

臨床試験の欠測データの取り扱いに関する 最近の展開と今後の課題について

- NAS レポート, EMA ガイドライン, estimand と解析方法の概説 -

2014 年 7 月（2025 年 2 月再掲）

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

タスクフォース 2

目次

1. 序文.....	6
1.1 はじめに.....	6
1.2 欠測データの影響.....	7
1.3 概要.....	9
1.3.1 欠測データの影響.....	9
1.3.2 NAS レポートと EMA ガイドライン.....	10
1.3.3 Estimand.....	11
1.3.4 欠測データの特徴づけ.....	12
1.3.5 感度分析.....	12
1.3.6 事前検討と事前明記.....	12
2. NAS レポートと EMA ガイドラインの対比と整理.....	14
2.1 背景.....	14
2.2 NAS レポートと EMA ガイドラインに共通するメッセージ.....	14
2.2.1 欠測データを最小限に抑える.....	14
2.2.2 試験治療中止後のデータを収集する.....	15
2.2.3 欠測データの影響を予測する.....	15
2.2.4 欠測データの取り扱いを事前明記する.....	16
2.2.5 感度分析.....	16
2.2.6 尤度に基づく方法について.....	17
2.3 NAS レポートと EMA ガイドラインで立場が異なる点.....	17
2.3.1 LOCF に対する受け入れ可能性.....	17
2.3.2 その他の相違点.....	18
2.4 日本の状況.....	19
3. Estimand と治療中止後のデータ収集.....	21
3.1 Estimand とは.....	21
3.2 試験治療中止後データの収集.....	24
3.3 まとめ.....	26
4. 欠測メカニズム.....	28
5. 解析方法.....	30
5.1 はじめに.....	30
5.2 Complete case analysis.....	30
5.3 Imputation-Based Approaches.....	30
5.4 Inverse Probability Weighting (IPW).....	32
5.5 Mixed Model for Repeated Measure (MMRM).....	32

5.6 Selection Model (SM)	32
5.7 Pattern Mixture Model (PMM).....	33
6. 主要な解析方法の選択と感度分析	34
6.1 はじめに.....	34
6.2 主要な解析方法の選択と感度分析の必要性	34
6.3 感度分析の方針.....	35
6.4 感度分析の解釈時の留意点.....	36
6.5 EMA ガイドラインでの欠測に関する情報の報告	37
6.6 結果のまとめ方の例.....	38
6.7 まとめ	41
7. 欠測データの発生を最小限にするための方策及び試験計画の検討	43
7.1 はじめに.....	43
7.2 試験計画の検討.....	43
7.2.1 Estimand の設定	43
7.2.2 欠測データの発生を最小限に抑える試験デザイン.....	43
7.2.3 サンプルサイズの検討	44
7.2.4 プロトコールへの明記	45
7.3 欠測データの発生を最小限にするための試験実施上の方策.....	46
7.3.1 スポンサーが実施できる方策	46
7.3.2 治験責任医師・分担医師や施設スタッフが実施できる方策.....	46
7.4 まとめ	47
8. 終わりに.....	48
9. 資料作成者	49

<用語>

プロトコール：治験実施計画書，臨床試験実施計画書

完全データ：計画された測定時点に欠測を含まず，すべて測定されたデータ

不完全データ：測定されているデータと欠測データが混在している状態のデータ

欠測メカニズム：症例・時点ごとの，測定されたデータ，欠測データ，（欠測やそもそも測定されることになっていないものも含む）共変量の影響を考慮した，欠測の発生の仕方..

欠測：予定された評価ポイントでデータが得られていないこと

中止・脱落：中止と脱落について，責任医師の判断や臨床試験実施計画書の規定による試験の中止を”中止”とし，試験薬とは無関係な試験の未了や原因を特定できない来院の途絶を”脱落”と区別する考え方がかつてあった．本報告書では，これらを区別せず，被験者が試験を中止することを”中止”と表現する．

Estimand：試験において，“何を推測したいのか”の定義

<略語>

AIPW：Augmented IPW

ANCOVA：Ananalysis of Covariance

BOCF：Baseline Observation Carried Forward

CHMP：Committee for Medicinal Products for Human Use；欧州医薬品委員会

CNSTAT：Committee on National Statistics；国家統計委員会

CPMP：Committee for Proprietary Medicinal Products；欧州医薬品委員会

CRF：Case Report Form

DMC：Data Monitoring Committee

EMA：European Medicines Agency；欧州医薬品庁

FAS：Full Analysis Set

FDA：Food and Drug Administration；アメリカ食品医薬品局

IOM：Institute of Medicine；米国医学研究所

IPW：Inverse Probability Weighting

ITT：Intention-to-Treat

LOCF：Last Observation Carried Forward

MAR：Missing at Random

MCAR：Missing Completely at Random

mITT：modified ITT

MMRM：Mixed Models for Repeated Measures

MNAR：Missing not at Random

NAS：National Academy of Sciences；全米科学アカデミー

NAE：National Academy of Engineering；全米技術アカデミー

NRC : National Research Council ; 全米研究評議会

PMDA : Pharmaceuticals and Medical Devices Agency ; 独立行政法人医薬品医療機器総合機構

PMM : Pattern Mixture Model

PP : Per Protocol

SM : Selection Model

wGEE : Weighted Generalized Estimating Equations

YMRS : Young Mania Rating Scale ; ヤング躁病評価尺度

1. 序文

1.1 はじめに

新薬開発のために実施される一般的な臨床試験では欠測データの発生が不可避である。欠測データを生じさせる最大の問題は、被験薬群と対照薬群の公平な比較が行えなくなることである。例えば、嚢胞性線維症を対象に開発されてきた Dry Powder Mannitol の事例では、2つの検証試験は効果を同わせるものであったが ($P < 0.001$, $P = 0.059$)、中止割合が対照薬群に比べて被験薬群で 5%以上高く (被験薬群 35%, 46%に対し、対照薬群 28%, 36%)、このことが評価に影響を及ぼしていないということを合理的に説明できなかったことを一因として、FDA は本薬を承認しなかった (2013 年; 付録の調査事例参照, LaVange 2014)。

欠測データの問題は ICH E9 ガイドライン (1997 年)においても"臨床試験において偏りを起こし得る代表的な原因である"とされてはいたが、同文書中でも、欠測データを扱う決定的な対処方法が存在しないこと、複数の解析方法によって結果の安定性を検討すべきであることが示されているのみであり、特に日本では近年まで last observation carried forward (LOCF)が標準的な方法として用いられてきた (Tanaka ら, 2014)。一方、海外では、米国において FDA からの要請によって 2010 年に NRC (National Research Council ; 全米研究評議会)が発表した「The prevention and treatment of missing data in clinical trials」(以下, NAS レポート¹)と、ほぼ同時期に EMA の CHMP から発表された「Guideline on Missing Data in Confirmatory Clinical Trial」(以下, EMA ガイドライン)において、LOCF による解析の問題点が指摘されてきた。これらを受けて近年では LOCF 以外の方法を主要解析としたり、複数の解析方法によって結果の安定性を検討したりしていることが公表資料から確認することができる (本報告書添付資料)。先に示した事例の Dry Power Mannitol においても、EMA は複数の解析結果から中止による不確実性は受け入れ可能とし、承認している (2012 年; 付録の調査事例参照)。

欠測データの取り扱いは、高度な知識、経験、リソースを要する複雑な解析を必要とするが、統計解析のみが解決できる問題ではない。欠測データを伴うデータに対する有効な解析方法を決定するためには、まず、その試験で示したいことは何かを特定する必要がある。特に estimand (試験において、“何を推測したいのか”の定義)と呼ばれる最近の考え方に応じた整理が有用である。選択される estimand によっては試験治療中止後のデータ収集が必要になるため、その試験で示すべきことの重要性について関係者と共有することが必要である。次に、欠測データに対する最も有効な方法は欠測データの発生を予防することなので、費用対効果も

¹ 略称は必ずしも統一されていないが、本報告書では LaVange 2014 に従い「NAS レポート」とした。

考えながら、欠測データ発生予防策を検討すべきである。欠測データの発生理由を明確に特定するよう努力することは解析のために重要であるため、臨床試験実施担当部門とデータの取得方法について事前に協議しておくことが必要である。また、解析方法は複雑で、複数の解析結果を総合的に判断する必要があるため、結果の解釈についても報告書作成部門との慎重な協議が必要となる。

本報告書では、まず NRC が発表した「The prevention and treatment of missing data in clinical trials」と CHMP から発表された「Guideline on Missing Data in Confirmatory Clinical Trial」を対比し整理することで、現在、欧米でどのようなことがコンセンサスであり、どのようなことが議論の段階であるのかを示す。次に、試験の目的を整理するための概念である **estimand** について説明する。続いて、欠測データを整理して理解するために必要な欠測メカニズムについて詳述し、具体的な解析方法について示す。そして、主要な解析方法の決定方針と結果を受け入れ可能にするための感度分析について示す。最後に、欠測データの発生を最小化するための方法と、説得力のある解析結果を示すために必要な計画書への事前明記について触れる。

本報告書で目指したことは、欠測データの問題は現在も研究が続いている高度な問題であるということと欠測データの問題への対処は統計解析だけの問題ではないということを経験担当者以外の試験関係者に知ってもらうこと、そして、欧米の方針に差異が存在することから日本の考え方をまとめたガイドラインが必要であるということを示すことである。

本報告書では、統計担当者以外の方に理解していただけるよう、数式や統計的な専門用語は極力使用しないような記載を可能な限り用いたが、理解が困難な部分は統計担当者にお問い合わせいただけると幸いである。なお、付録に、FDA、EMA 及び PMDA での承認審査の過程で、欠測データが議論された事例を調査した結果を添付したので参照して頂きたい。

本報告書内にて参照した英語資料について、公式な和訳が存在しないものについては、本報告書のために報告者らが和訳した。「should」については「べきである」、と和訳している点について留意して頂きたい。

1.2 欠測データの影響

欠測データが臨床試験の計画と結果に対してどのように影響するのかを説明するために、2 つの臨床試験の極端な仮想事例（表 1-1）を考えてみる。

表 1-1. 臨床試験の極端な仮想事例

試験 1			試験 2		
	試験治療	対照治療		試験治療	対照治療
有効	80	20	有効	80	20
無効	20	80	無効	20	80
欠測	10	10	欠測	50	50

どちらの試験も欠測なく効果を観測できたのは 100 例ずつであり、試験 1 では各群 10 例、試験 2 では 50 例ずつの欠測がある。このデータに対して、効果を観測できたデータだけを用いて解析した場合、有効割合の治療間差は両試験ともに 60.0% (48.9%~71.1%), $P<0.001$ である (括弧内は 95%信頼区間, 以下同様)。ここで欠測に 3 つの仮定をおいた時の解析結果を比較してみる。もし欠測が治療とは完全に独立に生じており、欠測データに対して、観測されているデータと同様の効果が観測できていたと仮定した場合を仮定①とする。欠測データはすべて無効であった (つまり、欠測するかどうか結果に依存している) と仮定した場合を仮定②とする。そして、欠測データは、試験治療ではすべてが無効で、対照治療ではすべてが有効である (つまり、最も保守的な状況) と仮定した場合を仮定③とする。以下にそれぞれ解析結果をまとめた。

表 1-2. 欠測に仮定をおいた下での解析結果

条件	試験 1	試験 2
観測データのみ	60.0% (48.9%~71.1%), $P<0.001$	60.0% (48.9%~71.1%), $P<0.001$
仮定①	60.0% (49.4%~70.6%), $P<0.001$	60.0% (51.0%~69.1%), $P<0.001$
仮定②	54.6% (43.5%~65.6%), $P<0.001$	40.0% (30.3%~49.7%), $P<0.001$
仮定③	45.5% (33.7%~57.2%), $P<0.001$	6.7% (-18.0%~4.6%), $P=0.248$

試験 1 では、仮定によって効果の大きさは小さくなっているが、欠測にいかなる仮定をおいても試験治療の効果は示すことができている (全て $P<0.001$)。一方、試験 2 では、仮定③の場合に効果を主張することができなくなっている ($P=0.248$)。これらの仮定は極端な仮定に思えるかもしれないが、「欠測データの全てが欠測でなかったとしたら、それぞれどのような値だったのか」の真実は知るすべがない。真実は、極端な仮定の中間のどこかにあると解釈できるように、極端な仮定が設定されているのである。

上記の検討では、欠測データに対する仮定をすべて同列に扱ったが、実際には各々の欠測データに対して欠測となった理由が存在する。例えば、効果不十分による中止、有害事象による中止、臨床試験とは関係の無い理由による中止などである。解析の際に、これらの理由別に取り扱いを変えることも一般的に行われており、このような対応をするためには欠測理由を収集できるように計画しておく必要がある。また、観測される効果が 2 値 (有

効と無効だけ)ではなく、血圧や血清コレステロールのように測定値そのものを評価する場合は、解析はさらに困難なものとなる。欠測データをどの程度の値とみなすのか、もしくは特定の値とはしないのかを決める必要があり、場合によっては試験の介入治療を中止した後のデータを得ておく必要がある場合もある。

この例を用いて伝えたいことは、欠測データの割合が少ない場合には、どのような仮定においても結果は安定しており、有効性の証拠として受け入れやすいということである。これに対して中止による欠測データが少なからず存在し、中止理由が多様に存在する場合には、結果を一意に表現することが困難になるということである。

1.3 概要

本節では、本報告書で議論している内容の概観を示すことによって、本報告書の全体像の把握を容易にし、以降の議論の理解の助けとしたい。

1.3.1 欠測データの影響

ほとんどの試験において、欠測データが生じる。欠測データが臨床試験の推論に与える影響は以下の3項目である。

1. 精度の低下 (検出力の低下)

推定値 (例えば変化率、有効率、生存割合など)の推定精度はデータ数が多くなれば高くなる (標準誤差が小さくなる)。即ち、欠測データが発生することによって、解析に利用できるデータ数が少なくなれば、推定値の精度が悪くなり (標準誤差が大きくなる)、検出力の低下に繋がりをえる。このことによって、信頼区間が広くなり、P値が大きくなる。場合によっては計画どおりの被験者数を収集し、想定通りの薬剤効果が観察されたとしても有意差が得られないということが起こり得る。

2. バイアスの発生

例えば、状態が悪くなると試験を中止する傾向が存在する場合に、試験に残された被験者のみで推測を行えば、その推定値は真実よりも良い結果となる (図 1-1.) このように結果が真実よりも系統的に偏っていることをバイアスが発生しているといい、無視できないバイアスの危惧がある場合には規制当局は結果を受け入れない可能性がある (1.1 Dry Power Mannitol の例参照)。

3. 試験結果の安定性が損なわれる

全ての欠測データの本当の発生理由はつきとめられない場合が多いし、欠測でなかったとしたらどのような値をとっていたかもわからない。だからこそ、さまざまな仮定を置いて「もしこの仮定が正しいならば、このような結果が得ら

れていたはずである」という推論を行うのである。どのような仮定を置いても結果が大きく異なるならば、結果の安定性は示せるだろう。しかし、仮定の置き方によって結果が大きく異なるのであれば、結果の安定性は損なわれ、結果の解釈において強い主張はしづらくなる。なお、安定性の評価では、仮定の置き方によってどの程度結果が異なるのか、置いた仮定のあり得なさ等を考慮すべきであり、安定性が保たれるか否かは程度問題である。

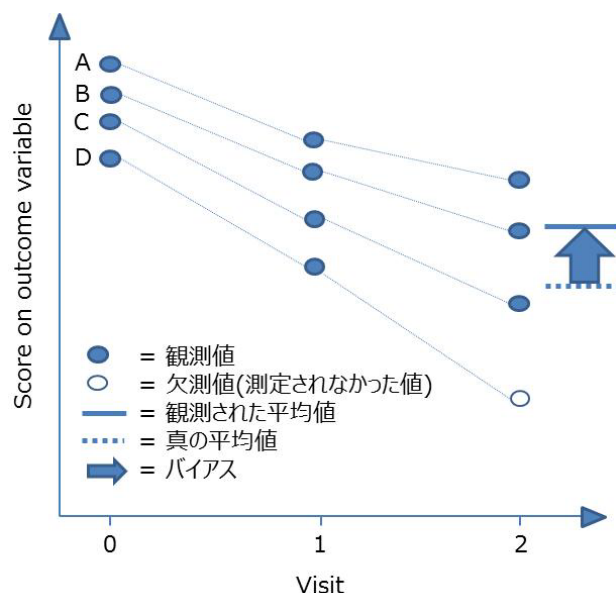


図 1-1. 欠測によるバイアスの例

症例 D が途中中止し、試験に残された被験者のみから得られる visit 2 の平均値は、真の平均値よりも高い値となっている。

1.3.2 NAS レポートと EMA ガイドライン

ICH E9 ガイドライン (1997 年)の本文中には、欠測データに関する記載として、“欠測値は、臨床試験において偏りを起こし得る代表的な原因である。”とあるものの、公表当時は欠測データの扱いが今日ほど注目されてはいなかった。しかし、自覚症状を伴う分野において欠測データの割合が軽視できないほど多かった事例の発生などから欠測データに対する取り扱いが注目され、米国では FDA からの要請により NRC が 2010 年に NAS レポートを公表し、ほぼ同時期に EMA の CHMP が EMA ガイドラインを公表した。これらの文書が述べていることは実際の審査の場においても重要な論点となっている。

二つの文書で共通している点は、欠測データの発生を最小化すべきであること、欠測データの種類と量を予測し、それらの扱いを事前明記すること、幅広い仮定に基づいて解析し結果の安定性を確認することなどである。一方で、欠測データを取り扱う一つの統計的方法である LOCF について、NAS レポートが科学的に正当化できるとき以外は否定的であ

るのに対して、EMA ガイドラインでは保守的であれば受け入れることもあり得るとしている違いがある。これは学術団体による「レポート」と規制当局による「ガイドライン」の違いが出ているところであろう。

1.3.3 Estimand

欠測データの取り扱いについて考えるためには、臨床試験において“何を推測したいのか”を明確にすることが有用である。この“何を推測したいのか”を整理した概念は近年 **estimand** (エスティマンド) と呼ばれている。臨床試験それぞれは、開発計画の中で、明確に位置づけられているはずである。その位置づけは試験の目的、主要評価項目、どのようにいつ/誰の評価項目が測定されるのか、介入効果の測定によって特徴付けられるが、それらは、“何を推測したいのか”が起源となっている。探索的であるか、検証的であるか以外にも、被験者に対する治療方針 (例えば、最初の治療法として選択する薬剤) としての有用性を知りたいのか、薬理学的な効果の大きさが知りたいのか、治療法の効果を知りたいのかなどが考えられる。

治療方針の効果を推測したいのであれば、なんらかの理由によって試験治療を中止しても試験自体は継続され、計画された時点のデータを得ることによって、治療方針としての効果を推測することができる。例えば効果は強いが毒性も強い場合には、治療法に耐えられる確率は高くなくとも、耐えられれば高い効果を得られ、そして、耐えられなかったとしても通常治療へ移行できるならばその治療法は通常治療よりも効果が高い (もしくは有用) と判断できる。この場合には試験治療を中止してもデータは欠測とされない設定が望ましい。例えば、新規抗がん剤について考えている。まず新規治療に挑戦し、治療に失敗した場合には通常治療を行うという場合には、新規治療に挑戦したために通常治療の効果が小さくなるということは考えられる。つまり、新規治療が継続不能となった場合に、打ち切り(censoring)として扱うのではなく、新規治療が終了した後の状況についてもデータを収集することによって、新規治療に挑戦するという行為に対する有用性を評価することができる。

一方で、承認前の臨床試験という限られた条件では、被験者に対する治療方針を推測することには限界があるとする考え方もある。この場合には、試験治療の効果を推測の対象とし、臨床現場の状況や公表されているあらゆるデータから、治療法としての有用性を推測することとなる。例えば、新規生活習慣病治療薬について考える。試験治療中止後のデータ取得について考えた場合、試験という実臨床に比較して特殊な状況であるため、試験中止後のデータを用いた推測が、そのまま試験治療の有用性に対する直接的な推定にならない可能性も考えられる。この場合には、試験治療中止後のデータを用いた推測だけでなく、割り付けた試験治療の効果を推定した後で、実臨床における効果を外部の情報を補完しながら評価する方法が考えられる。

1.3.4 欠測データの特徴づけ

欠測データを特徴付ける考え方に欠測メカニズムという考え方があり、次のような分類がある。欠測データを無視しても（つまり、どのような解析を行ったとしても）結果に偏りを生じない状況が MCAR (missing completely at random)、解析上の工夫によって偏りが無い推測が可能な状況が MAR (missing at random)、そして、解析上の工夫では偏りのない推測を行うことができない状況が MNAR (missing not at random)とよばれる。これらの分類は、主解析と感度分析の方針を決めるために有用である。欠測データがどの欠測メカニズムによって発生したかをデータから確認することはできない。したがって、欠測メカニズムを仮定し、適切なデザイン、主解析や感度分析の選択を行うのである。

1.3.5 感度分析

欠測データに対する感度分析では、異なる欠測メカニズムによる解析を行うことが通常行われる。例えば、最近 MMRM (mixed model repeated measure)と呼ばれる方法によって解析されることがある。この方法は MAR が仮定できる際にはバイアスの無い推測が可能であることが知られているが、先述のようにデータから欠測メカニズムを確認する方法は存在しない。従って、MAR とは異なる欠測メカニズムを仮定した場合での解析を行うことも必要であり、これが感度分析となる。一方で、欠測メカニズムに対する厳密な検討でなくとも、幅広い検討を行った結果の安定性を検討すべきであるという考え方も存在しており、例えば、LOCF などの汎用的に用いられてきた方法を感度分析の一部とすることも可能となる場合もある。

1.3.6 事前検討と事前明記

欠測データの影響は、1.3.1 に示したとおり、単純に利用できるデータが減ることだけではない。そこで、欠測データが試験結果に与える影響の可能性について検討し、それらが検討のとおり生じた場合の結果の安定性についても整理しておく必要がある。感度分析は多くの方法が考えられるため、事後的に感度分析の方法を選択すると、結果が変わらない分析のみを採用しているという疑いを持たれかねない。そのため、主解析における欠測の取扱い及び感度分析の方針は、盲検解除前に特定しておく必要があると考えられる。

また、欠測データへの対処は、解析段階で工夫を凝らすだけでなく、試験計画段階で欠測データの発生を最小限にする検討が重要である。

【参考文献】

- LaVange LM. The Role of Statistics in Regulatory Decision Making. *Therapeutic*

Innovation & Regulatory Science. 2014; 48 (1): 10-19.

- Tanaka S, Fukinbara S, Tsuchiya S, Suganami H, Ito MY, Current Practice in Japan for the Prevention and Treatment of Missing Data in Confirmatory Clinical Trials. *Therapeutic Innovation & Regulatory Science*. Published online before print April 16, 2014
- National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press; 2010.
- 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. Published July 2, 2010.

2. NAS レポートと EMA ガイドラインの対比と整理

2.1 背景

欠測データの問題は研究が活発な分野の一つであり、統計的な理論やその実践において近年多くの進歩があった。その結果、欠測データの問題に対処する基本的な原則やテクニックを提示した文書がいくつか作成されたが、中でも注目すべきは、NAS レポートと EMA ガイドラインである。

ここで、NAS レポートが作成された背景について簡単に説明する。NAS (National Academy of Sciences; 全米科学アカデミー)は、1863 年に民間の非営利団体として設立された米国の学術機関であり、NAE (National Academy of Engineering; 全米技術アカデミー), IOM (Institute of Medicine; 米国医学研究所), NRC (National Research Council; 全米研究評議会)とともに全米アカデミーズを構成している。NRC は全米アカデミーズにおける実務を担っており、さまざまな委員会を通じて議会や規制当局に対して専門的な助言や勧告を行っている。FDA の要請と出資を受けて NRC の CNSTAT (Committee on National Statistics; 国家統計委員会)は、欠測データの予防と取り扱いに関する専門家パネルを設立した。パネルは臨床試験経験のある大学関係の統計家など約 15 名により構成された。彼らは製薬業界や FDA の著名な統計家にインタビューを行い、そのインタビューや彼らの経験を基に 144 ページのレポートを作成し、2010 年に発表した。これが NAS レポートである。

一方、CHMP は EMA の科学委員会の 1 つであり、スポンサーが参照するための科学的または規制に関するガイドライン文書の作成等の活動を行っている。EMA ガイドラインは、CHMP の前身である CPMP が 2001 年に発表した「Points to Consider on Missing Data」に置き換わる形で 2010 年に発表された。

NAS レポートと EMA ガイドラインがいずれも最近の研究結果に基づいて作成されているため、多くの点で共通するメッセージを含んでいることは自然であるが、いくつかの点では立場が異なる。これは、NAS レポートがサイエンスの観点からの「レポート」であるのに対し、EMA ガイドラインはレギュラトリーの観点からの「ガイドライン」であることに起因すると思われる。以下に、共通するメッセージ及びスタンスが異なる点について取り上げる。

2.2 NAS レポートと EMA ガイドラインに共通するメッセージ

2.2.1 欠測データを最小限に抑える

無作為化臨床試験は、既知あるいは未知の要因について平均的にバランスを取ることで治療群間の比較可能性を確保するが、欠測データはこの比較可能性を損ない、治療効果の推定にバイアスを生じさせる可能性がある。この問題に対するアプローチの一つは、適切な解析方法を用いることであるが、現在考えられている不完全データ（測定されているデー

タと欠測データが混在している状態のデータ)に対する解析方法は、データからの検証が不可能な仮定を必要とするため、そこから得られた推論が適切であることを保証することは出来ない。従って、もう一つのアプローチである、欠測データを最小限に抑えることが最善の方法である。

NAS レポートではこの点に関する NAS の考え方が **Recommendation 2** として示されている。

Recommendation 2: 医師、スポンサー、規制当局は、結果データが集積されるまで、一貫してプロトコール規定の治療が維持される被験者が最大になることを目標に試験デザインを検討すべきである。

一方、EMA ガイドラインでも、5 章 **General Recommendation** の中で、欠測データを扱う際に考慮すべき点として、最初に「欠測データの回避」を挙げており、欠測データの存在を可能な限り避けることが極めて重要であると述べている。

2.2.2 試験治療中止後のデータを収集する

NAS レポートは **Recommendation 3** として、治療を中止した被験者の情報を収集し、解析に使用するべきであると提言している。

Recommendation 3: スポンサーは試験の途中でプロトコール規定の治療を中止した被験者に対する重要な結果情報の集積を継続して行うべきである。この情報は記録した上で解析に使われるべきである。

また、EMA ガイドラインも、一般的な提言として、欠測データを避けるために被験者が治療を中止した後にデータ収集を続けることを強く推奨している。

一方、O'Neill and Temple (2012)は、この提言を“意見が分かれるところかもしれない”と述べ、アウトカムスタディとそれ以外の試験で受け入れ可能性が異なると指摘している。また、Mallinckrodt ら (2014)は、治療中止後のデータを解析に含めるかどうかは *estimand* に依存すると述べ、これを整理している。

治療中止後のデータの取得と、*estimand* の関係については、3 章で詳しく述べることにする。

2.2.3 欠測データの影響を予測する

NAS レポートは試験デザインの段階だけでなく、試験実施の段階でも欠測データを減らすことができると述べている。**Recommendation 6** は、欠測データの影響を予測し、試験

デザイン及び試験実施の段階で欠測データを抑えるために取られる手立てについてプロトコールに記載すべきであると提言している。

Recommendation 6: スポンサーは起こり得る欠測データの問題を明確に予測すべきである。特にプロトコールに欠測データに関する章を設け、欠測データの影響をモニターし、制限するために試験デザイン・実施の段階で取られるステップについて記載すべきである。

EMA ガイドラインも、試験をデザインし、使用する統計手法を特定する際に欠測データの発生割合を見積もることは非常に重要であると述べている。

2.2.4 欠測データの取り扱いを事前明記する

試験デザイン及び試験実施の段階で欠測データを減らすための最善の努力をしたとしても、欠測データはしばしば避けられない。不完全データに対する解析方法は、仮定に応じた複数の解析方法が考えられ、仮定にも複数の仮定が考えられるため、主要解析が完全に特定できるように事前明記することは不可欠である。また、欠測のメカニズムとその対処方法について臨床家と議論し、実臨床の現場においてどのような仮定を設定しているのかということ共有することは、結果を解釈できるようにするために重要である。NAS レポートの **Recommendation 9** はこれらの点について提言している。

Recommendation 9: 欠測データを取り扱う統計手法について、スポンサーは欠測データの取り扱い方法をプロトコールに記載すべきである。また、設定した仮定は臨床家にも理解できるようにする必要がある。

EMA ガイドラインも、プロトコールの統計解析のセクションあるいは統計解析計画書に欠測データの取り扱い方法を事前明記し、その妥当性を説明すべきであると述べている。

2.2.5 感度分析

現在考えられている欠測データに対する解析方法は、データからの検証が不可能な仮定を必要とするため、解析結果の安定性を考慮に入れなければならない。結果の安定性を考慮する方法の一つに感度分析がある。感度分析については **Recommendation 15** で次のように提言されている。

Recommendation 15: 感度分析は臨床試験の主要な報告の一部であるべきである。特に欠測のメカニズムの仮定に対する感度分析は必須である。

EMA ガイドラインでも、欠測データが無視できないほど発生する場合には、感度分析を主要解析のサポートとして提示するべきであるとされている。

なお、感度分析についての詳細は 6 章で述べることとする。

2.2.6 尤度に基づく方法について

歴史的に、LOCF のような単一の値による補完方法は欠測データに対する主要なアプローチとして広く使われてきたが、後述するようにいくつかの問題点が指摘されている。そこで近年、代替アプローチとして MMRM をはじめとする尤度に基づく方法が提案されてきた。

NAS レポートでは MMRM という用語は用いていないものの、変量効果モデルは少数のパラメータを用いて高次の多変量分布を単純化するために非常に有用な手段となり得る一方、変量効果に関する仮定はしばしば計算に都合がよいように置かれるため注意が必要であると述べている。

Recommendation 11: パラメトリックモデル、特に変量効果モデルは、設定している仮定についてすべて説明し、正当化した上で注意して用いるべきである。パラメトリックな仮定に基づくモデルを使うときは、モデルの適合度を検討すべきである。

一方 EMA ガイドラインは、MMRM のような尤度に基づく方法は、MAR の仮定の下で、すべての被験者が完全に治療を続けた場合に観察される治療効果の推定値をバイアスなく推定するが、そのような推定値は complete case 解析による推定値と同様になる可能性があり、試験治療に有利なバイアスが入るかもしれないと指摘している。

EMA ガイドラインの方が尤度に基づく方法に対して、やや具体的な懸念を示しているが、欠測を含むデータの解析方法として MMRM のような尤度に基づく方法を無条件に適用するのは避けるべきであるというメッセージは共通している。

2.3 NAS レポートと EMA ガイドラインで立場が異なる点

これまでに説明したとおり、NAS レポートと EMA ガイドラインは多くの点で一致しているが、いくつかの点ではスタンスが異なる。

2.3.1 LOCF に対する受け入れ可能性

LOCF や BOCF 等、単一の値による補完方法は臨床試験の欠測データに対する主要なアプローチとして広く使われてきた。このアプローチは実施が容易であるが、一方でいくつかの問題点が指摘されてきた。

これらの問題点に関する認識は NAS レポートと EMA ガイドラインで共通している。具

体的には：

- （中止後に値が変わらないという）強い仮定を置いており，その仮定が妥当でない場合はバイアスが生じる．すなわち，MCAR, MAR, MNAR の条件だけでは，LOCF の適合の可否は判断すべきでない．
- バイアスの方向は保守的とは限らない
- LOCF の適合が妥当な条件下であっても，不確実性を無視することで標準誤差を過小評価する

NAS レポートはこれらの問題点を踏まえ，前提となる仮定が正当化できない限り主要解析として LOCF や BOCF を使用するべきではないと述べている．

Recommendation 10: LOCF や BOCF のような単一補完法（single imputation method）は，前提となっている仮定が科学的に妥当でない限り，欠測データに対する主要な方法として使うべきではない

一方 EMA ガイドラインは，明らかに保守的な主要解析に基づく結論は規制上の見地から説得力のあるエビデンスとなり得ると述べており，LOCF が保守的な推定値を導く状況は存在すると指摘している．

このように，問題点に対する認識が違わないのにスタンスが違ってくるように見えるのは興味深い，この違いは前述したように「レポート」と「ガイドライン」の違いによるものだろう．つまり，NAS レポートはサイエンスの観点から「仮定が科学的に正当化できること」を重視し，EMA ガイドラインはレギュラトリーの観点から「試験治療に有利なバイアスが入らないこと」を重視しているように見える．

2.3.2 その他の相違点

その他の相違点として，NAS レポートと EMA ガイドラインのどちらか一方でしか触れられていない要素がいくつかある．例えば，NAS レポートでは欠測データの割合のターゲットをプロトコールに設定することを提言しており（Recommendation 8），それによって医師を動機付けたり DMC（Data Monitoring Committee）がターゲットと実際の欠測の割合との差をモニタリングしたりすることが可能になると提案している．

Recommendation 8: プロトコールでは欠測データの量を最小化することの重要性を認識させるべきである．特に，過去の試験結果に基づき，主要な結果について完全なデータが少なくともどのくらいあればよいのか設定しておくべきである．

一方，EMA ガイドラインは，前述したように試験の計画時に欠測データの量を予測することは重要としつつも，プロトコールにおけるターゲットの設定については言及していな

い。逆に、EMA ガイドラインでは、治療群間で中止パターンが異なる場合に総括報告書において中止パターンを図示（例えば Kaplan-Meier プロット）するべきと述べているが、NAS レポートでは特にそのようなことには触れられていない。

2.4 日本の状況

日本においては、EMA ガイドラインのように欠測データの取り扱いに関する規制当局の考え方を示したガイドラインは存在しない。一方、第 34 回日本臨床薬理学会（2013 年 12 月）における「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」という演目の中で当時 PMDA に所属していた飛田英祐氏（当時の所属：新薬審査第三部（中枢神経領域））は、個人の経験に基づく個人的な見解として、以下の点に留意していると述べている。

<試験計画段階>

- 欠測に関する事前の情報の確認
 - － 疾患・薬剤の特徴等を踏まえた、欠測発現状況の想定をしているか
- 欠測を減らすための対策
 - － 試験デザイン（評価項目の数が多すぎないか、評価時期・間隔、初期用量は適切か、効果不十分の定義はあるか）
 - － 実施上の配慮（医師への説明、患者への同意説明は適切に実施予定か）
 - － 中止理由の詳細な収集（中止時期、主な理由以外の中止理由の聴取を予定しているか）
- 解析方法の選択と事前明記
 - － 主解析は、明らかに過大評価となる手法を避ける（保守的な場合には LOCF でも評価可能）
 - － 感度分析による検討を事前に計画しておく

<試験結果評価段階>

- 欠測状況の把握
 - － 群間での相違はあるか（中止割合、中止時期、理由の有無）
- 欠測データを含む解析の結果
 - － 事前に計画した解析か否か
 - － 結果の頑健性は確保されているか
 - － 頑健性に疑義がある場合、原因の追求はなされているか

また、2014 年 2 月 14 日に PMDA、統計数理研究所リスク解析戦略研究センター、日本製薬工業協会 医薬品評価委員会の共催により実施されたデータサイエンスラウンドテーブル会議では、以下のような議論がなされた。

- 欧米では NAS レポートやガイドラインが示されており、日本でも考え方の方向性を揃えるため、ガイドライン等の作成が望まれる。

- ・ 試験計画段階では、臨床的妥当かつ欠測データの少なくなる主要評価変数や試験デザイン、欠測データの原因、頻度の結果への影響の検討が必要である。これら検討には時間を要するが、試験を成功させるためにこの時間の確保が必要である。
- ・ 欠測データの解析では、これまで汎用的に使用された“Full Analysis Set (FAS)+ LOCF”を無条件に利用することはできず、状況に応じた解析方法の検討が必要である。また、感度分析の実施も重要である。

【参考文献】

- ・ O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clinical Pharmacology & Therapeutics*. 2012; 91 (3):550-554.
- ・ Mallinckrodt C, Roger J, Chuang-Stein C, Molenberghs G, O'Kelly M, Ratitch B, Janssens M, Bunouf P. Recent Developments in the Prevention and Treatment of Missing Data. *Therapeutic Innovation & Regulatory Science*. 2014; 48:68-80.
- ・ 飛田英祐, 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. 第 34 回日本臨床薬理学会要旨集. 2013.
- ・ データサイエンスラウンドテーブル会議 Missing data セッションのまとめ (2014 年 2 月 14 日開催).
http://www.pmda.go.jp/operations/shonin/info/report/2013datascience/file/RT1_Missing_Review.pdf

3. Estimand と治療中止後のデータ収集

3.1 Estimand とは

NAS レポートでは、estimand の推定が臨床試験における主要なゴールであり、欠測データによる混乱を避け、潜在的なバイアスを明確に評価するためには、estimand の設定、特に効果指標 (outcome measure) と対象集団を事前に明確化することが重要であると述べている。この estimand は、臨床試験で“何を推測したいのか”を整理した概念であり、NAS レポートでは 5 つの例をあげ estimand を説明している。また、これら 5 つの例は Little ら (2012)にも引用されている。

1) ランダム化されたすべての被験者でのアウトカムの改善 (の差)

割付けられた全ての被験者において、治療方針又は戦略の全体的な有用性を確認する、いわゆる'intention-to-treat'の estimand である。この estimand は患者集団における治療方針に係るものなので、観察された差は最初に割付けられた治療効果を、効果不足によって採用された代替治療も含めた上で反映することとなる。アウトカムは、試験治療を受けたかどうかに係らず、全ての被験者から収集することになる。この estimand を選択する場合には試験治療中止後も被験者からデータを収集する必要がある。

2) 初期の治療に耐えられた²被験者のアウトカムの改善 (の差)

初期の治療に耐えられた被験者において、アウトカム改善の程度を定量化する estimand である。この estimand は、ランダム化前に実薬投与期間を設定し、治療に耐えられた被験者のみに対してランダム化を行い、治療効果をみるものである。初期の治療に耐えられる部分集団を事前に設定することが困難であるため、確認された効果が実際よりも過大評価される可能性がある。

3) すべての被験者が治療を遵守したとみなした場合のアウトカムの改善 (の差)

全ての被験者がプロトコールで計画された研究期間にわたって試験治療を受けたという仮定の下で、アウトカム改善の程度を定量化する estimand である。この estimand は、プロトコールを遵守しなかった被験者に対して、仮に遵守していた場合にアウトカムがどうなるのか、を想定して治療効果を考える。この estimand の有用性は、避けられうる不遵守の程度に依存する。不遵守を回避できないことが多い状況では、この estimand は実行不可能な治療方針の効果を測定することとなる。アウトカムは全ての

² NAS レポートの原文では、「(Difference in) Outcome Improvement in Tolerators. This estimand quantifies the degree of outcome improvement in subjects who tolerated and adhered to a particular treatment. This estimand concerns the subject of the population who initially began treatment and tolerated the treatment.」とあり、初期の治療を遵守したと表現することもできる。

被験者から収集する。

4) 治療を遵守した期間でのアウトカムの曲面下面積 (の差)

この estimand はプロトコール遵守期間におけるアウトカムの曲線下面積の平均値を比較するものであり、全被験者におけるアウトカムと遵守期間の両方に対する治療効果を同時に定量化する。この estimand を利用する試験デザインでは、試験治療中止後または代替治療への変更後のアウトカムを収集する必要が無い。

5) 治療を遵守した期間でのアウトカム (の差)

この estimand は、試験開始時から試験終了時、又はプロトコールの遵守最終日までの平均アウトカムの差であり、全ての患者における試験治療の遵守期間と、アウトカム改善の両方を反映する。

これら 5 つの estimand のうち、1), 4), 5) は、effectiveness (有用性)、2), 3) は efficacy (有効性) に着目しており、それぞれの観点で解釈すべきである。特に 1), 4), 5) は有効性のみならず安全性、プロトコール遵守にも影響されることを留意すべきである。

従って estimand の選択は、臨床試験デザインを検討する上で極めて重大な課題であり、規制当局の意思決定にも影響を与えることに留意しなければならない。また、試験の条件により様々な estimand が考えられ、解析対象集団、試験デザイン及び欠測データの影響を受けることにも留意する必要がある。これら内容を踏まえ、NAS レポートでは 'Recommendation 1' として、以下の内容が示されている。

Recommendation 1: プロトコールには以下の項目 (a) 試験の目的、(b) 主要評価項目、(c) どのように、いつ、誰の評価項目が測定されるのか、(d) 介入効果の測定、すなわち、主要な目的の causal estimands を明確に定義すべきである。これらの測定はすべての試験の参加者にとって意味のあるもので、また最低限の仮定で推定可能であるべきである。後者に関して、プロトコールには欠測データの潜在的な影響とその扱いについて記載すべきである。

一方で、NAS で示される estimand の例は理解が困難であるため、estimand の理解には Mallinckrodt (2013) や Mallinckrodt ら (2014) が分かりやすい。Mallinckrodt ら (2014) は、NAS レポートで示された estimand について更に詳しく解説している。彼らは estimand の概念を“推定したいもの”としたうえで、経時測定の試験において estimand を構成するものを以下のように提示している。

- ・ パラメータ (例、治療群間の平均的变化の差)
- ・ 曝露期間 (例、8 週間)
- ・ 効果指標 (例、拡張期血圧)
- ・ 対象集団 (例、高血圧と診断された患者)

- ・ 割付けた試験治療中止後のデータの採用，不採用
- ・ 後治療の導入の採用，不採用

また，適切な estimand を議論する上では，efficacy (有効性)と effectiveness (有用性)のどちらに着目するかが重要であると述べている．ここで，efficacy (有効性)とは，指示通りに薬剤を使用した場合での効果であり，per-protocol estimand とも呼ばれる．一方，effectiveness (有用性)とは，実際に使用した場合の薬剤の効果であり，intention-to-treat (ITT)estimand とも呼ばれる．更に，estimand の選択が試験治療中止後のデータの解析での利用を左右するとも述べている．Mallinckrodt ら (2014)は estimand の例を表 3-1 のようにまとめている．

表 3-1. Estimand の例 (Mallinckrodt ら 2014 改変)

Estimand	推測	推論	対象集団	エンドポイント	試験治療中止後のデータの使用	NAS レポートでの estimand 例との対応*
A	有用性	治療方針	全被験者	計画されたエンドポイント	主要解析に含める (試験治療中止後データを利用)	1) ランダム化されたすべての被験者でのアウトカムの改善 (の差)
B	有効性	割付けた治療	全被験者	計画されたエンドポイント	主要解析に含めない	3) すべての被験者が治療を遵守したとみなした場合のアウトカムの改善 (の差)
C	有用性	割付けた治療	全被験者	計画されたエンドポイント	補完することが望ましい	—

* 本論文では，NAS レポートの estimand 2), 4), 5) に対応する estimand は紹介されていなかった．

各 estimand には明確な違いがある．Estimand A は臨床現場で利用された場合の因果効果を推定の対象としている．つまり，割付けた治療が失敗した場合に行われる代替治療の効果も推定の対象に含められる．言い換えると“最初に”割付けた治療を行った場合に，計画された時点で被験者がどのような状態になったかを推測しているということであり，割り付けた治療の効果を直接は推測していない．Estimand B は割り付けた治療を遵守した場合の効果を推測しており，割り付けた治療の効果を直接推測している．被験者の大多数がプロトコルを遵守している場合や，プロトコル不遵守を回避できる場合（副作用を防止する手段が利用可能な場合や，用量を制限することによって不遵守を回避できるような場合）には特に有用である．

Estimand C は，Estimand A と同様に有用性を重視しているが，試験治療中止後の代替

療法に制限を設けることや、試験治療中止後の観測値を試験治療中止前の観測値などから補完するなどして、計画された時点で割り付けた治療法がどのような因果効果を及ぼしたかを推測の対象としている。注意点として、試験治療中止後の代替治療の選択に制限を加えることの妥当性は倫理的な観点から検討される必要がある。図 6-1 に analytic roadmap(Mallinckrodt (2013) 図 12 を改変)にを示すように、Estimand B と Estimand C の両方を用いて推測する方法も提案されている。

なお、臨床試験は臨床現場に比べ実験的であり、限られた条件での推測となる点には留意したい。特に医薬品開発段階で、大規模臨床試験でない場合には、有用性よりも有効性評価に重点をおき、Estimand B のように割り付けた治療の効果を推測する estimand を設定して試験を実施し、有用性評価は臨床試験以外の情報も含めて考察する方策が考えられる。

3.2 試験治療中止後データの収集

NAS レポートでは、試験治療を中止した被験者から中止後のデータを収集し、解析に利用することを recommendation 3 及び 5 で推奨している。

Recommendation 3: スポンサーは試験の途中でプロトコル規定の治療を中止した被験者に対する重要な結果情報の集積を継続して行うべきである。この情報は記録した上で解析に使われるべきである。

Recommendation 5: プロトコルに規定された介入を受けたか受けなかったかに係らず、試験に組み込まれた全ての被験者に対してデータ収集、全ての関連する治療情報、及び重要な共変量を記録するべきである。

さらに NAS レポートでは、プロトコルに規定した治療を中止した場合に、代替治療を受けられるようにプロトコルで定めておくことを推奨している。

Recommendation 4: 試験デザイン検討チームはプロトコルで規定された治療を中止した被験者が、特定の代替治療を行う、または促すことを検討するべきである。代替治療はプロトコルに規定されるべきである。

また、EMA ガイドラインでも以下のような観点から試験治療中止後データの収集を推奨している。

- ・ 欠測データの数を最小限にするために、プロトコル遵守に関わらずデータを収集することは非常に重要である。
- ・ 試験治療中止後の継続したデータ（特定の臨床アウトカム）収集を強く勧める。
- ・ 試験治療中止後、更なる治療介入なしに被験者が回復した情報は、FAS の集団を最大限に構成し、通常、主解析のための十分な根拠となるだろう。試験期間全体における治療効果を評価する場合、この分析は特に重要である。

- ・ 試験治療中止後に代替治療を受けた被験者のデータも同様に収集すべきである。特に、代替治療の開始時期、期間、用量を含む全詳細を収集すべきである。
- ・ 被験者が試験を中止する場合、その中止理由を可能な限りすべて報告すべきである。これは、中止に至る極めて重要な原因の同定やこれらの被験者を欠測データ解析でどのように扱うかに関わってくるかもしれない。

さらに、White ら (2011)や White ら (2012)は、試験治療中止後の被験者の追跡を行わない試験では ITT の原則を主張することはできないとし、試験治療中止後の被験者の追跡を行わない場合の 4 つの不利益をあげている。

- 1) ITT の原則の主旨に反する
- 2) 試験治療中止後のデータを含めずに、試験治療中のデータのみで推定する治療効果は、ITT の下での治療効果と異なる
- 3) 試験治療中止後の追跡をしない場合、MAR の尤もらしさが損なわれる
- 4) たとえバイアスを誘導しなくても、治療効果が長い場合には検出力が低下する

Wittes (2009)は試験治療中止後のデータの取得を促進するための具体的な方法を示しており、同意・説明文書やプロトコールにおいて「試験治療の中止」と「試験の中止」を明確に区別することを推奨し、文章の事例を提示している。

同意・説明文書での記載例

この治験に参加するかどうかは、治験の内容をご理解いただいた上で、あなたの自由な意思によりお決めください。治験への参加を断ったとしても、あなたが不利益を受けることは決してありません。・・・本試験では、試験治療が中止された場合でも試験治療中止後の情報を調査するために来院していただくことをお願いいたします。その際、担当医師は試験治療中止後の情報を得るために問診を行います。もしこの試験への参加に同意いただいた後でも、あなたはいつでも、あなたに保証されているあらゆる利益への影響を受けることなく、試験への参加を中止することができます。

プロトコールでの記載例

試験治療の中止は、被験者が試験治療を永久に中止した際に発生する。試験の中止は、試験プロトコールの完了に先立ち、あらゆる状況によらず、被験者が死亡、追跡不能、または同意を撤回した場合に発生する。・・・以下の理由に該当する被験者については本試験を中止しなければならない。

- ・ 被験者自身又は、法的に認められた被験者の代表による試験中止の要求
- ・ 医師の見解において、試験の継続が被験者の福祉に対して有害とみなされる
- ・ 医師又は被験者によって判断された、容認できない有害事象の発生
- ・ 試験スケジュールに従って試験実施施設に来院しない
- ・ 持続的な服薬不履行

一方, Mallinckrodt ら (2014)は, 前述したように, 試験治療中止後のデータを解析へ含めるかどうかは *estimand* に依存し, 一つの試験の中で試験治療中止後のデータを解析に含める場合と含めない場合があるとも述べている. 特に臨床試験という限られた環境の中 (通常は二重盲検ランダム化試験)で, *estimand A* のような治療方針の検討が可能であるか否かについては慎重な議論が必要である. 例えば, サロゲートエンドポイントを利用した試験や試験規模が小さい試験の場合には, 治療方針の検討を第一の目的にするのではなく, 最初に割り付けた治療 (つまり試験治療)の効果을推定し, その一般化は外部の情報と合わせて検討したりするという方法もあり得る.

O'Neill and Temple (2012)は, この NAS レポートの提言 (試験治療終了後のデータ取得)を “意見が分かれるところかもしれない” と述べ, アウトカムスタディとそれ以外の試験で受け入れ可能性が異なると指摘している.

3.3 まとめ

欠測データによる諸問題を考慮する上で, 試験の計画段階において, *estimand* (臨床試験で “何を推測したいのか”)を明確に定めることは極めて重要である. 適切な *estimand* を設定するためには, 医薬品開発の各段階で何を示そうとしているのかを有効性と有用性という観点から整理しておくべきである. また, 選択される *estimand* によっては試験治療中止後のデータが必要となり, NAS レポートと EMA ガイドラインのいずれもそれを推奨している. 一方で, 市販される前の臨床試験という限られた条件での推測には限界があるため, 試験治療中止後のデータ収集については, *estimand* の選択と合わせて慎重に検討することが必要である. 試験治療中止後データの収集が必要な場合には, 「試験治療中止」と「試験中止」の違いをプロトコルで明確に定義し, さらに同意・説明文章で試験治療中止後データの収集の必要性について十分に説明することで, 試験実施責任者と被験者の理解を得られるように努める必要がある.

【参考文献】

- Little R, Cohen M, Dickersin K, Emerson S, Farrar J, Neaton J, Shih W, Siegel J, Stern H. The design and conduct of clinical trials to limit missing data. *Statistics in Medicine*. 2012; 31: 3433-3443.
- Mallinckrodt C. *Preventing and Treating Missing Data in Longitudinal Clinical Trials*. New York, NY: Cambridge University Press; 2013.
- Mallinckrodt C, Roger J, Chuang-Stein C, Molenberghs G, O'Kelly M, Ratitch B, Janssens M, Bunouf P. Recent Developments in the Prevention and Treatment of Missing Data. *Therapeutic Innovation & Regulatory Science*. 2014; 48:68-80.
- White I, Horton N, Carpenter J, Pocock SJ. Strategy for intention to treat

- analysis in randomised trials with missing outcome data. *BMJ*. 2011; 342: d40.
- White I, Carpenter J, Horton N. Including all individual is not enough: Lessons for intention-to-treat analysis. *Clinical Trials*. 2012; 9: 396-407.
 - Wittes J. Missing inaction: Preventing missing outcome data in randomized clinical trials. *Journal of Biopharmaceutical Statistics*. 2009; 19: 957-968.

4. 欠測メカニズム

欠測メカニズムという用語は、なぜデータが欠測になったのかという理由と、その理由とアウトカムとの関係を表している。ここでは、後での解析方法や感度分析を理解するために有用な概念として欠測メカニズムを整理する。Rubin (1976) 以降、欠測を取り扱う統計的な文献において、欠測メカニズムは、完全に無視可能なものと、工夫によって偏りの無い推測ができるもの (統計モデルが正しいという前提がないと正しい推測は得ることができない) と、偏りの無い推測を行うことができないものに分類され、これらは、それぞれ MCAR, MAR, MNAR と呼ばれている。

MCAR の典型的な例は、健康上以外の理由による転居である。この場合には欠測データの存在を無視しても偏りの無い推測が可能であり、この理由によって多くの欠測が生じるとは考えにくい。精度の低下も無視可能である。しかし、臨床試験における欠測データのすべてが MCAR によって生じているということは考えにくい。

MAR の仮定は多くの状況において受け入れられる可能性が高い。この仮定は、中止後の被験者の状態が中止前に観測されたデータから予測可能であることを示唆している。例えば、ある被験者が観測された一連の有効性データが悪い値であったことを受けて効果不十分を理由に中止した場合、この被験者に対して中止後の有効性データも悪い値を取ったであろうという仮定の下で解析することは受け入れられるだろう。

欠測メカニズムが MCAR でも MAR でもない場合、MNAR に分類される。この場合、観測されなかった値が解析に利用できない情報に依存することを意味するため、中止後の被験者の状態をバイアスなしに予測することは出来ない。例えば、観測された一連の有効性データが良い値であったにもかかわらず、その後、効果不十分を理由に中止した場合、観測されたデータのみに基づく統計モデルは中止後も良い値を予測するかもしれないが、そのような被験者がその治療からベネフィットを受け続けると仮定するのは、通常は妥当ではない。

以下に、欠測メカニズムの分類の概念図を示す。

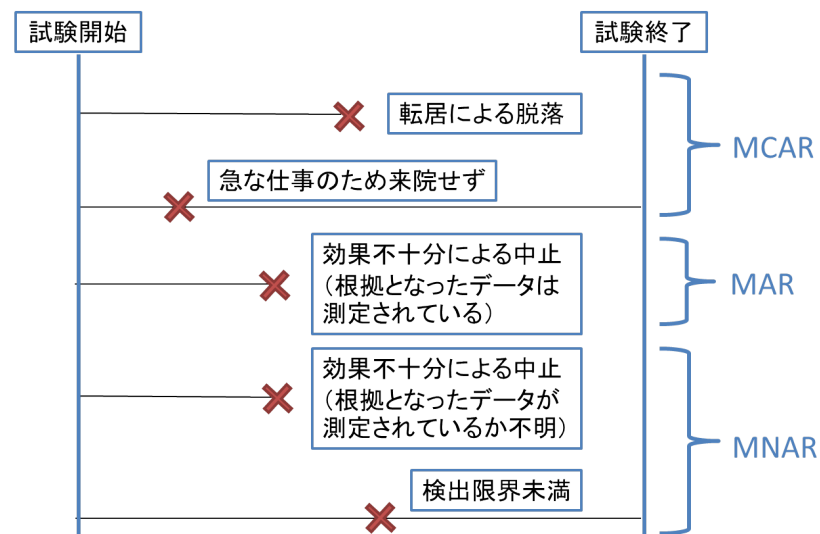


図 4-1. 欠測メカニズムの概念図

図中の X は欠測データが発生した時点を示す

一般的に MAR の仮定は MCAR よりも現実的であり、比較的簡便な解析方法が存在することからも実際の臨床試験において適用されることが多い欠測メカニズムであるが、あくまで仮定であり、実際のデータから検証することは不可能である。そのため、解析手法を選択する際には MAR や MCAR の仮定の下での性質のみに依存するのではなく、規制上の判断を下すために受け入れ可能な推定ができると考えられるかどうかを考慮すべきである。

【参考文献】

- Rubin, D.B. Inference and missing data. *Biometrika*, 1976; 63: 581-592.

5. 解析方法

5.1 はじめに

欠測メカニズムや欠測データを含む不完全データの解析は、以前から研究レベルでは議論されてきたが、臨床試験の実際の解析には、LOCF が汎用的に用いられていた (Tanaka ら, 2014).

不完全データの解析方法の分類方法として、不完全データから完全データを生成して解析を行う方法と観測データと欠測メカニズムをモデルで結合させることで解析的に計算する方法に大別される. 前者の完全データを生成する方法には、後述の Single Imputation, Multiple Imputation などがあり、不完全データの欠測部分をなんらかの値で補完するためバイアスの制御には強い制約が必要となる. 一方、後者のモデルによる解析は、欠測データを伴うデータに対して偏りが無い推測を行うために、欠測データの分布と欠測割合が必要となる. このため、欠測割合の推測の観点から、欠測データを MCAR, MAR と MNAR と分類することがある. また、欠測の原因別に分布が異なるとする考え方もあり、この分布を測定データから推測できる場合とできない場合がある. これらの組み合わせによって欠測データの状況を説明することができ、推定の目的別に選択できる解析方法が決まる. これらは研究が盛んな領域であるため、多くの手法があり、また欠測メカニズムに関する整理も様々な構成があるが、本稿では、NAS レポート及び EMA ガイドラインの記載に準じ、代表的な解析方法を紹介する.

5.2 Complete case analysis

Complete case analysis は、予定されたすべての測定がなされた対象者のみを解析対象とする方法である. この方法が妥当であるためには、欠測メカニズムが MCAR であることが必要である. すなわち、ここでは観測データに対する考慮のみで解析方法を規定することになる. 多くの臨床・疫学研究で、欠測メカニズムに MCAR を仮定できる状況は稀であり、たとえ MCAR を仮定することが合理的であったとしても、解析に寄与する対象者数が減るため、検出力の観点からも好ましい解析方法ではない (松山, 2004; 丹後, 上坂, 2006). EMA ガイドラインにおいても、検証的試験の主要な解析としては勧められないと述べられている.

5.3 Imputation-Based Approaches

欠測データを何らかの値で補完する方法であり、Single Imputation と Multiple Imputation (MI) がある. これらの代入法は、観測データに関する取扱いに規定されるものである. Single Imputation は欠測データを単一の値で置き換える方法であり、Multiple Imputation は複数回置き換えて、値を推定する方法である. 例えば計量値の評価変数に対しては、

Imputation method により値を補完して完全データを作成し、ANCOVA (analysis of covariance) のような解析を実施する。それぞれの例を以下で紹介する。

1) Single imputation Method

Single imputation とは欠測値を一つの値で補完する方法である。代表的な方法に LOCF 法と BOCF 法(baseline observation carried forward)がある。LOCF 法は、欠測データを最後に測定された値で補完する方法であり、BOCF 法は、欠測データをベースライン値で補完する方法である。これらの方法では欠測値が存在しないデータを作成した後で、解析を実施するため比較的理解しやすい。しかし、一般的な臨床試験で設定するような特定時点での効果を検討する場合に³、LOCF 法が妥当であるためには、「脱落後の結果変数の推移は最後に観察された値のまま変化しない」という非常に強い仮定を必要とする (松山, 2004; 丹後, 上坂, 2006)⁴。また BOCF 法が妥当であるためには、中止後の結果変数は治療前の状態に戻るという仮定が必要になる。

図 5-1 はアルツハイマー型認知症治療剤の有効性 (QOL Score)を評価するプラセボ対照臨床試験であり、LOCF を用いることによって生じるバイアスの概念図を示す (O'Neill RT, Temple R, 2012)。本疾患では、臨床的経過が継続的に増悪傾向を示すため、10 週での効果を検討したい場合に、途中中止時の 6 週時の値を 10 週時の値として補完すると、真値と仮定値には図で示す LOCF bias の大きさのずれが発生する。

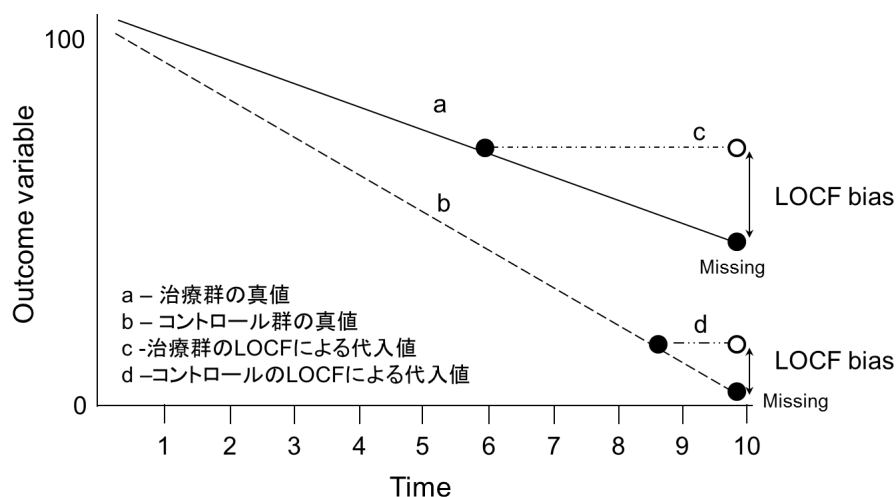


図 5-1. LOCF バイアスの概念図

³ 特定時点とは、ある 1 評価時点であり、試験で計画された服薬期間の最終時期とすることが多い (例えば、服薬期間が 12 週間や 3 か月間の場合の 12 週時や 3 カ月後時点)。

⁴ 特定時点での効果を検討するのではなく、各被験者の最終時点の測定値に臨床的な意義がある場合には最終時点を用いた解析を行えば良い。このことを最終評価時点における LOCF 法と表現することがあるが、実際には carried forward している訳ではなく、規定した特定時点の評価で carried forward をしている取り扱いとは異なる。

2) Multiple imputation Method (MI)

MI は、欠測データに複数回補完を行い複数個の完全データを作成し、作成した複数個の完全データでそれぞれ解析する。そして、得られた複数個の解析結果を併合して一つの結果を得る。補完の方法と併合の方法にはいくつかの方法があるので状況に応じて使い分ける必要がある。MI は、欠測値に対して複数個の補完を行うことでばらつきも考慮に入れているところが single imputation と異なる点である。なお、MI が Single imputation より優れた点として、MAR の欠測メカニズムのもとで妥当な推測ができることが挙げられる。また、欠測データの補完方法を工夫することによって、MNAR を仮定した解析でも利用可能な方法が提案されている。

5.4 Inverse Probability Weighting (IPW)

IPW は観測時点より前のデータや共変量から、観察されるデータ (例えば、完了被験者の最終時点) が観測される確率を推定し、観測されたデータの重みを調整して解析する。直感的な例では、測定される確率が 0.5 (= 50%) であれば、その観測値の重みは 2 となる。

IPW は、完了例のみが解析対象であり、すべての情報が有効に活用されておらず、観測確率 0 に近い場合には該当する被験者の重みが他と比べて極端に大きくなるという特徴がある。また、観測確率を正しく推定することは一般に難しく、モデルが誤っていた場合には、推定値にバイアスが入る。この弱点を緩和する方法として Augmented IPW (AIPW) 法がある。AIPW による推定は観測確率を算出するモデル、もしくは回帰モデルのどちらかが正しい場合には妥当な推定値を与えるという特徴があり、また対象被験者の重みが大きい場合には推定値の算出に制約をかけることも可能である。

IPW 法及び AIPW 法で妥当な結論を得るためには、MAR の仮定が必要である。

5.5 Mixed Model for Repeated Measure (MMRM)

MMRM は経時測定データの場合に、欠測値を補完することなく、応答に対して特定の確率分布と回帰モデルを仮定することによって解析を行う方法である。LOCF に変わる解析方法として汎用されているという報告がある (LaVange 2014)。MMRM で妥当な結論を得るためには、MAR の仮定が必要である。

5.6 Selection Model (SM)

SM は、経時測定データの解析で用いられる。MAR にも MNAR にも対処可能な解析方法であり後述の PMM と同様に感度分析として用いられている例がよく見られる。前述の MMRM は MAR を仮定した SM であると言える。また、MNAR を仮定した場合には、欠測確率を適切にモデル化することで、妥当な推測を行うことができる。ただし、モデルの適

切性は、データからは検証できないので、仮定の妥当性には特に注意が必要である。

5.7 Pattern Mixture Model (PMM)

PMM は、被験者の特徴により複数の分布（パターン）が混在するということを仮定する方法であり、SM と並び、MNAR にも対応可能な方法である。パターンは、例えば、中止時期によるパターン（試験早期に中止した被験者グループ、試験後半で中止した被験者グループ、試験中止せず試験完了した被験者グループ）や、中止理由に基づくパターンなどを設定し、それらのパターンごとに観測データを分類し、推測する。このパターンを特定するための情報の収集が必要となる。

【参考文献】

- Tanaka S, Fukinbara S, Tsuchiya S, Suganami H, Ito MY, Current Practice in Japan for the Prevention and Treatment of Missing Data in Confirmatory Clinical Trials. *Therapeutic Innovation & Regulatory Science*. Published online before print April 16, 2014
- Verbeke G, Molenberghs G. *Linear Mixed Models in Practice - A SAS - Oriented Approach*. New York, NY: Springer-Verlag; 1997. (編訳 松山裕, 山口拓洋. 医学統計のための線型混合モデル -SAS によるアプローチ-. サイエンティスト社; 2001.)
- 松山裕. 経時観察研究における欠測データの解析. *計量生物学*. 2004; 25 (2):89-116.
- 丹後俊郎, 上坂浩之. 臨床試験ハンドブック. 朝倉書店; 2006.
- O'Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clinical pharmacology and therapeutics*. 2012; 91 (3):550-554.
- LaVange LM. The Role of Statistics in Regulatory Decision Making. *Therapeutic Innovation & Regulatory Science*. 2014; 48 (1): 10-19.

6. 主要な解析方法の選択と感度分析

6.1 はじめに

臨床試験では、主要評価変数に関して主要な解析方法を一つ決定する。ただし、ICH E9 ガイドラインにもあるように、欠測データを伴う不完全データに対して普遍的に適用可能な解析方法はない。特に欠測データが多い場合には、欠測データに対処する方法により解析結果がどの程度変わるかを検討する必要がある。つまり、感度分析という形で異なる欠測メカニズムの仮定や解析方法に基づき推測を行い、解析結果の安定性を検討することが求められる。

以下、主要な解析方法の選択と感度分析の方針について概説する。なお、NAS レポートにもあるように、感度分析では、いくつかの異なるタイプのモデルに対する感度の評価が対象になるが、本報告書では欠測メカニズムに関する感度分析に焦点をあて論述する。また、本章で扱う話題は現在も活発な研究が行われている分野であり、明確なコンセンサスが得られていない点があることに留意頂きたい。

6.2 主要な解析方法の選択と感度分析の必要性

主要な解析方法の決定は、試験の目的や estimand を決めることと密接に関係する。3 章に示したように、estimand は試験の開発段階や目的、疾患領域などを考慮して設定する。そして、具体的な解析方法は、主要評価変数の特性（例えば、計量データ、分類データ、順序データなど）、欠測データの量、そして欠測メカニズムなどを考慮し選択することになる。例えば、3 章に示した estimand B 「全ての被験者が治療を遵守したとみなした場合のアウトカムの改善（の差）」を選択し、主要評価変数の欠測メカニズムに MAR が仮定できる場合には、MMRM や Multiple Imputation といった解析方法を選択することができる。多くの臨床試験の状況で、欠測メカニズムに MAR を仮定した解析を主要な解析方法に設定することが合理的であるという報告がある (Mallinckrodt, 2013)。一方で、疾患領域や過去試験から欠測メカニズムに MNAR を仮定した方が合理的な場合には、MNAR を仮定した解析方法を主要な解析方法に選択することもある。いずれにせよ、試験計画段階での十分に検討し、主要な解析方法をプロトコールに規定する必要がある。

5 章で示したいずれの解析方法も、データからの検証が不可能な仮定を必要とする。特に欠測データが多い場合には、主要な解析で仮定した欠測メカニズムが間違っている際にも、主要な解析方法の解釈（すなわち試験の結論）に大きな影響を与えないことを確認する必要がある。このための方法が感度分析であり、欠測メカニズムに関するいくつかのシナリオの下で結果の感度分析を行うことが重要になる。

6.3 感度分析の方針

欠測メカニズムに関する感度分析の方針，すなわち，どのような状況で，どの程度，どのような解析を実施すべきかについて，規制当局を含めて完全にコンセンサスが得られたルールはないことから，NAS レポートや EMA ガイドラインなどを参考に，個別に検討されているのが現状である．

NAS レポートでは，欠測メカニズムが MAR の下での妥当な解析方法を主要な解析方法として選択した場合には，Selection Model (SM) や Pattern Mixture Model (PMM) など欠測メカニズムに対する仮定を緩めた MNAR の仮定の下での様々な解析を行い，主要な解析結果からの乖離度合いを評価することについて記載している．そして，不完全データに対する推論を行う際に二つの仮定として (i) 検証不可能な仮定と (ii) 検証可能な仮定を考える必要があると述べている．ここで，検証不可能な仮定は欠測データの分布に基づく仮定，検証可能な仮定は観察データの分布に基づく仮定である．欠測メカニズムとして MAR が常に適切であると主張することは不可能であるため，MNAR の状況を含めた様々な条件の (i) 検証不可能な仮定の下で感度分析を行うことが必須であるとしており，(i) 検証不可能な仮定と (ii) 検証可能な仮定を結びつける手法である SM と PMM を用いた例を紹介している．

以下に NAS レポートの Recommendation 15 を再掲する．

Recommendation 15: 感度分析は臨床試験の主要な報告の一部であるべきである．特に欠測のメカニズムの仮定に対する感度分析は必須である．

EMA ガイドラインでも，欠測データが無視できないほど発生する場合には感度分析は主要解析のサポートとして提示されるべきであるとされており，また，欠測メカニズムの仮定に限らずより包括的なアプローチにより主要な解析結果に試験治療に有利なバイアスが働いていないかを確認することが記載されている．ただし，欠測メカニズムに関する感度分析に関しては，NAS レポートと EMA ガイドラインで大きな差は認められない．つまり，欠測メカニズムの仮定を変えた下で様々な感度分析が必要であるという共通のメッセージがある．

Mallinckrodt (2013) の 12 章では，エンドポイントが連続値の経時測定データの臨床試験で，主要な解析方法と感度分析を検討するためのフレームワークを Analytic road map という図を用いて示している (図 6-1)．詳細は Mallinckrodt (2013) の 12 章及び 14 章の事例を参照されたい．

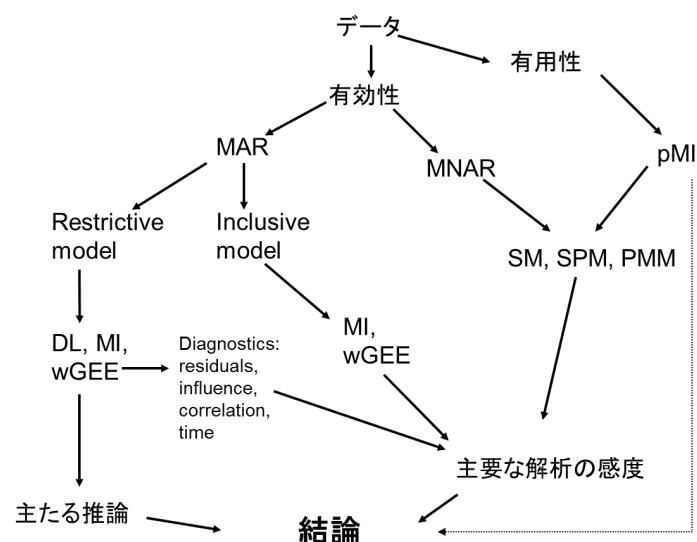


図 6-1. Analytic Road Map [Mallinckrodt (2013) 12 章 図 12.1 改変]

DL: Direct likelihood; MI: Multiple imputation; wGEE: Weighted generalized estimating equations; SM: Selection model; SPM: Shared-parameter model; PMM: Pattern-mixture model; pMI: Placebo multiple imputation; MAR: Missing at random; MNAR: Missing not at random

感度分析の方針は、試験計画段階に十分に検討することが重要であり、場合によってはその方針について規制当局と相談が必要になることもある。前述したように、欠測データが多く感度分析が重要な位置づけとなる場合には、感度分析の詳細をプロトコールに記載することもある。また、ICH E9 ガイドラインに「盲検下レビューの間に統計解析計画書を見直すことにより、欠測データを扱う方法が洗練されるであろう」とあるように、試験途中に適切にデータを検討することで、感度分析の方針を適宜見直すことも有用である。

欠測メカニズムに関する感度分析の目的は、主要な解析方法の欠測メカニズムの仮定への安定性を検討することである。例えば、欠測メカニズムに MAR を仮定した主要な解析方法を選択した場合には、6.6 節に示すアセナピンの事例のように異なる欠測メカニズムの下で複数の解析を実施することにより、MAR の仮定からの逸脱への結果の安定性を評価できる。一方、中止割合が非常に低く欠測データがほとんど認められない試験の場合は、欠測メカニズムに関する感度分析は不要であるかもしれない。このように、感度分析の方針は、試験ごとに個別に検討することになる。

6.4 感度分析の解釈時の留意点

感度分析の結果を解釈する際には、P 値に基づいて有意差だけを議論することは避けるべきである。まず、その感度分析が何を解析しているかについての特性を理解し、そして、P 値だけでなく得られた推定値を主要な解析結果と比較する。さらに、投与群ごとの欠測デ

ータの発生時期、割合や試験中止理由などの情報とあわせて総合的に考察する。その際には、次節に示すように表やグラフを効果的に用いることが適切な解釈に役立つ。

感度分析の結果が主要な解析結果と大きく異なり、試験結果の安定性に疑問が出る場合には、その原因を探索的に検討していくことが重要となる。例えば、試験の中止理由の詳細は欠測メカニズムの検討に役立つ。また、試験治療中止後のデータを収集する場合には、代替治療の情報が重要になる。よって、CRF (case report form)をデザインする際には、これらの情報が適切に収集できるような工夫が必要になる。

6.5 EMA ガイドラインでの欠測に関する情報の報告

EMA ガイドラインでは、欠測データに関して、総括報告書へ記載すべき情報として以下の内容を挙げている。なお、EMA ガイドラインは検証的試験を対象にした文書であるが、本内容は探索的試験でも参考になる。

表 6-1. 総括報告書に記載すべき情報 (EMA ガイドライン)

<u>総括報告書に記載すべき内容</u> ・ 欠測データの取り扱いに関する事前規定の詳細 ・ その計画からの変更及び変更が正当である理由 ・ 欠測データの数、タイミング、パターン、理由、予想される影響 - 欠測データの予想数からのズレに関する資料（事前に規定した解析や適切な感度分析の議論のため） - 欠測データのパターンに関する説明（解析におけるバイアスの方向を理解することに役立ち、欠測データの取り扱い方法の選択の評価として重要） ・ 脱落理由を含めた Kaplan-Meier 法等によるグラフィカルな要約（群間で脱落パターンが異なる場合） ・ 中止理由（中止に至る極めて重要な原因の同定やこれらの被験者を欠測データ解析でどのように扱うかに関わってくるかもしれない） ・ 感度分析の結果 - 試験治療に有利なバイアスが働いていないことを確認するために不可欠 - その信頼区間は推定された治療効果の不確実性を適切に反映するものであるべき ・ 各被験者の有効性データ（単一補完法を用いている場合は、補完された値が確認できるように） <u>総括報告書に記載が推奨される内容</u> ・ すべての関連因子における欠測データのアンバランスに関する調査 ・ 欠測データの有無による被験者背景の違いに関する情報 ・ 収集された脱落後のフォローアップ情報（脱落例の解析上の取り扱いを正当化することに役立つ）
--

6.6 結果のまとめ方の例

FDA 及び EMA の承認事例を参考として、主要な解析と感度分析の結果のまとめ方の一例を以下に紹介する。なお、これらの事例では NAS レポートや EMA ガイドラインで述べられている MNAR の下での感度分析について積極的に触れられていない。一つの理由は、これら事例が NAS レポートや EMA ガイドラインが公表された時期よりも以前の事例であることが考えられる。実際に結果をまとめる場合には、最新の事例を参考にして欲しい。

表 6-2 は、統合失調症及び双極性障害を対象として開発されたアセナピン（商品名：Sycrest）に対する EMA の審査報告書で、主要な解析と感度分析の結果が同一の表に示された例である。この表では、主要評価変数である 12 週時点での YMRS (Young Mania Rating Scale；ヤング躁病評価尺度)のベースラインからの変化量について、アセナピン群と対照薬

であるオランザピン群の差を示している。主要な解析と感度分析の結果を一つの表に要約しており、欠測メカニズムの仮定、解析方法、欠測データの取り扱い、主要な解析又は感度分析の位置づけ、主要評価変数の群間差の推定値とその95%信頼区間を表示している。

表 6-2. 主要な解析と感度分析の解析結果の表示例 1 (アセナピン EMA 審査報告書より)

欠測メカニズムの仮定	解析方法	欠測データの取り扱い	位置付け	群間差の推定値	95%信頼区間
MAR	MMRM	尤度に基づく	主要な解析	0.6	(-1.3, 2.5)
MAR	MI	多重補完	副次解析	1.0	(-0.5, 2.5)
MCAR ⁵	LOCF	単一補完	副次解析	3.1	(1.6, 4.7)
NA	OC	補完しない	副次解析	0.6	(-0.9, 2.2)

MCAR: Missing Completely at Random; MMRM: Mixed Model for Repeated Measures; LOCF: Last Observation Carried Forward; MI: Multiple Imputation; OC: Observed Case; NA: Not Applicable;

表 6-3. は、2 型糖尿病を対象として開発されたカナグリフロジン (商品名: Invokana) に対する FDA の統計審査官による審査報告書で、主要な解析と感度分析の結果が同一の帳票に示された例である。この表では、主要評価変数である 26 週時点での HbA1c のベースラインからの変化量について、プラセボ群に対するカナグリフロジン 100 mg 群及び 300 mg 群の差を示している。主要な解析である LOCF とともに、感度分析として MMRM と Per Protocol (PP) についても、各群での推定量を詳細に要約している。

表 6-3. 主要な解析と感度分析の解析結果の表示例 2 (カナグリフロジン審査報告書を改変)

Endpoint	Placebo		Canagliflozin 100 mg		Canagliflozin 300 mg	
HbA1c (%)	Nn		N		N	
Baseline mean ± SE	189	7.97 ± 0.07	191	8.06 ± 0.07	193	8.01 ± 0.07
Adj. Mean Change from baseline ± SE						
LOCF	189	0.14 ± 0.06	191	-0.77 ± 0.06	194	-1.03 ± 0.06
MMRM	182	0.05 ± 0.05	183	-0.78 ± 0.05	191	-1.03 ± 0.05
PP	121	-0.18 ± 0.07	165	-0.79 ± 0.06	171	-1.06 ± 0.06
Can-P, adjusted LS Mean (95% CI)						

⁵ LOCF の欠測メカニズムの仮定が MCAR とあるのは、EMA 審査報告書の原文通りの記載である。一方、NAS レポート (p.65) には、LOCF is sometimes mistakenly considered to be valid under MCAR or MAR, but in general it makes an MNAR assumption. すなわち、一般に LOCF は MNAR を仮定しているとの説明がある。

LOCF		-0.91 (-1.09, -0.73)	-1.16 (-1.34, -0.98)
MMRM		-0.83 (-0.97, -0.68)	-1.08 (-1.23, -0.94)
PP		-0.61 (-0.78, -0.44)	-0.88 (-1.05, -0.71)
Patient (%) achieving			
Hb1A1c < 7 ^{1,2}	28 (16%)	64 (38%)	91 (55%)
LOCF ²	32 (19%)	67 (39%)	95 (57%)
LOCF ³	39 (21%)	85 (45%)	121 (62%)

SE: Standard Error; LOCF: Last Observation Carried Forward; MMRM: Mixed Model for Repeated Measures; PP: Per Protocol; Cana: Canagliflozin; P: Placebo; LS Mean: Least-squares Mean

1 Based on patients at baseline with HbA1c>7%, placebo n=172, cana 100 mg n=170, and cana 300 mg n=166

2 Completers

3 All mITT patients, including those who had baseline HbA1c < 7%

また、カナグリフロジンの FDA 承認審査での諮問委員会の資料では、試験中止割合を、実施された一連の臨床試験データをまとめたうえで理由別に集計した表を示している (表 6-4)。この表は、欠測メカニズムを考察するうえで重要な情報になる。

表 6-4. 試験から脱落した被験者割合 (カナグリフロジン FDA 諮問委員会資料を改変)

	All Non- Cana (N=3262) n (%)	Cana 100 mg (N=3092) n (%)	Cana 300 mg (N=3085) n (%)	All Cana (N=6177) n (%)	Total (N=9439) n (%)
Subject Disposition Category					
Subjects who discontinued prior to the cutoff date	925 (28.4)	686 (22.2)	695 (22.5)	1381 (22.4)	2306 (24.4)
Primary reason for discontinuation					
Adverse event	156 (4.8)	170 (5.5)	217 (7.0)	387 (6.3)	543 (5.8)
Creatinine or eGFR withdrawal criteria	17 (0.5)	23 (0.7)	22 (0.7)	45 (0.7)	62 (0.7)
Death	30 (0.9)	21 (0.7)	17 (0.6)	38 (0.6)	68 (0.7)
Lack of efficacy on rescue therapy	33 (1.0)	8 (0.3)	6 (0.2)	14 (0.2)	47 (0.5)
Lost to follow-up	59 (1.8)	47 (1.5)	47 (1.5)	94 (1.5)	153 (1.6)
Noncompliance with study drug	21 (0.6)	19 (0.6)	9 (0.3)	28 (0.5)	49 (0.5)
Physician decision	39 (1.2)	30 (1.0)	22 (0.7)	52 (0.8)	91 (1.0)
Pregnancy	1 (<0.1)	1 (<0.1)	0	1 (<0.1)	2 (<0.1)
Protocol violation	22 (0.7)	24 (0.8)	14 (0.5)	38 (0.6)	60 (0.6)
Study terminated by sponsor	1 (<0.1)	1 (<0.1)	3 (<0.1)	4 (<0.1)	5 (<0.1)
Withdrawal of consent	145 (4.4)	90 (2.9)	102 (3.3)	192 (3.1)	337 (3.6)
Product quality complaint	0	1 (<0.1)	0	1 (<0.1)	1 (<0.1)
Unable to take protocol defined rescue therapy	38 (1.2)	17 (0.5)	15 (0.5)	32 (0.5)	70 (0.7)
Subject decided to discontinue early but agrees to be contacted	135 (4.1)	84 (2.7)	65 (2.1)	149 (2.4)	284 (3.0)
Other	228 (7.0)	150 (4.9)	156 (5.1)	306 (5.0)	534 (5.7)

さらに、本資料では各解析方法でのプラセボ群に対するカナグリフロジン 100 mg 群及び 300 mg 群の差をフォレストプロットとして示している (図 6-2)。各解析方法から得られた

結果を視覚的に比較できるため、フォレストプロットは有用である。

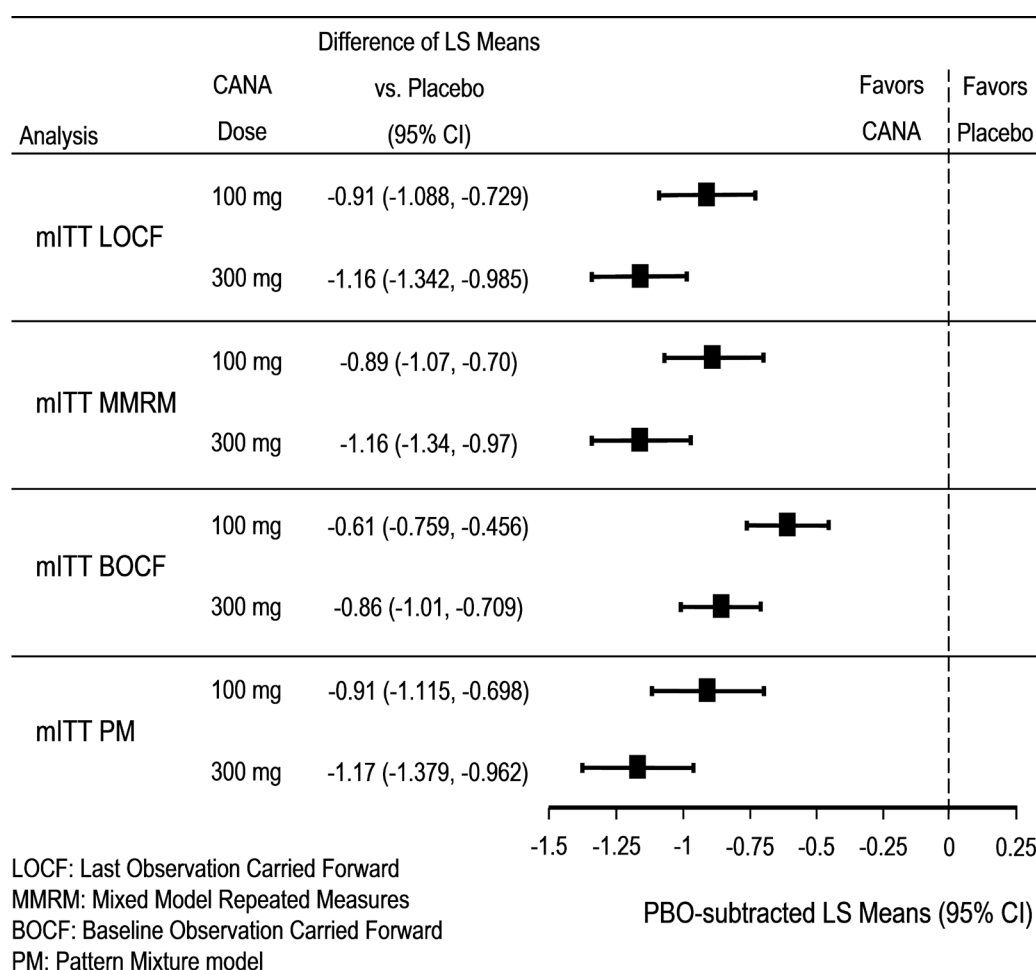


図 6-2. フォレストプロットによる主要な解析と感度分析の要約事例
(カナグリフロジン FDA 諮問委員会資料を改変)

PBO: Placebo, mITT: modified ITT

6.7 まとめ

欠測データを伴う不完全データに対して、普遍的に適用可能な解析方法はない。特に欠測データが多い場合には、感度分析という形で異なる欠測メカニズムや解析方法に基づき推測を行い、解析結果の安定性を検討し報告書に考察することが求められる。感度分析の方針や解釈について、規制当局を含めて完全にコンセンサスが得られたものはないが、欠測データが多い場合などには、感度分析が重要な位置づけとなるのは間違いない。感度分析の方針について、試験計画段階での十分な検討が必要であり、場合によっては、規制当局との相談が必要となることもある。さらに、試験途中の盲検下レビューにより、欠測データを扱う方法を洗練することも有用となる。そして、感度分析の結果を解釈する際には、

P 値に基づいて有意差だけを議論するのではなく、得られた推定値や、投与群ごとの欠測の発生割合などから総合的に考察する。

欠測メカニズムの仮定に対する感度分析の手法とその解釈は、現在も盛んに研究が行われている分野であり、最新の動向に留意されたい。

【参考文献】

- ICH E9 ガイドライン, 臨床試験のための統計的原則. (1998).
http://www.pmda.go.jp/ich/e/e9_98_11_30.pdf
- Mallinckrodt C, *Preventing and Treating Missing Data in Longitudinal Clinical Trials*. New York, NY: Cambridge University Press; 2013
- EMA, Sycrest Assessment report, EMA/CHMP/583011/2010
http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/001177/WC500096898.pdf
- FDA, Invoka Statistical Review (s)
http://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/204042Orig1s000StatR.pdf
- FDA, Invoka Advisory Committee Briefing Document
<http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM334551.pdf>

7. 欠測データの発生を最小限にするための方策及び試験計画の検討

7.1 はじめに

前述の通り、欠測データを伴う不完全データに対して普遍的に適用可能な唯一の方法は存在せず、感度分析によって主要な解析結果の安定性を検討することとなる。欠測データが多く含まれる場合には結果が不安定となる可能性が増し、解釈が困難となる可能性も高くなる。これらはサンプルサイズを増やすことだけでは対処することができないため、解析段階で工夫を凝らすよりも、まず欠測データの発生を最小限に抑えることが重要である。これは NAS レポートと EMA ガイダンスの共通するメッセージの一つでもある。以下に、欠測データの発生を最小限にするための試験デザインの工夫や、試験実施上のポイントを取り上げる。あわせて、試験計画時に考慮すべき事項を示す。

7.2 試験計画の検討

7.2.1 Estimand の設定

欠測データの影響を最小限に抑えるための第一ステップは、試験計画時に **estimand** (詳細は 3 章を参照) を詳細に検討し明確に設定することである。なぜなら、設定した **estimand** での欠測データの潜在的問題点を認識することが、欠測データ発生への対策を考えるうえで手がかりとなるからである。例えば、どういった欠測データが発生しうるか、欠測データがどのように評価へ影響するのか、試験治療中止後のデータ収集が必要か、などである。また、試験で示そうとしている証拠に対応する **estimand** が複数ある場合には、欠測データが生じにくいものを選択することも検討の対象となる。

7.2.2 欠測データの発生を最小限に抑える試験デザイン

NAS レポートでは、欠測データの大きな理由の一つである試験治療の中止を少なくするための試験デザインの例として、以下の 8 つを挙げている。

- 1) 試験治療に耐えうる、あるいは反応しうる被験者をランダム化前に特定するために **run-in** 期間あるいはエンリッチメントデザインを利用すること、
- 2) 被験者の状態に応じて投与量を適宜変更する **flexible dose (titration) study**,
- 3) ターゲットとする被験者集団を現在の治療で十分でないものに絞ること、
- 4) 効果のある既存治療へ上乗せする **add-on** デザイン、
- 5) 試験期間を短くすること、
- 6) 効果が無い場合にレスキュー治療の使用を許可すること、
- 7) 欠測データが多くなりそうなアウトカムを避けること、
- 8) ランダム化治療中止試験の利用、

である。これら試験デザインの詳細は、NAS レポートの 2 章や Mallinckrodt (2013) の 4 章を参照されたい。

ここで重要なことは、これら試験デザインを適用した場合に、他の側面へ負の影響があることを認識することである。例えば、run-in 期間を設定した試験や、ターゲットとする被験者集団を絞る試験の場合、欠測データの発生減少が期待できる一方、試験結果の一般化可能性は低くなる。既存治療への add-on デザインを用いた場合、投与方法 (dosing regimen) の薬効は検討できるが、当然ながら、被験薬単剤の薬効は検討できない。試験計画段階では、このようなトレードオフの関係を良く理解し、その試験の目的や estimand から適切な試験デザインを選択する。

7.2.3 サンプルサイズの検討

前述した通り、欠測データの問題として、治療効果の推定へのバイアスの発生や、解析結果の不安定さが増すことがあるが、それに加え、解析に寄与する被験者数減少に伴う検出力の低下がある。一般に、適切に設定した検出力を確保した臨床試験の実施が求められるが、従来、欠測データに伴う検出力の低下については単純な対応が多かった。例えば、サンプルサイズ計算で良く行われてきた方法は、欠測データを考慮せずにサンプルサイズを一旦計算し、そこで得られた被験者数から想定した中止割合を考慮し増加させる (i.e., 必要被験者 = 被験者数 / (1 - 想定する中止割合)) という単純なものである*。NAS レポートではこの点を指摘し、欠測メカニズムを考慮したシミュレーションによる検出力の検討を推奨している。この場合、例えば、各時点の治療効果や欠測データの割合、主要評価変数の時点間の関連性の強さ (例えば、今日と明日の臨床検査値の相関の強さ。LDL コレステロールは強く、中性脂肪は弱いことなどが知られている。など)、さらに欠測メカニズムなどを想定し、さまざまなシナリオの下で検出力を確認し、サンプルサイズを検討することになる。そして、シミュレーションを行うためのシナリオとして、過去試験データの再解析や類薬の文献調査など、念入りの事前準備も求められる。また、これらシナリオの妥当性の検討やシミュレーション結果の解釈の場面では、医師や臨床担当者と統計担当者などチーム内で綿密な議論が必要となる。

NAS レポートが言及するように、MAR や MNAR を仮定した場合のサンプルサイズ計算に関してはさらなる研究が必要である。なお、MMRM を適用した場合の検出力の計算に関しては、Kaifeng ら (2008) が参考になる。

* 例えば、次のようなサンプルサイズ計算が行われている。「被験薬とプラセボの変化量の差を 5 (標準偏差 10) と仮定した場合、 $\alpha=5\%$ (両側)、検出力=80% の条件下で、有意な差を得るためには 1 群あたり 64 例必要となる。中止割合 20% を想定し、本試験の目標被験者数は 1 群あたり 80 例 ($=64/(1-0.2)$) とした。」

7.2.4 プロトコールへの明記

NAS レポートでは、プロトコールに記載すべき内容を複数の **recommendation** の中で提言している。以下に代表的なものを示す (一部再掲)。

Recommendation 1: プロトコールには以下の項目：試験の目的、主要評価項目とその他の評価項目、評価項目の測定方法、治療効果の測定（つまり主要目的の **Causal Estimands**）を明確に定義すべきである。これらの測定はすべての試験の参加者にとって意味のあるもので、また最低限の仮定で推定可能であるべきである。後者に関して、プロトコールは欠測データの潜在的な影響とその扱いについて記載すべきである。

Recommendation 6: スポンサーは起こり得る欠測データの問題を明確に予測すべきである。特にプロトコールに欠測データに関する章を設け、欠測データの影響をモニターし、制限するために試験デザイン・実施の段階で取られるステップについて記載すべきである。

Recommendation 8: プロトコールでは欠測データの量を最小化することの重要性を認識させるべきである。特に、過去の試験結果に基づき、主要な結果について完全なデータが少なくともどのくらいあればよいのか設定しておくべきである。

Recommendation 9: 欠測データを取り扱う統計手法について、スポンサーは欠測データの取り扱いをプロトコールに記載すべきである。また、設定した仮定は臨床家にも理解できるようにする必要がある。

例えば、アメリカで実施される試験では、FDA から **estimand** や欠測データの取り扱い、及び感度分析の詳細をプロトコールに記載するよう要求されることがある。事後的に欠測データの取り扱いを決定することは、時に解析結果の受け入れを困難にする可能性がある。また、欠測データが多く、感度分析が重要な位置づけとなる場合には、その詳細をプロトコールに記載しておくことは価値があると考えられる。ただし、具体的にどのような内容をプロトコールに記載すべきかどうかは、医薬品開発の段階、試験の目的、さらには各規制当局の考えやスポンサーの判断によるのが实际的である。プロトコールへの記載方法以前に、ここであげられている事項を、試験計画時に良く検討することが本質的に重要である点を強調したい。

なお、上記の **recommendation 8** では、過去の試験結果に基づいて欠測データの割合についてベンチマークを設定することを推奨している。EMA ガイドラインでも、欠測データの割合を見積もることが望ましいとしている。例えば、欠測データの割合が想定よりも著しく高い場合に、試験デザインの変更を検討するための **DMC** の設置も考えられる。ただし、

NAS レポートでも、適切なベンチマークを設定するのが困難であることも示している。過去試験では欠測データの発生を抑えるためにどのような手段がとられたのか不明であり、その中止割合は当該試験に正しく当てはまらないかもしれない。従って、過去試験から欠測データの割合を想定し、試験開始時の検討に利用する場合には、広範囲な想定で検討することが必要になる。

7.3 欠測データの発生を最小限にするための試験実施上の方策

7.3.1 スポンサーが実施できる方策

NAS レポートでは、欠測データの発生を減らすためにスポンサーが実施できる具体的な方策が述べられている。例えば、被験者の負担や不便さを減らすために、1) visit や評価の回数を最小限にすること、2) 不要なデータを収集せず試験目的に必要な最小限の情報のみを収集すること、3) ユーザーフレンドリーな CRF を用いること、4) 可能な場合には、被験者の来院が不要なアウトカムを用いること、5) 評価のために十分な期間 (time window) を許容することが挙げられている。これらは試験計画時の検討事項になる。

また、NAS レポートでは、トレーニングの重要性を挙げている。試験の研究会などを通じて、治験責任医師・分担医師や施設関係者にデータの完全性の重要性を説明することが一つの方策となる。例えば試験治療中止後にデータを収集する試験では、なぜ必要であるのか、そして、試験治療中止と、試験中止の違いを強調して説明する必要がある。

データの完全性は、試験期間中のモニタリングでも改善できる。現在の臨床試験では EDC システムが広く使われており、随時入力されているデータを利用することで、試験治療の中止を含めた欠測データの発生状況を常にモニタリングすることが可能となる。これにより、欠測データの多い施設への再教育、場合によっては契約打ち切りといった柔軟な対策が可能となる。これは、近年注目されているリスクベースドモニタリングの考えと合致する。TransCelerate Biopharma 社のリスクベースドモニタリングに関するポジションペーパーでは、リスク指標の一つとして、ある施設の中止割合が全体の傾向と異なる場合にアクションを起こすことが例示されているので参考にされたい (TransCelerate, 2013)。

7.3.2 治験責任医師・分担医師や施設スタッフが実施できる方策

NAS レポートでは、治験責任医師・分担医師や施設スタッフが実施できる方策も述べている。興味深いのは、被験者へのトレーニングについても言及している点である。Mallinckrodt (2013)の 5 章では、被験者教育を高い質で一律に実施するためのアプローチの一例として、キーとなる手順やデータの完全性の重要性について説明するためのショートビデオの利用が述べられている。

臨床試験を実施する施設選定の際には、このような協力が得られる施設かどうか、プロトコルを適切に遵守できる施設かどうか、といった点を考慮する。例えば、プロトコー

ル遵守率について、日本の施設の方が海外に比べて高いと予想できる場合には、日本の施設を中心に臨床試験を実施していくことが考えられる。

なお、NAS レポートでは、金銭的なインセンティブによる被験者への働きかけがあげられているが、日本国内においては禁止されているため、日本で実施する臨床試験では適切ではない。金銭的なインセンティブが行われる国際共同試験に参加する場合には、特に留意する必要がある。

7.4 まとめ

欠測データの発生を最小限にするために、試験計画段階で綿密な検討をすることが重要である。試験の目的や *estimand* を明確にすること、さらに欠測メカニズムの想定、さまざまなシナリオの下での検出力確認、過去試験データの再解析や類薬の文献調査などが必要で、統計担当者と臨床担当者が協力して行うことが求められる。さらに、スポンサー、治験責任医師・分担医師や施設関係者ばかりでなく、被験者も含めて臨床試験の欠測データの問題を認識することで、欠測データの発生を最小限に抑えることが可能となる。

【参考文献】

- Mallinckrodt C. *Preventing and Treating Missing Data in Longitudinal Clinical Trials*. New York, NY: Cambridge University Press; 2013.
- Kaifeng Lu, Xiaohui Luo, Pei-Yun Chen. Sample Size Estimation for Repeated Measures Analysis in Randomized Clinical Trials with Missing Data. *The International Journal of Biostatistics*, 2008; 4 (1)
- TransCelerate Biopharma, Position Paper: Risk-Based Monitoring Methodology. 2013
<http://www.transceleratebiopharmainc.com/wp-content/uploads/2013/09/Risk-Based-Monitoring-Methodology-Position-Paper.pdf>

8. 終わりに

本報告書では、欧米規制当局の考え方の異同について触れ、現在のコンセンサスと議論の対象を明らかにした。そして、**estimand** を紹介することで、欠測データの取り扱いの決定には試験の位置づけが明らかにされる必要があることも示した。さらに、欠測データの種類と解析方法を示し、主要な解析方法の決定方法と感度分析の方法に触れることで、現実的にどのような問題が生じるかについても示した。本報告書をきっかけに、欠測データの取り扱いを真剣に考え、欠測データの扱いは事後的な解析方法で対処する問題ではなく、計画に織り込むべき問題であると認識頂けよう。

開発計画の中で、それぞれの試験で明らかにしようとすることに応じて、組み入れ基準や試験期間、用量、試験規模、主要評価変数、そして対応する解析方法が特定される。今後はこれらを決定することと同時に **estimand** を特定し、欠測データの影響を予測し、感度分析についても統計担当者と関係者で議論されるようになることを期待する。

9. 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 2013 年度タスクフォース 2

サブチーム 1. 本報告書担当

竹ノ内 一雅 アステラス製薬株式会社 (サブリーダー)
北川 忠行 武田薬品工業株式会社
佐伯 浩之 富士フイルム RI ファーマ株式会社
野村 真功 株式会社三和化学研究所
片岡 尚子 ノバルティスファーマ株式会社 (2013 年 10 月まで)

サブチーム 2. FDA, EMA, PMDA 事例調査担当

松岡 伸篤 ファイザー株式会社 (サブリーダー)
高木 弘毅 サノフィ株式会社
土川 克 ゼリア新薬工業株式会社
鵜飼 裕之 日本ベーリンガーインゲルハイム株式会社
笹川 裕次 Meiji Seika ファルマ株式会社
棚橋 昌也 興和株式会社

サブチーム 3. 統計方法論担当, SAS プログラミング開発担当

土居 正明 東レ株式会社 (サブリーダー)
藤原 正和 塩野義製薬株式会社
大江 基貴 株式会社大塚製薬工場
駒寄 弘 マルホ株式会社
高橋 文博 田辺三菱製薬株式会社
縄田 成毅 杏林製薬株式会社 (サブリーダー)
横溝 孝明 大正製薬株式会社
大浦 智紀 日本イーライリリー株式会社
横山雄一 持田製薬株式会社

タスクフォースリーダー兼, 推進委員

富金原 悟 小野薬品工業
菅波 秀規 興和株式会社
土屋 悟 大日本住友製薬株式会社

担当副部長

廣岡 秀樹 アステラス製薬株式会社

以上