1. 概 論

折 笠 秀 樹*

Handling of Missing Data in Clinical Trials: An Overview

Hideki ORIGASA*

* Division of Biostatistics and Clinical Epidemiology, The University of Toyama Graduate School of Medicine and Pharmaceutical Sciences, Japan

The European Medicines Agency issued a guideline on missing data in 2009. In 2012, the U.S. National Research Council published a book on the prevention and treatment of missing data in clinical trials, which has a great impact on medical fields. The most striking fact is that last observation carried forward (LOCF) is not recommended as a primary approach unless the assumption is confirmed. This article overviews some issues on missing data from statistical viewpoints. Prevention of missing data is very important in designing and conducting clinical trials. Once missing data are generated, an appropriate statistical method to handle the missing data is needed after ascertaining the missing mechanism as proposed by Rubin. Most statistical methods assume a mechanism of missing at random (MAR), where a systematic difference between complete and incomplete cases can be explained by a set of observed data. When there are confounding variables influencing missing data, mixed-effects models for repeated measures (MMRM) or other methods can be applied to adjust for confounding factors. Excluding all cases with missing data would usually induce a significant bias. Simple imputation such as LOCF will not be valid unless the underlying assumption is verified. Multiple imputation strategy may be valid under the assumption of MAR. Sensitivity analysis is necessary to confirm the consistency among results obtained by different approaches; for instance comparing intention-to-treat with per protocol set. More accurate reporting on missing data should be strengthened. Pre-specification of handling of missing data is also important in developing a study protocol.

Key words: missing data, guideline, missing mechanism, statistical methods, reporting

Jpn J Clin Pharmacol Ther. 2014; 45(4): 125-129

1. 最新事情

2009年には、欧州医薬品局(EMA)が欠測データの取扱いに関するガイダンスを発行した¹⁾。その後、2012年には米国学術会議(National Academy of Science)の中のNational Research Councilが欠測データの予防・処置という単行本を出版した²⁾。その中で18項目の勧告(Recommendations)が提唱された。中でも反響を呼んだ勧告はTable 1に示した3項目だろう。欠測データの補完(Imputation)に関し、現場で最も頻繁に使用されていたLOCF(Last Observation Carried Forward)という手法は、その仮定が満たされることを確認しないかぎり、主たる解析法としては使用すべきでないとされた。また、欠測データを扱う統計手法としてしばしば高価なモデルを用いるが、その際には仮定の検証が不可欠であるとした。さらに、欠測データの取扱いはプロトコルに明記すべきとした。それを受けて、New England Journal of Medicine誌は特別論文を掲載した³⁾。臨床試験の計画・実施段階において、欠測データを減らす手

Table 1 欠測データの対処法に関する National Research Council の報告書²⁾に示された勧告の中で反響を呼んだ3項目

R10.	Simple imputation methods like <u>LOCF</u> should <u>not</u> be used as a primary approach <u>unless</u> the assumptions are justified.
R11.	Parametric models should be used with caution. All their <u>assumptions</u> should be <u>clearly spelled out and justified</u> .
R9.	Statistical methods for handling missing data should be <u>specified in study protocol</u> .

段が示された。さらに、同誌のStatistical reviewerをしているハーバード大学のWare教授らからはコメントが提示された⁴⁾。

2. 偏りと精度

欠測データがあると、一般的には偏り(Bias)が発生するだろう。欠測データは、比較する両群で均等に発生しな

* 富山大学大学院医学薬学研究部バイオ統計学・臨床疫学

著者連絡先：折笠秀樹 富山大学大学院医学薬学研究部バイオ統計学・臨床疫学 〒930-0194 富山市杉谷 2630

E-mail: horigasa@las.u-toyama.ac.jp

ISSN 0388-1601 Copyright: ©2014 the Japanese Society of Clinical Pharmacology and Therapeutics (JSCPT)

Table 2 欠測のメカニズム

- **MCAR** (Missing completely at random)
 - Not depend on the observed and unobserved data.
 - 計測器が故障のため、今回欠測となった
- **MAR** (Missing at random)
 - Not depend on the unobserved data, but may depend on the observed data.
 - 若い患者での血圧データで、より多くの欠測例があった
 - ・欠測データは観測データよりも血圧値は低いだろう
 - ・患者が前回改善していたので、今回欠測となった
- **NMAR** (Not missing at random)
 - Depends on both the unobserved and observed data.
 - 高血圧患者ほど頭痛の為に来院せず、今回の血圧値が欠測となった
 - 患者の QOL が今回悪化したため、今回欠測となった

い可能性がある。そうすると交絡バイアスが混入し、比較可能性 (Comparability) が損なわれる。また、欠測データを除外することにより選択バイアスが混入し、それがデータの代表性 (Representativeness) を損なう。

一方、欠測データを除外することで解析対象例数が減少し、そのために検出力 (Power) は低下するだろう。しかし、欠測データとなる未完了者は極端な値を取りがちであり、そういった値を除外するとばらつきが減少する。そこで精度は上がるので検出力も上がると考えられる。このように、精度に関してその方向性はあいまいである。

3. 予防策

欠測データはできるかぎり発生しないように予防することが肝要である。計画段階と実施段階で予防策が考えられる。計画段階においては、試用期間 (Run-in period) を適用することで欠測データは防げるだろう。追跡期間 (Follow-up period) を短くすることでも欠測データは防げるだろう。エンドポイント選定においても、できるだけ欠測データが起りにくい項目を選ぶと良いだろう。一方、実施段階においては、収集するデータ項目数が多いと欠測が起りやすいので、調査項目は最小限にすることが望まれる。欠測データが生じないように、実施中に監視して注意喚起するのも有用である。計測ミスなどを防ぐために、計測法に関する訓練をすることも良いだろう。

4. 欠測メカニズム

Rubin 氏は欠測データを次の3つに分類した⁵⁾。すなわち、MCAR (Missing Completely At Random), MAR (Missing At Random), NMAR (Not Missing At Random) である (Table 2)。これは欠測理由の内容からというよりは、統計解析上の立場からの分類である。MCAR とは単純に偶然による欠測の場合である。すなわち、欠測値と観測値に系統的な違い (Systematic difference) は存在しない。一方、MAR

Figure 1 MCAR と MAR を見分けるための方法
(文献 6 より引用)

は観測値には依存するが、欠測値には依存しない。したがって、欠測値と観測値に系統的な違いは見られるものの、それは観測値で説明可能である。例を挙げよう。QOL 値が今回欠測となったのは、前回観測された QOL 値が悪かったからであり、今回の QOL (未観測) が悪かったからではない。このような例は MAR である。今回のデータ (欠測ではあるが) が原因で欠測となっていないことが大切である。最後の NMAR は今回のデータが原因で欠測となった場合である。欠測値と観測値には系統的な違いが見られ、それは観測値では説明できない場合が NMAR である。

MCAR と MAR を見分ける必要性は低いが、それを見分けるには、たとえば完了者 (非欠測例) と未完了者 (欠測例) を比較すればよい。Figure 1 のように系統的な違いが見られると MAR の可能性が高い⁶⁾。MCAR であれば系統的な違いを認めないからである。また、未完了者 (欠測例) には欠測となる原因 (MRSS 値が高かった) がグラフから読み取れる。このように、観測値から系統的な違いが説明される場合は MAR だろう。もし系統的な違いを説明するような観測値が見当たらないならば、それは NMAR であると思われる。

5. 対処法

計画・実施時に欠測データを極力減らすよう予防策を講じることが大切であるが、それでも欠測データは生じてしまう。その場合に欠測データをどのように取り扱うかという対処法 (Treatment) を考えないといけない。さまざまな対処法が目下利用可能である (Table 3)。単純なのは欠測データをすべて解析除外することだが、検出力が落ちるとともに、MCAR でないかぎり偏りも入りがちである。この対処法は Complete-Case Analysis (Available-Case Analysis) と呼ばれている。

第二は、除外ではなく補完 (Imputation) をする対処法である。補完には二つの手法があり、一つは単純補完 (Sim-

Table 3 欠測データに対する諸対処法 (Treatment)

- Exclusion
- Simple imputation
- Multiple imputation
- MMRM (Mixed-effects Models for Repeated Measures)
 - ↑ 連続データ
- GLMM (Generalized Linear Mixed Models)
 - ↑ カテゴリーデータ
- IPW (Inverse Probability Weighting)
- Challenges for the NMAR
- Sensitivity analysis

ple imputation) であり、もう一つは多重補完 (Multiple imputation) と呼ばれる手法である。単純補完にもいろいろある (Table 4)。LOCF (Last Observation Carried Forward) は臨床試験で頻繁に適用されるが、欠測の直前値で補完する方策である。BOCF (Baseline Observation Carried Forward) では初期値で補完する。Hot-deck では最悪値で補完する。また、欠測時点の平均値で補完したり、回帰式の予測値で補完したりする方策もある。LOCF は National Research Council の勧告でも敬遠されたが³²⁾、欠測時点が異なると偏りをもたらしることがある。アルツハイマー病やパーキンソン病のようにどんどん悪化する疾患では、Figure 2 のように評価値がどんどん悪化する⁷⁾。初期に脱落した場合 (Early escape) は LOCF により相当良く評価され、後期に脱落した場合 (Late escape) は LOCF により多少しか良く評価されない。これが偏りになる。

多重補完では補完を何回か繰り返す^{8~11)}。通常は 5 回ないしは 10 回繰り返すことが多い。回帰モデル等により単純補完するのではなく、モデルを変えたりして繰り返し補完する。そして、補完された複数個のデータセットを解析した結果を最後に統合する。多重補完は MAR の下では妥当な統計手法とされる。統計ソフト SAS では、PROC MI と PROC MIANALYZE で解析可能である。

MMRM (Mixed-effect Models for Repeated Measures) は正規分布したデータに適用される、モデルベースの統計手法である。反復測定データに対する混合モデルとして知られる。これも MAR の下で妥当な手法とされる。統計ソフト SAS では PROC MIXED で解析可能である。カテゴリーデータにおけるモデルベースな手法は GLMM (Generalized Linear Mixed Models) と呼ばれるが、これも MAR の下で妥当な手法とされる。

次に、IPW (Inverse Probability Weighting) では Complete cases には重みを大きくし、Incomplete cases には重みを小さくする¹²⁾。IPW regression model は GEE (Generalized Estimating Equations)¹³⁾を用いて推論する。ここでも MAR を仮定しているが、さらに観察データと未観察データは同じ分布に従うことも仮定している。

Table 4 さまざまな単純補完法 (Simple imputation)

- LOCF
 - Last observation carried forward
- BOCF
 - Baseline observation carried forward
- Hot-deck imputation
 - Worst case imputation
- Mean imputation
- Regression imputation

Figure 2 LOCF 法によるバイアスの発現状況について (文献 7 より引用)

これまで述べた手法はすべて MAR または MCAR を前提とするが、現実には NMAR が想定される欠測データもある。欠測例と完全例ではエンドポイントに違いが見られ、その違いが交絡変数で説明できない状況である。説明変数が特定できないので、上で述べたような回帰モデルの手法は使えない。目下のところ二つの手法が提案されている。一つは Selection models¹⁴⁾であり、もう一つは Pattern mixture models¹⁵⁾である。いずれも完璧とは言えない。これについては詳細を割愛させていただく。

欠測データの取扱いにはいろいろあり、手法によって結果が違ってくることがある。そこで、手法により結果に顕著な違いはないことを確認する必要がある。そのことを感度分析 (Sensitivity analysis) と呼んでいる。Figure 3 では、ITT (Intention-to-treat) 解析と PP (Per Protocol) 解析で結果は変わらないことを確認した¹⁶⁾。

6. 報告の現状

Figure 4 に報告の現状を示した¹⁷⁾。欠測データは除外するという Complete case analysis あるいは単純補完法が 95% を占めていた。IPW 法・多重補完・MMRM など高度な手法は、わずか 13% しか使用されていなかった。手法は重複使用されているので合計は 100% を超えている。Figure 5 からは、欠測メカニズムが論文でほとんど言及されていない現状が分かる¹⁷⁾。目立つ現象としては、近年多重補完を

	Adalimumab group	Tocilizumab group	Difference (95% CI)	p value
Primary endpoint (intention to treat, <u>last observation carried forward</u>)*	-1.8	-3.3	-1.5 (-1.8 to -1.1)	<0.0001
Sensitivity analysis (intention to treat, <u>observed data</u>)*	-2.2†	-3.8‡	-1.6 (-2.0 to -1.2)	<0.0001
Mixed-effects least square means estimate (SE; intention to treat) MMRM	-2.0 (0.1388)	-3.5 (0.1389)	-1.5 (-)	<0.0001
Sensitivity analysis (<u>per protocol</u> , last observation carried forward)	-2.0§	-3.3¶	-1.3 (-1.7 to -0.9)	<0.0001

If an erythrocyte sedimentation rate value was missing, we used the last observation carried forward. If erythrocyte sedimentation rate was 0, it was set to 1 mm/h to enable calculation of Disease Activity Score. For observed data, no imputation was done for missing data for change from baseline at week 24. *Means adjusted. †n=127. ‡n=131. §n=130. ¶n=143.

Table 2: Change from baseline in Disease Activity Score for 28 joints at week 24

Figure 3 関節リウマチ臨床試験での感度分析の実例（文献16より引用）
感度分析とは手法を変えて結果の整合性を確認すること。

Methods Reported	All Studies (n = 262) ^a %	Multi-item Instrument Studies (n = 62) ^b %
Complete-case analysis	81	79
Single imputation techniques	14	19
Including missing indicator as extra answer category	6	7
Inverse probability weighting	3	2
Multiple imputation	8	5
Maximum likelihood estimation	2	2
Unclear	13	13
No Information	2	3

^a67 studies reported multiple methods to handle their missing data.

^b16 studies reported multiple methods to handle their missing data.

Figure 4 欠測データの対処法に関する文献調査（文献17より引用）

Methods Applied or Reported	% of Studies
Selectivity of missing data examined (n = 262)	22
Method used to compare responders with nonresponders (n = 58)	
Differences in characteristics described in text	54
Descriptive statistics presented	22
Statistical test performed ^a	14
Unclear	10
<u>Reported mechanism</u> (n = 58)	
MCAR	0
<u>MAR</u>	14
MNAR	0
Unclear	2
<u>No Information</u> ^b	85

Figure 5 欠測データのメカニズムに関する報告状況（文献17より引用）

使用した事例が増えていることが挙げられる (Figure 6)¹¹⁾.

7. まとめ

欠測データの取扱いに関する現状について紹介した。まずは欠測データが発生しないような努力が望まれる。仮に欠測データが生じてしまったら、欠測メカニズムを踏まえたうえで、適切な統計手法による解析が求められる。また、欠測データの取扱いについては論文中へ適切に記載すべきである。

Conflict of Interest

著者は本稿作成に際し、開示すべき利益相反関係はない。

文 献

- 1) European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials (Draft). London: EMA-CHMP, 2009.
- 2) National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, D. C.: The National Academies Press, 2010.
- 3) Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012; **367**(14): 1355–60.
- 4) Ware JH, Harrington D, Hunter DJ, D'Agostino RB, Sr. Missing data. *N Engl J Med*. 2012; **367**: 1353–4.
- 5) Rubin DB. Inference and missing data. *Biometrika*. 1976; **63**(3): 581–92.
- 6) Wong WK, Boscardin WJ, Postlethwaite AE, Furst DE. Handling missing data issues in clinical trials for rheumatic diseases. *Contemp Clin Trials*. 2011; **32**(1): 1–9.
- 7) 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. *Clin Pharmacol Ther*. 2012; **91**(3): 550–4.
- 8) Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009; **338**: b2393.
- 9) Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. *Am Stat*. 2001; **55**(3): 244–54.
- 10) He Y. Missing data analysis using multiple imputation: getting to the

Figure 6 欠測データに多重補完法を使用した論文数の推移 (文献 11 より引用)

heart of the matter. *Circ Cardiovasc Qual Outcomes*. 2010; **3**(1): 98–105.

- 11) Mackinnon A. The use and reporting of multiple imputation in medical research: a review. *J Intern Med*. 2010; **268**(6): 586–93.
- 12) Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2011; **22**(3): 278–95.
- 13) Liang KY, Zeger SL. Inference based on estimating functions in the presence of nuisance parameters. *Stat Sci*. 1995; **10**(2): 158–73.
- 14) Diggle PJ, Kenward MG. Informative drop-out in longitudinal data analysis (with discussion). *Appl Statist*. 1994; **43**(1): 49–93.
- 15) Little RJA. Pattern-mixture models for multivariate incomplete data. *J Am Stat Assoc*. 1993; **88**(421): 125–34.
- 16) Gabay C, Emery P, van Vollenhoven R, Dikranian A, Alten R, Pavelka K, et al. Tocilizumab monotherapy versus adalimumab monotherapy for treatment of rheumatoid arthritis (ADACTA): a randomized, double-blind, controlled phase 4 trial. *Lancet*. 2013; **381**(9877): 1541–50.
- 17) Eekhout I, de Boer RM, Twisk JW, de Vet HC, Heymans MW. Missing data: a systematic review of how they are reported and handled. *Epidemiology*. 2012; **23**(5): 729–32.