

科学研究費助成事業 研究成果報告書

平成 26 年 4 月 17 日現在

機関番号：62603

研究種目：研究活動スタート支援

研究期間：2012～2013

課題番号：24800081

研究課題名(和文)階層ベイズモデルに基づく臨床研究・疫学研究の生物統計手法の開発

研究課題名(英文)Biostatistical methods for clinical and epidemiological researches using hierarchical Bayes models

研究代表者

野間 久史(Noma, Hisashi)

統計数理研究所・大学共同利用機関等の部局等・助教

研究者番号：70633486

交付決定額(研究期間全体)：(直接経費) 1,700,000円、(間接経費) 510,000円

研究成果の概要(和文)：近年の計算機環境の著しい進歩から、複雑な階層モデルを用いたベイズ流の統計手法が、医学研究のさまざまな応用において実用化され、実践でも広く用いられるようになった。本研究では、高次元のゲノム情報解析や分子診断法の開発、疫学研究の複雑な研究デザインにおける有効な推定量の構成方法、ネットワークメタアナリシスなどにおいて、新たなベイズ流の統計解析手法の開発を行った。

研究成果の概要(英文)：Recent developments of computational environments have facilitated theoretical and practical researches of Bayesian statistical methodologies. This study developed novel and useful Bayesian statistical methods for clinical and epidemiological researches. In particular, we developed efficient methodologies for high-dimensional genome data analyses and molecular diagnosis. In addition, we studied complex epidemiological study designs and their efficient estimating methods, and developed statistical methods for network meta-analyses.

研究分野：数物系科学

科研費の分科・細目：統計科学

キーワード：階層ベイズモデル 個別化医療 多重検定 ケースコホート研究 欠測データ セミパラメトリック統計理論 経験ベイズ法 ネットワークメタアナリシス

1. 研究開始当初の背景

近年の計算機環境の著しい進歩から、複雑な階層モデルを用いたベイズ流の統計手法が、医学研究のさまざまな応用において実用化され、実践でも広く用いられるようになった。米国医薬品食品局が、医療機器の臨床試験において、「ベイズ統計学の使用についてのガイドライン」を公表したことは、その重要性の顕れといえる。

近年では、さらに、基礎科学研究の著しい進歩や、増大する社会保障費に対する医療政策の課題、また、理論統計学の新たな進展などを受けて、生物統計のさまざまな研究課題における、ベイズ流の方法論の重要性は、ますます増しており、これらの問題解決への貢献が期待されている。

特に、ゲノム情報などの高次元データ解析や、個別化医療に向けての分子診断法、米国で打ち出された Comparative effectiveness research の方法論など、新規で困難な課題に対する、生物統計の方法論の確立と整備は、重要な課題である。また、従来からの方法論においても、不完全データの解析や疫学研究の複雑な研究デザインにおける有効な解析手法など、個別の応用問題ごとに、適切な手法の開発が必要とされる。

2. 研究の目的

本研究では、上述のような高次元のゲノム情報解析や分子診断法の開発、Comparative effectiveness research の方法論から、不完全データの解析、疫学研究の複雑な研究デザインにおける有効な推定量の構成方法などにおける、ベイズ流の統計手法の開発を行う。

3. 研究の方法

本研究は、医学研究を対象とした生物統計手法の開発が研究対象であり、新規な方法論の開発と数理的な論証、新たに開発した方法論における計算技法の開発(広く実践において使うことができるように、汎用性の高いプログラムを作成する)、コンピュータに基づくシミュレーション実験による評価、医学研究の実例による実証的な評

価などの過程を経て、方法論の開発と評価を行う。

4. 研究成果

(1) 高次元ゲノム情報解析における階層ベイズモデルを用いた統計解析手法の開発

ゲノム解析などの高次元データの解析で利用される FDR (false discovery rate) に基づく多重検定における、最強検定に対応する検定方式として、ODP (optimal discovery procedure) についての研究を行った。従来の方法論は、古典的な Neyman-Pearson の補題に基づく「個々の検定の検出力を最大化する」方法であり、必ずしも「多重検定における全体の検出力」を最大化するものとはなっていなかった。本研究で開発した手法は、理論上、この総合的な検出力を最大化する方法となっており、医学研究におけるゲノム情報解析などにおける効率的な情報解析に資するものになると考えられる。実際に、濾胞性リンパ腫の臨床研究 (Dave et al., New England Journal of Medicine 2004; 351: 2159-69) のデータを解析した結果が以下のものとなる。提案する方法は、従来の方法に比較し、より効率的に、臨床アウトカムと関連のある遺伝子を検出できることが示唆される。

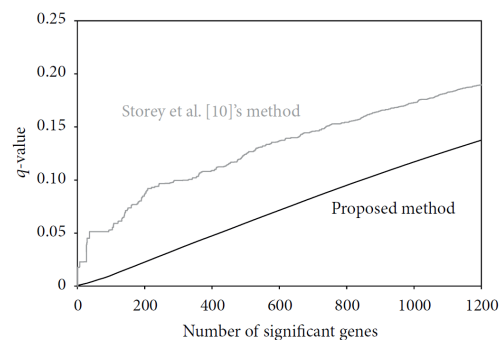


FIGURE 4: Results of the multiple testing for the lymphoma data [39].

図1 濾胞性リンパ腫の臨床研究データの解析結果。有意になった遺伝子の数と、それに対応する q-value (Noma et al., Computational and Mathematical Methods in Medicine 2013; 568480)。

また、経験ベイズ法の枠組みのもとでの階層ベイズモデルに基づく、最適な遺伝子ランキングの方法の研究を行った。特に、既存研究である Noma et al. (2010, *Biostatistics* 2010; 11: 281-289) の提案法を、柔軟なセミパラメトリックモデルを利用した方法に拡張した方法の提案を行った。こちらの方法論でも、より効率的かつ正確に、臨床アウトカムと関連の強い遺伝子を検出できることが可能となり、疾患に関わる分子生物学的機序の解明に資するものとなることが期待できる。本研究の成果は、生物統計学における一流ジャーナルである *Statistics in Medicine* 誌 (IF=2.044) に発表された。

(2) 疫学研究の2段階デザインの統計解析の方法

疫学研究・臨床研究における、効率的な研究デザインとして、実践でも広く利用されているネステッドケースコントロール研究、ケースコホート研究の統計解析の方法についての研究を行った。

特に、このデザインが、関心のあるターゲット集団からの2段階サンプリングに基づくデザインであることに着目し、これがMAR (missing at random) の欠測による不完全データとなっていることから、漸近有効な推定量を与える Multiple Imputation 法による解析を行うことを提案した。シミュレーション実験や実例に基づく数値的な検討を通して、海外のグループが提案したセミパラメトリックな推定量と同等、もしくはそれよりも良好な性能が確認され、その利便性からも、実践における有用性が示唆された。

一方で、2値アウトカムを対象としたケースコホート研究のサンプリング構造が、古典的な2段階ケースコントロール研究と一致することに着目し、これによる、ケースコホート研究のセミパラメトリック有効な推定方法などを系統的にまとめた。

これらの結果を踏まえて、ケースコホート研究の統計解析の方法論についての総説

論文を執筆した。

(3) ネットワークメタアナリシスの統計解析の方法

Comparative Effectiveness Researches の理念に基づく、システマティックレビューの新しい方法として、ネットワークメタアナリシスという方法論が、近年、臨床疫学・薬剤疫学の領域で、大きな注目を集めている。本研究では、このネットワークメタアナリシスの統計解析の方法論として、ネットワーク上の直接比較・間接比較の差の一致性を評価するための方法論についての研究を行った。まだ、具体的な研究成果の発表には至っていないが、ネットワーク全体の情報を利用することにより、より検出力の高い検定方式が期待できるものと思われる。

5. 主な発表論文等

[雑誌論文](計5件)

Noma, H. and Matsui, S. (2013). Empirical Bayes ranking and selection methods via semi-parametric hierarchical mixture models in microarray studies. *Statistics in Medicine* 32: 1904-1916. DOI: 10.1002/sim.5718.

Noma, H. and Matsui, S. (2013). An efficient semiparametric empirical Bayes method for the optimal discovery procedure in multiple significance testing. *Computational and Mathematical Methods in Medicine* 2013: Article ID 568480. DOI: 10.1155/2013/568480.

野間久史, 田中司朗, 田中佐智子, 和泉志津恵. (2013). Multiple Imputation 法によるネステッドケースコントロール研究, ケースコホート研究の解析. *計量生物学* 33: 101-124. DOI: 10.5691/jjb.33.101.

野間久史. (2014). ケースコホート研究の理論と統計手法. *統計数理* (印刷中).

竹内文乃, 野間久史. (2014). 観察研究におけるバイアスの感度解析. *統計数理* (印

刷中).

〔学会発表〕(計 6 件)

Noma, H. (2013). The optimal discovery procedure in multiple significance testing: an empirical Bayes approach. ISM-ISI-ISSAS Joint Conference 2013, Academia Sinica, Taipei, Taiwan, Feb., 2013.

Noma, H., Tanaka, S., Tanaka, S. and Izumi, S. (2013). Multiple imputation analysis of nested case-control and case-cohort studies. The 46th Society for Epidemiologic Research Annual Meeting, The Boston Park Plaza Hotel & Towers, Boston, Massachusetts, Jun., 2013.

野間久史. (2013). Case-Cohort Design : 調査・実験のコストと労力を節減するための新しい臨床研究・疫学研究のデザイン. SAS ユーザー会統計手法研究部会, 北海道大学, 札幌, 2013年6月.

野間久史, 田中司朗. (2013). ケースコホート研究の統計解析: 2 段階ケースコントロール研究との等価性と漸近有効な推定方式. 日本行動計量学会第 41 回大会, 東邦大学, 千葉, 2013年9月.

野間久史, 松井茂之. (2013). The optimal discovery procedure in multiple significance testing: An empirical Bayes method based on semiparametric hierarchical mixture models. 2013 年度統計関連学会連合大会, 大阪大学, 大阪, 2013年9月.

野間久史. (2014). ネットワークメタアナリシスの理論と応用. 第 5 回生物統計ネットワークシンポジウム, ステーションカンファレンス東京, 東京, 2014年3月.

〔図書〕(計 1 件)

Noma, H. and Matsui, S. (2014). Gene screening in the development of genomic signatures. In Design and Analysis of Clinical Trials for Predictive Medicine, Matsui, S., Buyse, M. and Simon, R. (eds.),

Boca Raton: Chapman and Hall/CRC, in press.

6 . 研究組織

(1)研究代表者

野間 久史 (Noma, Hisashi)

統計数理研究所・データ科学研究系・助教

研究者番号 : 70633486