総 説

臨床試験におけるヒストリカルコントロールデータの利用 Bayesian Approach to Utilize Historical Control Data in Clinical Trials

武田健太郎^{*1,*2}・大庭真梨^{*1}・柿爪智行^{*3}・坂巻顕太郎^{*1}・
田栗正隆^{*1}・森田智視^{*1,*4}

Kentaro Takeda^{*1,*2}, Mari Oba^{*1}, Tomoyuki Kakizume^{*3}, Kentaro Sakamaki^{*1},
Masataka Taguri^{*1} and Satoshi Morita^{*1,*4}

^{*1}横浜市立大学大学院医学研究科 臨床統計学・疫学

^{*2}アステラス製薬株式会社 開発本部 データサイエンス部

^{*3}ノバルティスファーマ株式会社 オンコロジー開発統括部
オンコロジー生物統計・DM 部

^{*4}京都大学大学院医学研究科 医学統計生物情報学

^{*1}Department of Biostatistics and Epidemiology,
Yokohama City University Graduate School of Medicine

^{*2}Biostatistics Group, Data Science, Global Development,
Astellas Pharma Inc.

^{*3}Oncology Biostatistics Group, Oncology Biometrics
and Data Management Department, Novartis Pharma K.K.

^{*4}Department of Biomedical Statistics and Bioinformatics,
Kyoto University Graduate School of Medicine
e-mail:kentaro.takeda@astellas.com

It is expected to develop new drug more efficiently by incorporating historical data into the current study data. Borrowing historical data which is sufficiently similar to the current data allows increasing power and improving the accuracy of the estimated treatment effect. On the other hand, if the historical data is not similar to the current data, there is a potential for bias and inflated type I error rate. Power prior and hierarchical model are widely known as the Bayesian approaches with borrowing strength from historical information. They have the advantage of deciding the amount of historical information continuously depending on the similarity between historical data and current data. Our goal is to introduce power prior and hierarchical model while showing some examples, and provide a review of points to keep in mind when these approaches are used in the clinical trials.

Key words: Historical control, Clinical trial, Bayesian approach, Power prior, Hierarchical model, Dynamic borrowing.

1. Introduction

新薬開発においては、まず早期の探索的試験で実際の患者における治療効果を探索した後、治療上の利益を証明するための検証的試験が実施される。その際、公表されたデータまたは既存試験の結果に基づき、新たな治療(以下、新治療)の既存治療に対する治療効果を見積り、事前に設定された有意水準および検出力を担保するような被験者数を設定する(厚生省, 1998)。このように、開発段階を問わず大部分の臨床試験では既存情報が新規試験のデザイン段階で利用されている。組み入れ可能な患者数が限られたり、対照群をおくことが難しい臨床試験では、既存試験において得られたデータ(以下、ヒストリカルデータ)を直接的に利用する例もある。例えば、早期臨床試験、具体的にはがん領域における第I相試験や第II相試験、一般薬のPOC(第IIa相)試験、あるいは医療機器の臨床試験である。しかしながら、ヒストリカルデータを新規試験データと直接的に併合して用いることは、特に本邦においては、現状ではあまり積極的に行われていない。

理論的には、ヒストリカルデータを新規試験データと組み合わせることにより、より効率の高い新治療法開発を行える可能性がある(Hobbs et al., 2013)。例えば、医療機器の開発では、対照治療となる既存の製品にはすでに臨床試験が実施されデータが十分に存在することが多い。これを新規試験の対照治療群のヒストリカルデータ(以下、ヒストリカルコントロールデータ)として用いることで、全体の被験者数を変えずに新治療群への割り付け比率を上げることができる。この場合、新治療群に割り当てられる患者数が増加することにより、新治療に関する安全性および有効性情報をより多く得られるという利点がある。新規試験と試験デザインあるいは患者背景が類似した既存試験から得られるヒストリカルデータを適切に選択して併合することができれば、新規試験において検出力の増大や治療効果の推定精度の向上が期待される。一方で、新規試験データと類似性のないデータを用いると、第一種の過誤が増大したり、治療効果の推定値にバイアスが生じたりする可能性があることが大きな問題となる。

ヒストリカルコントロールデータを利用する例として、新治療の開発を目的とした臨床試験開始前に、対照治療に関する臨床試験が実施されており、そこでは対照治療に関して100例中60例の奏効例が見られた状況を考える。ヒストリカルコントロールデータを利用する単純なアプローチとして、pooling法とtest-then-pool法が挙げられる(Viele et al., 2014)。Pooling法は、ヒストリカルコントロールデータを新規試験の対照治療群データと同等とみなし、そのまま対照治療群データと併合する方法である。例えば、新規試験で対照治療群200例中140例の奏効が見られた場合にpooling法を用いると、併合奏効率は $(140 + 60) / (200 + 100) = 0.667$ となる。一方、test-then-pool法は、新規試験での対照治療群の奏効率とヒストリカルコントロールの奏効率が同じかどうかを仮説検定により判定し、帰無仮説が棄却されなければ併合し、棄却された場合は新規試験データのみを用いる方法である。これらの頻度論的アプローチは単純であるため利用しやすい反面、ヒストリカルコントロールデータの利用法についていくつかの懸念がある。Pooling法は、結果の類似性に関わらずヒストリカルコントロールデータを全て利用するため、ヒストリカルコントロールデータが極端に良い(あるいは悪い)結果だった場合、全体の結果がそれに大きく影響を受けて誤った意思決定をしてしまうリスクがある。Test-then-pool法は、新規試験とヒストリカルコントロールデータとの類似性を判定してからヒストリカルコントロールデータを利

用する分, pooling 法と比べてリスクの少ない方法ではあるが, 仮説検定の結果による類似性の有無のみに着目し, ヒストリカルコントロールデータが類似していれば pooling 法と同様全てのデータを利用する一方, 類似していなければ全く利用しないという, 極端な利用法である. また, 類似性の有無を判定する有意水準の設定も判断が難しい.

Bayes 流アプローチは, 事前情報に基づき事前分布を設定し, 当該試験中に得られた情報によって逐次的に事前分布を更新し, 算出される事後分布に従って統計的推測を行う方法である. Bayes 流アプローチにおけるパラメータ推定の過程は尤度更新とも呼ばれ, データを複数回加えて尤度更新を行った場合, パラメータの最終的な事後分布は最初の事前分布とそれぞれのデータの尤度の積に比例する. ヒストリカルデータを利用する Bayes 流アプローチの利点の 1 つは, 複数のヒストリカルデータの信頼性およびバラツキを考慮しながら, それらの情報を事前分布として柔軟に取り込むことができる点である (Spiegelhalter et al., 2004; Gsteiger et al., 2013). 例えば, ヒストリカルデータをそのまま使うのではなく, ヒストリカルデータの信頼性に応じて情報を割り引いた形で用いることができる (Hueber et al., 2012). ヒストリカルデータの情報量を割り引いて用いる理由は, ヒストリカルデータと新規試験データの間に潜在的な不均一性が存在する可能性に起因する. さらに, 先に述べた test-then-pool 法と同様にヒストリカルデータと新規試験の類似性を考慮しながら, 類似性の程度に応じて連続的にヒストリカルデータから利用する情報量を決定する (以下, 動的利用 (dynamic borrowing)) ことができる. 本論文で紹介する power prior (Chen and Ibrahim, 2000) および階層モデル (Neuenschwander et al., 2010) は, どちらも情報量を割り引く, もしくは, データに依存した動的利用を行うことができる Bayes 流アプローチとして広く知られている (Viele et al., 2014). また, 尤度関数を通してヒストリカルデータを事前分布として利用する明示的な方法として体系的に発展しているのはこれら 2 つであるため, 本論文では, power prior と階層モデルに焦点を当てる.

本レビュー論文では, 臨床試験においてヒストリカルコントロールデータの情報を有効に活用するための Bayes 流アプローチについて議論し, 事例紹介を行うことを目的とする. 第 2 章では, 本論文で用いる事例を紹介する. 第 3 章では, power prior の諸法について論じる. 第 4 章では, 階層モデルの諸法について論じる. 第 5 章では, いくつかの例示とともに臨床試験におけるヒストリカルコントロールデータの利用方法全般における留意点に言及し, 結びに代える.

2. 事例: 胎児心拍数モニタリングと代謝性アシドーシス (Rietbergen et al., 2011)

胎児心拍数モニタリングは, 妊婦の体腹壁に陣痛計を装着し胎児の状態を評価する検査である. 胎児心拍数モニタリングにより胎児心拍数陣痛図 (cardiotocography; CTG) が得られる. CTG によって胎児の心拍数そのものと, 胎動や母体の子宮収縮に対する胎児の心拍数の変化を併せて確認し, 胎児の状態を総合的に評価することができる. Westerhuis らは, CTG に心電図の ST セグメントのモニタリングを加えることで, CTG 単独よりも代謝性アシドーシスの発生リスクを抑えることができるかどうかを検討するランダム化比較試験を行った (Westerhuis et al., 2009). それ以前に 1991 年から 2006 年にかけて同様のランダム化比較試験が 4 つ行われた (表 1). これらの 5 試験は規模もイベント数も互いに異なるものであった. 本論文では, Westerhuis 試験を新規試

表 1. 試験特性とリスク比の推定値および 95 %信頼区間

	試験	登録最終年	N	試験群	対照群	リスク比	95 % CI
新規試験	Westerhuis	2007	5667	CTG+ST	CTG	0.67	(0.38, 1.18)
既存試験	Vayssiere	2006	799	CTG+ST	CTG	1.62	(0.53, 4.86)
	Ojala	2004	1436	ST	CTG	2.45	(0.86, 6.85)
	Amer-Wählin	2000	4238	CTG+ST	CTG	0.46	(0.25, 0.86)
	Westgate	1991	2434	CTG+ST	CTG	0.38	(0.14, 1.07)

表 2. 事例の解析で用いるデータ

	試験	試験治療群	対照治療群
新規試験	Westerhuis	20 / 2827 (0.71 %)	30 / 2840 (1.06 %)
既存試験	Vayssiere	—	148 / 400 (37.0 %)
	Ojala	—	5 / 722 (0.69 %)
	Amer-Wählin	—	31 / 2079 (1.49 %)
	Westgate	—	13 / 1215 (1.07 %)

験, それ以前の 4 つの試験を既存試験と見なし, 紹介する各手法の解析事例として用いる. 表 2 に各試験の症例数および代謝性アシドーシスの発生数の詳細を示す.

3. Power Prior

3.1 概 要

通常の尤度更新では初期事前分布 (initial prior: ヒストリカルデータを得る前の事前分布) とヒストリカルデータの尤度の積から事後分布を算出する. 一方, power prior は初期事前分布とヒストリカルデータの尤度の積から事後分布を求める際に, ヒストリカルデータの尤度をべき乗パラメータ (power parameter) $a_0 \in [0, 1]$ でべき乗することにより, ヒストリカルデータの情報量を割り引くことができる. Power prior は初期事前分布と情報量を割り引かれたヒストリカルデータの尤度から得られた事後分布を指し, 新規試験データを解析する際にヒストリカルデータの情報を有した事前分布として用いる.

本章では二値データにおける power prior の適用法に焦点を当て, これまでに提案されている power prior の諸法を紹介する. なお, 簡単のため, 以降の議論において事前分布は全て proper である (積分が発散しない) と仮定する.

3.2 Conditional Power Prior

Ibrahim and Chen (2000) は, 既知の定数である a_0 を用いてヒストリカルデータの情報を割り引く conditional power prior を提案した. a_0 に分布を仮定せずに既知の定数を与えるため, conditional power prior と呼ばれており, θ を興味のあるパラメータ, D_0 をヒストリカルデータ, その尤度を $L(\theta | D_0)$ と表し, 初期事前分布を $\pi_0(\theta)$ とすると, conditional power prior $\pi^{CPP}(\theta | D_0, a_0)$ は以下の式で与えられる.

$$\pi^{CPP}(\theta | D_0, a_0) \propto L(\theta | D_0)^{a_0} \pi_0(\theta) \quad (1)$$

D_0 において, n_0 例中 x_0 例の反応例および $y_0 = n_0 - x_0$ 例の無反応例が観察された場合,

$$\pi^{CPP}(\theta | D_0, a_0) \propto \theta^{a_0 x_0} (1 - \theta)^{a_0 y_0} \pi_0(\theta) \quad (2)$$

となる. D を D_0 と同じ構造を持つ新規試験データとすると, $\pi^{CPP}(\theta | D_0, a_0)$ は D における事前分布になる. D において n 例中 x 例の反応例および $y = n - x$ 例の無反応例が観察された場合, θ の条件付き事後分布 $q^{CPP}(\theta | D_0, D, a_0)$ は,

$$q^{CPP}(\theta | D_0, D, a_0) \propto L(\theta | D) \pi^{CPP}(\theta | D_0, a_0) = \theta^{a_0 x_0 + x} (1 - \theta)^{a_0 y_0 + y} \pi_0(\theta) \quad (3)$$

と表される.

べき乗パラメータ a_0 はヒストリカルデータから得られる情報量を調整している. 例えば, $a_0 = 0$ の場合はヒストリカルデータを全く用いずに初期事前分布をそのまま事前分布に用い (no borrowing), $a_0 = 1$ の場合はヒストリカルデータをそのまま全て用いて算出された事後分布を事前分布に用いる (full borrowing). また, べき乗パラメータ a_0 とヒストリカルデータの症例数 n_0 の積 $a_0 n_0$ は, ヒストリカルデータから新規試験データに組み入れられる実質的な情報量を表していると考えられる (Hobbs et al., 2011).

べき乗パラメータ a_0 の決定方法として, 専門家の意見に基づき設定する方法 (Neelon and O'Malley, 2010), あるいは, 複数の異なる a_0 の下で算出された conditional power prior の中で事後分布に関する偏差情報量基準 (Deviance Information Criterion, DIC) が最も小さいものを選ぶ方法 (Ibrahim, Chen and Chu, 2012) などが提案されている. また, prior effective sample size を用いて, conditional power prior の情報量を人数に換算して評価する方法も提案されている (Morita et al., 2008).

3.3 Joint Power Prior

Conditional power prior ではべき乗パラメータ a_0 に既知の定数を用いたが, Bayes 流アプローチにおける自然な発想として, a_0 に事前分布 $\pi(a_0)$ を与える方法も提案されている (Ibrahim and Chen, 2000 など). Ibrahim and Chen (2000) は conditional power prior と独立するべき乗パラメータ a_0 の分布の積から構成される joint power prior $\pi^{JPP}(\theta, a_0 | D_0)$ を提案した. $\pi^{JPP}(\theta, a_0 | D_0)$ は, 以下の式で与えられる.

$$\pi^{JPP}(\theta, a_0 | D_0) \propto L(\theta | D_0)^{a_0} \pi_0(\theta) \pi(a_0) \quad (4)$$

ここで, $\pi(a_0)$ は a_0 の事前分布を表し, $0 \leq a_0 \leq 1$ であることから, 通常はベータ分布が用いられる. また, 事前分布の尤度関数は, 2 項係数を含むものと含まないものが用いられている.

Joint power prior は比較的広く用いられているが, いくつかの研究で a_0 の事後分布が 0 付近に集まりやすいことが指摘されている (Duan et al., 2006; Neuenschwander et al., 2009). これは, joint power prior が新規試験データの推論にヒストリカルデータを活用するための枠組みであるにもかかわらず, 実際にはヒストリカルデータが θ の事後分布にほとんど影響を与えていない, すなわち, ヒストリカルデータを十分に活用できていないことを意味する. この傾向は, 事前分布の尤度関数に 2 項係数を含まない joint power prior において顕著となる (Neuenschwander et al., 2009).

さらに, joint power prior は, 尤度関数 $L(\theta | D_0)$ に正の定数を乗じた際に, 異なる joint power prior および事後分布が得られるという特徴を持ち, 尤度原理を満たしていないことが指摘されている (Duan et al., 2006). この尤度原理を満たさないという性質も joint power prior の問題点として挙げられることがある.

3.4 Modified Power Prior (Normalized Power Prior)

Duan et al. (2006) および Neuenschwander et al. (2009) は, 尤度原理を満たした上でべき乗パラメータ a_0 を確率変数として取り扱うために, 基準化 (normalization) された conditional power prior とべき乗パラメータ a_0 の事前分布 $\pi(a_0)$ の積から構成される modified power prior (normalized power prior) $\pi^{MPP}(\theta, a_0 | D_0)$ を提案した. $\pi^{MPP}(\theta, a_0 | D_0)$ は, 以下の式で与えられる.

$$\pi^{MPP}(\theta, a_0 | D_0) \propto \pi(\theta | D_0, a_0) \pi(a_0) = \frac{L(\theta | D_0)^{a_0} \pi_0(\theta) \pi(a_0)}{\int L(\theta | D_0)^{a_0} \pi_0(\theta) d\theta} \quad (5)$$

ここで, $\int L(\theta | D_0)^{a_0} \pi_0(\theta) d\theta$ は基準化定数を表す.

D_0 において, n_0 例中 x_0 例の反応例および $y_0 = n_0 - x_0$ 例の無反応例が観察された場合, $\pi_0(\theta)$ がベータ分布 $Beta(c, d)$ とすると,

$$\pi^{MPP}(\theta, a_0 | D_0) \propto \frac{\Gamma(a_0 n_0 + c + d)}{\Gamma(a_0 x_0 + c) \Gamma(a_0 y_0 + d)} \theta^{a_0 x_0 + c - 1} (1 - \theta)^{a_0 y_0 + d - 1} \pi(a_0) \quad (6)$$

となる. また, D において n 例中 x 例の反応例および $y = n - x$ 例の無反応例が観察された場合, a_0 の周辺事後分布 $q^{MPP}(a_0 | D_0, D)$ は,

$$\begin{aligned} q^{MPP}(a_0 | D_0, D) &= \int \frac{\Gamma(a_0 n_0 + c + d)}{\Gamma(a_0 x_0 + c) \Gamma(a_0 y_0 + d)} \theta^{a_0 x_0 + x + c - 1} (1 - \theta)^{a_0 y_0 + y + d - 1} \pi(a_0) d\theta \\ &= \frac{\Gamma(a_0 n_0 + c + d) \Gamma(a_0 x_0 + x + c) \Gamma(a_0 y_0 + y + d)}{\Gamma(a_0 x_0 + c) \Gamma(a_0 y_0 + d) \Gamma(a_0 n_0 + n + c + d)} \pi(a_0) \end{aligned} \quad (7)$$

となる.

Modified power prior は尤度原理を満たす. また, 事前分布 $\pi(a_0)$ にベータ分布 $Beta(a, b)$ を用いる場合, 定数パラメータ a, b を用いてヒストリカルデータから得られる情報量を調整することができる. 例えば, $(a = 10, b = 1)$ ならばヒストリカルデータから得られる情報量は大きく, $(a = 1, b = 10)$ ならばヒストリカルデータから得られる情報量は小さくなる. さらに, a_0 の事後分布が極端に 0 に近い値を取る傾向もないので, joint power prior のようにヒストリカルデータの情報量を過小評価することも避けられる (Duan et al., 2006; Neuenschwander et al., 2009).

シミュレーション研究によると, modified power prior には, ヒストリカルデータと新規試験データの性質が類似している場合, べき乗パラメータ a_0 の初期事前分布の設定に従ってヒストリカルデータを利用し, ヒストリカルデータと新規試験データの性質が異なっている場合, べき乗パラメータ a_0 の事後平均は適応的に 0 方向に変化し, ヒストリカルデータから得られる情報量は小さくなる, という特徴がある (Hobbs et al., 2011; 武田, 2014).

ただし, modified power prior を用いる場合, MCMC サンプラー内で a_0 が更新されるたびに基準化定数 $\int L(\theta | D_0)^{a_0} \pi_0(\theta) d\theta$ を数値積分により求める必要があり, 計算が非常に高負荷にな

るという問題が生じる。また、 θ が多次元の場合、数値積分は 1 次元の場合に比べて格段に難しくなる。SAS による modified power prior の算出については、武田 (2014) に詳しい。

3.5 Commensurate Power Prior

Hobbs et al. (2011) は、Joint power prior および modified power prior の課題として、ヒストリカルデータと新規試験データの類似性を直接パラメータ化していない点を、共通性(commensurability)を考慮していないという表現を用いて指摘し、ヒストリカルデータ D_0 と新規試験データ D の共通性に基づいてヒストリカルデータから得られる情報量を適応的に調整する方法として、location commensurate power prior $\pi^{LCP}(\theta, a_0, \lambda | D_0)$ を提案した。Location commensurate power prior では、 D_0 の位置パラメータ θ_0 と D の位置パラメータ θ を別々に考え、 θ の事前分布の平均を θ_0 、精度を λ としている。この際、 λ が D_0 と D の共通性を測る尺度となっている。 $\pi^{LCP}(\theta, a_0, \lambda | D_0)$ は、以下の式で与えられる。

$$\pi^{LCP}(\theta, a_0, \lambda | D_0) \propto \int \frac{L(\theta_0 | D_0)^{a_0} \pi_0(\theta_0)}{\int L(\theta_0 | D_0)^{a_0} \pi_0(\theta_0) d\theta_0} \pi_0(\theta | \theta_0, \lambda) d\theta_0 \pi(a_0 | \lambda) \pi(\lambda) \quad (8)$$

ここで、べき乗パラメータ a_0 の事前分布 $\pi(a_0 | \lambda)$ に $Beta(\lambda, 1)$ もしくは $Beta(\log(\lambda), 1)$ を用いる。これにより、ヒストリカルデータから得られる情報量を共通性の程度 λ に応じて調整できるようになる。例えば、 λ が 0 に近い(共通性が低い)ときは D_0 から得る情報量を小さくし、 λ が大きい(共通性が高い)ときは D_0 から得る情報量を大きくすることができる。Hobbs et al. (2011) では $\log(\lambda)$ の事前分布にコーシー分布 $Cauchy(0, 30)$ を用いている。さらに、べき乗パラメータ a_0 を用いずに D_0 と D の共通性を利用してヒストリカルデータを適応的に組み込む方法も提案されている(Hobbs et al., 2011; 2012)。

3.6 複数のヒストリカルデータにおける Power Prior

ここまで単一のヒストリカルデータを対象とした power prior について紹介したが、新規試験実施時に複数の既存試験から得られた複数のヒストリカルデータが存在する場合もある。こうした状況に対応するために、power prior の適用範囲を複数のヒストリカルデータに拡張する方法が提案されている。

Ibrahim and Chen (2000) は joint power prior を複数のヒストリカルデータに一般化した。 H 個の独立なヒストリカルデータを仮定し、 D_{0h} を h 番目のヒストリカルデータとする ($D_0 = (D_{01}, \dots, D_{0H})$)。さらに、 δ_h を D_{0h} における重み付けパラメータとし、 δ_h に互いに独立な事前分布 $Beta(a, b)$ を与え、 $\delta_h = (\delta_1, \dots, \delta_H)$ とすると、複数のヒストリカルデータにおける joint power prior $\pi^{JPP}(\theta, \delta | D_0)$ は、以下の式で与えられる。

$$\pi^{JPP}(\theta, \delta | D_0) = \left(\prod_{h=1}^H [L(\theta | D_{0j})]^{\delta_h} \pi(\delta_h | a, b) \right) \pi(\theta) \quad (9)$$

ただし、複数のヒストリカルデータにおける joint power prior においても、3.3 節で紹介したように、尤度原理を満たさず、ヒストリカルデータを過小評価するといった問題は残る。そのため、Duan et al. (2006) は複数のヒストリカルデータに対応できるように、modified power prior を拡張した。複数のヒストリカルデータにおける modified power prior $\pi^{MPP}(\theta, \delta | D_0)$ は、以下の式で与えられる。

$$\pi^{MPP}(\theta, \delta | D_0) = \frac{\left(\prod_{h=1}^H [L(\theta | D_{0h})]^{\delta_h} \pi(\delta_h | a, b) \right) \pi(\theta)}{\int_{\Theta} \left(\prod_{h=1}^H [L(\theta | D_{0h})]^{\delta_h} \right) \pi(\theta) d\theta} \quad (10)$$

複数のヒストリカルデータはそれぞれの試験内では均一と考えられるが、試験間では異なる施設・異なる測定時期などに起因する潜在的な不均一性が存在する。ここで紹介した枠組みは、試験毎のヒストリカルデータに互いに独立な重み付けパラメータを与え、新規試験内で各ヒストリカルデータを動的利用することにより、複数のヒストリカルデータ間の潜在的な不均一性に対応し、ヒストリカルデータを適切に利用することができると考えられる。なお、power prior を複数のヒストリカルデータに拡張する際の D_{0h} における重み付けパラメータ δ_h には確率変数が与えられることが多く、既知の定数を与える conditional power prior はあまり用いられていない。

3.7 適用例

表2の事例を power prior を用いて解析した。既存試験には新規試験と規模の近い Amer-Wählin 試験を採用した。代謝性アシドーシスの発生のデータは2項分布に従うことを仮定し、初期事前分布には $Beta(1,1)$ を用いた。表3に conditional power prior, joint power prior 1 (2項係数を含まない joint power prior), joint power prior 2 (2項係数を含む joint power prior), modified power prior を用いた下での新治療群と対照治療群のリスク比 RR およびべき乗パラメータ a_0 の事後平均と95%信用区間を示した。また、ヒストリカルデータから新規試験データに組み入れられる情報量の目安として、べき乗パラメータ a_0 の事後平均(conditional power prior の場合は既知の定数)とヒストリカルデータの症例数の積 $a_0 n_0$ を示した。 RR および a_0 の事後分布の特徴を視覚的に検討するために、図1に RR の事後分布のカーネル密度推定の結果を示し、図2に a_0 の事後分布および初期事前分布のカーネル密度推定の結果を示した。

Conditional power prior では、 a_0 の値が大きくなるにつれて、 RR の事後平均は小さくなり、

表 3. 各種 power prior におけるリスク比 RR およびべき乗パラメータ a_0 の事後分布

方法	超事前分布	リスク比 RR			べき乗パラメータ a_0			$a_0 n_0$
		事後平均	標準偏差	95 % CI	事後平均	標準偏差	95 % CI	
Conditional Power Prior	$a_0=0$	0.702	0.203	(0.420, 1.072)	—	—	—	0.0
	$a_0=0.2$	0.667	0.184	(0.407, 0.998)	—	—	—	415.8
	$a_0=0.4$	0.643	0.173	(0.398, 0.956)	—	—	—	831.6
	$a_0=0.6$	0.625	0.164	(0.389, 0.920)	—	—	—	1247.4
	$a_0=0.8$	0.611	0.156	(0.385, 0.892)	—	—	—	1663.2
	$a_0=1.0$	0.599	0.151	(0.381, 0.870)	—	—	—	2079.0
Joint Power Prior 1	$a_0 \sim Be(1,1)$	0.702	0.203	(0.420, 1.076)	0.006	0.006	(0.000, 0.018)	12.7
	$a_0 \sim Be(10,1)$	0.691	0.197	(0.417, 1.052)	0.062	0.019	(0.034, 0.097)	128.0
	$a_0 \sim Be(1,10)$	0.702	0.203	(0.418, 1.074)	0.006	0.006	(0.000, 0.017)	12.0
Joint Power Prior 2	$a_0 \sim Be(1,1)$	0.668	0.189	(0.405, 1.012)	0.237	0.216	(0.013, 0.702)	492.8
	$a_0 \sim Be(10,1)$	0.607	0.156	(0.382, 0.889)	0.880	0.104	(0.673, 0.992)	1830.0
	$a_0 \sim Be(1,10)$	0.692	0.197	(0.416, 1.052)	0.067	0.065	(0.004, 0.199)	139.5
Modified Power Prior	$a_0 \sim Be(1,1)$	0.635	0.173	(0.392, 0.952)	0.525	0.273	(0.091, 0.953)	1091.1
	$a_0 \sim Be(10,1)$	0.605	0.155	(0.381, 0.884)	0.908	0.084	(0.739, 0.995)	1887.0
	$a_0 \sim Be(1,10)$	0.683	0.193	(0.411, 1.033)	0.120	0.087	(0.017, 0.292)	248.6

図 1. リスク比 RR の事後分布

$a_0 = 0.2$ 以上で RR の 95%信用区間の上限は 1 を下回った. $a_0 = 0.2$ の場合, ヒストリカルデータから約 415 例分の情報を利用して, RR の事後分布は a_0 の値が大きくなるにつれて, 左方向にシフトしていた(図 1 (a)).

Joint power prior 1 では, a_0 の初期事前分布の設定に関わらず, a_0 の事後平均は 0 に近い値をとり, ヒストリカルデータをほとんど利用していないことが示唆された. RR の事後平均は 0.70 前後であり, 95%信用区間の上限は 1 を上回った. $a_0 \sim \text{Beta}(1,1)$ の場合, ヒストリカルデータから約 13 例分の情報しか利用していなかった. RR の事後分布は, a_0 の初期事前分布の設定に関わらず, ほぼ同じ形状をしていた(図 1 (b)). a_0 の事後分布は, a_0 の初期事前分布の設定に関わらず, 0 付近に集中した(図 2 (b)).

Joint power prior 2 では, a_0 の事後平均は joint power prior 1 よりも大きな値となった. べき乗パラメータを $a_0 \sim \text{Beta}(10,1)$ と設定し, ヒストリカルデータから得られる情報量を大きくした場合に RR の事後平均は約 0.69 となり, 95%信用区間の上限は 1 を下回った. $a_0 \sim \text{Beta}(1,1)$ の場合, ヒストリカルデータから約 493 例分の情報を利用して, RR の事後分布は, a_0 の初期事前分布の設定においてヒストリカルデータから得られる情報量を大きくした場合に, 初期事前分布に比べて左方向にシフトしていた(図 1 (c)). a_0 の事後分布は, やや左方向にシフトして

図 2. べき乗パラメータ a_0 の事後分布および初期事前分布

いた(図 2 (c)).

Modified power prior では, べき乗パラメータを $a_0 \sim \text{Beta}(1,1)$ および $a_0 \sim \text{Beta}(10,1)$ と設定した場合に, a_0 の事後平均は約 0.5 および約 0.9 となり, RR の 95%信用区間の上限は 1 を下回った. $a_0 \sim \text{Beta}(1,1)$ の場合, ヒストリカルデータから約 1091 例分の情報を利用していた. RR の事後分布は, a_0 の初期事前分布の設定においてヒストリカルデータから得られる情報量を大きくした場合に, 左方向にシフトしていた(図 1 (d)). a_0 の事後分布は初期事前分布とほぼ同じ形状をしていた(図 2 (d)). 本事例においては, ヒストリカルデータと新規試験の対照治療群のデータが非常に類似していたため, べき乗パラメータ a_0 の事後平均は初期事前分布の平均値とほぼ同一となり, 初期設定に従いヒストリカルデータから得られる情報を利用していった.

Joint power prior 1, joint power prior 2, modified power prior において, 同じ初期事前分布の下で a_0 の事後平均を比較すると, joint power prior 1, joint power prior 2, modified power prior の順に大きくなった. 同様の順に, RR の事後平均は小さくなり, 95%信用区間はこの順に狭くなった. 以上より, joint power prior 1, joint power prior 2, modified power prior の順にヒストリカルデータから得られる情報をより大きく利用していることがわかる.

以上より, 新規試験データのみを用いた場合は RR の 95%信用区間の上限が 1 を上回っていた

が、ヒストリカルデータから得られる情報を一定以上利用した場合は RR の 95 % 信用区間の上限が 1 を下回った。そのため、ヒストリカルデータを適切に利用することで新規治療群の治療効果を説明できる可能性が示唆された。

4. 階層モデル

4.1 概 要

本章では新規試験の対照治療群のパラメータと既存試験の対照治療群のパラメータが同じ分布からのランダムサンプルとみなす階層モデルを用いる方法を紹介する。新規試験の対照治療群のパラメータと既存試験の対照治療群のパラメータを同じ分布からのランダムサンプルとみなすことで、完全に同一でも独立でもないという柔軟(中間的)な仮定を置くことができる。以下は正規分布に従うデータで記述するが、他の分布に従うデータでも要約統計量を正規近似して同様の議論が可能である。

4.2 既存試験における予測分布を用いる方法 (Meta-Analytic Predictive; MAP (Schmidli et al., 2014))

利用可能なヒストリカルデータが H 個存在する場合を考える。既存試験 h の対照治療群の要約統計量およびパラメータを Y_h ($h = 1, \dots, H$), θ_h とし、新規試験の対照治療群の要約統計量およびパラメータを Y^* , θ^* とする。既存試験のデータから θ^* に関連する情報を抽出するために次のようにサンプリングモデルと初期事前分布を考える。データは正規分布に従い、初期事前分布も正規分布の場合、

サンプリングモデル

$$Y_h | \theta_h, \sigma_h \sim N(\theta_h, \sigma_h^2) \quad (11)$$

初期事前分布

$$\theta_1, \dots, \theta_H, \theta^* \sim N(\mu, \tau^2) \quad (12)$$

となる。ここで τ^2 は試験間分散、 σ_h^2 は試験内分散である。初期事前分布はすべてのパラメータが共通の分布に従うことを表している。既存試験のパラメータと新規試験のパラメータを関連付ける初期事前分布にはいくつかの方法がある (Pocock, 1976; Turner et al., 2009) が、ここでは最もシンプルな上記の初期事前分布を用いる。

完全 Bayes の方法では、さらにパラメータ μ, τ^2 に不確実性を考慮し、以下のような超事前分布を仮定する。ここでは、分散パラメータの超事前分布として形状パラメータ a_ϕ 、尺度パラメータ b_ϕ の逆ガンマ分布を記載しているが、正の定義域を有する他の確率分布も適用可能である (Gelman, 2006; Gelman et al., 2014)。

超事前分布

$$\begin{aligned} \mu &\sim N(\mu_\phi, \tau_\phi^2), \\ \tau^2 &\sim IG(a_\phi, b_\phi) \end{aligned} \quad (13)$$

階層モデルにおいて、ヒストリカルデータの利用の程度と関連するパラメータは、試験間のばらつきを表す τ である。 $\tau \approx 0$ はパラメータ θ が試験間でほぼ同じであることを意味し、ヒストリカルデータを積極的に利用して興味あるパラメータを推定する。 τ が大きければ試験結果が互いに大きく異なることを意味し、ヒストリカルデータを利用せず、新規試験のみを使った解析となる。 τ の超事前分布の設定については後述する。なお、試験内分散 σ_h^2 は本章では既知とする。 σ_h^2 を未知として事後分布を推定することも可能だが、同一試験内の比較的十分な人数によって推定されるため、その不確実性を無視しても興味のある結果への影響が少ないとされている (Gelman, 2006)。

新規試験のデザイン段階で用意する事前分布として予測分布 $p(\theta^* | Y_1, \dots, Y_H)$ を利用することができる (Neuenschwander et al., 2010; Gsteiger et al., 2013)。 $p(\theta^* | Y_1, \dots, Y_H)$ は、以下の式で与えられる。

$$p(\theta^* | Y_1, \dots, Y_H) = \int p(\theta^* | \mu, \tau) \prod_h L(\theta_h | Y_h) p(\theta_h | \mu, \tau) p(\mu, \tau) d\theta_1 \cdots d\theta_H d\mu d\tau \quad (14)$$

ここで、 τ^2 および σ_h^2 が既知の場合、 μ の超事前分布に一樣分布を仮定すれば、

$$\theta^* | Y_1, \dots, Y_H, \tau \sim N \left(\frac{\sum w_h Y_h}{\sum w_h}, \frac{1}{\sum w_h} + \tau^2 \right) \quad (15)$$

と表せる。ここで $w_h = (\sigma_h^2 + \tau^2)^{-1}$ である。予測分布の分散は全体平均 μ の事後分散に、新規試験のパラメータ θ^* の不確実性を追加した形で表現され、既存試験間の分散の影響が大きいことがわかる。

4.3 事後分布を用いる方法 (Meta-Analytic Combined; MAC)

ヒストリカルデータにおける予測分布を事前分布とする方法は、新規試験データがまだ存在しない段階で準備できるため、前向き (prospective) な活用法といえる。しかし、新規試験のパラメータ θ^* と既存試験のパラメータ $\theta_1, \dots, \theta_H$ の近さに基づいて、ヒストリカルデータを動的利用したい場合には、新規試験のデータ Y^* を得た後にこれを含めた θ^* の事後分布 $p(\theta^* | Y_1, \dots, Y_H, Y^*)$ を直接推定する必要がある。 $p(\theta^* | Y_1, \dots, Y_H, Y^*)$ は、以下の式で与えられる。

$$p(\theta^* | Y_1, \dots, Y_H, Y^*) = \int L(\theta^* | Y^*) p(\theta^* | \mu, \tau) \prod_h L(\theta_h | Y_h) p(\theta_h | \mu, \tau) p(\mu, \tau) d\theta_1 \cdots d\theta_H d\mu d\tau \quad (16)$$

なお、式 (15) と同様の条件下で、すなわち τ^2 および σ_h^2 が既知の場合、 τ 条件付きの事後分布は

$$\theta^* | Y_1, \dots, Y_H, Y^*, \tau \sim N \left(\frac{\frac{Y^*}{\sigma^{*2}} + \frac{\sum w_h Y_h}{1 + \tau^2 \sum w_h}}{\frac{1}{\sigma^{*2}} + \frac{\sum w_h}{1 + \tau^2 \sum w_h}}, \frac{1}{\frac{1}{\sigma^{*2}} + \frac{\sum w_h}{1 + \tau^2 \sum w_h}} \right) \quad (17)$$

となり、式 (15) を新しい新規試験のデータ Y^* で更新した形で表される。ここでも試験間分散 τ^2 が影響することがわかる。予測分布を用いる方法との違いは、試験間分散 τ^2 の事後分布の推定に Y^* の情報が加味されることであり、その点が動的利用と呼ばれる特徴となる。この事後分布を用

いる方法は、基本的には、新規試験データが既に得られた後に適用される後ろ向き (retrospective) な手法である。前向きに新規試験の計画に用いる場合には、想定される結果のデータを発生させ、事後推定値の挙動を確認するというインテンシブなシミュレーションによる検討が必要となる (4.8 節の事例参照)。

4.4 試験間ばらつきの事前分布

MAP, MAC とともに、試験間ばらつきの指標 τ は既存試験の結果に基づき推定されるが、利用可能な既存試験は少ないことが多いため、正確な推定が難しく、超事前分布の影響を受けやすいことが指摘されている (Lambert et al., 2005; Gelman, 2006)。一般的には、分散に対して共役分布である逆ガンマ分布、あるいは、分散の対数または標準偏差に対して一定の範囲内の値が等確率であることを意味する一様分布などが用いられることが多い。しかし、試験間ばらつきに関する事前情報が少ない場合や試験数が少ない場合には、複数の異なる超事前分布を用いた感度解析が必要である。多くは無情報事前分布になるように超事前パラメータを設定することが多い (Gelman, 2006)。また、感度解析の一つとして“現実的な超事前分布” (plausible prior) を利用することが議論されている (Spiegelhalter et al., 2001)。例えば対数オッズのスケールで 2 を超える状況は、試験間の不均一性がとても大きいことを意味し、各試験を独立に解析した場合と同等な結果を導く。これを踏まえて、例えば既存試験と新規試験がある程度類似性があると考える場合に、「超事前分布に切断標準正規分布を用いる」ことができる。切断標準正規分布では試験間の異質性が大きい確率が 5% ($P(\tau > 2) \approx 5\%$) となるためである。適用例 (4.7) の解析では試験間分散 τ^2 あるいは標準偏差 τ の超事前分布を複数設定し、新規試験における治療効果の推定値の挙動を評価する。

4.5 既存データの利用の程度

階層モデルでは、ヒストリカルデータからどれくらいの情報量を利用するかは試験内分散 σ_h^2 、試験間分散 τ^2 、および試験間ばらつきの超事前分布に依存する。多重積分が必要なため power prior のような明示的な表現は困難だが、分散比 (Pennello et al., 2008; Neuenschwander et al., 2010) や prior effective sample size (Morita et al., 2012) を用いて、事前分布の情報量を評価する方法が提案されている。

例えば予測分布を用いる場合に分散比で考えると、 σ_h^2 および τ^2 で条件付けた下では、次のように評価できる。既存試験のパラメータの期待値が全く同一、すなわち $\tau = 0$ の場合の新規試験の事前分散は、式 (15) に $\tau^2 = 0$ を代入し、

$$V(\theta^* | \tau = 0) = 1 / \sum \sigma_h^{-2} \quad (18)$$

となる。これを、ヒストリカルデータの全ての情報が新規試験の事前分布に用いられている状態と考える。一方、既存試験のパラメータの期待値にばらつきが生じる場合 ($\tau > 0$)、事前分散は式 (15) の分散である。これらの比

$$\frac{V(\theta^* | \tau = 0)}{V(\theta^* | \tau > 0)} = \frac{1 / \sum \sigma_h^{-2}}{1 / \sum (\sigma_h^2 + \tau^2)^{-1} + \tau^2} \quad (19)$$

により、試験間のばらつきの違いによってどの程度情報が割り引かれて事前分布に利用されているかを相対的に評価することができる。既存試験同士で等分散 ($\sigma_1^2 = \dots = \sigma_H^2 = \sigma^2$) および等被験者数のとき、(19) 式は、

$$\lambda = \frac{\sigma^2}{\sigma^2 + N \left(1 + \frac{1}{H}\right) \tau^2} \quad (20)$$

に簡略化できる. τ^2 が大きいほどヒストリカルデータの情報の利用が少ないこと, 既存試験の全症例数 N および τ^2 を与えた下では, 試験数が多い方がヒストリカルデータの情報をより大きく利用していることがわかる.

4.6 Power Prior との関係

Power prior では a_0 がヒストリカルデータを利用する程度(重み)を表し, 事前分布によってその値を調整する. 階層モデルでは, 試験間分散 τ^2 がヒストリカルデータと新規試験データの共通性を表すとともに, 直接的にヒストリカルデータを利用する程度を決定するパラメータでもある. Power prior における a_0 と事後分布を用いる方法における τ^2 の関係について, Chen and Ibrahim (2006) は一般線形モデルの場合に解析的な対応関係を示している. n_0 人の既存試験データ D_0 に基づく尤度 $L(\theta | D_0)$ および初期事前分布 $\pi_0(\theta)$ で構成する power prior $\pi(\theta | D_0, a_0)$ は, 以下で与えられる.

$$\pi(\theta | D_0, a_0) \propto L(\theta | D_0)^{a_0} \pi(\theta) \quad (21)$$

データが従う分布および事前分布が正規分布で既存試験が 1 つの場合に, power prior の初期事前分布を $\pi_0(\theta) \propto 1$ とし, 階層モデルの試験間分散 τ^2 を既知, μ の超事前分布パラメータを無情報 ($\mu_\phi = 0, \tau_\phi^2 \rightarrow \infty$) とすると, べき乗パラメータ a_0 は階層モデルの試験間分散および試験内分散の関数となり,

$$a_0 = \frac{1}{\frac{2\tau^2 n_0}{\sigma^2} + 1} \quad (22)$$

と表せる. Chen and Ibrahim (2006) はさらに一般化線形モデルの場合, 既存試験が複数ある場合についても漸近的に power prior のべき乗パラメータ a_0 と階層モデルの試験間分散および試験内分散の関係を導いており, 階層モデルの試験間分散および試験内分散を power prior のべき乗パラメータに換算して考察することを可能としている.

4.7 適用例

第 3 章と同様, 表 2 の事例を用いる. 新治療群, 対照治療群それぞれにおけるイベント発生対数オッズ θ を正規近似し, 新治療群と対照治療群のオッズ比 OR と試験間分散 τ^2 の事後平均, 標準偏差および 95% 信用区間を MAP の方法と MAC の方法それぞれで推定した.

対照治療群の平均 μ の超事前分布は以下のように設定した.

$$\mu \sim N(0, 100^2) \quad (23)$$

試験間ばらつきには, 異なる複数の超事前分布を設定した. まず, τ^2 の超事前分布に逆ガンマ分布を用いた 3 通り, τ の超事前分布に一樣分布を用いた 2 通り, 最後に τ の超事前分布に切断標準正規分布を用いた 1 通りである.

$$\begin{aligned}
\tau^2 &\sim IGa(10, 0.01) \\
\tau^2 &\sim IGa(1, 0.01) \\
\tau^2 &\sim IGa(0.01, 0.01) \\
\tau &\sim U(0, 1) \\
\tau &\sim U(0, 5) \\
\tau &\sim \text{half } N(0, 1)
\end{aligned}
\tag{24}$$

新治療群の対数オッズの事前分布は無情報 $N(0, 100^2)$ とした。なお新規試験単独で推定されるオッズ比の最尤推定値は 0.667 (95 %信頼区間 0.378, 1.178), 新規試験と Amer-Wåhlin 試験を併合して得られるオッズ比は 0.567 (95 %信頼区間 0.342, 0.942) である。

MAP を用いた結果を表 4 に示した。既存試験が 1 つの場合は τ^2 の情報が得られないため既知として与えた。 τ^2 の値が小さいほどプールした結果に近くなり、95 %信用区間の上限が 1 を下回った。

表 4. オッズ比 OR の事後分布および試験間分散 τ^2 の事後分布 (MAP を用いた方法)

超事前分布	オッズ比 OR			試験間分散 τ^2		
	事後平均	標準偏差	95 % CI	事後平均	標準偏差	95 % CI
1 つの 既存試験						
$\tau^2=0.001$	0.586	0.153	(0.344, 0.936)			
$\tau^2=1$	0.696	0.202	(0.381, 1.170)			
$\tau^2=5$	0.699	0.204	(0.382, 1.176)			
複数の 既存試験						
$\tau^2 \sim IGa(10, 0.01)$	0.661	0.191	(0.364, 1.107)	0.504	0.175	(0.263, 0.934)
$\tau^2 \sim IGa(1, 0.01)$	0.695	0.203	(0.381, 1.170)	3.986	4.846	(0.902, 14.535)
$\tau^2 \sim IGa(0.01, 0.01)$	0.698	0.204	(0.382, 1.175)	11.428	39.501	(1.250, 55.663)
$\tau \sim U(0, 1)$	0.678	0.197	(0.372, 1.138)	0.863	0.109	(0.597, 0.995)
$\tau \sim U(0, 5)$	0.698	0.203	(0.382, 1.174)	8.231	5.754	(1.510, 22.547)
$\tau \sim \text{half } N(0, 1)$	0.694	0.202	(0.380, 1.167)	3.040	1.639	(1.011, 7.207)

事後分布の形状を視覚的に検討する目的で、 OR の事後分布を図 3 に、 τ^2 の事後分布を図 4 に示した。既存試験が 1 つの MAP (図 3 (a)) では、 τ^2 が小さいほど OR の分布が左方向にシフトした。

既存試験が複数の場合の結果は τ^2 の超事前分布をヒストリカルデータで更新した分布を示している。 τ^2 の事前分布において 0 周辺の密度が高いものほど OR の事後平均は小さく、無情報なほど OR の事後平均は大きかった。 τ^2 の事後平均および標準偏差は事前分布の影響を強く受けた。図 3 (b) においてヒストリカルデータをより多く利用している場合に OR の分布が左にシフトした。

MAC を用いた結果を表 5 に示した。ここで示す τ^2 の事後分布は、超事前分布をヒストリカルデータおよび新規試験のデータで更新した分布である。既存試験が 1 つの場合、 τ^2 の事後分布は超事前分布の影響を強く受け、大きくばらついた。 OR の事後平均は τ^2 の事後平均が小さいほどプールした結果に近づいた。

表 5. オッズ比 OR の事後分布および試験間分散 τ^2 の事後分布 (MAC を用いた方法)

超事前分布		オッズ比 OR			試験間分散 τ^2		
		事後平均	標準偏差	95 % CI	事後平均	標準偏差	95 % CI
1 つの 既存試験	$\tau^2 \sim IGa(10, 0.01)$	0.585	0.152	(0.341, 0.937)	0.001	0.0004	(0.001, 0.002)
	$\tau^2 \sim IGa(1, 0.01)$	0.622	0.173	(0.352, 1.024)	0.040	0.180	(0.003, 0.213)
	$\tau^2 \sim IGa(0.01, 0.01)$	0.667	0.192	(0.367, 1.112)	61.761	1884.860	(0.007, 51.031)
	$\tau \sim U(0, 1)$	0.662	0.191	(0.368, 1.117)	0.251	0.268	(0.000, 0.911)
	$\tau \sim U(0, 5)$	0.679	0.197	(0.375, 1.137)	3.773	5.826	(0.001, 21.392)
複数の 既存試験	$\tau \sim half N(0, 1)$	0.667	0.193	(0.370, 1.124)	0.516	0.902	(0.001, 3.105)
	$\tau^2 \sim IGa(10, 0.01)$	0.662	0.192	(0.363, 1.111)	0.533	0.176	(0.287, 0.968)
	$\tau^2 \sim IGa(1, 0.01)$	0.694	0.203	(0.378, 1.174)	3.197	3.506	(0.877, 10.240)
	$\tau^2 \sim IGa(0.01, 0.01)$	0.695	0.203	(0.380, 1.160)	6.345	11.826	(1.137, 26.117)
	$\tau \sim U(0, 1)$	0.676	0.196	(0.372, 1.127)	0.866	0.108	(0.604, 0.996)
	$\tau \sim U(0, 5)$	0.697	0.203	(0.383, 1.173)	6.659	5.063	(1.340, 20.876)
	$\tau \sim half N(0, 1)$	0.692	0.204	(0.380, 1.168)	2.791	1.476	(0.983, 6.576)

既存試験が複数の場合, OR の事後平均は MAP とほぼ同じであったが, τ^2 の事後平均は MAP に比べやや小さくなり, 標準偏差と信用区間も狭まった. 図 3 (c), 図 3 (d) においてもヒストリカルデータをより多く利用している場合に OR の分布が左にシフトした.

図 3. オッズ比 OR の事後分布

図 4. 試験間分散パラメータ τ^2 の分布

τ^2 の事後分布について、図 4 (c)、図 4 (d) では、事前分布が 0 周辺の密度が高いものはその影響をのこして $[0, 1]$ に高い密度をもち、その他は幅の広い事後分布となった。

本事例は、各既存試験でイベント発生割合が大きく異なるため、MAP、MAC いずれも複数の既存試験を使った際の τ^2 が無視できない大きさとなった。既存試験の OR 推定への影響はわずかで、治療効果について新規試験のみの解析と結論が変わらなかった。

以上より、既存試験が 1 つの場合、試験間分散 τ^2 が群間差の推定に影響すること、既存試験が複数存在しても試験間で結果が大きく異なると、ヒストリカルデータから利用される情報量は大きくならないこと、また、 τ^2 の評価には既存試験が 4 つでは不十分であることが示された。なお、 τ^2 あるいは τ の超事前分布に何を用いるかについては、逆ガンマ分布よりも一様分布あるいは切断標準正規分布の方が範囲や分散が直観的に理解しやすく、分布の値の設定がしやすいという利点があると考えられた。

4.8 事例：ST 上昇型心筋梗塞患者を対象とした臨床試験

臨床試験に階層モデルが用いられた実例として、AMIHOT II 試験 (Stone et al., 2009) を紹介する。AMIHOT II 試験の前に行われた AMIHOT I 試験は、ST 上昇型心筋梗塞患者に対し、カテーテルを用いて内科的に治療を行う経皮的冠動脈形成術を行った後、冠動脈内に過飽和酸素を

図 5. AMIHOT I および II 試験における梗塞面積の左心室に対する割合(%)の分布(左から AMIHOT I 試験, AMIHOT II 試験および二つの試験より階層モデルによって推定された事後分布. Stone et al., 2009 より引用).

Adapted with permission from Lippincott Williams and Wilkins/Wolters Kluwer Health: Stone, G. W., Martin, J. L., de Boer, M. J., Margheri, M., Bramucci, E., Blankenship, J. C., et al., Effect of supersaturated oxygen delivery on infarct size after percutaneous coronary intervention in acute myocardial infarction, *Circulation: Cardiovascular Interventions*, 2009.

注入することで心室機能を改善できるかどうかを検討したランダム化比較試験である(William et al., 2007). 本試験では, 主要評価項目の 1 つである 14 日後の梗塞面積の左心室に対する割合の中央値[四分位範囲]は, 試験群で 13% [0 – 39], 対照群で 11% [0 – 38] で, 統計学的有意差は認められなかった($p = 0.30$). しかしながら, 事後的なサブグループ解析において心筋梗塞の兆候を認めてから 6 時間以内の患者に限った場合は, 統計的有意差が認められた($p = 0.030$; 図 5 左). この結果を受けて, AMIHOT II 試験(Stone et al., 2009)は, 心筋梗塞の兆候を認めてから 6 時間以内の患者を対象に実施され, 主要評価項目の解析に Bayes 流の階層モデルが用いられた.

試験 h ($h = \text{I}$: AMIHOT I, $h = \text{II}$: AMIHOT II) の, サブグループの水準 j ($j = 1, 2, 3, 4$; $h = \text{II}$ では $j = 1$ のみ)に対する 14 日後の梗塞面積の左心室にする割合の期待値が以下のようにモデル化された.

$$\begin{aligned}\mu_{hj}^C &= \mu_0 + \omega_j^C + \gamma_h^C, \\ \mu_{hj}^T &= \mu_{hj}^C + \delta + \omega_j^T + \gamma_h^T\end{aligned}\quad (25)$$

ここで μ_0 は対照群の全体平均, δ は介入効果の平均を表す. ω_j^C は対照群のサブグループ j ($j = 1, 2$) の効果, γ_h^C は対照群の試験 h ($h = \text{I}, \text{II}$) の効果を表す. ω_j^T と γ_h^T はサブグループおよび試験に対する試験群の上乗せ効果である. 各群における k 番目の患者に対する結果変数 Y_{hjk}^C および Y_{hjk}^T について,

$$\begin{aligned} Y_{hjk}^C &\sim N(\mu_{hj}^C, \sigma_C^2), \\ Y_{hjk}^T &\sim N(\mu_{hj}^T, \sigma_T^2) \end{aligned} \quad (26)$$

が仮定された。各パラメータに関する事前分布は、ほぼ無情報となるように (27) の通り定められた。

$$\begin{aligned} \mu_0 &\sim N(3.2, 0.7^2) & \sigma_0 &\sim N(0, 0.7^2) \\ \omega_j^C &\sim N(0, \varphi_\omega^2) & \gamma_h^C &\sim N(0, \varphi_\gamma^2) \\ \omega_j^T &\sim N(0, \tau_\omega^2) & \gamma_h^T &\sim N(0, \tau_\gamma^2) \\ \varphi_\omega &\sim U(0.01, 0.67) & \varphi_\gamma &\sim U(0.01, 0.1) \\ \tau_\omega &\sim U(0.01, 0.67) & \tau_\gamma &\sim U(0.01, 0.1) \\ \sigma_C &\sim U(0.01, 2) & \sigma_T &\sim U(0.01, 2) \end{aligned} \quad (27)$$

統計的推測は、AMIHOT II 試験 ($h = \text{II}$) における心筋梗塞の兆候を認めてから 6 時間以内の患者 ($j = 1$) に対する群間差、すなわち $\mu_{\text{II}1}^T - \mu_{\text{II}1}^C = \delta_0 + \omega_1^T + \gamma_{\text{II}}^T$ の事後分布に基づいて行われた。サンプルサイズ設計はシミュレーションに基づいて行われ、その結果を参考に試験群：対照群で 2.8 : 1 の不均等なランダム割付が行われた。計画時に行われた検出力計算では、試験群 224 例に対して対照群 80 例のもとで、5.0 % の差を見出すための検出力は階層モデルでは 85.4 % だったのに対し、AMIHOT II 試験のみに基づいた場合は検出力 73 %、AMIHOT I 試験の結果と AMIHOT II 試験の結果を完全にプールした場合の検出力は 93 % であった。一方、群間に差がないという帰無仮説のもとでは、第一種の過誤はそれぞれ 5 %、5 %、17 % であった。

結果として試験群 222 例、対照群 79 例が解析対象となった。AMIHOT I 試験のデータのみ、AMIHOT II 試験のデータのための解析結果および、階層モデルを用いて併合した事後分布をそれぞれ左から図 5 の中央および右側に示す。試験群が対照群を上回っている Bayes 流事後確率は 96.9 % で、有意差が認められた。

この事例は承認申請にも認められた検証試験であり、第一種の過誤を名義水準に保ちながらヒストリカルデータを計画時から効率的に用いることができた 1 つの事例である。

5. Discussion

本論文では、臨床試験においてヒストリカルデータの情報を有効に活用するための方法論について論述し、事例紹介を行った。ヒストリカルコントロールデータを用いて治療効果に対するバイアスのない推測を行うためには、ヒストリカルコントロールデータと新規試験の対照治療群（および新治療群）との高い比較可能性が要求される。また、治療効果の評価方法や患者に対するケアの違いなどもバイアスの原因になり得るため大きな注意を要する。Pocock (1976) は、新規試験の対照治療群の一部として利用するデータを提供可能な既存試験の条件として、以下の 6 つを挙げている。

1. 詳細に定義された、新規試験の対照治療群と同じ治療を受けている。
2. 新規試験と同じ適格基準を用いた最近実施された臨床試験である。

3. 治療の評価方法が新規試験と同じである.
4. 重要な患者背景の分布が新規試験と同様である.
5. 新規試験とほぼ同じ組織で実施されている.
6. 新規試験との結果の違いを生むと予想される他の要因が存在しない. 例えば, 新規試験において症例登録スピードが予想よりも上がらないとき, 適格基準をみたすぎりぎりの症例が登録されることも起こりうる. そういった場合, 新規試験と既存試験で共通の適格基準を用いたとしても登録患者の背景が二つの試験間で異なってくる可能性がある.

Pocock は, 以上の 6 条件をすべて満たせばヒストリカルコントロールデータを新規試験の一部として安全に利用できるが, ひとつでも条件が満たされないときは, 治療法の比較に大きなバイアスが生じる可能性を否定できないと指摘している. Pocock の 6 条件が満たされる例として, 4.8 節で紹介した AMIHOT I・II 試験が挙げられる. しかしながら, 現実的には全ての条件を満たすヒストリカルデータが存在することは稀である. そこで, 全ての条件が厳密に満たされていない状況でヒストリカルデータの情報を活用する有効な方法を考える必要がある. 本論文でまとめた Bayes 流アプローチのうち, 3.4-3.6 節および 4.3 節で紹介した modified power prior および階層モデルは, ヒストリカルデータと新規試験データの試験結果の近さ(類似度)に応じてヒストリカルデータから得られる情報量の大きさを制御する方法であるため, Pocock の条件が厳密に満たされない状況でも過度のバイアスを避けるために有用だと考えられる. 近年, FDA は時間の節約および患者数の削減のために, 医療機器の承認申請においてヒストリカルデータを用いた臨床試験の実施を認めている (Campbell, 2011; Bonangelino et al., 2011).

本論文では, Bayes 流アプローチとして power prior および階層モデルを紹介した. Power prior の利点の 1 つとして, 個々のヒストリカルデータにおける情報量をどの程度割り引いたかがべき乗パラメータの値から容易に理解できる点が挙げられる. 階層モデルにおいては, ヒストリカルデータに与える重みは治療効果に関する分散パラメータで制御される. Bayes 流アプローチ, ヒストリカルデータの割り引きの程度, または事前分布の設定を変えることで事後分布がどう変化するかを感度解析することが必要である. 事前分布の影響は, power prior ではべき乗パラメータ a_0 の事後分布, 階層モデルでは分散パラメータ τ^2 の事後分布により評価できる. Chen and Ibrahim (2006) は, 正規線形モデルにおいては conditional power prior のべき乗パラメータと階層モデルの分散パラメータの間に単調な負の関連がある事を解析的に示した. 彼らは, 一般化線形モデルの場合にも同様の関係を漸近的に示せると論じている. したがって, 少なくとも Chen and Ibrahim (2006) の考えた状況においては, 階層モデルの分散パラメータの値をべき乗パラメータに置き換えることで, 当該データに対するヒストリカルデータの情報を利用する程度について解釈することができる. べき乗パラメータおよび分散パラメータに分布を仮定する場合には両者の関係は解析的に示すことができず, 現在のところ定量的な性能比較も行われていない. 特徴を質的に対比すると, power prior はヒストリカルデータの情報を利用する程度の解釈が容易なこと, 一つの既存試験の利用を中心に発展してきたことが挙げられ, 階層モデルは利用可能な既存試験が複数ある場合を中心に発展してきたことが挙げられる. 今後さらに事例を重ねることで実践上の特徴を明らかにしていくことが期待できる. さらに, 両者のハイブリッド法として, 階層モデルにおい

て、ヒストリカルデータの尤度にべき乗パラメータを乗じることでヒストリカルデータから得られる情報量を調整する方法なども提案されている (Neelon and O'Malley, 2010). また、ヒストリカルデータの情報量を直感的に理解するための他の方法として、prior effective sample size を用いることも提案されている (Morita et al., 2008, 2012; Pennello et al., 2008; Neuenschwander et al., 2010).

ヒストリカルデータの利用が有効な状況として、市販後調査・小児などの臨床試験を行うことが難しい集団、あるいは興味のあるサブグループに対する検討などが考えられる (Wiens et al., 2014). 具体的に、市販後調査では市販前および市販後のデータを統合することが考えられる (Campbell, 2011). また、疾病経過や治療効果などが成人と小児で似ていると考えられる場合は、Pocock の条件を満たしてはいないものの、成人のデータを小児のヒストリカルデータとして用いることができる可能性がある (Schoenfeld et al., 2009; Campbell, 2011). 既存の臨床試験における探索的なサブグループ解析の結果から新しい検証的試験を計画する場合は、既存の臨床試験の結果をヒストリカルデータとして用いることが考えられる (Bonangelino et al., 2011). 複数のサブグループ解析を実施している場合には、デザイン段階で平均値への回帰に対する十分な配慮が必要と考えられるが、4.8 節で紹介した AMIHOT I・II 試験の事例 (Stone et al., 2009) はその示唆に富む例である.

特に医療機器の臨床試験においては、Food and Drug Administration (FDA) のガイダンス (FDA, 2010) で示されているように、ヒストリカルコントロールを用いた Bayes 流アプローチが適している. その理由として、薬剤と違って医療機器は作用のメカニズムが局所的でよく知られており、患者の反応性が予見しやすく、ヒストリカルデータと新規試験データの類似性が見通しがつきやすいことが挙げられる. 例えば、新たな機器が既存の機器にわずかな改良を加えたものであれば、既存の機器の開発と評価に用いた情報は新しい機器の安全性および有効性のある程度予測可能であろう (Pennello and Thompson, 2007).

ヒストリカルデータの情報を Bayes 流に用いる際、データの類似度に応じてヒストリカルデータの重みを決定する動的利用を行わないときは、どの程度ヒストリカルデータの情報を事前分布に反映させるかが決定的に重要な問題となる. 米国の MD アンダーソンがんセンターでは、早期臨床試験においてヒストリカルデータに基づき情報を有した事前分布を設定しているが、その際ヒストリカルデータの情報をある程度割り引いて利用している. 例えば、100 人中 30 人のレスポンスが得られたというヒストリカルデータがあった場合、その 10% 分の情報を有するベータ分布 $Beta(3,7)$ を事前分布として設定する. この例のように、20% 程度の利用上限を目安としている (Biswas et al., 2009). Hueber et al. (2012) は、クローン病患者を対象とした secukinumab の有効性および安全性を評価する臨床試験において、新規試験と適格基準が同様な患者集団を対象とした 6 つの既存の臨床試験から得られたプラセボ群 671 例の結果から、20 例分の情報を持つようにプラセボ群の事前分布を設定した. これにより、新規試験の対象患者 60 例を、secukinumab 群 40 例、プラセボ群 20 例と 2:1 に配分することができ、結果的に secukinumab に関する情報をより多く集めることができた. 事前分布を決定する方法の 1 つとして専門家の意見を利用することが考えられるが (Kass and Wasserman, 1996)、事前分布が専門家の選び方あるいは専門家

の意見を利用する方法に依存するため、客観的な正当化が難しいという問題点が指摘されている (Bonangelino et al., 2011). このため、FDA の医療機器ガイダンス (FDA, 2010) では、専門家の意見によって事前分布を決定する方法を推奨していない。また、prior effective sample size を利用して事前分布の情報量を評価し、適切な情報量を有した事前分布を設定する方法も提案されている (Takeda and Morita, 2015)。

ヒストリカルデータの情報源が複数存在する場合には、新規試験のヒストリカルコントロールとしてどのデータを用いるかの選択も重要になる。Pocock の基準以外にも、非劣性試験で用いるための一致性の評価 (Cuffe, 2011), あるいはメタアナリシスで用いられる試験選択の推奨に従うこと (Gsteigeret et al., 2013) などが提案されている。また、ヒストリカルデータを有効活用するためのデータベースの構築も重要である。Pfizer 社ではプラセボを用いた 1134 の臨床試験のデータをデータベース化している (Desai et al., 2013)。Pfizer 社のみならず、企業間で共通のデータベースをすることによって、より有効な薬剤開発ができるようになるとの指摘がある (Desai et al., 2013)。

実際の臨床試験でヒストリカルデータを用いるためには、計画段階において様々な要因を考慮に入れる必要がある。例えば、用いるヒストリカルデータの情報が、対照治療群の情報だけなのか、治療効果に関する情報も利用できる可能性があるのか、という点を考えなければならない。また、試験実施に必要な資金などの資源の制約、あるいは試験目的 (内部的な意思決定か、承認申請か、論文報告か、など) も考慮する必要がある (Neuenschwander et al., 2010)。Bayes 流アプローチを実際に用いるには、当然ながら研究者側に Bayes 流統計学に精通した統計家が必要である。同時に、ヒストリカルデータあるいは解析手法の選択といった、試験デザインに関してもより高度な専門性が要求される (Gsteigeret et al., 2013)。Bayes 流アプローチを利用した臨床試験結果が規制当局に受け入れられた前例がない場合、承認申請の際に問題が生じる可能性がある (Wiens et al., 2014)。例えば、小児や希少疾患を対象とした臨床試験のように、検証的な試験においてもヒストリカルデータを積極的に活用したい状況も考えられるが、本論文で紹介した方法を検証的な試験で利用する際には、課題の一つとして第一種の過誤の厳密な統制が挙げられる。そのため、Bayes 流アプローチの臨床試験への実装に際しては、規制当局との共通理解が必要である。

本論文では議論していないが、ヒストリカルデータを新規試験データと統合する際、全試験共通で測定した共変量情報を用いることが適切な推測に有用である。Pocock の 4 番目の条件が満たされていない場合、すなわちヒストリカルデータの対象患者の背景と新規試験の対象患者の背景が異なる場合、共変量を考慮しない解析では治療効果の推定値に大きなバイアスが生じる可能性がある。調整解析を行う方法としては、回帰モデルの利用が考えられる (例えば Hobbs et al., 2012)。また、傾向スコア (Rosenbaum and Rubin, 1983) を Bayes 流に利用することも考えられよう。より単純な共変量の利用方法としては、試験群、対照治療群それぞれの期待値に対してではなく、治療効果に関するヒストリカル情報を事前情報として用いることが考えられる。階層モデルを用いる場合、試験治療群および対照治療群それぞれの期待値に関する試験間のバラツキは非常に大きい可能性があるが、治療効果は比較的一定に近い可能性が考えられる。この場合、階層モデルを用いた解析の効率は治療効果に関する交換可能性を仮定したモデルの方が高くなる (Pennello

and Thompson, 2007).

Bayes 流の計算を行う際に利用可能なソフトウェアについて簡単に述べる. マルコフ連鎖モンテカルロ法 (Markov Chain Monte Carlo; MCMC) を実装する統計ソフトは SAS の MCMC プロシジャ, BUGS (Bayesian inference Using Gibbs Sampling) が多く用いられているようである. なお BUGS は現在, WinBUGS の開発が終了し OpenBUGS がオープンソースとして提供されている (Lunn, 2009). BUGS と同様ギブスサンプリングを行う JAGS (Just Another Gibbs Sampler), C++ で実行する STAN といったソフトも使用可能である. いずれも R から呼び出して解析することができる.

本論文で紹介したように, ヒストリカルデータと新規試験データを適切に組み合わせることにより, より効率の良い新治療法開発を行える可能性がある. ヒストリカルデータの情報を安全かつ効果的に用いることができる Bayes 流アプローチのさらなる開発と利用に関する議論が, 今後ますます重要になってくるだろう.

謝 辞

本稿について, 丁寧な査読を通じて貴重なご意見をいただきました査読者および編集者の先生方に心より感謝申し上げます.

参考文献

- Amer-Wåhlin, I., Hellsten, C., Norén, H., Hagberg, H., Herbst, A., Kjellmer, I., et al. (2001). Cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram for intrapartum fetal monitoring: a Swedish randomised controlled trial. *Lancet* **358**, 534–538.
- Biswas, S., Liu, D. D., Lee, J. J. and Berry, D. A. (2009). Bayesian clinical trials at the University of Texas MD Anderson cancer center. *Clinical Trials* **6**, 205–216.
- Bonangelino, P., Irony, T., Liang, S., Li, X., Mukhi, V., Ruan, S., et al. (2011). Bayesian approaches in medical device clinical trials: a discussion with examples in the regulatory setting. *Journal of Biopharmaceutical Statistics* **21**, 938–953.
- Campbell, G. (2011). Bayesian statistics in medical devices: innovation sparked by the FDA. *Journal of Biopharmaceutical Statistics* **21**, 871–887.
- Chen, M. H. and Ibrahim, J. G. (2000). Power prior distributions for regression models. *Statistical Science* **1**, 46–60.
- Chen, M. H. and Ibrahim, J. G. (2006). The relationship between the Power Prior and hierarchical models. *Bayesian Analysis* **1**, 551–574.
- Cuffe, R. L. (2011). The inclusion of historical control data may reduce the power of a confirmatory study. *Statistics in Medicine* **30**, 1329–1338.

- Desai, J. R., Bowen, E. A., Danielson, M. M., Allam, R. R. and Cantor, M. N. (2013). Creation and implementation of a historical controls database from randomized clinical trials. *Journal of the American Medical Informatics Association* **20**, e162–e168.
- Duan, Y., Ye, K. and Smith, E. P. (2006). Evaluating water quality using power priors to incorporate historical information. *Environmetrics* **17**, 95–106.
- Duan, Y., Smith, E. P. and Ye, K. (2006). Using power priors to improve the binomial test of water quality. *Journal of Agricultural, Biological and Environmental Statistics* **11**, 151–168.
- Food and Drug Administration (2010). Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials (released February 5, 2010).
<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm> (accessed Aug 8, 2014)
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**, 515–534.
- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2014). Bayesian Data Analysis. 3rd edition. Chapman & Hall/CC: Boca Raton, Florida.
- Gajewski, B. J. (2010). Comments on ‘A note on the power prior’ by Neuenschwander B, Branson M and Spiegelhalter DJ. *Statistics in Medicine* **29**, 708–710.
- Gsteiger, S., Neuenschwander, B., Mercier, F. and Schmidli, H. (2013). Using historical control information for the design and analysis of clinical trials with overdispersed count data. *Statistics in Medicine* **32**, 3609–3622.
- Hobbs, B. P., Carlin, B. P., Mandrekar, S. J. and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **67**, 1047–1056.
- Hobbs, B. P., Sargent, D. J. and Carlin, B. P. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* **7**, 639–674.
- Hobbs, B. P., Carlin, B. P., and Sargent, D. J. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials* **10**, 430–440.
- Hueber, W., Sands, B. E., Lewitzky, S., Vandemeulebroecke, M., Reinisch, W., Higgins, P. D., et al. (2012). Secukinumab, a human anti-IL-17A monoclonal antibody, for moderate to severe Crohn’s disease: unexpected results of a randomised, double-blind placebo-controlled trial. *Gut* **61**, 1693–1700.
- Ibrahim, J. G. and Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science* **15**, 46–60.

- Ibrahim, J. G., Chen, M. H. and Chu, H. (2012). Bayesian methods in clinical trials: a Bayesian analysis of ECOG trials E1684 and E1690. *BMC Medical Research Methodology* **12**, 183.
- Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association* **91**, 1343–1370.
- 厚生省医薬安全局審査管理課 (1998). 「臨床試験のための統計的原則」について. 医薬審第 1047 号 (平成 10 年 11 月 30 日).
- Lambert, P. C., Sutton, A. J., Burton, P. R. and Abrams, K. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* **24**, 2401–2428.
- Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine* **28**, 3049–3067.
- Morita, S., Thall, P. F. and Mueller, P. (2008). Determining the effective sample size of a parametric prior. *Biometrics* **64**, 595–602.
- Morita, S., Thall, P. F. and Muller, P. (2012). Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis* **7**, 591–614.
- Neelon, B. and O'Malley, J. (2010). Bayesian analysis using power priors with application to pediatric quality of care. *Journal of Biometrics and Biostatistics* **1**, 103.
- Neuenschwander, B., Branson, M. and Spiegelhalter, D. J. (2009). A note on the power prior. *Statistics in Medicine* **28**, 3562–3566.
- Neuenschwander, B., Capkun-Niggli, G., Branson, M. and Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials* **7**, 5–18.
- Ojala, K., Vääräsmäki, M., Mäkilä, K., Valkama, M. and Tekay, A. (2006). A comparison of intrapartum automated fetal electrocardiography and conventional cardiotocography—a randomised controlled study. *BJOG: An International Journal of Obstetrics and Gynaecology* **113**, 419–423.
- Pennello, G. and Thompson, L. (2008). Experience with reviewing Bayesian medical device trials. *Journal of Biopharmaceutical Statistics* **18**, 81–115.
- Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases* **29**, 175–188.
- Rietbergen, C., Klugkist, I., Janssen, K. J., Moons, K. G. and Hoijtink, H. J. (2011). Incorporation of historical data in the analysis of randomized therapeutic trials. *Contemporary Clinical Trials* **32**, 848–855.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.

- Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D. and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, in press.
- Schoenfeld, D. A., Zheng, H. and Finkelstein, D. M. (2009). Bayesian design using adult data to augment pediatric trials. *Clinical Trials* **6**, 297–304.
- Spiegelhalter, D. J. (2001). Bayesian methods for cluster randomized trials with continuous responses. *Statistics in Medicine* **20**, 435–452.
- Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). Bayesian approaches to clinical trials and health-care evaluation. John Wiley & Sons Ltd: Chichester, UK.
- Stone, G. W., Martin, J. L., de Boer, M. J., Margheri, M., Bramucci, E., Blankenship, J. C., et al. (2009). Effect of supersaturated oxygen delivery on infarct size after percutaneous coronary intervention in acute myocardial infarction. *Circulation: Cardiovascular Interventions* **2**, 366–375.
- Takeda, K. and Morita, S. (2015). Incorporating Historical Data in Bayesian Phase I Trial Design: The Caucasian-to-Asian Toxicity Tolerability Problem. *Therapeutic Innovation & Regulatory Science* **49**(1), 93–99.
- 武田 純 (2014). MCMC プロシジャを用いた Normalized Power Prior の実用的な実装. SAS ユーザー総会 2014 論文集, 539–552.
- Turner, R. M., Spiegelhalter, D. J., Smith, G. C. S. and Thompson, S. G. (2009). Bias modelling in evidence synthesis. *Journal of Royal Statistical Society series A* **172**, 21–47.
- Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., et al. (2014). Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceutical Statistics* **13**, 41–54.
- Westerhuis, M. E., Moons, K. G., van Beek, E., Bijvoet, S. M., Drogdrop, A. P., van Geijn, H. P., et al. (2007). A randomised clinical trial on cardiotocography plus fetal blood sampling versus cardiotocography plus ST-analysis of the fetal electrocardiogram (STAN) for intrapartum monitoring. *BMC Pregnancy and Childbirth* **7**, 13.
- Westgate, J., Harris, M., Curnow, J. S. and Greene, K. R. (1993). Plymouth randomized trial of cardiotocogram only versus ST waveform plus cardiotocogram for intrapartum monitoring in 2400 cases. *American Journal of Obstetrics and Gynecology* **169**, 1151–1160.
- Wiens, B. L., Lystig, T. C. and Berry, S. M. (2014). Recent statistical contributions to medical device development. *Therapeutic Innovation & Regulatory Science* **48**, 90–97.