

## 科学研究費助成事業 研究成果報告書

平成 27 年 6 月 3 日現在

機関番号：13103

研究種目：若手研究(B)

研究期間：2012～2014

課題番号：24730530

研究課題名(和文)階層的線形モデルを用いた単一事例実験のメタ分析における公表バイアスに関する研究

研究課題名(英文)Publication bias in meta-analysis of single-case research designs based on hierarchical linear models.

研究代表者

奥村 太一 (Okumura, Taichi)

上越教育大学・学校教育研究科(研究院)・准教授

研究者番号：90547035

交付決定額(研究期間全体)：(直接経費) 2,300,000円

研究成果の概要(和文)：本研究では、階層的線形モデル(HLM)によって単一事例実験によって得られたデータをメタ分析した場合、公表バイアスがどのような要因によってどの程度発生するかを検証した。その結果、明らかになったことは以下の3点である。(1)国内で発表された単一事例実験の効果量はいずれも大きく歪んで分布しており、公表バイアスの存在が疑われる。(2)処置前値によって対象者をスクリーニングした場合、平均への回帰によって本来は存在しない処置効果が検出されることがしばしばある。(3)統計的に有意かどうかによって公表の有無が左右されていた場合、メタ分析によって本来は効果のない処置が相当高い効果があると評価される可能性がある。

研究成果の概要(英文)：This study examined the factors and degrees of publication bias when single-case research design data are synthesized by the hierarchical linear models (HLM). Results can be summarized into the following three points: (1) Distributions of effect sizes of single-case designs that were reported in Japan were distorted, implying that the publication biases existed in these studies. (2) When subjects are screened by their pretest scores, treatment effects tend to be overestimated by the regression to the mean effect, and even non-effective treatments can be evaluated as largely effective. (3) When publications of the single-case studies depend on the statistical significance of the treatments and only published cases are used for meta-analyses by the HLM, originally non-effective treatments can be evaluated as if they were considerably effective ones.

研究分野：心理統計学

キーワード：階層的線形モデル 単一事例実験 メタ分析

1. 研究開始当初の背景

(1) 教育心理学と単一事例実験

教育心理学の実験的研究は、複数の被験者を対象として行われることが一般的である。一方で、一人の被験者を対象として行われる単一事例実験も介入の効果を検証するために有用な手法として利用されてきた (Kratochwill, 1985; Wacker, Steege, & Berg, 1988)。単一事例実験の特徴は、個人内の変化をその事例の特殊性と絡めて詳細に記述できることである。

(2) メタ分析による単一事例実験の統合

単一事例実験に対しては、介入に伴う個人の変化に関する詳細な記述ができる一方で、その介入のもたらす効果の全体像を描くことが難しいという、結果の一般化可能性に関する批判がある。

近年、こうした問題を解決するため、メタ分析によって過去に公表された単一事例実験の結果を統計的に統合するための方法論が提案され、活発に議論されるようになった。例えば、国内では高橋・山田・小笠原 (2009) が過去 40 年間の「特殊教育学研究」誌に掲載された単一事例実験の結果をメタ分析の手法にもとづいて統合する試みを行っている。

こうした手法の中でも、近年特に注目されてきているのが、階層的線形モデル (HLM: Hierarchical Linear Models) を応用したメタ分析手法である (Sheu & Suzuki, 2001; Van den Noortgate, & Onghena, 2003)。HLM は、もともと学級など集団単位で得られたデータや個人内変化の個人差を分析するのに有効な手法として、多くの教育心理学の研究で利用されてきたものである (Raudenbush & Bryk, 2002)。

HLM を応用した単一事例実験のメタ分析手法は、以下の三点において有用である。第一に、事例によってなぜ介入の効果が違いが生じるのか、交互作用を生じさせる調整要因を統計的に検証できる。第二に、観測時点が比較的少なくとも、(ある程度の事例数が確保されていれば) 繰り返し測定に伴うデータの時系列性を考慮できる。第三に、モデルのデータに対する当てはまりの程度を統計的に評価できる

(3) 単一事例実験と公表バイアス

HLM を応用した単一事例実験のメタ分析は優れた手法であるが、解決すべき問題点がいくつか残されている。最も重要なものの一つに、公表バイアス (publication bias) の問題がある。これは、実施された研究のうち公表されたもののみを統合した結果が、公表されていないものも含めた全研究を統合した場合の結果と食い違う現象のことである。

一般に、統計的に有意な結果が得られている研究はそうでない研究に比べ公表されやすい傾向がある (丹後, 2002)。従って、公

表された研究結果のみを対象にメタ分析を行うと、その結果はより研究者にとって望ましい方向に偏ることになる。

単一事例実験のメタ分析における公表バイアスには、有意な結果が得られたかどうか以外にも、事例研究ならではの要因も考えられる。一般的に、事例研究は研究者がその特定の事例に興味を持ったからこそ、あるいは介入する必要性を感じたからこそ研究の対象とされるものである (Van den Noortgate & Onghena, 2003)。従って、単一事例実験のメタ分析においては、事例がランダムに選ばれるという一般的に想定されている前提が成り立っているとは考えにくい。

メタ分析における公表バイアスを出したりその影響を補正したりする方法は、これまでも提案されてきている (Rothstein, Sutton, & Borenstein, 2005)。また、研究間の非独立性をモデル化する方法についても検討が行われてきている (Stevens & Taylor, 2009)。しかし、これらはいずれも複数の被験者に対して行われた研究のメタ分析に関するものであって、単一事例実験に関するものではない。

2. 研究の目的

本研究の目的は、教育心理学の事例研究で実施される単一事例実験について、メタ分析により結果を統合する際の公表バイアスの影響を評価することである。メタ分析については、すでに述べたように HLM を用いて実施することを想定している。

本研究期間中は、特に次の 3 つの点に注目して検証を実施することにした。

- (1) わが国で過去に行われた単一事例実験が、実際に公表バイアスの影響を受けているか検証する (研究 1)
- (2) 研究者にとって興味のある事例や介入を特に必要とする事例が特にピックアップされて公表されるといった事例研究に特有の現象が、公表バイアスをどの程度生じさせるのか検証する (研究 2)
- (3) 各事例において算出された効果量が有意であるかどうかによって公表の是非が左右される傾向にあるとした場合、HLM によるメタ分析で公表バイアスがどの程度生じるのか検証する (研究 3)

3. 研究の方法

(1) 研究 1: わが国で行われた単一事例実験における公表バイアスの検証

国内では、高橋・山田・小笠原 (2009) が「特殊教育学研究」誌に掲載された過去 40 年間の単一事例実験の結果が、種々の効果量とともに報告されている。これらの結果には、デザイン (ABAB、多重ベースライン、処置交替など) や障害の種類、標的行動、介入方法などに関して様々なものが混在している。しかし、もし公表バイアスが存在しないのであれば、これらの効果量の分布はおおむね左右

対称となるであろうことが予想される。

そこで、研究 1 では高橋・山田・小笠原 (2009) らの報告結果から効果量に関する情報を取り出して検討することとした。高橋・山田・小笠原 (2009) では、各事例につき 3 つの効果量を算出している (PND、ES\_BS2、ES\_C) ため、これら全てを検討の対象とすることとした。また、多重ベースラインデザインが他と区別してコーディングされていることから、全体での分布と多重ベースラインデザインによる研究のみを取り上げた場合の分布を比較することとした。

(2) 研究 2: 介入する価値や必要性のある事例のみが公表された場合のバイアスの検証

#### ① 平均への回帰

事例研究にとって介入する価値や必要性を左右する要因の一つは、その事例のベースラインにおける問題の深刻さの度合いであると考えられる。仮にある望ましい行動を介入によって増やしたいと考えるのであれば、多数の潜在的な対象者の中から、そうした行動が特に少ない者をスクリーニングして実践の対象とすることが現実的であろうし、また倫理的にも重要であろうと考えられる。そこで、本研究では、処置前値によってスクリーニングがなされ、介入対象となった事例のみが公表される状況にあった場合、介入効果の評価にどのようなバイアスが生じるか、シミュレーションにより検証することとした。言うまでもなく、このバイアスを生じさせる原因の一つはいわゆる「平均への回帰」と呼ばれる現象である。

#### ② シミュレーションモデル

ここでは、AB デザインにもとづき、被験者  $i$  の処置前値  $X_i$  と処置後値  $Y_i$  が、

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_{Xi} \\ \mu_{Yi} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix} \right)$$
$$\begin{pmatrix} \mu_{Xi} \\ \mu_{Yi} \end{pmatrix} \sim N \left( \begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} \tau^2 & 0 \\ 0 & \tau^2 \end{pmatrix} \right)$$

なる単純な正規分布に従うモデルを設定した。シミュレーションにおいては、処置前値と処置後値の全体平均を  $\mu = 50$  と設定し、個人内分散  $\sigma^2$  と個人間分散  $\tau^2$  の合計を 100 とした。その上で、 $X_i$  と  $Y_i$  に共通の級内相関係数 (ICC) ( $= \tau^2 / (\sigma^2 + \tau^2)$ ) が 0.2 もしくは 0.5 となるよう分散を 2 つに分割した。また、 $X_i$  と  $Y_i$  の相関係数  $\rho$  については、0.2、0.5、0.8 の 3 条件を設けた。

対象者の選抜は処置前値にもとづいて行われ、その平均が  $\bar{X}_i < c$  を満たす個人が最終的に選抜されるものとした (比較のために選抜が行われない条件も設けた)。選抜の基準値は  $c = 60$  とした。選抜が行われない完全データについては、その規模を 50 人  $\times$  20 時点とした。

#### ③ 分析モデル

スクリーニングの結果、処置前値と処置後値の両方が得られた事例について、以下のモデルを当てはめた。

$$D_{ai} = \delta_i + r_{ai}$$

$$\delta_i \sim N(\delta, \tau_\delta^2)$$

$$r_{ai} \sim N(0, \sigma_\delta^2)$$

処置の効果は  $\delta$  であるが、データ発生モデルにおいては  $\delta = 0$  である。制約付き最尤法により推定値を求め、 $\delta$  については Snijders & Bosker (2012) にもとづき、5%水準の Wald 検定を行った。シミュレーションでは、各条件につき 5,000 セットのデータを発生させ、それぞれのデータに上記のモデルを当てはめて  $\delta$  に関する推定と検定を行った

(3) 研究 3: 効果量の検定結果にもとづいて公表が左右された場合のバイアスの検証

#### ① 効果量の検定結果にもとづく公表

本研究では、統計的に有意な効果量のみが公表された場合のメタ分析におけるバイアスについて検証することとした。ここでは、AB デザインによる研究が行われたことを想定し、効果量として Busk & Serlin (1992) による標準化された平均値差  $d$  を対象とすることとした。これは、高橋・山田・小笠原 (2009) における ES\_BS2 と同等である。また Raudenbush & Liu (2000) よって提案された HLM による縦断的データ分析のための効果量  $d$  と同等であるので、これにもとづいて効果量の検定や推定を行うことができる。 $d$  の母数を  $\delta$  とし、シミュレーションにおいてはこの値を 0 もしくは 1.0 と設定した。これは、Parker, Brossart, Vannest, Long, De-Alba, Baugh, Sullivan (2005) による、単一事例実験の効果量を Cohen (1988) による集団比較研究のそれと対応させた結果から導いたものである。この換算では、 $\delta = 1.0$  は Cohen (1988) の基準でいうところの「小 (small)」に相当する。

また、単一事例実験の特性に鑑み、ラグ 1 の系列相関 AR(1) を、また、メタ分析が行われることに鑑み、研究 2 と同じく級内相関を操作することとした。AR(1) については、Parker et al. (2005) のレビューを参考に 0 と 0.2 を、級内相関については Moeyaert, Ferron, Beretvas, & Van den Noortgate (2014) を参考に 0.1 と 0.2 を設定した。観測時点数については、Shadish & Sullivan (2011) を参照に AB 各フェーズで 5 時点または 10 時点と設定した。以上 2 水準からなる 4 要因の組み合わせで計 16 条件が構成された。

#### ② データの発生と分析

データは、以上の条件設定のもとで各 5,000 ケース発生させ、ケースごとに効果量  $d$  が有意であったものだけを残した。その後、残ったケースを HLM により統合し、効果量の推定値  $\delta$  を得た。全データを対象にした場合

の効果量の推定値 $\widehat{\delta}_0$ との差  $\widehat{\delta} - \widehat{\delta}_0$ を公表バイアスの大きさとして評価した。

#### 4. 研究成果

##### (1) 研究 1

図 1 は、高橋・山田・小笠原 (2009) による効果量の分布を示したものである。3 つの効果量の分布は、いずれも大きく歪んでいた。

PND については、70%を超えたケースを確かな効果あり、50%~70%では不確実だが効果ありとする基準がある (Scruggs et al., 1987)。これに従えば、公表されている事例のほとんどが介入効果を支持していることになる。

また、ES\_BS2 や ES\_C にもとづいても、ネガティブな効果が報告されているケースはほとんどないことがわかる。同様の傾向は、デザインを多重ベースラインに限定しても観察された。これは、少なくともこれらの事例について何らかの公表バイアスが生じている可能性を示すものである。

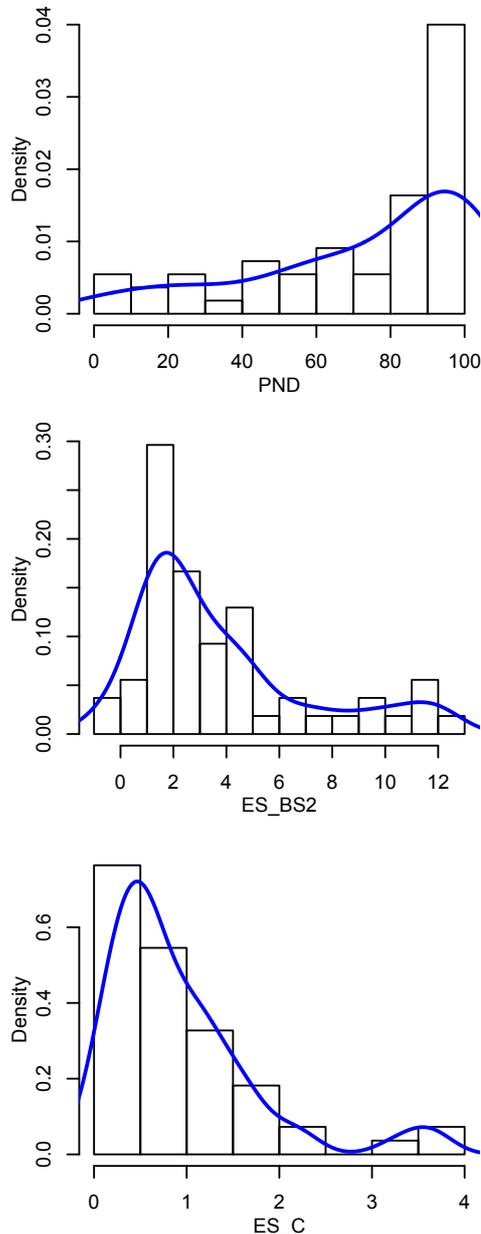


図 1 : 効果量の分布

##### (2) 研究 2

図 2 は、AB デザインにおける処置前値の平均でスクリーニングが行われた場合のメタ分析結果における効果量のバイアスを示したものである。

これを見ると、効果量のバイアスは級内相関の高さに大きく左右され、特に級内相関が  $\rho = .5$  の場合には、全く介入の効果がないにも関わらずメタ分析の結果では  $d = 1.2$  程度という非常に大きい効果を見いだしてしまうことを示している。

処置前値と処置後値との間の相関関係の強さも効果量のバイアスに多少は寄与していたが、級内相関に比べるとその影響は微々たるものであった。

以上の結果より、介入が特に必要とされる対象を選んで実践を行うという事例研究に特有の現象が「平均への回帰」として作用し、メタ分析の結果、本来は存在しない介入の効果があたかも非常に大きいものであるかのように評価される危険性が明らかになったと言える。

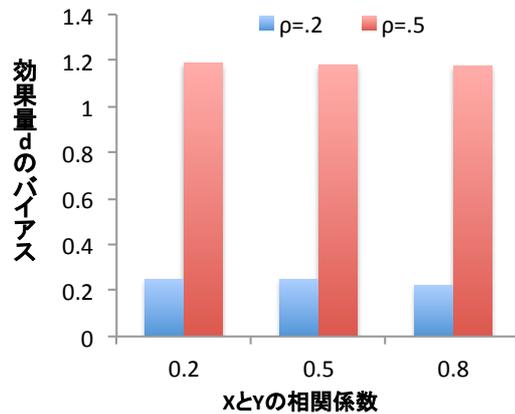


図 2 : 処置前値にもとづくスクリーニングによる効果量のバイアス

##### (3) 研究 3

表 1 は、効果量の検定結果が 5%水準で有意となった事例のみが公表され、その結果を HLM によりメタ分析した場合の効果量のバイアスを、設定した条件ごとに示したものである。

その結果、 $\delta = 0, n = 10$  の組み合わせの場合が最もバイアスが大きく、 $\widehat{\delta}$  と  $\widehat{\delta}_0$  の差は 2.0 程度もあった。一方、 $\delta = 1, n = 20$  の場合が最もバイアスが小さかったが、それでも  $\widehat{\delta}$  と  $\widehat{\delta}_0$  の差は 0.5 程度と無視できない大きさであった。さらに、他の条件が等しければ、級内相関および系列相関の値が大きいほどバイアスは大きい傾向にあった。しかし、これらの影響は  $\delta$  や  $n$  のそれに比べればわずかなものであった。

以上の結果は、効果量が統計的に有意であるかどうか事象公表の条件となっている場合、介入の効果が大幅に過大評価されかね

ないことを示している。特に、本当は介入の効果が小さくとも存在する場合 ( $\delta = 1$ ) よりも、むしろ介入の効果が全く存在しない場合 ( $\delta = 0$ ) の場合の方がバイアスは大きいという結果は注目に値するものである。

表 1：効果量の検定結果によって公表の是非が決まった場合のバイアス

条件				効果量		有意な事例の割合	バイアス
$\delta$	$\rho$	AR(1)	n	全て	有意のみ		
0	0.1	0.0	10	-0.012	1.899	0.031	1.911
0	0.1	0.0	20	0.003	1.145	0.071	1.142
0	0.1	0.2	10	0.004	1.959	0.056	1.956
0	0.1	0.2	20	0.009	1.173	0.099	1.164
0	0.2	0.0	10	0.001	2.014	0.047	2.013
0	0.2	0.0	20	-0.001	1.215	0.110	1.216
0	0.2	0.2	10	0.005	1.995	0.074	1.991
0	0.2	0.2	20	-0.004	1.222	0.129	1.227
1	0.1	0.0	10	1.015	2.114	0.284	1.099
1	0.1	0.0	20	1.002	1.381	0.610	0.379
1	0.1	0.2	10	0.938	2.131	0.290	1.193
1	0.1	0.2	20	0.974	1.435	0.587	0.461
1	0.2	0.0	10	1.006	2.160	0.286	1.154
1	0.2	0.0	20	1.001	1.468	0.590	0.467
1	0.2	0.2	10	0.975	2.173	0.319	1.198
1	0.2	0.2	20	0.987	1.489	0.588	0.502

#### (4) 総括

本研究によって明らかになったことは主に次の三点である。

第一に、わが国で行われた単一事例実験のレビューから、効果量の分布は大きく歪んでいることがわかった。「特殊教育学研究」という特定の雑誌に限定しており、また研究の特性を細かく分類して検討したわけではないが、実際にこの分野の事例研究において公表バイアスが存在していると推察される結果であると言える。

第二に、研究者や実践者が介入の必要性を感じるような事例について選択的に介入を行い、その結果が公表されることで、メタ分析による介入効果の評価に無視できないバイアスが生じる可能性のあることがわかった。すなわち、「平均への回帰」により、本来効果が存在しないはずの介入についても、非常に効果的であるかのように評価される危険性のあることがわかった。この傾向は、特に観測値の級内相関が強いほど顕著であった。ただし、ここで設定した級内相関のうち、 $\rho = .5$ というのは Moeyaert et al. (2014) のレビューに従えばいささか大きすぎる値であり、ここまでの状況が現実的であるかは別途検討する必要があるだろう。

第三に、効果量が統計的に有意であるかどうかをもとに公表の是非が決まるような状況では、メタ分析によって統合された効果量の推定値はポジティブな方向に偏ることがわかった。しかも、この傾向は、少しは効果的である介入についてよりも、全く効果的な

い介入について評価を行う場合の方が深刻であった。

以上の結果を総括すると、単一事例実験の対象者をいかに選定するかというところから始まり、どのような基準にもとづいて結果が公表されるのか、またメタ分析の際に公表バイアスをいかに考慮するかといったことが、介入効果を正確に評価するために非常に重要な検討事項であると言えるだろう。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 1 件)

- ① Okumura, T. (2013). Regression to the mean in pretest-posttest comparisons with two-level selection of subjects. *Bulletin of Joetsu University of Education*, 32, 87-92. (査読無)

#### 6. 研究組織

##### (1) 研究代表者

奥村 太一 (OKUMURA, Taichi)

上越教育大学・大学院学校教育研究科・准教授

研究者番号：90547035