に $Y = 0$ の対象者 6291 例のうち $Y^* = 0$ の症例数は 4403 例となり感度と特異度がどちらも約 70%となっていることがわかる.

表 6：真のアウトカムと誤分類のあるアウトカムのクロス表

	$Y^* = 1$	$Y^* = 0$	合計
$Y = 1$	6,097	2,612	8,709
$Y = 0$	1,888	4,403	6,291
合計	7,985	7,015	15,000

Neuhaus(1999)はアウトカムに誤分類がある場合に，感度および特異度を用いて誤分類を尤度に組み込み，真のアウトカムに対するロジスティック回帰モデルのパラメータを推定する方法を提案している．いま，誤分類の生じていないアウトカムに対するモデルを

$$p_i = P(Y_i = 1|X_i) = \frac{1}{1+\exp(-X_i^T\beta)} = expit(X_i^T\beta) \quad \cdots (1)$$

とする．感度および特異度が処置変数や共変量に依存しないと仮定すると，それぞれ

$$sen = P(Y_i^* = 1|Y_i = 1), \qquad spe = P(Y_i^* = 0|Y_i = 0)$$

となり，誤分類があるアウトカムを用いた場合の尤度は以下のようになる．

$$L = \prod_{i=1}^{n} P(Y_i^* = y_i^*|X_i = x_i) = \prod_{i=1}^{n}\left(\sum_{y_i=0}^{1} P(Y_i^* = y_i^*|Y_i = y) \times P(Y_i = y_i|X_i = x_i)\right) \quad \cdots (2)$$

ここで，(2)式の最後の総和の部分に注目すると，

$$\sum_{y_i=0}^{1} P(Y_i^* = y_i^*|Y_i = y) \times P(Y_i = y_i|X_i = x_i)$$

$$= [P(Y_i^* = 1|Y_i = 1) \times \frac{1}{1+\exp(-X_i^T\beta)} + P(Y_i^* = 1|Y_i = 0) \times \left(1 - \frac{1}{1+\exp(-X_i^T\beta)}\right)]^{y_i^*}$$

$$\times [P(Y_i^* = 0|Y_i = 1) \times \frac{1}{1+\exp(-X_i^T\beta)} + P(Y_i^* = 0|Y_i = 0) \times \left(1 - \frac{1}{1+\exp(-X_i^T\beta)}\right)]^{1-y_i^*}$$

となる．$P(Y_i^* = y_i^*|Y_i = y)$ の部分はそれぞれ感度や特異度と 1 からそれらの値をひいたものであり，これらを既知とすると，未知パラメータは(1)式で定義した真のアウトカムに対するロジスティック回帰のパラメータ β のみとなる．したがって，真のアウトカム Y の値が不明であっても誤分類の可能性があるアウトカム Y^* と感度および特異度を用いて(2)式の尤度を最大化することで，モデルや感度，特異度が正しいなどの仮定のもとでは，真のアウトカムに対するロジスティック回帰のパラメータ β の一致推定量を得ることができる．(2)式の対数尤度を微分して 0 とおいた推定方程式は以下となる．

$$\sum_{i=1}^{n}\left[\frac{y_i^*(sen+spe-1)}{sen \times p_i + (1-spe) \times (1-p_i)} - \frac{(1-y_i^*)(sen+spe-1)}{(1-sen)p_i + spe(1-p_i)}\right] \times p_i(1-p_i)X_i = 0$$

この手法を 4 節の G-computation のステップ 1 に適用することで，時間依存性交絡が存在する状況における誤分類考慮したリスクを推定してみよう．すなわち，ステップ 1 を以下のように修正する

修正したステップ 1：

真のアウトカム Y に対して A_1, L_2, A_2 とそれら全ての交互作用項を説明変数とするロジスティック回帰モデル(飽和モデル)を仮定する．観測された誤分類があるアウトカム Y^*と，既知の感度および特異度を用いて(2)式の誤分類を考慮した尤度を最大化することにより，真のアウトカムモデルのパラメータ β を推定する．

このステップで真のアウトカムに対するロジスティック回帰のパラメータが推定できるため，以降のステップは 4 節と同様になり，全てのステップを積み上げた推定方程式は以下となる．4 節との違いは 1 つ目の式のみである．

$$\psi(\theta) = \sum_{i=1}^{n}\begin{pmatrix}\left[\frac{y_i^*(sen+spe-1)}{sen \times p_i + (1-spe) \times (1-p_i)} - \frac{(1-y_i^*)(sen+spe-1)}{(1-sen)p_i + spe(1-p_i)}\right] p_i(1-p_i)X_i \\ Z_i\left(expit\left(X_i^{0,0,T}\beta\right) - Z_i^T\gamma\right) \\ Z_i^{0,T}\gamma - \mu\end{pmatrix} = 0$$

証明は省略するが，この推定方程式は，感度や特異度が正しい，各ステップでの回帰モデルが正しいなどの仮定や，解が一意に存在するなどの正則条件の下では，真のパラメータにおいて期待値が 0 となる推定方程式となるため，2 節の議論から推定方程式を解くことで得られる推定量 $\hat{\theta}$ は一致性と漸近正規性をもつ．また 3 節と同様にサンドイッチ分散を計算できる．

実装のためのコードは 4 節とほとんど同じであるため，誤分類の生成と推定方程式の定義部分のみのコードを以下に示す．

```
/*表3のデータから誤分類の生成*/
proc sort data = dummy;
 by A1 L2 A2 Y;
run;

data dummy_mis;
 set dummy; by A1 L2 A2 Y; retain count; if first.Y then count = 0; count + 1;
 if Y = 1 then do; /* Y=1：30%を0へ誤分類 ＊簡易化のため乱数は使用しない*/
  if count <= round(N1 * 0.30) then Y_mis = 0;
  else Y_mis = 1;
 end;
 else if Y = 0 then do; /* Y=0：30%を1へ誤分類 ＊簡易化のため乱数は使用しない*/
  if count <= round((N - N1) * 0.30) then Y_mis = 1;
  else Y_mis = 0;
 end;
run;

/*誤分類を考慮して時点1，時点2でともに治療を受けなかった場合のリスクの推定*/
proc iml;
 use dummy_mis; /*dummy_misデータから使用する変数の抽出*/
  read all var {Y_mis A1 L2 A2} into data;/*記載順にdata[i,k]のk列目にデータを格納*/
 close dummy_mis;
 n = nrow(data); /*レコード数の保持*/
 p = 11; /*パラメータ数の指定*/
 /*推定関数の総和の計算*/
 do i = 1 to n;
 pi = ~ ; /*pi, linear, pi_0_0, pi_0は4節のコードと同様*/
 linear = ~ ;
 pi_0_0 = ~ ;
 pi_0 = ~ ;
 A = (1-0.7)*(1-pi) + 0.7*pi; /*推定方程式の[]内の第1項の分母*/
```

```
  B = 0.7*(1-pi) + (1-0.7)*pi; /*推定方程式の[]内の第2項の分母*/

   f1 = f1 + ((((data[i,1]*(0.7 + 0.7 - 1)) / A) - (((1-data[i,1])*(0.7 + 0.7 -1))/B))*pi*(1-pi));/*ステップ1のロジスティック回帰の切片項*/
   f2 = f2 + data[i,2]*((((data[i,1]*(0.7 + 0.7 - 1)) / A) - (((1- data[i,1]) *(0.7 + 0.7 - 1))/B))*pi*(1-pi)) /*ステップ1のロジスティック回帰のA1の項*/

   /*f3~f8までのそれぞれf1, f2と同様に4節のコードを修正しロジスティック回帰の各項を記載*/

   f9 = ~ ; /*f9~f11は4節のコードと同様*/
  end;
/*以降の初期値の設定などやサンドイッチ分散の計算も3節および4節のコードとほとんど同様*/
quit;
```

表 7 の”真のアウトカムを用いた場合の推定値”は 4 節のコードに追加でサンドイッチ分散を算出し, 95%信頼区間を計算した結果である. この信頼区間は Shinozaki and Suzuki (2020) による周辺構造モデルの結果とほとんど同じ値となっている. 表 7 の“誤分類を考慮しない場合の推定値” は 4 節の手法とコードを誤分類のあるアウトカムに対して適用した結果であり, “誤分類を考慮した場合の推定値”が誤分類のあるアウトカムに対して上記の修正した手順とコードを適用した結果である. なお, それぞれの結果は小数点第 3 桁で四捨五入を行った. 表 7 から, 誤分類を考慮しない場合はリスクをうまく推定できていないが, 提案法を用いて誤分類を考慮することで, 表 3 の潜在リスク 0.66 と0.48 が正しく推定できていることがわかる. 一方で, 真のアウトカムに対して 4 節の G-computation で推定した結果と比べ, 信頼区間が広くなっていることがわかる. 本稿では, 比較的シンプルな状況の仮想的なコホートデータに対してのみ検討を行ったが, 誤分類が処置変数や共変量に依存する状況などより複雑な状況において, シミュレーションによりバイアスや推定精度を評価することを今後の課題としたい.

表 7：誤分類を考慮しない場合と考慮した場合のリスクの推定値と 95%信頼区間

推定対象	真のアウトカムを用いた場合の推定値(95%信頼区間)	誤分類を考慮しない場合の推定値(95%信頼区間)	誤分類を考慮した場合の推定値(95%信頼区間)
$Y^{0,0}$ のリスク	0.66 (0.64, 0.68)	0.56 (0.54, 0.59)	0.66 (0.60, 0.72)
$Y^{1,1}$ のリスク	0.48 (0.46, 0.50)	0.46 (0.43, 0.49)	0.48 (0.43, 0.53)

6. まとめ

本稿では, 修正ポアソン回帰, IPW, G-computation といった既存の手法に加え, ICE 法に基づく誤分類を考慮する G-computation を提案し, IML プロシジャを用いて推定方程式を直接解くことで実装した. 推定方程式の視点から各手法を捉えることで, 一見異なる解析手法の共通構造を理解し, 推定量の導出や仮定について体系的に考察することが可能となる. また, 推定量全体を推定方程式として記述することで, 複数の推定過程

を含む手法の分散推定への拡張や，独自の補正項・推定対象を組み込んだ推定量の構築にもつながる．IML プロシジャは，このような推定方程式を直接実装するための柔軟な基盤として，統計手法の理解と新たな推定量の実装の両面で有用であると考えられる．

参考文献

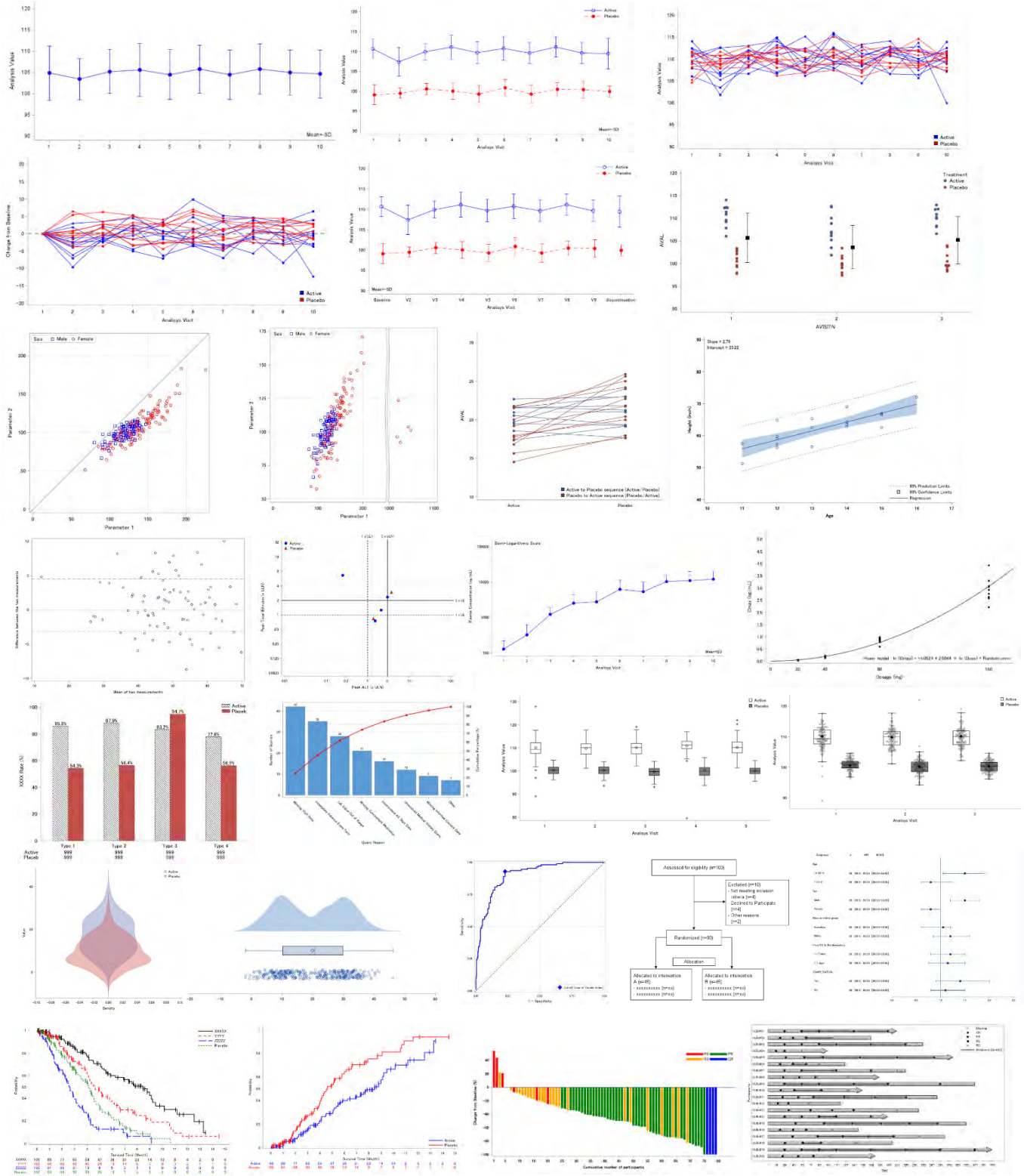
1. Edwards JK, Cole SR, Moore RD, Mathews WC, Kitahata M, Eron JJ. Sensitivity analyses for misclassification of cause of death in the parametric g-formula. *Am J Epidemiol*. 2018;187(8):1808–1816.
2. Neuhaus JM. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*. 1999;86(4):843–855.
3. Ross RK, Zivich PN, Stringer JSA, Cole SR. M-estimation for common epidemiological measures: introduction and applied examples. *Int J Epidemiol.* 2024;53(2):dyae030.
4. Shinozaki T, Suzuki E. Understanding marginal structural models for time-varying exposures: pitfalls and tips. *J Epidemiol.* 2020;30(9):377–389.
5. Stefanski LA, Boos DD. The calculus of M-estimation. *Am Stat.* 2002;56(1):29–38.
6. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol.* 2004;159(7):702–706.
7. 篠崎智大，萩原康博，田栗正隆，松山裕．疫学・臨床研究のための因果推論―Robins の g-methods による現実問題へのアプローチ―．東京：朝倉書店；2026.
8. 堀本隆一．IML プロシジャで理解する Cox 回帰．SAS ユーザー総会；2025.

SASグラフ100景

森岡　裕 Yutaka Morioka

（イーピーエス株式会社）EPS Corporation.

臨床試験でよく見かける以下のプロットについて描画方法を解説する.



多時点の相関を解析する - TQT試験での検討事例 -

○仲家 諒[1]

([1]武田薬品工業株式会社)

2009年に本邦においてICH E14が発出されて以降、新薬候補化合物の催不整脈作用の潜在的可能性を臨床段階において評価するためのThrough QT試験が多く実施されてきた。Through QT試験では、QT時間の日内変動を考慮するために日内で多時点の心電図測定を行い治験薬投与後の最大のQT延長作用を評価するが、試験デザイン（並行群間、クロスオーバー）やベースラインの設定方法など様々な検討が必要になる。加えて、解析段階においては時点間の相関の考慮も必要となる。本発表では、当時の試験立案時の試験デザイン、解析のモデル、相関構造などの検討内容をまとめ、ICH E14発出移行のQT延長作用の評価に関する規制の変遷をまとめる。あわせて、時点間相関を可視化するためのヒートマップ作成プログラムのSASパッケージCorrも紹介する。

非臨床研究データの統計解析の注意点
（British Journal of Pharmacology の実験計画と解析のガイドラインを中心に）

○原田　淳

（第一三共・生物統計＆データマネジメント部）

本文

生物学研究において様々な研究領域で多くの論文が発表されているが、それらの論文で報告された新規な知見の内容を第三者が追試すると再現性が低い論文が多数存在することが問題となっている。それらは、科学研究の信頼性を損ねてその健全な発展を阻害するだけでなく、その研究成果を基にして研究を行った他の研究者の労力の無駄使いにもつながっている。

そのような問題によって生じる研究費の空費を憂慮した機関が医学研究助成金の交付を担う英国の医学研究審議会 (MRC) や米国の国立衛生研究所 (NIH) であり、2013年に科学論文雑誌の編集者と共同して、研究成果の再現性・信頼性を高めるための検討を行い、その結果は例えば動物実験に関する ARRIVE ガイドラインに反映され、また各種科学雑誌もそれに準じたガイドラインを制定している。またそれらの内容は研究者からのフィードバックや編集者間の議論をもとに改訂が行われてきた。

研究成果の再現性の悪さの原因は必ずしもデータの捏造や改ざんのような研究不正に限らない。いろいろ実験した中から都合の良い結果だけを示すいいとこ取りや、十分な検出力を確保しない等の不適切な研究計画、結果が出た後で様々な解析法を試して最もよい結果を選ぶ (P-hacking) 等の不適切なデータ解析をはじめとして様々な原因が考えられる。そのため、ガイドラインでは適切な研究計画とデータ解析を行うための注意点が述べられており、それらの項目は投稿された論文の審査の基準にも使われている。

本発表では、国際的に著名な科学雑誌である Nature や薬理学系の British Journal of Pharmacology (BJP) および Journal of Pharmacology and Experimental Therapeutics (JPET) のガイドラインの内容を中心に、非臨床研究で求められる信頼性や妥当性を担保するための適切な研究計画とデータ解析の方法について紹介する。

RでPROC COMPAREっぽいことをやってみた

○関根　暁史

（藤本製薬株式会社）

本発表では、RでSASのPROC COMPAREに相当するデータ比較手法として、arsenalパッケージのcomparedf関数の活用を紹介する。値の差分確認に加え、by変数を用いた比較(SASのID値相当)や、tolオプションによる丸め誤差の無視(SASのcriterion=相当)など、実務で有用な機能を解説する。また、比較結果をCSVファイルとして出力する方法についても示す。さらに、all.equal関数を併用することで、主に属性差分の確認が可能である点にも触れる。これらの機能を、5行5列のダミーデータを用いて分かりやすく紹介する。

SASは次のステージへ ～ AI・オープンソース・高速分析を実現する最新Viya アップデート～

土生敏明 (SAS Institute Japan)

生成AIの普及やオープンソース技術の拡大により、データ分析を取り巻く環境は大きく変化しています。SAS Viyaもこうした変化に対応し、分析者の生産性向上とデータ活用の高度化を支援する機能を続々と提供しています。 本セッションでは、３つの機能を中心に紹介いたします。
①SASとRの連携を大きく進化させるPROC Rをご紹介します。Viya環境上でRプログラムを実行し、SASデータとの連携を容易に行えるため、既存のR資産を活用した分析基盤の構築が可能になります。 ②近年注目を集めるDuckDBとの連携機能をご紹介します。Parquetなどのオープンフォーマットを活用しながら、高速な分析処理とストレージコスト削減を実現する新しいアプローチについて、具体例を交えて解説します。 ③SAS Viya Copilotによる生成AI活用についてもご紹介します。自然言語による分析支援やコード生成、データ探索など、分析者の業務を大きく変える最新機能をデモを交えながら解説します。近年のViyaアップデートで利用範囲が拡大しているCopilotの方向性と今後の展望についても触れます。

非臨床統計の現在地
― 見えている課題から考える　関わり方とつながり ―

○本田　主税
（小野薬品工業株式会社）

非臨床統計は、薬理、毒性・安全性、CMC（Chemistry, Manufacturing and Controls：品質・製造）など複数の分野にまたがって関与する一方、その全体像や役割は必ずしも十分に共有されていない。近年では、毒性試験の背景データを活用して対照群を仮想化するバーチャル対照群（Virtual Control Groups：VCG）の導入、in vivo QT 試験に代表される高頻度・経時データの解析、動物実験代替法（New Approach Methodologies：NAMs）における予測性能の評価、CMC における統計手法の高度化など、非臨床分野で統計の関与が求められる課題は拡大している。本講演では、これらの課題を手がかりに、非臨床統計の現在地を俯瞰する。

はじめに、非臨床統計が関わる分野の広がりを概観する。薬理・効力評価、毒性・安全性、安全性薬理における高頻度データ、薬物動態、CMC、データ標準、さらに実験計画法や多重比較、生存時間解析といった分野横断的な手法まで、その裾野は広い。同じ統計手法を用いる場合でも、試験の目的、データの構造、意思決定が行われる場面は分野ごとに異なる。

次に、欧米における企業内統計職能と外部コミュニティの横断的な接続と、日本における分野別の実装との違いを整理する。欧米では、規制上の評価区分が企業内の統計職能や専門コミュニティの構成にも反映されやすく、非臨床統計が横断的な専門領域として認識されている。一方、日本では分野ごとの課題解決型の取り組みが中心であり、非臨床の統計を支えてきた担い手は確かに存在するものの、その役割が「非臨床統計」という職能として外部からは見えにくい状況にある。

そのうえで、課題が明確であれば場は作れる一方、それを支える担い手をどう育て、つなぐかという課題について考察する。実際に、規制上の要請が明確な領域では、研究会や研究班を通じて実務家と統計の専門家が具体的な議論を重ねている例もあり、分野を越えた議論が成立する条件について示唆が得られる。

さらに、本セッションを分野を越えた接点の一つとして位置づけ、異なる分野からの取り組みを共有することの意義を示す。本セッションでは、試験計画と報告の信頼性、毒性試験における VCG とデータ標準、in vivo QT 試験の解析、CMC における工程モデルの統合、生存時間解析における多重比較法の性能評価と、対象も手法も異なる発表が一つの場に並ぶ。それぞれの分野で直面している課題や工夫を並べて眺めることで、個々の分野に閉じていては見えにくい共通性と差異が浮かび上がることを期待している。

OSSコミュニティに参加しよう！

○仲家 諒[1]

([1]武田薬品工業株式会社)

OSS（Open Source Software）は、多くの人の参加によって成り立ち、継続していく活動です。一方で、「OSSコミュニティへの参加は難しそう」「GitHubに詳しくないと参加できない」と感じ、興味があっても一歩を踏み出せない方も多いのではないでしょうか。本発表では、PharmaForestの例をもとにOSSコミュニティへの参加を「興味を持つ」「応援する」「利用する」「報告する」「情報共有する」「自身のパッケージを共有する」「活動・レビュー・テストなどに貢献する」「運営に参加する」の8段階に細分化して紹介します。最初からコードを書く必要はありません。ニュースレターを読む、ステッカーを貼る、Starを付けるといった小さな一歩も立派な参加です。日本からOSS活動に参加する人を増やすため、まずは自分に合った無理のない関わり方を見つけてもらい、OSS参加へのハードルを下げることを目指します。

Robustな解析手法のSAS実装　-KL-divergenceやβ-divergence-

笹原 大敬

（イーピーエス株式会社）

実臨床で収集されるデータにはしばしば外れ値が含まれ、このデータ汚染が統計的推定結果に深刻なバイアスをもたらす場合がある。従来、こうした外れ値の影響を和らげる代表的なロバスト統計手法として、残差の大きいデータの重みを制御する「M推定」などが挙げられる。

これに対し本発表では、個々のデータ点の制御にとどまらず、確率分布全体の形状に着目した数理的アプローチについて紹介する。具体的には、最尤推定と数理的に同等であり、外れ値の混入によって結果が歪みやすい「KLダイバージェンス（Kullback-Leibler divergence）」の概念をまず紹介する。その上で、密度関数のべき乗を重みとして利用することで外れ値へのロバスト性を高めた「β-divergence（Density Power divergence）」などのロバストダイバージェンスに基づく統計的推定手法を紹介する。

分布間の幾何学的距離を数理的にコントロールするこのアプローチが、データ汚染環境下において、外れ値に頑健な推定精度を持つことを説明する。また、M推定（PROC ROBUSTREG等）とロバストダイバージェンスの手法について、それぞれの特性を反映した具体的なSASプログラムについても詳しく紹介する。

[参考文献]

・ 藤澤洋徳. ロバストな推定を導く ダイバージェンス. 第41回情報論的学習理論ワークショップ https://ibisml.org/wp-content/uploads/2020/10/ibisml41-fujisawa.pdf

・ 金森敬文, 藤澤洋徳. ダイバージェンスによる統計的推論. 日本統計学会誌, 2017, 47(1), 1-18. https://www.jstage.jst.go.jp/article/jjssj/47/1/47_1/_pdf/-char/ja

アダプティブデザインの実践

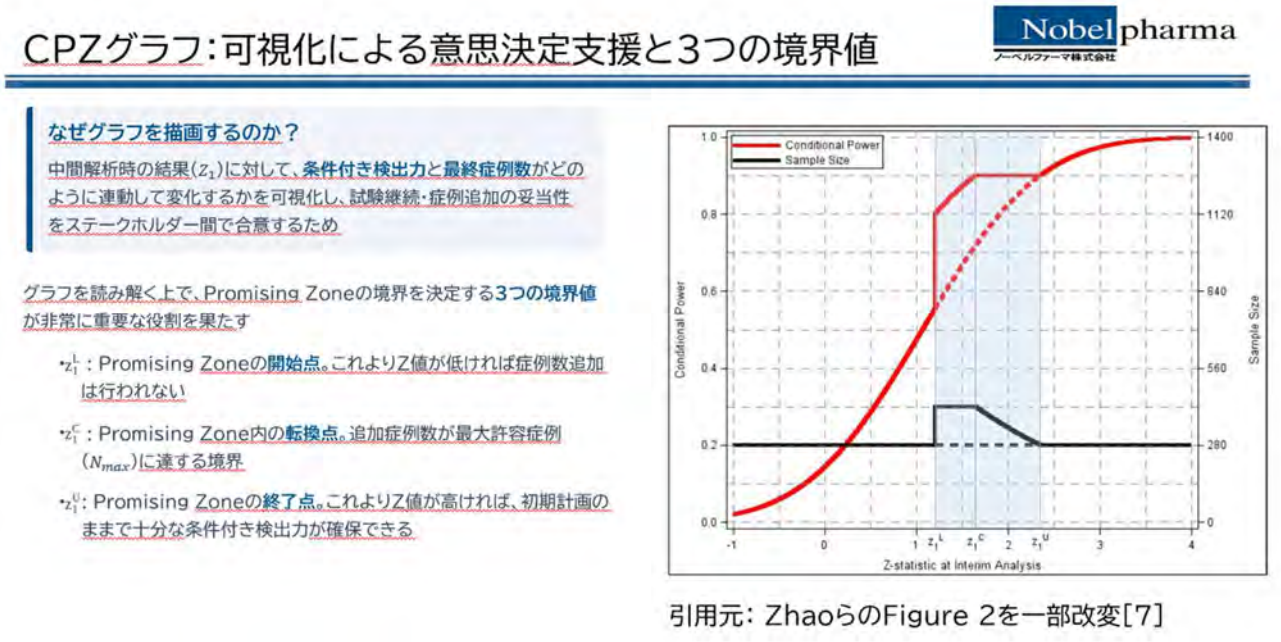
～CPZデザインの計算・評価・可視化を支えるSAS実装～

○内藤　寿輝、太田　裕二、森田　祐介

（ノーベルファーマ株式会社　データサイエンス部）

本発表では、Constrained Promising Zone（CPZ）デザインを用いた症例数再推定について、SASによる実装手法を紹介する。

昨年度の発表ではデザインの概念的紹介に留まったが、本年はSASプログラムの具体的なロジック解説にフォーカスを当てる。具体的には、条件付き検出力に基づく症例数再推定から、動作特性評価、そして総合的な意思決定を支援する「CPZグラフ」の可視化まで、一連のプロセスを実現するSASコードの要点を提示しながら解説を行う。



[7] Zhao Yang, Shivani Nanda. Promising Zone Designs for Sample Size Re-estimation in Clinical Trials: Graphical Approaches Using SAS. PharmaSUG 2025, San Diego, CA, 2025.

小さな製薬企業における統計解析プロセスと効率化の紹介

～試行錯誤の段階にある、一担当者によるリアルな事例共有～

○太田　裕二、森田　祐介、内藤　寿輝

（ノーベルファーマ株式会社　データサイエンス部）

小さな製薬企業における統計解析業務は、限られた社内リソースのもと全面的にCROへ委託して進められることが多い。特にノーベルファーマ株式会社では、多岐にわたる疾患領域や臨床担当による要望の多様化に加え、社内の統計解析担当者の経験や委託先CROの独自仕様による多様化が重なり、プロセスの標準化が困難な構造が存在する。

本発表は統計解析プロセスの成功事例の報告とはならないが、検討段階のリアルな事例としてこの「2つの多様化」による問題点を共有する。さらに社内における効率化のための標準統計解析帳票の作成とマクロ開発の試みについて紹介し、今後の改善に向けた一担当者視点での展望を考察する。

数式なしのPROC MCMC入門

○大山 暁史

（イーピーエス株式会社）

近年、臨床試験におけるベイズ流解析の活用が進展しており、米国FDAおよび欧州EMAにおいても関連文書の整備が進められている。ベイズ流解析の理論を解説した有用な参考書は数多く存在するが、その一方で、解析計画を実際のSASプログラムへ落とし込むイメージを持つことは、初学者にとって依然として容易ではない。

本発表では、SASのPROC MCMCを用いたベイズ流解析の基礎について、数式を用いずに直感的な理解を目指した解説を行う。まず、事前分布、尤度、事後分布といったベイズ流解析の基本概念を説明した後、MCMC（Markov Chain Monte Carlo）の構成要素であるマルコフ連鎖およびモンテカルロ法の概念を具体例とともに紹介する。

また、PROC MCMCの基本的な記述方法として、PARMS、PRIOR、MODELなどの主要ステートメントや、NBI、NMC、THINといった代表的なオプションの役割を説明する。加えて、事後平均、HPD信用区間、等裾信用区間などの推定結果の読み方や、トレースプロット、コレログラム、密度プロットを用いたMCMCの挙動の確認方法についても整理する。さらに、General関数による任意の尤度関数の実装、階層ベイズモデルにおけるハイパーパラメータの設定、ならびに動作特性の評価など、実務で活用される発展的な機能についても簡単に紹介する。

本発表が、ベイズ統計の理論とSASプログラミングとの橋渡しとなり、初学者が臨床試験におけるベイズ流解析の実装イメージを深める一助となれば幸いである。

製薬協 非臨床 一般毒性・DX 活動紹介 1
－バーチャル対照群を用いた非臨床毒性試験評価法の概要－

○黒岡 貴生[1, 2]

（[1]EA ファーマ株式会社、[2]日本製薬工業協会 医薬品評価委員会 専門チーム ET1-5）

バーチャル対照群（Virtual Control Group : VCG）は、過去試験の対照群データを背景データ（Historical Control Data : HCD）として蓄積し、実装対象試験と比較可能な個体データを選択することで、同時対照群（Concurrent Control Group：CCG）を仮想的に補完または代替する手法である。VCG には、毒性学的判断の信頼性を維持しながら使用動物数を削減し、3Rs、とりわけ Reduction に貢献することが期待されている。

非臨床毒性試験では、臨床病理、器官重量、病理組織学的検査など多数の評価項目から、変化の程度、用量反応性、背景値との関係および関連所見との整合性を考察し、総合的な毒性解釈が行われる。さらに、動物種、系統、性、週齢、飼育環境、投与条件、検体採取法および測定法など、多様な要因が検査値や所見に影響する。このため、VCG の利用に際しては、単に過去の対照群データを集積して用いるだけでなく、実装対象試験との比較可能性を確保するためのフィルタリング条件、個体選択方法および評価手順を事前に規定する必要がある。また、その妥当性は、統計学的な群間差の再現性だけでなく、毒性変化の検出力や毒性学的判断の一致性を含めて検証することが重要である。

Standard for Exchange of Nonclinical Data（SEND）は、個体情報、体重、臨床病理検査結果、器官重量および病理所見などを共通の構造と標準化された用語で格納できるため、複数試験を横断した HCD の統合、比較、抽出および可視化を支える、VCG 実装に向けた有用なデータ基盤の一つと位置づけられる。

本発表では、臨床試験と非臨床毒性試験における外部対照の位置づけの違いを踏まえ、VCG の基本概念、HCD から VCG を構築する流れ、国際的な議論の動向および検証事例を紹介する。さらに、VCG 実装における SEND データ活用の意義を概説し、次演題における実務的検討につなげる。

製薬協 非臨床 一般毒性・DX活動紹介2

-バーチャル対照群導入に向けたSENDの活用支援-

○吉池 通晴[1, 2]

（[1]日本製薬工業協会 非臨床評価専門部会 一般毒性DX 、[2]第一三共株式会社 安全性研究所）

前発表者の継続演題として、バーチャル対照群導入に向けたSEND（Standard for Exchange of Nonclinical Data）の活用支援についての発表を行う。バーチャル対照群導入の観点において、SENDはドメインと言われる毒性試験における一般的なエンドポイントをモデル化する為に必要な変数を含んだデータセットから構成されている。どのドメインのどの変数にどのようなルールでデータ格納すればバーチャル対照群導入に向けて効率的なSENDデータセットが作成できるか紹介する。

SASとRにおけるFirth法の実装差の整理と数値的評価

○西川　寛来[1]、後藤　駿輔[2]

([1]協和キリン株式会社、[2]ノバルティス ファーマ株式会社)

【背景・目的】Firth法は、ロジスティック回帰の小標本バイアスを低減し、分離への対処にも用いられる。SASとRの実装差はCAMISで報告されているが、小標本・少数イベントの状況で、各設定の違いが結果に及ぼす影響は十分に整理されていない。本研究では、信頼区間・検定方式、標準誤差、計算アルゴリズムおよび収束基準の違いを整理し、結果の差とその要因を数値的に評価した。

【方法】SAS PROC LOGISTIC、SASマクロ%FL、Rのlogistfおよびbrglm2を比較した。2×2表では、複数の総症例数と群配分の組み合わせについて網羅的に設定した。共変量モデルでは、連続共変量間の相関の強さを段階的に変えた条件を設定した。両実装の出力について、信頼区間が対数オッズ比0を含むか、および有意水準に基づく判定が一致するかどうかと、数値差を評価した。

【結果】PROC LOGISTICのWald区間とlogistfのプロファイル罰則付き尤度に基づく区間（PL区間）とで、一部の表において判定が異なった。両者をPL区間に揃えると、これらの判定差は解消した。一方Wald区間同士を比較すると、少数の表で判定差が残ったが、通常の情報行列から求めた共通の標準誤差で区間を再計算すると、判定差は解消した。PROC LOGISTICとlogistfのPL区間、およびPROC LOGISTICとbrglm2のWald区間は判定が一致した。

【結論】今回検討した条件では、信頼区間の方式と標準誤差の計算方法の違いが判定差の要因となった。実装間の結果を比較する際は、区間・検定方式と標準誤差の定義を確認し、数値差、出力可否、点推定・区間計算それぞれの収束状態を併せて評価する必要がある。

USDMからTLFsまで ― メタデータ駆動によるEnd-to-End SASプログラム自動生成への挑戦

○魚田 孝志[1]、森岡 裕[1]

([1]イーピーエス株式会社)

臨床試験における統計解析業務では、プロトコルや統計解析計画書（SAP）をもとにADaM仕様書や解析帳票仕様書が作成され、SASプログラマはこれらを参照してADaMおよびTables、Listings、Figures（TLFs）を作成するプログラムを実装している。この一連の工程では、上流で定義された情報を人が繰り返し解釈し、仕様書やプログラムへ反映する作業が多く存在する。一方、CDISC Unified Study Definitions Model（USDM）やAnalysis Results Standard（ARS）など、臨床試験や解析に関する情報を機械可読なメタデータとして表現する標準化が進められている。

本検討では、これらの情報を活用し、プロトコルからADaM、TLFsまでをEnd-to-EndでつなぐSASプログラム生成の実現可能性を検討する。第一歩として、ADSLを用いたDemography Tableを最小ユースケースとし、Mock Tableから帳票作成に必要な解析対象集団、解析変数、群、解析手法、出力統計量、表示方法等の情報を洗い出す。これらを実行用メタデータとしてJSON形式で構造化し、あらかじめ用意した解析パターンと組み合わせることで、TLF作成用SASプログラムの自動生成を試みる。

次に、実行用メタデータの各項目について、ARS、ADaM metadata、SAP、USDM等のどこから取得可能であるかを整理し、既存標準で表現できる情報と追加の構造化が必要な情報を検討する。また、TLF作成に必要なADSL変数を起点として、SDTMからの導出ルールを構造化し、同様のメタデータ駆動型アプローチによるADSL作成用SASプログラムの生成を試みる。

本発表では、最小ユースケースをEnd-to-Endで実装する過程を通じて、USDM、SAP、ADaM metadata、ARS等に分散する情報をどのようにつなげればSASプログラム生成に利用できるのかを整理し、プロトコルからADaM、TLFsまでをつなぐメタデータ駆動型自動化の可能性と課題について考察する。

レポーティングにおけるRとSASの共存

- 単一ソースによる実行と文書化 -

○後藤　駿輔[1]、西川　寛来[2]

([1]ノバルティスファーマ 株式会社、[2]協和キリン株式会社)

本文

QuartoはMarkdownベースのレポーティングツールであり、プログラムコードとその実行結果を同一文書内で管理しながら、レポートを作成することができる。この仕組みにより、解析から報告までの一連のプロセスを文書化でき、再現性の高いレポーティングが可能となる。近年の統計プログラミング業務ではオープンソースソフトウェアの活用が拡大し、SASとRを併用する機会が増えている。一方で、Quartoなどのレポーティングツールで複数言語を同時に扱う場合、環境が複雑になりやすく、コード管理や実行環境の統一が課題となる。

そこで、本発表ではSASとRを連携したレポーティング手法として、QuartoのSASコードブロック、SASmarkdown、およびsasquatchの実行方式と必要なSAS実行環境を整理し、特徴の比較を行った。

ローカル環境にSASが導入されている場合、Quarto単体でも容易にSASとRを統合したレポーティングが実現可能であった。SASmarkdownもローカル環境のSASを用いることで、ログやHTML出力の制御機能を提供し、より柔軟なレポーティングが可能であった。一方、sasquatchはSASPyを介してリモートSAS環境と接続することで、ローカルにSASが存在しない環境においてもSASを実行することが可能であり、加えてRとSAS間のデータ受け渡しにも優れていた。

利用環境や目的に応じて適切な手法を選択することで、SASとRの両言語を用いる場合でも、再現性の高いレポーティングと効率的なコード管理が可能であることが示唆された。

PROC SGPLOTのPOLYGONステートメントによる RWD可視化の実践

○塩田 早恵、三浦 伸也

（シミック株式会社）

リアルワールドデータ（RWD）を用いたデータベース研究では、患者の薬剤使用の推移を表す治療パターンや階層構造など、複数のカテゴリ間の関係を可視化してデータの特徴を把握する機会が増えている。特に、治療の開始・変更・追加などにより経時的に分岐・遷移する複雑な治療パターンは、集計表や一般的な棒グラフのみでは全体像を捉えにくく、Sankey Diagramなどによる視覚的な表現が有用である。一方、SASには、これらのグラフを直接作成するための専用ステートメントが必ずしも用意されておらず、目的とする可視化に応じた実装上の工夫が必要となる。

PROC SGPLOTのPOLYGONステートメントは、データセットに格納されたX座標およびY座標を順番に結び、任意の多角形を描画する機能である。表現したいグラフを長方形、扇形、曲線で構成される帯などの図形として捉え、その輪郭を座標データへ変換することで、SASに専用機能がないグラフも作成できる。

本発表では、Meyersらが報告したPOLYGONステートメントと数式を組み合わせる作図手法に基づき、RWDにおける治療パターンや階層構造の可視化を題材として、座標設計とSASによる実装方法を紹介する。はじめにPOLYGONステートメントの基本的な考え方を説明し、その応用例としてMosaic Plot、Icicle Plot、Sunburst PlotおよびSankey Diagramを取り上げる。Icicle Plotでは長方形の頂点座標、Sunburst Plotでは三角関数による円弧座標、Sankey Diagramではカテゴリ間を接続する帯の輪郭を算出し、各グラフを描画する。

これらの例を通じて、外観や用途が異なるグラフであっても、「描画対象を図形として捉え、その輪郭を座標データへ変換する」という共通の考え方で実装できることを示す。POLYGONステートメントは、個別の作図手順にとどまらず、SASにおける可視化の選択肢を拡張するための汎用的なアプローチとなり得る。

参考文献

Meyers J. Combining Functions and the POLYGON Plot to Create Unavailable Graphs Including Sankey and Sunburst Charts. PharmaSUG 2024, Paper DV-155.

AIエージェントとOSSから考える今後のコーディングとSAS

○中松 建

（個人）

2026年にはコーディングエージェントと呼ばれるツール・サービスの普及が進み、コーディングの大部分をAIに任せることも可能になってきている。 その背景には、モデル自体の性能向上に加えて、それを制御するハーネスの普及があり、Agent Skillsは複数のベンダーで採用が進む共通規格で、ユーザー側からモデルの動作を制御するうえで非常に重要な技術のひとつとなっている。SASにおけるAgent Skillsの利用例として、SPF（SAS Packages Framework）用に作成し、PharmaForestで管理しているSkill[1]と、関連する取り組みとしてドキュメント作成用のSPFパッケージのSASDoGs[2]を紹介する。また、このようなハーネスで制御されたAIエージェントを中心とした開発を行うことで、これまで個人や少人数では扱うことが難しかった規模のプロジェクトにも取り組めるようになってきている。SASに関連する例として、Pythonのデータフレームを SASの記法で操作するライブラリとして開発しているlimulus[3]を紹介する。

一方で、生成AIの登場以降ではAI Slopと呼ばれる低品質なコードがOSSコミュニティにおいて問題となっており、SAS関連でも、多数のリポジトリを無差別的に対象としてAIにより生成されたプルリクエストが送られるという事例が発生した。 この実際の事例をもとに、生成されたコード等をレビューすることで現在のAIによる生成物の品質を確認するとともに、OSSへの貢献、ライセンス・規約、マナーといった観点から問題点を整理し、OSSに限らず一般的な組織でAIを利用する際の課題等を検討する。

参考資料

[1] PharmaForest「spf-skills」：https://github.com/PharmaForest/spf-skills

[2] PharmaForest「SASDoGs」：https://github.com/PharmaForest/SASDoGs

[3] k-nkmt「limulus」：https://github.com/k-nkmt/limulus

SASによるCDISC ARD実装とTable化の試み

～SASパッケージ化に向けて～

山野辺 浩己

（第一三共株式会社 生物統計&データマネジメント部）

CDISCが提唱するAnalysis Results Standard（ARS）では、解析結果を標準化して保持するためのAnalysis Results Dataset（ARD）が定義されている。ARDは、ARS Metadataと関連付けて統計解析結果を構造化することで、解析結果のトレーサビリティや再現性の向上、レポーティングの自動化に寄与するものと期待されている。ARDはCDISC標準の中でも比較的新しい概念であり、現時点ではR言語のcardsパッケージをはじめ、いくつかの先進的な実装例がコミュニティ内で公開・議論されている。本発表では、これら既存の実装アプローチを参考にしつつ、SASにおけるARD実装についての議論を深めていくことを目的に、検討した課題を報告する。

SASにおけるARDの具体的な実装方法は、まだ確立されたものが少なく、実務での活用に向けては試行段階にあるのが現状である。SASは長年ADaMやTFL作成において独自のワークフローを培ってきたため、既存のR実装をそのまま適用するのではなく、SASの特性に合わせた再設計が必要であると考えられる。

本発表では、既存のARDパッケージの構造やコンセプトを概観し、そこから得られた知見をもとに、SASを用いたARDの実装案を提示する。具体的には、実際の統計解析結果（記述統計、群間比較検定等）の情報をSASデータセットへ構築するプロセスについて、SASコードを交えながら解説する。さらに、構築したARDデータセットから最終的な報告用Table（TFL）を生成するためのTable化のロジックについても検討し、SASを用いた実装例を紹介する。ARDを経由したTable化により解析結果とレポート様式を分離できるメリットや、その過程で直面した課題（変数名の標準化、メタデータの持たせ方等）についても言及する。

本取り組みで作成したSASコードは、再利用可能な形に整理し、SASパッケージとして作成する予定である。パッケージ化により、ARDの導入をより簡便にし、SASユーザー間での共有・改善を促進することを目指す。本発表を通じて、ARDの実務導入における課題や工夫点を共有し、今後のSASパッケージとしての発展や、SASコミュニティ・PharmaForestを含めたパッケージ共有基盤での活用可能性について議論する。

PharmaSUG 2026 U.S.レポート

堺映奈

（協和キリン株式会社）

PharmaSUG 2026 U.S.に参加し、製薬・ライフサイエンス業界におけるデータ標準、統計プログラミング、生成AIなどの最新動向に触れる機会を得た。本発表では、学会の概要や現地で得られた知見を紹介するとともに、特に実務への影響が大きいと考えられるSDTM関連の2演題について報告する。

1つ目は、同一被験者に複数回のスクリーニングや登録が生じる場合のSDTM実装である。DMドメインにおける1被験者1レコードの原則を維持しながら、同一被験者に単一のUSUBJIDを付与し、各参加を異なるSUBJIDおよびDCドメインで管理する考え方が提案された。また、DMドメインへの情報集約やPinnacle 21で想定されるValidationメッセージなど、実装上の留意点についても紹介する。

2つ目は、SDTMIG v4.0に向けた主要な変更点である。Baseline Flagの廃止、Multiple Subject Instancesへの対応、SUPPQUALからNon-Standard Variablesへの移行を中心に、新規ドメインや既存ドメインの改訂内容を紹介する。

海外学会へ参加して最新動向や実装事例に触れることで、今後優先的に準備・改善すべき課題を考える契機となった。また、活発なQ&Aや参加者同士の議論から、学会は情報を得る場にとどまらず、実務上の課題や経験を共有しながら相互に知見を深める場であることを強く感じた。初めての海外学会参加を通じて得られた気づきと日常業務における変化についても併せて共有する。

正しく推定し、正しく理解するために

○中川 雄貴

（イーピーエス株式会社）

本発表では、臨床試験で用いられる連続型データ、2値型データ、生存時間データを題材に、解析結果を適切に解釈するために理解しておくべき統計手法の特徴と解釈上の注意点ついて整理する。近年、ICH E9(R1) によりEstimandの考え方が導入され、「何を推定したいか」を明確に定義したうえで試験デザインや解析手法を選択することの重要性が強調されている。しかし、得られた推定値を適切に理解するためには、解析手法の前提やデータの性質を踏まえた解釈も不可欠である。連続型データでは共変量調整が推定値や推定精度に与える影響を示し、2値型データでは二値化による情報損失や閾値設定の影響を考察する。さらに生存時間データでは、区間打ち切りや観察間隔が生存時間の推定に及ぼす影響を示す。これらの事例を通じて、統計解析から得られた数値をそのまま解釈するのではなく、データ収集方法、試験デザインおよび解析手法の特性を踏まえて解釈することの重要性について考察する。

SASによるベイズ流単群Poissonイベント発現率における症例数設計

Bayesian Sample-Size Determination for a Single Poisson Event Rate using SAS

○張　方紅

（バイオジェン・ジャパン株式会社）

1．背景　近年，ベイズ流デザインは，事前情報と試験データを統合し意思決定を事後確率として表現できることから，単群第II相試験や製造販売後調査（PMS）において注目されている．パラメータが関心ある区間に属する事後確率を用いて，Go（次の開発相に進む），No Go（開発中止）の基準を定義し，または，リサーチ・クエスチョンを設定して，ベイズ流症例数設計法が提案されている．二項データを対象としたベイズ流症例数設計法は数多く提案されている一方，ポアソン発現率を対象とした方法論は限られている．最近，単群ポアソン発現率に対するベイズ流症例数設計法が提案された（張，2025）．本発表では，その内容をもとにSASによる実装方法を紹介する．

2．方法　n例中発生したイベント件数XはPoisson($n\lambda$)に従うと仮定し，λに対してGamma(a, b)事前分布を設定すると，事後分布はGamma($x+a, n+b$)に更新される．関心のある事象は臨床的に好ましくない状態を反映するものと仮定する．このため，発現率が低いほど望ましい．λが閾値発現率λ_0を下回る場合を「有望」，対立閾値λ_1を上回る場合を「非有望」とし，Pr($\lambda < \lambda_0 | n, x$)≥1-$\alpha$を有効，Pr($\lambda$>$\lambda_1 | n, x$)≥1-$\beta$を無効の判定基準とする（$1-\alpha$，$1-\beta$はいずれも1に近い閾値確率）．ベイズ流症例数設計は，有効判定基準を満たす症例数と境界イベント数（有効基準を満たす最大の整数）の組合せ(n, u)を探索する問題として統一的に理解できる．Whitehead法は有望領域と非有望領域が隣接するように，Zaslavsky法は両領域が一点で重なるように症例数を探索する．Tan and Machin法は閾値発現率よりεだけ小さい仮想的な発現率の下で有望の判定基準を満たす最小の症例数を求め，改良手法はεの代わりに事後分布の平均が目標発現率λ_tと一致する制約を用いる．

3．結果と考察　Zaslavsky（2012）の血栓塞栓症の例（$\lambda_0 = 0.22$，$\lambda_1 = 0.148$，Gamma(2,1)，1-α=0.95，1-β=0.90）に4手法を適用したところ，必要症例数（境界イベント数u）はWhitehead法277例（u=47），Zaslavsky法307例（u=53），Tan and Machin法277例（u=48，ε=0.05），改良手法413例（u=75，λ_t=0.184）となった．真の発現率を固定した場合の動作特性を確認すると，Whitehead法とZaslavsky法は妥当な誤り確率と検出力（0.038/0.85，0.040/0.88）を示し，Tan and Machin法は誤り確率がやや大きく（0.068/0.91），改良手法はより高い検出力を与えた（0.063/0.97）．二項データ向けの手法がポアソン発現率データにも適用可能であることが示された．SASでは，Gamma事前分布とPoisson尤度に基づく事後分布計算，症例数探索アルゴリズム，動作特性評価を実装できる．

4．まとめ　本発表では，二項データを対象としたベイズ流症例数設計法をポアソン発現率データへ拡張し，判定基準，症例数設計（最近提案された改良手法を含む），動作特性評価までをSASで実装する方法を紹介する．単群第II相試験やPMSなど，カウントデータをエンドポイントとする単群試験への応用可能性についても議論する．

参考文献：張方紅(2025)．日本計算機統計学会第39回シンポジウム ／ 張方紅・他(2021)．計量生物学，42 ／Whitehead(2008). Stat. Med. 27 ／ Zaslavsky(2012). Stat. Biopharm. Res., 4 ／ Tan&Machin (2002). Stat. Med., 21.

千葉大学医学部附属病院臨床試験部における
統計解析業務の現状と課題
―少人数体制での効率化・標準化に向けた実践―

○髙橋 康平[1]、高橋 杏奈[1]、小澤 義人[1]、稲葉 洋介[1]

([1]千葉大学医学部附属病院　臨床試験部　生物統計室)

【背景・発表の目的】

千葉大学医学部附属病院 臨床試験部 生物統計室において生物統計家は、医師主導治験や特定臨床研究等、年間数十試験の統計解析支援を担っている。限られたリソースの中での高品質な解析と迅速なレポーティングを提供するためには、解析業務の効率化と品質管理の標準化が不可欠である。しかし実際の解析手順やプログラム記述は属人化しやすく、室員間でのノウハウ共有や継承に問題を抱えてきた。そこで当室はSASマクロの共有など、プログラム標準化を目指した取り組みを実施している。本発表では現在実施している取り組みや直面している課題を共有するとともに、数年内で目指す現実的なゴールについて紹介する。

【施設体制】

当室は、所属する生物統計家4名が10年以上のベテランと3年未満の若手に二極化している特徴がある。多数の研究支援を遅滞なく進行しつつ技術継承と解析品質を両立させるため、当室では実務分担を行っている。具体的にはベテランがマクロの作成や配布を担い、若手はそのマクロを使用して解析実装する運用を取っている。また、効率的に解析結果の品質を確認するため、セルフチェックを含む目視チェックに加え、重要な項目を中心にダブルプログラミングによる詳細なQC（Quality Control）を実施している。

【課題】

解析プログラム記述の共通化によるQC時間の短縮や、解析帳票・Mockupの定型作業化による記載ミス削減を目指し、SAS共通ライブラリの構築に取り組んでいる。将来的なデータセットの標準化を見据えつつ、まずは利用頻度の高い背景情報と有害事象の集計を対象に記述共通化に向けた導入を開始した。一方で、マクロのバージョン管理や作成ルールの定着が懸念事項であり、GitHub導入によるバージョン管理の体系化や、AIを活用したプログラミング記述支援など最適な運用のあり方を模索している。

【結語・今後の展望】

少人数かつ経験差の大きいチームにおいて、マクロ等のツール整備をはじめとした業務の標準化は、解析作業の効率化と若手育成を同時に推進する手段となる。本発表では、現場での試行錯誤をオープンに共有し、アカデミアにおける持続可能な統計解析基盤の構築について、他施設の参加者と深く議論したい。

臨床試験のデータ解析とOSSの最近の動向

堀田真一

（協和キリン株式会社）

臨床試験のデータ解析へのオープンソースソフトウェア、特にRの導入が進んでいると言われている。同時に、Rの導入に際して、RとパッケージのCSV（Computerized System Validation）が課題になっているとも言われている。2026年7月に行われた「第10回データサイエンスラウンドテーブル会議」では、普段は臨床試験のデータ解析を主に担当している参加者が、CSVのみならず、ソフトウェアの品質保証の新しいフレームワークであるCSA（Computer Software Assurance）についても学び、Rとパッケージの品質保証をどのように行うべきかをディスカッションした。この際の議論の結果（https://www.pmda.go.jp/files/000281996.pdf）を基に、臨床試験のデータ解析におけるRの現状と、今後の展望を考察する。

Logistic一般化線形混合モデルにおける周辺平均の推定

○王春弋

（エイツーヘルスケア株式会社）

背景

一般化線形混合モデル（Generalized Linear Mixed Model：GLMM）は、反復測定データやクラスター化データなどの相関構造を有する非正規データの解析に広く用いられている。Logistic一般化線形混合モデル（Logistic Generalized Linear Mixed Model: Logistic GLMM）では、処置群の固定効果の推定値はランダム効果が与えられたもとでの条件付き平均であり、集団全体における処置群の平均（周辺平均）とは一般に一致しない。また、臨床試験や疫学研究では処置間の差の周辺平均の評価が重要となり、その推定には一般化推定方程式（GEE）も用いられるが、それに対して、GLMMは被験者内相関や欠測データに容易に対応できる利点を有する。このため、本研究ではGLMMの枠組みにおける周辺平均の推定に着目する。

方法

Logistic GLMMにおける条件付き平均と周辺平均の乖離は非可縮性（non-collapsibility）に起因する。周辺平均を得るにはランダム効果を積分消去する必要があるが、一般に閉形式解は得られない。そこで本研究では、Monte Carlo積分およびZeger近似による推定法を取り上げ、周辺平均確率、周辺オッズ比（OR）、周辺リスク差（RD）及び信頼区間を推定する。さらに、反復測定の2群比較のシミュレーションを通じてランダム効果分散、処置成功/不成功の二項確率水準（低、中、高の3領域）およびサンプルサイズが推定性能に与える影響を評価し、推定結果のバイアスおよびカバレッジ確率（CP）を算出した。また、条件付き平均も算出し、条件付き平均と周辺平均の違いを明示した。

結果

Monte Carlo積分法およびZeger近似法のいずれにおいても、周辺平均確率およびRDは真値に近い推定値を示したが、全ての確率領域において、Monte Carlo積分法のバイアスはZeger近似法よりも小さかった。ORには真値からの乖離が認められ、特に成功率が高い条件下でバイアスが増大する傾向がみられた。また、サンプルサイズの増加に伴い、そのバイアスは縮小した。ORとRDのCPはいずれの条件でも概ね95%付近であった。さらに、条件付き確率と周辺平均確率を比較した結果、低確率領域では条件付き確率が周辺平均確率を下回り、高確率領域では上回る傾向が認められた。

結論

どちらの方法においても高確率領域においてORのバイアスは増大した一方、RD推定に対しては限定的であり、RDはより頑健な指標であることが示唆された。また、Monte Carlo積分法による推定値は真値と良好に一致し、Zeger近似法より高い推定精度を示した。

SASPy応用: 新機能の紹介から実践例まで

○中松 建

（個人）

Pythonは近年で特に人気の高いプログラミング言語のひとつであり、機械学習やAI関連の多くのライブラリやツールがPythonを使用していることから、これらの分野への対応のためにも、Pythonの利用が選択肢となる場面は以前より増加しているものと考えられる。PROC PYTHONがSAS9.4 M9のhotfixでサポートされるなど、Pythonとの新たな連携方法も増えつつあるが、複数の接続方法に対応しPython側からSASを呼び出せることから、依然としてPythonとの連携においてSASPyは有力な選択肢のひとつである。

SASPyに関してはこれまで入門的な紹介が多かったことから、本発表ではSASPyを使用したことがある、またはすでにPythonをある程度利用しているユーザーを対象として、応用的な内容を中心に扱う。前半ではSASPyに適したケースや設定に関するこれまでの変更点など、利用環境や設定に関する内容を紹介する。後半ではstreamlitによる簡易的なアプリケーションの構築、およびAIエージェントとの連携で重要なMCP(Model Context Protocol)サーバーの構築を通して、発表者のプルリクエストによってv5.105.0 で取り込まれた機能と、実践的な利用におけるポイントを紹介する。

参考資料

[1] SASPy v5.105.0 リリースノート：https://github.com/sassoftware/saspy/releases/tag/v5.105.0

[2] サンプルコード公開リポジトリ：https://github.com/k-nkmt/saspy_advanced

3群以上で使える共変量バランス評価方法

○中倉 章祥

（久留米大学バイオ統計センター）

群間比較では、共変量の分布が群間で均衡していることが重要である。分布に偏りがみられる場合、傾向スコアなどを使った調整が試みられるが、調整前後の共変量バランスを評価することは不可欠である。2群比較では標準化平均差（SMD）が広く用いられており、一般に0.1や0.2以下が望ましいとされる。一方、3群以上の比較ではSMDはすべての群の組み合わせに対して算出する必要があり、群数の増加に伴って解釈が複雑化する。SMDの絶対値の最大値や平均値を代表値として確認することもあるが、単一の指標では十分にバランスを評価できない場合があり、確認の仕方も定まっていない。本発表ではSMDを用いた共変量バランス評価方法の問題点を整理し、Population Standardized Bias（PSB）の考え方を用いて、3群以上でも解釈しやすい共変量バランス評価方法を検討する。

PSBは解析対象集団の平均値と各群の平均値との差を、解析対象集団の標準偏差で標準化した指標である。SMDが2つの群の距離の指標として解釈される一方、PSBは解析対象集団から各群への距離の指標として解釈される。PSBは解析対象集団を基準として各群の偏りを評価できるため、群数が増えた場合にも各群のバランスを同じ基準で確認しやすい。PSBの絶対値の平均値を用いて全体的なバランス評価を行い、各群のPSBによって群ごとの詳細な検討が可能である。これらの特徴を活かして、実際に傾向スコアによる3群の共変量調整が行われた研究を題材に、直感的に理解しやすい図表の示し方を提案する。

日本発！眼科領域における TAUG 開発

○山崎 文寛[1]

([1]武田薬品工業株式会社)

医薬品開発では、新薬申請時に提出する臨床試験データについて、CDISC 標準と呼ばれる国際的なデータ標準の利用が進んでいる。Therapeutic Area User Guide（TAUG）は、CDISC 標準を疾患・治療領域ごとのデータにどのように適用するかを示したガイドであり、これまでに約 50 種類が公開されている。しかし、眼科領域を対象とした TAUG は存在しない。そこで、CDISC Japan User Group（CJUG）-SDTM および CJUG-ADaM による合同プロジェクト「チームいちがん」は、米国 CDISC 本部との協力のもと、日本に拠点を置く製薬・医療機器企業および CRO などのメンバーがそれぞれの専門性を持ち寄り、完全ボランティアで眼科領域 TAUG の開発に取り組んでいる。日本主導での TAUG 開発は初の試みであり、データ収集から解析までを一貫したストーリーとして整理しながら検討を進めている。

活動は 2 段階のスキームで進めており、ADaM チームの活動として、第 1 段階として CJUG 内で日本語版の TAUG 案を作成した。現在は第 2 段階として英訳および米国 CDISC 本部との連携による CDISC のウェブサイトでの公開準備に移行している。第 1 段階では、ClinicalTrials.gov 等を用いた過去の眼科試験の評価項目・解析手法の調査、およびリリース済 TAUG の構成・詳細度の横断比較を実施し、そのうえで ADaM パートを連続値とカテゴリ値の 2 本立てで構成し、Statistical Endpoint から SDTM、ADaM、Analysis Results Metadata までを一貫して例示した日本語ドラフトを作成した。

本発表では、これらの取り組みを通じて得られた知見と課題を共有する。限られたリソースの下では「どこまでを対象とするか」を早期に宣言することが決定的に重要であること、SDTM と ADaM が同じ議論の場に立つことで収集から解析まで一貫したストーリーを描けること、そして役割を明確化することがボランティア活動の推進力となることを、実例を交えて紹介する。あわせて、CDISC Operating Procedure COP-001 Standards Development に基づく標準開発プロセスにおける現在地と、米国 CDISC 本部との主な論点についても報告する。

再生医療等製品の開発におけるベイズ流拡大試験の提案と適用事例

上村鋼平 (東京大学)

再生医療等製品の開発では、条件期限付き承認後の本承認を目指した市販後のエビデンス構築の仕方が課題となっている。本発表では、このような現状を受け、条件期限付き承認時の有効性データを事前情報として動的借用を行うベイズ流拡大試験の方法論を提案し、その適用事例を紹介する。提案法における事前情報は、外部対照とは異なり、有効性の比較データを想定している。適用事例に関して規制当局との議論では、第一種の過誤確率の制御が大きな論点となったが、提案法は、ハイブリッドコントロール法における動的借用の方法論をうまく適用することにより、第一種の過誤確率を制御しつつ、試験の成功確率を上昇させることが可能である。本発表では、適用事例の背景を説明し、提案法について説明を行う。シミュレーション実験により提案法の動作特性を調べた結果を示し、今後の課題等について議論する。

Integrated Analysis Data Reviewer’s Guide Completion Guidelinesの解説と実務におけるポイント

○長又　理緒

（アストラゼネカ株式会社）

Integrated Analysis Data Reviewer's Guide Completion Guidelines（以下、iADRGガイドライン）は、ISS (Integrated Summary of Safety) やISE (Integrated Summary of Efficacy) などの統合解析データセットに対するReviewer’s Guideの作成指針を示した文書である。iADRGは規制当局の審査官に対して統合解析データセットの作成方針や背景情報を提供し、Define-XMLや統合解析計画書だけでは十分に伝わりにくい統合時の判断やデータの関係性を補完する役割を担う。統合解析では複数試験のデータを統合する特性上、標準や辞書のバージョン統一、試験間で異なる治療変数や解析対象集団の標準化など、個別試験のADRGにはみられない固有の課題が発生する。これらの課題に対する対応方針をiADRGで適切に説明することにより、審査官によるデータの理解が促進されレビューの効率化につながることが期待される。

iADRGガイドラインでは、統合解析に使用した標準や辞書、ソースデータ、トレーサビリティ、統合解析に含めた試験の要約、解析上の考慮事項、統合データセットの作成手順、データセットの説明、バリデーション結果、および提出プログラムに関する情報を記載することが求められている。一方で、同ガイドラインは英語で提供されていることから、日本の承認申請における実務的な解釈や具体的な記載方法について情報を整理する意義があると考えられた。

2024年度CJUG ADaM テーマ5では、英語で提供されているiADRGガイドラインを日本の実務担当者が簡便に理解できるよう日本語の解説書を作成した。解説書ではガイドラインの各セクションの要約に加えて、具体的な記載例や原文だけでは理解が難しい箇所に対する補足解説を掲載した。さらに、チームメンバーの実務経験や関連文献に基づく提案に加え、技術的ガイドなどPMDAの電子データ提出関連通知との整合性を踏まえた日本の申請実務への適用上の留意事項についても整理した。

本発表では、2024年度CJUG ADaM テーマ5で作成した日本語解説書の概要を紹介するとともに、iADRGの構成とセクションごとの主要な考え方を解説する。さらに、統合解析特有の標準化、トレーサビリティ確保、SASプログラム提出時の留意点など、検討チームの経験に基づいた提案を紹介しPMDA申請を見据えたiADRG作成の効率化と品質向上に資する知見を共有する。

SASを用いた抗体医薬品の各製造工程をつなぐ統合モデルの構築と統合モデルを用いた規格設定のリスク評価

伊藤勇弥 (小野薬品工業株式会社)

抗体医薬品の製造プロセスでは、複数の工程が連続し、それぞれの工程条件のばらつきが最終製品の品質に伝播する。規格の妥当性を説明する際には、最終品質特性が正規分布に従うことを前提とした統計的評価が広く用いられる。しかし実際には、工程間の非線形な関係、複数のばらつき要因の重なり等の影響により、最終品質特性は歪度や裾の厚さを持つ分布となり得る。このような場合、単一工程または最終試験データのみを対象とした評価では、規格逸脱確率、特にテール領域のリスクを適切に捉えられない可能性がある。 本発表では、各工程のProcess Characterization（PC）Studyで得られたモデルを連結し、入力から最終品質特性までを一貫して表現するIntegrated Process Model（IPM）の考え方及び、IPMによって推定される最終品質特性の分布の形状、規格逸脱リスクを評価することで科学的根拠に基づく規格の妥当性説明について紹介する。

RMST尺度における因果媒介分析

○石原 政佳、今村 拓未、渡辺 秀章、長谷川 貴大

（塩野義製薬株式会社　解析センター）

臨床試験では，関心のあるアウトカムに対する治療効果を評価するだけでなく，その効果がどのような作用機序を介して発現したかを理解することも重要である．因果媒介分析は，治療効果を直接効果と間接効果に分解し，バイオマーカーやPK/PD指標などを介した効果を評価するための統計的手法である．生存時間アウトカムに対する因果媒介分析については，製薬協データサイエンス部会の成果物 (2024) において，ハザード比に基づく手法が紹介されている．しかし，ハザード比は比例ハザード性を前提とする効果尺度であり，非比例ハザード下では適用や解釈に課題がある．一方，Restricted Mean Survival Time (RMST) は，臨床的に解釈しやすい効果尺度として近年注目されている．Chernofskyら (2025) は，pseudo-value法を用いてRMST尺度上で治療効果を直接効果および間接効果に分解する因果媒介分析手法を提案した．しかし，因果媒介分析の枠組みにおいて，RMST尺度に基づく手法の推定性能や適用上の特性については，既存手法との比較を含め，十分な検討がなされていない．

本研究では，Chernofskyら (2025) が提案したRMST尺度に基づく因果媒介分析をSAS上で実装するとともに，数値実験を通じてその推定性能および適用上の特性を評価した．具体的には，生存関数の推定結果に基づいてpseudo-valueを算出し，媒介変数モデルおよびpseudo-valueモデルを構築した．さらに，これらのモデルを用いて反実仮想下のRMSTを推定し，直接効果，間接効果および治療効果を推定する解析手順を実装した．数値実験では，治療から媒介変数への効果および媒介変数からアウトカムへの効果の大きさに加え，比例ハザード性の成立の有無別に複数のシナリオを設定した．比例ハザード性が成立する条件下では，RMST尺度に基づく手法が既存のハザード比に基づく手法と比較して劣らない推定性能を有するか否かを評価した．また，比例ハザード性が成立しない条件下では，RMST尺度に基づく手法の推定性能を評価し，ハザード比に基づく手法では適用が難しい状況における有用性を検討した．本発表では，RMST尺度に基づく因果媒介分析の考え方およびSASによる実装方法を紹介する．さらに，数値実験の結果を通じて，本手法の推定特性ならびに有効な適用場面について考察する．

参考文献

- 日本製薬工業協会 医薬品評価委員会 データサイエンス部会 2023年度タスクフォース3-1 (2024). “解釈のための因果推論～因果媒介分析編～”，医薬品評価委員会の成果物一覧. https://www.jpma.or.jp/information/evaluation/results/allotment/ds_202406_causal-ia.html (参照2026-03-05)
- Chernofsky A, Lok JJ. Causal mediation analysis for time-to-event outcomes on the Restricted Mean Survival Time scale: A pseudo-value approach. PLoS One. 2025 Apr 9;20(4):e0319074. doi: 10.1371/journal.pone.0319074. PMID: 40203275; PMCID: PMC11981657.

SASプログラムの副作用を考える

○平井健太

（株式会社タクミインフォメーションテクノロジー）

SASプログラムでは、データ加工や解析、帳票作成などの処理そのものに注目しがちであるが、プログラム実行によってSASセッションの実行環境が変更されることにも注意が必要である。例えば、WORKライブラリに作成された一時データセット、グローバルマクロ変数、SASシステムオプション、TITLE及びFOOTNOTE、ODS設定などが実行後も残存することで、後続するプログラムの処理や出力に意図しない影響を与える可能性がある。

このような実行環境への「副作用」は、個々のプログラムを単独で確認しただけでは発見しにくい。実行順序や実行前のセッション状態によって結果が変化する場合、プログラムの再現性や独立性が損なわれ、不具合発生時の原因特定を困難にする要因にもなり得る。また、複数のプログラムを連続して実行する業務や、独立したプログラムによる検証を行う場面では、特に留意すべき問題である。

本発表では、SASプログラムが実行環境に残す代表的な副作用を整理し、実行前の状態が後続プログラムの結果に影響する具体例を紹介する。さらに、不要なWORKデータセットやマクロ変数の管理、変更したシステムオプションの確認など、副作用を抑えるための実装方法について検討する。

プログラムの品質を「期待した結果を正しく生成すること」だけでなく、「実行環境に意図しない影響を残さないこと」という観点から捉え、再現性、独立性および保守性の高いSASプログラムを作成するための考え方を提示する。

SAS及びRによるノンコンパートメント解析（NCA）の実装とWinNonlinとの結果比較：統合的な薬物動態解析プロセスへの展開

○飯泉 寛

（Meiji Seikaファルマ株式会社）

医薬品開発における薬物動態（PK）評価では、ノンコンパートメント解析（NCA）が標準的に用いられ、その実施にはPhoenix WinNonlin（以下、WinNonlin）が広く利用されている。一方、SDTM、ADaM、TFL等の成果物は主にSASで作成されるため、PK評価を伴う臨床試験の解析プロセスにはWinNonlinが外部ソフトウェアとして介在する。このようなプロセスでは、複数部門の担当者や外部ベンダーが関与することも多く、データや解析結果の受け渡しに伴う作業負担、遅延および人的エラーのリスクが課題となり得る。

そこで、CJUG-ADaM臨床薬理チームでは、よりシームレスなNCAプロセスを志向する場面を想定し、WinNonlinを用いずにPKパラメータを算出する方法を検討した。現状の統計解析業務が主としてSASで実施されることから、SASによるNCAを主要な検討対象とした。また、近年のオープンソースソフトウェア（OSS）の活用状況を踏まえ、参考としてRでのNCAも検討した。

本発表では、SASおよびRによるNCAの実装方法を提示し、各方法による算出結果をWinNonlinと比較する。SASによるNCAには「PK-NCA SASプログラム」を、RによるNCAには「aNCA」を用いた。「PK-NCA SASプログラム」は、公開されているWinNonlinのNCAアルゴリズムを参考に、CJUG-ADaM臨床薬理チームがSASで独自に実装したプログラムである。「aNCA」はpharmaverseが公開している、PKNCAを計算基盤とするNCA用Rパッケージであり、PKパラメータの算出およびSDTM PPやADaM ADPP形式の出力を支援する。各ソフトウェア・パッケージのバージョンは、SAS 9.4、R 4.6.0、WinNonlin Version 8.6、aNCA Version 0.1.0を用いた。WinNonlinとの比較には、複数の投与条件（経口投与または静脈内点滴投与、単回投与または反復投与）および多様な濃度推移を含む仮想データを用いた。その結果、「PK-NCA SASプログラム」では、比較対象としたすべてのPKパラメータについて、すべての仮想データでWinNonlinと算出結果が一致した。一方、「aNCA」では、一部の投与条件において、WinNonlinと一致しないPKパラメータが認められた。本発表では、これらの比較結果を報告するとともに、SASまたはRを活用した統合的なPK解析プロセスへの適用可能性および今後の検討課題を論じる。

ダミーデータを使った時間依存性共変量の解析

○矢田 徹

（株式会社セブントゥワン）

2群間で治療効果を比較する際には、性別や年齢など治療効果に影響を及ぼすと考えられる因子について群間の偏りを考慮する必要がある。非介入観察研究では、傾向スコアを用いた調整により、治療開始時点で測定可能な因子について群間のバランスを改善することができる。一方で、このような調整は治療開始時点の情報に基づくため、治療開始後に変化する因子については十分に考慮できない。

例えば、肝疾患に対する2つの治療の効果を生存時間で評価する場合、治療開始後に実施される肝移植の頻度が群間で異なることがある。このような因子は治療開始時には存在せず、かつ予後に影響する可能性があるため、その取扱いが治療効果の評価に影響を及ぼす。

治療開始後に変化する因子を考慮する方法の一つとして、時間依存性共変量を組み込んだCox比例ハザードモデルがある。本手法では、ある事象の発生前後でハザードが変化すると仮定し、その影響を解析に反映することができる。

本発表では、ダミーデータを用いて時間依存性共変量を考慮したCox比例ハザード解析を実施し、時間依存性共変量を考慮しない解析結果との比較を行う。ダミーデータは、対象疾患の発症後に重症度や診断・治療状況が時系列で変化するモデルを想定し、追加治療による有効性および安全性への影響に加え、治療開始後の移植実施が予後に影響する状況をシミュレーションして作成した。解析にはSASを使用した。

なお、本発表の目的は時間依存性共変量を考慮したCox回帰モデルの基本的な考え方の紹介である。肝移植のような治療開始後事象については、時間依存性交絡を考慮した因果推論手法やEstimandの枠組みに基づく検討も重要となるが、これらについては本発表の対象外とする。

SASの世代データセット機能やAudit機能を利用する

森岡　裕 Yutaka Morioka

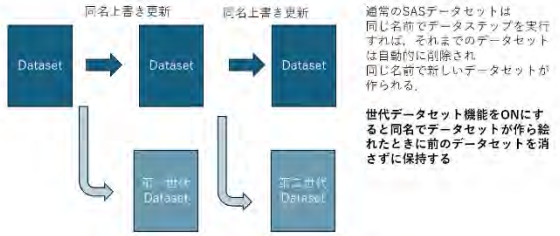
（イーピーエス株式会社）EPS Corporation.

SASでは、同名データセットの上書きやデータ更新により、処理前の状態や変更内容の追跡が難しくなる。本発表では、世代データセット機能および監査証跡（Audit Trail）機能を用いて、データ操作の再現性と変更履歴の可視性を高める方法を紹介する。

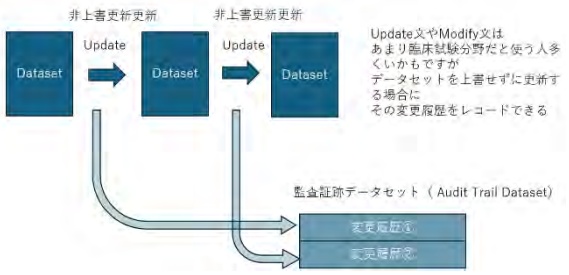
世代データセットでは、DATAステップのデータセットオプションgenmax=で保持世代数を指定し、同名で更新するたびに更新前データセットを世代として自動保存する。参照時にはgennum=で過去世代を指定できるため、更新前後の比較や復元が容易となる。

監査証跡では、PROC DATASETSのAUDIT文で監査を開始し、UPDATE文やMODIFY文等による変更を記録する。監査証跡データセットからは、操作日時（_ATDATETIME_）、対象オブザベーション番号（_ATOBSNO_）、実行ユーザー（_ATUSERID_）、操作種別（_ATOPCODE_）を取得できる。

世代データセット機能（ Generation Data Set ）



監査証跡データセット（ Audit Trail Dataset）



【結果・考察】

世代データセットは、条件やモデルを変えながら同名データセットを繰り返し作成する場面で、手動バックアップを省力化し、差分確認を迅速にする。change_tracer_dog パッケージ（https://github.com/PharmaForest/change_tracer_dog）を用いることで、世代間比較をより簡便に実施できる。

監査証跡は、データカットオフ時のレコード削除、日時の置換、対象者除外、nullフラグ処理等の詳細な変更記録に有効であり、datacut_sasパッケージ（https://github.com/PharmaForest/datacut_sas）では複数のカットオフ操作と変更内容をまとめたExcelレポートの作成を支援する。

【結論】

既存のSAS機能を適切に組み合わせることで、データハンドリングにおけるバックアップ、比較、変更履歴管理を実用的に強化できる。

臨床試験業務に使えるオープンソースSAS Package紹介

○森岡 裕

（イーピーエス株式会社）

【背景・目的】

臨床試験のデータ標準化、解析、帳票作成では、再現性・標準準拠・効率性を高める共通基盤が求められる。Rではpharmaverseが広く利用されている一方、SASにおいてもオープンソースのパッケージを共有・活用するためのエコシステムが整備されつつある。本発表では、製薬・臨床試験向けSASパッケージ群であるPharmaForestと、その実務利用を紹介する。

【方法】

PharmaForestはSAS Package Framework（SPF）を基盤とし、SASコードをパッケージとしてインストール・ロード・管理する仕組みを提供する。パッケージ保管先を設定しSPFinit.sasを実行した後、%installPackageおよび%loadPackageにより利用できる。ネットワーク接続が制限された環境では、パッケージを事前にダウンロードしてローカル保管先からロードする方法も利用可能である。

【結果・考察】

紹介するパッケージは、SDTM・ADaMの整合したダミーデータを生成するsas_faker、ハッシュオブジェクトによる結合・キー照合を支援するsashash、Rolling join等を実装するmergex、CTCAE v5.0 gradingを行うctcae5_grader、ADaM導出を支援するAdamskiを含む。さらに、ログ・データセット比較、Dataset-JSON／XPT変換、カットオフ処理の履歴化、CDISC ARSに基づくARD作成、集計・RTF帳票・オンコロジーグラフ・症例フロー図の作成までをカバーする。これらを組み合わせることで、臨床試験解析プログラミングの各工程を共通化できる。

【結論】

PharmaForestは、SASを用いる臨床試験業務において、オープンソースの知見を再利用可能なパッケージとして取り込む基盤となる。導入時にはライセンス、組織内での検証・バリデーション、解析結果の確認に留意しつつ活用することで、品質と生産性の向上が期待できる。

Rolling joinをSASで考えてみる

森岡　裕 Yutaka Morioka

（イーピーエス株式会社）EPS Corporation.

【背景・目的】

時系列データを扱う実務では、結合キーの一部が完全一致しない場合に、直前・直後または最も近い値を持つレコードを対応付ける処理が頻繁に必要となる。この処理はRのdata.tableでRolling joinとして明確に定義され、その後、as-of join等の名称で一般化されてきた。本発表では、Rolling joinの考え方を整理し、SASで再利用可能な形に実装する方法を示す。

【方法】

ID等の一致キー内で、時刻・日付等の順序キー（rollvar）について最も近い候補を選択する処理をRolling Matchとして定義した。マッチ方向は、現在値以前の最も近い値を選ぶBACK、以後を選ぶFORWARD、前後を比較するNEARESTの3種類とし、過去・未来方向の許容幅および等距離時の優先方向を指定可能とした。実装では、取得元データにオブザベーション番号を付与したビューを作成し、重複キーを許容するハッシュオブジェクトのDO_OVERメソッドで候補を探索する。候補レコードの特定と値の取得を別ハッシュに分けることで、1対多によるレコード増加を防ぎつつ必要な変数を取得する。

【結果・考察】

作成した%rolling_matchマクロは、master、key、rollvar、rolltype、取得変数等の指定により、直前投与量の参照、時点に応じた情報付与、最近傍値の選択といった処理を統一的に記述できる。さらに、重複キーを扱うDO_OVERメソッドおよび、キー制約を超えて順序付きハッシュ内を走査できるハッシュ反復子オブジェクトは、条件表との照合、期間の集約、総当たりによる最近傍探索にも応用できる。

【結論】

日常的な条件付き結合にRolling joinという名称と操作仕様を与え、マクロとして実装することで、複雑なデータハンドリングを可読性・再現性の高い処理へと整理できる

SAS環境で始めるR活用

- LSAFにおけるRパッケージ利用とpharmaverse導入事例

○田中　優司、堀田　真一

（協和キリン株式会社　バイオメトリックス部）

LSAF（Life Science Analytics Framework）は、SAS社が提供するクラウドベースの臨床試験データ統合解析環境であり、SASに加えてRを用いた解析も実行できる。一方で、標準環境では利用可能なRパッケージが限定されており、必要なパッケージの導入や管理が課題となっている。そこで弊社では、SAS社が提供するR Package Installation機能を活用し、LSAF内で新たなRパッケージを導入・共有できる環境を構築した。本発表では、LSAFにおけるRの基本的な利用方法を概説するとともに、miniCRANを用いた依存関係を考慮したパッケージ取得から、R Package Installationによる導入手順までを紹介する。また、導入事例としてpharmaverseパッケージ群の一つであるSDTM.oakを取り上げ、臨床試験データからSDTM形式データセットを効率的に作成する活用例を示す。これらの事例を通じて、LSAF環境におけるR活用の可能性と運用上の留意点について考察する。

Analysis Results Standard の解説

○福田　裕章
（MSD 株式会社）

Analysis Results Standard（以下，ARS）は，解析結果の標準として CDISC にて開発された。従来の解析結果は，各社でレイアウトが異なる静的な RTF/PDF ファイルとして作成されることが多く，再利用やトレーサビリティの確保，作成プロセスの自動化に課題がある。また，Analysis Results Metadata には解析手法や結果の表示に関する情報が含まれるものの，レイアウト等の情報が限定的であり，自動化に必要なメタデータとしては不十分である。ARS を利用して解析結果のメタデータを事前に定義することにより，解析の標準化，一貫性およびトレーサビリティの向上が期待される。

ARS v1.0 として，Analysis Results Standard User Guide（以下，ARS-UG），Analysis Results Standard Model Documentation，Reporting Events and Supporting Files が公開されている。このうち ARS-UG は ARS の実装ガイドであり，ARS の目的や使い方の事例も含めた実装方法の詳細が記載されている。

2025 年度 CJUG ADaM テーマ 2 では，ARS-UG の内容を基に ARS の各種構成要素について解説資料を作成した。あわせて，LinkML や YAML といった ARS-UG で用いられている理論モデルや記述言語をはじめとした用語の解説，ならびに ARD generator や eTFL portal といった関連情報を整理した。また，公開されているプログラムおよび資料を用いて実際に ARS を作成し，作成時の課題を検討するとともに，利用可能な R パッケージおよび参考となるツールについて調査した。さらに，ARS-UG の解説を通して，ARS 利用におけるメリットおよびデメリットについても考察した。

本発表では，2025年度CJUG ADaMテーマ2で作成したARS-UGの解説資料を基に，ARS開発の背景，ARSの構成概要，ならびにARSで利用されている理論モデルであるLinkMLについて解説する。さらに，本テーマで検討したARSの作成フローを紹介するとともに，実際に作成することで得られた課題やARS作成時の参考情報についても紹介する。あわせて，ARS利用のメリットおよびデメリットに関する考察内容を共有する。

生成AIは統計プログラマーの業務をどこまで支援できるか

～CJUG-ADaMテーマ５での2025年の試行錯誤から～

○栗栖 章
（MSD株式会社）

近年、生成AIの急速な発展により、統計解析業務への活用可能性が注目されている。しかし、実務において生成AIがどの程度有効であり、どのような課題が存在するかについては十分な知見が蓄積されていない。

CJUG-ADaMテーマ5では、製薬企業および関連企業のメンバーが参加し、「生成AIを活用した業務効率化」をテーマとして1年間の共同検証を実施した。活動では、解析結果帳票やe-Submission文書の整合性確認などのQC作業、統計解析計画書（SAP）やADaM仕様書等のドキュメント作成支援、SASおよびRプログラム作成支援、論文調査支援など複数のユースケースについて検討を行った。

検証の結果、生成AIは情報整理やドキュメントのドラフト作成、プログラムコードの生成やレビュー支援において有用であり、初期作業の効率化に寄与することが示唆された。一方で、出力の再現性の不安定さやハルシネーション、複雑な整合性確認における限界も確認された。そのため、生成AIの出力をそのまま採用するのではなく、利用者によるレビューと妥当性確認が不可欠であることが示された。

本発表では、CJUG-ADaMにおける実践的な検証事例を紹介するとともに、統計プログラマーが生成AIを活用する際の有効な活用方法および留意点について報告する。

生成AI支援による統計解析実務の有用性と課題

中島誉也，佐藤俊太朗(長崎大学

ある臨床研究の統計解析において、独立した解析担当者2名で解析用データセットを作成し，一人が統計解析報告書を作成した。解析用データセット作成はダブルプログラミングでおこない、解析主担当者（生物統計家16年目、R使用歴20年）が人力、解析副担当者（生物統計家1年目、R使用歴7年、発表者）が生成AIを用いた。個人情報保護のため生データは外部生成AIへ入力せず、データセット仕様書と統計解析計画書を参照させ、データセット作成・解析コードを作成した。仕様が明確な箇所では、生成AIは忠実で、ケアレスミスは少なく、人力との一致度も高かった。一方、仕様書の軽微な誤りや生データの不整合にも従うため不一致が生じた。人力でコードを作成した解析者は原因を迅速に特定できたが、生成AIが作成したコードを十分に把握していない副担当者では修正に時間を要した。解析コードには、解析主担当者から誤りは指摘されなかった。本事例から、生成AI活用の成否は、生データのクリーニング、仕様書・SAPの完成度、利用者のコード理解と検証体制に依存すると考えられた。今後はローカルLLMでの実装や、サンプルデータによるコード開発を検討する。

九州大学病院における統計解析業務の体制と今後の展望

○廣瀬 将行、田川 公士朗、岸本 淳司、田中 司朗

（九州大学病院 ARO次世代医療センター）

【発表要旨】

本発表では、当センターにおける統計解析業務について、特に医師主導治験の支援に関する業務フローを紹介する。また現在直面している課題に触れ、それらの解決に向けた方向性の展望を述べたい。

【発表概略】

九州大学病院 ARO次世代医療センター（以下、当センター）に初めて生物統計家が着任して20年ほどが経ち、現在では学内で実施される臨床研究の研究デザインおよび統計解析に関する支援体制は十分に整備されてきたと考えられる。一方で、慢性的な人員不足や、臨床研究に必要とされる研究方法論・統計学的方法論の高度化、組織の効率的な運営への要求など、難しい課題も多くあるのが現状である。今後は現在までに構築してきた業務基盤をもとに安定した業務を行いつつ、同時に新たな課題への対応も求められる。

本発表では最初に、当センターが主に支援の対象とする「医師主導治験」における、研究デザイン・統計解析に関する支援の現状を紹介する。続いて、現在当センターが抱えている課題に触れる。業務量が近年増加傾向にあるため、課題解決に向けて十分に時間と労力を割けていないのが現状であるが、生成AIの有効活用を軸にした業務効率化・品質向上を重要テーマの一つとして位置づけ、具体的な実装方法を模索している。最後にこれらを踏まえて今後の展望について整理する。当日は発表に加え、可能であれば来場者の方々との意見交換も活発に行いたい。

【ページ設定】

用紙：　A4・縦（1ページ）

余白：　上/下/左/右　25mm/25mm/20mm/20mm

行送り：　36行、18pt送り

文字間隔：　標準

■フォント

日本語：　MS明朝

English：　Times New Roman

数字：　Times New Roman (0123456789)

■フォントサイズ・配置

演題名：　16ポイント・中央揃え

演者名：　11ポイント・中央揃え

※発表する演者の氏名の前に○をつける

※演者の所属が複数になる場合、各演者の氏名の後に所属を表す上付き数字を記載

演者所属名：　11ポイント・中央揃え　※全体をカッコで囲う

抄録本文：　10ポイント、両端揃え

■提出ファイル形式

・Microsoft Office Word 2013 / 2016 / 2019 / 2021

・PDF

提出する際は、こちらのテンプレートを使用して作成した Word データと、保存形式を PDF にして保存したファイル両方（2 ファイル）をご提出ください。
抄録は 1 ページ以内で作成してください。
ご不明点等ございましたら、ユーザー総会事務局までお問合せください。

【お問合せ先】

SASユーザー総会事務局

E-mail: info@sas-user2026.ywstat.jp

SAS-MCP-Server：AI エージェントと SAS Viya の連携による分析の民主化

○長縄 俊生
（SAS Institute Japan 株式会社）

近年、生成 AI は、質問応答を中心とするチャットツールから、複数のシステムやツールを活用し、計画立案、情報取得、処理実行、結果整理までを担う「AI エージェント」へと発展している。解析業務においても、自然言語による指示を起点に、解析内容の整理、プログラム作成、実行、結果解釈までを連続的に支援する活用が期待される。特に、SAS が広く活用されているライフサイエンス領域、なかでも臨床試験に関わる解析業務では、規制対応、標準データ構造、解析手順の再現性が重要であり、生成 AI と決定論的な分析基盤の連携が有効なテーマとなる。一方で、生成 AI は確率的に応答を生成するため、精度、再現性、解析手順の一貫性が求められる業務領域では、SAS Viya のような決定論的な分析基盤との連携が重要となる。

本発表では、臨床試験の解析業務を想定し、AI エージェントと SAS Viya を MCP（Model Context Protocol）により接続する「SAS-MCP-Server」を題材とし、自然言語を起点として解析業務を遂行する新たな作業プロセスをデモンストレーションにより示す。まず、解析に関する章を含むプロトコルを AI エージェントに読み込ませ、試験目的、評価項目、解析対象、解析方法等を抽出し、実施すべき解析内容を整理する。次に、整理された解析要件と過去の類似解析で使用された SAS コードを参照し、解析計画書（SAP）のドラフトを作成する。さらに、SAP に基づき、必要な Tables、Listings、Figures を想定した Mock TLFs を作成し、解析実行前に出力内容と解析要件の整合性を確認する。

続いて、解析に必要な変数やデータ構造を確認し、SDTM や ADaM 等の標準的なデータ構造を活用しながら、解析に適した形式へデータを成形する。SAP、Mock TLFs、データ構造、過去の SAS コードを基に SAS プログラムを生成し、SAS-MCP-Server を介して SAS Viya 上で実行する。得られたログ、表、リスト、グラフを確認し、AI エージェントとの対話を通じて、結果の要点および追加で確認すべき事項を整理する。

本アプローチにおいて、AI エージェントは解析者の判断を代替するものではなく、文書読解、要件整理、既存資産の検索、ドラフトおよびプログラム生成を支援する役割を担う。解析者は、データの理解、仮説立案、解析計画の妥当性確認、生成コードの検証、結果の統計的・業務的考察に注力する。SAS-MCP-Server により解析工程と既存資産を接続することで、業務効率化、品質の均一化、解析ノウハウの再利用、属人化の軽減、技術的障壁の低減を実現し、「分析の民主化」に寄与する可能性を示す。

TRIにおける統計解析業務の体制と効率化への取り組み

○山川 茜、加藤 直樹、岡田 悠、岡本 里香、川本 篤彦
（神戸医療産業都市推進機構　医療イノベーション推進センター）

【発表要旨】

医療イノベーション推進センターにおける、生物統計チームと統計解析プログラミングチームの連携による統計解析業務フローの詳細、実務の効率化のための取り組み、LLM の活用事例や今後の展望について紹介する。

【発表概略】

当機関は、アカデミアにおける初のデータセンター・解析センターとして 2003 年に神戸市により創設された。2018 年以降、自機関内に病院機能を有さない ARO となり、神戸医療産業都市内はじめ国内外機関の医療研究開発を支援している。

統計解析業務は、生物統計と統計解析プログラミングの 2 チームが連携し、生物統計チームでは統計解析計画書(SAP)及び図表見本案の作成、統計解析プログラミングチームでは解析データセットの仕様書や解析プログラム、帳票作成を行っている。

原則として解析用データセットを ADaM-like な構造で作成することにより、異なる試験間でもデータ構造を理解しやすく、解析プログラムの共通化も容易となり、業務効率化に繋がっている。

2026 年度より、全職員の PC に Copilot Studio が導入され、自機関の環境下における LLM の活用が可能となり、統計解析業務の効率化が期待されている。

本発表では、当機関における統計解析業務フローの詳細や、LLM の具体的な活用事例等、組織での取り組みについて紹介する。

AROにおける特定臨床研究解析の事例紹介（神戸大学）

○今井匠[1]

（[1]神戸大学医学部附属病院臨床研究推進センター）

臨床研究における統計解析では、解析結果だけでなく、データ処理から解析、結果の報告に至る過程を再現可能かつ透明な形で管理することが重要である。神戸大学AROでは、観察研究から治験まで、さらに論文化やリバイスへの対応まで、多様な研究ニーズに柔軟に応えることを目指している。解析結果の修正やデータの差し替えなどが実務上起こり得ることを前提とすると、個々の作業を属人的な手順に依存させず、再現可能な形で解析業務を設計することが重要となる。近年、医学研究における再現性向上のため、データクリーニングや解析上の判断を含む研究プロセスをコードと一体的に記録することが提案されており、R Markdownのようなexecutable documentもその手段の一つとして紹介されている[1]。また、Clinical Study Reportにおける表・図・listingについては、成果物の量ではなく、研究結果を適切に伝えるという観点から、その内容を設計することの重要性が議論されている[2]。

本発表では、神戸大学AROにおける特定臨床研究の統計解析業務を例に、データの受領から解析、結果の検証、報告書作成に至るまでの実務上の工夫を紹介する。解析から成果物の作成までを一貫したプロセスとして設計し、手作業による転記や個別の加工を抑えることで、データから解析結果、成果物に至るまでの対応を追跡できるようにしている。解析結果の検証についても、可能な範囲でプログラムによる検証を行うことで、確認作業の再現性を高めている。また、修正や再解析、データの差し替えが生じた場合にも、必要な変更を反映したうえで、担当者の変更や時間の経過に左右されることなく、同一のプロセスから報告書を速やかに再生成できることを重視している。さらに、AROにおける統計解析では、解析結果を報告して業務を完了するのではなく、その後の研究者による結果の解釈や論文化までを見据えて成果物を作成することが重要である。そのため、報告書では研究者による解釈・論文化に必要な情報を中心に結果を提示し、表のFootnoteや図のlegendについても、報告書の段階から論文でそのまま利用できる形に整えるなど、解析結果の確認から論文化までを円滑につなぐ工夫を行っている。

本発表では、こうした実務上の工夫を具体的な事例とともに紹介し、アカデミアにおける統計解析業務を進める上で、どのような作業思想が重要となり得るのかを検討・共有する。

1. Streiber AM, Hoepel SJW, Blok E, van Rooij FJA, Neitzel J, Labrecque J, et al. Improving reproducibility of data analysis and code in medical research: 5 recommendations to get started. BMJ Open. 2025;15(10):e104691.
2. Li N, Zhang Y, Ting N. Tables, listings and figures in a clinical study report - quality or quantity? J Biopharm Stat. 2025 online ahead of print.

SAS Proc FCMPによる行列演算処理を用いた野球の打撃力評価モデルOERAの適用事例…対戦チームを考慮したOERAスコアの考察

○荒井遠樹、小池沙織、岡野かほる、榎統那、北村凌士、山下敏治
（株式会社セルネサイエンス）

はじめに

OERA(Offensive Earned-Run Average)は野球における打者の打撃力を評価するために考案されたマルコフ連鎖に基づく確率過程モデルである。ノーアウトランナー無からツーアウト満塁までの24の状態と打者の打撃成績、及びOERAの慣習に基づく得点行列から推移確率行列を構成し、１試合当たりの期待得点値を計算する。OERAの算出には行列演算が必要であり、SASで行列演算を行う方法としてSAS/IMLがあるが利用にはライセンスが必要となる。一方、SAS/IMLを使用しない方法として、Proc FCMP を用いた行列演算も可能である。また対戦毎にOERAモデルを適用することで対戦チームとの打撃力評価が可能である。本発表では公開されている2026年のNPB打撃データの一部を用いてProc FCMPによりOERAを算出し、その計算過程及び算出結果を示すとともに、得られた結果について考察する。

本論

以下の流れで発表する。(1) OERAモデルの概略、Proc FCMP、及び利用したNPB打撃データの説明。(2) Proc FCMPを使ったOERA計算過程の説明。(3) 算出したOERAの結果の説明。(4) スポーツデータサイエンス及びプログラムコーディングの観点からのまとめと考察、今後の展望。なお、Proc FCMPで用いる行列操作ルーチン、及びOERA計算に用いた打撃データ配置、実際の対戦の打撃データ例、算出されたOERAを下記に示す。

Proc FCMP: zeromatrix, identity, substractmatrix, inv, mult

	打撃データ							打撃データ（率）						OERA
	打数	単打	二塁打	三塁打	本塁打	四死球	打点	凡打	単打	二塁打	三塁打	本塁打	四死球	
チーム_i	ni	ni1	ni2	ni3	ni4	nib		Pi0	Pi1	Pi2	Pi3	Pi4	Pib	OERAi
チーム_j	nj	nj1	nj2	nj3	nj4	njb		Pj0	Pj1	Pj2	Pj3	Pj4	Pjb	OERAj
A	36	5	1		3	6	6	0.643	0.119	0.024	0.000	0.071	0.143	6.928
B	34	9	1	1	1	6	7	0.550	0.225	0.025	0.025	0.025	0.150	10.103

参考文献

1) T.M. Cover, C.W. Keilers, “An Offensive Earned-Run Average for Baseball.” Opns. Res. 25, 729-740(1977)
2) 木下栄蔵「野球における打者・投手の評価」日本オペレーションズリサーチ学会(1987年10月号)
3) 森岡裕「IMLライセンスを利用せずにProc FCMPで基本的な行列計算を行う方法について」SASユーザー総会2025論文集

統計解析プロセス効率化のその先にあるもの

○魚住 龍史[1,2]
[1] 東京科学大学工学院経営工学系
[2] Total Health Design Visionary Initiative, Institute of Science Tokyo
Email: uozumi.r.ab@m.titech.ac.jp

臨床研究のデータ解析における統計解析プロセスの効率化として，生成AIの利活用や解析プログラムや表の標準化が挙げられるのではないだろうか．しかしながら，演者はこれまで個人で実施する統計解析プロセスを簡略化できるよう，被験者背景のような頻用される表の計算プログラムのマクロを作成しているものの，組織として標準化を実施した経験がない．本企画セッション『アカデミアでの統計解析業務はどう進めるべきか ―ガラパゴス化を越えて，実務知を共有する―』では，ARO (Academic Research Organization) における統計解析プロセスが主題となるが，果たしてAROにおいて標準化などが必要だろうか．

本講演では，企画セッションで実務知を共有することを目標に，セッションのオープニングにて演者作成のGoogle Formでアンケートを収集し，アンケート結果を聴衆に共有することで，AROにおける統計解析プロセスの効率化の動機付けについて議論する．演者は，これまで国内製薬企業と大学病院における統計解析に従事してきた．国内製薬企業と大学病院での生物統計家の役割の違いについても，演者の個人的な見解を踏まえて，AROにおける生物統計家の役割を議論する．

以下，事後的な共有であるが，37名にアンケートに回答いただいた．オープニングにてアンケート収集したため，参加者の半分以下のデータしか集まらなかったことが残念である．内訳として，ARO等アカデミア勤務 (n = 15)，国内製薬企業 (n = 12)，CRO勤務 (n = 10) であった．勤務歴については大きな偏りはなかった．演者が最も関心のあった質問項目として，統計解析プロセスの効率化の動機であるが，最も回答者が多かったのは「仕事する時間を減らしたいから」であった．さらに，最も時間を費やしているタスクとして，臨床研究の統計解析と回答した者が23名と最上位であり，本セッションの参加者が統計解析に四苦八苦していることがわかる．次いで，臨床研究のデザインに時間を費やすいている者 (主に，AROあるいは国内製薬企業勤務) が多い結果であった．一方，生物統計の研究・勉強というタスクも選択肢に設けていたものの，ARO勤務かつ勤務歴が短い者であっても，本タスクに時間をあまり使えていない結果が共有され，生物統計の研究・勉強の時間を確保できるような職場環境を整えることが求められる．

本セッションに参加したARO所属の生物統計家には，本講演を含めたセッション内の各演題内容を持ち帰り，自身の所属する組織の常識が業界の常識と齟齬がないか確認すべきである．そして，必要に応じてSOP (Standard Operating Procedure) 改訂などに尽力することが，最適な統計解析プロセス効率化になるのではないだろうか．

主催：SAS ユーザー会 世話人会

代表世話人 伊藤 陽一 北海道大学病院
世話人(50 音順)
井上 貴博 協和キリン株式会社
上村 鋼平 東京大学
岸本 淳司 九州大学 ARO 次世代医療センター
菅波 秀規 興和株式会社
土生 敏明 SAS Institute Japan 株式会社
林 行和 エイツーヘルスケア株式会社
森岡 裕 イーピーエス株式会社

企画委員(50 音順)
佐藤 俊太朗 長崎大学病院
徳田 芳稀 エイツーヘルスケア株式会社
本田 主税 小野薬品工業株式会社

SAS ユーザー総会 2026 についての問い合わせ先：info@sas-user2026.ywstat.jp

協賛(50 音順)

イーピーエス株式会社
株式会社インテリム
エイツーヘルスケア株式会社株式会社
エス・シー・エー
SAS Institute Japan 株式会社
コムチュア株式会社
シミック株式会社
スタットコム株式会社
株式会社タクミインフォメーションテクノロジー
株式会社ピープロ・ジャパン