



欠損値の補完方法【第48回生物統計学】

1 はじめに

臨床試験が開始されてから終了するまで、被験者が途中脱落などの理由で少しずつ減っていくがあります。その場合、試験を脱落した被験者のデータは、脱落した時点以降欠損値となってしまいます。今回は、そのような場合の欠損値を補完する方法のうちBOCF（Baseline Observation Carried Forward）とLOCF（Last Observation Carried Forward）について紹介します。

2 欠損値の影響

まず、欠損値が与える影響について説明します。欠損値が臨床試験の推論に与える影響は以下の3項目になります。

1. 制度の低下（検出力の低下）

推定値の推定制度はデータ数が多くなれば、標準誤差が小さくなるため高くなります。すなわち、欠損値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くなれば、標準誤差が大きくなることによって推定値の制度が悪くなり、検出力の低下に繋がります。このことによって、信頼区間が広くなり、P値が大きくなります。場合によっては、想定通りの介入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えます。

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いい結果となる（図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ます。

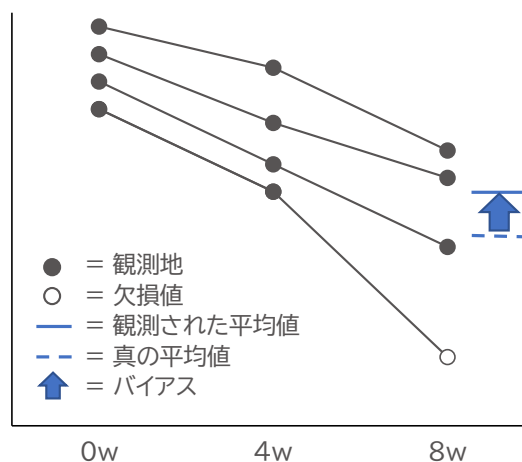


図1. 欠損によるバイアスの例



3. 試験結果の安定性が損なわれる

全ての欠損値の本当の発生理由は突き止められない場合が多いし、欠損でなかったとしたらどのような値をとっていたかわかりません。だからこそ、様々な仮定を置いて「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と思われます。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ります。

このように欠損値が少ない場合には、結果は安定し、有効性の証拠として受け入れやすくなります。これに対し欠損値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す。

3 欠損値の補完方法

3-1. BOCF

BOCF は、欠損値をベースライン値で補完する方法です。これは「治療後、ベースラインの状態に戻る」という前提のもと、欠損値をベースライン値で補完して解析する方法です（図2）。BOCF は、試験を中止後、ベースライン時に戻ると臨床的に考えられる場合に、保守的な方法として推奨されます。

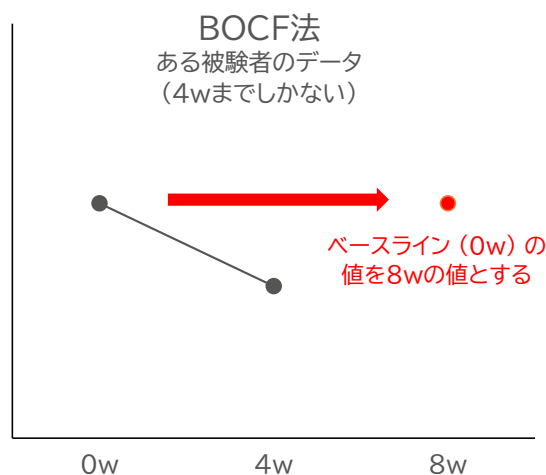


図 2. BOCF 法



3-2. LOCF

LOCF は、欠損値を直近の先行観測データで補完する方法です。これは「4w の状態が中止後も 8w まで続く」と仮定できる場合に、先行観測データで補完して解析する方法です（図 3）。臨床試験では頻繁に適用されています。

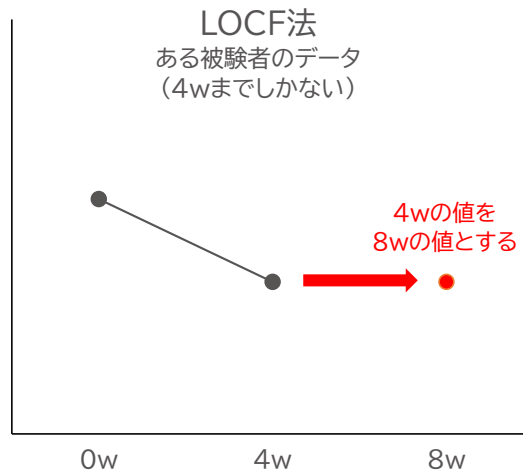


図3. LOCF 法

4 まとめ

今回は欠損値の補完方法について紹介しました。欠損値が多い状況では結果への影響が甚大なものとなってしまいます。欠損値に対する絶対的な対処法は存在しないため、欠損が発生しないよう予防策を検討することが最も重要です。仮に欠損値が発生してしまった場合は、欠損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ます。