

2018年3月9日提出

ICH E9(R1):「臨床試験のための統計的原則補遺 臨床試験における estimand と感度分析(案)」へのパブリックコメント

京都大学医療統計 佐藤俊哉

1. ガイドライン全般 1

E9 本文が 38 ページに対し、補遺が 30 ページというのはあまりにも分量が多すぎると思います。本補遺における「統計的原則」とはなにかを整理していただき、不要なところは削除して、コンパクトな補遺にまとめていただくようお願いします。

「estimand」という概念を中心とした新しいフレームワークは立派な統計的原則だと思います。しかし一度 estimand を定義してしまえば、その estimand を推定可能とするために中間事象をどう取り扱うかは(当局との合意は必要ですが)スポンサーの責任であり、統計的原則ではありません。

また、「A.7. 一般的な例」には「この例は規制上の推奨事項と捉えるべきではなく(641行)」と書かれているように「医薬品規制調和国際会議」のガイドラインに加えるべき内容ではないと考えます。これらは、ケース・スタディのような形で、本補遺とは別文書にして公表されてはいかがでしょうか。

というのも、本補遺のタイトルが「臨床試験における estimand と感度分析」であるにも関わらず、「A.5.2. 感度分析」は 2 ページ半しかなく、estimand の記述とくらべて分量に差がありすぎると思います。estimand に関する部分は短くして、感度解析の記述を増やすべきではないでしょうか。

2. ガイドライン全般 2

タイトルで「estimand」が英語のままとなっています。これまで ICH E シリーズの翻訳でタイトルに英語が含まれているものはなかったと思います。「estimand」は本補遺の中心ですし、コメント 8 に述べますように「estimand」は統計家以外の臨床試験関係者にも知悉してもらう必要があることから、適切な日本語訳をお願いします。

3. 「A.1. 目的と適用範囲」、39-45 行

「第一に、ICH E9 は、治療方針の効果、すなわち予定した試験治療のコースの遵守状況に関わらない、ベースライン時に最初に割付けられた治療の効果と関連して、intention-to-treat (ITT) の原則を導入し、ランダム化の維持が統計的検定の確固とした基盤を与えることを示した。ランダム化は比較試験の基礎であり、解析はランダム化の利点を最大限に活用することを目指すべきであることについては、議論の余地がない。しかし、治療方針の効果を理解することが常に規制上及び臨床上の意志決定に最も関連する治療効果を目的としているものなのかという疑問が残る。」

ICH E9 では「治療方針の効果」という記載があるのは用語集のみであり、本文中では用いていません。E9 は FAS (ITT) にもとづく治療効果の推定を中心にまとめています。それは、本補遺で言うところの中間事象がある場合、治療効果(平均因果効果)を偏りなく推定する方法に関するコンセンサスが、E9 作成当時の臨床試験コミュニティにはまだなかったからです。(当時は操作変数法に関する論文が、臨床試験領域で発表され始めた時期でした。)

このため E9 では、中間事象がある場合でも多くの場合に妥当なアルファレベル検定を与える「ITT の原則」を述べたものです。これは決して「治療方針の効果」を推定することを推奨しているのではなく、治療効果の推定にはバイアスは入るものの、検定のアルファレベルは保たれる方法を積極的に採用したものです。

(当時の E9 EWG メンバーの一人、Susan Ellenberg ペンシルバニア大教授に 2017 年の Joint Statistical Meeting でこの点について確認したところ、Ellenberg 先生も同様の見解でした。)

4. 「A.3.2. 中間事象に対応するためのストラテジー」全般

A.3.2 では「中間事象に対応するためのストラテジー」が estimand を規定するように書かれていますが、これらのストラテジーは(別の) estimand に対する偏りのある推定量を与えるものと考えられます。コメント 3 で述べたように、「治療方針ストラテジー」は「仮想ストラテジー」にもとづく estimand の偏りのある推定量を与えている、とも解釈できます。

しかし、本補遺では「特定の estimand の偏りのある推定量」(たとえば、治療方針ストラテジーもとづく推定量は仮想ストラテジーにもとづく estimand の偏りのある推定量)という視点がみられません。このため、target of estimation である estimand と「偏りのある推定量の期待値(仮想ストラテジーにもとづく estimand の推定に治療方針ストラテジーにもとづく偏りのある推定量を使用しているのにも関わらず、偏りのある推定量の漸近期待値を「治療方針ストラテジーにもとづく estimand」と解釈すること)」が混乱して使われていますので、区別して記述してください。

5. 「A.3.2. 中間事象に対応するためのストラテジー」、177-178 行

「その他の estimand の要素とともに、どのように中間事象に対応するかを選択することで、対象とする治療効果を記述する。」

本補遺では、6 ページの図にまとめられているように、estimand を定義してそれに沿って立案、計画、実施、解析を整合させる(alignment は「整合させる」よりも「順序立てる」でしょうか)、としていながら、「どのように中間事象に対応するかを選択することで、対象とする治療効果を記述する。」と estimand と中間事象の対応の順序が逆転している記載が随所にみられ、混乱させられてしまいます。あくまでも最初に estimand を決め、次いでその estimand が正しく推定できるように中間事象の取り扱いを選択する、ということが統計的原則であると考えますので、本補遺の記述もそのように統一してください。

また、コメント4で述べたように、estimandを正しく推定できる中間事象の取り扱いがない、または取り扱いが難しい場合には、別な estimand に変えるのではなく、もともとの estimand に対し偏りのある推定量を与える中間事象の取り扱いを選択することもありうると思います。

6. 「A.3.2. 中間事象に対応するためのストラテジー」、複合ストラテジー(203-222行)全般 複合ストラテジーは本補遺の統計的原則にそぐわないと思います。

複合ストラテジーは中間事象そのものの取り扱いではなく、主要な変数の定義を変えることでもとの estimand では中間事象であったものを、中間事象ではなくしてしまう方法です。主要な変数の定義を変えることから、必然的に estimand 自身も変わってしまうため、はじめに適切な estimand を定義するという6ページの図に整合しないため削除するか、本補遺に残すのであれば、何度も強調していますように、特定の estimand の偏りのある推定量を与えるという観点から複合ストラテジーを位置づけるべきだと思います。

7. 「A.3.2. 中間事象に対応するためのストラテジー」、216-222行

「中間事象と考えられる事象自体が、関心のある治療効果を定量化するために測定され得る最も意味のある変数である場合もある。死亡が起きる場合はこれに相当し得る。被験者が死亡したという事実は死亡前の観測値よりもはるかに意味がある可能性があり、また死亡後の観測値は存在しないからである。例えば、心筋梗塞に注目した試験において、死亡した被験者に心筋梗塞が起きていたか、又は起きたであろうかを調べることは必ずしも可能ではないが、変数を死亡又は心筋梗塞を複合したものと定義すれば、この変数は完全に特定できるかもしれない。」

死亡が「死亡前の観測値よりもはるかに意味がある」場合は死亡自体が評価変数であり中間事象ではないと考えます。このため、この段落は単に複合エンドポイントについて述べたものとなっています。

8. 「A.3.3.1. 一般的留意事項」、296-297行

「どのような臨床試験における estimand の構成も、臨床試験の計画及び実施に関与する臨床家、統計家及びその他の専門分野の者も含む他分野にわたる作業となる。」

本補遺中のこの記述は非常に重要だと考えます。しかし、estimand に関する記載が本補遺内のみでは、この点が統計家以外の臨床試験関係者には伝わらないのではないかと危惧します。少なくとも「estimand」という用語と考え方があることは、ICH E6ないしはE8にも記載がないと、統計家以外の臨床試験関係者にその重要性が伝わらないと思いますので、日本から世界に向けてE6、E8へのestimandの記載を提案していただければと思います。また、統計家以外の臨床試験関係者に正しく理解してもらうためにも、「estimand」の適切な日本語訳は必要だと思いますので、検討をお願いいたします。

9. 「A.3.3.1. 一般的留意事項」、312-314 行

「適切な試験デザインの策定や、特定の estimand に対する信頼できる推定値の導出に重大な問題が存在する場合には、別の estimand、試験デザイン、及び解析手法について検討する必要があるだろう。」

別な estimand を検討するのでは最初に設計した試験本来の目的が達成できないことになり、「別の estimand、試験デザイン、及び解析手法」を用いた臨床試験が承認申請のための重要な試験となるとは考えにくいと思います。

コメント 3 でも述べたように、「特定の estimand」に対する偏りのある推定量が「別の estimand」等よりも意味がある場合も考えられますが、本補遺ではそのような記載が欠如しているように思われます。

10. 「A.3.3.2. 治療背景及び試験の実験的側面に関する留意事項」、355-356 行

「このように連続量のスコアを二分化することは、結果として estimand を変えることになる。」

本補遺では随所に estimand がストラテジーで決まるという記述がみられます。本補遺の統計的原則を考えると、「連続量のスコアを二分化して estimand を変える」ことが許容される場合がありうるのか、理解できませんでした。この点はコメント 6 で述べた複合ストラテジーの大きな問題を端的に表しています。どういう estimand になるのかはわかりませんが、「そういう estimand を推定することが試験本来の目的であるから、連続量のスコアを二分化する」であるべきであり、それはもはや中間事象の取り扱いに関する「複合ストラテジー」ではないと思います。

11. 「A.4. 試験デザイン及び実施への影響」、440-443 行

「規制当局と治験依頼者との間の協議では、提案された導入期が対象集団を特定するのに適切であるかどうか、またそれに続く試験デザインの選択（例えば、ウォッシュアウト期間、ランダム化）が対象とする治療効果の推定とそれに関連する推測の裏付けとなるかどうかを検討する必要がある。」

なぜ「ランダム化」が選択の対象となるのでしょうか？

12. 「A.4. 試験デザイン及び実施への影響」、465-469 行

「複数の目的や評価項目のある試験は、検定の多重性の問題を生じさせる可能性があるが、原則としてこの問題は複数の estimand がある場合にも同様に適用される。特に、一つ以上のものを主要とし、他のものを副次とする場合におけるこれらの問題に対処するために用いられる手法は、estimand に対しても同様に適応することができる。」

ここでいっているのは、複数の estimand がある場合、多重性の調整が必要、ということなのでしょう。具体的にどのように estimand の多重性を調整するのか理解できませんでしたし、463-465 行に述べられているように、他の estimand に対して試験デザインが適切でないので

あれば、単に統計的な多重性だけの問題ではなくバイアスの問題となるように思われます。

13. 「A.5.3. 補足的解析」、596-597 行

「したがって、このような解析の役割は、治験実施計画書の違反や逸脱の程度が試験結果の信用度を損なわせていないかどうかを評価することに限定される。」

中間事象が存在する場合の解析方法のひとつとして操作変数法がありますが、ランダム化臨床試験では治療のランダム化が操作変数となり、治療効果の上限と下限を求めるところができます。(たとえば、Robins and Greenland, JASA 1996; 91: 456-458, Greenland, International Journal of Epidemiology 2000; 29: 722-729)

さらに、FAS (ITT) にもとづく解析と PPS にもとづく解析を治療効果の上限と下限として解釈できる仮定もあるので、FAS (ITT) の解析と PPS の解析を感度解析のひとつととらえることが可能であり、この記述は強すぎると思います。Chiba, Sato, and Greenland, Statistics in Medicine 2007; 26: 5125-5135 を参照してください。