



臨床試験における estimand と感度分析 に関する文献調査

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

2018 年度 タスクフォース 4

Estimand／因果推論チーム

2019 年 5 月

目次

1. はじめに.....	4
1.1 背景と目的.....	4
1.2 文献調査方法の概要.....	4
1.3 レポートの構成.....	5
2. レポートの要旨.....	6
3. 文献調査結果のまとめと考察.....	8
3.1 ITTの原則に基づく治療方針の効果とその他の治療効果.....	8
3.1.1 調査対象詳細.....	8
3.1.2 文献調査方法の詳細.....	8
3.1.3 文献調査結果.....	10
3.1.4 文献調査結果からの考察.....	21
3.1.5 参考文献.....	26
3.2 データの取扱い及び欠測データに含めて検討されている問題.....	27
3.2.1 調査対象詳細.....	27
3.2.2 文献調査方法の詳細.....	29
3.2.3 文献調査結果.....	30
3.2.4 文献調査結果からの考察.....	37
3.2.5 参考文献.....	41
3.3 解析対象集団の定義と役割.....	42
3.3.1 調査対象詳細.....	42
3.3.2 文献調査方法の詳細.....	44
3.3.3 文献調査結果.....	45
3.3.4 文献調査結果からの考察.....	48
3.3.5 参考文献.....	49
3.4 安定性と感度分析.....	50
3.4.1 調査対象詳細.....	50
3.4.2 文献調査方法の詳細.....	52
3.4.3 文献調査結果.....	54
3.4.4 文献調査結果からの考察.....	61
3.4.5 参考文献.....	65
4. 全体を通しての考察とまとめ.....	67
4.1.1 参考文献.....	70
5. 資料作成者.....	72
6. 付録 調査対象文献一覧.....	73

略語一覽

AD	Alzheimer's Disease
AE	Adverse Event
AT	As Treated
ANCOVA	Analysis of Covariance
BE	Bioequivalence
CACE	Complier Average Causal Effect
CHMP	Committee for Medicinal Products for Human Use
CPMP	Committee for Proprietary Medicinal Products
COPD	Chronic Obstructive Pulmonary Disease
EMA	European Medicines Agency
EMEA	European Medicines Evaluation Agency
EWG	Expert Working Group
FAS	Full Analysis Set
FEV ₁	Forced Expiratory Volume in One Second
FDA	Food and Drug Administration
GFR	Glomerular Filtration Rate
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IP	Inverse Probability
ITT	Intention-to-Treat
IV	Instrumental Variable
JPMA	Japan Pharmaceutical Manufacturers Association
MAR	Missing at Random
MCAR	Missing Completely at Random
MCI	Mild Cognitive Impairment
MHLW	Ministry of Health, Labour and Welfare
MMRM	Mixed Models for Repeated Measures
MNAR	Missing not at Random
NRC	National Research Council
PP	Per-Protocol
PPS	Per Protocol Set
QOL	Quality of Life
SACE	Survivor Average Causal Effect
TIRS	Therapeutic Innovation & Regulatory Science

1. はじめに

1.1 背景と目的

2014年10月に final concept paper が承認された ICH E9 (R1)は、1) estimand の定義を明確にし、適切に選択するためのガイダンスを提供すること、及び2) 感度分析の計画・実施・解釈のフレームワークに関するガイダンスを提供することの2点を大きな柱として活動が開始された。その後、2017年6月に Step2 文書が合意され、日本では2017年9月～2018年3月に Step 2 文書に対する意見募集が行われたところである。また、現在の予定では2019年6月に Step 4 に到達することが期待されている(2019年4月時点)。

一方、医薬品評価委員会 データサイエンス部会では、2013年度以降、欠測データ及び estimand に関する検討を継続して行ってきた。そこで、タスクフォース4 Estimand/因果推論チームでは、ICH E9 (R1)の発出が間近に迫っていることを踏まえて、ガイドラインの記載内容に対する理解を深め、臨床試験をより適切に計画・実施するために、ICH E9 (R1) step2 文書で示されている以下の4つの論点について、文献調査を通じて整理・考察することとした。

1. ITT の原則に基づく治療方針の効果とその他の治療効果
2. データの取扱い及び欠測データに含めて検討されている問題
3. 解析対象集団の定義と役割
4. 安定性と感度分析

調査対象として設定したこれらの内容は、ICH E9 (R1)の検討対象である estimand と感度分析に強く関連する事項であるとともに、ICH E9 (R1) step 2 文書において、ICH E9 の内容を「明確にし、拡張する」対象として記載されているものである。また、ICH E9 (R1)は1998年に発出された ICH E9 の補遺との位置付けである。ここで、ICH の SOP "Standard Operating Procedure of the ICH Working Groups" (Version 7.0, 2018年11月13日)において、「no amendments to the content of the existing Guideline are necessary but there is a need to provide further complementary guidance (既存のガイドラインの内容の修正はないが、補足的なガイダンスを提供する必要がある)」の場合に補遺を作成すると定められており、ICH E9 (R1)は ICH E9 で記載されている原則を変更するものではないことから、ICH E9 において解釈の余地がある、又はより深い理解が必要とされると考えられるこれら4つの事項に対して文献調査を実施し、調査結果をまとめたうえで考察を行うことにより、統計担当者のみならず、医薬品開発に関わる担当者が ICH E9 (R1)に対する理解をさらに深めることに役立つことができると考え、これらの調査対象を設定した。

1.2 文献調査方法の概要

文献調査の対象とする文献は、原則として ICH E9 が発出された1998年以降に掲載又は公開された英語又は日本語で記載されている文献とし、英語文献は PubMed 及び TIRS の web ページ、日本語文献は医中誌 Web にて検索を行った。その際、検索時のキーワード等

の設定に関しては、各文献調査の対象項目に対して設定する調査事項詳細に基づいてそれぞれ決定した(文献調査方法の詳細については「3. 文献調査結果のまとめと考察」を参照)。また、上記の検索にて特定された文献に加え、検索結果には含まれない重要な文献、書籍、ガイドライン等についても、各調査担当者の判断にて調査対象文献に含めることとした。

1.3 レポートの構成

本レポートの構成は以下の通りである。

まず2章において本文献調査の内容と結果の要約を記載した。

3章では、文献調査を実施した4つの調査対象について、それぞれ3.1節～3.4節において文献調査結果のまとめと考察を記載している。これには、各調査対象項目に対して設定した調査事項詳細に関する説明及び検索キーワード等の文献調査方法の詳細に関する記載も含まれる。また3章で文献からの引用を行う際には、原則として英文の原文をそのまま記載するとともに、読者の理解を助けるため、英文の直後に日本語での要約も付与した。

4章では、文献調査全体を通した結果のまとめと考察を記載し、5章には資料作成担当者の一覧を示した。また本レポートの付録として、各文献調査において調査対象とした文献の一覧を掲載した。

なお、調査対象文献には含まれないものの、3章において調査事項詳細に関する説明等で使用した参考文献については3.X.5節に、また4章の参考文献は4.1.1節に記載した。

2. レポートの要旨

本レポートでは、文献調査を通じて大きく以下の4つの論点について整理・考察した。

1. ITT の原則に基づく治療方針の効果とその他の治療効果
2. データの取扱い及び欠測データに含めて検討されている問題
3. 解析対象集団の定義と役割
4. 安定性と感度分析

「1. ITT の原則に基づく治療方針の効果とその他の治療効果」では、ICH E9 及びそれ以前のガイドラインにおける推定したい治療効果，並びにこれまでに考えられてきた推定したい治療効果をより詳細な論点と設定し，整理・考察を行った。調査の結果，ICH E9 及びそれ以前のガイドラインからは推定したい治療効果を明確に読み取ることが困難であった。また，2010 年以前は各種文献において *effectiveness* や *efficacy* として表現されることが多かった推定したい治療効果について，2010 年以降は *estimand* という概念を用いてより詳細に表現されるようになったものの，提案された推定したい治療効果については年代によらずにいずれも利点や欠点が指摘されており，直面している状況に応じて推定したい治療効果を適切に設定することの重要性が改めて浮き彫りとなった。調査結果を踏まえて，結果の一般化可能性と比較可能性，推定のためのデータの利用可能性，比較の形式(優越性又は非劣性/同等性)，臨床試験終了後の意思決定といった点が，今後 *estimand* を選択する際に留意すべき事項になると考えた。

「2. データの取扱い及び欠測データに含めて検討されている問題」では，欠測データの定義，欠測データを最小にするための方法論と適用時の注意点，及び中間事象発現後のデータの取り扱い方法をより詳細な論点と設定し，整理・考察を行った。調査の結果，死亡後の QOL データ等の「存在しないデータ」を欠測データとみなして仮想的な状況を想定することの問題点とともに，*estimand* を検討する際に「存在しないデータ」への対応方法を注意深く検討することの重要性が明らかになった。また欠測データを最小にするために文献上で提案されている種々の方策については，その内容に応じて適用時に留意すべき事項もあるものの，いくつかの方策についてはより積極的に検討・活用すべきと考えられた。さらに，中間事象発現後のデータの取り扱い方法について検討した文献が存在し，解析方法を検討する際には中間事象の発現状況を適切に考慮することが必要である一方で，適用する手法の複雑さや必要とされる仮定の強さについても今後検討することが重要になると思われた。

「3. 解析対象集団の定義と役割」では，これまでに利用されてきた各解析対象集団の役割，並びに治験実施計画書に適合した対象集団(*Per Protocol Set: PPS*)の問題点及びその対処方法をより詳細な論点と設定し，整理・考察を行った。調査の結果，*PPS* を利用した解析や *as treated* 解析についてはランダム化された被験者を除外することに起因するバイアスが問題点として広く指摘されており，またこの問題点を解決するために主要層解析や *trimmed mean* を利用する解析等が提案されていた。しかしながら，これらの解析手法は解析の目的や必要な仮定がそれぞれ異なる場合があるため，試験の目的やデザイン，設定する *estimand*

に応じて適切な手法を選択することが重要であると考える。

「4. 安定性と感度分析」では、適切な安定性評価の考え方、安定性評価の解釈の方法、及び提案されている感度分析の手法をより詳細な論点と設定し、整理・考察を行った。調査の結果、安定性評価を行うために実施する解析の範囲は必ずしも主要な解析の仮定や結果に限定すべきではないものの、単に異なる解析を感度分析として実施するのではなく、必要な解析については注意深く検討することの必要性が改めて浮き彫りとなった。また安定性評価と意思決定の関連について議論している文献は今回の調査では見つからなかった、これは望ましい意思決定に必要とされる結論の度合いが状況に応じて異なるためと考えられたものの、感度分析の結果とその後の意思決定の問題は **estimand** の選択にも影響を与えうるものであることから、今後のさらなる議論が待たれる。さらに、調査した文献においては多数の感度分析の手法が提案されていたが、感度分析において対処すべき課題は対象とする疾患領域や治療の特性、試験デザイン等を踏まえて注意深く分析する必要があると思われた。

最後に本文献調査全体のまとめとして、**estimand** の選択に際しては、推定したい治療効果と欠測データ、及び解析対象集団の選択の3つの問題を混同せず、それぞれの関係を整理して検討することが特に重要になると考えられた。

3. 文献調査結果のまとめと考察

3.1 ITT の原則に基づく治療方針の効果とその他の治療効果

3.1.1 調査対象詳細

ICH E9 (R1) step 2 文書では、下記のように、治療方針の効果だけでなく、様々な治療効果を議論するための基盤として、計画、デザイン、実施、解析及び解釈を整合させるフレームワーク及び estimand という概念が提案された。

第一に、ICH E9 は、治療方針の効果、すなわち予定した試験治療のコースの遵守状況に関わらない、ベースライン時に最初に割付けられた治療の効果と関連して、intention-to-treat (ITT)の原則を導入し、ランダム化の維持が統計的検定の確固とした基盤を与えることを示した。

ランダム化は比較試験の基礎であり、解析はランダム化の利点を最大限に活用することを目指すべきであることについては、議論の余地がない。

しかし、治療方針の効果を理解することが常に規制上及び臨床上の意志決定に最も関連する治療効果を目的としているものなのかという疑問が残る。

本補遺で概説するフレームワークは、その他の治療効果、及び意志決定のために信頼できる治療効果の推定値を与えるための臨床試験のデザイン及び解析の留意点について、議論するための基盤を与える。

上記で、疑問として取り上げられているが、どのような場合に、治療方針の効果の理解が規制上及び臨床上の意志決定に最も関連する治療効果と紐づかなくなるのか？ ICH E9 で ITT の原則が導入された当時、治療方針の効果を含めてどのような治療効果を推定しようと考えられていたのだろうか？また、それらの考えはどのように変遷し、今回の ICH E9 (R1) step 2 文書につながってきたのだろうか？本節では、ITT の原則を中心とした文献調査を通じて以下を特定し、整理する。

- ICH E9及びそれ以前のガイドラインにおける推定したい治療効果
- これまでに考えられてきた推定したい治療効果(治療方針の効果の再考も含む)

3.1.2 文献調査方法の詳細

本文献調査では、2018年8月に、PubMed, TIRS 及び医中誌 Web の3つの文献検索サイトを利用し、設定した調査対象に関連する記載があることが期待される文献を検索した。なお、本調査項目は ICH E9 以前の議論も重要と考え、1998年以前の文献も調査対象としている。

文献検索時のキーワード

- PubMed (検索式)
Search 1: Intention to treat [Title] OR Intent to treat [Title] OR ITT [Title]
Search 2: Estimand OR Estimands
- TIRS
Search 3: タイトルに”Intention to treat”, ”Intent to treat”, ”ITT”のいずれかを含む論文

Search 4: "Estimand", "Estimands"のいずれかを含む論文

- 医中誌 (検索式)¹

Search 5: "Intention to treat"/TA or "Intent to treat"/TA or "ITT"/TA

Search 6: "Estimand"/AL or "Estimands"/AL

また、ICH E9の発出以前のガイドラインも、同様に重要と考えたため、ICH E9が発出された1998年以前ではあるが、下記のガイドラインについても調査対象とした。

- Lewis et al. (1995) Biostatistical methodology in clinical trials - a European guideline, *Statistics in Medicine*, 14, 1655-7.
- CPMP Working Party on Efficacy of Medicinal Products. (1995) Biostatistical methodology in clinical trials in applications for marketing authorizations for medicinal products, *Statistics in Medicine*, 14, 1659-1682.
- Food and Drug Administration. (1988) Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications, FDA, U.S. Department of Health and Human Services, Rockville, MA, U.S.A.
- Ministry of Health and Welfare. (1992) Guideline for the Statistical analysis of clinical trials, MHLW pharmaceutical Affairs Bureau, Tokyo, Japan (in Japanese)

さらに、臨床試験において推定したい治療効果に関して記載され始めていることから、過去1年間に発出された下記European Medicines Agency (EMA)の臨床開発ガイドラインについても調査対象とした。なお、文献調査と同様に、2018年8月までを対象とした。

- EMA (2018) Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease
- EMA (2018) Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus[draft]
- EMA (2018) Guideline on clinical evaluation of vaccines[draft]
- EMA (2017) Guideline on the clinical investigation of medicines for the treatment of chronic heart failure
- EMA (2017) Guideline on the evaluation of anticancer medicinal products in man
- EMA (2017) Guideline on the clinical development of medicinal products for the treatment of Autism Spectrum Disorder (ASD)
- EMA (2017) Guideline on the Clinical Investigation of Medicinal Products for the Treatment of Axial Spondyloarthritis

¹ TA:タイトル+抄録, AL: All Fields

- EMA (2017) Guideline on the clinical investigation of recombinant and human plasma-derived factor VIII products [draft]
- EMA (2017) Guideline on the clinical evaluation of medicinal products indicated for the prophylaxis or treatment of respiratory syncytial virus (RSV) disease [draft]
- EMA (2017) Guideline on clinical investigation of medicinal products for the treatment of rheumatoid arthritis

最後に、解析対象集団に関する問題を取り上げている下記の和書にて引用されている論文・文献についても、調査対象とした。

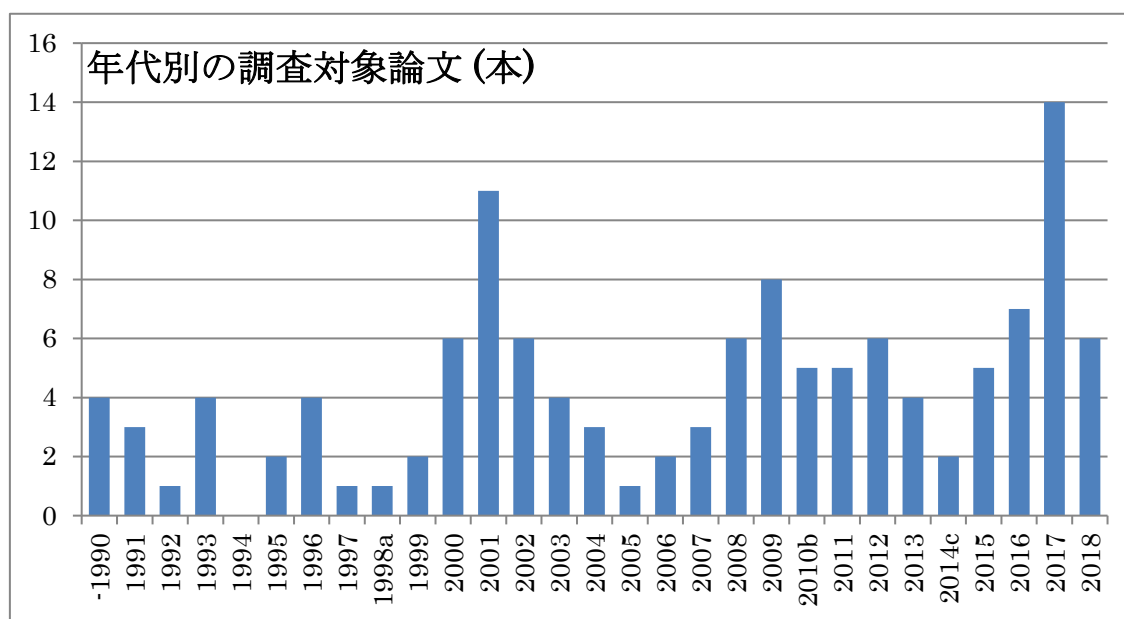
- 佐藤・上坂 (2006) ICH E9ガイドライン 「臨床試験のための統計的原則」の理解を助ける文献リスト[丹後・上坂編 (2006) 臨床試験ハンドブックより]

上記の方法により特定された文献[合計:610本(PubMed: 493本, TIRS: 32本, 医中誌: 23本, ガイドライン: 15本, ハンドブック引用: 47本), ただし重複あり]について, 本文献調査の担当でタイトル及び抄録を確認し, 「調査対象の文献」または「論文本体の内容を確認する必要がある文献」となった138本の文献(PubMed:93本, TIRS: 22本, 医中誌: 0本, ガイドライン: 15本, ハンドブック引用: 8本)を調査対象文献(文献一覧は6.付録(1)参照)とした。なお, 調査対象文献に対するコメント論文で重要と判断した3本も調査対象に追加した。

3.1.3 文献調査結果

3.1.3.1 調査対象とした文献の内訳

調査対象の文献が発表された年代の頻度集計の結果を示す(ガイドライン15本は除く)。ICH E9が発表された1998年直後, NRC (2010)の前後, また, ICH E9 (R1) final concept paper (2014)からICH E9 (R1) step 2文書の発出(2017)に合わせて, 増加している傾向が認められた。また, 特にITTの原則に関しては, ICH E9以前も一定の議論がなされていたことが見受けられる。



a: ICH E9 (1998), b: NRC (2010), c: ICH E9 (R1) final concept paper (2014)

上記のように、推定したい治療効果の議論は、NRC (2010)、ICH E9 (R1) final concept paper (2014)を踏まえて加速していることから、調査対象文献の年代を2010年以前と2010年以降に分けて調査結果を整理した。まず、出発地点として、ICH E9及びそれ以前のガイドラインにおけるITTの原則及び推定したい治療効果について概観する(3.1.3.2項)。次に、ITTの原則の利点および欠点/注意すべき点を整理する(3.1.3.3項)。その後、NRC (2010)以前の推定したい治療効果とNRC (2010)以降の推定したい治療効果について言及する(それぞれ、3.1.3.4項、3.1.3.5項)。

3.1.3.2 ICH E9 及びそれ以前のガイドラインにおける ITT の原則と推定したい治療効果

3.1.3.2.1 ICH E9

まず、ICH E9 における ITT の原則と推定したい治療効果に関する調査結果を示す。

Intention-To-Treat の原則は、ICH E9 の用語集にて下記のように示された。

Intention-To-Treat の原則 Intention-To-Treat Principle

治療に用いる治療方針により得られる効果は、実際に受けた試験治療ではなく、被験者を治療しようとする意図 (予定した試験治療規定)に基づくことにより最もよく評価できる、ということを主張する原則。この原則から、一つの試験治療グループに割付けられた被験者は、予定した試験治療のコースを遵守したかどうかにかかわらず、割付けられたグループのまま追跡され、評価され、解析されるべきであることが導かれる。

一方、ランダム化が行われた全被験者を完全に追跡することが難しいことを踏まえ、intent-to-treat という理想に可能な限り完全に近づけた解析対象集団として「最大の解析対象集団 (Full Analysis Set: FAS)」が定義された。また、偏りを導入する可能性がない場合

にのみ、主要な登録基準を満たしていない、試験治療を一回も受けていない、ランダム化後のデータがない等の被験者を「最大の解析対象集団」から除外できることも提示された。さらに、ランダム化後に起こる治験実施計画書違反は、特にそれらの発生が試験治療の割付と関連しているときに、データ及び結論に影響するおそれがある点も言及された。

また、治験実施計画書の基礎となっている科学的なモデルを最もよく反映する集団として、最大の解析対象集団の被験者のうち、治験実施計画書をより遵守している一部の被験者による「治験実施計画書に適合した対象集団 (Per Protocol Set: PPS)」も定義された。しかし、対応する仮説検定と試験治療の効果の推定値は試験により保守的となる場合もならない場合もある点、また、治験実施計画書の遵守状況が試験治療及び臨床結果と関連している場合には、生じる偏りは重大なものとなるおそれがある点も言及された。

以上のように、「最大の解析対象集団」を利用することによって、ランダム化による比較可能性を最大限担保することを目指すものの、それによって推定したい治療効果が治療方針の効果であるかどうかは明確ではないと考えられた。その理由として、以下の 2 点が挙げられる。

- 1) ICH E9 の 5.2.1 節(最大の解析対象集団)では、ランダム化が行われた全被験者を含める ITT を「理想」としており、この理想に可能な限り完全に近づけた解析対象集団として FAS が提示されている。さらに、用語集での Intention-To-Treat の原則の記述において「治療方針により得られる効果 (effect of treatment policy)」と記載されている。以上より、臨床試験では FAS を用いて治療方針の効果を推定することと解釈できる。
- 2) 一方、ICH E9 では、主として「試験治療の効果／治療効果(treatment effect)²⁾」という表現が用いられている。偏りの定義も試験治療の効果の推定値と真の値の系統的な差と定義されている³⁾。また、「治療方針の効果 (effect of treatment policy)」は、「治療効果(treatment effect)」と別の表現が用いられており、ICH E9 の 3.5 節(必要な被験者数)でも『解析対象集団が「最大の解析対象集団」である場合、効果の大きさに関する見積値は、治験実施計画書に適合した対象集団(用語集参照)の場合に比べて小さくする必要があろう。これは、試験治療を中止した被験者又は服薬遵守状況の悪い被験者を解析に含めることにより、試験治療の効果が薄められることを考慮するためである。』とある。以上より、治療方針の効果は、主として用いら

²⁾ 試験治療の効果 Treatment Effect (用語集の記載)：臨床試験における試験治療の効果。ほとんどの臨床試験では、関心のある試験治療の効果は、二つ以上の試験治療間の比較(又は対比)である。

³⁾ 偏り(統計的及び運営上の) Bias (Statistical & Operational) (用語集の記載)：臨床試験の計画、実施、解析及び結果の評価と関連した因子の影響により、試験治療の効果の推定値と真の値に系統的な差が生じること。試験実施中の逸脱によって起こる偏りを「運営上の」偏りとよび、それ以外の原因によって起こる偏りを「統計的」偏りとよぶ。

れている試験治療の効果とは異なり、試験治療の効果が薄められたものと解釈できる。

また、基礎となる科学的モデルに従った試験治療の効果を指して、「治験実施計画書に適合した解析対象集団」が定義された。なお、治験実施計画書に適合した対象集団の解析の意味と役割についての詳細は、本報告書の 3.3 節でとりまとめている。

「最大の解析対象集団」については、佐藤・上坂(2006)でも「E9 では intention-to-treat (ITT)解析対象集団を定義してない。ITT という言葉をめぐって EWG では多くの議論があった。ITT 解析対象集団の規定は様々であることは早い段階で意見が一致していた。結局、ITT は概念であり、解析対象集団を定義する言葉としては適切ではないとして、ITT の概念を反映する言葉をどう定義するかが最後まで議論され、full analysis set が採用された。しかし、full analysis set の内容の議論は十分に合意されないままに文書ができ上がり、重大な試験計画違反や脱落後のデータの取扱いに関しては記述が不明瞭になったままである」と言及されている。

3.1.3.2.2 ICH E9 以前のガイドライン

次に、ICH E9 以前のガイドラインにおける ITT の原則と推定したい治療効果に関する調査結果を示す。

ICH E9 の出発点であったヨーロッパのガイドライン CPMP Working Party on Efficacy of Medicinal Products (1995) は、study population として、ITT population と per protocol population を提示した。ただし、本ガイドラインの ITT population は、ICH E9 の FAS に近い記載であり、事前明記に基づくバイアスを含まない被験者の除外を認めていたり、データ収集に関しても ICH E9 と同様の懸念を示すのみであった。ICH E9 同様、それによって推定したい治療効果が治療方針の効果であるかどうかは明確ではなかった。また、per protocol population は治験実施計画書に適合した解析対象集団と同様の記載であった。

FDA (1988)は、data sets analyzed の節にいくつかの例示 (eg., all patients with any effectiveness observation, only patients completing the trial, and so on)が示されている中で、必ず intention-to-treat analysis を実施することが言及されていた。一方で、推定したい治療効果に対する記載はなかった。

MHLW (1992)は、実践的な試験と研究的な試験という分類を用いて、実践的な試験では比較にとりあげた試験処置のうち、現実の治療にどちらを選ぶかの決定に力点があり、試験処置を開始した被験者は、処置の長短には無関係に解析に含めること (intent-to-treat)、研究的な試験では限定した条件での処置法の比較、薬理学的な特徴の検索、説明に力点があり、適格な被験者でしかも計画通りに試験処置を完了した被験者だけを解析に含めること (protocol-compatible)を示した。そして、通常の比較臨床試験はこれらの中間的な性格をもつことが多いとした。

以上を踏まえ、ICH E9 以前のガイドラインと ICH E9 における推定したい治療効果を表

Aにまとめた。

表 A. ICH E9 以前のガイドラインと ICH E9 における推定したい治療効果

ガイドライン名	推定したい治療効果	解析対象集団 / 解析
ICH E9 (1998)	明確な記載は読み取れない*	Full analysis set
	基礎となる科学的モデルに従った試験治療の効果	Per protocol set
CPMP (1995)	明確な記載は読み取れない*	ITT population
	基礎となる科学的モデルに従った試験治療の効果	Per protocol population
FDA (1988)	明確な記載は読み取れない	Intention-to-treat analysis**
MHLW (1992)	現実的な意思決定に用いる治療効果***	Intent-to-treat
	限定した条件で、薬理学的な特徴が示される治療効果***	Protocol-compatible

* 保守的で日常診療における効果に近い点に言及するものの、治療方針の効果という明確な記載はない

** なお、data sets analyzed の節に ITT 以外の例示 (eg., only patients completing the trial, only patients with a specified degree of compliance, and so on)も示されている

*** 通常の比較臨床試験はこれらの中間的な性格をもつことが多いと記載されている

3.1.3.3 ITT の原則の利点と欠点 / 注意すべき点

文献調査に基づく、ITT の原則の利点と欠点 / 注意すべき点は下記のとおりであった。なお ICH E9 の FAS に対する意見も ITT の原則として含めた。

ITT の原則の利点

- ランダム化を維持することで偏りを防ぐことができる(選択バイアスを避ける)
[ICH E9 (1998) 【5.2.1, p25】 , Newell (1992), 他]
- 統計的検定の強固な基盤を与える [ICH E9 (1998) 【5.2.1, p24】 , 他]
- 保守的 [ICH E9 (1998) 【5.2.1, p24】 , 他]
- 日常診療の効果を反映する可能性が高い [ICH E9 (1998) 【5.2.1, p24】 , Gibaldi et al. (1997), Heritier et al. (2003), Soares et al. (2002), 他]
- 特定の状況(進行性の疾患)では、試験治療を十分に受けることが出来た集団に絞った efficacy 解析よりも検出力が高い [Lachin (2000)]
- 臨床試験において唯一の解釈可能な分析 [Chêne et al. (1998)]
- 服薬不遵守や中止・脱落によって解析から除外されないため、計画されたサンプルサイズが保持される [Gupta (2011)]

ITT の原則の欠点 / 注意すべき点

- ランダム化後におこる事象のデータ及び結論への影響 [ICH E9 (1998) 【5.2.1, p25】 , Newell 1992, 他]
 - 真の治療効果が過小推定される場合や検知できない場合がある [Gibaldi et al. (1997), Whitney (2001), Gupta (2011), 他]

- ◇ 同等性試験又は非劣性試験，安全性評価では保守的ではない[ICH E9 (1998) 【5.2.3, p27】， Lewis et al. (1993)， Hernán et al. (2012)， 他]
- 知りたいものとずれており，直観的ではない
 - ◇ non-compliance のパターンが同じでないとい一般化できない[Ten Have et al. (2008)， Lavori et al. (2001)]
 - ◇ 個々の患者への効果を知りたい医師も，治療方針に従う患者も，non-compliance を含む全体的な結果には興味がない[Soares et al. (2002)]
 - ◇ 治療を検討している人が知る必要がある true clinical effectiveness からほど遠いものとなる[Welsh (2013)].
- データが収集できないと，ITT 解析として実施できない[Greenland et al. (2008)， 他]
- 選択・除外基準を逸脱した不適切な集団が影響してしまう[Gillings et al. (1991)]
- 規制当局の要請を満たすための試験である場合は，pragmatic だけでは満たせないかもしれない[Lavori et al. (2001)]

上記のように，ITT の原則は，ランダム化による比較可能性を維持できる利点を持つ一方で，ランダム化後におこる事象のデータ及び結論への影響が避けられない。利点として挙げられる保守的である点は，主に服薬遵守の観点でとらえられており，計画されたとおりに服薬遵守できないことから，治療効果は減弱されると考えられている[Gibaldi et al. (1997)， 他]。また，服薬遵守のみならず，併用薬違反等のプロトコル違反といった中間事象の影響も含まれることから，より実臨床に近い結果であり，effectiveness や治療方針の効果を示していると考えられている[Gibaldi et al. (1997)， Soares et al. (2002)， 他]。

この保守的という側面は，優越性試験における有効性評価に対する方向性に基づくものであり，非劣性試験や同等性試験，または安全性評価においては逆に働くこととなる[ICH E9 (1998) 【5.2.3, p27】， Lewis et al. (1993)， Hernán et al. (2012)， 他]。

また，effectiveness や治療方針の効果という側面は，中間事象の状況が同じでないとい一般化できない(例えば，当該試験と比較して，服薬がより遵守される，または，より遵守されないといった状況の場合，結果が一般化できない) [Ten Have et al. (2008)， Lavori et al. (2001)]，臨床試験下は eligibility, monitoring, blinding という観点から日常診療への一般化が困難である[Hernán et al. (2012)]，そもそも推定したいものではない[Soares et al. (2002)， Welsh (2013)]という点で，必ずしも利点としてとらえられない場合もあった。

3.1.3.4 NRC (2010)以前の推定したい治療効果

NRC (2010)以前の推定したい治療効果は大きく分けて，effectiveness, efficacy の2つであった。なお，推定したい治療効果が明確でない論文，解析対象集団・解析方法のみに言及する論文も多かった。

Gibaldi et al. (1997)では，effectiveness を「the actual effect of the treatment when

given under usual clinical circumstances, including persons who fail to comply, who change treatment, or who discontinue treatment altogether (治療の不遵守, 変更, 中止などを踏まえた通常の臨床環境下での実際の治療の効果)], efficacy を「the effect of the treatment under optimal circumstances, that is, whether a treatment can have an effect on outcome (最適な環境下での治療の効果, 治療がアウトカムに直接的に影響するかどうか)」としている。また, pragmatic trial に関して「seeks to describe the consequences, both good and bad, of attempting to treat an illness in a certain way and to determine whether treatment works under usual clinical circumstances [疾患に対してある決まった方法で治療を行った結果(良し悪し含む)を示し, 通常の臨床環境下で治療が効くかどうかを決定することを求める]」, explanatory trial に関して「seeks to describe how a treatment produces its effects and to determine whether it can work, usually under restricted or ideal circumstances (治療がどうやれば効果を発揮するかを示し, 制限された理想的な環境下で治療が効くかどうかを決定することを求める)」と説明したうえで, pragmatic trial と effectiveness 及び explanatory trial と efficacy の関連がそれぞれ示されている。

試験において effectiveness を推定することの利点としては, ITT の原則による解析で知ることができると考えられていること [Newell (1992), Gibaldi et al. (1997)] があげられており, またその他の点については ITT の原則を利用することの利点と同様であった。一方で, efficacy を推定することの利点としては, 下記の内容が指摘されていた。

- 理想的な状況下での治療効果を特定する手助けとなる [Whitney (2001)]
- 特に降圧薬の試験などで治療の変更が生じやすい状況では, 治療中のデータのみを用いる解析の方が clinical practice をより反映している [Elliott (1995); Gibaldi et al. (1997) より引用]
- 医療上の意思決定には割付けられた治療の効果よりも実際に受けた治療の効果のほうが重要である [Sheiner et al. (1995)]

なお, efficacy については, per-protocol 解析や as treated 解析 (実際に治療された投与群を用いて比較する方法) によって推定を行う場合, これらの解析では特定の間事象が発生した被験者を対象集団から除外するため, 比較可能性が担保できないという点が問題となる。この点を踏まえ, per-protocol 解析や as treated 解析ではなく, g-estimation, 主要層解析や操作変数法等による因果推論の手法を用いた解析により efficacy の推定を行うことが提案されている [Sheiner et al. (1995), Bang et al. (2007), Greenland et al. (2008), Taylor et al. (2009), Sjölander et al. (2009), Matsuyama (2010)]。

3.1.3.5 NRC (2010)以降の推定したい治療効果

3.1.3.5.1 論文で取り上げられた治療効果

NRC (2010) 以前に effectiveness, efficacy など数種類の呼び方で表現されていた臨床試験で推定したい治療効果が, 2010年以降, estimand という概念を用いて, より多種多様

にきめ細かく表現できるようになった。ICH E9 (R1) step 2文書でも、対象集団、変数・評価項目、中間事象の取り扱い、集団レベルでの要約方法、という4要素が提示されたように、その組み合わせによって表現されるestimandの数は計り知れない。ここでは、調査結果を以下の2つに分けてまとめるが、estimandの種類はこれらに限定するものではない。

- Mallinkrodt et al. (2012): Estimand 1/ 3 /6
- 部分集団における efficacy の観点から定義された推定したい治療効果
まず, Mallinkrodt et al. (2012)は, NRC (2010)を踏まえ, 推定したい治療効果 / estimandとして6つの例を取り上げた。そのうちの代表的な3つを以下に記載した。
- Estimand 1: ランダム化されたすべての患者における(中間事象の影響も含んだ)効果
Difference in outcome improvement at the planned endpoint for all randomized participants
- Estimand 3: すべての患者が治療を完了した場合の治療効果
Difference in outcome improvement if all subjects tolerated or adhered
- Estimand 6: ランダム化されたすべての患者における(ランダム化された初期の)試験治療に帰属する治療効果
Difference in outcome improvement in all randomized patients at the planned endpoint of the trial attributable to the initially randomized medication

これらの違いは、中間事象の取り扱いによるものであり、estimandの考え方やそれぞれの利点・欠点を示すために様々な論文で取り上げられている(表B参照)。また、具体的な事例として、Holzhauer et al. (2015)は、レスキュー薬を用意している2型糖尿病の臨床試験におけるestimandを”Estimand 6”として捉え、その解析方法についてまとめている。

表B. Mallinckrodt (2012)のestimandとその引用

Mallinckrodt et al. (2012)	Estimand 1	Estimand 6	Estimand 3
Hypothesis	<i>de facto</i> (effectiveness)	<i>de facto</i> (effectiveness)	<i>de jure</i> (efficacy)
中間事象の取り扱い	試験治療中止やレスキュー薬開始に関わらず、すべて用いる	試験治療中止やレスキュー薬開始後について、無治療を仮定	試験治療中止やレスキュー薬開始後について、試験治療継続を仮定
Mallinckrodt (2013)*	Estimand 1	Estimand 6	Estimand 3
Mallinckrodt et al. (2014)	Estimand A	Estimand B	Estimand C
Leuchs et al. (2015)	Difference for All Randomized Patients	Difference for All Randomized Patients Attributable to the Initially Randomized Treatment	Difference if All Patients Adhered
Mallinckrodt et al. (2017)	Estimand 1	Estimand 2	Estimand 3
Phillips et al. (2017)	Estimand 1	Estimand 2	Estimand 3

* Mallinckrodt (2013)は調査対象ではなかったが、関連する文献なので、参考として追記した。

一方、前節でも触れたように、推定したい治療効果を、部分集団におけるefficacyの観点から定義するものもあった。Shrier et al. (2014)は、下記のようなclinical questionに対して、どのような解析手法がそれぞれの疑問に対する適切な因果効果を提供するかを議論した。

- **Question 1: Policy.** Will the active treatment improve health in the population if I fund the program? (方針：治療の導入が想定する集団全体の健康増進に寄与するか?)
- **Question 2: Clinical Practice.** Will changing treatment practices improve the health of my patients in general? (治療：治療変更が患者の健康増進に一般的に寄与するか?)
- **Question 3: Clinician recommendation to compliant patient.** Will the active treatment improve the health of my patients who would follow my advice? (服薬遵守者への医師の推奨：治療が、医師の推奨に従って服薬を遵守する患者の健康増進に寄与するか?)
- **Question 4: Individual patient receiving treatment.** Will active treatment improve the health of my patients who will take the active treatment? (治療を受ける個々の患者：治療が、医師が推奨するしないに関わらずその治療を受ける患者の健康増進に寄与するか?)

特にQ3に対しては、主要層解析や操作変数法によってComplier Average Causal Effect (CACE, どの治療を推奨されたとしても適切に遵守できるであろう集団での因果効果) を推定することで対応できること、また、Q4に対する解析手法は開発されていないことを言及した。また、Lou et al. (2019)は、非劣性・同等性におけるestimandとして、Survivor Average Causal Effect (SACE) Estimandを取り上げた。さらに、Akacha et al. (2017)は、臨床試験における治療方針の効果は、実臨床における治療効果とは言い切れず⁴、また、信じがたい結論を導く場合があると指摘し、患者、規制当局、医療費用支払者が知りたいものとして、推定したい治療効果を下記の3つの疑問にわたるtripartite frameworkを提案した。

- Non-adherence to treatment due to safety
(安全性によってどれだけの患者が治療を遵守できないか?)
- Non-adherence to treatment due to lack of efficacy
(有効性の欠如によってどれだけの患者が治療を遵守できないか?)
- Effects in adherers
(治療を遵守できる集団における有効性と安全性のプロファイルは?)

この中で、effects in adherersについては、因果推論が応用できる点、およびそのためには臨床統計家に対するフォーマルなトレーニングと実際の応用経験が必要である点を言及した(Ruberg et al. 2017も参照)。Akacha et al. (2017)に対して、Rockhold (2017)は、処方箋の3分の1が調剤されていないといったデータやいつでも添付文書に従うわけではないことを踏まえ、perfect adherers (完全に服薬遵守できる集団)を中心に考える点について懸念を示した。また、ITTの原則が不明瞭というAkacha et al. (2017)の意見に対し、ITTの原則以上に複雑なアプローチ⁵を臨床医や患者が解釈できるのかについて、懸念を示した。一方、ITTの原則は、いくつかの問題はあるものの、結果を科学や医療のコミュニティに取り込むためのよく確立された信頼できる科学的コンセプトであると言及した。また、Permutt (2017)は、効果の定義が曖昧である点、ランダム化臨床試験が観察研究と同等と考えることになる点、因果推論の場合、用いられるモデルに対するstakeholderの信頼度合に影響される点について、懸念を示した。一方、treatment policy estimandであれば、ロバストに推定できる点を言及した。Bauer (2017)は、すべてのstakeholderのニーズを満たすただ一つのestimandはなく、それぞれに利点・欠点があるため、estimandを比較することを避けたいとした。また、tripartite approachは、適切な有効性の推定をどう

⁴ ランダム化している、盲検下で行われている、定期的フォローアップされる、通常以上の検査や介入が行われる、試験治療の用量変更やレスキュー薬に対して厳格なルールがあるといったことから、通常の治療環境下と異なることを指摘し、実臨床における治療効果を知りたい場合は pragmatic study のような実臨床に合致した試験デザインを用いるべきと言及した。

⁵ Akacha et al. (2017)が触れた因果推論に基づくアプローチのことと読み取れる

やって構築するかという複雑さを後々の意思決定(例：承認するかしないか)に先延ばしする点を言及した。

この節で取り上げた2つの分類[Mallinkrodt et al. (2012)の分類，部分集団における efficacyの観点から定義されたもの]以外として，Permutt (2016)は，2つ目の分類と同様に，因果推論の観点から様々なestimandについて言及した。JPMA (2017)「臨床試験の estimand を理解するための因果推論」はその詳細をまとめている。

3.1.3.5.2 EMA の各種疾患ガイドラインで取り上げられた治療効果

ICH E9 (R1) step 2 文書の発出と前後して，EMA の各疾患に対するガイドラインでも，その疾患に特化した治療効果が示されるようになってきた。調査対象のうち，推定したい治療効果が明記されていた3つのガイドラインを以下にまとめた。

Guideline on clinical evaluation of vaccines (draft) (2018)は，下記のように，主たる興味によっては，ランダム化されたすべての被験者に基づく解析ではなく，一部の部分集団に基づく解析も適切であることを示した。

- When the major interest is to estimate the vaccine efficacy that could be expected in routine use, the primary analysis may be conducted in all randomised subjects who receive at least one dose of assigned treatment

(主たる興味が通常時に期待されるワクチンの有効性の推定の場合，主要な解析は，ランダム化され少なくとも一回は割り当てられた治療を受けた被験者で行われるかもしれない)。

- When the major interest is to obtain a best-case estimate of vaccine efficacy, the primary analysis may be conducted in subjects who received all the allocated doses within pre-defined windows. For some vaccines and infectious diseases, it may also be acceptable that the primary analysis is confined to those subjects who were seronegative or had no ongoing infection with the target pathogen at trial baseline

(主たる興味が一番良い状況におけるワクチンの有効性の推定の場合，主要な解析は，事前に定義された期間内に割り当てられたすべての治療を受けた被験者で行われるかもしれない。いつかのワクチンや感染症においては，主要な解析を，ベースラインにおいて血清陰性であった又は標的病原体に継続的に感染していなかった被験者に限定することも許容されるかもしれない)。

Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus (draft) (2018)は，HbA1c を用いた評価におけるいくつかの治療効果に言及した。例えば，有害事象等で試験治療が続けられない場合，血糖値のコントロールに対するベネフィットが期待できないため，試験治療中止後に観測された HbA1c や試験治療中止を踏まえたモデルに基づく HbA1c を用いて，治療遵守状況を反映した治療効果を推定すべきとした。また，レスキュー薬を含む HbA1c に影響する薬剤の変更や使用は，有効性評価に影響する

ため、それらが導入されなかったと仮定した場合の治療効果も取り上げた。さらに、追加治療なしで試験終了時点における HbA1c の基準値を満たし、かつ治療遵守した場合を治療成功と考える複合ストラテジーも示している。

Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease (2018)は、ICH E9 (R1) step 2 文書と同様に、試験デザイン、データ収集、統計解析の選択は、試験目的によって提起される興味ある科学的な質問に沿うべきであり、中間事象の取り扱いに関するストラテジーを含んだ estimand の詳細な規定が必要であるとした。また、対象とする疾患段階[Alzheimer's Disease (AD) Dementia, Prodromal AD/Mild cognitive impairment (MCI) due to AD, Preclinical AD]によって推定すべき対象は異なってくるが、どの段階においても、実際の投与状況を推定すべき対象に反映させるべき (例えば、試験治療中止という中間事象に対して treatment policy を適用)とした。

AD Dementia においては、試験治療中止後にベネフィットが継続することが期待されないため、試験治療中止を反映した治療効果が通常、主たる推定対象となる。また、レスキュー薬としての symptomatic treatment の開始・用量変更が考えられるため、そのようなレスキュー薬がなかった場合といった仮想的なシナリオに基づく estimand も適切と考えられる。試験が長期間となる場合は、estimand における死亡に対する取り扱いも考える必要がある。

Prodromal AD/MCI due to AD においても、レスキュー薬としての symptomatic treatment の開始が中間事象として考えられ、AD Dementia 同様、そのようなレスキュー薬がなかった場合を考える estimand も適切と考えられる。

Preclinical AD においては、長期間の試験が計画され、試験治療中止が想定されることから、実際の投与状況が推定すべき対象に反映されなければならない。また、一部の集団における効果、例えば、十分な試験治療が実施される患者集団における治療効果の理解も、興味の対象となるが、信頼できる推定値を得るための試験デザインや統計解析は明らかではない。ただし、observed case タイプのアプローチは不十分である。また、AD や治療には関連しなくとも、アウトカムに関連するであろう幅広い中間事象について、estimand の中で、その取り扱いを適切に示す必要がある。

3.1.4 文献調査結果からの考察

調査結果を踏まえ、ITT の原則と治療方針の効果、efficacy の対象集団、及び estimand に対する今後の留意事項の 3 点で考察する。

3.1.4.1 臨床試験における ITT の原則と治療方針の効果に対する再考

選択バイアスを避け、偏りのない比較を行うために、ランダム化が行われ、それを維持するために ITT の原則が示された[ICH E9 (1998)]. 調査結果(3.1.3.3 項 ITT の原則の利点と欠点 / 注意すべき点)で示したように、ITT の原則に基づいて推定される効果には、ランダ

ム化後の試験治療の中止，レスキュー薬の開始，併用禁止薬の使用などの中間事象の影響をすべて含むという特徴がある．この単純だが幅広い特徴のため，推定される効果の一義的な意味づけが困難になる場合もある．例えば，試験治療中止後のデータを用いることで，服薬状況を加味した治療効果が推定できそうだが，臨床試験における服薬状況が実臨床環境下で再現されるかは，また別の話である[Ten Have et al. (2008), Lavori et al. (2001)]．また，そもそも臨床試験下は **eligibility, monitoring, blinding** という観点から日常診療への一般化が困難という意見もある[Hernán et al. (2012)]．さらに ITT の原則は，解析すべき対象集団や中間事象の取り扱いが含まれるものの，試験デザインや評価項目について定めたものではない．非常に統制され，服薬遵守状況も非常に高い臨床試験が実施された場合，ITT の原則に基づいて推定される効果の意味づけは明確であり，**efficacy** に近いものと考えられる．このように，ITT の原則に従う場合であっても，治療方針の効果や **effectiveness** という表現だけでは不明確であり，選択基準や除外基準から決まる対象集団，試験治療の遵守状況，併用薬やレスキュー薬の利用可能性，評価項目，起こりうる中間事象やその起こりやすさ等を加味してはじめて，どのような効果となるかが明確になる．

3.1.4.2 Efficacy として推定した治療効果の対象集団は？

ICH E9 では，最適な環境下での治療効果[Gibaldi et al. (1997)]，限定した条件で，薬理学的な特徴が示される治療効果[MHLW (1992)]を推定し，科学的なモデルを最もよく反映する集団として，PPS を提示した．また，ICH E9 では触れられていないが，ランダム化されたとおりに実施した治療だけで比較する **per-protocol** に対応する形で，ランダム化された治療によらず実施した治療に基づいて比較するアプローチ(**as treated approach**)も議論されてきた[Newell (1992), Elenberg (1996), Gibaldi et al. (1997)]．これらは，中間事象に基づいて，ランダム化された集団から一部を除外することから，比較可能性が担保されず，バイアスが入ることが問題となってきた[ICH E9(1998)他]．したがって，ICH E9 で，PPS の役割として，PPS 単体の解析により意味のある治療効果を推定することを意図しているかは不明であるという意見もある．この問題に対して，調査結果 (3.1.3.4 項 NRC (2010)以前の推定したい治療効果や 3.1.3.5 項 NRC (2010)以降の推定したい治療効果)で示したように，統計的因果推論をベースとして，全体集団を対象とする解析手法[Inverse probability (IP) weighting: Hernán et al. (2012), g-estimation: Greenland et al. (2008), Matsuyama (2010), Hernán et al. (2012)]と一部の集団を対象とする解析手法[complier average causal effect に対する主要層解析や操作変数法: Hernán et al. (2012), Shrier et al. (2014)]が提案されてきた．ここでまず，立ち戻りたいのが，**estimand** の要素の一つである対象集団である．私たちは，どのような対象集団における治療効果を推定するために，PPS を用いたり，**efficacy** と表現したりしていたのだろうか？

例えば，Mallinckrodt et al. (2012)の **estimand 3** は，すべての患者が治療を完了した場合の治療効果を見ているものであり，**efficacy** としての治療効果を推定していた．しかしな

がら、一部の疾患領域においては、実際には様々な理由(例えば、有害事象、有効性の欠如等)によって治療が完了できなかった患者も存在するにも関わらず、すべての患者が治療を完了できたとするこの *estimand* は、薬剤の有効性を検討する上で適切な *estimand* ではないという意見もあった(例えば、EMA のアルツハイマー病ガイドライン)。推定したい治療効果に服薬遵守状況を加味する観点には、Akacha et al. (2017)の *tripartite framework* でも強調されている。

このような状況も踏まえ、ICH E9 (R1) step 2 文書では、「本補遺では、治験実施計画書に適合した対象集団の解析の意味と役割について再考する」と記載されたと考えられる。

また、何が推定したいかという観点の他に、どのような仮定の下でそれが推定可能か、また、どのようなデータを収集する必要があるかという点についても、注目する必要がある。例えば、Bang (2007)では、操作変数法の使用について、操作変数に対する正確な情報とデータから推定可能とするための仮定が必要である点に、また、Hernán et al. (2012)も「*IP weighting and g-estimation require untestable assumptions and time-varying data on confounders and adherence. IV estimation requires a dose-response model and time-varying data on adherence.* (逆確率重み付け法や *g-estimation* は検証できない仮定及び交絡因子や服薬遵守に対する時期ごとのデータを必要とする。また、操作変数法は用量反応モデル及び服薬遵守に対する時期ごとのデータを必要とする)」と言及している。

なお、治験実施計画書に適合した対象集団の解析の意味と役割についての詳細は、本報告書の 3.3 節でとりまとめている。

3.1.4.3 *Estimand* に対する今後の留意事項

調査結果 (3.1.3.5 項 NRC (2010)以降の推定したい治療効果)で示したように、NRC (2010)、ICH E9 (R1) final concept paper (2014)、そして ICH E9 (R1) step 2 文書を踏まえ、推定したい治療効果は何か?という観点で研究された論文が増えてきている。そして、それらの中には、臨床試験において実臨床での治療方針の効果を推定することが適切でないと考えられたため、治療方針の効果とは異なる治療効果を目的とするものも多く見受けられた。例えば、実際にはレスキュー薬が使用された場合に、レスキュー薬が使用されなかったと仮定した場合の治療効果が挙げられる[Mallinckrodt et al. (2012)他]。EMA 疾患ガイドライン(糖尿病及びアルツハイマー病)においても、他治療の影響を除外した治療効果を検討する旨が明記されていた。一方で、治療を選択する医師の観点から、様々な服薬遵守状況によって試験治療の効果が薄められていない、医師の指示に従って治療が完了できる場合の治療効果も検討されていた[Shrier et al. (2014)]。

これらの治療効果は、治療方針の効果よりも規制上及び臨床上的意思決定に関連する治療効果かもしれない。ただし、文献調査の結果、(1)結果の一般化可能性、(2)結果の比較可能性、(3)推定のためのデータの利用可能性、(4)データが存在しない場合、(5)非劣性・同等性や安全性評価の *estimand*、(6) *estimand* と意思決定、といった点が、今後の留意事項に

なると考えた。以下に詳細を述べる。

まず、結果の一般化可能性についてである。ICH E9 では、検証的試験において、「意図している患者集団への一般化の根拠が理解でき、説明できることは重要である」としている。ITT の原則に基づく場合、様々な中間事象の影響を受けた効果となる。これが「日常診療の効果を反映する可能性が高い」という利点とされたり、「状況が同じでない日常診療における効果として一般化できない」という欠点とされたりする。一方、様々な中間事象の影響を除いた特定の治療効果や特定の集団を目指した治療効果において、ITT の原則に基づく効果よりも一般化可能性が高いかどうかは明確ではない。なお、per-protocol effect を推定して、必要に応じて中間事象の割合を考慮することで、一般集団の場合の効果を推測するという主張も考えられる。どのように結果を一般化したいのかについても、推定したい治療効果と同様に、実施可能性も踏まえた上で、事前に十分に議論される必要がある。

次に、結果の比較可能性についてである。Leuchs et al. (2017)は、ITT の原則を以下の3点に分け、うち上の2つ(割付, 集団)を守ることをITT の原則と再定義し、これにより選択バイアスを避け比較可能性を担保することを示した。

- **Allocation:** Treatment groups are defined by the randomized treatment allocation and not by the treatment actually received.
(割付：治療群は、実際に受けた治療ではなく、ランダム化された治療に基づいて定義される)
- **Population:** All randomized patients are included.
(集団：ランダム化された全ての患者を含める)
- **Follow-up:** All patients are followed up irrespective of any protocol deviations (ie, completely observed data).
(フォローアップ：プロトコルの逸脱に関わらず全ての患者をフォローアップし、データを入手する)

この定義は推定したい治療効果を検討する上での出発点として考えられ、NRC (2010)以降の多くの治療効果が、Leuchs et al. (2017)が再定義したように、上記の3つのうち上2つを遵守できるように構築されている。

さらに、適切に試験治療の効果を推定できるかという観点である。どのような効果を選ぶにしても、その効果が推定できるデータが得られるかは非常に重要である。例えば、ITT の原則に基づく効果の場合、どのような中間事象が起きたとしても、計画された最終時点における測定値を収集する必要がある。Efficacy に基づく治療効果についても、inverse probability (IP) weighting や g-estimation, 操作変数法といった因果推論に基づく場合、交絡因子や服薬遵守状況、場合によっては中間事象後の結果変数のデータを収集する必要がある。仮にレスキュー薬を使用しなかった場合の治療効果においても、使用しなかった状況をモデル化するためのデータが必要であり、ほとんどの患者がレスキュー薬を使用することが予想される場合は適応が難しい。一方、estimand によっては、逆にデータを取る必

要が無くなることもあり得る。例えば、中間事象自体をエンドポイントに組み込む複合ストラテジーを用いた場合、中間事象が起こった後のデータについては、収集する必要はない。

データの利用可能性という点でいえば、死亡といった中間事象後のデータがそもそも存在しない場合もある。調査した論文の中で、推定したい治療効果として、これらを考慮しているものは限られていた。例えば、ICH E9 では、最大の解析対象集団の被験者全てについて、主要変数の測定値が利用できることを保証するための方法の一つとして、単純化した臨床結果(例えば、成功か失敗か)が挙げられており、ICH E9 (R1) step 2 文書でも複合ストラテジーとして提示されている。アルツハイマー病に対する EMA ガイドラインでは、「試験が長期間となる場合は、estimand における死亡に対する取り扱いも考える必要がある」と触れられていた。Permutt (2016)では、死亡に対して、主要評価項目と同じスケールの値 (utility)を与える utility estimand が提案されていた。Lou et al. (2019)は、生存者に限定する形で estimand を構築した[非劣性・同等性における estimand として、Survivor Average Causal Effect (SACE)]。どのような estimand を選ぶにしても、事象発現後のデータが存在しない死亡のような中間事象が多く発現し、治療効果の評価に対する影響が大きいと考えられる場合は、取り扱いを明確にする必要がある。

また、非劣性試験や安全性評価における estimand についても言及する。服薬遵守の観点から ITT の原則に基づく場合、試験治療の効果が減弱すると考えられ、この性質が非劣性試験や安全性評価においては、逆に保守的ではないと考えられてきた。これに対して、例えば、Matsuyama (2010)では、非劣性試験における生存時間解析における、より適切な推定方法として、g-estimation に基づく治療効果を取りあげ、ITT や per-protocol approach と比較検討した。つまり、服薬遵守が適切な場合の治療効果を検討することで、より適切な評価をしようという考えである。一方で、非劣性試験の中には、同等の有効性を担保した上での安全性の改善を目的とした試験も存在する。そのような場合、服薬遵守が適切な場合を仮定した治療効果が試験目的に即しているかどうか注意する必要がある。

最後に、意思決定と estimand の関係である。Leuchs et al. (2015)や Phillips et al. (2017)で言及されたように、臨床試験は様々な stakeholder が関与するため、必ずしも一つの estimand で十分というわけではない。Akacha et al. (2017)は、それらをまとめる tripartite framework を提案したが、これまで中心的に用いられてきた ITT の原則の方が確立された信頼できる科学的コンセプトというコメントもあり、議論は始まったばかりである。

以上のように、estimand を設定するにあたっての留意事項は、estimand で対応する範囲よりもはるかに幅広い。今一度、ICH E9 (R1) step 2 文書のフレームワークを紐解くと、estimand の前に試験の目的がある。一部の解析手法によって estimand が暗に決まっていたように、estimand によって試験目的が無意識のうちに決まらないように注意しなければならない。

3.1.5 参考文献

- Elliott W. J. (1995), Glucose and cholesterol elevations during thiazide therapy: intention-to-treat versus actual on-therapy experience. *The American Journal of Medicine* (99), 261-269.
- Mallinckrodt (2013) *Preventing and Treating Missing Data in Longitudinal Clinical Trials: A Practical Guide*, Cambridge University Press.
- JPMA (2017). 臨床試験の estimand を理解するための因果推論. Available at: <http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/estimand.html> [Accessed 30 May 2019]

3.2 データの取扱い及び欠測データに含めて検討されている問題

3.2.1 調査対象詳細

ICH E9 及び ICH E9 (R1) step 2 文書では、欠測データに関連して主に以下の記載がされている。

ICH E9 (5.2) :

・ 治験実施計画書には、十分な解析を損なうと予想される研究実施中の変則的な事例のすべて(様々な種類の治験実施計画書違反、試験治療の中止、欠測値を含む)を最小にするための手続きも明記すべきである。

ICH E9 (R1) step 2 文書(A.1.) :

・ 中間事象は、場合によっては、その発現後に変数の測定値を収集することが可能であっても、その測定値を意味がない又は解釈が困難なものにしてしまうことがある。死亡の場合には、被験者が死亡した後の測定値は存在しない。

・ Estimand を明確にすることにより、どのデータを収集する必要があるか、つまり、どのデータが収集されなかった場合に対処すべき欠測データの問題があるかを計画するための根拠が得られる。

ICH E9 (R1) step 2 文書 (用語集) :

・ 欠測データ：規定した estimand の解析に対して意味があると考えられるが、収集されなかったデータ。存在しないデータや中間事象の発現により意味があると思えないデータとは区別されるべきである。

・ 中間事象：治療開始後に発現し、変数を観測できなくする、又は変数の解釈に影響を与える事象。

これらの記載を踏まえ、「データの取扱い及び欠測データに含めて検討されている問題」に関して詳細に検討する調査対象として「欠測データの定義」、「欠測データを最小にするための方法論と適用時の注意点」及び「中間事象発現後のデータの取り扱い方法」の3点を設定した。

3.2.1.1 欠測データの定義

欠測値または欠測データという用語は、臨床試験に限らず一般的に用いられる用語である。例えば、統計学辞典(1989)では、自然科学的実験・調査における欠測データとして「実験・調査の計画段階において、実験・調査者は、得られる測定値の全体像、すなわち変数の数と性質、変数の取りうる値の範囲と分布、測定値の個数、にたいしておおよそのイメージを持っている。調査・実験が完了したときに、このイメージのもとで得られていたはずの測定値が得られていなかったとき、これを欠測値という」としている。また医学統計学ハンドブック(1995)では、「欠測値とは、本来得られるはずが実際には得られなかったデータ」と記載されている。いずれの定義においても、「事前段階では収集可能と考えていたものの、実際には収集されなかったデータ」という点において共通していると考えられる。

一方で、臨床試験においては、欠測データという用語は頻繁に使用されており、ICH E9 において、十分な解析を損なうと予想される研究実施中の変則的な事例の一つの例として欠測値が記載されている。しかし、ICH E9 には欠測データそのものの定義は明確に記載されていない。欠測データの定義は、ICH E9 (R1) step 2 文書の用語集で明記されており、「規定した estimand の解析に対して意味があると考えられるが、収集されなかったデータ。存在しないデータや中間事象の発現により意味があると思えないデータとは区別されるべきである。」とある。すなわち、ICH E9 (R1) step 2 文書の欠測データの定義では、事前段階において、必ずしも全例が定められた治療を完了するという理想的な状況を考えることを前提としておらず、むしろ、種々の可能性を考慮して事前に収集すべきデータを明確に想定していると考えられる。

本調査では、「存在」と「収集」の概念を明確にするため、ICH E9 (R1) step 2 文書で使用されている用語を次のように読み替えている。具体的には、「規定した estimand の解析に対して意味があると考えられるが、収集されなかったデータ」を「(1)収集すべき、かつ存在するが、収集されなかったデータ」, 「存在しないデータ」を「(2)値自体が存在しないため収集できないデータ」, また「中間事象の発現により意味があると思えないデータ」を「(3)値自体は存在するが、中間事象の発現により実施する解析に対して意味があると思えないため、収集しないデータ」と言い換えた。

欠測データの定義に関連する調査では、上記の(1)~(3)のデータの考え方を整理することを目的とした。

3.2.1.2 欠測データを最小にするための方法論と適用時の注意点

ICH E9 では、十分な解析を損なうと予想される研究実施中の変則的な事例の一つとして欠測データがあげられており、これら変則的な事例の発生頻度を最小化することの重要性が明記されている。

ここでは、欠測データの発生頻度を最小化するために実施しうる具体的な方法を試験計画、試験デザイン、試験実施に区別して整理する。一方で、これらの工夫を行うことによって、欠測データの発生頻度以外においても様々な影響が生じることが想定されることから、欠測データを最小にする工夫を実施する際に留意すべき事項に関しても検討を行う。

3.2.1.3 中間事象発現後のデータの取り扱い方法

ICH E9 (R1) step 2 文書では、中間事象とは、「治療開始後に発現し、変数を観測できなくする、又は変数の解釈に影響を与える事象のことであり、その発現後に変数の測定値を収集することが可能であってもその測定値を意味がない又は解釈が困難なものにしてしまうことがある」と記載され、一例として、「中間事象には特に、代替治療(レスキュー薬、治験実施計画書における併用禁止薬、又は後続ラインの治療)の使用、治療の中止、治療の切り替え、状況によっては死亡といった終末事象が含まれる。」とある。過去には、これらの事象の発現後にはデータの収集を中止したり、又はデータを収集していても解析には利用せずに、欠測データの問題として対処していた事例も多く、製薬協レポート「欠測のある連続量経時データに対する統計手法について」(2016年4月)の Appendix 2 (欠測のあるデータに関する感度分析の事例)においてもいくつかの事例が紹介されている。

ICH E9 (R1) step 2 文書では、中間事象に対応するためのストラテジーの例がいくつか紹介されている。例えば、治療方針ストラテジーの場合は、「中間事象の発現は問題としないものとする、つまり、中間事象の発現の有無に関わらず、関心のある変数の値を用いる。」と記載されており、中間事象が発現した後であっても収集された実際のデータを用いて、治療効果の推定を行うストラテジーと解釈することもできる。すなわち、関心のある変数が全て収集されていれば、その値をそのまま用いて解析されるため、中間事象発現後に収集されたデータの取扱いに関して検討する事項はそれほど多くないと考えられる。しかし、治療中止後のデータが収集されていないなど、中間事象発現後に収集されなかったデータの場合は、中間事象発現後のデータの取扱いに関する検討に注意が必要になると考えられる。

また、仮想ストラテジーの場合は、「中間事象が発現しなかった状況を想定する。関心のある科学的疑問を反映する値は、定義された仮想的な状況において得られたであろう変数の値であるとする。」と記載されており、何かしらの仮想のデータを明示的もしくは非明示的な方法で補完する必要があると考えられる。そのため、実際には存在しない仮想的なデータの取扱いに関する検討が必要になると考えられる。

以上より、本調査では、中間事象発現後に収集されなかったデータ、収集できないデータ、収集しないデータの取扱い方法の方針に関連する論文を中心に調査し、整理することを目的とした。なお、多くの論文では中間事象という用語は使用されていないため、ICH E9 (R1) step 2 文書で記載されている例に従い、中間事象を治療中止や脱落などに置き換えて調査した。

3.2.2 文献調査方法の詳細

本文献調査では、2018年8月に、PubMed, TIRS 及び医中誌 Web の3つの文献検索サイトを利用し、設定した調査対象に関連する記載があることが期待される文献を検索した。この際、欠測データに加えて、脱落や死亡等の中間事象に関する議論を臨床試験の文脈において主に取り扱っている文献が検索可能となるように検索時のキーワードを設定した。その後、検索結果として表示された文献について、本文献調査の担当者である3名がそれぞれ文献のタイトル及び抄録を確認し、少なくとも1名以上が調査対象に関連する文献である可能性があるかと判断した場合に、当該文献の本文を確認する手順とした。

文献検索時のキーワード

- PubMed (検索式)
 - Search 1: (Missing[title]) AND (Data[title] OR Outcome[title]) AND (clinical[title] OR randomi*[title] OR regulatory[title] OR dropout[title] OR discontinu*[title] OR additional[title] OR death[title] OR prevent*[title] OR treat*[title]) NOT (meta[title]) NOT (sensitivity[title]) NOT (cluster[title])
 - Search 2: Missing and estimand
 - Search 3: (ICH E9 [Title] OR ICH-E9 [Title] OR E9 [Title]) AND (clinical trial [MeSH Major Topic])
- TIRS (検索式)

Search 4: Missing data[Title]

- 医中誌 (検索式)

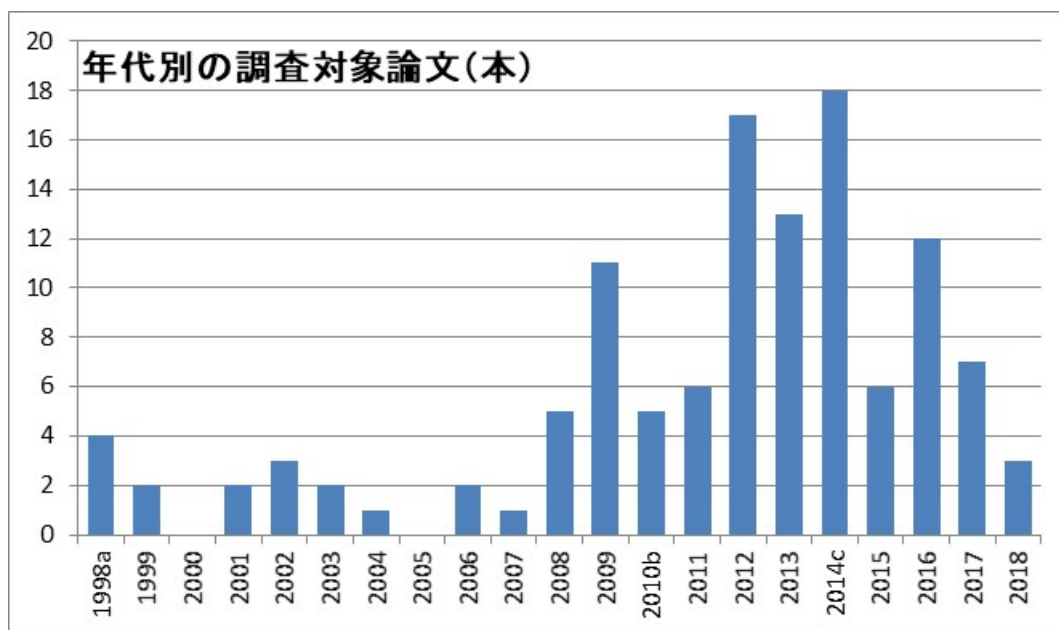
Search 5: 欠測値 or 欠測データ

上記により検索結果として表示された文献[合計: 339 本(PubMed: 280 本, TIRS: 10 本, 医中誌: 49 本)]について, 本文献調査の担当で文献のタイトル及び抄録を確認し, 調査対象に関連する文献である可能性があると判断された 120 本の文献(PubMed: 99 本, TIRS: 10 本, 医中誌: 11 本)を調査対象文献(文献一覧は 6.付録(2)参照)とした。

3.2.3 文献調査結果

3.2.3.1 調査対象とした文献の内訳

120 本の調査対象文献について, 調査対象文献の年代の分布は, NRC (2010)発出以降の論文が 87 本であり, 2009 年以前の論文は 33 本であった。NRC (2010)発出前の 2008 年頃から論文数が急激に増加している様子が伺える。



a: ICH E9 (1998), b: NRC (2010), c: ICH E9 (R1) final concept paper (2014)

3.2.3.2 欠測データの定義

3.2.3.2.1 欠測データの定義に関する記載

調査対象文献のうち, 7つの文献において欠測データの定義が明記されていた。今回の調査では, 臨床試験において広く統一的に用いられている欠測データの定義に関する記載は見つからなかったものの, 欠測データの定義として「収集されなかったデータ」[Lavori et al. (2008), Ibrahim et al. (2012), Bannon (2015)]としているものと, 「解析に対して意味があると考えられるが収集されなかったデータ」[NRC (2010), Little et al. (2012b), Kang (2013), LaVange et al. (2016)]の大きく 2つに分けられると考えた。特に NRC (2010)では,

「whether missingness of a particular value hides a true underlying value that is meaningful for analysis (特定の値の欠如により、解析に意味のある真の潜在的な値が隠されているか)」を特定することが必要であるとしている。

3.2.3.2.2 「存在しないデータ」と仮想的な状況を想定することの妥当性

LaVange et al. (2016)では、「unobserved data and nonexistent data are not the same problem, but the two situations have often been confused (観測されなかったデータと存在しないデータは異なる問題であるが、しばしば混同される)」とされており、欠測データの問題を考えるうえで、「存在するが収集されなかったデータ(unobserved data)」と「存在しないデータ(nonexistent data)」の違いを考えることは特に重要と思われる。ここでは、文献において「存在しないデータ」に関連して記載されていた内容を以下に示す。

「存在しないデータ」の例示として多くの文献で紹介されているのが死亡後の QOL の測定値である[Shih et al. (1997), Myers (2000), Lee et al. (2010), NRC (2010), Kurland et al. (2012), Little et al. (2012b), LaVange et al. (2016)]. これらの「存在しないデータ」が発生する状況において各文献で共通する指摘は、「もし死亡等の事象が発生しなかったら」という仮想的な状況を想定するような、「存在しないデータ」を通常の欠測データと同様に扱ったうえでの推測を行うべきでないという点であった[Myers (2000), Lavori et al. (2008), Soon (2009), Permutt (2016)]. この仮想的な状況を想定するという点に関連して、Lavori et al. (2008)では、「One might object that they have "potential" reality (what would the symptoms be if the patient had not died), but then they lack definiteness because we have not specified how the patient came not to die (起こり得たかもしれない状況を想定するという考えがあるかもしれないが、その場合にはその想定する状況に対する明確性を欠いている)」, 及び「One would expect the symptom status to depend on how the suicide was averted (症状は自殺をどのように回避したかに依存する)」として、「もし死亡が発生しなかったら」という仮想的な状況を正確に想定することの困難さを指摘している。なお Lavori et al. (2008)では、仮想的な状況を想定することが困難な他の事例として治療中止後のスコアに関する記載をしており、治療を継続したという仮想的な状況を想定して解析することに対する留意点として「The imputed post-discontinuation scores lack a definite unobserved underlying value. Because no one would imagine continuing the drug in the clinic in such a patient, there is no real value in knowing "what would have happened" if it had been continued (試験治療の継続が想定できない状況では試験治療を継続したときの実際の値は存在せず、補完された試験治療中止後のスコアに対しては観測されなかった潜在的な値というものが存在しない)」と指摘している。

一方で、死亡のみならず、腎不全で腎移植が適応となった場合の GFR 値[Shih et al. (1997), Myers (2000)]や喘息発作に対するレスキュー薬使用後の FEV₁ 値[Shih et al. (1997)]について、これらの例では GFR 値や FEV₁ 値自体は存在するものの、死亡後の QOL の測定値と同様の文脈で例示がされており、同様の留意点が与えられていた。

3.2.3.2.3 欠測データの分類

欠測データの解析を検討する際には、欠測メカニズムに基づく分類[Missing Completely at Random (MCAR), Missing at Random (MAR)又は Missing not at Random (MNAR) : 例えば Little et al. (2002)]が用いられることが多いが、ここでは欠測メカニズムに基づく分類以外の欠測データの分類方法に関する記載をしていた Wei (2011)を紹介する。

Wei (2011)では、薬剤固有の性質に関連するか否かで欠測データを区分し、薬剤固有の性質に関連する欠測(intrinsic missing)とそれ以外の欠測(extrinsic missing)に分類している。

- If a patient discontinues a clinical trial due to intolerance of the drug or lack of efficacy or any other drug-related reasons other than protocol- or trial-related reasons, the missing data so caused is called intrinsic missing (試験治療に不耐等の試験治療に関連する理由で臨床試験を中止した場合の欠測データを intrinsic missing と呼ぶ)
- Any missing data caused by external factors such as protocol violations, AE not related to drug, etc. are classified as extrinsic missing (試験治療と関連しない理由により生じた欠測データを extrinsic missing と呼ぶ)

この分類は推定対象とする興味のある母集団パラメータを定義するために有益であり、前者の intrinsic missing は興味のある母集団パラメータを定義する際に考慮に入れるべきであるが、後者の extrinsic missing は考慮に入れるべきでないとされている。

3.2.3.3 欠測データを最小にするための方法論と適用時の注意点

NRC (2010)の解説をしている LaVange et al. (2016)では「They laid heavy stress on the issues of design and conduct, going so far as to put ‘prevention’ before ‘treatment of missing data’ in the title. (タイトルで「欠測データの予防」を「欠測データの取り扱い」より前にすることで、試験デザイン及び実施に関する問題に重きを置いている。)」とあるように「欠測データの取り扱い」より「欠測データの予防」が重要であることが言及されている。実際、NRC (2010)には、「The panel’s first set of recommendations emphasized the role of design and trial conduct to limit the amount and impact of missing data. (最初の recommendation は、欠測データの量及び影響を最小にするための試験デザイン及び実施の役割を強調している。)」と記載されている。また欠測データを最小にする利点の1つとして、Soon (2009)では

「minimizing the extent of missing data will reduce reliance on unrealistic and unverifiable assumptions on the mechanism of missing data, and it will lead to more credible and robust conclusions. (欠測データを最小にすることで欠測メカニズムに対する非現実的で検証不可能な仮定に頼ることが少なくなるだろう。)」とあるように、欠測データを最小にすることで解析実施時の仮定に対する信頼性が高まり、信頼性のある、ロバストな結論に繋がるということが記載されている。

3.2.3.3.1 欠測データを最小にするための方法

本節では、今回の調査対象となった文献における欠測データを最小にするための方法を「試験計画」、「試験デザイン」及び「試験実施」に分けてまとめる。

(1) 試験計画時の欠測データ最小化の方法

試験計画時において、欠測データを最小にする方法として『試験実施 (screening procedures; training requirements; methods of communication; implementation of the treatment; and data collection, entry, and editing procedures) に関する文書の作成』が数報あった[Wisniewski et al. (2006), Dziura et al. (2013), Kang (2013)]. 文書作成の例として、治験実施計画書に統計解析で欠測データを適切に取り扱うことが可能などと誤解するような記載を避けること[Fleming (2011)]や現実的で許容可能な欠測データの割合を設定し、進捗状況をモニタリングすること[Little et al. (2012a, 2012b)]があった。後者の方法のように『許容可能な欠測データの割合の設定又は欠測データを起こす因子の特定』を欠測データを最小にする方法として記載していた文献は他にもあった[Dziura et al. (2013), Gemmen (2015), Kang (2013), Scharfstein et al. (2012), Witkiewitz et al. (2014)]. また患者の登録・観察・データ収集等に優れている『試験実施施設の選択』[Dziura et al. (2013), Little et al. (2012a, 2012b), NRC (2010), Scharfstein et al. (2012), Singhal et al. (2014)]や、試験実施手順(手順書やデータ収集方法の明確さ、試験における患者負担など)をテストすることを目的としたパイロット試験(pilot study)を実施することが記載されている文献もあった[Dziura et al. (2013), Kang (2013), Johnston et al. (2016), Wisniewski et al. (2006)].

(2) 試験デザイン検討時の欠測データ最小化の方法

試験デザインを検討する際の欠測データを最小にする方法として、『適切な試験デザインの選択』[Flyer et al. (2009), Liu et al. (2006)]があるが、いくつかの文献にはランダム化治療中止デザイン(randomized withdrawal design)⁶やクロスオーバーデザイン(cross-over design)のような具体的な試験デザインが記載されていた[Dziura et al. (2013), LaVange et al. (2016), Little et al. (2012a, 2012b), NRC (2010), O'Neill et al. (2012)]. その他に以下の方法が記載されていた。

- レスキュー薬の使用
 - Gewandter et al. (2014), Kurland et al. (2012), Little et al. (2012a, 2012b), NRC (2010), Tanaka et al. (2014)
- 柔軟な試験治療の設定(有害事象に応じて投与量を変更するなど)
 - Gewandter et al. (2014), Little et al. (2012a, 2012b), NRC (2010), Tanaka et al. (2014), 飛田ら(2014)
- 来院回数制限
 - Dziura et al. (2013), Gewandter et al. (2014), Hollestein et al. (2017), Singhal et al. (2014), Tanaka et al. (2014), 飛田ら(2014)
- 試験治療中止患者の継続観察

⁶ 同様の目的として試用期間(run-in period)を設定することも記載されていた[Dziura et al. (2013), Johnston et al. (2016), Little et al. (2012a, 2012b), Myers (2000), NRC (2010), Tanaka et al. (2014), 折笠 (2014)].

- Fairclough et al. (1998), Fleming (2011), Gewandter et al. (2014), LaVange et al. (2016), Li et al. (2014), Little et al. (2012b), Myers (2000), O'Neill et al. (2012), Singhal et al. (2014), Soon (2009), White et al. (2011)
- 試験参加者への経済的援助
 - Dziura et al. (2013), Gewandter et al. (2014), Hughes et al. (2012), Little et al. (2012a, 2012b), NRC (2010), Scharfstein et al. (2012), Singhal et al. (2014)
- データ収集における患者負担の軽減
 - Biswas, K. (2012), Dziura et al. (2013), Hollestein et al. (2017), Kang (2013), Little et al. (2012a, 2012b), Lewis et al. (2012), NRC (2010), Scharfstein et al. (2012), Singhal et al. (2014), 折笠(2014), 飛田ら(2014), 松井(2014)
- 適切な評価項目の選択
 - NRC (2010), LaVange et al. (2016), Li et al. (2014), Little et al. (2012a, 2012b), Soon (2009), Tanaka et al. (2014), 折笠(2014)
- 適切な試験期間の設定
 - Little et al. (2012a, 2012b), NRC (2010), 折笠(2014)

(3) 試験実施に関する欠測データ最小化の方法

試験実施前又は試験実施中における、欠測データを最小にする方法として、多かったのが『試験参加者(施設, 患者)への十分な説明』であった。その説明の対象となる内容としては以下があった。

- (同意取得時などで)試験を完了することの重要性の説明
 - Dziura et al. (2013), Fleming (2011), LaVange et al. (2016), Little et al. (2012a, 2012b, 2016), Myers (2000), NRC (2010), Scharfstein et al. (2012), Singhal et al. (2014), Soon (2009), Tanaka et al. (2014), Wittes (2009)
- データ収集方法及びデータ入力方法の説明
 - Dziura et al. (2013), Wisniewski et al. (2006)
- 欠測データが試験結果や解釈に与える影響の説明
 - Tanaka et al. (2014), 飛田ら(2014)

試験参加者(施設, 患者)への十分な説明の他に以下の方法が記載されていた。

- 欠測データの量を定期的にモニタリングし、再トレーニングの必要性を検討
 - Biswas, K. (2012), Dziura et al. (2013), Fleming (2011), Hughes et al. (2012), Li et al. (2014), Little et al. (2016), NRC (2010), Singhal et al.

- (2014), Tanaka et al. (2014), Wisniewski et al. (2006), 折笠(2014), 藤後(2014)
- 患者とのコミュニケーション(例えば, 未来院を防ぐための電話によるリマインダー)
 - Dziura et al. (2013), Gewandter et al. (2014), Kang (2013), Hughes et al. (2012), Little et al. (2012b), Scharfstein et al. (2012), Singhal et al. (2014), Soon (2009), Sullivan et al. (2017), Tanaka et al. (2014), Wisniewski et al. (2006)

3.2.3.3.2 欠測データを最小にする方法を適用する際の注意点

欠測データを最小にする方法を適用する際の注意点として, Hughes et al. (2012)では, 過去の複数の試験の結果をもとに脱落する可能性が高い患者集団を特定し, 次の試験計画に活かした実例が紹介されているが, その特定をするために使用するデータはそれらの試験で収集されたデータに制限されてしまうこと, 評価項目が主観的であるような試験では試験継続を促す方策が評価に影響を与えてしまう可能性があること, そして次の試験の対象集団や試験デザインが完全に同じであることは少なく, 過去の試験の調査結果が必ずしも次の試験の欠測データの最小化につながるとは言えないことが指摘されている. また脱落する可能性が高い患者集団を除外することで試験結果の一般化可能性(*generalizability*)を失わないようにすべきであることを指摘している. この一般化可能性に関する留意点については, Hollestein et al. (2017), Little et al. (2012a), NRC (2010)でも指摘されている.

また適切な評価項目の選択が欠測データの予防の1つであることを記載したが, これに関してはNRC (2010)では「Although it is important to define clinical endpoints that are measurable for as many participants as possible in order to reduce the impact of missing data, in doing so one must also consider the impact on the relevance of the primary endpoint. (欠測データの影響を最小にするために可能な限り多くの被験者で測定可能である評価項目を定義することが重要であるが, 主要評価項目との関連性の影響も考慮する必要がある.)」のように欠測データを最小にすることのみに重点を置く欠測データの影響が小さい評価項目を選択することが考えられるが, その評価項目が本当に評価すべき評価項目であるかは確認する必要があると注意をしている. 同様の記載は, Fleming (2011)にも「Changing the definition of a primary end point to reduce the risk for missing data would be inappropriate if such a change meaningfully compromises the endpoint's clinical relevance. (欠測データの影響を最小にするために主要評価項目を変更しても, 臨床的関連性が失われるようであればその変更は不適切であるだろう.)」とある.

最後に試験デザインの選択に関しては, LaVange et al. (2016)で *randomized withdrawal design* が紹介されたが, ランダム化された対象集団が何であるか, また得られる結果の解釈などの注意点が記載されている.

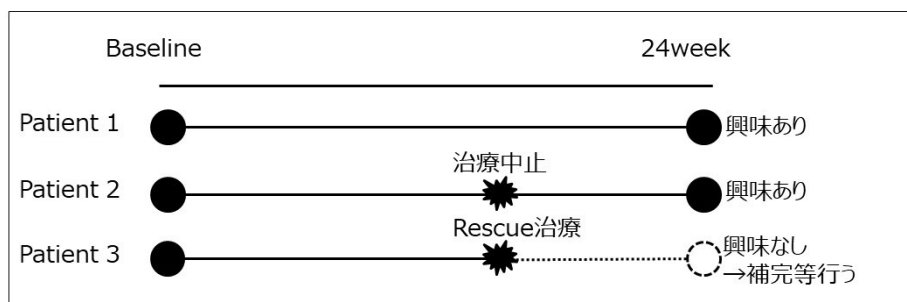
3.2.3.4 中間事象発現後のデータの取り扱い方法

「存在するが収集されなかったデータ」「存在するが収集しないデータ」「存在しないため収集できないデータ」別に, 主な調査結果をまとめた.

存在するが収集されなかったデータ，すなわち，治療中止や脱落后のデータに興味があるにも関わらず収集されなかったデータは，治療中止や脱落后に興味のあるデータが収集されている別の被験者の情報を用いて推測する方針が適切であると記載している論文が数報あった[Flyer and Hirman (2009), Witkiewitz et al. (2014)]. 例えば，Flyer and Hirman (2009) には，「If there are sufficient data collected on other subjects following discontinuation of treatment due to an AE (this is unlikely unless the protocol requires this), the subjects with data can be used to model what would have been observed for subjects lacking data. (有害事象により治療中止後のデータが収集されなかった被験者のデータを，有害事象による治療中止後のデータが収集された他の被験者のデータから推測する)」と記載されている。なお同時に，治療中止後のデータが収集されなかった被験者と治療中止後のデータが収集された被験者のデータの推移が似ているという仮定が正しい場合にものみ，適切なデータの取扱いとなるとしている。

McEvoy (2016)では，治療中止後のデータの取扱いに関連した FDA による審査の一例が紹介されていた。なお，FDA 審査時には，本試験の estimand が明確には議論されていなかった事例であるものの，同論文の introduction によれば，ITT effect に興味がある状況が想定されていたようである。なお，本試験の主要評価項目は 56 週時におけるベースラインからの体重変化率，申請者による主要な解析は，LOOT-CF (last observation on-treatment carried forward)を用いた ANCOVA であった。一方，FDA による解析は「The approach used by FDA incorporated these observations by using the experiences from RD to inform what the weight at week 56 for those with missing week 56 data may have been if their weight was measured (56 週以前において治療中止し 56 週のデータが収集されなかった被験者のデータを，56 週以前に治療中止したが 56 週のデータが収集された被験者のデータから推測する)」であった。RD とは「Retrieved Dropout」の略で，本事例の場合は，56 週以前に治療中止したが 56 週時のデータが収集された被験者と解釈できる。なお，引き続いて，56 週以前に治療中止したが 56 週時のデータが収集された被験者(RD)は，56 週以前において治療中止し 56 週のデータが収集されなかった被験者(non-RD)を代表していない可能性もあるが，試験途中で治療中止しているという点は同じであるため，特に ITT の視点においては，RD は重要な役割を果たすであろうと記載されていた。

存在するが収集しないデータ，すなわち，中間事象に対して仮想ストラテジーを想定するようなケースを考える。一部の中間事象に対して「仮想」の状況下における治療の効果に興味があると論文にて明記されていた Mehrotra (2017)を紹介する。Mehrotra (2017)にて取り扱っている estimand は「the estimand of interest is the between-treatment difference in population means of a clinical endpoint that is free from the confounding effects of "rescue" medication (e.g., HbA1c change from baseline at 24 weeks that would be observed without rescue medication regardless of whether or when the assigned treatment was discontinued). (治療中止の取扱い：治療方針ストラテジー(下図, Patient 2), rescue 治療の取扱い：仮想ストラテジー(下図, Patient 3)のケース)」である。



MMRMにおけるMARの仮定は、試験途中までの測定値の経時推移が似ていれば、その後の経時推移も似ているという仮定であり、この仮定が妥当な場面としては「This assumption may be reasonable in the context of projecting what would have been observed with complete follow-up if all patients could have remained on their assigned therapy despite experiencing inefficacy or intolerability (全員が割り付けられた治療を継続していると仮定した際に観察されるであろう結果に興味がある場合)」と記載されている。一方で、本文献にて興味のある治療効果は、特に治療中止した場合は、(仮に治療継続したということを想定しているのではなく)、その治療中止したという情報も考慮した条件下における効果である。そのため、筆者らはMMRMを用いるにあたり、「We caution that the commonly used mixed-effects model repeated measures analysis with the embedded missing at random assumption can deliver an exaggerated estimate of the aforementioned estimand of interest. This happens, in part, due to implicit imputation of an overly optimistic mean for "dropouts" (i.e., patients with missing endpoint data of interest) in the drug arm. (MMRMにより推定される治療効果が過大評価となる可能性がある)」のように注意を促している。

存在しないため収集できないデータ、特に、死亡後のデータの取扱いについては、最悪値または0で補完する方法が提案されている[Flyer and Hirman (2009), Shih (2002)]. また、NRC (2010)では、Approaches to missing data due to death (死亡により発生する欠測データに対する取り扱い)として、以下の3つが紹介されていた。

- To consider cause-specific death as a primary endpoint (e.g., death related to cardiovascular event). Death for other reasons (e.g., not related to the clinical study) may properly be treated as a censoring event (ある特定の死因を primary endpoint として取り扱う。その他の死亡は打ち切り扱い)
- To fold death into another outcome to form a composite outcome (複合アウトカムとして取り扱う)
- Approach related to principal stratification on a postrandomization event (when death is related to the outcome of interest) (principal stratification の利用)

3.2.4 文献調査結果からの考察

3.2.4.1 欠測データの定義

ICH E9 (R1) step 2 文書の用語集における欠測データの定義を基に、本調査では下記の3種類のデータに対する考え方を整理することを目的とした。

- (1) 収集すべき、かつ存在するが、収集されなかったデータ
- (2) 値自体が存在しないため収集できないデータ

(3) 値自体は存在するが、中間事象の発現により実施する解析に対して意味があるとみなせないため、収集しないデータ

まず、上記の(2)に該当するデータ(例えば、死亡後の QOL の測定値)と(3)に該当するデータ(例えば、喘息発作に対するレスキュー薬使用後の FEV₁ 値)の大きな違いは、評価項目に関する何かしらの値が実在するかという点である。つまり、(2)では被験者が死亡しているため当該被験者の QOL の測定値は実在しないが、(3)では喘息発作に対するレスキュー薬使用後の FEV₁ 値を測定することは可能であり、よって FEV₁ 値は実在する。しかしながら、レスキュー薬を使用していない状況での FEV₁ 値を評価することが目的である場合、本来の評価に使用したい FEV₁ 値は現実に存在しないこととなる。このため、調査対象文献において(2)又は(3)に該当するデータについては同様の文脈で記載されており、同様の留意点を与えられているものと考えられる。したがって、(2)又は(3)に該当するデータについては、広い意味での「存在しないデータ(nonexistent data)」として対応する必要があるだろう。これには、治療を継続している状況下での評価項目の値に興味がある状況での、副作用による試験治療中止後のデータも該当することとなるであろう。

次に、(1)に該当する、収集すべき、かつ存在するが、収集されなかったデータと、(2)又は(3)に該当する広い意味での存在しないデータとの違いであるが、(1)に該当するデータについては「収集されていたら」という状況を想定することが比較的容易である一方で、存在しないデータについては「存在していたら」という状況をかなり注意深く検討する必要があるという点である。例えば、死亡や副作用による治療中止が発現した場合、これらの事象が発生せずに治療を継続したという仮想的な状況を考えることが適切か否かは、その仮想的な状況を厳密かつ臨床的に意味のある形で記述可能であるかに依存するだろう。具体的には、死亡せずに治療を継続したという仮想的な状況を想定する際には、「どのようにして死亡を回避し、生存し続けているのか?」や「生存していた場合には当該治療以外にどのような治療が行われるのか?」などといった点について検討する必要がある。またそれらを想定する状況が臨床的に意味のあるものであるかについても考慮が必要となる。一般にはこれらの仮想的な状況を、適切かつ意味のあるものとして想定することが困難であることから、多くの文献において存在しないデータは「解析に対して意味があると考えられない」とされているのであると思われる。仮想的な状況を想定することが困難である場合には、死亡を含む複合エンドポイントとして評価項目の定義を変更することや、副作用による治療中止後のデータを収集して解析に用いる等の対応をする(つまり解析の目的を変更する)ことが必要となるかもしれない。一方で、存在しないデータについて、仮想的な状況を想定して解析することが適切と考えられる状況では、想定する状況に応じて適切に欠測データに対して提案されている解析手法を利用すればよいと考えられる。しかしながら、当該データの発生割合が大きい場合には、解析結果の安定性に懸念が生じる可能性があることから、仮想的な状況を想定して解析を行うことが本当に適切であるかについては、臨床的な妥当性の観点のみでなく、統計手法の性能の観点からの評価も必要になる場合があるだろう。

(1)の収集すべき、かつ存在するが、収集されなかったデータについては、臨床試験における発生頻度を最小とすることが望ましいと考えられる。また、欠測データが発生した場合においても、実際には存在するデータが何らかの理由で観測されなかったもの(検査の未

実施等)であるため、欠測データに対して提案されている解析手法を適切に用いることによって、多くの場合に適切な推測を行うことができると考えられる。

ここで、Wei (2011)の *intrinsic/extrinsic missing* の考え方との関連も見てみると、*extrinsic missing* は治療と関連しない要因(プロトコル違反や治療と関連しない AE など)による欠測データであるため、その多くは(1)のデータに該当するか、又は(2)若しくは(3)のデータのうち仮想的な状況を想定可能なデータに該当するものと考えられる。一方、*intrinsic missing* は治療に起因して生じる欠測データであるため、基本的に(2)又は(3)のデータに該当し、その際には仮想的な状況を想定可能かどうかについては慎重な検討が必要になるであろう。

なお、本調査では、臨床試験において広く統一的に用いられている欠測データの定義に関する記載は見つからなかった。例えば、計画していたものの収集されなかったデータすべてを欠測データと定義している文献もあれば、死亡後の QOL の測定値は存在し得ないデータであり解析に対して意味があるとは考えられないことから、欠測データではないとしている文献もあった。しかし、ここで重要なのは、欠測データの正確な定義を特定することではなく、収集されない、収集できない又は存在しないデータが発生しうる状況において如何に適切な推測を行うかという点である。そのため、欠測データの定義に合致しない場合には対応が不要なわけではなく、そのような場合においても、適切な推測を行うためには何かしらの対処が必要となることは上記の議論からも明らかであろう。

適切な推測を行うために試験計画時に慎重に検討する必要がある事項として、推定対象とする興味のあるパラメータを適切に定義することがあげられる。これは ICH E9 (R1) step 2 文書における *estimand* を定義することに相当する。*Estimand* を定義する際には、(2)の存在しないため収集できないデータを踏まえる必要があるのみでなく、中間事象の考慮の仕方によっては(3)に該当する、値自体は存在するが収集しないデータが発生しうることに留意する必要があるだろう。また、臨床試験で主要な *estimand* に加えて副次的な *estimand* を設定しているなど、複数の *estimand* が存在する場合、ある *estimand* では(3)に該当するデータが別の *estimand* では(1)に該当するデータとなる可能性があることも留意すべき点と考えられる。

最後に、(3)に関連する内容として、ある中間事象による影響を含めた治療効果に関心がない場合であっても、当該中間事象発現後のデータを収集しなくてよいとは限らない点については留意が必要である。例えば、CACE の推定の際に ITT 解析で推定された結果を利用する場合などがあげられる [Taguri et al. (2012)]. 臨床試験で収集すべきデータの範囲は、*estimand* のみならず、*estimand* に対する推定量(*estimator*)まで考慮したうえで決定する必要があるだろう。

3.2.4.2 欠測データを最小にするための方法論と適用時の注意点

欠測データは統計解析及び試験結果の解釈に影響を与える潜在的な原因の一つであるため、欠測データがないことが理想的であることは自明である。ただし、臨床試験という性質上、倫理的な理由を含め、様々な原因(試験中止、未来院、未測定など)により、解析において重要なデータが欠測してしまうことが想定される。今回の文献調査では、欠測データを最小にするための方法論を紹介したが、それらは試験計画時に検討すべき方法もあれ

ば、試験実施中に実施すべき方策もあった。これらの方策は試験結果の解釈や estimand の設定に影響を与えることなく、欠測データを最小にすることができ試験の質を保証することができるため積極的に実施すべきと考える。一方、ICH E9 (R1) step2 文書に「被験者を試験に留まらせる対策をとることはできるが、実臨床に対する試験の外的妥当性が保持されるように注意すべきである。例えば、毒性の影響を軽減するための、試験の対象集団の選択又は漸増法や併用薬の使用は、それが実臨床においても行われるものでなければ適切ではないだろう。」と記載されているように、欠測データを過剰に予防する方策により得られたデータに基づく試験結果が、本当に求めるべきものとは異なる偏った結果となる可能性を指摘している文献がいくつかあった。また ICH E9 (R1) step 2 文書では中間事象としてレスキュー薬の使用が例として記載されているが、レスキュー薬の使用は患者の試験継続のための一つの方法として考えられている。そして、試験デザインの工夫に関する方策が紹介されていたが、これに対して ICH E9 (R1) step2 文書には「特定の estimand では、導入期を用いたデザインやエンリッチメントデザイン、ランダム化治療中止デザイン又は漸増デザインといった標準的ではない試験デザインが必要となる、又はそれらデザインの恩恵を受けるかもしれない。しかし、このような標準的ではないデザインでは、その実施やその後の統計的推測に関して特別な注意が必要となるだろう。」と記載されている。これら点からも、試験計画時に試験の目的を明確にし、estimand を設定したうえで、ICH E9 (R1) step 2 文書で定義された欠測データが何であるかを特定し、estimand との一貫性及び統計的な推論の妥当性を確保するために、その欠測データを最小とする方策を臨床担当者 と検討し、実施することが今後より一層重要になると考える。

3.2.4.3 中間事象発現後のデータの取り扱い方法

本調査の対象とした文献中において、治療中止や脱落后に収集されていない、又は収集できないデータの取扱いに関する記載がされている文献は 2009 年以前にも発出されており、NRC (2010)の発出以前から、本テーマに関して議論されていたことが確認できた。Holzhauer (2015)には、NRC (2010)発出前の FDA の本テーマに対する考え方として「Prior to the FDA commissioned National Academy of Sciences report on the prevention and treatment of missing data in clinical trials, the FDA recommended to consider HbA1c data after the initiation of rescue medication as missing and suggested to impute such data, as well as other missing data, by carrying forward the last pre-rescue observation. (NRC (2010)発出前、FDA は LOCF 流の欠測補完方法を提案していた)」であったと記載されており、FDA の本テーマに関する考え方も 2000 年代に大きく変化したと考えられる。また、White (2011)にも「The main analysis should be chosen to be valid under a plausible assumption about the missing data (欠測データに関する適切な仮定を考慮し、妥当な解析を選択すべき)」と記載されているように、欠測データを適切な仮定の下で取り扱っていくことがより重要になってきたと考えられる。中間事象は変数の解釈に影響を与えるものであることから、例えば欠測データを取り扱う際には、中間事象の発現状況が類似している被験者のデータを用いてモデルを構築すること等が必要になると考える。

一方で、欠測データに対して適切な仮定を置いた解析手法が重要とはしながらも、その解析手法が非常に複雑になり、統計量を構成するためには、非常に多くの仮定が必要にも

なってくる場合がある。例えば、ICH E9 (R1) step 2 文書に中間事象に対するストラテジーの一例として記載されている、主要層ストラテジーに対応した解析を行う場合には、治療効果を推定するために、多くの検証不能な仮定を置くだけでなく、必要とされる統計解析手法が複雑になってしまう可能性が高い。今回の調査に含めた死亡例を考慮する解析(3.2.3.4 項)もその一つである。また、仮想ストラテジーに基づく解析も同様である。そのため、臨床試験の目的に応じた estimand を設定し、統計量を検討していく流れが原則である一方で、適切な推測ができる統計量を構築できるかという視点も estimand の選択の際には、実務上では、必要になってくると考えられる。また、3.4 章で触れるように、仮定に対する感度分析も重要となってくる。

3.2.5 参考文献

- JPMA (2016). 欠測のある連続量経時データに対する統計手法について.
<http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/statistics.html> [Accessed 30 May 2019]
- Little, RJA. and Rubin, DB. (2002). *Statistical analysis with missing data. 2nd edition*. John Wiley & Sons. New York.
- Taguri, M. and Chiba, Y. (2012). Instruments and bounds for causal effects under the monotonic selection assumption. *The International Journal of Biostatistics*, 8(1).
- 宮原英夫, 丹後俊郎編 (1995). 医学統計学ハンドブック. 朝倉書店
- 竹内啓 他編 (1989). 統計学辞典. 東洋経済新報社

3.3 解析対象集団の定義と役割

3.3.1 調査対象詳細

ICH E9 では「最大の解析対象集団」(以下, FAS), 「治験実施計画書に適合した対象集団」(以下, PPS)の 2 つの解析対象集団及びその役割について, 以下のように言及されている. FAS は原則ランダム化が行われた全被験者から構成される一方, PPS は FAS の被験者のうち治験実施計画書をより遵守している一部の被験者から構成される. PPS を用いることは, 治験実施計画書の基礎となっている科学的なモデルを最もよく反映することになるが, 治験実施計画書の遵守状況が試験治療及び臨床結果と関連している場合には, 生じる偏りは重大なものとなるおそれがあることが言及されている. また, 2 つの解析対象集団の役割として, 各解析対象集団の解析結果が本質的に同じ結論に達する場合, 試験結果の信用度は高くなること, 試験の特性に応じてその役割は十分慎重に考慮すべきであることが言及されている.

- 5.2.1 最大の解析対象集団
 - Intention-to-treat の原則は, ランダム化が行われた全被験者を主要な解析に含めるべきであると主張するものである.
 - 本ガイドラインでは「最大の解析対象集団」という用語を用いて, ランダム化が行われた全被験者を含める intention-to-treat という理想に可能な限り完全に近づけた解析対象集団を表すことにする.
 - ランダム化が行われた被験者を最大の解析対象集団から除外することになる状況は限られている.

- 5.2.2 治験実施計画書に適合した対象集団
 - 治験実施計画書に適合した対象集団は, 最大の解析対象集団の被験者のうち治験実施計画書をより遵守している一部であり, 以下のような基準から特徴づけられる.
 - (i) 事前に定められた最低限の試験治療規定を完了していること
 - (ii) 主要変数の測定値が利用可能であること
 - (iii) 登録基準違反などの重大な治験実施計画書違反がないこと
 - 治験実施計画書に適合した対象集団を用いることは, 解析において新しい試験治療が有効性を示す可能性を最も大きくするといってもよく, 治験実施計画書の基礎となっている科学的なモデルを最もよく反映することになる.
 - 治験実施計画書の遵守状況が試験治療及び臨床結果と関連している場合には, 生じる偏りは重大なものとなるおそれがある.

- 5.2.3 二つの異なる解析対象集団の役割
 - 最大の解析対象集団の解析と治験実施計画書に適合した対象集団の解析が本質的に同じ結論に達する場合、試験結果の信用度は高くなる。しかしその場合でも、最大の解析対象集団の相当の割合を治験実施計画書に適合した解析で除かなければならない場合には、試験全体の妥当性が疑われることに注意する必要がある。
 - 優越性試験では、最大の解析対象集団から得られる有効性の推定値は、治験実施計画書に適合した解析ほど過度に楽観的ではないため、最大の解析対象集団が(例外的な状況を除いて)主要な解析に用いられる。なぜならば、最大の解析対象集団に含まれる遵守状況の悪い被験者は、推定される試験治療の効果を一般に小さくするからである。しかし、同等性試験又は非劣性試験では、最大の解析対象集団を使用することは一般に保守的ではないため、その役割は十分慎重に考慮すべきである。

ICH E9 (R1) step 2 文書では PPS について、以下で言及されているとおり、PPS の解析の意味と役割については再考された。PPS に基づく解析では、異なる治療を受けている同様の被験者集団を比較することにならないため、治療を遵守する被験者集団における効果を推定するという目標を達成しないと、より偏りが少なく、より解釈可能な方法で対応することができるかについて考慮することとなった。主要層ストラテジーが一例として取り上げられている。

- A.1. 目的と適用範囲
 - 本補遺では、治験実施計画書に適合した対象集団の解析の意味と役割についても再考する。特に、治験実施計画書に対する違反と逸脱の影響を調べる必要がある際に、治験実施計画書に適合した対象集団による単純な解析と比較してより偏りが少なく、より解釈可能な方法で対応することができるかについて考慮する。
- A.5.3. 補足的解析
 - 本補遺で示したフレームワークにおいては、特定の治療群に割付けられた上で治験実施計画書を遵守した被験者の部分集団に基づく解析を実施することはできるが、それ自体は関心のある治療効果を明確に定義するものではない。上述したとおり、治験実施計画書を遵守した被験者集団のデータに基づく解析では、治療を遵守する被験者集団における効果を推定するという目標を達成しない。なぜなら、異なる治療を受けている同様の被験者集団を比較することにならないからである。したがって、このような解析の役割は、治験実施計画書の違反や逸

脱の程度が試験結果の信用度を損なわせていないかどうかを評価することに限定される。治験実施計画書の違反や逸脱には、中間事象として扱われるものもある。中間事象の多くが estimand の構成によって取り扱われる場合には、残された治験実施計画書の違反や逸脱の数は少なく、PPS の解析は追加の考察を加えることにならないかもしれない。

本節では以上の経緯を踏まえて、文献調査を通じて「解析対象集団」に関する議論を以下の2つの観点から整理することを目的とする。まず、ITTの原則及び解析対象集団である FAS、PPS の役割について、どのような議論が行われてきたかを整理する。次に、ICHE9 (R1) step 2 文書で役割を再考することとなった PPS の問題点及びその対処方法を整理する。

1. 各解析対象集団の役割
2. PPS の問題点及びその対処方法

3.3.2 文献調査方法の詳細

本文献調査では、2018年8月に、PubMed、TIRS 及び医中誌 Web の3つの文献検索サイトを利用し、設定した調査対象に関連する記載があることが期待される文献を検索した。本節の目的は対象集団、特に「治験実施計画書に適合した対象集団」に関する情報の収集であるため、以下のキーワードを設定した。

文献検索時のキーワード

- PubMed (検索式)
Search 1: “per protocol” [title]
Search 2: “full analysis” [title]
Search 3: “analysis population” [title]
Search 4: Intention to treat [Title] OR Intent to treat [Title] OR ITT [Title]
- TIRS
Search 5: “analysis population” or “per protocol”
- 医中誌 Web
Search 6: ((対象集団/AL) or (FAS/AL and (PPS/AL)))

上記のキーワードで選択された文献から次に示す手順により調査対象の文献を特定した。さらに、上記のキーワードで選択することができなかったが、明らかに本調査の目的と関連すると考えられる文献を調査対象の文献に追加した。

1. 文献名から明らかに目的と関連しないと考えられる文献を除外した。まず、2人が独立に文献の採否を判断し、除外することで意見が一致した文献を除外した。次に、意見が異なった場合は、互いに判断の根拠を説明し、除外することで意見が一致した文献を除外した。

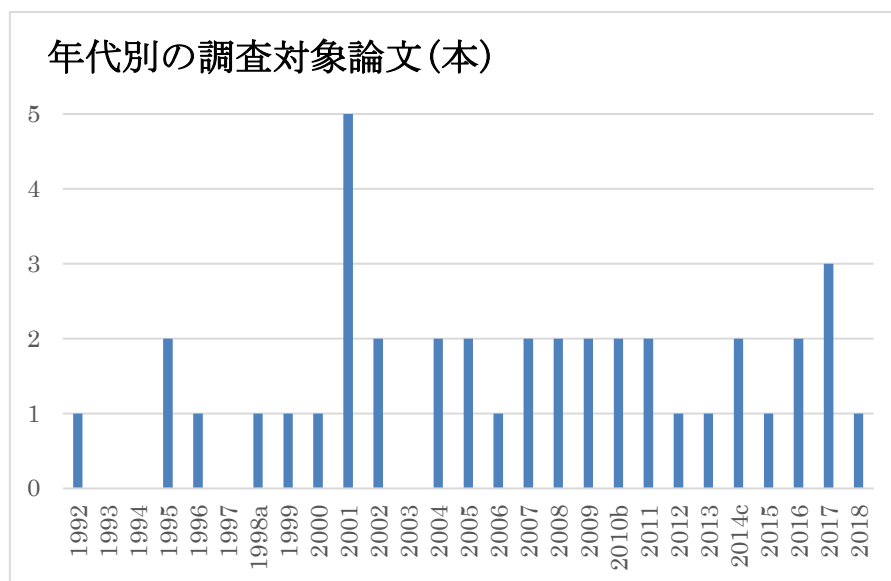
2. 1.で採用となった文献の abstract を確認し、明らかに目的と関連しないと考えられる文献を除外した。まず、2人が独立に文献の採否を判断し、除外することで意見が一致した文献を除外した。次に、意見が異なった場合は、互いに判断の根拠を説明し、除外することで意見が一致した文献を除外した。
3. 2.で採用となった文献を調査対象の文献とした。

3.3.3 文献調査結果

3.3.3.1 調査対象とした文献の内訳

キーワードにより収集された文献は合計 798 本(PubMed: 318 本, TIRS: 77 本, 医中誌: 403 本)であった。そのうち、文献名または abstract によって調査対象となった文献は合計 36 本(PubMed: 22 本, TIRS: 14 本, 医中誌: 0 本)であった。さらに、明らかに目的と関連すると考えられる文献(「その他」カテゴリー)を追加し、最終的に調査対象となった文献は合計 40 本(PubMed: 22 本, TIRS: 14 本, 医中誌: 0 本, その他: 4 本)であった。

また、調査対象の文献が発表された年代の頻度集計の結果を示す。年代の最小値は 1992 年, 最大値は 2018 年であった。年代に大きな偏りは見られなかった。



a: ICH E9 (1998), b: NRC (2010), c: ICH E9 (R1) final concept paper (2014)

3.3.3.2 各解析対象集団の役割

各解析対象集団の役割について報告された文献を調査した結果を以下に紹介する。FASに関する文献を今回の検索条件で見つけることができなかったが、ITT 及び PPS に加えて、modified ITT (mITT) 及び実際に投与された群を解析に用いる as treated (AT) が解析対象集団として文献に挙げられていた。

- ITT

(特徴)

- ランダム化が行われた全被験者から構成される ITT は、一般化の対象となる集団に近い集団であることが好ましい[Williams (2001)].

(利点)

- ランダム化が保持されているため比較の点では妥当な推測が可能である[Keene (2011)].
- 優越性試験において有効性の推定値が過度に楽観的にならない[Matilde et al. (2006)].

(欠点)

- 中間事象を含めた推定値の解釈に対する課題が存在し、例えば、脱落後のレスキュー治療がなされた場合の群間差を用いて治療の有効性を解釈することが適切でない場合がある[Keene (2011)].
- 有効性の推定値が過度に楽観的にならない、という優越性試験での利点は、非劣性試験では利点ではなくなる[Matsuyama (2010)]. ただし、この性質は同等性試験や安全性の評価でも同様であると考えられる。

- PPS

(特徴)

- 治療の逸脱によって曖昧にされない、純粋な効果に対する情報を提供することを目的としている[Leuchs (2017)].

(欠点)

- ランダム化後の被験者を除外するため、群間で背景因子がバランスしていると考えられることは出来ず、被験者数の減少による検出力の減少を招く[Ranganathan et al. (2016)]
- 同等性試験において、ITT は保守的ではなくなるため、PPS の使用が考えられるが、PPS はランダム化後の中間事象である欠測値や non-compliance によって規定され、選択バイアスを導く可能性があるため、「PP analysis is generally not causal.」である[Lou et al. (2018)]. この性質は同等性試験に限らず、他試験でも同様であると考えられる。

- mITT

(特徴)

- ITT からの様々な逸脱により定義されるため、試験間で共通の定義を持たないが、例えば **treatment** (処置がされた場合のみ mITT に含める), **post-baseline assessment** (ランダム化後のデータが存在する場合のみ mITT に含める), **combined** (前者 2 つのどちらも合致する場合に mITT に含める) 及び **other** の 4 つのカテゴリによって mITT を分類することができる [Dossing et al. (2014)].

(欠点)

- ランダム化後に被験者を除外することがバイアスの原因になりうる [Montedori et al. (2011)].

- AT

(欠点)

- ランダム化後の被験者の除外による交絡とタイプ I エラーの増大が生じる [Bang et al. (2007)].

以上より、PPS, FAS, mITT 及び AT は、ランダム化後に被験者を除外するため、ランダム化への影響が評価できない場合、同様の欠点を持つ可能性があると考えられる。なお、FAS や mITT は、ランダム化後に被験者を除外するものの、ランダム化によって得られる比較可能性を可能な限り担保できるようにしていることから、特に、PPS や AT において、この欠点は重大と考えられる。

3.3.3.3 PPS の問題点及びその対処方法

PPS の問題点はランダム化後の被験者を除外することに起因するバイアスが存在することであった。この問題点は ICH E9 及び ICH E9 (R1) step 2 文書でも言及されていた。この問題点を解決するための方法として、主要層及び **trimmed mean** を用いた方法が本調査によって確認された。これらの方法を簡単に紹介する。

Lou et al. (2018) は同等性試験において **co-primary causal estimand** を提案した。主要層の考え方を用いた **Survivor Average Causal Effect (SACE) estimand** と **protocol 遵守に関する proportion estimand** から構成される **co-primary causal estimand** により、同等性が成り立つかを判断する。この論文で定義された主要層はランダム化前の変数と考えることができるため、主要層内での治療効果は因果効果と考えることができ、「遵守状況が試験治療及び臨床結果と関連している場合に生じる偏り」は生じない。

Permutt (2017) は試験治療を中止した症例に対して試験治療中止後の潜在的なアウトカム変数を考えるのではなく、試験治療を中止したという事実を利用する **trimmed mean** を提

案した。この解析の目的は、試験治療中止を含む悪いアウトカムを一律に扱うことで、悪くないアウトカムをもつ集団における効果を推定することであり、ランダム化された症例を対象とする解析であるため、「遵守状況が試験治療及び臨床結果と関連している場合に生じる偏り」は生じない。

また、3.1.3.4 項でも言及されているとおり、主要層を用いた方法の他、g-estimation や操作変数法等による因果推論の手法を用いた解析により推定を行うことが提案されている。

3.3.4 文献調査結果からの考察

ICH E9 で PPS は「治験実施計画書の基礎となっている科学的なモデルを最もよく反映する」と言及されていた。一方で、「FAS と PPS の解析が本質的に同じ結論に達する場合、試験結果の信用度は高くなる」と言及されており、PPS の役割は FAS の結果と比較することにより試験結果の信用度を評価することであった。したがって、ICH E9 で、PPS の役割として、PPS 単体の解析により意味のある治療効果を推定することを意図しているかは不明であった。しかしながら、文献調査では「PPS の解析は治療の逸脱によって曖昧にされない、純粋な効果に対する情報を提供することを目的としている」等の治験実施計画書を遵守した場合の治療効果を推定することを PPS の解析の目的としている文献が確認された。ICH E9 (R1) step 2 文書でも「治験実施計画書を遵守した被験者集団のデータに基づく解析では、治療を遵守する被験者集団における効果を推定するという目標を達成しない」と言及しており、同様の目的を意図していると考えられる。また、「治験実施計画書を遵守した場合の治療効果」を「per-protocol effect」と定義している文献も報告されている[Swanson et al. (2015)].

しかし、治験実施計画書を遵守した場合の治療効果を推定するために、PPS の解析を実施する場合は、ランダム化後の被験者を除外することに起因するバイアスが存在するという問題があった。ICH E9 で PPS の解析は「治験実施計画書の遵守状況が試験治療及び臨床結果と関連している場合には、生じる偏りは重大なものとなるおそれがある。」、「試験治療グループ間で、これらの問題の発生頻度及び発生時間のパターンを評価することはよい対処法である。」と言及されていたが、それ以上の対処方法の記載はなかった。また、当時の文献にも偏りが存在することの報告は確認されたが、PPS に基づく解析として、この偏りに対処できる方法に関する文献は今回の検索条件では見つけることができなかった。統計的な方法以外の対処方法としては、試験の質を向上する(服薬遵守率を高くする、脱落率を低くする等)ことが考えられる。

治験実施計画書を遵守した場合の治療効果をバイアスなく推定する方法として、主要層及び trimmed mean 等を用いた方法が本調査によって確認された。これらは PPS を使用しない方法であった。主要層を用いた方法は Complier Average Causal Effect (CACE) [Lou et al. (2018)]又は Survivor Average Causal Effect (SACE) [Tchetgen (2014)]を estimand にした場合の解析方法として知られている。Trimmed mean を用いた方法は Perumutt (2016)で紹介された better-half estimand の解析方法の一種であると考えられる(JPMA (2017)も参照のこと)。もう

一つの方法として、各群の脱落状況等を踏まえて MAR と仮定できるか検討したうえで、治療中止後のデータを使わない MAR を仮定した MMRM も考えることができるかもしれない。以上の方法は、目的や必要な仮定が異なる場合があるため、試験のデザイン、目的及び estimand に応じて、その試験に適した手法を選択することが大切であると考えられる。

なお、非劣性試験・同等性試験の場合は、ICH E9 で「最大の解析対象集団を使用することは一般に保守的ではないため、その役割は十分慎重に考慮すべきである。」と言及されている。一方で、前述したとおり、PPS の解析は偏りが重大なものとなるおそれがある。したがって、どちらの解析対象集団も不適切である場合が存在すると考えられる。これに対して、例えば、Matsuyama (2010) では、非劣性試験の生存時間解析における、より適切な推定方法として、g-estimation に基づく治療効果を取りあげ、ITT や per-protocol approach と比較検討した。つまり、服薬遵守が適切な場合の治療効果を検討することで、より適切な評価をしようという考えである。一方で、非劣性試験の中には、同等の有効性を担保した上での安全性の改善を目的とした試験も存在する。そのような場合、服薬遵守が適切な場合の治療効果が目的に即しているかは注意しなければならない。

PPS の解析の実施について、ICH E9 (R1) step 2 文書で「治験実施計画書の違反や逸脱には、中間事象として扱われるものもある。」、「中間事象の多くが estimand の構成によって取り扱われる場合には、残された治験実施計画書の違反や逸脱の数は少なく、PPS の解析は追加の考察を加えることにならないかもしれない。」と言及されており、試験で設定された estimand によっては、PPS の解析を実施することは不要になる可能性がある。治験実施計画書を遵守した場合の治療効果に関する estimand の文献は多くなかったため、今後更なる調査が期待される。

3.3.5 参考文献

- Permutt, T. (2016). A taxonomy of estimands for regulatory clinical trials with discontinuations. *Statistics in Medicine*, 35(17), 2865-2875.
- Tchetgen Tchetgen, E. J. (2014). Identification and estimation of survivor average causal effects. *Statistics in Medicine*, 33(21), 3601-3628.
- JPMA (2017). 臨床試験の estimand を理解するための因果推論. Available at: <http://www.jpma.or.jp/medicine/shinyaku/tiken/allotment/estimand.html> [Accessed 30 May 2019]

3.4 安定性と感度分析

3.4.1 調査対象詳細

ICH E9 において、「安定性(ロバストネス)とは、データ、仮定及び解析方針についての様々な制限に対して全体の結論がどの程度変わり易いかに関連した概念である。安定性は、異なる仮定又は異なる解析方針に基づいて解析を行った場合でも、試験治療の効果と試験の主要な結論は大きく影響されないことを意味している。⁷⁾」と説明されている。この「どの程度変わり易いか」の部分は ICH E9 の英語版では **sensitivity** という表現が使用されている。また、ICH E9 の 3 章、5 章及び 7 章の各項目内にて、種々の論点に関して「安定性を実証する必要がある」、「安定性を示すことも重要」、「どの程度変わり易いか調べることは重要」、「どの程度変わり易いかを、更に探索する計画を立てた方が望ましい場合もあろう」等と記述されている。以上のとおり、ICH E9 では、安定性評価を検討すべきであることは示されているものの、それを評価するための解析として一般的に用いられる「感度分析(sensitivity analysis)」の項や用語はなく、どのように感度分析を行い、解釈すれば良いかは示されていない。

一方で、2010 年に EMA の CHMP から発表された **Guideline on Missing Data in Confirmatory Clinical Trial** では「**sensitivity analyses can be defined as a set of analyses where the missing data are handled in a different way as compared to the primary analysis.** (感度分析とは、欠測データを主要な解析とは異なる方法で取り扱った一連の解析である。)」及び「**Obtaining similar results from a range of methods that make similar or the same assumptions does not constitute an adequate set of sensitivity analyses.** (類似または同一の仮定による同様の結果が得られることは、適切な感度分析とはいえない。)」と記載があり、試験結果の安定性の判断基準として、以下のように試験治療の効果が著しく小さくないなど、一貫した結果が得られていることが言及されている。

- **The consistency of results from a broader range of sensitivity analyses will give greater reassurance on the robustness of the trial results.** (広範な感度分析の一貫した結果は、試験結果の安定性の確認になるだろう。)
- **if they produce inconsistent results (e.g. a markedly smaller estimate of treatment effect), their repercussions on the conclusions of the trial must be discussed.** (感度分析の結果が一貫していない場合(例えば、試験治療の効果が著しく小さい推定値)、試験の結論への影響を検討する必要がある。)

また、ほぼ同時期に米国で発表された NRC (2010)においても、「**sensitivity analyses that quantify the robustness of inferences to departures from underlying assumptions** (感度分析とは仮定からのずれに対する推測の安定性を評価するものである。)」とし、そ

⁷⁾ 英語の ICH E9 では「**Robustness implies that the treatment effect and primary conclusions of the trial are not substantially affected when analyses are carried out based on alternative assumptions or analytic approaches.**」

の原則及び方法が記載されている章が設けられており、「It is possible to assess model sensitivities of several different types, including sensitivity to: (a) distributional assumptions for the full data, (b) outlying or influential observations, and (c) assumptions about the missing data mechanism. (モデルの感度を評価するものとして、(a)完全データの分布の仮定に対する感度、(b)外れ値や影響が大きい観測値に対する感度、(c)欠測メカニズムの仮定に対する感度がある。)」と、感度分析の内容やその結果の解釈に関する記載がされている。

上記の通り、既にガイドライン等で「安定性」や「感度分析」に関する記載があるものの、感度分析の範囲や程度について、ICH E9に記載されている「安定性(ロバストネス)」を評価するための十分なガイドラインが必ずしもなかった。感度分析の現状として、ICH E9 (R1) step 2 文書のトレーニング資料 (Module 2.1 – Introduction)には、以下のように広い意味の解析として用いられていることが言及されている。

- Analyses currently labelled as ‘sensitivity analyses’ can in fact have different targets of estimation (estimands), so that consistent results between analyses should not necessarily be expected. (現在「感度分析」と分類されている解析は、実際は異なる推定の対象(estimand)であるため、一貫した結果が必ずしも期待できない。)
- Requirements that a broad range of sensitivity analyses all give consistent results might unnecessarily increase the hurdle for demonstration of therapeutic efficacy. (広範な感度分析がすべて一貫した結果を与えるという要件は、試験治療の効果を証明するハードルを不必要に高くしているかもしれない。)

そのため ICH E9 (R1) step 2 文書では、「感度分析という項目の下で、安定性(ロバストネス)の概念について拡張した議論を行う。特に、特定の解析における特定の仮定に対する推測の感度と、より広い意味での解析手法の選択に対する感度の区別をする。⁸⁾ (A.1. 目的と適用範囲)」と示され、用語集に「感度分析 Sensitivity Analysis」の定義が追加された。

感度分析 Sensitivity Analysis

主とする推定量の、モデル化における仮定からのずれとデータの限界に対する推測の安定性を調べるために実施される、同じ estimand を対象として、異なる仮定を用いた一連の解析。

この定義は「特定の解析における特定の仮定に対する推測の感度」を検討する方法であると考えられ、「より広い意味での解析手法の選択に対する感度」を検討する方法として「補足的解析 Supplementary Analysis」という新たな用語も追加された。

⁸ Finally, the concept of robustness is given expanded discussion under the heading of sensitivity analysis. In particular, a distinction is made between the sensitivity of inference to the particular assumptions of a particular analysis and the sensitivity to the choice of analytic approach more broadly.

補足的解析 Supplementary Analysis

主とする解析及び感度分析に加えて、治療効果の理解に追加の考察を与えるための解析の総称。この用語は感度分析よりも広い種類の解析を指す。

ICH E9 (R1) step 2 文書で感度分析の定義は示されたが、「同じ estimand を対象として」や「異なる仮定を用いた」をどの程度の範囲と考えるか、また補足的解析との明確な区別などは、ICH E9 (R1) step 2 文書で初めて登場した estimand や中間事象などとあわせて整理する必要がある。

これらの背景を踏まえ、本節では、文献調査を通じて以下の3つを特定し、整理する。

- 適切な安定性評価の考え方
 - 臨床試験における安定性評価として、どのような目的で実施すればよいか、また必要となる解析の範囲をどのように検討すればよいか？ ICH E9 (R1) step 2 文書で定義された感度分析のみを実施すればよいのか？
- 安定性評価の解釈の方法
 - 安定性評価のための解析の実施後に、それらの結果をどのように解釈したらよいか？
- 提示されている感度分析の手法
 - 具体的にどのような手法が提案されているか？ 主となる推定量や対処すべき課題は何か？ 評価項目の型(連続量, 2値, Time-to-eventなど)でどのような違いがあるか？

3.4.2 文献調査方法の詳細

上述した3つの項目を整理するにあたり、「感度分析(sensitivity analysis)」、ICH E9 に記載され本節の主題でもある「安定性(robustness)」,そして各キーワードを含む用語及び ICH E9 に関する文献で安定性に関する記載がある可能性がある文献を調査対象とした。

本文献調査では、2018年8月に、PubMed, TIRS 及び医中誌 Web の3つの文献検索サイトを利用し、設定した調査対象に関連する記載があることが期待される文献を検索した。

文献検索時のキーワード

- PubMed (検索式)⁹
 - Search 1: sensitivity [Title] OR sensitivity analysis [Title] OR sensitivity analyses [Title]
 - Search 2: robustness [Title] OR robust [Title] OR robust inference [Title]

⁹ MeSH : Medical Subject Headings

Search 3: (sensitivity analysis [Title] OR sensitivity analyses [Title]) AND clinical trial [MeSH Major Topic]

Search 4: (ICH E9 [Title] OR ICH-E9 [Title] OR E9 [Title]) AND (clinical trial [MeSH Major Topic] OR sensitivity [MeSH Major Topic] OR robustness [MeSH Major Topic])

なお Search 1 及び Search 2 については、対象分野が多くなることを避けるため、方法論に関する文献に注目する目的で、対象雑誌を限定する以下の検索式を追加した¹⁰。

- Search 1 で追加したジャーナルの制約
 - Biometrics [Journal] OR Biostatistics [Journal] OR BMC Medical Research Methodology [Journal] OR Contemporary Clinical Trials [Journal] OR Controlled Clinical Trials [Journal] OR Journal of Biopharmaceutical Statistics [Journal] OR Journal of the American Statistical Association [Journal] OR Pharmaceutical Statistics [Journal] OR Statistics in Biopharmaceutical Research [Journal] OR Statistics in Medicine [Journal] OR International Journal of Epidemiology [Journal]
- Search 2 で追加したジャーナル
 - Biometrics [Journal] OR Biostatistics [Journal] OR BMC Medical Research Methodology [Journal] OR Contemporary Clinical Trials [Journal] OR Controlled Clinical Trials [Journal] OR Journal of Biopharmaceutical Statistics [Journal] OR Journal of the American Statistical Association [Journal] OR Pharmaceutical Statistics [Journal] OR Statistics in Biopharmaceutical Research [Journal] OR Statistics in Medicine [Journal]
- TIRS

Search 5: タイトルに”sensitivity analysis”, ”sensitivity analyses”のいずれかを含む論文

Search 6: タイトルに”robustness”, ”robust”, ”robust inference”のいずれかを含む論文

Search 7: タイトルに”ICH E9”, ”ICH-E9”, ”E9”のいずれかを含む論文

Search 8: キーワードに”sensitivity analysis”, ”sensitivity analyses” のいずれかを含む論文
- 医中誌 (検索式)¹¹

Search 9: (安定性/TA and 臨床試験/TA) not 安定性試験/TA

Search 10: (感度分析/TA not 感度と特異度/TA) and 臨床試験/AL

¹⁰ 実際、各検索式のみで検索したところ、Search 1 では 896,55 本、Search 2 では 15,027 本の文献が選択された。

¹¹ TA: タイトル+抄録, AL: All Fields

Search 11: 統計的原則/TA

上記のキーワードで選択された文献[合計:661本(PubMed:539本, TIRS:23本, 医中誌:99本)]及び上記のキーワードを用いた検索では抽出されなかったが重要と判断した文献(TIRS:1本)のうち, 本文献調査の対象外とした下記の項目がタイトルから確認できる論文を除いた.

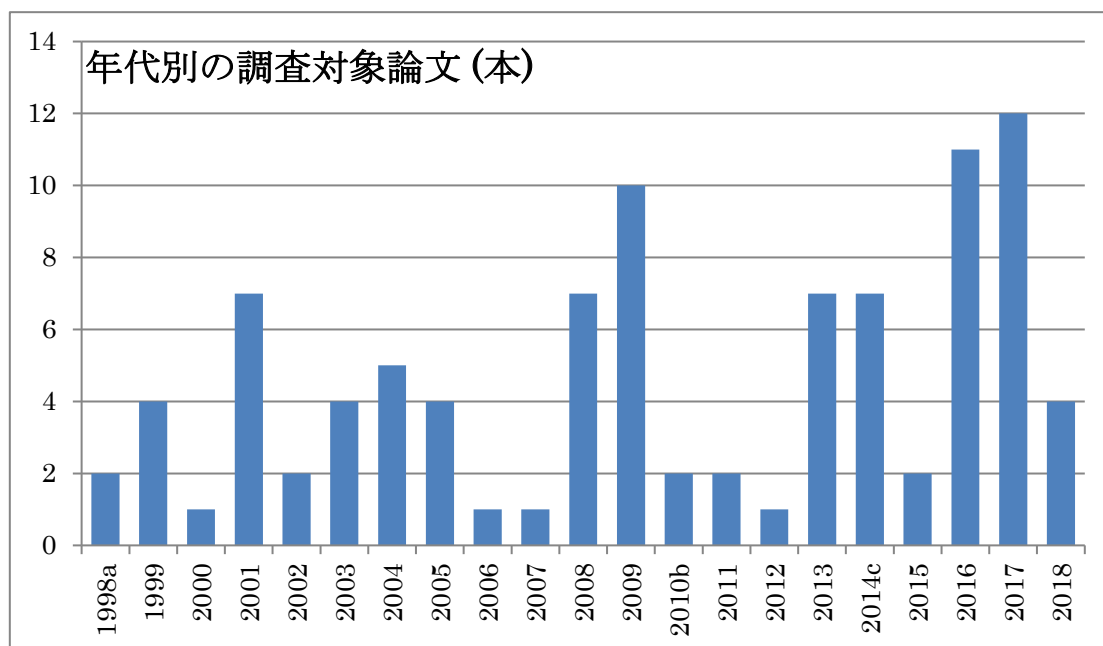
- 疫学または観察研究
- 薬物動態学に関する感度分析が記載されている文献
- 試験デザインに関する文献
- 検査の精度を評価する際の尺度の一つである感度(sensitivity)

除外されなかった文献[合計:138本(PubMed:104本, TIRS:12本, 医中誌:22本)]について, 本調査の担当者全員で文献のタイトル及び抄録を確認し, 調査対象に関連する文献である可能性ありと判断された計96本の文献(PubMed:74本, TIRS:8本, 医中誌:14本)を調査対象文献(文献一覧は6.付録(4)参照)とした.

3.4.3 文献調査結果

3.4.3.1 調査対象とした文献の内訳

調査対象の文献が発表された年代の頻度集計の結果を示す. ICH E9 が発表された1998年直後に比べ, EMEA の Points to Consider on Missing Data の発表の2001年以降は, 検証が困難な仮定を設定した欠測データの取り扱いに関する議論が多くなり, その仮定からのずれを評価する感度分析に関連する論文が増えてきている.



a: ICH E9 (1998), b: NRC (2010), c: ICH E9 (R1) final concept paper (2014)

なお調査結果では、文献内の「sensitivity analysis」または「sensitivity analyses」という記載をすべて「感度分析」としている。

3.4.3.2 適切な安定性評価の考え方

Thabane et al. (2013)及び de Souza et al. (2015)では、感度分析は「a method to determine the robustness of an assessment by examining the extent to which results are affected by changes in methods, models, values of unmeasured variables or assumptions (手法やモデル、仮定等の変化によって、結果にどの程度影響があるかを検討することで、評価の安定性を判断する手法)」と定義されており、さらに「SAs play a crucial role in assessing the robustness of the findings or conclusions (結果や結論の安定性の評価において感度分析は重要な役割を有する)」[de Souza et al. (2015)]と記載されている(SA: sensitivity analysis)。また、Akacha et al. (2017a)においても、「the role of sensitivity analysis to assess the robustness of conclusions to untestable assumptions (検証できない仮定に対する結論の安定性を評価するための感度分析の役割)」との記載があり、これらの文献において、感度分析は「結論の安定性を評価するための解析」と位置付けられている。このように、感度分析の定義とその目的に関してはいくつかの文献においても明確に示されている一方で、感度分析における問題点に関する指摘として、以下のような記載がされていた。

- The strategies of worst case and best case analysis are too extreme for a meaningful sensitivity analysis (worst/best case analysisは意味のある感度分析としては極端すぎる) [Unnebrink et al. (1999)]
- additional analyses that are conducted solely for purposes of finding a less favorable result are problematic (単により好ましくない結果を得るためだけに実施される追加の解析は問題である) [LaVange (2014)]
- Sensitivity analyses supporting the primary analysis for the primary endpoint can also be misaligned with respect to the primary objectives (主要な解析をサポートする感度分析と臨床試験の主要な目的が対応していないことがある) [Phillips et al. (2017)]
- there is commonly a lack of alignment between the reported sensitivity analysis and the corresponding estimand of interest (感度分析と興味のあるestimandが対応していない状況が一般に存在する) [Mehrotra et al. (2016)]

これらの指摘は、上記に該当するような感度分析を多数実施した場合、「result in a proliferation of P values that are difficult to interpret and ... cause confusion when clarity for decision making is most needed (解釈が困難な p 値の蔓延につながり、明確な意思決定が必要な際に混乱をもたらす)」[LaVange (2014)]となる可能性があることであり、本来の目的である「結論の安定性を評価する」ことが適切に実施できないことに加え

て、意思決定において悪影響をもたらすリスクが生じることが指摘されている。また、「Other than embarking on a quest for the optimal sensitivity analysis, it is more important for the researcher to consider a number of well-chosen sensitivity analysis routes and then to assess stability of the conclusions reached across these (最適な感度分析を求めるよりも、いくつかの適切な感度分析を選択し、それによって得られた結論の安定性を評価することが重要)」[Molenberghs (2009b)]や「the choice of sensitivity analyses should be based on the circumstances at hand (感度分析の選択は直面している状況に応じてなされるべき)」[Phillips et al. (2017)]との記載もあり、状況に応じて適切な感度分析の実施内容を検討する必要があることが示唆されている。しかしながら、具体的にどのような感度分析が安定性を評価するために適切であるかに関して言及している文献はほとんどなく、今回の調査において該当する記載がみつけれられたのは Morris et al. (2014)のみであった。

Morris et al. (2014)では、感度分析を「'Sensitivity analysis' aims to investigate whether the results of important analyses are sensitive or robust to violations of the assumptions by performing analyses addressing a specific clinical question under contrasting assumptions (感度分析は、重要な解析結果が仮定からのずれに対して安定しているかどうかを調査する目的で実施される)」とした上で、以下の3つの質問に対する回答がすべて Yes であることが意義のある感度分析であるとしている。

- 1) Does the proposed sensitivity analysis address the same question as the primary analysis? (感度分析において、主要な解析と同じ課題に対応しているか)
- 2) Is it possible for the proposed sensitivity analysis to arrive at a different conclusion to the primary analysis? (感度分析の結果が主要な解析の結果と異なる可能性があるか)
- 3) If the proposed sensitivity analysis leads to a different conclusion to the primary analysis, is there a genuine degree of uncertainty as to which will be believed? (感度分析の結果と主要な解析の結果に差異が生じた場合、いずれの結果を信じるべきかに関して不確実性が存在するか)

具体的な例として文献中に挙げられているものとして、主要な解析として ITT による解析を実施した場合に、CACE 解析や per-protocol 解析は、1)に対する回答が No であり、上記の定義では感度分析には該当しないとされている。また、主要な解析として完了した被験者に対する t 検定を実施した場合に、脱落例に関して主要な解析における仮定と同一の仮定を用いて多重代入法を利用した解析は 2)に対する回答が No となり、さらに、cluster-randomized trial において主要な解析でクラスターを考慮した解析を実施した場合に、クラスターを考慮しない解析は 3)に対する回答が No となり、これらの解析も感度分析には該当しないとされている。

一方、Leuchs et al. (2015)では、estimand の観点から感度分析を二つの種類に分類し

で考えている。その一つ目は「addresses the robustness of the primary estimation method of the primary estimand with respect to model assumptions and deviations from them (モデルの仮定や仮定からのずれに関して、主要な estimand に対する主要な推定方法の安定性を評価)」のための解析で主要な解析と同一の estimand を対象としたものであり、二つ目は「addresses alternative estimands (ie, addresses robustness with respect to the generalizability of the treatment effect, eg, to populations other than that relevant to the primary estimand) (主要な estimand とは異なる estimand を評価)」のための解析である。Leuchs et al. (2015)は、一つ目の感度分析は臨床試験の internal validity に関連し、二つ目の感度分析は臨床試験結果の credibility and external validity に関連するとともに、試験治療に関するより明確な全体像を与えるものであるとしている。

3.4.3.3 安定性評価の結果の解釈の方法

NRC (2010)では、欠測メカニズムの仮定に対する感度分析の手法として感度パラメータを用いた解析を紹介した上で、意思決定の際に利用可能と考えられる3つの方法を提示している。

- 1) One possibility is to specify a plausible region for the sensitivity parameters and report estimates of the lower and upper bounds from this range (感度パラメータについて尤もらしい範囲を定め、その中での推定値の下限と上限を報告)
- 2) A second possibility is to carry out inference under MAR and determine the set of sensitivity parameter values that would lead to overturning the conclusion from MAR (MARの仮定の下での結論と異なる結論が得られる感度パラメータの値を報告)
- 3) The third possibility is to derive a summary inference that averages over values of the sensitivity parameters in some principled fashion (各感度パラメータの値から得られる結果について何らかの方法で平均をとる)

なお、NRC (2010)においては、「there is no scientific consensus on how to synthesize information from a sensitivity analysis into a single decision about treatment effect (感度分析から得られる情報を、試験治療効果に関する一つの意思決定に統合する方法についてのコンセンサスはない)」との記載がある。

一方で、Permutt (2016)では上記のNRC (2010)における記載について、「No mention is made of significance testing. At least some of the panelists may have felt that significance testing is overemphasized in the regulatory environment, but approval of new drug applications involves some more or less binary decisions. In any case, NRC's second proposed method requires some operational definition of when conclusions are 'changed', which implies some notion of binary decision-making; and, given the well-known equivalence of confidence intervals and tests, the first method can also be

understood as referring to testing in an essentially identical way. Therefore, while we will focus on significance testing as the most natural form in which to view the problem in the regulatory environment, we believe that our ideas are consonant with those of the NRC panel. (NRC (2010)では検定に関する言及はないものの、提案されている1つ目と2つ目の方法は検定と密接な関連があるものであり、検定の結果に強い関心をもつという規制当局の考えとNRC panelの考えに大差はない)と感度分析の結果に基づく意思決定を行う際には、有意性検定の結果に注目するとしている。

de Souza et al. (2015)はランダム化比較試験における感度分析に関する文献であり、感度分析の解釈の方法について以下の見解を述べている。

- When the primary analysis and all SAs provide similar estimates of the magnitude and direction of treatment effect, similar P values, and similar widths of CIs, one can be reasonably confident that the finding of the trial is robust. (主要な解析と感度分析で類似した結果が得られた場合、臨床試験から得られた結論が安定していると確信できる)
- When the primary analysis and SAs differ, the situation is more difficult. ... The assessment of robustness is often based on the magnitude, direction, or statistical significance of the estimates. A sensitivity analysis cannot be used to choose an alternative conclusion to a study. Authors should state the conclusion on the basis of the primary analysis plan and use the presentation of one or more sensitivity analyses as a way of assessing the robustness and validity of the primary analysis. If the SA suggests that the primary analysis is not robust, it may point to the need for future research that might address the source of the inconsistency. (主要な解析と感度分析で異なる結果が得られた場合、臨床試験の結論は主要な解析の結果を基にして述べるべきであり、主要な解析の安定性と妥当性を評価するための方法として感度分析の結果を提示すべきであって、感度分析の結果によって臨床試験の結論を変えるべきではない)

de Souza et al. (2015)では、安定性を評価するための判断基準(効果の大きさ・向き、統計的有意性)については明確な主張はしていないものの、試験の結論は計画された主要な解析に基づいて主張するべきであって、感度分析の結果を試験の結論を変更するために利用することはできないとしている。

3.4.3.4 感度分析の方法

本項では治療の有効性を評価するランダム化比較試験における感度分析に関する文献の調査結果から、感度分析の方法に関してまとめる。

感度分析はICH E9 (R1) step 2 文書で「主とする推定量の、モデル化における仮定からのずれとデータの限界に対する推測の安定性を調べるために実施される、同じ estimand を

対象として、異なる仮定を用いた一連の解析。」と定義されている。どのような状況でどのような限界を課題とした感度分析かという観点で、調査対象の文献において特に多く見られたのは、経時観察研究における *monotone missing*、規定時点の評価項目の欠測、事後的に集団が特定される状況の *selection bias*、*time-to-event data* における *informative censoring* であった。

- 経時観察研究における *monotone missing*

一定期間観察が繰り返され、かつ途中時点の観察結果にも一定の意味がある状況で脱落が中間事象として生じ、データが *monotone missing* となる課題を想定した感度分析を扱っている文献を本項にまとめた。脱落は具体的には治療の中止が見られていた(試験治療中止後に効果が継続しないことが明確で中止後の観察が意味を持たない疾患領域で、治療の中止と観察の中止を分けて考えられていない模様)。想定されている主要な解析は一部示されていない文献があったが、示されている場合は *complete case analysis* や、*MAR*、*MCAR* を想定した解析であった。

評価項目のデータの型が連続値の場合を扱っている文献として Scharfstein et al. (2018), Daniels et al. (2000), Ayele et al. (2014), Curran et al. (2004), Matsuyama (2003), Cresswell et al. (2014), Moreno-Betancur et al. (2016), Scharfstein et al. (2009)が見られた。取り扱われていた感度分析の方法は、*pattern mixture model* 又は何らかのモデル化を伴う *imputation* [Cresswell et al. (2014), Curran et al. (2004), Daniels et al. (2000), Moreno-Betancur et al. (2016), Scharfstein et al. (2009), Scharfstein et al. (2018)], *placebo or control based imputation* [Ayele et al. (2014)], *Inverse Probability of Censoring Weighted (IPCW)* [Matsuyama (2003)]法であった。欠測過程をモデル化する際に *sensitivity parameter* を設定し、*plausible* な範囲で動かした結果を示しているものも多く見られた。

評価項目が 2 値等のカテゴリカルデータの場合を扱っている文献として Tang (2018), Kaciroti et al. (2009), Kenward (1998)が見られた。取り扱われていた感度分析の方法はそれぞれ、*control-based imputation*, *pattern mixture model*, *selection model* であった。

評価項目が *recurrent data* の場合を扱っている文献として Akacha et al. (2016), Gao et al. (2017), Keene et al. (2014)が見られた。取り扱われていた感度分析の方法はいずれも *control-based imputation* であった。

1 件のみ *non-monotone missing* を取り上げている文献があった [Minini et al. (2004a)]。取り扱われていた感度分析の方法は *pattern mixture model* であった。

- 規定時点の評価項目の欠測

ランダム化後のある規定の時期に評価項目を観察する状況で、データが欠測となる課題を想定した感度分析を扱っている文献を本項にまとめた。評価項目のデータの型が連続値

の場合を扱っている文献として Andrea et al. (2001), Minini et al. (2004b)が、2 値の場合を扱っている文献として Li et al. (2017), Liublinska et al. (2014), Hollis (2002)が見られた。想定されている主要な解析は一部示されていない文献があったが、示されている場合は complete case analysis であった。感度分析の方法は、いずれの文献でも sensitivity parameter が用いられていた。Li et al. (2017)は imputation で、Minini et al. (2004b)は selection model で sensitivity parameter を用いていた。

- 事後的に集団が特定される状況の selection bias

ランダム化後、評価項目の観察までに特定の条件を満たしている集団が興味の対象となる状況で、除外の原因となる中間事象の発生を課題とした感度分析を扱っている文献を本項にまとめた。

Shepherd et al. (2011)では評価項目は time to event, 中間事象は死亡, 主要な解析は survival causal effect で, sensitivity parameter を用いた感度分析が取り上げられていた。

Fox et al. (2005)では評価項目は 2 値, 中間事象には評価項目の欠測, 併用治療薬の使用, 組み入れ基準違反を想定した。これらによるバイアスを除いた結果を得るため, 欠測については共変量を含むロジスティックモデルで推定した成功確率に, 併用治療薬の使用と組み入れ基準違反については先行研究の結果に基づくデータの reconstruction を用いた感度分析が取り上げられていた。

Gilbert et al. (2013) では評価項目は time to event, 中間事象はワクチン治療の不完遂又は完遂以前のイベント発現, 主要な解析は complete case analysis で, selection bias が存在する状況に相当する複数の sensitivity parameter を用いた感度分析が取り上げられていた。

Mehrotra et al. (2017) では評価項目は連続値, 中間事象はレスキュー薬の使用, 主要な解析は MAR-based imputation で, sensitivity parameter を用いた imputation が感度分析として取り上げられていた。

Lu et al. (2013) では試験中に感染イベントが生じた集団を興味の対象とするワクチンの試験や, 試験中にがんと診断された集団を興味の対象とするがん予防の試験を想定し, 主要層を用いた感度分析が取り上げられていた。

Egleston et al. (2010) は禁煙を評価する試験での介入の不遵守を中間事象とし, 主要な解析は observed case analysis で, sensitivity parameter を用いた感度分析が取り上げられていた。

- Time-to-event dataにおける informative censoring

評価項目が time to event の研究で informative censoring を想定した感度分析である。主要な解析では打ち切りを non-informative として扱う解析が想定される。調査対象の文

献内で使用されていた感度分析の方法は、欠測以降の time to event データの補完又は sensitivity parameter の利用により効果の推定値を得るもの [Huang et al. (2008), Lipkovich et al. (2016), Liu (2017), Ruan et al. (2008), Siannis et al. (2005), Zhao et al. (2014), Zhang (2009)], index of local sensitivity to nonignorability を用いて推定値付近の sensitivity を定量化する方法 [Liu et al. (2012)] であった。

- その他

上記の他では、以下の課題を対象とした感度分析が確認された。

(欠測関係)

- Competing risks の cause of failure の欠測 [Moreno-Betancur et al. (2015)], matched-pair design の pair 内で missing mechanism が identical でない可能性 [Imai et al. (2018)]

(治療関係)

- 群間で治療の switching が存在する状況で ITT の主要な解析に対し as-treated 解析を考慮 [Wright et al. (2003)]
- Withdraw する人としめない人で服薬遵守率が異なる可能性を考慮 [Salim et al. (2008)]

(中間事象と関連しないもの)

- 不適切なランダム化 [Marcus (2001)], 調査対象にしていない共変量の imbalance を考慮 [Lin et al. (2013)], local influence [Lee et al. (2004), Shen et al. (2006), Verbeke et al. (2001a)], PFS の評価の subjectivity, 評価タイミング [Bhattacharya et al. (2009)]

3.4.4 文献調査結果からの考察

3.4.4.1 適切な安定性評価の考え方

ICH E9 において、「安定性とは、データ、仮定及び解析方針についての様々な制限に対して全体の結論がどの程度変わり易いかに関連した概念」とされており、また「安定性は、異なる仮定又は異なる解析方針に基づいて解析を行った場合でも、試験治療の効果と試験の主要な結論は大きく影響されないことを意味している」とある。日本製薬工業協会の部会資料(1999年3月)では、感度分析を「結論の安定性を評価するため」のすべての解析としていたが、今回の文献調査においても同様の定義を用いている文献が存在していた [Thabane et al. (2013), de Souza et al. (2015)]. さらに、「there is commonly a lack of alignment between the reported sensitivity analysis and the corresponding estimand of interest (感度分析と興味のある estimand が対応していない状況が一般に存在する)」 [Mehrotra et al. (2016)]等の指摘があることを踏まえると、感度分析の定義を明確には記載していないものの、同様の考えに基づいて解析を実施している事例は多いものと考えら

れる。一方、ICH E9 (R1) step 2 文書における感度分析の定義は、「主とする推定量の、モデル化における仮定からのずれとデータの限界に対する推測の安定性を調べるために実施される、同じ estimand を対象として、異なる仮定を用いた一連の解析」とされており、前述の感度分析の定義と比較して限定されたものになっていると考えることができる。またこれは、Leuchs et al. (2015)の感度分析の分類における一つの種類である、臨床試験の internal validity に関連する感度分析に一致するものと思われる。なお、今回調査した範囲においては、明確に ICH E9 (R1) step 2 文書における感度分析と同一の定義を用いている文献は見つからなかった。これは、ICH E9 (R1) step 2 文書で estimand という概念を新たに明記したためと考えられる。

次に、臨床試験において安定性を評価すべき範囲について考察する。ICH E9 における安定性はあくまでも「全体の結論」に関する概念として記載されているが、ICH E9 においても他の項では「推定値の安定性」や「対応する解析結果の安定性」と記載されている箇所も存在し、また「Sensitivity analyses should also be incorporated into analysis plans to specifically assess the robustness of results to key assumptions of the primary analysis (主要な解析の重要な仮定に対する結果の安定性を評価するために、感度分析を解析計画に含めるべき)」[Mallinckrodt et al. (2009)]や「Our guidance in this situation is to remember that the goal of an SA is not to select the best results. Rather, the aim is to assess the robustness or consistency of the results (感度分析の目的は最良の結果を選択することではなく、結果の安定性や一貫性を評価すること)」[de Souza et al. (2015)]とあるように、特定の解析の仮定や結果に関する文脈で安定性という用語を使用している場合もあるなど、その用語の利用のされ方は幅広い。しかしながら、臨床試験で実施する一連の解析を計画する際には、当該臨床試験の目的に対応する全体の結論に関する評価を可能な限り明確に得ることができるような計画とすることが望ましいことを踏まえれば、感度分析自体をどのように定義するかに関わらず、安定性を評価する範囲に関しては必ずしも主要な解析の仮定や結果に限定するべきではないだろう。

例えば、Morris et al. (2014)では、評価項目の定義が明確に定義されていない場合に、評価項目の定義を変更して実施する解析を感度分析と呼んでおり、また Thabane et al. (2013)では、部分集団間での結果の類似性を確認するための解析を subgroup "sensitivity" analyses と呼んでいる。こういった解析は主要な解析と異なる estimand に対する解析であるため、ICH E9 (R1) step 2 文書での感度分析には該当せず、補足的解析として分類されるものの、「結論の安定性を評価するため」には必要な解析である状況が存在すると考えられる。このことは、ICH E9 (R1) step 2 文書における「異なる定義の反応例に基づく解析が、その定義に対して結果が安定しているかを評価するために実施されるかもしれない。補足的解析の必要性と有用性は、それぞれの試験において判断すべきである」との記載につながるところであろう。

また、ICH E9 (R1) step 2 文書での感度分析は、主とする推定量の、モデル化における

仮定からのずれとデータの限界に対する推測の安定性を調べるために実施されるものとされているが、この感度分析の定義に該当する解析の中においても、感度分析の対象とする仮定やデータの限界の内容によって、安定性の評価に必要となる解析の範囲は異なる可能性がある。例えば、レスキュー薬の使用という中間事象に対して仮想ストラテジーを適用し、レスキュー薬を使用しなかったという状況での治療効果を推定する状況での解析手法の一つとして **reference-based imputation** を利用することが考えられるが[Fujiwara (2017)], この際にレスキュー薬使用後における「レスキュー薬を使用しなかったと想定した場合」の測定値の推移に関する仮定からのずれに対しては、特にレスキュー薬の使用例が多いと想定される場合に十分に安定性を評価する計画を立てる必要があるだろう。一方、ICH E9 (R1) step 2 文書での定義における欠測データは、基本的にその発生頻度を最小とすることが望ましいと考えられるものであることから(3.2.4.1 項参照)、特に欠測データの発生頻度が低いと想定される場合に、欠測メカニズムの仮定からのずれに関する安定性の評価は限定的な内容で十分な状況も考えられるだろう。

最後に、安定性を評価するために実施する解析は、その目的の達成に適切に貢献しうるものに限定すべきである。無計画に多数の解析を実施し、その結果を単純に併記することは、意思決定において無用な混乱を生じさせる危険性があることに留意する必要があるだろう。例えば、臨床的な面からの根拠が弱い極端な仮定を用いた解析や、主要な解析と手法は異なるものの同様の仮定を用いた解析等では、試験の結論の安定性評価が適切に実施できないことに加えて、試験結果の解釈、さらにはその後の意思決定をも困難にしてしまう可能性がある。ここで、具体的にどのような感度分析や補足的解析が必要であり、有用であるかについては、使用する統計手法の性質や治療効果の評価方法に関連する事項の臨床的な観点からの検討に加え、Morris et al. (2014)で提案された3つの質問に当てはまるか否かを考えることにも意義があると考えられる。ただし、Morris et al. (2014)の1つ目の質問の中で記載されている「question」の定義に関しては曖昧さもあると考えられるため、この点に関しては注意をしようでの検討が必要であろう。

3.4.4.2 安定性評価の結果の解釈の方法

安定性が「全体の結論がどの程度変わり易いかに関連した概念」であるとされていることを踏まえると、全体の結論自体は安定性の評価を実施しなくとも得られているものであり、安定性の評価を加えることで、導かれた全体の結論の強さや妥当性を示すことになると考えられる。これは de Souza et al. (2015)において、「A sensitivity analysis cannot be used to choose an alternative conclusion to a study. Authors should state the conclusion on the basis of the primary analysis plan and use the presentation of one or more sensitivity analyses as a way of assessing the robustness and validity of the primary analysis. (臨床試験の結論は主要な解析の結果を基にして述べるべきであり、主要な解析の安定性と妥当性を評価するための方法として感度分析の結果を提示すべきであって、感

度分析の結果によって臨床試験の結論を変えるべきではない」とされていることと一致している。しかしながら、試験の結論自体は変わらなくとも、その結論の強さや妥当性に応じて、試験の結論を受けての意思決定や取りうる行動に影響を与える可能性があるだろう。

NRC (2010)や Permutt (2016)は感度分析と意思決定の関係を論じており、NRC (2010)では感度分析により得られた情報を統合し一つの意思決定につなげる方法のコンセンサスはないとしている一方で、Permutt (2016)では有意性検定の結果に基づいて意思決定することを推奨している。なお、NRC (2010)及び Permutt (2016)ともに、欠測メカニズムの仮定に対する感度分析に限定した議論であることには注意が必要であろう。欠測メカニズムの仮定は検証不可能な仮定であり、ICH E9 (R1) step 2 文書でも「欠測データは感度分析において特に注意が必要」とされていることから、当該感度分析から得られる結果が意思決定に与える影響の大きさは、他の感度分析により得られる結果とは異なる可能性がある。

今回の調査では、より一般的な、欠測メカニズムの仮定に限定しない感度分析の結果と意思決定との関連について議論をしている文献は見つからなかった。しかし、試験の結論を踏まえて意思決定を行う際には、主要な解析や感度分析で利用した各仮定の尤もらしさに加え、意思決定の内容とその重要性、また、必要に応じて当該試験での安全性に関する成績、他の臨床試験の成績、対象疾患の重篤度や既存治療の有無等、種々の情報を基にした総合的な判断が必要になると考えられるため、望ましい意思決定に必要とされる結論の安定性の度合いは状況に応じて異なると考えられる。そのため、普遍的に適用可能な意思決定の方法を示すことは困難であり、個別の状況に応じて判断をしなければならないというのが実情であろう。Permutt (2016)においても、感度分析に関するスポンサーと規制当局による事前の議論と合意の重要性が示唆されているが、安定性を評価すべき範囲及びその具体的な方法、並びに結果の解釈の方法に関して、事前に規制当局と議論し合意を得ることが有益な状況もあると考えられる。

3.4.4.3 感度分析の方法

ICH E9 の安定性の定義において対処すべき事項は「データ、仮定及び解析方針についての様々な制限」と記述されており、ICH E9 (R1) step2 文書では感度分析に関して「規制上の関心としては特定の解析に関する仮定からのずれ及びデータの限界に対する感度に焦点を絞ることができる」と記述されている。ここで、データの限界と解析に関する仮定からのずれは独立したものではないことに留意が必要である。例えば、経時観察研究における **monotone missing** では、予定していたすべての観察結果が不足しているという点でデータの限界であり、主要な解析で **MAR** 又は **MCAR** を仮定した解析をおこなっている場合、当該仮定からのずれにより結果や結論がどれだけ影響を受けるかが感度分析の焦点となる。死亡、治療不遵守、規定時点の評価項目の **missing** 等による **selection bias** では、観察時期、

それまでの治療状況等による規定を満たした観察結果が不足しているという点でデータの限界であり、主要な解析では観察された症例のみを対象としてもバイアスが無いと仮定している場合、当該仮定からのずれにより結果や結論がどれだけ影響を受けるかが感度分析の焦点となる。試験により感度分析でどのような課題に対処すべきであるかは、当該試験の疾患領域や治療特性等はどうなものか、どのような評価項目か、どの中間事象がどのようなデータの限界をもたらすか、どのような主要な解析か、その主要な解析はどのような想定に基づいているか等を順に分析していく必要があると思われる。これは ICH E9 (R1) step2 文書で示されているフレームワークと近い。ただし、ICH E9 (R1) step2 文書の感度分析の定義の「モデル化における仮定からのずれとデータの限界に対する推測の安定性を調べるため」の記述は、中間事象に起因した課題と限定されてはいない。今回調査した文献でも、中間事象に起因しない課題に対処した感度分析が確認された(例：不適切なランダム化、調査対象にしていない共変量、PFS の評価の subjectivity 及び評価のタイミング)。中間事象はデータの限界をもたらすが、データの限界は中間事象に起因するものに限らない点に、一定の注意が必要と思われる。

3.4.5 参考文献

- European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP). Guideline on missing data in confirmatory clinical trials. EMA/CPMP/EWP/1776/99 Rev. 1. (http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf) [Accessed 30 May 2019]
- Fujiwara, M. (2017). 6. 感度分析 2 (Reference-based imputation). 日本製薬工業協会シンポジウム 臨床試験の estimand に対する最近の議論と、欠測のあるデータに対する基本的解析手法について. (<http://www.jpma.or.jp/medicine/shinyaku/tiken/symposium/>) [Accessed 30 May 2019]
- ICH E9 (R1) Expert Working Group (2018). Draft (Step 2) guideline ICH E9 (R1) Estimands and Sensitivity Analysis in Clinical Trials. Training material. (https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/E9-R1EWG_Modules1-3_Step2_COMPILATION_TrainingMaterial_2018_0703.pdf) [Accessed 30 May 2019]
- JPMA (1999). 「臨床試験のための統計的原則」に関する問題点の解説. (http://www.jpma.or.jp/information/evaluation/publishing_center/) [Accessed 30 May 2019]
- National Research Council. (2010). The prevention and treatment of missing data

in clinical trials. National Academy Press.

4. 全体を通しての考察とまとめ

ICH E9 (R1)の final concept paper では、以下の2点が Issues to be Resolved として記載された。

- Incorrect choice of estimand and unclear definitions for estimands lead to problems in relation to trial design, conduct and analysis and introduce potential for inconsistencies in inference and decision making.
- Absence of a framework for planning, conducting and interpreting sensitivity analyses may lead to inconsistencies in inference and decision making within and between regulatory regions.

1998年にICH E9で臨床試験における統計的原則が提示されて以降、臨床試験における統計的課題に関するより詳細な事項を記述した種々のガイダンス・ガイドラインが日米欧の規制当局から発出されてきているが、estimandを適切に選択するガイダンスはなく、また、感度分析で何をすべきか、範囲や程度について、必ずしも十分なガイダンスはなかったとされている[Tsuchiya (2014)]。実際、estimandと感度分析に関する明確なガイダンスが存在しないことによる影響をうかがい知ることができる事例として、COPDの治療薬である anoro ellipta の FDA statistical review の事例[Miyazaki (2017)]がある。Anoro elliptaの検証試験の主要な解析において、試験治療中止後の FEV₁ 値に対して MAR を仮定した mixed effect model を適用していることに対し、FDA statistical review では、

- If the estimand of interest is the hypothetical effectiveness of the assigned treatment if all patients could tolerate and adhere to the combination product, then the estimated treatment effect from the MMRM model may provide a reliable estimate of the truth. (すべての患者が割り付けられた治療を継続できたとしたときの仮想的な有用性が興味のある estimand である場合、MMRM で推定される試験治療効果は信頼のできる値を提供するかもしれない)
- if the estimand of interest is the effectiveness of the assigned treatment in all randomized participants, at real world achievable adherence and tolerability, the MMRM model likely does not produce a reliable estimate of the truth. (興味のある estimand が割り付けられた治療の有用性である場合、MMRM で推定される治療効果は信頼できるものではないだろう)

としており、またこの結果として「we gave importance to a sensitivity analysis that multiply imputed missing data under the assumption that dropouts on all treatment arms would have had outcomes similar to those that were observed among completers in the control group (試験治療群で中止した被験者は対照薬群で完了した被験者と同様のアウトカムを有するという仮定を用いた感度分析を重視)」と記載している。これらの記述からは、当該試験での主要な estimand が明確でなく、また estimand に関するスポンサーと規制当局間での事前の合意がなかったこと、さらに重要視すべき感度分析の結果も事前に明

確になっていなかったことが読み取れ、薬剤の承認可否という意味決定に影響を与えかねないと思われるこれらの事項が承認審査の段階になって顕在化したものと推測される。こうした事例の再発を回避するためにも、ICH E9 (R1)において estimand と感度分析に関する明確なガイダンスが記述されることが望まれる。

一方、ICH E9 及び ICH E9 (R1)は統計的原則について記述されるものであることから、その内容を理解したうえで臨床試験を適切に計画・実施するためには、ガイドラインの記載事項の背景にある知識及び情報が重要であると考えられる。このような理由から、タスクフォース 4 Estimand/因果推論チームの 2018 年度の活動では、estimand と感度分析に関する文献調査を実施し、その結果及び結果から得られた考察を本レポートにまとめた。文献調査の対象とした「ITT の原則に基づく治療方針の効果とその他の治療効果」、「データの取扱い及び欠測データに含めて検討されている問題」及び「解析対象集団の定義と役割」は estimand との関連が強い事項であり、また「安定性と感度分析」は感度分析に対応するものである。

ここでは、本文献調査を通して、4つの調査対象から横断的に得られた以下の3点の知見についてまとめることとする。

1点目は、推定したい治療効果と欠測データ、及び解析対象集団の選択の3つの問題を混同せずに、それぞれ整理して検討する必要があるという点である。ここで特に注目すべきは、ICH E9 の ITT の原則では、解析対象集団及び ICH E9 (R1) step 2 文書に記載されている estimand の構成要素のうち中間事象の取扱いは定めているが、他の構成要素である対象集団、評価項目及び変数の要約については定めていない点である。ITT と effectiveness を紐付けて議論している文献も存在したが、3.1.4.1 項での記載の通り、非常に統制され、服薬遵守状況も非常に高い臨床試験が実施された場合、ITT の原則に基づいて推定される治療効果は efficacy に近いものと考えられるであろう。この点に関しては Lavori et al. (2008) でも、「Instead of trying to eke out inference related to “what if” questions from the observed data, we investigate ways to reframe the treatment comparison a priori (in the design) so that the resulting design provides a better fit to the investigator's research goals when the resulting data are analyzed by ITT (観測されたデータから“もしこうだったら”という課題に対応しようとするのではなく、得られたデータを ITT で解析した際に、研究の目的に合致する回答を得られるような試験デザインを事前に検討すべき)」や「the idea is to convert trials whose ITT analysis is unsatisfactory into new designs that are suited to ITT (ITT 解析では試験の目的と一致しない臨床試験を、ITT での解析に適したデザインに変更する)」とあり、ITT の原則に従った解析によって興味のある治療効果を推定するために、試験デザインの変更により対応することが提案されている。このことから、「特定の試験における治療方針の効果」は、対象集団、評価項目、その他の試験デザイン及び実施計画への遵守状況によって変わりうるものであり、「日常診療における治療方針の効果」とは必ずしも一致しないことを認識したうえで、これが試験の目的に合致し、適用可能であ

るかという観点からも、事前に検討することが重要であると考えられる。次に、推定したい治療効果と欠測データの関連に関してであるが、ICH E9 (R1) step 2 文書の欠測データの定義から、estimand の選択が欠測データに該当するデータの範囲に影響を与えることとなる。また一方で、これまで欠測データの問題の一部として考えられてきた、3.2.4.1 項で記載した広い意味での「存在しないデータ」への対処方法について深く検討することが、estimand の適切な選択に寄与すると考えられる。このような関連を整理することで、欠測データの問題を取り上げている文献において記載されている内容を適切な estimand の選択に生かすことができると考えられる。最後に、解析対象集団については、estimand の構成要素の一つとされている対象集団(科学的疑問の対象となる患者)とは異なることに注意が必要であるとともに、estimand を推定するためにどのような estimator を利用するかによって適切に設定していく必要があるだろう。これは、同一の estimand を対象としている場合であっても、利用する解析手法、つまり estimator によって、解析に必要なデータが異なる可能性があるからである。

2 点目は、一般化可能性についてである。検証試験においては基本的に、比較可能性を担保した上で治療効果に関する検証を行い、またこれによって治療効果が検証された場合に、それが意図している集団に対して一般化可能であるかが検討される。このため、検証試験では、比較可能性と一般化可能性を可能な限り高めるべきであることに疑問の余地はないであろう。しかし、3.2.4.2 項で記載されているように、中間事象が発現しにくい患者層のみを臨床試験に組み入れたり、服薬遵守状況を非常に高く保つための施策を実施するなど、比較可能性を高めるための方策が一般化可能性に影響を与える可能性がある点に注意が必要である。また、3.1.4.1 項や 3.1.4.3 項で記載されているように、特定の estimand が一般化可能性の面で常に優れているかは明確ではない。例えば、ITT の原則に基づいて推定される治療効果に対しても一般化が困難という意見がある一方で、efficacy を推定するための手法として提案されている主要層解析に関しても、ICH E9 (R1) step 2 文書では主要層ストラテジーについて「集団を事前に特定することが困難である」ことを指摘している。これは、主要層解析により得られた結果がどのような集団に一般化可能であるかに関しても不明瞭となる可能性を示唆していると考えられる。「strictly defined treatments on a relatively arbitrary class of patients (比較的あいまいな患者集団における、厳密に定義された治療)」[Schwartz et al. (1967)]が医薬品の承認審査において一定の意義を有するか否かについては、今後の議論が待たれるところであろう。最後に、一般化可能性に関しては、検証試験での対象集団に加えて、部分集団解析の結果や他の臨床試験結果、対象疾患や試験治療の特徴なども踏まえて検討されるべき事項である。そのため、検証試験で担保すべき一般化可能性の程度はそれぞれの状況によって個別に検討する必要がある。このことから、estimand の選択に際しては、臨床担当者との協議に加えて、薬事担当者や規制当局との協議も重要となってくるであろう。

3 点目は、estimand に対する臨床的な興味と estimator に対する統計的な性能面からの

評価に関してである。ICH E9 (R1) step 2 文書では、「特定の estimand に基づく推測は、データの限界と、主とする推定量に対する統計モデルで用いる仮定からのずれに対して安定しているべきである。この安定性は感度分析によって評価される。」と記載されており、感度分析によって安定性を評価することを求めている。このため、どれほど臨床的な面から興味が高い estimand であったとしても、その estimand に対応して安定した推定が可能な estimator が構成できない場合には、当該試験で選択する estimand を再検討する必要があるかもしれない。このことは、ICH E9 (R1) step 2 文書においても「Estimand の構成には、手順の繰り返しが必要となるかもしれない・・・適切な試験デザインの策定や、特定の estimand に対する信頼できる推定値の導出に重大な問題が存在する場合には、別の estimand、試験デザイン、及び解析手法について検討する必要があるだろう。」に通じる内容である。しかしながら、文献調査の調査対象文献においても、感度分析の結果とその後の意思決定の問題に関して論じている文献は少なく(3.4.3.3 項)、現時点では個別の状況に応じて判断する必要があると考えられた。この感度分析の結果とその後の意思決定の問題は estimand の選択に影響を与えるうるものでもあることから、今後の更なる議論が待たれるところである。

今回実施した文献調査では、英語文献の検索を PubMed 又は TIRS、また日本語文献の検索を医中誌 Web のみで実施した。また、設定した調査詳細事項に応じて、各担当者にて検索時のキーワードや他の制約条件を設定した。検索結果に含まれない文献についても、重要と思われる文献については適宜調査対象文献に追加したものの、今回の調査対象に関連する記述がされている文献が調査対象文献に含まれていない可能性があることには注意が必要である。

今後、ICH E9 (R1)が step 4 に到達することに伴って、estimand と感度分析に関する議論がますます盛り上がるとともに、注目を集めるものになると考える。本レポートが、ICH E9 (R1)の理解、及び今後なされるであろう estimand と感度分析に関する議論の理解に少しでも役立ち、より適切な医薬品開発を支援できれば幸いである。

4.1.1 参考文献

- Lavori, PW., Brown, CH., Duan, N., Gibbons, RD., and Greenhouse, J. (2008). Missing data in longitudinal clinical trials part A: design and conceptual issues. *Psychiatric Annals*, **38**(12), 784-792.
- Miyazaki, Y. (2017). 1. 背景及び基本事項と本シンポジウムの概説. 日本製薬工業協会シンポジウム 臨床試験の estimand に対する最近の議論と、欠測のあるデータに対する基本的解析手法について.
<http://www.jpma.or.jp/medicine/shinyaku/tiken/symposium/> [Accessed 30 May 2019]
- Schwartz, D. and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases*. **20**(8), 637-648.

- Tsuchiya, S. (2014). ICH E9 (R1). 第 31 回 ICH 即時報告会.
<http://www.jpma.or.jp/information/ich/sokuji/ich141211.html> [Accessed 30 May 2019]

5. 資料作成者

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 2018年度タスクフォース4

Estimand/因果推論チーム

兼清 道雄	エーザイ株式会社	3.1 節, 3.3 節担当
黒田 晋吾*	武田薬品工業株式会社	1 章, 2 章, 3.2 節, 3.4 節, 4 章, 5 章担当
高橋 文博	田辺三菱製薬株式会社	3.1 節, 3.3 節調査サポート
高本 幹大**	旭化成ファーマ株式会社	3.2 節担当
竹ノ内 一雅#	アステラス製薬株式会社	3.1 節, 3.3 節調査サポート
谷口 隆司	株式会社ヤクルト本社	3.4 節担当
土川 克	ゼリア新薬工業株式会社	3.1 節担当
中田 陽介	大日本住友製薬株式会社	3.3 節担当
中西 豊支	興和株式会社	3.1 節, 3.3 節調査サポート
横山 雄一	持田製薬株式会社	3.1 節, 3.3 節調査サポート
吉田 和生	小野薬品工業株式会社	3.1 節, 3.3 節担当
渡邊 大丞**	サノフィ株式会社	3.2 節, 3.4 節担当

Estimand/因果推論チームリーダー

* 文献調査チームリーダー

** 文献調査チームサブリーダー

以上

6. 付録 調査対象文献一覧

(1) ITT の原則に基づく治療方針の効果とその他の治療効果

- 1 Akacha, M., Binkowitz, B., Claggett, B., Hung, H. M. J., Mueller-Velten, G. & Stockbridge, N. (2018). Assessing Treatment Effects That Capture Disease Burden in Serious Chronic Diseases. *Therapeutic Innovation & Regulatory Science*, published online.
- 2 Akacha, M., Bretz, F. & Ruberg, S. (2017). Estimands in clinical trials - broadening the perspective. *Statistics in Medicine* 36 (1), 5-19.
- 3 Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P. C., Lang, T. & for the CONSORT Group (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Annals of internal medicine* 134 (8), 663-694.
- 4 Atkins, D. C. (2009). Clinical trials methodology: randomization, intent-to-treat, and random-effects regression. *Depression and Anxiety* 26 (8), 697-700.
- 5 Bang, H. & Davis, C. E. (2007). On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? *Statistics in Medicine* 26 (5), 954-964.
- 6 Bang, H. & Davis, C. E. (2009). Author's reply. *Statistics in Medicine* 28 (14), 1967-1968.
- 7 Barczyński, M. (2013). Intention to Treat Analysis: Are We Really Doing It?: Reply. *World Journal of Surgery* 37 (5), 1183-1184.
- 8 Bauer, P. (2017). Comments on 'Estimands in clinical trials - broadening the perspective'. *Statistics in Medicine* 36 (1), 22-23.
- 9 Berger, V. W. (2011). Intent-to-treat: further insight. *Research in Nursing & Health* 34 (1), 4.
- 10 Bhaskar, E. (2009). Intention to treat analysis from a clinician's viewpoint. *Open medicine* 3 (3), e120.
- 11 Boulware, D. R. (2010). Safety, Censoring, and Intent-to-Treat Analysis: Dangers to Generalizability. *Clinical Infectious Diseases* 51 (8), 985-986.
- 12 Capri, S., Ceci, A., Terranova, L., Merlo, F. & Mantovani, L. (2001). Guidelines for Economic Evaluations in Italy: Recommendations from the Italian Group of Pharmacoeconomic Studies. *Drug Information Journal* 35 (1), 189-201.
- 13 Chêne, G., Morlat, P., Leport, C., Hafner, R., Dequae, L., Charreau, I., Aboulker, J.-P., Luft, B., Aubertin, J., Vildé, J.-L. & Salamon, R. (1998). Intention-to-Treat vs. On-Treatment Analyses of Clinical Trial Data. *Controlled Clinical Trials* 19 (3), 233-248.
- 14 Cipriani, A. & Barbui, C. (2008). What is an intention to treat analysis? *Epidemiologia e Psichiatria Sociale* 17 (1), 26-27.
- 15 Comstock, G. W. (2000). Letter to the editor. *Controlled Clinical Trials* 21 (5), 501.
- 16 CPMP Working Party on Efficacy of Medicinal Products (1995) Biostatistical methodology in

- clinical trials in applications for marketing authorizations for medicinal products. *Statistics in Medicine* 14 (15), 1659-1682.
- 17 Dai, J. Y., Gilbert, P. B. & Mâsse, B. R. (2012). Partially Hidden Markov Model for Time-Varying Principal Stratification in HIV Prevention Trials. *Journal of the American Statistical Association* 107 (497), 52-65.
 - 18 Day, S. (2017). Editorial: Estimands in clinical trials - a challenge to intention to treat? *Statistics in Medicine* 36 (1), 3-4.
 - 19 Dixon, D. (2015). Food and Drug Administration guidance on progression-free survival and the intention-to-treat principle. *Clinical Trials* 12 (2), 179
 - 20 Ellenberg, J. H. (1996). Intent-to-Treat Analysis versus As-Treated Analysis. *Drug Information Journal* 30 (2), 535-544.
 - 21 EMA (2018) Guideline on clinical evaluation of vaccines (draft)
 - 22 EMA (2018) Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus (draft)
 - 23 EMA (2018) Guideline on the clinical investigation of medicines for the treatment of Alzheimer's disease
 - 24 EMA (2017) Guideline on clinical investigation of medicinal products for the treatment of rheumatoid arthritis
 - 25 EMA (2017) Guideline on clinical investigation of medicinal products for the treatment of chronic heart failure
 - 26 EMA (2017) Guideline on the clinical development of medicinal products for the treatment of autism spectrum disorder (ASD)
 - 27 EMA (2017) Guideline on the clinical evaluation of medicinal products indicated for the prophylaxis or treatment of respiratory syncytial virus (RSV) disease(draft)
 - 28 EMA (2017) Guideline on the clinical investigation of medicinal products for the treatment of axial spondyloarthritis
 - 29 EMA (2017) Guideline on the clinical investigation of recombinant and human plasma-derived factor VIII products (draft)
 - 30 EMA (2017) Guideline on the evaluation of anticancer medicinal products in man
 - 31 Farrington, C. P. (1993). Intention-to-treat analyses in clinical trials and cohort studies. *International journal of epidemiology* 22 (3), 566.
 - 32 Fay, M. P., Brittain, E. H., Shih, J. H., Follmann, D. A. & Gabriel, E. E. (2018). Causal estimands and confidence intervals associated with Wilcoxon-Mann-Whitney tests in randomized experiments. *Statistics in Medicine* 37 (20), 2923-2937.

- 33 FDA (1988). Guideline for the Format and Content of the Clinical and Statistical Sections of New Drug Applications.
- 34 Feinman, R. D. (2009). Intention-to-treat. What is the question? *Nutrition & Metabolism* 6 (1), 1.
- 35 Fergusson, D., Aaron, S. D., Guyatt, G. & Hébert, P. (2002). Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 325 (7365), 652-654.
- 36 Fletcher, C., Tsuchiya, S. & Mehrotra, D. (2017). Current Practices in Choosing Estimands and Sensitivity Analyses in Clinical Trials. *Therapeutic Innovation & Regulatory Science* 51 (1), 69-76.
- 37 Frangakis, C. E. & Rubin, D. B. (2002). Principal Stratification in Causal Inference. *Biometrics* 58 (1), 21-29.
- 38 Gail, M. H. (1985). Eligibility exclusions, losses to follow-up, removal of randomized patients, and uncounted events in cancer clinical trials. *Cancer treatment reports* 69 (10), 1107-13.
- 39 Garrett, A. (2015). Choosing Appropriate Estimands in Clinical Trials (Leuchs et al). *Therapeutic Innovation & Regulatory Science* 49 (4), 601-601.
- 40 Gibaldi, M. & Sullivan, S. (1997). Intention-to-Treat Analysis in Randomized Trials: Who Gets Counted? *The Journal of Clinical Pharmacology* 37 (8), 667-672.
- 41 Gilbert, P. B., Hudgens, M. G. & Wolfson, J. (2011). Commentary on "Principal Stratification – a Goal or a Tool?" by Judea Pearl. *The International Journal of Biostatistics* 7 (1), 1-15.
- 42 Gillings, D. & Koch, G. (1991). The Application of the Principle of Intention-to-Treat to the Analysis of Clinical Trials. *Drug Information Journal* 25 (3), 411-424.
- 43 Goetghebeur, E. (2002). Beyond Intention to Treat. *Epidemiologic Reviews* 24 (1), 85-90.
- 44 Greene, T., Joffe, M., Hu, B., Li, L. & Boucher, K. (2013). The Balanced Survivor Average Causal Effect. *The International Journal of Biostatistics* 9 (2), 291-306.
- 45 Greenland, S., Lanes, S. & Jara, M. (2008). Estimating effects from randomized trials with discontinuations: the need for intent-to-treat design and G-estimation. *Clinical Trials* 5 (1), 5-13.
- 46 Gupta, S. (2011). Intention-to-treat concept: A review. *Perspectives in Clinical Research* 2 (3), 109.
- 47 Helms, R. W. (2016). Precise definitions of some terminology for longitudinal clinical trials: subjects, patient populations, analysis sets, intention to treat, and related terms. *Pharmaceutical Statistics* 15 (6), 471-485.

- 48 Herigon, J. C. & Newland, J. G. (2012). The Role of Intention-to-Treat Analyses in Randomized Trials. *Infection Control & Hospital Epidemiology* 33 (02), 207-208.
- 49 Heritier, S. R., Gebiski, V. J. & Keech, A. C. (2003). Inclusion of patients in clinical trial analysis: the intention-to-treat principle. *The Medical journal of Australia* 179 (8), 438-40.
- 50 Hernán, M. A. & Hernández-Díaz, S. (2012). Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials* 9 (1), 48-55.
- 51 Hernán, M. A. & Scharfstein, D. (2018). Cautions as Regulators Move to End Exclusive Reliance on Intention to Treat. *Annals of Internal Medicine* 168 (7), 515.
- 52 Hodges, J. S. (2008). Intention-to-treat has implications for study planning and execution, not just subject retention and follow-up. *Journal of Clinical Periodontology* 35 (8), 683-684.
- 53 Hogan, J. W. & Laird, N. M. (1996). Intention-to-Treat Analyses for Incomplete Repeated Measures Data. *Biometrics* 52 (3), 1002.
- 54 Holzhauer, B., Akacha, M. & Bermann, G. (2015). Choice of estimand and analysis methods in diabetes trials with rescue medication. *Pharmaceutical Statistics* 14 (6), 433-447.
- 55 Horvitz-Lennon, M., James O'Malley, A., Frank, R. G. & Normand, S.-L. T. (2005). Improving traditional intention-to-treat analyses: a new approach. *Psychological Medicine* 35 (7), 961-970.
- 56 Houck, P. R., Mazumdar, S., Mulsant, B. H., Pollock, B. G., Dew, M. A. & Reynolds, C. F. (2003). An intent-to-treat method for enhancing analysis of clinical trials with rescue medication: a mixed-model approach. *Psychopharmacology bulletin* 37 (1), 79-89.
- 57 Hudgens, M. G. & Halloran, M. E. (2006). Causal Vaccine Effects on Binary Postinfection Outcomes. *Journal of the American Statistical Association* 101 (473), 51-64.
- 58 Joffe, M. (2011). Principal Stratification and Attribution Prohibition: Good Ideas Taken Too Far. *The International Journal of Biostatistics* 7 (1), 1-22.
- 59 Jones, D. R. (1993). Improving Statistics in License Applications: Handle with Sensitivity? *Drug Information Journal* 27 (3), 833-836.
- 60 Kannel, W. B. (1990). Coronary artery surgery study revisited. Limitation of the intent-to-treat principle. *Circulation* 82 (5), 1859-1862.
- 61 Karrison, T. & Kocherginsky, M. (2018). Restricted mean survival time: Does covariate adjustment improve precision in randomized clinical trials? *Clinical Trials* 15 (2), 178-188.

- 62 Keene, O. N. (2011), Intent-to-treat analysis in the presence of off-treatment or missing data. *Pharmaceutical Statistics* 10 (3), 191-195.
- 63 Koch, G. G., Amara, I. A., Forster, J., McSorley, D. & Peace, K. E. (1993). Statistical Issues in the Design and Analysis of Ulcer Healing and Recurrence Studies. *Drug Information Journal* 27 (3), 805-824.
- 64 Koch, G. G. & Wiener, L. E. (2016). Commentary for the Missing Data Working Group's perspective for regulatory clinical trials, estimands, and sensitivity analyses. *Statistics in Medicine* 35 (17), 2887-2893.
- 65 Kohara, N., Kaneko, M. & Narukawa, M. (2019). Investigation of the Study Characteristics Affecting Clinical Trial Quality Using the Protocol Deviations Leading to Exclusion of Subjects From the Per Protocol Set Data in Studies for New Drug Application: A Retrospective Analysis. *Therapeutic Innovation & Regulatory Science* 53 (1), 45-51.
- 66 Korhonen, P., Loeys, T., Goetghebeur, E. & Palmgren, J. (2000). Vitamin A and infant mortality: beyond intention-to-treat in a randomized trial. *Lifetime data analysis* 6 (2), 107-21.
- 67 van der Laan, M. J. & Petersen, M. L. (2007). Causal Effect Models for Realistic Individualized Treatment and Intention to Treat Rules. *The International Journal of Biostatistics* 3 (1), 1-51.
- 68 Lachin, J. M. (2000). Statistical Considerations in the Intent-to-Treat Principle. *Controlled Clinical Trials* 21 (3), 167-189.
- 69 Lavori, P. W. & Dawson, R. (2001). Designing for Intent-to-Treat. *Drug Information Journal* 35 (4), 1079-1086.
- 70 Leuchs, A.-K., Brandt, A., Zinserling, J. & Benda, N. (2017). Disentangling estimands and the intention-to-treat principle. *Pharmaceutical Statistics* 16 (1), 12-19.
- 71 Leuchs, A.-K., Zinserling, J., Brandt, A., Wirtz, D. & Benda, N. (2015). Choosing Appropriate Estimands in Clinical Trials. *Therapeutic Innovation & Regulatory Science* 49 (4), 584-592.
- 72 Lewis, J. A., Jones, D. R. & Rühmel, J. (1995). Biostatistical methodology in clinical trials — a European guideline. *Statistics in Medicine* 14 (15), 1655-1657.
- 73 Lewis, J. A. & Machin, D. (1993). Intention to treat—who should use ITT? *British journal of cancer* 68 (4), 647-50.
- 74 Little, R. & Kang, S. (2015). Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials. *Statistics in Medicine* 34 (16), 2381-2390.

- 75 Lou, Y., Jones, M. P. & Sun, W. (2019). Estimation of causal effects in clinical endpoint bioequivalence studies in the presence of intercurrent events: noncompliance and missing data. *Journal of Biopharmaceutical Statistics* 29 (1), 151-173.
- 76 Lui, K.-J. (2007). Notes on Test Equality in Stratified Noncompliance Randomized Trials. *Drug Information Journal* 41 (5), 607-618.
- 77 Lui, K.-J. (2009). Comments on 'On estimating treatment effects under non-compliance in randomized clinical trials: Are intent-to-treat or instrumental variables analyses perfect solutions?' by Heejung Bang and Clarence E. Davis, *Statistics in Medicine* 2007; 26 :954-964. *Statistics in Medicine* 28 (3), 531-534.
- 78 Lynch, C. J. & Lachenbruch, P. A. (1996). Statistical Issues in Biologics Submissions to the FDA. *Drug Information Journal* 30 (4), 921-932.
- 79 Mahaniah, K. J. & Rao, G. (2004). Intention-to-treat analysis: protecting the integrity of randomization. *The Journal of family practice* 53 (8), 644.
- 80 Mak, C., Heft, S. M., Amkraut, H. & Ling, M.-H. (2001). Compliance as an Explanatory Variable in Hepatitis-C. *Drug Information Journal* 35 (4), 1351-1361.
- 81 Mallinckrodt, C., Molenberghs, G. & Rathmann, S. (2017). Choosing estimands in clinical trials with missing data. *Pharmaceutical Statistics* 16 (1), 29--36.
- 82 Mallinckrodt, C., Roger, J., Chuang-Stein, C., Molenberghs, G., O'Kelly, M., Ratitch, B., Janssens, M. & Bunouf, P. (2014). Recent Developments in the Prevention and Treatment of Missing Data. *Therapeutic Innovation & Regulatory Science* 48 (1), 68-80.
- 83 Mallinckrodt, C. H., Lin, Q., Lipkovich, I. & Molenberghs, G. (2012). A structured approach to choosing estimands and estimators in longitudinal clinical trials. *Pharmaceutical Statistics* 11 (6), 456-461.
- 84 Matsuyama, Y. (2010). A comparison of the results of intent-to-treat, per-protocol, and g-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial. *Statistics in Medicine* 29 (20), 2107-2116.
- 85 Matsuyama, Y. & Morita, S. (2006). Estimation of the average causal effect among subgroups defined by post-treatment variables. *Clinical Trials* 3 (1), 1-9.
- 86 Mehrotra, D. V., Hemmings, R. J. & Russek-Cohen, E. (2016). Seeking harmony: estimands and sensitivity analyses for confirmatory clinical trials. *Clinical Trials: Journal of the Society for Clinical Trials* 13 (4), 456-458.
- 87 Mehrotra, D. V., Liu, F. & Permutt, T. (2017). Missing data in clinical trials: control-based mean imputation and sensitivity analysis. *Pharmaceutical Statistics* 16 (5), 378-392.

- 88 Ministry of Health and Welfare (1992) Guideline for the Statistical analysis of clinical trials (in Japanese)
- 89 Mohanty, D., Garg, P. K. & Kumar, A. (2013). Intention To Treat Analysis: Are We Really Doing It? *World Journal of Surgery* 37 (5), 1181-1182.
- 90 Moher, D. (2001). The CONSORT Statement: Revised Recommendations for Improving the Quality of Reports of Parallel-Group Randomized Trials. *JAMA* 285 (15), 1987.
- 91 Molenberghs, G. (2009). Incomplete Data in Clinical Studies: Analysis, Sensitivity, and Sensitivity Analysis—Rejoinder. *Drug Information Journal* 43 (4), 447-448.
- 92 Newell, D. J. (1992). Intention-to-treat analysis: implications for quantitative and qualitative research. *International journal of epidemiology* 21 (5), 837-41.
- 93 Nich, C. & Carroll, K. M. (2002). 'Intention-to-treat' meets 'missing data': implications of alternate strategies for analyzing clinical trials data. *Drug and Alcohol Dependence* 68 (2), 121-130.
- 94 Pearl, J. (2011). Principal Stratification – a Goal or a Tool? *The International Journal of Biostatistics* 7 (1), 1-13.
- 95 Peduzzi, P., Detre, K., Wittes, J. & Holford, T. (1991). Intent-to-treat analysis and the problem of crossovers. An example from the Veterans Administration coronary bypass surgery study. *The Journal of thoracic and cardiovascular surgery* 101 (3), 481-7.
- 96 Permutt, T. (2016). A taxonomy of estimands for regulatory clinical trials with discontinuations. *Statistics in Medicine* 35 (17), 2865-2875.
- 97 Permutt, T. (2017). Comments on 'Estimands in clinical trials - broadening the perspective'. *Statistics in Medicine* 36 (1), 20-21.
- 98 Peto, R., Pike, M. C., Armitage, P., Breslow, N. E., Cox, D. R., Howard, S. V., Mantel, N., McPherson, K., Peto, J. & Smith, P. G. (1976). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *British Journal of Cancer* 34 (6), 585-612.
- 99 Petroff, D. (2017). A simple correction to completer analyses and improvement on baseline observation carried forward. *Biometrics* 73 (4), 1366-1378.
- 100 Peugh, J. L., Strotman, D., McGrady, M., Rausch, J. & Kashikar-Zuck, S. (2017). Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer. *Journal of School Psychology* 60, 7-24.
- 101 Phillips, A., Abellan-Andres, J., Soren, A., Bretz, F., Fletcher, C., France, L., Garrett, A., Harris, R., Kjaer, M., Keene, O., Morgan, D., O'Kelly, M. & Roger, J.

- (2017). Estimands: discussion points from the PSI estimands and sensitivity expert group. *Pharmaceutical Statistics* 16 (1), 6-11.
- 102 Polit, D. F. & Gillespie, B. M. (2010). Intention-to-treat in randomized controlled trials: Recommendations for a total trial strategy. *Research in Nursing & Health* 33 (4), 355-368.
- 103 Ranstam, J., Turkiewicz, A., Boonen, S., Van Meirhaeghe, J., Bastian, L. & Wardlaw, D. (2012). Alternative analyses for handling incomplete follow-up in the intention-to-treat analysis: the randomized controlled trial of balloon kyphoplasty versus non-surgical care for vertebral compression fracture (FREE). *BMC Medical Research Methodology* 12 (1), 35.
- 104 Rochon, J. (1999). Issues in Adjusting for Covariates Arising Postrandomization in Clinical Trials. *Drug Information Journal* 33 (4), 1219-1228.
- 105 Rockhold, F. (2017). Comments on 'Estimands in clinical trials - broadening the perspective'. *Statistics in Medicine* 36 (1), 24-26
- 106 Ruberg, S. J. & Akacha, M. (2017). Considerations for Evaluating Treatment Effects From Randomized Clinical Trials. *Clinical Pharmacology & Therapeutics* 102 (6), 917-923.
- 107 Sabin, C. A., Lepri, A. C. & Phillips, A. N. (2015). A practical guide to applying the intention-to-treat principle to clinical trials in HIV infection. *HIV clinical trials* 1 (2), 31-8.
- 108 Sainani, K. L. (2010). Making Sense of Intention-to-Treat. *PM&R* 2 (3), 209-213.
- 109 Salim, A., Mackinnon, A., Christensen, H. & Griffiths, K. (2008). Comparison of data analysis strategies for intent-to-treat analysis in pre-test–post-test designs with substantial dropout rates. *Psychiatry Research* 160 (3), 335-345.
- 110 Schiffner, R., Schiffner-Rohe, J., Gerstenhauer, M., Hofstadter, F., Landthaler, M. & Stolz, W. (2001). Differences in efficacy between intention-to-treat and per-protocol analyses for patients with psoriasis vulgaris and atopic dermatitis: clinical and pharmaco-economic implications. *British Journal of Dermatology* 144 (6), 1154-1160.
- 111 Seals, D. R. & Tanaka, H. (2000). Manuscript peer review: a helpful checklist for students and novice referees. *Advances in Physiology Education* 23 (1), S52-58.
- 112 Sheiner, L. B. (2002). Is intent-to-treat analysis always (ever) enough? *British Journal of Clinical Pharmacology* 54 (2), 203-211.
- 113 Sheiner, L. B. & Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials. *Clinical Pharmacology & Therapeutics* 57 (1), 6-15.

- 114 Shepherd, B. E., Liu, Q., Mercaldo, N., Jenkins, C. A., Lau, B., Cole, S. R., Saag, M. S. & Sterling, T. R. (2016). Comparing results from multiple imputation and dynamic marginal structural models for estimating when to start antiretroviral therapy. *Statistics in Medicine* 35 (24), 4335-4351.
- 115 Shih, W. J. (2016). Comments on the three papers by the FDA/CDER research team on the regulatory perspective of the missing data problem. *Statistics in Medicine* 35 (17), 2880-2886.
- 116 Shrier, I., Steele, R. J., Verhagen, E., Herbert, R., Riddell, C. A. & Kaufman, J. S. (2014). Beyond intention to treat: What is the right question? *Clinical Trials* 11 (1), 28-37.
- 117 Shrier, I., Verhagen, E. & Stovitz, S. D. (2017). The Intention-to-Treat Analysis Is Not Always the Conservative Approach. *The American Journal of Medicine* 130 (7), 867-871.
- 118 Sjölander, A., Humphreys, K., Vansteelandt, S., Bellocco, R. & Palmgren, J. (2009). Sensitivity Analysis for Principal Stratum Direct Effects, with an Application to a Study of Physical Activity and Coronary Heart Disease. *Biometrics* 65 (2), 514-520.
- 119 Skoglund, L. A. (1996). Paracetamol-codeine combinations versus paracetamol alone. Intention to treat approach is not popular. *BMJ* 313 (7066), 1209.
- 120 Soares, I. & Carneiro, A. V. (2002). Intention-to-treat analysis in clinical trials: principles and practical importance. *Portuguese journal of cardiology* 21 (10), 1191-8.
- 121 Sorbello, A., Komo, S. & Valappil, T. (2010). Noninferiority Margin for Clinical Trials of Antibacterial Drugs for Nosocomial Pneumonia. *Drug Information Journal* 44 (2), 165-176.
- 122 Stewart, W. H. (2004). Basing Intention-to-Treat on Cause and Effect Criteria. *Drug Information Journal* 38 (4), 361-369.
- 123 Streiner, D. & Geddes, J. (2001). Intention to treat analysis in clinical trials when there are missing data. *Evidence-Based Mental Health* 4 (3), 70-71.
- 124 Svensson, K. (1995). How Many Populations Must be Analyzed and How Should They be Defined (Intention-To-Treat, Eligible, per Protocol Population, Etc.)? *Drug Information Journal* 29 (2), 475-477.
- 125 Taylor, L. & Zhou, X. H. (2009). Multiple Imputation Methods for Treatment Noncompliance and Nonresponse in Randomized Clinical Trials. *Biometrics* 65 (1), 88-95.
- 126 Ten Have, T. R., Normand, S.-l. T., Marcus, S. M., Brown, C. H., Lavori, P. & Duan, N. (2008). Intent-to-Treat vs. Non-Intent-to-Treat Analyses Under Treatment Non-

- Adherence in Mental Health Randomized Trials. *Psychiatric Annals* 38 (12), 772-783.
- 127 Tillmann, H. C., Sharpe, N., Sporer, G. & Wehling, M. (2001). Does intention-to-treat analysis answer all questions in long-term mortality trials? Considerations on the basis of the ANZ trial. *International journal of clinical pharmacology and therapeutics* 39 (5), 205-12.
- 128 Tsiatis, A. (1990). Methodological issues in AIDS clinical trials. Intent-to-treat analysis. *Journal of acquired immune deficiency syndromes* 3 Suppl 2, S120-3.
- 129 Tufanaru, C. (2017). Demystifying intention-to-treat analyses. *JBI Database of Systematic Reviews and Implementation Reports* 15 (4), 853-854.
- 130 Unnebrink, K. & Windeler, J. (1999). Sensitivity Analysis by Worst and Best Case Assessment: Is it Really Sensitive? *Drug Information Journal* 33 (3), 835-839.
- 131 Unnebrink, K. & Windeler, J. (2001). Intention-to-treat: methods for dealing with missing values in clinical trials of progressively deteriorating diseases. *Statistics in Medicine* 20 (24), 3931-3946.
- 132 Vlad, S. C. & LaValley, M. P. (2008). Intention-to-treat analysis may better represent the actual efficacy. *Archives of internal medicine* 168 (11), 1228.
- 133 Welsh, A. W. (2013). Randomised controlled trials and clinical maternity care: moving on from intention-to-treat and other simplistic analyses of efficacy. *BMC pregnancy and childbirth* 13 (1), 15.
- 134 White, I. R., Carpenter, J. & Horton, N. J. (2012). Including all individuals is not enough: Lessons for intention-to-treat analysis. *Clinical Trials* 9 (4), 396-407.
- 135 Whitney, J. (2001). Using intent-to-treat analysis in clinical studies. *Journal of WOCN* 28 (6), 267-269.
- 136 Winnock, M., Rancinan, C., de Ledinghen, V., Couzigou, P. & Ch[ecirc]ne, G. (2001). What hides behind an intention-to-treat analysis? *Hepatology* 33 (4), 1014.
- 137 Woolley, P. & Seiler, W. (2001). Training Medical Writers in Today's Environment 3: How to Please a Medical Writer. *Drug Information Journal* 35 (1), 259-268
- 138 広津千尋. (2004) 医学・薬学データの統計解析,, 東京大学出版会, 2004.
- 139 上坂浩之,森川敏彦,魚井 徹. (2003) 非劣性試験におけるデザインと試験の質の諸問題., 計量生物学 2000 ; 20 : 201-215.
- 140 丹後俊郎. (2003) 無作為化比較試験-デザインと統計解析-,医学統計シリーズ 5, 朝倉書店 2003.
- 141 厚生労働省医薬局審査管理課 (1998) 臨床試験のための統計的原則.,

(2) データの取扱い及び欠測データに含めて検討されている問題

1. Akl, E. A., Shawwa, K., Kahale, L. A., Agoritsas, T., Brignardello-Petersen, R., Busse, J. W., . . . Guyatt, G. H. (2015). Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide. *BMJ Open*, 5(12), e008431. doi:10.1136/bmjopen-2015-008431
2. Alshurafa, M., Briel, M., Akl, E. A., Haines, T., Moayyedi, P., Gentles, S. J., . . . Guyatt, G. H. (2012). Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. *PLoS One*, 7(11), e49163. doi:10.1371/journal.pone.0049163
3. Anderson, S. F. (2018). Missing Data in Randomized Pretest Posttest Studies. *Multivariate Behav Res*, 53(1), 125-126. doi:10.1080/00273171.2017.1404893
4. Bannon, W., Jr. (2015). Missing data within a quantitative research study: How to assess it, treat it, and why you should care. *J Am Assoc Nurse Pract*, 27(4), 230-232. doi:10.1002/2327-6924.12208
5. Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., . . . Hurny, C. (1998). Missing quality of life data in cancer clinical trials: serious problems and challenges. *Stat Med*, 17(5-7), 517-532.
6. Biswas, K. (2012). Prevention and management of missing data during conduct of a clinical study. *Shanghai Arch Psychiatry*, 24(4), 235-237. doi:10.3969/j.issn.1002-0829.2012.04.008
7. Brown, D., Day, S., Hemmings, R., & Wright, D. (2008). Assessing the impact of ICH E9. *Pharm Stat*, 7(2), 77-87. doi:10.1002/pst.337
8. Brummel, S. S., & Gillen, D. L. (2013). On the Use of Local Assessments for Monitoring Centrally Reviewed Endpoints with Missing Data in Clinical Trials. *Open J Stat*, 3(4a), 41-54. doi:10.4236/ojs.2013.34A005
9. Carpenter, J., Pocock, S., & Lamm, C. J. (2002). Coping with missing data in clinical trials: a model-based approach applied to asthma trials. *Stat Med*, 21(8), 1043-1066.
10. Carpenter, J. R., Roger, J. H., & Kenward, M. G. (2013). Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *J Biopharm Stat*, 23(6), 1352-1371. doi:10.1080/10543406.2013.834911
11. Dziura, J. D., Post, L. A., Zhao, Q., Fu, Z., & Peduzzi, P. (2013). Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med*, 86(3), 343-358.
12. Fairclough, D. L., Peterson, H. F., & Chang, V. (1998). Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med*, 17(5-7), 667-677.
13. Fleming, T. R. (2011). Addressing missing data in clinical trials. *Ann Intern Med*, 154(2), 113-117. doi:10.7326/0003-4819-154-2-201101180-00010
14. Flyer, P., & Hirman, J. (2009). Missing data in confirmatory clinical trials. *J Biopharm Stat*, 19(6),

- 969-979. doi:10.1080/10543400903242746
15. Frangakis, C. E., Rubin, D. B., An, M. W., & MacKenzie, E. (2007). Principal stratification designs to estimate input data missing due to death. *Biometrics*, *63*(3), 641-649; discussion 650-662. doi:10.1111/j.1541-0420.2007.00847_1.x
 16. Franks, A. M., Csardi, G., Drummond, D. A., & Airoidi, E. M. (2015). Estimating a structured covariance matrix from multi-lab measurements in high-throughput biology. *J Am Stat Assoc*, *110*(509), 27-44. doi:10.1080/01621459.2014.964404
 17. Gallo, P., & Chuang-Steiny, C. (2009). A Note on Missing Data in Noninferiority Trials. *Drug Information Journal*, *43*(4), 469-474.
 18. Gemmen, E. (2015). Prevention and Treatment of Missing Data in Real World Research. *Value Health*, *18*(7), A732. doi:10.1016/j.jval.2015.09.2799
 19. Gewandter, J. S., McDermott, M. P., McKeown, A., Smith, S. M., Williams, M. R., Hunsinger, M., . . . Dworkin, R. H. (2014). Reporting of missing data and methods used to accommodate them in recent analgesic clinical trials: ACTION systematic review and recommendations. *Pain*, *155*(9), 1871-1877. doi:10.1016/j.pain.2014.06.018
 20. Gilbert, P. B., & Hudgens, M. G. (2008). Evaluating candidate principal surrogate endpoints. *Biometrics*, *64*(4), 1146-1154. doi:10.1111/j.1541-0420.2008.01014.x
 21. Godlee, F., & Loder, E. (2010). Missing clinical trial data: setting the record straight. *BMJ*, *341*, c5641. doi:10.1136/bmj.c5641
 22. Greene, T., Joffe, M., Hu, B., Li, L., & Boucher, K. (2013). The balanced survivor average causal effect. *Int J Biostat*, *9*(2), 291-306. doi:10.1515/ijb-2012-0013
 23. Groenwold, R. H., Donders, A. R., Roes, K. C., Harrell, F. E., Jr., & Moons, K. G. (2012). Dealing with missing outcome data in randomized trials and observational studies. *Am J Epidemiol*, *175*(3), 210-217. doi:10.1093/aje/kwr302
 24. Harel, O., & Demirtas, H. (2011). Re: Joint modeling of missing data due to non-participation and death in longitudinal aging studies by K. B. Rajan and S. E. Leurgans, *Statistics in Medicine* 2010; 29:2260-2268. *Stat Med*, *30*(21), 2663-2665; author reply 2666-2668. doi:10.1002/sim.4171
 25. Hirman, J., & Flyer, P. (2012). Missing data in clinical trials. *N Engl J Med*, *367*(26), 2557; author reply 2557-2558. doi:10.1056/NEJMc1213388
 26. Hollestein, L. M., & Carpenter, J. R. (2017). Missing data in clinical research: an integrated approach. *Br J Dermatol*, *177*(6), 1463-1465. doi:10.1111/bjd.16010
 27. Holzhauser, B., Akacha, M., & Bermann, G. (2015). Choice of estimand and analysis methods in diabetes trials with rescue medication. *Pharm Stat*, *14*(6), 433-447. doi:10.1002/pst.1705
 28. Hughes, S., Harris, J., Flack, N., & Cuffe, R. L. (2012). The statistician's role in the prevention of missing data. *Pharm Stat*, *11*(5), 410-416. doi:10.1002/pst.1528

29. Ibrahim, J. G., Chu, H., & Chen, M. H. (2012). Missing data in clinical studies: issues and methods. *J Clin Oncol*, *30*(26), 3297-3303. doi:10.1200/jco.2011.38.7589
30. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. (1999). *Stat Med*, *18*(15), 1905-1942.
31. Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. *BMC Med Res Methodol*, *17*(1), 162. doi:10.1186/s12874-017-0442-1
32. Jo, B., Ginexi, E. M., & Ialongo, N. S. (2010). Handling missing data in randomized experiments with noncompliance. *Prev Sci*, *11*(4), 384-396. doi:10.1007/s11121-010-0175-4
33. Johnston, B. C., & Guyatt, G. H. (2016). Best (but oft-forgotten) practices: intention-to-treat, treatment adherence, and missing participant outcome data in the nutrition literature. *Am J Clin Nutr*, *104*(5), 1197-1201. doi:10.3945/ajcn.115.123315
34. Kang, H. (2013). The prevention and handling of the missing data. *Korean J Anesthesiol*, *64*(5), 402-406. doi:10.4097/kjae.2013.64.5.402
35. Kaushal, S. (2014). Missing data in clinical trials: Pitfalls and remedies. *Int J Appl Basic Med Res*, *4*(Suppl 1), S6-7.
36. Kelleher, T., Thiry, A., Wilber, R., & Cross, A. (2016). Missing Data Methods in Hiv Clinical Trials: Regulatory Guidance and Alternative Approaches. *Drug Information Journal*, *35*(4), 1363-1371.
37. Kistin, C. J. (2014). Transparent reporting of missing outcome data in clinical trials: applying the general principles of CONSORT 2010. *Evid Based Med*, *19*(5), 161-162. doi:10.1136/eb-2014-101797
38. Kurland, B. F., & Egleston, B. L. (2016). For Health-Related Quality of Life and Other Longitudinal Data, Analysis Should Distinguish Between Truncation by Death and Data Missing Because of Nonresponse. *J Clin Oncol*, *34*(36), 4449. doi:10.1200/jco.2016.69.1220
39. Kurland, B. F., Johnson, L. L., & Diehr, P. H. (2012). Accommodation of missing data in supportive and palliative care clinical trials. *Curr Opin Support Palliat Care*, *6*(4), 465-470. doi:10.1097/SPC.0b013e328358441d
40. LaVange, L. M., & Permutt, T. (2016). A regulatory perspective on missing data in the aftermath of the NRC report. *Stat Med*, *35*(17), 2853-2864. doi:10.1002/sim.6840
41. Lavori, P. W., Brown, C. H., Duan, N., Gibbons, R. D., & Greenhouse, J. (2008). Missing Data in Longitudinal Clinical Trials Part A: Design and Conceptual Issues. *Psychiatr Ann*, *38*(12), 784-792.
42. Lee, K., Daniels, M. J., & Sargent, D. J. (2010). CAUSAL EFFECTS OF TREATMENTS FOR INFORMATIVE MISSING DATA DUE TO PROGRESSION/DEATH. *J Am Stat Assoc*, *105*(491), 912-929. doi:10.1198/jasa.2010.ap08739.

43. Lehman, R., & Loder, E. (2012). Missing clinical trial data. *BMJ*, *344*, d8158. doi:10.1136/bmj.d8158
44. Leuchs, A. K., Zinserling, J., Schlosser-Weber, G., Berres, M., Neuhauser, M., & Benda, N. (2014). Estimation of the treatment effect in the presence of non-compliance and missing data. *Stat Med*, *33*(2), 193-208. doi:10.1002/sim.5924
45. Lewis, H. D., Jr. (2012). Missing data in clinical trials. *N Engl J Med*, *367*(26), 2557; author reply 2557-2558. doi:10.1056/NEJMc1213388
46. Lewis, J. (2008). 10 years of ICH E9. *Pharm Stat*, *7*(1), 1-3. doi:10.1002/pst.320
47. Lewis, J., Louv, W., Rockhold, F., & Sato, T. (2001). The impact of the international guideline entitled Statistical Principles for Clinical Trials (ICH E9). *Stat Med*, *20*(17-18), 2549-2560.
48. Lewis, J. A. (1999). Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Stat Med*, *18*(15), 1903-1942.
49. Li, T., Hutfless, S., Scharfstein, D. O., Daniels, M. J., Hogan, J. W., Little, R. J., . . . Dickersin, K. (2014). Standards should be applied in the prevention and handling of missing data for patient-centered outcomes research: a systematic review and expert consensus. *J Clin Epidemiol*, *67*(1), 15-32. doi:10.1016/j.jclinepi.2013.08.013
50. Li, X., Wang, W. W., Liu, G. F., & Chan, I. S. (2011). Handling missing data in vaccine clinical trials for immunogenicity and safety evaluation. *J Biopharm Stat*, *21*(2), 294-310. doi:10.1080/10543406.2011.550111
51. Liao, J. M., & Stack, C. B. (2017). Annals Understanding Clinical Research: Implications of Missing Data Due to Dropout. *Ann Intern Med*, *166*(8), 596-598. doi:10.7326/m17-0195
52. Little, R., & Kang, S. (2015). Intention-to-treat analysis with treatment discontinuation and missing data in clinical trials. *Stat Med*, *34*(16), 2381-2390. doi:10.1002/sim.6352
53. Little, R. J., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Neaton, J. D., . . . Stern, H. (2012a). The design and conduct of clinical trials to limit missing data. *Stat Med*, *31*(28), 3433-3443. doi:10.1002/sim.5519
54. Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., . . . Stern, H. (2012b). The prevention and treatment of missing data in clinical trials. *N Engl J Med*, *367*(14), 1355-1360. doi:10.1056/NEJMsr1203730
55. Little, R. J., Wang, J., Sun, X., Tian, H., Suh, E. Y., Lee, M., . . . Mohanty, S. (2016). The treatment of missing data in a large cardiovascular clinical outcomes study. *Clin Trials*, *13*(3), 344-351. doi:10.1177/1740774515626411
56. Liu, M., Wei, L., & Zhang, J. (2006). Review of guidelines and literature for handling missing data in longitudinal clinical trials with a case study. *Pharm Stat*, *5*(1), 7-18. doi:10.1002/pst.189
57. Liublinska, V., & Rubin, D. B. (2012). Re: "dealing with missing outcome data in randomized trials and observational studies". *Am J Epidemiol*, *176*(4), 357-358; author reply 358-359.

doi:10.1093/aje/kws215

58. Mallinckrodt, C., Molenberghs, G., & Rathmann, S. (2017). Choosing estimands in clinical trials with missing data. *Pharm Stat*, 16(1), 29-36. doi:10.1002/pst.1765
59. Mallinckrodt, C. H., Lin, Q., Lipkovich, I., & Molenberghs, G. (2012). A structured approach to choosing estimands and estimators in longitudinal clinical trials. *Pharm Stat*, 11(6), 456-461. doi:10.1002/pst.1536
60. Mallinckrodt, C., Roger, J., Chuang-Stein, C., Molenberghs, G., O'Kelly, M., Ratitch, B., Janssens, M., & Bunouf, P. (2013). Recent Developments in the Prevention and Treatment of Missing Data. *Therapeutic Innovation & Regulatory Science*, 48(1), 68-80.
61. Marston, L., & Sedgwick, P. (2014). Randomised controlled trials: missing data. *BMJ*, 349, g4656. doi:10.1136/bmj.g4656
62. Matthews, J. N., & Henderson, R. (2013). Two-period, two-treatment crossover designs subject to non-ignorable missing data. *Biostatistics*, 14(4), 626-638. doi:10.1093/biostatistics/kxt009
63. McEvoy, B. W. (2016). Missing data in clinical trials for weight management. *J Biopharm Stat*, 26(1), 30-36. doi:10.1080/10543406.2015.1094814
64. Mealli, F., Imbens, G. W., Ferro, S., & Biggeri, A. (2004). Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. *Biostatistics*, 5(2), 207-222. doi:10.1093/biostatistics/5.2.207
65. Mehrotra, D. V., Hemmings, R. J., & Russek-Cohen, E. (2016). Seeking harmony: estimands and sensitivity analyses for confirmatory clinical trials. *Clin Trials*, 13(4), 456-458. doi:10.1177/1740774516633115
66. Mehrotra, D. V., Liu, F., & Permutt, T. (2017). Missing data in clinical trials: control-based mean imputation and sensitivity analysis. *Pharm Stat*, 16(5), 378-392. doi:10.1002/pst.1817
67. Mendelsohn, A. B., Dreyer, A. D., Mattox, P. W., Su, Z., Swenson, A., Li, R., Turner, J. R., & Velentgas, P. (2009). Characterization of Missing Data in Clinical Registry Studies. *Drug Information Journal*, 43(4), 469-474.
68. Myers, WR. (2016). Handling Missing Data in Clinical Trials: An Overview. *Drug Information Journal*, 34(2), 525-533.
69. National Research Council Panel on Handling Missing Data in Clinical, T. (2010) *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington (DC): National Academies Press (US)
Copyright 2010 by the National Academy of Sciences. All rights reserved.
70. Newgard, C. D., & Lewis, R. J. (2016). Accounting for Missing Data in Clinical Research--Reply. *JAMA*, 315(5), 518. doi:10.1001/jama.2015.16470
71. O'Neill, R. T., & Temple, R. (2012). The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clin Pharmacol Ther*, 91(3), 550-

554. doi:10.1038/clpt.2011.340
72. Permutt, T. (2016). A taxonomy of estimands for regulatory clinical trials with discontinuations. *Stat Med*, 35(17), 2865-2875. doi:10.1002/sim.6841
73. Permutt, T., & Pinheiro, J. (2009). Editorial: Dealing with the Missing Data Challenge in Clinical Trials. *Drug Information Journal*, 43(4), 403-408.
74. Petroff, D. (2017). A simple correction to completer analyses and improvement on baseline observation carried forward. *Biometrics*, 73(4), 1366-1378. doi:10.1111/biom.12661
75. Phillips, A., & Haudiquet, V. (2003). ICH E9 guideline 'Statistical principles for clinical trials': a case study. *Stat Med*, 22(1), 1-11; discussion 13-17. doi:10.1002/sim.1328
76. Potkin, S. G., & Siu, C. O. (2009). Dropouts and missing data in psychiatric clinical trials. *Am J Psychiatry*, 166(11), 1295; author reply 1295-1296. doi:10.1176/appi.ajp.2009.09070959
77. Powney, M., Williamson, P., Kirkham, J., & Kolamunnage-Dona, R. (2014). A review of the handling of missing longitudinal outcome data in clinical trials. *Trials*, 15, 237. doi:10.1186/1745-6215-15-237
78. Rabideau, D. J., Nierenberg, A. A., Sylvia, L. G., Friedman, E. S., Bowden, C. L., Thase, M. E., . . . Schoenfeld, D. A. (2014). A novel application of the Intent to Attend assessment to reduce bias due to missing data in a randomized controlled clinical trial. *Clin Trials*, 11(4), 494-502. doi:10.1177/1740774514531096
79. Raboud, J. M., Singer, J., Thorne, A., Schechter, M. T., & Shafran, S. D. (1998). Estimating the effect of treatment on quality of life in the presence of missing data due to drop-out and death. *Qual Life Res*, 7(6), 487-494.
80. Rajan, K. B., & Leurgans, S. E. (2010). Joint modeling of missing data due to non-participation and death in longitudinal aging studies. *Stat Med*, 29(21), 2260-2268. doi:10.1002/sim.4010
81. Rockhold, F. W., & Ruberg, S. J. (2008). ICH-E9 reflections and considerations. *Pharm Stat*, 7(4), 233-235. doi:10.1002/pst.345
82. Scharfstein, D. O., Hogan, J., & Herman, A. (2012). On the prevention and analysis of missing data in randomized clinical trials: the state of the art. *J Bone Joint Surg Am*, 94 Suppl 1, 80-84. doi:10.2106/jbjs.1.00273
83. Shih, W. (2002). Problems in dealing with missing data and informative censoring in clinical trials. *Curr Control Trials Cardiovasc Med*, 3(1), 4.
84. Shih, W. J. (2016). Comments on the three papers by the FDA/CDER research team on the regulatory perspective of the missing data problem. *Stat Med*, 35(17), 2880-2886. doi:10.1002/sim.6895
85. Singhal, R., & Rana, R. (2014). Intricacy of missing data in clinical trials: Deterrence and management. *Int J Appl Basic Med Res*, 4(Suppl 1), S2-5. doi:10.4103/2229-516x.140706
86. Smith, L. J., Wiessinger, D., West, D., & Pittman, T. (2014). Murky definitions, missing data

- prevent meaningful conclusions. *Pediatrics*, 134(6), e1776. doi:10.1542/peds.2014-2794A
87. Sobel, M. E., & Muthen, B. (2012). Compliance mixture modelling with a zero-effect complier class and missing data. *Biometrics*, 68(4), 1037-1045. doi:10.1111/j.1541-0420.2012.01791.x
 88. Soon, G. (2009). Editorial: missing data--prevention and analysis. *J Biopharm Stat*, 19(6), 941-944. doi:10.1080/10543400903331226
 89. Soon, G. (2009). Minimizing Missing Data in Clinical Trials: Design, Operation, and Regulatory Considerations. *Drug Information Journal*, 43(4), 485-492.
 90. Sridhara, R., Mandrekar, S. J., & Dodd, L. E. (2013). Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clin Cancer Res*, 19(10), 2613-2620. doi:10.1158/1078-0432.ccr-12-2938
 91. Stack, C. B., Butterworth, T., & Goldin, R. (2018). Designed Learning: Missing Data in Clinical Research. *Ann Intern Med*, 168(10), 744. doi:10.7326/m18-0534
 92. Stevens, J. R., Suyundikov, A., & Slattery, M. L. (2016). Accounting for Missing Data in Clinical Research. *JAMA*, 315(5), 517-518. doi:10.1001/jama.2015.16461
 93. Streiner, D., & Geddes, J. (2001). Intention to treat analysis in clinical trials when there are missing data. *Evid Based Ment Health*, 4(3), 70-71.
 94. Sullivan, T. R., Yelland, L. N., Lee, K. J., Ryan, P., & Salter, A. B. (2017). Treatment of missing data in follow-up studies of randomised controlled trials: A systematic review of the literature. *Clin Trials*, 14(4), 387-395. doi:10.1177/1740774517703319
 95. Tanaka, S., Fukinbara, S., Tsuchiya, S., Suganami, H., & Ito, Y.M. (2014). Current Practice in Japan for the Prevention and Treatment of Missing Data in Confirmatory Clinical Trials. *Therapeutic Innovation & Regulatory Science*, 48(6), 717-723.
 96. Taylor, L., & Zhou, X. H. (2009). Multiple imputation methods for treatment noncompliance and nonresponse in randomized clinical trials. *Biometrics*, 65(1), 88-95. doi:10.1111/j.1541-0420.2008.01023.x
 97. Tchetgen, E. J. T. (2009). A Commentary on G. Molenberghs's Review of Missing Data Methods. *Drug Information Journal*, 43(4), 433-435.
 98. Touloumi, G., Pocock, S. J., Babiker, A. G., & Darbyshire, J. H. (2002). Impact of missing data due to selective dropouts in cohort studies and clinical trials. *Epidemiology*, 13(3), 347-355.
 99. Troxel, A. B., Fairclough, D. L., Curran, D., & Hahn, E. A. (1998). Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat Med*, 17(5-7), 653-666.
 100. van der Laan, M., & Lorenz, T. J. (2011). Addressing missing data in clinical trials. *Ann Intern Med*, 155(2), 135. doi:10.7326/0003-4819-155-2-201107190-00016
 101. Vickers, A. J., & McCarney, R. (2003). Use of a single global assessment to reduce missing data in a clinical trial with follow-up at one year. *Control Clin Trials*, 24(6), 731-735.
 102. Wang, M. D., Liu, J., Molenberghs, G., & Mallinckrodt, C. (2018). An evaluation of the trimmed

- mean approach in clinical trials with dropout. *Pharm Stat*, 17(3), 278-289. doi:10.1002/pst.1858
103. Wei, L. (2011). An alternative way to classify missing data mechanism in clinical trials--a dialogue on missing data. *J Biopharm Stat*, 21(2), 355-361. doi:10.1080/10543406.2011.550115
104. White, I. R., Horton, N. J., Carpenter, J., & Pocock, S. J. (2011). Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ*, 342, d40. doi:10.1136/bmj.d40
105. Wisniewski, S. R., Leon, A. C., Otto, M. W., & Trivedi, M. H. (2006). Prevention of missing data in clinical research studies. *Biol Psychiatry*, 59(11), 997-1000. doi:10.1016/j.biopsych.2006.01.017
106. Witkiewitz, K., Falk, D. E., Kranzler, H. R., Litten, R. Z., Hallgren, K. A., O'Malley, S. S., & Anton, R. F. (2014). Methods to analyze treatment effects in the presence of missing data for a continuous heavy drinking outcome measure when participants drop out from treatment in alcohol clinical trials. *Alcohol Clin Exp Res*, 38(11), 2826-2834. doi:10.1111/acer.12543
107. Wittes, J. (2009). Missing inaction: preventing missing outcome data in randomized clinical trials. *J Biopharm Stat*, 19(6), 957-968. doi:10.1080/10543400903239825
108. Yan, W., Hu, Y., & Geng, Z. (2012). Identifiability of causal effects for binary variables with baseline data missing due to death. *Biometrics*, 68(1), 121-128. doi:10.1111/j.1541-0420.2011.01653.x
109. Zhang, J. (2009). Sensitivity Analysis of Missing Data: Case Studies Using Model-Based Multiple Imputation. *Drug Information Journal*, 43(4), 475-484.
110. 折笠 秀樹. (2013). 欠測データの取り扱い 最新の動向 欠測データの取り扱い 概論. *臨床薬理*, 44(Suppl), S199
111. 折笠 秀樹. (2014). 【臨床試験における欠測データの取扱い:最新の動向】 概論. *臨床薬理*, 45(4), 125-129.
112. 藤後 修. (2013). 欠測データの取り扱い 最新の動向 欠測データの取り扱い インダストリーの立場から. *臨床薬理*, 44(Suppl), S200.
113. 藤後 修. (2014). 【臨床試験における欠測データの取扱い:最新の動向】 インダストリーの立場から. *臨床薬理*, 45(4), 135-140.
114. 飛田 英祐. (2013). 欠測データの取り扱い 最新の動向 欠測データの取り扱い 最新の動向 レギュラトリーの立場から. *臨床薬理*, 44(Suppl), S200.
115. 飛田 英祐. 原 綾子. 安藤 友紀. (2014). 【臨床試験における欠測データの取扱い:最新の動向】 レギュレーションの立場から. *臨床薬理*, 45(4), 141-146.
116. 松井 研一. (2014). 【臨床試験における欠測データの取扱い:最新の動向】 事後対応と事前対策. *臨床薬理*, 45(4), 147-149.
117. 松山 裕. (2013). 【循環器病学における臨床研究-いかに確実に臨床に還元するか】 臨床研究をデザインするうえで知っておきたい知識・最新トピックス 欠測データにどのように対応するか. *医学のあゆみ*, 244(13), 1258-1262.

118. 松山 裕. (2014). 糖尿病と実践臨床統計学 欠測データの取り扱い. *Diabetes Frontier*, 25(4), 483-489.
119. 森田 智視. (2013). 欠測データの取り扱い 最新の動向 経時測定データ解析と欠測データ対応. *臨床薬理*, 44(Suppl), S199.
120. 森田 智視. (2014). 【臨床試験における欠測データの取扱い:最新の動向】 アカデミアの立場から. *臨床薬理*, 45(4), 131-134.

(3) 解析対象集団

- 1 Abraha, I., & Montedori, A. (2010). Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ*, *340*(jun14 1), c2697-c2697.
- 2 Aras, G. (2001). Superiority, Noninferiority, Equivalence, and Bioequivalence—Revisited. *Drug Information Journal*, *35*(4), 1157-1164.
- 3 Bang, H., & Davis, C. E. (2007). On estimating treatment effects under non-compliance in randomized clinical trials: are intent-to-treat or instrumental variables analyses perfect solutions? *Statistics in Medicine*, *26*(5), 954-964.
- 4 Berman, R. S. (1992). Format and Content of a Statistical Data Item. *Drug Information Journal*, *26*(2), 187-190.
- 5 Brittain, E., & Lin, D. (2005). A comparison of intent-to-treat and per-protocol results in antibiotic non-inferiority trials. *Statistics in Medicine*, *24*(1), 1-10.
- 6 Brown, D., Day, S., Hemmings, R., & Wright, D. (2008). Assessing the impact of ICH E9. *Pharmaceutical Statistics*, *7*(2), 77-87.
- 7 Chassany, O., Sagnier, P., Marquis, P., Fullerton, S., & Aaronson, N. (2002). Patient-Reported Outcomes: The Example of Health-Related Quality of Life—A European Guidance Document for the Improved Integration of Health-Related Quality of Life Assessment in the Drug Regulatory Process. *Drug Information Journal*, *36*(1), 209-238.
- 8 Dodd, S., White, I. R., & Williamson, P. (2012). Nonadherence to treatment protocol in published randomised controlled trials: a review. *Trials*, *13*(1).
- 9 Dossing, A., Tarp, S., Furst, D. E., Gluud, C., Beyene, J., Hansen, B. B., ... Christensen, R. (2014). Interpreting trial results following use of different intention-to-treat approaches for preventing attrition bias: a meta-epidemiological study protocol. *BMJ Open*, *4*(9), e005297-e005297.
- 10 Dossing, A., Tarp, S., Furst, D. E., Gluud, C., Wells, G. A., Beyene, J., ... Christensen, R. (2016). Modified intention-to-treat analysis did not bias trial results. *Journal of Clinical Epidemiology*, *72*, 66-74.
- 11 Ellenberg, J. H. (1996). Intent-to-Treat Analysis versus As-Treated Analysis. *Drug Information Journal*, *30*(2), 535-544.
- 12 Ferner, U. J., & Clarke, M. R. (1995). American, Japanese, and European Regulatory Guidelines with Respect to Biostatistics: a Personal View. *Drug Information Journal*, *29*(4), 1337-1343.
- 13 Giorgi Rossi, P. (2014). Screening: the information individuals need to support their decision: per protocol analysis is better than intention-to-treat analysis at quantifying potential benefits and harms of screening. *BMC Medical Ethics*, *15*(1).

- 14 Given, B., Given, C. W., Sikorskii, A., You, M., McCorkle, R., & Champion, V. (2009). Analyzing Symptom Management Trials: The Value of Both Intention-to-Treat and Per-Protocol Approaches. *Oncology Nursing Forum*, *36*(6), E293-E302.
- 15 Hartung, J., & Cottrell, J. E. (2005). Intention-To-Treat Analysis Should Be Supplemented by On-Treatment Analysis. *Journal of Neurosurgical Anesthesiology*, *17*(2), 63-64.
- 16 Hauck, W. W., & Anderson, S. (1999). Some Issues in the Design and Analysis of Equivalence Trials. *Drug Information Journal*, *33*(1), 109-118.
- 17 Hauschke, D. (2001). Choice of Delta: A Special Case. *Drug Information Journal*, *35*(3), 875-879.
- 18 Hernán, M. A., & Robins, J. M. (2017). Per-Protocol Analyses of Pragmatic Trials. *New England Journal of Medicine*, *377*(14), 1391-1398.
- 19 Keene, O. N. (2011). Intent-to-treat analysis in the presence of off-treatment or missing data. *Pharmaceutical Statistics*, *10*(3), 191-195.
- 20 Lange, S. (2001). The All Randomized/Full Analysis Set (ICH E9)—May Patients Be Excluded from the Analysis? *Drug Information Journal*, *35*(3), 881-891.
- 21 Leuchs, A., Brandt, A., Zinserling, J., & Benda, N. (2017). Disentangling estimands and the intention-to-treat principle. *Pharmaceutical Statistics*, *16*(1), 12-19.
- 22 Lou, Y., Jones, M. P., & Sun, W. (2018). Estimation of causal effects in clinical endpoint bioequivalence studies in the presence of intercurrent events: noncompliance and missing data. *Journal of Biopharmaceutical Statistics*, *29*(1), 151-173.
- 23 Matilde Sanchez, M., & Chen, X. (2006). Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat. *Statistics in Medicine*, *25*(7), 1169-1181.
- 24 Matsuyama, Y. (2010). A comparison of the results of intent-to-treat, per-protocol, and g-estimation in the presence of non-random treatment changes in a time-to-event non-inferiority trial. *Statistics in Medicine*, *29*(20), 2107-2116.
- 25 McNamee, R. (2009). Intention to treat, per protocol, as treated and instrumental variable estimators given non-compliance and effect heterogeneity. *Statistics in Medicine*, *28*(21), 2639-2652.
- 26 Montedori, A., Bonacini, M. I., Casazza, G., Luchetta, M. L., Duca, P., Cozzolino, F., & Abraha, I. (2011). Modified versus standard intention-to-treat reporting: Are there differences in methodological quality, sponsorship, and findings in randomized trials? A cross-sectional study. *Trials*, *12*(1).
- 27 Permutt, T., & Li, F. (2017). Trimmed means for symptom trials with dropouts.

Pharmaceutical Statistics, 16(1), 20-28.

- 28 Phillips, A., Ebbutt, A., France, L., & Morgan, D. (2000). The International Conference on Harmonization Guideline “Statistical Principles for Clinical Trials”: Issues in Applying the Guideline in Practice. *Drug Information Journal*, 34(2), 337-348.
- 29 Porta, N., Bonet, C., & Cobo, E. (2007). Discordance between reported intention-to-treat and per protocol analyses. *Journal of Clinical Epidemiology*, 60(7), 663-669.
- 30 Ranganathan, P., Pramesh, C., & Aggarwal, R. (2016). Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. *Perspectives in Clinical Research*, 7(3), 144.
- 31 Rees, J. (2002). Are we going OTT about ITT? *British Journal of Dermatology*, 146(2), 331-331.
- 32 Rosenkranz, G. K. (2013). Analysis Sets and Inference in Clinical Trials. *Therapeutic Innovation & Regulatory Science*, 4(4), 455-459.
- 33 Scott, P. E., & Campbell, G. (1998). Interpretation of Subgroup Analyses in Medical Device Clinical Trials. *Drug Information Journal*, 32(1), 213-220.
- 34 Stewart, W. C., Jackson, A. L., & Jenkins, J. N. (2004). Dropout rates for intent-to-treat and per protocol analyses. *American Journal of Ophthalmology*, 137(4), 639-645.
- 35 Stewart, W. H. (2004). Basing Intention-to-Treat on Cause and Effect Criteria. *Drug Information Journal*, 38(4), 361-369.
- 36 Svensson, K. (1995). How Many Populations Must be Analyzed and How Should They be Defined (Intention-To-Treat, Eligible, per Protocol Population, Etc.)? *Drug Information Journal*, 29(2), 475-477.
- 37 Swanson, S. A., Holme, Ø., Løberg, M., Kalager, M., Bretthauer, M., Hoff, G., ... Hernán, M. A. (2015). Bounding the per-protocol effect in randomized trials: an application to colorectal cancer screening. *Trials*, 16(1).
- 38 Van der Weijden, F., Slot, D. E., Rosema, M., & Timmerman, M. (2008). ITT in respect to GCP: a matter of diligence. *Journal of Clinical Periodontology*, 35(8), 681-682.
- 39 Wiens, B. L. (2001). Something for Nothing in Noninferiority/Superiority Testing: A Caution. *Drug Information Journal*, 35(1), 241-245.
- 40 Williams, H. (2001). Are we going OTT about ITT? *British Journal of Dermatology*, 144(6), 1101-1102.

(4) 安定性

- 1 Akacha, M., Bretz, F., & Ruberg, S. (2017a). Estimands in clinical trials-broadening the perspective. *Statistics in medicine*, 36(1), 5-19.
- 2 Akacha, M., & Kothny, W. (2017b). Estimands: A More Strategic Approach to Study Design and Analysis. *Clinical Pharmacology & Therapeutics*, 102(6), 894-896.
- 3 Akacha, M., & Ogundimu, E. O. (2016). Sensitivity analyses for partially observed recurrent event data. *Pharmaceutical statistics*, 15(1), 4-14.
- 4 Andrea, R., Scharfstein, D., Su, T. L., & Robins, J. (2001). Methods for conducting sensitivity analysis of trials with potentially nonignorable competing causes of censoring. *Biometrics*, 57(1), 103-113.
- 5 Ayele, B. T., Lipkovich, I., Molenberghs, G., & Mallinckrodt, C. H. (2014). A multiple-imputation-based approach to sensitivity analyses and effectiveness assessments in longitudinal clinical trials. *Journal of biopharmaceutical statistics*, 24(2), 211-228.
- 6 Bhattacharya, S., Fyfe, G., Gray, R. J., & Sargent, D. J. (2009). Role of sensitivity analyses in assessing progression-free survival in late-stage oncology trials. *Journal of clinical oncology*, 27(35), 5958-5964.
- 7 Brown, D., Day, S., Hemmings, R., & Wright, D. (2008). Assessing the impact of ICH E9. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, 7(2), 77-87.
- 8 Cresswell, L., & Mander, A. P. (2014). How to use published complete case results from weight loss studies in a missing data sensitivity analysis. *Obesity*, 22(4), 996-1001.
- 9 Curran, D., Molenberghs, G., Thijs, H., & Verbeke, G. (2004). Sensitivity analysis for pattern mixture models. *Journal of biopharmaceutical statistics*, 14(1), 125-143.
- 10 Daniels, M. J., & Hogan, J. W. (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics*, 56(4), 1241-1248.
- 11 de Souza, R. J., Eisen, R. B., Perera, S., Bantoto, B., Bawor, M., Dennis, B. B., Samaan, Z., Thabane, L. (2015). Best (but oft-forgotten) practices: sensitivity analyses in randomized controlled trials. *The American journal of clinical nutrition*, 103(1), 5-17.
- 12 Egleston, B. L., Cropsey, K. L., Lazev, A. B., & Heckman, C. J. (2010). A tutorial on principal stratification-based sensitivity analysis: application to smoking cessation studies. *Clinical Trials*, 7(3), 286-298.
- 13 Fletcher, C., Tsuchiya, S., & Mehrotra, D. V. (2017). Current Practices in Choosing Estimands and Sensitivity Analyses in Clinical Trials: Results of the ICH E9 Survey. *Therapeutic Innovation & Regulatory Science*, 51(1), 69-76.
- 14 Flyer, P. A. (2009). Discussion: "Incomplete data in clinical studies: Analysis, sensitivity, and sensitivity analysis" by Geert Molenberghs. *Drug Information Journal*, 43(4), 437-439.

- 15 Fox, M. P., Lash, T. L., & Hamer, D. H. (2005). A sensitivity analysis of a randomized controlled trial of zinc in treatment of falciparum malaria in children. *Contemporary clinical trials*, 26(3), 281-289.
- 16 Gao, F., Liu, G. F., Zeng, D., Xu, L., Lin, B., Diao, G., Golm, G., Heyse, J. F., Ibrahim J. G. (2017). Control-based imputation for sensitivity analyses in informative censoring for recurrent event data. *Pharmaceutical statistics*, 16(6), 424-432.
- 17 Gilbert, P. B., Shepherd, B. E., & Hudgens, M. G. (2013). Sensitivity analysis of per-protocol time-to-event treatment efficacy in randomized clinical trials. *Journal of the American Statistical Association*, 108(503), 789-800.
- 18 Gordon, K. (2008). Who wants to be an ICH Euro billionaire?. *Pharmaceutical statistics*, 7(1), 4-8.
- 19 Hirotsu, C., & Hothorn, L. A. (2003). Impact of the ICH E9 Guideline statistical principles for clinical trials on the conduct of clinical trials in Japan. *Drug information journal*, 37(4), 381-395.
- 20 Hollis, S. (2002). A graphical sensitivity analysis for clinical trials with non-ignorable missing binary outcome. *Statistics in Medicine*, 21(24), 3823-3834.
- 21 Huang, X., & Zhang, N. (2008). Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach. *Biometrics*, 64(4), 1090-1099.
- 22 Ibrahim, J. G. (2009). Comment: Incomplete data in clinical studies: analysis, sensitivity, and sensitivity analysis. *Drug Information Journal*, 43(4), 431-432.
- 23 Imai, K., & Jiang, Z. (2018). A sensitivity analysis for missing outcomes due to truncation by death under the matched-pairs design. *Statistics in medicine*, 37(20), 2907-2922.
- 24 International Conference on Harmonisation E9 Expert Working Group. (1999). Statistical principles for clinical trials: ICH harmonized tripartite guideline. *Statistics in Medicine*, 18(15), 1905-1942.
- 25 Jo, B. (2002). Model misspecification sensitivity analysis in estimating causal effects of interventions with non-compliance. *Statistics in Medicine*, 21(21), 3161-3181.
- 26 Kaciroti, N. A., & Raghunathan, T. (2014). Bayesian sensitivity analysis of incomplete data: bridging pattern-mixture and selection models. *Statistics in medicine*, 33(27), 4841-4857.
- 27 Kaciroti, N. A., Schork, M. A., Raghunathan, T., & Julius, S. (2009). A Bayesian sensitivity model for intention-to-treat analysis on binary outcomes with dropouts. *Statistics in medicine*, 28(4), 572-585.
- 28 Keene, O. N., Roger, J. H., Hartley, B. F., & Kenward, M. G. (2014). Missing data sensitivity analysis for recurrent event data using controlled imputation. *Pharmaceutical statistics*, 13(4), 258-264.

- 29 Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in medicine*, 17(23), 2723-2732.
- 30 Koch, G. G., & Wiener, L. E. (2016). Commentary for the Missing Data Working Group's perspective for regulatory clinical trials, estimands, and sensitivity analyses. *Statistics in medicine*, 35(17), 2887-2893.
- 31 LaVange, L. M. (2014). The role of statistics in regulatory decision making. *Therapeutic Innovation & Regulatory Science*, 48(1), 10-19.
- 32 Lee, A. H., Xiang, L., & Fung, W. K. (2004). Sensitivity of score tests for zero-inflation in count data. *Statistics in medicine*, 23(17), 2757-2769.
- 33 Leuchs, A. K., Zinserling, J., Brandt, A., Wirtz, D., & Benda, N. (2015). Choosing appropriate estimands in clinical trials. *Therapeutic innovation & regulatory science*, 49(4), 584-592.
- 34 Leuchs, A. K., Brandt, A., Zinserling, J., & Benda, N. (2017). Disentangling estimands and the intention-to-treat principle. *Pharmaceutical statistics*, 16(1), 12-19.
- 35 Lewis, J. (2008). 10 years of ICH E9. *Pharmaceutical statistics*, 7(1), 1-3.
- 36 Lewis, J., Louv, W., Rockhold, F., & Sato, T. (2001). The impact of the international guideline entitled Statistical Principles for Clinical Trials (ICH E9). *Statistics in medicine*, 20(17 - 18), 2549-2560.
- 37 Lewis, J. A. (1999). Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Statistics in medicine*, 18(15), 1903-1942.
- 38 Li, S., Koch, G. G., Preisser, J. S., Lam, D., & Sanchez-Kam, M. (2017). Sensitivity analysis for missing dichotomous outcome data in multi-visit randomized clinical trial with randomization-based covariance adjustment. *Journal of biopharmaceutical statistics*, 27(3), 387-398.
- 39 Lin, N. X., Logan, S., & Henley, W. E. (2013). Bias and sensitivity analysis when estimating treatment effects from the Cox model with omitted covariates. *Biometrics*, 69(4), 850-860.
- 40 Lipkovich, I., Ratitch, B., & O'Kelly, M. (2016). Sensitivity to censored-at-random assumption in the analysis of time-to-event endpoints. *Pharmaceutical statistics*, 15(3), 216-229.
- 41 Liu, M., Daniels, M. J., & Perri, M. G. (2015). Quantile regression in the presence of monotone missingness with sensitivity analysis. *Biostatistics*, 17(1), 108-121.
- 42 Liu, T., & Heitjan, D. F. (2012). Sensitivity of the discrete - time Kaplan–Meier estimate to nonignorable censoring: Application in a clinical trial. *Statistics in medicine*, 31(25), 2998-3010.
- 43 Liu, Y. (2017). Sensitivity analyses for informative censoring in survival data: A trial example. *Journal of biopharmaceutical statistics*, 27(4), 595-610.
- 44 Liublinska, V., & Rubin, D. B. (2014). Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. *Statistics in medicine*, 33(24), 4170-4185.

- 45 Lu, X., Mehrotra, D. V., & Shepherd, B. E. (2013). Rank-based principal stratum sensitivity analyses. *Statistics in medicine*, 32(26), 4526-4539.
- 46 Mallinckrodt, C., Roger, J., Chuang-Stein, C., Molenberghs, G., O'Kelly, M., Ratitch, B., Janssens, M., Bunouf, P. (2014). Recent developments in the prevention and treatment of missing data. *Therapeutic Innovation & Regulatory Science*, 48(1), 68-80.
- 47 Mallinckrodt, C. H., & Kenward, M. G. (2009). Conceptual considerations regarding endpoints, hypotheses, and analyses for incomplete longitudinal clinical trial data. *Drug Information Journal*, 43(4), 449-458.
- 48 Mallinckrodt, C. H., Lin, Q., & Molenberghs, M. (2013). A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. *Pharmaceutical statistics*, 12(1), 1-6.
- 49 Marcus, S. M. (2001). A sensitivity analysis for subverting randomization in controlled trials. *Statistics in medicine*, 20(4), 545-555.
- 50 Matsuyama, Y. (2003). Sensitivity analysis for the estimation of rates of change with non-ignorable drop-out: an application to a randomized clinical trial of the vitamin D3. *Statistics in medicine*, 22(5), 811-827.
- 51 Mehrotra, D. V., Hemmings, R. J., Russek-Cohen, E., & ICH E9/R1 Expert Working Group. (2016). Seeking harmony: estimands and sensitivity analyses for confirmatory clinical trials. *Clinical Trials*, 13(4), 456-458.
- 52 Mehrotra, D. V., Liu, F., & Permutt, T. (2017). Missing data in clinical trials: control-based mean imputation and sensitivity analysis. *Pharmaceutical statistics*, 16(5), 378-392.
- 53 Minini, P., & Chavance, M. (2004a). Sensitivity analysis of longitudinal binary data with non-monotone missing values. *Biostatistics*, 5(4), 531-544.
- 54 Minini, P., & Chavance, M. (2004b). Sensitivity analysis of longitudinal normal data with drop-outs. *Statistics in medicine*, 23(7), 1039-1054.
- 55 Molenberghs, G. (2009a). Incomplete data in clinical studies: analysis, sensitivity, and sensitivity analysis. *Drug Information Journal*, 43(4), 409-429.
- 56 Molenberghs, G. (2009b). Incomplete data in clinical studies: analysis, sensitivity, and sensitivity analysis—Rejoinder. *Drug Information Journal*, 43(4), 447-448.
- 57 Moreno-Betancur, M., & Chavance, M. (2016). Sensitivity analysis of incomplete longitudinal data departing from the missing at random assumption: Methodology and application in a clinical trial with drop-outs. *Statistical methods in medical research*, 25(4), 1471-1489.
- 58 Moreno-Betancur, M., Rey, G., & Latouche, A. (2015). Direct likelihood inference and sensitivity analysis for competing risks regression with missing causes of failure. *Biometrics*, 71(2), 498-507.

- 59 Morris, T. P., Kahan, B. C., & White, I. R. (2014). Choosing sensitivity analyses for randomised trials: principles. *BMC medical research methodology*, *14*(1), 11.
- 60 Permutt, T. (2016). Sensitivity analysis for missing data in regulatory submissions. *Statistics in medicine*, *35*(17), 2876-2879.
- 61 Phillips, A., Abellan-Andres, J., Soren, A., Bretz, F., Fletcher, C., France, L., Garrett, A., Harris, R., Kjaer, M., Keene, O., Morgan, D., O'Kelly, M., Roger, J. (2017). Estimands: discussion points from the PSI estimands and sensitivity expert group. *Pharmaceutical statistics*, *16*(1), 6-11.
- 62 Phillips, A., & Haudiquet, V. (2003). ICH E9 guideline 'Statistical principles for clinical trials': a case study. *Statistics in medicine*, *22*(1), 1-11.
- 63 Rezvan, P. H., White, I. R., Lee, K. J., Carlin, J. B., & Simpson, J. A. (2015). Evaluation of a weighting approach for performing sensitivity analysis after multiple imputation. *BMC medical research methodology*, *15*(1), 83.
- 64 Rockhold, F. W., & Ruberg, S. J. (2008). ICH - E9 reflections and considerations. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, *7*(4), 233-235.
- 65 Ruan, P. K., & Gray, R. J. (2008). Sensitivity analysis of progression-free survival with dependent withdrawal. *Statistics in medicine*, *27*(8), 1180-1198.
- 66 Salim, A., Mackinnon, A., & Griffiths, K. (2008). Sensitivity analysis of intention-to-treat estimates when withdrawals are related to unobserved compliance status. *Statistics in medicine*, *27*(8), 1164-1179.
- 67 Scharfstein, D., McDermott, A., Díaz, I., Carone, M., Lunardon, N., & Turkoz, I. (2018). Global sensitivity analysis for repeated measures studies with informative drop-out: A semi-parametric approach. *Biometrics*, *74*(1), 207-219.
- 68 Scharfstein, D. O., & Achy-Brou, A. (2009). Global Sensitivity Analysis for Randomized Trials With Informative Dropout: A Semiparametric Perspective. *Drug Information Journal*, *43*(4), 441-446.
- 69 Shaffer, M. L., & Chinchilli, V. M. (2007). Including multiple imputation in a sensitivity analysis for clinical trials with treatment failures. *Contemporary clinical trials*, *28*(2), 130-137.
- 70 Shardell, M., & El-Kamary, S. S. (2009). Sensitivity analysis of informatively coarsened data using pattern mixture models. *Journal of biopharmaceutical statistics*, *19*(6), 1018-1038.
- 71 Shen, S., Beunckens, C., Mallinckrodt, C., & Molenberghs, G. (2006). A local influence sensitivity analysis for incomplete longitudinal depression data. *Journal of biopharmaceutical statistics*, *16*(3), 365-384.

- 72 Shepherd, B. E., Gilbert, P. B., & Dupont, C. T. (2011). Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected postrandomization and relaxing monotonicity. *Biometrics*, *67*(3), 1100-1110.
- 73 Siannis, F., Copas, J., & Lu, G. (2005). Sensitivity analysis for informative censoring in parametric survival models. *Biostatistics*, *6*(1), 77-91.
- 74 Tang, Y. (2018). Controlled pattern imputation for sensitivity analysis of longitudinal binary and ordinal outcomes with nonignorable dropout. *Statistics in medicine*, *37*(9), 1467-1481.
- 75 Thabane, L., Mbuagbaw, L., Zhang, S., Samaan, Z., Marcucci, M., Ye, C., Thabane, M., Giangregorio, L., Dennis, B., Kosa, D., Borg Debono, V., Dillenburg, R., Fruci, V., Bawor, M., Lee, J., Wells, G., Goldsmith, C. H. (2013). A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC medical research methodology*, *13*(1), 92.
- 76 Unnebrink, K., & Windeler, J. (1999). Sensitivity analysis by worst and best case assessment: is it really sensitive?. *Drug information journal*, *33*(3), 835-839.
- 77 US Food and Drug Administration, International Conference on Harmonisation (ICH). (1998). Guidance on statistical principles for clinical trials; availability. *Federal Reg*, *63*, 49583-98.
- 78 Van Steen, K., Curran, D., & Molenberghs, G. (2001). Sensitivity analysis of longitudinal binary quality of life data with drop-out: an example using the EORTC QLQ - C30. *Statistics in medicine*, *20*(24), 3901-3920.
- 79 Vansteelandt, S., & Goetghebeur, E. (2005). Sense and sensitivity when correcting for observed exposures in randomized clinical trials. *Statistics in Medicine*, *24*(2), 191-210.
- 80 Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E., & Kenward, M. G. (2001a). Sensitivity analysis for nonrandom dropout: a local influence approach. *Biometrics*, *57*(1), 7-14.
- 81 Verbeke, G., Lesaffre, E., & Spiessens, B. (2001b). The practical use of different strategies to handle dropout in longitudinal studies. *Drug Information Journal*, *35*(2), 419-434.
- 82 Wang, C., & Daniels, M. J. (2011). A note on MAR, identifying restrictions, model comparison, and sensitivity analysis in pattern mixture models with and without covariates for incomplete data. *Biometrics*, *67*(3), 810-818.
- 83 Wolff, S. N. (2010). Applying Sensitivity Analyses to Overall Survival in Cancer Clinical Trials. *Journal of Clinical Oncology*, *28*(19), e323-e323.
- 84 Wright, C. C., & Sim, J. (2003). Intention-to-treat approach to data from randomized controlled trials: a sensitivity analysis. *Journal of clinical epidemiology*, *56*(9), 833-842.
- 85 Xie, H., & Heitjan, D. F. (2004). Sensitivity analysis of causal inference in a clinical trial subject to crossover. *Clinical Trials*, *1*(1), 21-30.
- 86 Zhang, J. (2009). Sensitivity analysis of missing data: Case studies using model-based multiple imputation. *Drug Information Journal*, *43*(4), 475-484.

- 87 Zhao, Y., Herring, A. H., Zhou, H., Ali, M. W., & Koch, G. G. (2014). A multiple imputation method for sensitivity analyses of time-to-event data with possibly informative censoring. *Journal of biopharmaceutical statistics*, 24(2), 229-253.
- 88 Zhao, Y., Saville, B. R., Zhou, H., & Koch, G. G. (2016). Sensitivity analysis for missing outcomes in time-to-event data with covariate adjustment. *Journal of biopharmaceutical statistics*, 26(2), 269-279.
- 89 佐藤 俊哉. (1999). これからの臨床試験のための統計的原則. *臨床評価*, 27, 73-81.
- 90 田栗 正隆. (2017). 統計トピック: ICH E9(R1)臨床試験におけるestimandと感度分析-最新ガイドライン案の理解のために- 因果効果の定義と推定について. 第38回日本臨床薬理学会学術総会(抄録).
- 91 土屋 悟. (2017). 統計トピック: ICH E9(R1)臨床試験におけるestimandと感度分析-最新ガイドライン案の理解のために- ICH E9(R1)ガイドラインが与える臨床試験の実務への影響. 第38回日本臨床薬理学会学術総会(抄録).
- 92 中林 哲夫. (2018). 臨床試験のデータを読み解く (第4回) 臨床試験の有効性データの確認ポイント (1) 結果の安定性. *臨床精神薬理*, 21(5), 687-696.
- 93 原 綾子. (2017). 統計トピック: ICH E9(R1)臨床試験におけるestimandと感度分析-最新ガイドライン案の理解のために- Estimandの構成にあたっての留意点. 第38回日本臨床薬理学会学術総会(抄録).
- 94 広津 千尋. (2001). 科学技術としての統計的方法. *臨床精神薬理*, 4(6), 763-773.
- 95 富金原 悟. (2017). 統計トピック: ICH E9(R1)臨床試験におけるestimandと感度分析-最新ガイドライン案の理解のために- Estimandと中間事象について. 第38回日本臨床薬理学会学術総会(抄録).
- 96 松井 茂之. (2005). 「薬剤疫学における統計的視点」統計的原則, 今後の課題. *臨床評価*, 32 (2•3), 389-391.