

2020年6月

多重性とは?【第13回生物統計学】

1. 第1種の過誤と第2種の過誤

機能性表示食品（例えばEPA含有サプリメント）に効果があることを証明する場合、以下のように統計解析を行う必要があります。

Step1 「仮説を設定する」

帰無仮説 (H_0): EPA含有サプリメントとプラセボのLDL-コレステロールの変化率に差がない
対立仮説 (H_1): EPA含有サプリメントとプラセボのLDL-コレステロールの変化率に差がある

Step2 「検定統計量を算出する」

Step3 「有意確率を算出する」

Step4 「有意差の有無を判定する」

有意水準を5%未満に設定した場合、 $P < 0.05$ で有意差ありとし、対立仮説 (H_1) を採用する (※)

※臨床的意義のある変動かどうかとも考慮したほうが良い

機能性表示食品に効果があるか否かについての真実は分からない下で検定を用いて意思決定を行うので、仮説検定で誤った判断をする可能性が常に存在します。

- 本来は効果がない機能性表示食品を効果があると判断する誤り（第1種の過誤 / α エラー）
- 本来は効果がある機能性表示食品を効果がないと判断する誤り（第2種の過誤 / β エラー）

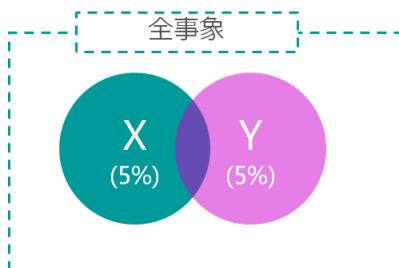
上述の2点は、いずれも避けることが理想ではありますが、第2種の過誤は機会損失にはなるものの現状に変化を生じさせません。一方で、第1種の過誤は誤った選択肢を増やすことで消費者を危険に晒すことで、社会的損失を招く恐れがあります。また、効果がない機能性表示食品を購入（選択）されることに起因する「機能性表示食品の本来の目的」の達成から遠ざかることとなります。

仮説	真理	
	帰無仮説を棄却 (有意差あり)	帰無仮説を採択 (有意差なし)
帰無仮説を棄却 (有意差あり)	正しい	α エラー
帰無仮説を採択 (有意差なし)	β エラー	正しい

第2種類の過誤を小さくすることには限度があるため、消費者危険を重視する立場から第1種の過誤を許容可能な範囲まで小さくする事が求められており、その考え方にに基づき、検証試験においては第1種の過誤を5%以下、第2種の過誤を10~20%に抑えることが通例となっています。

2. 多重性の問題

主要な評価項目 (X と Y) を2つ設定した場合の概念図を以下に示しました。



上図のような状況かつ「本来は効果がない」という状況の場合、検定結果は以下の4通りが考えられます。

- ① 1つ目の検定のみが誤って効果ありとする場合 (X)
- ② 2つ目の検定のみが誤って効果ありとする場合 (Y)
- ③ いずれの検定も誤って効果ありとする場合 ($X \cap Y$)
- ④ いずれの検定も正しく効果なしとする場合 ($\overline{X \cup Y}$)

この時、「効果がない」と正しく判断できているのは④であり、他の結果は、「効果がある」と誤って判断しています。「効果がない」機能性表示食品を「効果がある」と判断してしまう確率を5%以下に抑える必要がありますが、2つの円のそれぞれが5%となるように正しい検定方法を用いたとしても、①+②+③の領域の合計は1つの円の大きさを超過しているため、5%を越えてしまいます。よって、XとYでそれぞれ有意水準を5%となるように検定を実施しても、検定を複数回実施することによって、全体の第1種の過誤確率が上昇することがわかっていきます。

実際の有意水準を計算すると・・・

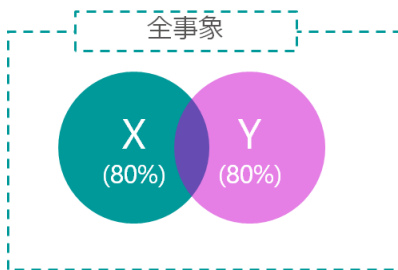
$$1 - (1 - 0.05)^2 = \frac{39}{400} = 0.0975 > 0.05$$



となり、有意水準を5%に設定していたはずなのに9.75%に上昇しています。
そのため、有意水準または算出された有意確率を調整する必要があります。

3. 第2種の過誤確率の制御と検出力の確保

1 (100%) から第2種の過誤確率 (真実は「差がある」にも係わらず差がないと判断してしまう確率) を引いたものを統計学的検出力 (「差がある」場合に正しく「差がある」と判断する確率) と呼びます。主要な評価項目 (X と Y) を2つ設定した場合に対立仮説が正しい下で検定を2回実施した際の検出力の概念図概念図を以下に示しました。



上図のような状況かつ「本来は効果がない」という状況の場合、検定結果は以下の4通りが考えられます。

- ① 1つ目の検定のみが正しく効果ありとする場合 (X)
- ② 2つ目の検定のみが正しく効果ありとする場合 (Y)
- ③ いずれの検定も正しく効果ありとする場合 ($X \cap Y$)
- ④ いずれの検定も誤って効果なしとする場合 ($\overline{X \cup Y}$)

①と③で構成される円はXの検定の検出力、②と③で構成される円はYの検定の検出力を示しています。円が重複する③の部分は双方で効果を正しく判断できる状態であり、円の外側である④の部分はいずれの検定でも効果のありなしの判断を誤ってしまう状態を示しています。この場合、いずれかの検定で「効果がある」と正しく判断できるのは①、②、③であり、いずれも「効果がない」と誤って判断してしまう場合は④となります。しかし、①と②は一方の仮説を正しく判断できないため、2つの仮説を同時に検証したい場合は、③の部分を一定の検出力を保てるように症例数を設定する必要があります。

各検定における評価項目間に関連がなく独立であり、それぞれの検定の検出力を80%とする場合、③の確率は $80\% \times 80\% = 64\%$ となり、2つの検定に対する検出力は個々の検定の検出力より小さくなることに注意が必要です。2つの検定における検出力を80%程度に確保したい場合は、個々の検定の検出力を

90%とすることで $90\% \times 90\% = 81\%$ となることから考慮することができます。この場合は個々の検定の検出力を90%に上げる必要があるため、目標症例数が増大することに注意してください。

4. まとめ

今回は多重性の基本的な概要を紹介しました。これらは機能性表示食品の臨床試験でも重要なポイントとなります。機会がありましたら、どのような場合に多重性の問題が発生するのか?多重性への対策はあるのか?どのような手法で考慮できるのか?を紹介していきたいと思います。

臨床試験は是非、オルトメディコまで!

5. 参考文献

- 臨床試験のための統計的原則 [<https://www.pmda.go.jp/files/000156112.pdf>]