

何があっても割り付けどおりに解析する

佐藤俊哉¹, 松岡 淨²

1 京都大学大学院医学研究科社会健康医学系専攻医療統計学

2 藤沢薬品工業株式会社

(2001年9月15日日本行動計量学会にて講演)

佐藤: みなさん、おはようございます。京都大学医療統計の佐藤です。

松岡: おはようございます。藤沢薬品の松岡です。

佐藤: 今日は、松岡さんと一緒に、医学における典型的な実験データである、「臨床試験」について、わたしたちがふだんどんなことを考えているのか、どのようなアプローチをとっているのかをお話したいと思います。

今日のおはなし

- ◎ 臨床試験とガイドライン
- ◎ Intention-to-treat の原理
- ◎ 欠測データについて
- ◎ まとめ

2

佐藤: 今日のおはなしの構成ですが、まず最初に、松岡さんから、臨床試験っていったいどんなことをしているのか、その概略を説明してもらいます。

松岡: そんな大それたお話しはできませんが、佐藤先生よりはだいぶ臨床試験の経験がありますので、かいつまんで説明させていただきます。

佐藤: よろしくお願ひします。

それから、臨床試験データで非常にひんぱんに起きる2つの問題についてお話します。

ひとつは intention-to-treat と書きましたが、この「intention-to-treat」という用語を聞いたことのある方はどのくらいいらっしゃいますか、ちょっと手を挙げてみてください。

Intention-to-treat というのは、医療統計、生物統計を専門とされている方でも、きちんと理解している人は多くないのではないかと思います。どんな考え方なのかは後のお楽し

みとして、最後に欠測データのことをお話したいと思います。

それでは、松岡さんに臨床試験の概略を説明していただきましょう。

臨床試験

- ◎ 人に対する実験
 - ↳ 肉体的侵襲あり
- ◎ 科学的であると同時に
- ◎ 倫理的であることが要求される
- ◎ 様々なガイドライン、規制、法律

3

松岡：臨床試験がほかの分野、工業ですとか農学ですとかの実験と一番異なっている点は、人に対する実験だということだと思います。

もちろん、心理学ですとか、行動科学領域の実験でも人間を対象といたしますけれども、臨床試験ではこれに肉体的な侵襲が加わるわけです。

ですから、科学的な臨床試験を実施しなければならないのはもちろんなのですが、それと同時に試験に参加していただく患者さんの人権やプライバシーに配慮した、倫理的な実験であることも要求されるわけです。

このため、臨床試験を実施する際に守らなければならない様々なガイドラインですとか、規制、法律が定められています。

今日お話しするのは、臨床試験の中でも「ランダム化臨床試験」という、最後の段階で実施する、もっとも証拠としての質が高い試験ですが、その前に新しい薬や治療法を開発する場合に必要とされる実験の流れについて簡単に説明いたします。

臨床試験

- ◎ 非臨床試験
 - ➔ 細胞、組織、動物を使った実験（一般毒性、変異原性、催奇形性、がん原性）
- ◎ (第I相)試験
 - ➔ 健常成人(薬物動態、最大許容量...)
- ◎ (第II相)試験
 - ➔ 患者(有効性、安全性、用量－反応)

4

松岡：新しい薬を開発するためには、最終的に患者さんで実験を行って、有効性と安全性を確認しなければならないことはいうまでもありません。ですけれども、まだ海のものとも山のものともつかない「薬の候補」をいきなり人間に使うわけにはいきませんから、長い間に出来上がってきた手順があります。

最初は、「非臨床試験」と呼ばれる、細胞や組織、動物などを使った実験を行って、安全性に問題がないかどうかを詳しく調べます。この段階で、人間に使っても安全性に問題がないと確認されて、はじめて臨床試験に進むわけです。

臨床試験に進んでも、いきなり体の弱った患者さんに「薬の候補」を使うわけにはいきません。まず、健康な成人で体の中でどのくらい吸収され、分解され、どのくらいの時間で体外に出て行くのか、どのくらい使ったら有害な作用が出るのか、といったことを確認します。この試験は「第I相試験」とよばれています。

第I相試験が終了して、はじめて実際の患者さんに薬の候補を使うこととなります。この段階では、どのくらいの効果がみられるか、安全性に大きな問題はないか、使う量を増やすと効果も大きくなる、いわゆる用量－反応関係がみられるかどうか、などを確認いたします。

こうしてようやく...

ランダム化臨床試験

- ◎ (第III相) 試験
 - 治療をランダムに割り付ける
 - プラセボの使用
 - 二重マスク化
 - 検証的試験
- ◎ 市販後(第IV相)臨床試験

5

松岡: 新しい薬を開発するための最終段階である、第 III 相試験に進みます。

第 III 相試験は、「薬の候補」の有効性を科学的に検証する、という目的で実施します。このため、数百名から場合によっては数千名の患者さんを対象とします。

有効性を科学的に検証するために、通常は患者さんに「薬の候補」とコントロール治療をランダムに割り付けます。日本では標準的に使われている薬をコントロールとして使うことが多かったのですが、欧米では「プラセボ」とよばれる、有効成分を抜いてしまった「薬もどき」がコントロールとしてよく使われています。

プラセボを使うことで、治療をしているお医者さんも、治療を受けている患者さんも、どっちの治療を受けているのかわからない、という「二重マスク」あるいは「二重目かくし」の状態にすることができます。こうすることで、できるだけ心理的な影響が評価に入らないようにすることができるわけです。

実際には、臨床試験は第 III 相で終わるわけではありません。第 III 相試験までは、「薬の候補」を「新しい薬」として世の中に出していかどうかを調べるだけですから、今度は世の中に出た「新しい薬」がほかの薬と比べてどんな特徴を持っているのか、ということをもっとたくさんの患者さんで調べる市販後臨床試験というものが行われます。

以上が、臨床試験の大まかなまとめです。

倫理的な要求

- ◎ ヘルシンキ宣言
- ◎ 日米欧による International Conference on Harmonization の活動 (ICH)
 - 医薬品の臨床試験の実施の基準に関する省令 (GCP)
 - 臨床試験のための統計的原則について (E9) (統計ガイドライン)
 - 臨床試験における対照群の選択とそれに関連する諸問題 (E10)

6

佐藤: 松岡さん、どうもありがとうございます。

いま松岡さんに説明していただいたように、医薬品の開発では非常に厳密な手順を踏んで、人体への安全性をワンステップずつ確認しながら、慎重に試験を進めていきます。

このような手順が確立するまでには、ナチスによる人体実験にはじまる、研究に名を借りた、たくさんの非人道的な行為がありました。

そういった悲惨な経験から、世界医師会は人に対する実験を行う際に守るべき原則をまとめた「ヘルシンキ宣言」を公表しています。このヘルシンキ宣言は、国際統計協会やアメリカ統計協会の倫理ガイドラインにも、医学領域で働く統計家が守るべきガイドラインとして挙げられています。

松岡: あのを、ヘルシンキ宣言というのは、いってみれば大原則をうたった憲法のようなものなんでしょう。でも、それだけでは実際に臨床試験を実施するときに、具体的にどうすれば患者さんの権利を守って、科学的な臨床試験を行えるか、ということには答えてくれないわけです。

最近、日本とアメリカとヨーロッパの間で、いい薬を早く患者さんの元に、を合い言葉に、新しい薬の審査を迅速に行う国際協調のための会議が開かれているんです。そこでは、共通の規準で臨床試験を実施できるようにするため、たくさんのガイドラインが公表されています。

佐藤: 日本でもアメリカでヨーロッパでも、共通のガイドラインにしたがって臨床試験を行っているんですね。

松岡: そうなんです。わたくしどもは通称 ICH といっていますが、インフォームド・コンセントなどの倫理的な配慮や、試験実施計画書を作るといった基本的なことがらが書かれている GCP と呼ばれるガイドラインや、統計に関するガイドライン、対照群に関するガイドラインなどたくさんものがあります。

ガイドライン

- ◎ ヘルシンキ宣言
 - ➔ 日本医師会ホームページ
 - ➔ www.med.or.jp/wma/index.html
- ◎ ICHガイドライン
 - ➔ 医薬品食品衛生研究所ホームページ
 - ➔ www.nihs.go.jp/dig/ich/ichindex.htm

7

松岡: それで、いま紹介したガイドラインなんですが、もちろん出版されているものもあるんですが、それは少々お高いので、こちらのホームページからダウンロードされたほうが、いいと思います。

佐藤: きめ細かいご配慮、どうもありがとうございます。

松岡さんに、臨床試験に関する説明をしていただきました。

それではいよいよ本題の、臨床試験から得られるデータにはどういった特徴があるか、その特徴をどのように考慮して、解析をしなければならないか、についてお話ししたいと思います。

臨床試験から得られるデータ

- ◎ 試験実施計画書の作成
 - ➔ 実験であるからには当然だし、結果の信頼性のレベルは計画の信頼性のレベルで保証される(統計ガイドライン、1.2節)
- ◎ 試験実施計画書に沿った試験の実施
 - ➔ 計画書にどれだけ沿っているか
 - ➔ 違反・逸脱が多い試験は科学的、倫理的に問題

8

佐藤: みなさんの中で物理、化学、生物をとわず、「実験」をしたことのある方はどのくらいいらっしゃいますか。

それでは実験をされたことのある方のなかで、正式に文書にしたかどうかは別にして、研究ノートのような形でも、実験を行う前に計画を立てた方はどのくらいいらっしゃいますか。

か。

ほとんどの方が計画を立てていますよね。さきほど、松岡さんから、臨床試験をほかの分野の実験と分けているのは、「人に対する実験である」という点だ、というお話がありました。しかし、人に対する実験であろうが、人以外のものに対する実験であろうが、科学的な「実験」である以上、きちんとした計画を作る必要があります。

松岡: それは佐藤先生のおっしゃる通りで、ICH 統計ガイドラインにも、結果が信頼できるかどうかは、最終的に得られた結果がいい結果だったかどうかではなくて、そもそもの計画が信頼できるかどうかにかかっている、という記述があります。

佐藤: ということはつまり、試験計画書に定めたことを守らなかった、あるいは守れなかった試験は、科学的にも、また倫理的にも問題となりますから、試験から得られた結果の信頼性を増すためには、最初に立てた計画は、どんなことがあっても守らなければならない、ということが要求されるわけですね。

松岡: あの一、こういうところで佐藤先生に反対の意見をのべるのはたいへん勇気のいることなのですが、実際に臨床試験をやっておきますと、試験を倫理的に行うとすればするほど、当初の計画を守れない、ということが現実起こってくるんです。

計画書に沿った実施

- ◎ 計画書は必ず守らなければならない!
- ◎ 有害作用で苦しんでいる患者さんに、同じ治療を続けられますか?
- ◎ 倫理的に問題
- ◎ 倫理的な問題が生じる場合に限っては、計画書からの逸脱が正当化(GCP第46条)
 - ➡薬の使用を止める

9

佐藤: だけど松岡さん、試験計画というものは、きちんと計画を守って、はじめて科学的な評価ができるように設計するわけですね。

そうしたら、計画を守らなかった、たとえば一日 3 回薬の候補を飲んでもらう、という計画なのに一日 1 回しか飲まなかった患者さんがいたら、有効性の評価ができなくなってしまっ、困るじゃないですか。

松岡: 先生はそうおっしゃいますが、薬の候補を一日 3 回飲まなかったといっても、ついうっかり忘れてしまった人もいれば、吐き気とか頭痛とか有害な作用が出てしまって飲めなくなってしまった方もいるわけです。

現実には吐き気や頭痛で苦しんでいる患者さんに、「計画通り薬の候補を飲んでくださ

い」ということは、倫理的に問題だと思うんです。

実際に、臨床試験の倫理的な問題をカバーしている GCP には、試験に参加している患者さんに倫理的な問題が生じる場合に限っては、計画書を守れなくても仕方がない、ということが書かれています。

これこそが人に対する実験が守るべき大原則だと思います。

臨床試験のデータ解析

- ◎ 薬の候補を使う、使わない、をランダムに割り付け
 - ◎ 倫理的な理由から
 - ➔ 薬の候補の使用を止める
 - ➔ 薬の候補を使う
- ことがある

どうやって解析したらいいんですか？

10

佐藤: なるほど、臨床試験は人に対する実験だから、科学性よりも倫理性が優先する、というわけですか。

確かに、松岡さんがいうように、「わたしたちは倫理的な配慮を十分に行って臨床試験をしています」、というのは、一般受けはいいと思います。

でも現実問題として、わたしたちは、科学的に妥当な結果を出すために、臨床試験という実験の場でしか行えない、ランダム化という非常に強い操作を行っているわけです。

それなのに、計画を守れなかった参加者が出てきた場合、たとえば薬の候補を飲むはずだった人が、飲まなくなってしまうたり、その逆にプラセボを飲むはずだった人が薬の候補を飲んでしまった場合、その人たちは解析ではどう扱ったらいいんですか。

松岡: 先生、それこそわたしが先生にうかがいたかったことなんですよ。

臨床試験データ

	実際に使ったのは A	実際に使ったのは B
割り付けは A	Aa	Ab
割り付けは B	Ba	Bb

11

佐藤: みなさんも一緒に考えてみてください。

臨床試験では、試験に参加していただいた患者さんを、ランダムに治療 A、または治療 B に割り付けます。

しかし、実際には、治療 A を割り付けられて、その通りに治療 A を使った人もいれば、さまざまな理由から実際には治療 B を使ってしまった。また、治療 B を割り付けられて、その通りに治療 B を使いました、あるいは割り付けとはことなって治療 A を使ってしまった、なんていう人が出てきてしまうわけです。

割り付けられた治療を大文字の A と B、実際に使った治療を小文字の a と b で書きましょう。

さて、松岡さん、こんなデータとなった場合、一体どのグループとどのグループを比較して解析したらいいんでしょう？

松岡: それは先生、臨床試験は実験ですから、先生もおっしゃいましたように、試験計画書を守ったかどうかが一番重要なわけです。

ですから、試験計画書に書いてあるように、割り付けをきちんと守った、Aa の人たちと Bb の人たちを比較すればいいに決まっているじゃありませんか。

誰と誰を比較

- ◎ 計画書を守った人たち
 - ➡ Aa vs Bb

12

佐藤: なるほど、計画書通りに、割り付けられた治療を正しく守った人たちだけを解析すればいい、というわけですか。

みなさんはどうですか。正しく計画書を守った Aa と Bb のグループを比較して解析すればいい、と思われる方、手を上げてください。

松岡さん、どうでしょう。

松岡: まことに申し上げにくいんですが、今手を上げた人たちは残念ながら間違っています。計画書通りに割り付けられた治療を守った人たちだけを解析したのでは、割り付けを守らなかった Ab と Ba の人たちの貴重なデータを無駄にってしまうじゃないですか。

せっかく、臨床試験に協力してくださった患者さんのデータを無駄にするなんて、どうかしています。

佐藤: えっ、それじゃあどのグループとどのグループを比較したらいいんですか？

臨床試験データ

	実際に使ったのは A	実際に使ったのは B
割り付けは A	Aa	Ab
割り付けは B	Ba	Bb

13

松岡: 臨床試験に参加してくださった患者さんのデータを無駄にするなんて、それこそ非倫理的な解析だと思います。

薬や治療というものは、実際にその薬や治療受けて、はじめて効果があるかないかをうんぬんできるものですから、実際に治療 A を使った Aa と Ba を足したグループと、実際に治療 B を使った Ab と Bb を足したグループを比較すべきだと思います。

佐藤: えーっと、先ほどとだいぶ違ったご意見ですが、みなさんどうでしょうか?

誰と誰を比較

- ◎ 計画書を守った人たち
 - ➡ Aa vs Bb
- ◎ 実際に薬を使ったか、使わなかったか
 - ➡ Aa+Ba vs Ab+Bb

14

佐藤: 今の松岡さんの説明どおり、実際に使った治療にもとづいて、Aa と Ba を足した治療 A を使ったグループと、Ab と Bb を足した治療 B を使ったグループを比較すればいいと思われる方、手を上げてください。

さて、松岡さん、今度はどうでしょう。

松岡: まことに申し上げにくいんですが、今手を上げた人たちは、残念ながら間違ってます。

佐藤: えーっ、またですか。そしたら一体、どのグループとどのグループを比較したらいいんですか?

松岡: みなさん意外に思われるかもしれませんが、...

臨床試験データ

	実際に使ったのは A	実際に使ったのは B
割り付けは A	Aa	Ab
割り付けは B	Ba	Bb

15

松岡: みなさん、「計画書を守った」ということを誤解されているんじゃないかと思います。

計画書に書いてあることは、「A と B をランダムに割り付ける」ということで、治療が割り付けられた後に、割り付けを守るかどうかは倫理的な理由から「守れないこともある」ことを許しているわけです。

ですから、「計画書通りに解析する」ということは、割り付けを正しく守った人たちだけを解析することでもなければ、実際に受けた治療にもとづいて解析することでもなく、そもそものランダムに割り付けられた治療である、Aa と Ab を足したグループと、Ba と Bb を足したグループを比較すべきなんです。

佐藤: なんだかすごいことになってきましたが、…。計画書を守った人たちの比較でもなく、実際に受けた治療にもとづいた比較でもなく、ランダムに割り付けた治療にもとづいて比較しろ、というんですか。

さて、みなさんはどう思われるでしょうか。松岡さんのいうように、割り付け通りに比較して解析すべきだと思われる方、手を上げてください。

誰と誰を比較

- ① 計画書を守った人たち
 - Aa vs Bb
- ② 実際に薬を使ったか、使わなかったか
 - Aa+Ba vs Ab+Bb
- ③ 実際の使用は無視、割り付け通りに比較
 - A vs B

16

佐藤: それでは、正解をみてみましょう、正解は、「割り付け通りに比較する」でした。驚きましたね。

松岡: あの一、佐藤先生。今まで台本どおりに話してきたんですが、実はだんだん不安になってきたんです。

それというのも、実際に受けた治療がなんであったか、ということ無視してしまって、もともとの割り付け通りに解析するのが正解だとおっしゃるんですが、ほんとうにそれでいいのでしょうか。

佐藤: さてみなさん、今日のタイトルを思い出してみてください。

「何があっても割り付け通りに解析する」でしたよね。この「割り付け通りに解析する」ことを、...

Intention-to-treat の原理

- ① なぜ割り付け通りに比較するんですか?
- ② 割り付けを守らなかった(守れなかった)のには理由がある
 - ランダムではない
- ③ 実際の使用にもとづいた比較にはバイアスが...(臨床試験のための統計的原則、5.2.1節、用語集)

17

佐藤: わたしたちは intention-to-treat の原理とよんでいます。

なぜこんな、奇妙な解析がまかり通っているのかというと....

松岡さんにかがいたいんですが、臨床試験で割り付けられた治療を守れなかった人、
というのはどういう人たちでしょう。

松岡: それはですね、いろいろな理由があると思いますが、典型的なのは「治療が効かないから止めた」という理由や、その反対に「治療が効いて良くなったから止めた」ということだと思います。

佐藤: そうすると、割り付けられた治療を守れなくなってしまう人というのは、決して偶然守れなくなってしまうのではなく、治療の有効性に関係した理由があって、割り付けられた治療を続けられない、ということですね。

松岡: もちろん、うっかり飲み忘れたとか、交通事故にあって入院してしまって治療を続けられなくなった、とか、治療の有効性とは関係のない理由で割り付けを守れないこともあると思いますが、多くの場合は治療が効かなかったから止めた、または効いたから止めた、という特別な理由があって止めた人たちだと思います。

佐藤: そうすると、そういう人たちを解析から省いてしまったり、治療が効いたとか効かなかったとかという理由によって、治療を変更した人たちを最終的に受けた治療にもとづいて比較したらおかしい結果になってしまいますよね。

割り付け通りに比較したって...

- ◎ 治療を受けなかったのに「受けた」、治療Bを受けたのに「治療Aグループ」として解析するのはおかしいじゃないですか！
- ◎ 帰無仮説「治療を受けても、受けなくても効果は同じ」のもとでは...
- ◎ 治療を受けても受けなくても、受けて途中で止めても、途中から受けたとしても、効果は一緒(効かない)

18

松岡: でも先生、実際には治療を受けていないのに、割り付けでは治療を受けるグループだったから「治療を受けた」と思って解析したり、治療Aを割り付けられたのに、治療Bを受けてしまって、それでも「治療Aグループである」と解析するのは、どうも納得がいかないんですが...

佐藤: みなさんはどうでしょう。松岡さんと同じように、何がなんでも割り付けどおりに解析する、というのは気持ち悪いと思われる方、手を上げてください。

みなさんの気持ちの悪さというのは、「効果が出るのは実際に治療を受けたからであって、割り付けられただけで効果が出るわけではない。それなのに、実際に受けた治療ではなく、割り付けられた治療にもとづいて比較するのはへんだ」ということでしょうか？

どうもみなさん、「薬の候補は有効だ」という場合しか考えていないようですが、もしほんとうは「薬の候補は効かなかった」、としたらどうでしょうか。

薬の候補に効果がない場合には、その薬の候補を飲んでも、飲まなくても、効果がないことには変わりありませんし、飲んでいた薬の候補を途中でやめてしまっても、反対に途中から薬の候補を飲んだとしても、どっちにしろそれは効かないんですから、一緒ですよ。

Intention-to-treat

- ◎ 対立仮説のもとでは
 - ➔ 実際の使用: 治療をよく見せがち
 - ➔ 割り付け通り: 治療の効果を薄める
- ◎ 帰無仮説のもとでは
 - ➔ 実際の使用: 治療をよく見せがち
 - ➔ 割り付け通り: 正しく「効果なし」

19

佐藤: つまり、「プラセボにくらべて治療効果がある」という仮説が正しい場合には、実際に受けた治療にもとづいた比較や、計画書を守った人たちだけの比較では、多くの場合治療効果を過大評価しがちとなります。

一方、割り付け通りに intention-to-treat の解析を行うと、今度は治療をした人やしなかった人が混ざったグループを、「治療を割り付けたグループ」として解析しますから、薄まった治療効果をみていることとなります。

松岡: 対立仮説のもとではどんな解析を行ってもバイアスが入ってしまうわけですね。

佐藤: そうなんです。ですけれども、「治療効果がない」という仮説が正しい場合、実際に受けた治療にもとづいた比較や、計画書を守った人たちだけの比較では、やっぱり治療効果をよくみせてしまうのに対して、intention-to-treat にもとづく比較では、正しく「治療効果がない」といえるわけです。

松岡: 帰無仮説のもとでは intention-to-treat にもとづく解析では名義水準通りタイプ I エラーを守れるけれども、そうでない解析ではタイプ I エラーが守れないんですね。

そうしますと治療効果の推定にはいくつか選択の余地はあるとしても、帰無仮説の検定は、いつも intention-to-treat にもとづいて行わなければならないんですね。

Intention-to-treat の要求

- ◎ ランダム化された参加者
 - ➡ すべてを解析に用いる
 - ➡ すべてを漏れなく追跡
 - ➡ すべてについて結果を記録

20

佐藤: 今、松岡さんにまとめていただいたように、「治療効果はない」という帰無仮説の検定は、少なくとも intention-to-treat にもとづいて行う必要があります。

Intention-to-treat 解析ではランダム化された試験参加者を全員解析するので、当然といえば当然なのですが、intention-to-treat で解析しようと思ったら、ランダム化された試験参加者全員を、割り付けられた治療を守ったか守らないかに関らず、漏れなく追跡して、漏れなく必要な測定を行わなければならないんです。

一見たいへんなように聞こえますが、臨床試験は人に対する実験ですから、一度試験に参加した患者さんは、途中の経緯がどうであれ、試験の最後まで責任を持って面倒を見るのは当然のことだと思います。

Intention-to-treat の意味

- ◎ 帰無仮説のもとでは
 - ➡ 常に妥当な α レベル検定
- ◎ 対立仮説のもとでは
 - ➡ 薄まった治療効果を推定しているが...
 - ➡ 実際的な治療効果

21

佐藤: こうしてランダム化された試験参加者を全員解析することによって、帰無仮説のもとではタイプ I エラーを守った、妥当なアルファレベル検定をすることができます。

また、さきほどは「治療効果の推定にはいくつか選択の余地がある」といいましたが、

実は治療効果の推定にも intention-to-treat 解析を行うことが多くなっています。

というのも、intention-to-treat 解析では薄まった治療効果をみていることになるのですが、その治療法や薬の候補が実際に世の中に出て、日常診療で使われるようになったとき、仮に臨床試験のときにその治療が効かなくて止めてしまった人が診察に来て、日常診療の場でその治療を使われたとしたら、松岡さん、どうなると思いますか。

松岡: それはやはり、その人には効かないでしょうから、その治療は止める、ということになるのではありませんか。

佐藤: そうですね。たぶん臨床試験で起こったことは日常診療の場でも起こるであろうと考えるわけです。

このため、intention-to-treat による治療効果は薄まった治療効果ではあるんですが、その治療を実施しようとした患者さんに対して、治療が続けられなくなったということもマイナスの治療効果だと判断しよう、というのが治療効果の推定にも intention-to-treat を使う理由となっています。

Intention-to-treat の要求

- ◎ ランダム化された参加者
 - ➔ すべてを解析に用いる
 - ➔ すべてを漏れなく追跡
 - ➔ すべてについて結果を記録

そんなの無理じゃないですか

22

松岡: 先生のお話をうかがっていると、intention-to-treat というのはいいことばかりのようなんですけど、実は実際の臨床試験をやっているとそうはうまくいかないじゃないかと思うんです。

といいますのも intention-to-treat 解析を行うためには、ランダム化された試験参加者をすべて解析する、全員を漏れなく追跡する、全員の結果を測定する、ということだったんですが、現実にはこんなこと無理だし意味がないと思うんです。

ランダム化された全員を解析

- ⊙ 間違って登録されてしまった人
 - ⊙ 薬を間違って投与された人
 - ⊙ 一回も薬を飲んでいない人
 - ⊙ まったくデータがない人(亡くなられた人)
- も解析するんですか？

統計的には正しくても、医学的に
解釈できるんですか？

23

松岡：たとえば、臨床試験をいくら厳密に行っても、なんらかのミスで本来はその試験には参加するはずではなかった人が登録されてしまうことがあります。いくら注意していても、わたしたち人間が試験を運営している以上仕方のないことなんですが、そんな「間違って登録されてしまった患者さん」まで解析しなければならないのでしょうか。

また、登録センターから割り付けを連絡するときに、本当は A という割り付けでしたのに、間違って B と連絡してしまった、なんてこともまれにあるわけです。こんな場合でも「割り付けが A だったんだから、A として解析する」ということでいいのでしょうか。

割り付けられた後に一度も薬を飲まなかった人や、お亡くなりになってしまって割り付け後のデータが全然ない人もおりますし、こういった人たちを無理矢理 intention-to-treat だといって解析してしまうことは、統計的には正しいかもしれませんが、それで本当に医学的に正しく解釈できるのかどうか心配なんです。

最大の解析対象集団

- ⊙ 臨床試験のための統計的原則
 - ➡ 間違って登録されてしまった人
 - ➡ 一回も薬を飲んでいない人
 - ➡ まったくデータがない人
- は解析から除いてもよい

24

松岡: 実は厚生労働省が出している ICH 統計ガイドラインには、ランダム化した参加者全員を解析しなさいといっても、間違っで登録されてしまった人や、割り付けの後一度も薬を飲んでいないような人、さらに、割り付けの後一度もデータが取れなかった人などは仕方がないので、解析から省いてもいい場合がありますよ、と書いてあります。

この考え方は、解析できるせいっぱいの人を解析しようという考え方で、「最大の解析対象集団」とよばれています。

Intention-to-treat の要求

- ◎ ランダム化された参加者
 - ➔ すべてを解析に用いる
 - ➔ すべてを漏れなく追跡
 - ➔ すべてについて結果を記録

病院に来られなくなってしまった
参加者はしかたないじゃないですか

25

松岡: 今お話ししたのは、おもに試験の最初の段階でのことですが、試験の途中でだっでいろいろな理由で病院に来られなくなってしまう患者さんはたくさんいます。そういう人たちは重要な結果の測定ができなくなってしまうわけなんです、...

欠測、追跡不能

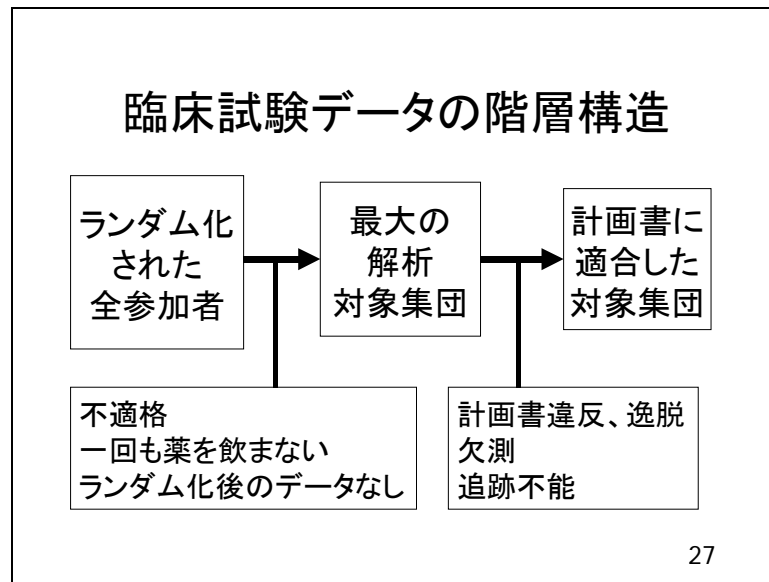
- ◎ 人に対する実験である以上
 - ➔ 忙しくて病院に来られなかった
 - ➔ 試験を続けるのがいやになった
 - ➔ 引っ越してしまったといった理由でデータがない!?

26

松岡: 臨床試験がまさしく人に対する実験であることから、仕事が忙しくてそれ以上病院にこられなくなってしまう方もいると思いますし、試験を続ける意義が疑問となっで、もう協力したくないという方だっでいらっしやると思います。

そういった人たちに、これは実験なんだからなんとしてでも病院に来て、結果を測定しなさいと命令することはできません。

そんなことしたら非倫理的な試験になりかねないと思います。



佐藤: 松岡さんの説明をまとめるとこういうことになります。まずわたしたちは、理想として「ランダム化された試験参加者全員」のデータを解析することが要求されているわけですが、もともと試験に入るべきではなかった人、適格ではなかった人が誤って登録されてしまった場合ですね、それから割り付けられた後一度も薬を飲まなかった人、割り付けの後のデータがまったくない人、こういう人たちは「治療効果の評価」という観点から、ほとんど情報を持っていないはずですから、解析から除外してもいい場合もあると思います。

ICH 統計ガイドラインでは、こういった人たちを除いた対象集団を、「最大の解析対象集団」と呼んでいます。

さらに、最大の解析対象集団から、計画書に違反した人、データが途中で欠測となった人、追跡できなくなってしまった人を除いた集団を、「試験計画書に適合した解析対象集団」と呼んでいます。

どの集団を解析すべきか？

- ① ランダム化された全参加者
 - ➔ ITTが要求する集団 ← 欠測の問題は
- ② 最大の解析対象集団 ← どうするんですか？
 - ➔ 統計ガイドラインが許容する集団
- ③ 計画書に適合した集団
 - ➔ バイアスのかかった集団

28

佐藤: 「ランダム化された試験参加者全員」は intention-to-treat が要求する対象集団ですし、「最大の解析対象集団」は統計ガイドラインが許容している対象集団です。

最後の「試験計画書に適合した解析対象集団」はたぶん治療効果をよくみせる方向にバイアスが入っているはずですが、重要なことは、解析から除外しなければならない人をできる限り少なくして、「ランダム化された試験参加者全員」と「最大の解析対象集団」と「試験計画書に適合した解析対象集団」ができるだけ同じになるような努力をし、どの解析対象集団の結果も変わらないような臨床試験を行うです。こういう努力をすることが、質の高い臨床試験を実施することにつながるんだと思います。

松岡: 先生、すべての患者さんについてデータがちゃんとそろっているのは「試験計画書に適合した解析対象集団」の人たちだけですよね。

真ん中にある「最大の解析対象集団」を解析する場合でも、欠測値をどう扱うかという問題は避けて通れないと思うのですが...

解決しなければならないのは

- ① Intention-to-treat の原則から参加者を完全に追跡することが要請されるのは理解できませんし、ランダム化がその鍵とは思いますが...
 - ➔ 臨床試験では欠測あるいは脱落は必ず起きるので大きな問題となるのでは？
 - ➔ 統計解析における欠測データへの対処の仕方はどうするんですか？

29

佐藤: できれば避けて通ろうと思っていたのですが...

松岡: そうはいきませんよ、最初に欠測の問題もお話しすると約束してしまったんですから。

うかがいたいのは、さきほどもいいましたが、臨床試験では人間を対象としていますから欠測があつたり脱落が起きたり、ということはしょっちゅうあるわけです。それが解析にどういう影響をおよぼすのか、ということと、実際に解析を行う場合欠測データはどう扱ったいいのでしょうか。

欠測が問題となる場合

- ① 欠測が完全にランダム (Missing Completely at Random) でなく
- ② グループ間で欠測の割合が異なる
 - Ex. 死亡をみたい試験で、プラセボグループから高齢者が選択的に追跡不能
- ③ どちらかを満たさなければ、問題なし

30

佐藤: 欠測や追跡不能は、もちろん数が多ければ試験自体の質が疑われてしまうので問題なのですが、解析上は、欠測が起きたこと自体が問題となるわけではないんです。

欠測が問題となるのは、欠測がランダムなメカニズムで起きているのではなく、さらに、治療グループ間で欠測の割合がことなっている、この両方を満たす場合に限って、治療効果の推定にバイアスが入ります。

まあ臨床試験の場合はたいていなんらかの理由がありますから、欠測がランダムに起きるということは、ほとんどありえないのですが、スライドにしめましたように、薬剤が死亡を減らすかどうかを調べている臨床試験で、プラセボグループから高齢者が選択的に追跡不能となったとすると、追跡できた人だけを解析すると、プラセボグループの結果がだいぶ良くなってしまふのがわかります。

しかし、かりに高齢者に追跡不能となる人が多かったとしても、薬の候補のグループ、プラセボグループとで追跡不能となる割合が変わらなければ、欠測や追跡不能といってもそれほど問題とはなりません。

欠測があつた場合に、どういった扱いをするのかについては、経験豊富な松岡さんに教えていただきたいのですが...

欠測データの非統計的処理

- ◎ 最悪値を代入
 - ➡ グループ間の差が大きくなる場合あり
- ◎ 直前値を代入 (Last Observation Carried Forward, LOCF)
 - ➡ 分散を過小評価
- ◎ グループ間の差が縮まる値を代入
 - ➡ 過度に保守的

31

松岡: 欠測を作る経験は豊富なんですが...

欠測の扱いについては、古来からほんとにさまざまな考え方があると思います。

たとえば、欠測したところには最悪の値を入れておくとか、経時的に観察している場合には直前の値を入れるというのもよくやります。また、タイプ I エラーを守るという意味では、グループ間の治療効果の差が縮まるような値を入れる、ということだって考えられています。

ただ、それぞれのやり方には当然欠点もあります。最悪値を使ってもそれによって治療効果をよくみせる方向にバイアスが入ることだってありますし、直前の値を使うのは同じ値を使うことになりますから、分散を小さくすることになってタイプ I エラーが守れなくなることがあります。

また、治療効果がなくなるような値を入れるのは、過度に保守的になってしまい、実際に行うのには抵抗があります。

こういった方法とは別に、解析のときに欠測データを取り扱う方法にはどんなものがあるのでしょうか。

欠測データの統計的処理

- ④ Multiple Imputation (Rubin, JASA, 1996)
- ④ Inverse Probability of Censoring Weighting (Robins et al. JASA, 1995)
- ④ 欠測メカニズムが、共変量で説明できる「Missing at Random」ならOK
- ④ 欠測メカニズムが Nonignorable だったらどうするんですか？

32

佐藤: 欠測データ処理の一般的な方法論として、Rubin の multiple imputation 法が有名です。もともとは標本調査の際の欠測データ処理のために考え出された方法ですが、最近では臨床試験データへの応用もあるようです。

また、データが観察されるメカニズムをモデル化して、データが観察される確率を推定します。その観察確率の逆数でデータを重み付けて解析する、inverse probability of censoring weight 法というものが提案されています。

これは、松岡さんが観察される確率が4分の1だとすると、松岡さんと同じような人で、実際には観察されなかった人がもう3人いる、と考えて松岡さんのデータを4倍にふくらませて解析することになります。

松岡: Multiple imputation も観察確率の逆数で重み付けるのも、どちらもすばらしい方法だと思いますが、どちらの方法も欠測の起きるメカニズムが、測定されているデータで説明できる、missing at random と呼ばれているものでないとだめだと思うんです。

ところがわたしたちは、欠測が起こるほんとうのメカニズムなんてわかるわけがありませんから、missing at random のように欠測メカニズムが「無視できる」場合ばかりならいいんですが、「無視できない」nonignorable な欠測メカニズムだってありうると思うんです。

そういう場合にはどんな方法があるのでしょうか。

Nonignorable Missing

- ◎ 現在のところ決定版はなし
 - ➡ いかんともしがたい
- ◎ 高々できることは
 - ➡ Sensitivity Analysis
- ◎ どうやって?
 - ➡ またの機会に

33

佐藤: 松岡さんは答えられない質問をされるので、嫌いなんですけど、欠測メカニズムが無視できない場合には、今のところ解析の決定版はないといってもいいと思いますし、たぶん将来にわたっても決定版はでないんじゃないかと思います。

じゃあわたしたちは何もできないのか、というと、わたしたちにできることは高々 sensitivity analysis ではないかと思います。

Sensitivity analysis というのは、欠測メカニズムが無視できないのだとすると、無視できない程度というものを数値化して、無視できない程度の値を動かして解析すると、治療効果の推定結果がどのくらい変わるか、を調べるんですが、ちょっとややこしいので具体的なやり方はまたの機会に、ということにいたしましょう。

さて、そろそろ時間となりましたので、まとめたいと思います。松岡さん、まとめをよろしくお願いします。

まとめ

- ◎ 臨床試験では、治療法と結果との因果関係を調べることが目的
- ◎ そのためにはランダム化が必須の要素
- ◎ ランダム化(比較可能性の確保、確率計算の基礎)を守るためには、ランダム化された全参加者を解析することが必要
 - ➡ Intention-to-treat の原理

34

松岡: わたしがまとめるんですか...。ひどいなあ...

本日はわたしたちの講演を長々とお聞きいただき、ありがとうございました。今日、わたしたちが強調したかったことは何かと申しますと、医学における典型的な実験である臨床試験では、治療と結果との間の因果関係を調べるのが一番大きな目的であって、そのためにはランダム化を行うということが必須である、ということです。

このランダム化にもとづいて、今日お話ししました、intention-to-treat の原理などが導かれるわけです。

まとめ 2

- ④ ITTの原則を守るためにはランダム化された全参加者を徹底して追跡する必要あり
- ④ それでも人に対する実験である以上、欠測、追跡不能は必ず起こり、しかも欠測には理由がある
- ④ 欠測に対処する適切な方法と、Sensitivity Analysis が重要

35

松岡： このintention-to-treatの原理を守るためには、ランダム化された試験参加者全員を漏れなく追跡して、結果を測定し、全員を解析する必要があるのですが、現実の臨床試験でそんなとんでもないことはできないわけです。

そのため臨床試験のデータ解析では、欠測に対処する適切な統計的方法も必要となります。現在のところ、欠測に対処する統計的方法は、欠測のメカニズムが無視できる場合にしか適用できませんので、「無視できない欠測メカニズム」であった場合を想定した感度解析を行うのが重要なんだ、ということになります。

佐藤： 松岡さん、どうもありがとうございました。

松岡： どうもありがとうございました。