

Individual Participant Dataに基づく メタアナリシス

野間 久史[†]

(受付 2014 年 4 月 30 日 ; 採択 7 月 2 日)

要 旨

Individual Participant Data (IPD) に基づくメタアナリシスは、複数の研究から得られた個人レベルのデータを統合することによって行われるメタアナリシスである。利用可能な最も精確な情報を用いた、メタアナリシスにおける 1 つのゴールドスタンダードであるともいわれる。IPD メタアナリシスは、従来の文献ごとの要約指標に基づくメタアナリシスに比べて、多くの方法論的利点があることが知られており、過去 10 年間ほどで、医学研究におけるレビューの数も大幅に増加している。一方で、近年の方法論的研究によって、IPD メタアナリシスには特有のさまざまなバイアスがあることも報告されるようになってきている。本稿では、近年までに発展した IPD メタアナリシスの方法論的研究のレビューを行い、具体的な事例を交えながら、これらの利点と注意点についての解説を行う。

キーワード：システムティックレビュー、メタアナリシス、Individual Participant Data, Aggregate Data, バイアス。

1. はじめに

メタアナリシス (meta-analysis) とは、過去に行われた臨床試験の結果を統合し、総合的な治療効果の評価を行うための方法論である (Higgins and Green, 2008; Whitehead, 2002)。一般的なメタアナリシスでは、出版された論文で報告されたオッズ比やハザード比などの要約指標 (Aggregate Data; AD) を、適切な統計手法によって統合し、平均的な治療効果の推定が行われるが、Individual Participants Data (IPD) によるメタアナリシスとは「対象となる試験における個人ごとの生データを利用するメタアナリシス」である。IPD メタアナリシスでは、従来から普及している AD によるメタアナリシスに比べて、データを要約することによる情報の損失を避けることができ、より精確な情報に基づく治療効果の推定ができる。実際、Stewart and Parmar (1993) は、IPD メタアナリシスを「メタアナリシスにおける 1 つのゴールドスタンダード (a gold standard for meta-analysis)」であると述べている。近年までの方法論的研究によって、IPD メタアナリシスには、さまざまな方法論的利点があることが明らかになっており、実践においても、過去 10 年間ほどで、IPD によるレビューの数は大きく増加している (Riley et al., 2010)。

しかしながら、近年の方法論的研究では、AD メタアナリシスではあまり考えられないような特有のバイアスが、IPD メタアナリシスには起こる可能性があるという報告もされたりなどしており、IPD メタアナリシスにおいても、その特性を踏まえた慎重な評価を行うことが重要と

[†] 統計数理研究所：〒 190-8562 東京都立川市緑町 10-3

表 1. IPD メタアナリシスのデータのレイアウト. 対象となるすべての試験において, 個人レベルのデータが得られている. Wang et al.(2005)による仮想的なデータ.

試験 ID	個人 ID	年齢 (歳)	性別 (男性 : 1, 女性 : 0)	治療群 (試験薬 : 1, 対照薬 : 0)	治療前の 収縮期血圧 (mmHg)	治療後の 収縮期血圧 (mmHg)
1	1	46	1	1	137	111
1	2	35	1	0	143	133
...
1	1520	62	0	0	209	219
2	1	55	0	1	170	155
2	2	38	1	1	144	139
...
2	368	44	1	0	153	129
3	1	51	1	1	186	166
3	2	39	0	1	201	144
...
3	671	54	0	0	166	141
...
10	1	71	0	1	149	128
10	2	59	1	0	168	169
...
10	978	63	0	1	174	128

なる. 本稿は, 近年までの IPD メタアナリシスの方法論的研究における最新の成果をまとめ, IPD メタアナリシスに関する有益な解説を提供することを目的としている.

本稿の構成は, 以下の通りである. まず, 2 節において, IPD メタアナリシスの統計解析の方法についての解説を行い, 3 節でいくつかの事例を交えながら IPD メタアナリシスの方法論的な利点を紹介する. 続いて, 4 節で, IPD メタアナリシスで注意すべき潜在的なバイアスの問題についての解説を行う. 最後に, 5 節で, 本稿のまとめを行う.

2. IPD メタアナリシスの統計解析

表 1 に, Wang et al.(2005)による IPD メタアナリシスのデータセットの仮想的な例を示した. IPD によるデータは, その名の通り, 対象となるすべての試験において, 個人ごとのデータが, 結果変数, 治療群に関する変数, 共変量などについて得られており, 原則として, 個々の試験の原論文の結果を再現できるだけの生データが得られていることとなる. メタアナリシスでは, この IPD のデータを用いて, 対象となる試験の結果を統合し, 総合的な治療効果を推定することになる.

統計解析の方法は, 大きく分けて 2 つの方法があり, 個人レベルのデータを一旦 AD に要約してから統合を行う 2 段階法(two-stage method)と, 個人レベルのデータを直接モデル化して統合を行う 1 段階法(one-stage method)がある. 本節では, 具体例を交えながら, これらの 2 つの方法についての解説を行う.

2.1 2段階法

2段階法とは、その名の通り、IPDのデータセットから、2段階のステップを経てメタアナリシスを行う方法である。具体的には、以下のような2段階の手順を経る方法となる。

Step 1. 試験ごとのIPDのデータを、個別に解析し、試験ごとのオッズ比・ハザード比などの要約統計量(AD)を算出する。

Step 2. 上記のステップで得られたADを、標準的なメタアナリシスの方法(DerSimonian-Laird法など)によって統合する。

一般的には、個人レベルのデータが、元の解析を再現できる状態で揃っているため、Step 1の過程では、対象となる試験すべてにおいて、標準化した方法での統計解析が可能となる。往々にして、まったく同じデータセットの解析でも、モデルで調整する共変量の組が異なることで、解析結果が大きく異なることはあり得る。加えて、ほとんどすべての臨床試験には、欠測データが含まれており、これをどのように取り扱うかによっても、結果が異なってくる。IPDメタアナリシスでは、個人レベルのデータから、個々の試験のデータを再解析することができるため、これによって、すべての試験で標準化した方法でADを統合することができる。また、Step 2では、標準的なADメタアナリシスの解析手法をすべて用いることができるため、試験間の異質性の評価(Q統計量、 I^2 統計量； Higgins and Thompson, 2002; Higgins et al., 2003)、Funnelプロットによる出版バイアスの評価(Higgins and Green, 2008)などを行うこともできる。また、対象となる試験の中で部分的にIPDが得られなかった場合、ADとIPDが混在したデータセットで統合を行うことになるが、2段階法を用いれば、Step 2の過程にADの試験を加えることで、比較的容易にAD/IPDの統合を行うことができる。共変量の効果の修飾の問題についても、試験ごとの共変量の要約統計量を用いることにより、メタ回帰を用いた評価を行うことができる。ただし、これについては、エコロジカルバイアスと、交互作用検定の検出力の問題などから、次に述べる1段階法での解析が勧められる(詳細については、3節の例3、4を参照)。

2.2 1段階法

2段階法に対して、1段階法とは、途中でIPDのデータをADに要約するという過程を経ずに、直接、個人ごとのデータをモデル化して解析を行うという方法である。IPDを直接モデル化することから、2段階法に比べて、より柔軟なモデルでの解析が可能であり、共変量の影響や交互作用の評価においても、詳細な検討を行うことが可能となる。

具体的な解析方法を、生存時間解析のCox回帰モデル(Cox, 1972)を例として紹介する。まず、すべての試験において、共通のハザード比を仮定した固定効果モデルとして、以下のモデルを考える。

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp(\beta_1 x_{1ij})$$

$\lambda_{ij}(t)$ は、 j 番目の試験($j = 1, 2, \dots, J$)における i 番目の対象者($i = 1, 2, \dots, n_j$)のハザード関数であり、これに対して、試験を層とした層別比例ハザードモデルを仮定する。 x_{1ij} は治療群を表す指示変数、 β_1 はすべての試験を通じて共通の対数ハザード比、 $\lambda_{0j}(t)$ は、 j 番目の試験におけるベースラインハザード関数である。一般的に、IPDメタアナリシスでは、試験ごとに患者背景やセッティング、施設や地域が異なるため、試験ごとのベースラインリスクが異なることが想定される。したがって、この試験ごとの効果を適切にモデル化したもとの解析を行わなくては、治療効果の推定・検定の妥当性が保証されなくなる(Smith et al., 2005)。しかしながら、具体的な解析方法は、多施設臨床試験の統計解析(Glidden and Vittinghoff, 2004; Yamaguchi et al., 2002)とまったく同じであり、これらをそのまま適用すればよいことになる。上記の層別比

例ハザードモデルでは、試験ごとのベースラインハザード関数が異なるモデルを仮定することによって、このベースラインリスクの異質性をモデル化しており、一般的な統計ソフトウェアで、容易に共通ハザード比の推定値を得ることができる。

同様にして、層別比例ハザードモデルの β_1 を変量効果でモデル化することによって(生存時間解析では、変量効果のことを Frailty と呼ぶことが多い)、試験間の異質性を考慮した、DerSimonian-Laird 法 (DerSimonian and Laird, 1986) と同様のモデル化を行うことができる。

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp(\beta_{1j}x_{1ij})$$

$$\beta_{1j} \sim N(\beta_1, \tau^2)$$

先述の固定効果モデルと異なるのは、試験ごとで対数ハザード比 β_{1j} が異なることをモデル化しているという点であり、Frailty モデルによって、これが平均治療効果 β_1 を中心にして分布するという仮定を置く。 τ^2 は、DerSimonian-Laird 法の分散成分に対応するものであり、試験間での治療効果の異質性の大きさを表すパラメータとなる。推定値の計算は、高度な数値計算を要することとなるが、最近では、SAS、R や STATA などの一般的な統計ソフトウェアでも、標準的な解析モジュールとして実装されており、比較的簡便な手順によって解析を行うことができる。

固定効果モデル、変量効果モデルのいずれにおいても、その他の共変量 Z_1, Z_2, \dots, Z_p を同時にモデル化することが可能であり、

$$\lambda_{ij}(t) = \lambda_{0j}(t)\exp(\beta_{1j}x_{1ij} + \gamma_1z_{1ij} + \gamma_2z_{2ij} + \dots + \gamma_pz_{pij})$$

$$\beta_{1j} \sim N(\beta_1, \tau^2)$$

などといったモデルでの解析も、同様に行うことができる。当然ながら、交互作用項をモデル化することも可能である。また、 $\gamma_1, \dots, \gamma_p$ を Frailty でモデル化することも可能である。ここまで紹介した一連の解析方法は、例えば、SAS の PROC PHREG で実行することが可能である。

生存時間解析以外でも、ベースライン効果の異質性を適切にモデル化した解析を用いれば、1 段階法での解析は、比較的容易に実施することができる。試験ごとのベースライン効果を、変量効果でモデル化して解析を行うことも可能であり、例えば、連続量の結果変数には、線形モデル (Higgins et al., 2001)、2 値結果変数には、ロジスティック回帰モデル (Turner et al., 2000)、名義変数には、比例オッズモデル (Whitehead et al., 2001) などによる変量効果モデルでの解析方法が提案されている。これらの計算においても、一般化線形混合モデル (generalized linear mixed model; GLMM) の計算モジュールが、近年になって、多くの統計ソフトウェアに実装されるようになっており (例えば、SAS の PROC GLIMMIX)、比較的容易に解析を行うことが可能となっている。

2.3 事例：てんかんの治療薬のメタアナリシス

ここでは、具体的な事例として、Smith et al. (2005) によるてんかんの治療薬の IPD メタアナリシスの解析例を紹介する。Smith et al. (2005) では、Carbamazepine (CBZ) と Sodium Valproate (SV) を比較した 5 つの試験 ($N = 1225$) の IPD を 1 段階法で統合する解析例が紹介されており、前節で示した Cox 回帰モデルでの解析が行われている。

まず、1 段階法のモデルに基づく、治療効果の異質性の検定では、 $P = 0.02$ ($I^2 = 66\%$) となっており、試験間での治療効果の異質性が示唆される結果となっている。固定効果モデル、Frailty モデルによる単変量モデルでのハザード比の推定の結果は、表 2 の (a), (b) となる。いずれにおいても、有意な結果は得られていないが、SV のほうが若干有効であるという結果になっている。

表 2. Smith et al. (2005)によるてんかんの IPD メタアナリシスの解析結果.

	比例ハザードモデルの推定値 (SE)					
	β_1	τ^2	年齢	病型	log (発作数)	年齢×治療群
単変量モデル						
(a) 固定効果モデル	-0.132 (0.073)	—	—	—	—	—
(b) Frailty モデル	-0.098 (0.125)	0.0484 (0.055)	—	—	—	—
多変量モデル						
(c) 固定効果モデル	0.162 (0.129)	—	0.005 (0.003)	-0.186 (0.088)	-0.192 (0.031)	-0.009 (0.004)
(d) Frailty モデル	0.163 (0.139)	0.006 (0.027)	0.005 (0.003)	-0.185 (0.088)	-0.192 (0.031)	-0.009 (0.004)

また, Frailty モデルの分散成分の推定値は, 0.0484 となっており, 95%予測区間の幅は 0.862 となる. Frailty モデルでの平均治療効果の推定値は, 対数ハザード比のスケールで, -0.098 なので, 試験間の異質性は, これに比べても数倍近く大きいという結果になっている.

試験間の異質性が起こり得るひとつの原因には, 試験ごとの対象者集団の背景要因の相違があり, IPD メタアナリシスの 1 段階法の利点は, これらを直接, モデル化できるという点にある. Smith et al. (2005) は, さまざまな効果の修飾の要因を考慮して, 交互作用項まで含めた候補モデルから, ステップワイズ法などの複数のモデル選択法によって, 適合度の高いモデルを選択している. その結果, 考慮したすべての方法で, 共通のモデルが選択され, 年齢, てんかんの病型, log(発作数), 年齢×治療群(交互作用項)が残ったとされている. この選択されたモデルによる結果を, 表 2 の (c), (d) にそれぞれ示している. Frailty モデルの分散成分は 0.006 と小さくなっており, 調整した共変量・交互作用項によって, 試験間の異質性はかなりの割合で説明できたことが示唆される. 今回のモデルでも, 固定効果モデル, Frailty モデルのいずれにおいても主効果の検定は有意になっていないが, 年齢×治療群の交互作用項についての検定はいずれも有意になっている. つまり, 年齢によって, どちらの薬剤が効かが異なる可能性があることが示唆される. 図 1 に, 単変量モデル, 多変量モデルの Frailty モデルから推定された, 年齢とハザード比の関係を示す. 交互作用を考慮すると, ハザード比が 1 より大きければ CBZ のほうが有効であり, 1 未満であれば SV のほうが有効であるという解釈となる. 概ね 18 歳前後を境に, 治療効果の優劣が逆転するという結果になっている.

この結果は, 当然ながら, 事後的に得られた「仮説」程度のものでしか解釈はできないが, IPD メタアナリシスは, このような探索的な解析にも利用することができる. 一般的に, 個々の臨床試験は「すべての対象者に平均的に有効かどうか」を検証するために設計されており, このような交互作用の評価を十分な検出力で行うことは難しい. しかし, IPD メタアナリシスでは, 個々の臨床試験では十分な検出力による評価が難しい設定であっても, 複数の試験を統合して, より大きな情報量による高い検出力のもとで, 治療と共変量の交互作用の評価を行うことが可能となる (Lambert et al., 2002).

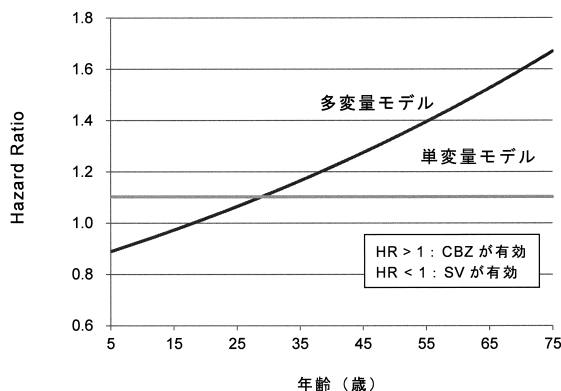


図1. てんかんの IPD メタアナリシスにおける年齢と治療効果の関係(Smith et al., 2005).
年齢・治療の間に、有意な交互作用が検出されており、年齢によって、どちらの治療が有効であるかが異なることが示唆されている。

2.4 統計解析の実際

Simmonds et al.(2005)では、少し古い情報ではあるが、1999年から2001年までに報告された44報のIPDメタアナリシスの文献調査を行っている。このうち、36報(82%)の試験が2段階法を利用しており、1段階法を用いた解析結果を提示しているのは14報(32%)であった。また、1段階法を用いた14報のうち、試験ごとのベースライン効果を見逃した解析を行っているのは3報(21%)、試験間の治療効果の異質性を無視した固定効果モデルによる解析を行っていたのが10報(71%)であった。単純には、この時期には、まだCox回帰のFrailtyモデルやGLMMのソフトウェアが開発の途上であったため、変量効果モデルでの解析を実装するのが難しかったということが理由としては考えられる。近年になって、これらの計算モジュールの普及は急速に進んでいるため、この問題は改善されつつある。

3. IPD メタアナリシスの実践上の利点

従来からのADに基づくメタアナリシスに対して、なぜIPDメタアナリシスが‘gold standard’であるのか、Riley et al.(2010)は、表3のような14の利点をまとめている。従来から知られている公表バイアスに加えて、出版された論文においても、ADには、しばしば統計的に有意な結果や、研究者に都合のよい結果が選択的に報告されるというバイアスがあることが知られており(厳しい査読を経て公表された論文でも、掲載されている結果そのものに、既に、研究者にとって都合のよい主観に基づくバイアスが含まれている可能性がある)、また、論文ごとに結果に提示される情報が異なったり、不十分な情報しか提示されていないことも往々にしてあり得る。統計解析の方法も、対象となる研究を通して、標準化もしくは統一されていることはまずあり得ない(先述の通り、調整される変数の組が異なったり、欠測がある場合の対処法も研究ごとにほぼ確実に異なる)。IPDメタアナリシスでは、こうして生まれるさまざまな潜在的なバイアスを回避することができ、①すべての試験で統計解析の方法を標準化することが可能であること、②個人ごとの生データをもとに詳細な検討が可能であること、③統計的・臨床的な有意性や、個々の研究の実施者の主観に基づく結果の報告バイアスを避けて再解析ができること、④公表された論文で考慮されていないアウトカムについての解析が可能であること、また、長期間の追跡データが利用できる可能性があること、などが代表的な利点であるといえる。

表 3. Riley et al. (2010)による IPD メタアナリシスの 14 の利点.

1. 研究ごとに、一貫した適合基準・除外規準を用いることができる。また、元の解析で除外された対象者を、解析対象集団に含めることができる。
2. 欠測データを、個人レベルで把握することができる。また、これを考慮した解析ができる。
3. 元の研究論文で報告された結果を再解析することで確認することができる（原論文の解析で使われた個人のデータが、IPD メタアナリシスのデータとも対応するものとして）。
4. 元の研究論文で用いられた追跡情報よりも、より長期間の追跡のデータが利用可能となることもある。
5. 対象となる研究間で、重複する対象者集団がある場合には、それを識別することができる。
6. 原論文で報告されていなかった、もしくは、十分に報告されていなかったアウトカムについての結果を計算することができ、解析に含めることもできる。これにより、選択的な結果の報告のバイアスを軽減することができる。
7. 公表されていない研究の結果を、解析に含めることができる（ただし、IPD が入手できた場合）。出版バイアスを軽減することができる。
8. 研究ごとの統計解析の方法を標準化することができる。必要であれば、より適切な方法、あるいは、高度な解析手法を利用することもできる。
9. それぞれの研究で用いられる統計手法やモデルの仮定を評価することができる。例えば、Cox 回帰における比例ハザード性のチェックや、時間依存性効果のような複雑な関連の評価を行うことができる。
10. 原論文の結果が、共変量の調整を行っていない解析しかしていなくても、ベースライン共変量（予後因子）の調整を行った解析を行うことができる。これにより、統計的な検出力を改善できる可能性があり、また、潜在的な交絡要因の調整を行うこともできる。
11. 研究間で共通のベースライン共変量（予後因子）の調整を行うことができる。
12. 特定の患者集団のサブグループにおけるメタアナリシスを行うことができる（例えば、特定の治療の併用を行ったサブグループや、特定のバイオマーカーがある閾値以上にあるサブグループなど）。これにより、個人ごとの治療効果の異質性についての評価を行うことができたり、研究間の異質性を軽減することもできる。
13. 予後モデル（あるいは、リスクスコア）を作製したり、そのバリデーションをしたりすることができる。特に、複数の個人レベルの要因を、その組み合わせまで含めて調べることができる（例えば、複数のバイオマーカー、遺伝的要因、それらの交互作用など）。
14. 複数の時点における結果を用いた解析を行う場合（例えば、経時的繰り返し測定データの解析）、これらの相関を評価することができる。

このような点からも、IPDに基づくメタアナリシスは、利用可能な最も正確な情報を用いることができ、メタアナリシスにおける‘gold standard’であるといえる。また、このようなさまざまなバイアスが結果に影響することを考慮すると、同じ目的で行われたメタアナリシスでも、ADに基づくメタアナリシスとIPDに基づくメタアナリシスでは、まったく異なる結果が得られる可能性も十分に起こり得る。以下では、これらのバイアスの影響が考慮される事例として、Riley et al.(2010)によって紹介された4つの例を紹介する。

例1. McCormack et al.(2004)では、鼠径ヘルニアの外科的治療で、腹腔鏡下手術と開腹手術を比較したランダム化比較試験のIPDメタアナリシスを行っており、腹腔鏡下手術のほうが有意に持続性疼痛のリスクを軽減する(オッズ比:0.54, 95%信頼区間:0.46, 0.64)という報告を行っている。しかし、これよりも早く報告されたADによるメタアナリシスでは、まったく逆の結果が得られており、開腹手術のほうが持続性疼痛のリスクを有意に軽減する(オッズ比:2.03, 95%信頼区間:1.03, 4.01)という報告が行われていた。この結果の矛盾の原因は、早期のメタアナリシスが、当時、公表されていた論文のうち、このアウトカムについての結果が提示されていた3つの試験のデータしか用いてなかったためである。McCormack et al.(2004)では、この他に、実際には17試験のデータが利用可能であったと述べており、これらのデータも追加して、IPDメタアナリシスを行っている。いわゆる結果の選択的な報告バイアスが大きな影響を与えた事例といえる。また、IPDによる再解析の結果、1つの試験では、公表された結果とは大きく異なる結果が得られたとも報告されている。

例2. Jeng et al.(1995)では、父系の白血球免疫療法によって再発性流産のリスクが低減できるかどうかを調べたメタアナリシスにおいて、ADとIPDに基づくメタアナリシスの比較を行っている。生児出生におけるリスク比は、出版された4試験の論文の結果に基づくADのメタアナリシスでは1.29(95%信頼区間:1.03, 1.60)となっており、有意な結果が得られていた。一方、すべての試験で標準化された解析を行ったIPDメタアナリシスでは、これよりも控えめの結果しか得られず、1.17(95%信頼区間:0.97, 1.37)となっており、有意な結果も得られていなかった。このメタアナリシスでは、公表された論文において、一部、解析に含まれなかった対象者がいたとされており、Jeng et al.(1995)のIPDメタアナリシスでは、これを加えた再解析が行われている。上記の結果の相違は、この対象者集団の不一致が原因であった。また、この他にも出版されていない4つのUnpublishedな試験が存在しており、これをIPDメタアナリシスで統合すると、1.01(95%信頼区間:0.74, 1.28)となったという。典型的な出版バイアスがあったことも示唆されている。

例3. Berlin et al.(2002)は、腎移植患者における抗リンパ球抗体療法の有効性を評価した5つの試験のIPDメタアナリシスを行っている。このIPDメタアナリシスでは、効果の修飾の評価において、パネル反応性抗体が20%以上増加している患者とそうでない患者で、治療効果が異なるかどうかの評価を行っており、このサブグループ間で、強い効果の修飾があったことが示唆されている(対数オッズ比の差:-1.33, $P=0.01$)。一方で、ADメタアナリシスにおけるメタ回帰では、対数オッズ比の差は-0.014($P=0.68$)となり、まったく有意な結果は得られていなかった。原因のひとつとして考えられるのが、IPDレベルでの交互作用検定と、ADレベルでのメタ回帰の検出力の違いである。

Lambert et al.(2002)は、この検出力の差異の問題を、広範な設定のもとでのシミュレーション実験によって評価しており、類似した設定($N=5$)のもとで、IPDレベルでの交互作用検定の検出力は90.8%、ADレベルでのメタ回帰では14.8%と、大きく異なる結果が得られたと報告している。これは、単純には、IPDレベルの交互作用検定が、数百~数千人(あるいは、それ以

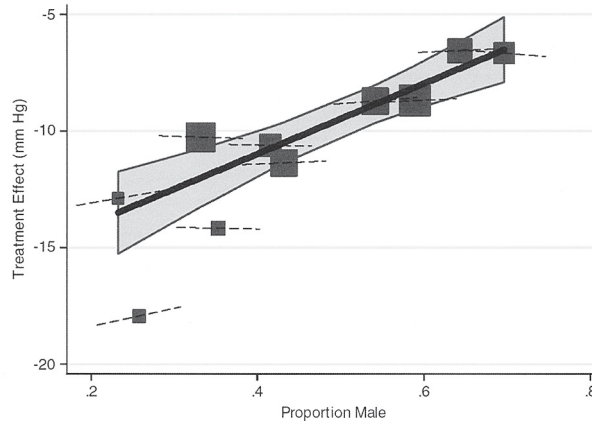


図 2. AD に基づくメタ回帰におけるエコロジカルバイアスの例。Riley et al.(2008, p.1882)より転載。太線がメタ回帰による回帰直線、各試験のプロットに付記されている破線は、試験ごとの IPD による交互作用の評価の推定値。

上)の対象者の個々人の共変量をもとにした検定を行っているのに対して、メタ回帰は、試験ごとに要約された「パネル反応性抗体が20%以上増加している患者の割合」についての関連を評価しているだけであるからと考えられる(この事例でも、ただだか $N = 5$ である)。このように、IPD メタアナリシスでは、1段階法を用いることによって、個々の患者が持つ特性が治療効果をどのように修飾するかについて、高い精度による評価が可能となる(Thompson and Higgins, 2005)。これによって、例えば、どのような特性を持つ患者のサブグループに大きな治療効果が期待できるかなどといった、探索的な評価も可能となる。

例 4. Riley et al.(2008)は、降圧薬の臨床試験について、10 試験の結果を統合したメタアナリシスを行っている。男性の割合と収縮期血圧についての平均的な降圧効果についてのメタ回帰を行ったところ、女性のほうが 15.10 mmHg (95%信頼区間: $8.78, 21.41$) も平均的な降圧効果が大きいという結果が得られた(図 2)。しかし、IPD メタアナリシスの1段階法で、男女間の降圧効果の差についての交互作用の評価を行うと、 0.89 mmHg (95%信頼区間: $0.07, 1.30$) という結果となり、統計的にはかろうじて有意ではあるものの、臨床的には意味のない差しか認められなかった。個々の試験ごとに、性別と降圧効果についての交互作用を評価したものが、図 2 の試験ごとのプロットに示されている「破線の傾き」であり、これによると、ほとんどすべての試験において、個人レベルでは、性別と降圧効果の間に差がなかったという結果になっている。これは、いわゆるエコロジカルバイアス(ecological bias)の典型的な例である。そもそもメタ回帰は「試験ごとの『収縮期血圧の平均値の差』と『男性の割合』についての相関」を評価しているだけであり、これは「個人レベルの降圧効果の性別間の差」を評価しているわけではない。メタ回帰の結果が、実際に治療効果と共変量の交互作用を反映したものとなるには、この2つが一致するということが前提となるが、それをデータから確認する術はなく(少なくとも、ランダムなメカニズムで部分的な IPD が得られていなくては)、また、一般的に、これらが一致しているという保証はない。

この他にも、IPD を用いることにより、複数の個人レベルでの要因が与える影響の評価が可能であり、例えば、バイオマーカーや遺伝的要因の組み合わせによる臨床アウトカムの予測モデルの評価なども可能である(Ioannidis et al., 2002)。近年、議論になっている、治療が有効な

ターゲットとなるサブグループの探索など(stratified medicine; Trusheim et al., 2007)にも活用できる可能性がある。また、個人レベルでのベースライン共変量を調整することによって、治療効果についての検定の検出力を改善することができる可能性もある(Hernandez et al., 2006)。観察研究のメタアナリシスにおいては、交絡要因の調整を行う上でもメリットがある。

複数の研究のデータを用いた、予後モデルやリスクスコアの構築についての検討も行われている(Altman et al., 2009)。IPDに基づく情報があれば、複数の試験にまたがって、モデルの構築が可能であり、また、モデルのバリデーションを行う際にも、古典的な Split-Sample 法や Cross-Validation 法とは異なり、対象となる研究の一部によって予後モデルを構築して、残りの研究のデータによって予測精度を評価するという方法を採用することもできる(Royston et al., 2004)。これにより、従来の方法では難しかった、外的妥当性を考慮した評価も可能となる。IPD メタアナリシスによる予後モデルの構築については、Ahmed et al.(2014)による総説が参考になる。

4. IPD メタアナリシスにおける潜在的なバイアス

3節でも述べてきた通り、IPDに基づくメタアナリシスは、利用可能な最も正確な情報を用いることができる、メタアナリシスにおけるひとつの‘gold standard’といえる。しかし、当然ながら、従来のADに基づくメタアナリシスと同様に、さまざまな潜在的なバイアスのリスクが存在する。これらを考慮しなくては、誤った結果の解釈を生んでしまう可能性がある。本節では、このバイアスの問題についての解説を行う。

4.1 出版バイアス(publication bias)

IPD メタアナリシスでも、従来のメタアナリシスと同じく、出版バイアスの影響は起こり得る。PubMedなどの文献データベースに登録されているような、英語圏で一定以上の水準にあるジャーナルに出版された論文であれば、ほぼ間違いなく文献検索で候補に挙がることになるが、そうでない「グレイ」な文献(結果が公表されなかった試験、英語以外の言語によるジャーナルの論文、会議の抄録・Book Chapter・Letterにのみ報告されている結果など)については、最善を尽くして系統的な文献検索を行ったとしても、候補に挙がってこない可能性がある。システマティックレビューは、これらのすべてを含めることが原則であり、そうでない場合、IPDメタアナリシスでも、治療効果の推定には、バイアスが起る可能性がある(一般的には、治療効果を過大評価する方向へのバイアスが起る)。

Burdett et al.(2003)による文献レビューでは、対象となった11報のIPDメタアナリシスのうち、9報がこの「グレイ」な文献を加えることで、治療効果の推定値が弱まったという報告がされており、出版バイアスの影響が示唆されている。同様に、Ahmed et al.(2012)でも、2007年から2009年に報告された31報のIPDメタアナリシスの文献調査が行われており、このうち、主要な解析にグレイな文献まで考慮していたのは9報(29%)のみであったとされている。また、論文中で、公表バイアスの評価や考察を行っていたのは、31報中10報(32.2%)のみであったという。

4.2 データの利用可能性に関するバイアス(data availability bias)

IPDメタアナリシスでは、対象となるすべての試験において、結果を再現できるだけの個人レベルのデータセットが入手できることが前提となるが、そう安々と個々の試験の実施者がデータの提供に応じてくれるはずがない。IPDが入手できない試験があれば、「個人レベルのデータにアクセスできない試験」が、部分的に生じてしまうこととなり、このIPDの利用可能性が、試験の結果と関連するものとなると、当然ながら、治療効果の推定にもバイアスが生じること

になる。IPD が提供されない理由には、①試験のデータが破棄された、もしくは、なんらかの理由で失われた、②論文の著者への連絡がとれない(所属機関の変更があったり、退職・死亡しているなど)、③なんらかの理由でデータが送付できない、④共同研究への参加拒否(データ提供の拒否)などといったものが考えられる。データ利用可能性バイアスは、試験の結果が、実施者にとって望ましいものであった場合にも、望ましくないものであった場合にも起こり得るため、出版バイアスとは異なり、バイアスの方向さえも予測がつかないという難点がある。

Riley et al.(2007)による文献調査では、199 報のメタアナリシスのうち、102 報(58%)では、対象となった試験の 90%以上から IPD が入手できていたが、51 報(29%)では、IPD が入手できた試験が 80%未満であったとされている。同様に、Ahmed et al.(2012)では、2007 年から 2009 年に発表された 30 報の IPD メタアナリシスのうち、16 報(53%)が、対象となった試験から部分的にしか IPD を入手できていなかったと報告しており、そのうち、10 報(33%)では、IPD が入手できた試験が 80%未満であったと述べている。これらの実態からもわかる通り、すべての対象となる試験から、IPD を揃えることはそもそも困難である。

対象となるすべての試験において IPD が揃わない場合、次善の策として、IPD が得られなかった試験の結果が最終的な結果にどのような影響を与えるか、感度解析を行うべきである。IPD が得られなかった試験についても、実際には、公表された論文などに AD の結果は報告されていることがあり(グレイな文献であっても、会議報告などで入手できることがある)、2 段階法を用いることで、IPD データに基づく結果と統合した解析を行うことはできる。感度解析として、これらの試験を含めた場合の結果が、どのような影響を受けるか、確認しておくことは重要である。また、技術的には高度なものとなるが、Riley et al.(2008)は、IPD と AD を統合するためのメタアナリシスの方法を提案している。

4.3 選択バイアス(selection bias)

IPD メタアナリシスでは、先述の通り、IPD の入手がそもそも難しいということに関連して、はじめから、研究者が IPD の入手が見込まれる範囲に限定して、データの収集が行われることがある。しかし、これは大きなバイアスの原因になる可能性がある。出版バイアスの項でも述べた通り、システマティックレビューの原則は、実際に行われたすべての試験を対象とすることであり、メタアナリシスは、単に「既存のデータを統合する」ための統計手法ではない。研究者が、自身の知る範囲での研究仲間や知人、コンタクトがとれる範囲のみに、データの収集範囲を限定してしまうと、それだけで、恣意的な選択バイアスが入ってしまう。この場合、仮に、試験の結果を統合した解析を行ったとしても、そもそもシステマティックレビューとは言えない。

選択バイアスも、極めて複雑な要素が組み合わさるため、どのような影響が出るか、まったく予測不能である。バイアスの要因としては、研究者の関心の範囲、その領域での知識、コンタクトのとれる範囲、共同研究者、また、研究者の主観的な意見にも左右されることとなる。Ahmed et al.(2012)の文献調査によると、31 報の IPD メタアナリシスのうち、9 報(29%)に系統的な文献検索がされていたかどうか明記されておらず、恣意的な研究の選択がされたリスクがあったという。IPD メタアナリシスでも、システマティックレビューの原則に則った、系統的な文献検索を怠ってはならない。

4.4 事例：中悪性度リンパ腫のメタアナリシス

Greb et al.(2008)では、中悪性度リンパ腫の第一選択療法の一部として、高用量化学療法・自家幹細胞移植が、患者の生存を改善するかについての IPD メタアナリシスが行われている。対象となったのは、高用量化学療法と従来の化学療法を比較した 15 試験となっている。このメタアナリシスを事例に、ここまで述べたバイアスの影響についての検討を行う。

まず、選択バイアスについては、対象の15試験は、系統的な文献検索を経て選ばれており、恣意的な選択によるバイアスは含まれていない。次に、出版バイアスについても、グレイな文献についての調査が適切に行われており、実施されたすべての試験が出版されていたことが確認されている。問題なのは、データ利用可能性についてのバイアスで、このメタアナリシスでも、15試験中10試験(67%)しかIPDが入手できなかったとされている。IPDが入手できた10試験では、ハザード比が1.14(95%信頼区間:0.98, 1.34)となっており、高用量化学療法が少し死亡のリスクを高めてしまうという結果になっている。一方で、IPDが入手できなかった試験の影響を評価するため、可能な限りADの情報を収集したところ、4試験(29%)からハザード比の推定値が得られている。図3に、これらの試験の結果についてのForestプロットを示している。視覚的にも明らかであるが、IPDが入手できなかった4試験の結果は、IPDが得られた10試験とはまったく異なる傾向が見て取れる。要約ハザード比も0.86(95%信頼区間:0.69, 1.09)となっており、高用量化学療法のほうが少し有効であるという結果となっている。つまり、このメタアナリシスでは、結果に依存した偏りのあるバイアスが生じてしまっている。両者を統合した場合にも、ハザード比は1.05(95%信頼区間:0.92, 1.19)となり、全体としての結果は、ほとんど治療効果に差がないことを示唆している。先述の通り、IPDがすべての試験で入手できなかった場合にも、このようなバイアスが生じてしまう可能性があるため、最善を尽くして感度解析を行い、結論に影響し得る重要な結果が出た場合には、必ず報告されるべきである。

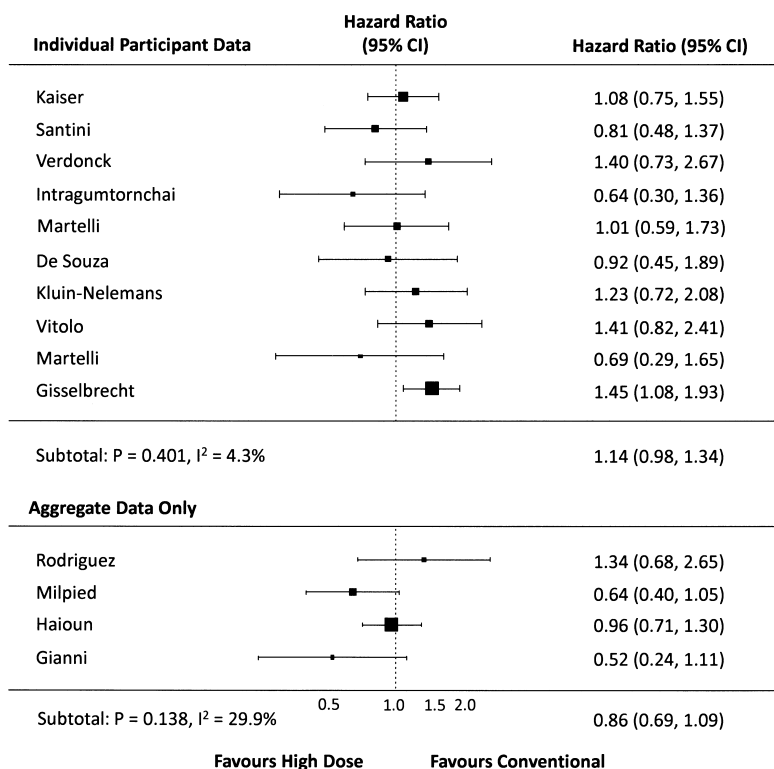


図3. Greb et al. (2008)による中悪性度リンパ腫のメタアナリシス。

5. まとめ

IPD メタアナリシスは、本稿でも紹介してきた通り、利用可能な最も精確な情報を用いた、メタアナリシスにおける‘gold standard’として、今後とも実践で用いられていくことになると思われる。加えて、治療と共変量の交互作用の評価や、予後モデルの構築など、探索的な解析も含めて、応用の範囲も広まっていくことが期待できる。ただし、4節で述べた通り、さまざまな要因によるバイアスは避けられないため、適切な防止策と評価方法を用いて、十分な検討を行う必要がある。

コクラン共同計画でも、IPD メタアナリシスについての方法論のグループができており (<http://ipdmamg.cochrane.org/welcome-ipd-meta-analysis-methods-group>), ホームページ上に、解析プログラムや研究資源などが公開されている。昨今の臨床試験データ公開に関わる議論などからも、IPD メタアナリシスは、将来的にさらに大きな重要性を持つ研究手法となることも予想される。

謝 辞

本稿を執筆するにあたって、総合研究大学院大学複合科学研究科の米岡大輔氏には、多くの有益なコメントをいただきました。ここに、謝意を表します。本解説を執筆するにあたり、科学研究費補助金(課題番号: 26670314)の助成を受けました。

参 考 文 献

- Ahmed, I., Sutton, A. J. and Riley, R. D. (2012). Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: A database survey, *British Medical Journal*, **344**, d7762.
- Ahmed, I., Debray, T. P., Moons, K. G. and Riley, R. D. (2014). Developing and validating risk prediction models in an individual participant data meta-analysis, *BMC Medical Research Methodology*, **14**, 3.
- Altman, D. G., Vergouwe, Y., Royston, P. and Moons, K. G. (2009). Prognosis and prognostic research: Validating a prognostic model, *British Medical Journal*, **338**, b605.
- Berlin, J. A., Santanna, J., Schmid, C. H., Szczech, L. A. and Feldman, H. I. (2002). Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: Ecological bias rears its ugly head, *Statistics in Medicine*, **21**, 371–387.
- Burdett, S., Stewart, L. A. and Tierney, J. F. (2003). Publication bias and meta-analyses: A practical example, *International Journal of Technology Assessment in Health Care*, **19**, 129–134.
- Cox, D. R. (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.
- DerSimonian, R. and Larid, N. M. (1986). Meta-analysis in clinical trials, *Controlled Clinical Trials*, **7**, 177–188.
- Glidden, D. V. and Vittinghoff, E. (2004). Modelling clustered survival data from multicentre clinical trials, *Statistics in Medicine*, **23**, 369–388.
- Greb, A., Bohlius, J., Schiefer, D., Schwarzer, G., Schulz, H. and Engert, A. (2008). High-dose chemotherapy with autologous stem cell transplantation in the first line treatment of aggressive non-Hodgkin lymphoma (NHL) in adults, *Cochrane Database of Systematic Reviews*, CD004024.
- Hernandez, A. V., Eijkemans, M. J. and Steyerberg, E. W. (2006). Randomized controlled trials with

- time-to-event outcomes: How much does prespecified covariate adjustment increase power?, *Annals of Epidemiology*, **16**, 41–48.
- Higgins, J. P. and Green, S. (eds.) (2008). *Cochrane Handbook for Systematic Reviews of Interventions*, Wiley-Blackwell, Chichester.
- Higgins, J. P. and Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis, *Statistics in Medicine*, **21**, 1539–1558.
- Higgins, J. P., Whitehead, A., Turner, R. M., Omar, R. Z. and Thompson, S. G. (2001). Meta-analysis of continuous outcome data from individual patients, *Statistics in Medicine*, **20**, 2219–2241.
- Higgins, J. P., Thompson, S. G., Deeks, J. J. and Altman, D. G. (2003). Measuring inconsistency in meta-analyses, *British Medical Journal*, **327**, 557–560.
- Ioannidis, J. P., Rosenberg, P. S., Goedert, J. J. and O'Brien, T. R. (2002). Commentary: Meta-analysis of individual participants' data in genetic epidemiology, *American Journal of Epidemiology*, **156**, 204–210.
- Jeng, G. T., Scott, J. R. and Burmeister, L. F. (1995). A comparison of meta-analytic results using literature vs individual patient data. Paternal cell immunization for recurrent miscarriage, *Journal of the American Medical Association*, **274**, 830–836.
- Lambert, P. C., Sutton, A. J., Abrams, K. R. and Jones, D. R. (2002). A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis, *Journal of Clinical Epidemiology*, **55**, 86–94.
- McCormack, K., Grant, A., Scott, N. and EU Hernia Trialists Collaboration (2004). Value of updating a systematic review in surgery using individual patient data, *British Journal of Surgery*, **91**, 495–499.
- Riley, R. D., Simmonds, M. C. and Look, M. P. (2007). Evidence synthesis combining individual patient data and aggregate data: A systematic review identified current practice and possible methods, *Journal of Clinical Epidemiology*, **60**, 431–439.
- Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L. and Bouillon-Boutitie, F. (2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data, *Statistics in Medicine*, **27**, 1870–1893.
- Riley, R. D., Lambert, P. C. and Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting, *British Medical Journal*, **340**, c221.
- Royston, P., Parmar, M. K. and Sylvester, R. (2004). Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer, *Statistics in Medicine*, **23**, 907–926.
- Simmonds, M. C., Higgins, J. P., Stewart, L. A., Tierney, J. F., Clarke, M. J. and Thompson, S. G. (2005). Meta-analysis of individual patient data from randomized trials: A review of methods used in practice, *Clinical Trials*, **2**, 209–217.
- Smith, C. T., Williamson, P. R. and Marson, A. G. (2005). Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes, *Statistics in Medicine*, **24**, 1307–1319.
- Stewart, L. A. and Parmar, M. K. (1993). Meta-analysis of the literature or of individual patient data: Is there a difference?, *Lancet*, **341**, 418–422.
- Thompson, S. G. and Higgins, J. P. (2005). Treating individuals 4: Can meta-analysis help target interventions at individuals most likely to benefit?, *Lancet*, **365**, 341–346.
- Trusheim, M. R., Berndt, E. R. and Douglas, F. L. (2007). Stratified medicine: Strategic and economic implications of combining drugs and clinical biomarkers, *Nature Reviews Drug Discovery*, **6**, 287–293.
- Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H. and Thompson, S. G. (2000). A multilevel model framework for meta-analysis of clinical trials with binary outcomes, *Statistics in Medicine*, **19**,

3417–3432.

- Wang, J. G., Staessen, J. A., Franklin, S. S., Fagard, R. and Gueyffier, F. (2005). Systolic and diastolic blood pressure lowering as determinants of cardiovascular outcome, *Hypertension*, **45**, 907–913.
- Whitehead, A. (2002). *Meta-analysis of Controlled Clinical Trials*, John Wiley & Sons, Chichester.
- Whitehead, A., Omar, R. Z., Higgins, J. P., Savaluny, E., Turner, R. M. and Thompson, S. G. (2001). Meta-analysis of ordinal outcomes using individual patient data, *Statistics in Medicine*, **20**, 2243–2260.
- Yamaguchi, T., Ohashi, Y. and Matsuyama, Y. (2002). Proportional hazards models with random effects to examine centre effects in multicentre cancer clinical trials, *Statistical Methods in Medical Research*, **11**, 221–236.

Meta-analysis Based on Individual Participant Data

Hisashi Noma

The Institute of Statistical Mathematics

Meta-analysis based on individual participant data (IPD) involves obtaining and then synthesizing raw individual level data from multiple studies. IPD meta-analysis uses the most precise available information and is called a ‘gold standard’ of meta-analysis. The use of individual participant data has various advantages in both statistical and clinical aspects, and has been increasingly applied in systematic reviews in the past 10 years. However, recent methodological researches have pointed out that several particular biases are involved in the IPD reviews. This article provides a comprehensive review for statistical methodologies and several meaningful applications of IPD meta-analysis.