シリーズ

FDAの一室から

米食品医薬品局医療機器・電磁波製品審査センター
循環器医療機器審査部審査官
内田毅彦

第11回　医療機器の臨床試験（2）

試験のデザイニング（最近の話題）

●ベイズ統計学

近年，臨床試験により多く使われる傾向にあるのがベイズ統計学である。これは，臨床試験のダウンサイジングにもつながる統計学的な手法の１つで，このベイズ統計学に対して，通常用いられている統計学は古典的あるいは頻度論的統計学などと呼ばれている。両者の詳細な説明やよしあしについては他の専門書に譲る。

頻度論的統計学では，知りたいことの本当の答（真の値）は，全対象集団のなかに確かに固定値として存在していることを前提にしている。実際にはすべての対象者，例えば臨床試験の適応・除外規準を満たすすべての人を試験に組み入れて結果（真の値）を出すことは非現実的であるから，今試験のために選ばれた（抽出された）対象の値を利用してその真の値を推論するわけである*。

一方，ベイズ統計学では以前に得られた情報（過去の試験結果など）を用い，現在，得られた試験結果から知りたい値があると考えられる範囲をベイズの定理に基づいて確率的に求めていくというものである。

頻度論的なデザインでも，以前の試験データを新しい試験に組み入れて試験をデザインすることが可能ではあるが，データ間の整合性などにより厳しい制約を受ける。この点，ベイズ統計学流の試験デザインでは，より柔軟に以前のデータを有効利用することで被験者数を少なくすることが容易にできるメリットがある。また，頻度論では第一種の過誤のコントロールの問題で単純に試験の途中で被験者数を増やすことはできないが（*），ベイズ統計学では，試験の途中で被験者数を増やすことが可能であるというメリットもある。また，頻度論では，まず仮説ありきであって，あらかじめ描いた試験デザインが終了して初めて結果を評価できるのであるが，ベイズ統計学では試験データの積み重ねにより，逐次知りたい値を確率的に入手できるので，途中結果によって，ランダム化の方法を変更したり，非劣勢の試験デザインを優位性に変更することも許されうる。となると，ベイズ統計学のほうが常によいのではということになりそうだが，必ずしもそうではない。

前述の通り，ベイズ統計学は以前のデータ（事前の確率分布）と得られた試験結果から，結論（事後の確率分布）を得るが，以前のデータの選び方次第で結論が変わりうるから，その選び方に議論の余地がある。また，同じ事前情報と新たに得られた試験結果を用いても，結論に達する方法（確率モデル）は単一でなく，方法によって得られる結果が変わりうるから，確率モデルの選択法についても評価の対象となる。さらに，ベイズ統計学を用いた臨床試験は，上述の試験デザインに関する問題のほかにも解析方法の複雑さなどから，

9月末に３日間の日程で開かれたFDA/Industry Statistics Workshopの会場。FDAは産との連携プロジェクトにはとても積極的で，フットワークも軽い

準備がたいへんであるうえに，生物統計家以外の臨床試験スタッフが理解しにくいなどの問題がある。

しかしながら，医療機器は医薬品よりも開発のスパンが短く，市場に登場した後も短命であるため，以前のデータを効率的に利用し，試験の規模を縮小させる目的とその柔軟性がためにベイズ統計学を利用した臨床試験が増えつつある。医療機器・電磁波製品審査センター（CDRH）では，ベイズ統計学を用いた臨床試験に関して，事前情報と確率モデル（第一種の過誤をコントロールするためのシミュレーション）については慎重に審査を行っている。また，その複雑さゆえに，試験開始前にデザインなどについて相談することを強く勧めている。なお，最近CDRHからベイズ統計学についてのガイドラインが出たので付記する。統計学は専門家でもない限り“よくわからない”と避けて通る医療関係者も多いと思うが，このガイドラインは難解な数式を避けてわかりやすく書かれていると思う（http://www.fda.gov/cdrh/osb/guidance/ 1601.html）。

●Adaptive Trial Design

まだ適切な日本語を見かけないこの“順応試験デザイン”は，先日行われたFDA/Industry Statistics Workshopでもベイズ統計学と並んで２大トピックの１つであった。実際に臨床試験を開始してみると，開始前の予想よりも当該治療法の治療効果が高かったり，低かったりする。

前述の通り，頻度論的な試験デザインでは第一種の過誤のコントロールの問題から試験の途中で被験者数を増やすことはできないため，一度試験を開始すると仮に途中で最終結果がある程度予測できても，みすみす試験を開発上，意味のないものにしてしまうこともありうる。医薬品の第Ⅲ相試験の実に50％は予想に反する結果が出たために，申請に使えなかったという報告がある。

しかし，統計学を歴史的に見ればつい最近になって（1990年代半ばから），数学的に第一種の過誤の問題を克服する方法が考えられるようになり，無意味になりそうな試験をそのまま継続できるかもしれないということから，この“順応試験デザイン”が脚光を浴びるようになった。ただ，試験の途中で被験者を増やすことが許されると，費用を削減するために，少なくスタートして必要なだけ後から増やすケースも考えられるが，そう単純ではない。

例えば，プロトコルに“結果によっては途中で被験者数を増やすという変更を加える”とあって，実際にそうなったらどうであろうか。試験を実施している医療スタッフは，「予想より治療効果がよくないのでは」と不審に思うもしれない。医療機器の臨床試験では，特に盲験化が難しいことから，該当治療に対するその後の評価にバイアスが生じる可能性が高くなる。また，それほど治療効果がないと思えば，医師は試験プロトコルを無視して別の治療法を追加しようと考えることもあり，被験者を集めにくくするという問題も起こりうる。さらに，予想を下回った治療効果でも確かに被験者数を増やすことで有意差が出ることになりうるわけだが，この場合の治療効果が臨床的に意義があるかどうかも重要である。

“順応性”については，被験者数だけでなく，エンドポイントや患者の適応・除外基準の変更なども範疇に入る。そして，これらの“順応”のために中間解析の際に，盲験を解除しなくてはならない状況もあり，中間解析と判断がだれによって行われるかによって，さらなるバイアスが試験に加えられる可能性も出てくることも考慮が必要である。

これらの“順応”は試験デザインに対して順応，すなわちその順応性が試験デザインのなかに組み込まれているべきであって，むやみに試験の途中で試験デザインを変更することを意味するのではない。あらかじめプロトコルに“順応”のルールについて可能な限り詳細に定めておくことが望ましい。FDAは“順応試験デザイン”のプロトコルを受け付けているが，今後この種の臨床試験が広まるには，上述のようなさまざまな問題点を克服するためのしっかりとした議論が必要なことは言うまでもない。“順応試験デザイン”は同じように柔軟な試験デザインのなかでも，試験の途中で試験を打ち切るタイプの群逐次デザインに劣るとの見解もが，それでもFDAがこうした試験デザインを完全に否定しないのには，臨床開発や医学の進歩について，経済効率なども踏まえて大きな視野で見ているからにほかならない。開発コストの削減が，薬価や医療機器価格の抑制につながると考えれば，医療費の抑制が重要課題である今，FDAのこうした姿勢は現実的で，リーズナブルであると思う。

* 抽出された対象の値は，抽出の度合いに変わりうるため，それがどれくらい真の値に近そうか推論し，仮説を検証する必要がある。仮説検証（A＝B）においては，仮説は正しいというところからスタートする。そして，実際の試験ではAとBで差が生じた際，この差が今回示された確率（これがP値）を治療効果の差（被験者数）を踏まえた仮説のもとで計算する。これが５％以下なら，もはやこの差は偶然とは言えず，それはそもそも仮説が正しくなかったからだということになる。したがって，試験の途中で被験者を増やすことは，仮説自体をゆがめるので許されないのである。

一方，ベイズ流の考え方では，以前から知っている情報に，新しい情報（試験結果）が加わることで，知りたい結果についての確率を直接求めることができるため，考え方がよりわれわれの日常生活に近いと言える。

■訂正　10月26日号45ページの第10回のPropensity スコア解析の注釈〔注３）〕に一部誤りがありました。おわびして以下の通り訂正します。

◇

Propensityスコア法は，非ランダム化比較試験 において，治療群と対照群の間に生じるバイアス（結果に影響を及ぼす他の因子の両群間での偏り）の軽減を図る方法である。

冠動脈ステント施行例で，ランダム化比較試験による従来型ステントAのデータと新しいステントBを比較する場合を考える。まず，知りたいエンドポイント（再狭窄率など）の結果を除き，ステントAの患者は１，Bは０として（逆も可能），新たに従属変数をつくり，糖尿病歴や病変長などの重要な独立変数を含んだ多重回帰分析モデルを用いて，どのような因子を持っている人が，どれくらいステントA（またはB）で治療されるかについての確率を求める。この確率がpropensityスコアであり，確率なので０～１の値を取る。１対１でランダム化すると理論上，すべての被験者のスコアが0.5となる。このスコアをA群とB群で比較することによって，両群で生じているバランスの程度について知ることができる。そして，このスコアに基づいて，両群からスコアが似ている患者同士をマッチングして比較する方法や，スコアを５分割に層別化して比較する方法，再狭窄率を従属変数とした多重ロジスティック回帰分析のモデルにpropensityスコア自体を組み入れて比較したりする方法，さらにはそれらを組み合わせて比較する方法を用いて，知りえたエンドポイントについてA群とB群を比較する。