

臨床試験における検定の多重性 ～ガイドンス・ガイドラインと実際～

東京大学大学院医学系研究科 生物統計情報学講座
生物統計学シンポジウム

坂巻頭太郎
sakamaki@m.u-tokyo.ac.jp

統計的有意性とP値に関するASA声明

1. はじめに
近年の定量的な科学研究の増加と膨大で複雑なデータセットの激増は、統計的方法の応用範囲を拡張しつつある。このことは科学の進歩に新たな道を開いたが、一方でデータから導かれる結論に対する懸念をもたらした。科学的結論の妥当性が、その再現性も含めて依存しているのは、統計的方法だけにとどまらない。適切に選択された技術、適正になされた解析、および統計的結果の正しい解釈も、結論がしっかりしたものであり、かつ結論の不確実性が適正に述べられていることを確かなものにするために中心的な役割を果たしている。

公表された多くの科学的結論の土台となっているのは「統計的有意性」という概念であり、通常P値と呼ばれる指標で評価される。P値は有用な統計指標ではあるが、誤用と誤解がまかり

育、公共政策、ジャーナリズム、法律にも影響を与える。この声明は、健全な統計学の実践に関連したすべての問題を解決しようとするものでもなければ、土台となる論争を解決するためのものでもない。この声明では、統計コミュニティで広く認められたコンセンサスに基づいて、専門用語を使わずに、定量的な科学研究の実施や解釈を改善するべきの原則を述べる。

2. P値とは?
おおざっぱにいうと、P値とは特定の統計モデルのもとで、データの統計的要約(たとえば、2グループ比較での標本平均の差)が観察された値と等しいか、それよりも極端な値をとる確率である。

3. 原則

<http://www.biometrics.gr.jp/news/all/ASA.pdf> 1

判断における問題

- P値を用いる判断(仮説検定)で退けたいこと
 - 治療効果がない
 - 第1種の過誤の制御
 - データ数が少なくて判断できない
- 臨床的な判断は総合的な問題
 - 治療効果の大きさ、その他の治療方法の有無、治療費用などを用いて総合的に治療効果の評価は行われる

2

1つの試験から何を言う?

THE NEW ENGLAND JOURNAL OF MEDICINE

REVIEW ARTICLE

THE CHANGING FACE OF CLINICAL TRIALS
Jeffrey M. Drazen, M.D., David F. Hawthorn, Ph.D., John L. McNamee, M.D., James H. Ware, Ph.D., and Janet Woodcock, M.D., Editors

The Primary Outcome Is Positive — Is That Good Enough?

Stuart J. Pocock, Ph.D., and Gregg W. Stone, M.D.

THERE IS A NATURAL TENDENCY TO SIMPLIFY THE FINDINGS OF A CLINICAL trial into a binary conclusion: “Was there a positive outcome—or not?” In order to address this question with some objectivity, attention is typically focused on whether the prespecified measure of success for the primary outcome has been met—that is, whether a P value of less than 0.05 has been achieved for the difference in treatments. In reality, a more nuanced interpretation requires a thorough examination of the totality of the evidence, including secondary end points, safety issues, and the size and quality of the trial. In this article, which focuses on the evaluation of “positive” studies—as in our previous article¹ which focused on the appraisal of “negative” studies—our intent is to facilitate a more sophisticated and balanced interpretation of trial evidence. Again, we make our points using examples from trials involving cardiovascular disease (our area of expertise, but the messages can be easily applied to other subject areas.

THE NEW ENGLAND JOURNAL OF MEDICINE

REVIEW ARTICLE

THE CHANGING FACE OF CLINICAL TRIALS
Jeffrey M. Drazen, M.D., David F. Hawthorn, Ph.D., John L. McNamee, M.D., James H. Ware, Ph.D., and Janet Woodcock, M.D., Editors

The Primary Outcome Fails — What Next?

Stuart J. Pocock, Ph.D., and Gregg W. Stone, M.D.

A WELL-DESIGNED TRIAL DERIVES ITS CREDIBILITY FROM THE INCLUSION of a prespecified, a priori hypothesis that helps its authors avoid making a potentially false positive claim on the basis of an exploratory analysis of the data. Nevertheless, an unreasonable yet widespread practice is the labeling of all randomized trials as either positive or negative on the basis of whether the P value for the primary outcome is less than 0.05. This view is overly simplistic. P values should be interpreted as a continuum wherein the smaller the P value, the greater the strength of the evidence for a real treatment effect.^{1,2} Confidence intervals are also useful in indicating the range of uncertainty around the estimated treatment effect. Moreover, the interpretation of any trial should depend on the totality of the evidence (i.e., the primary, secondary, and safety outcomes), not just a single end point.

Pocock SJ, et al., *NEJM*, 2016 Pocock SJ, et al., *NEJM*, 2016 3

(“良い”という判断ではなく)
 Data dredgingをどう防ぐかという問題

4

検定の多重性の調整

- 第1種の過誤を制御することが目的
 - 規制
 - “効かない”(効く判断できない)薬剤を承認しない
 - 企業や研究者
 - Phase IIIの失敗などを抑える
 - 再現可能性(reproducibility)を担保する
- Data dredgingに対処するための1つの考え方
 - 客観的に(数値を用いて)示すことも重要

5

2017年のトピック


Multiple Endpoints in
 Clinical Trials
 Guidance for Industry
 DRAFT GUIDANCE

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the Federal Register of the notice announcing the availability of the draft guidance. Original comments are to be submitted to CDER.Docket@FDA.gov. Other comments, including those received in response to the public notice, will be posted to the public docket on www.fda.gov/oc/ohrt. All comments should be identified with the docket number listed in the notice of availability that publishes in the Federal Register.

For questions regarding this draft document contact: Scott Gardner at 301-796-3303 or CDER.Docket@FDA.gov, Office of Communications, Outreach, and Development, 301-401-4709 or 240-401-8022.

U.S. Department of Health and Human Services
 Food and Drug Administration
 Center for Drug Evaluation and Research (CDER)
 Center for Biologics Evaluation and Research (CBER)
 January 2017
 Draft of Guidance

FDA, 2017



1 13 December 2016
 2 13 December 2016
 3 Committee for Human Medicines (CHMP)

4 Guideline on multiplicity issues in clinical trials
 5 Draft

Draft agreed for Medicines Working Party (MWP)	November 2016
Approved by CHMP for release for consultation	13 December 2016
Start of public consultation	04 April 2017
End of consultation (deadline for comments)	30 September 2017

6 This guideline relates the "needs to consider on multiplicity issues in clinical trials"
 7 (commentary)

8 Comments should be provided using this [template](#). The completed comments form should be sent to: PublicAffairs@ema.europa.eu

9 Keywords: Multiplicity, Reproducibility, Type I error, subgrouping, responder, estimation, confidence interval.

EMA, 2017

6

講演内容

- 諸外国のガイダンス・ガイドライン
 - 小山暢之(第一三共株式会社)
- 多重性調整方法に関する最近の話題 (Graphical approachとMCP-mod)
 - 坂巻頭太郎(東京大学)
- 多重性調整に関する現在の実践
 - 吉田征太郎(中外製薬株式会社)
- 審査の経験を踏まえての事例紹介
 - 浅野淳一(PMDA)

7