

心理学研究におけるリサーチデザインの理想

平 井 啓

大阪大学

Ideals of research design in Japanese psychological research

Kei HIRAI

Osaka University

From my academic experience of research design in a clinical study conducted in the interdisciplinary area of health psychology, social psychology and clinical medicine, I discuss the reproducibility and the “ideals” of research design in Japanese psychology research. To realize the “ideals” of research design in psychological research to reveal the universal nature of psychological phenomenon as complex and probable, it will be necessary to establish a robust theory through considerable theoretical studies, critical discussions, and experiments and surveys based on small hypotheses, and to conduct an empirical study using experimental design after registration in a psychological research registry, in which researchers should disclose their research design including estimated effect-size and calculated sample-size in advance.

Key words: reproducibility, research design, research question, study protocol, effect-size, sample-size, research registry

キーワード：再現可能性, リサーチデザイン, リサーチ・クエスチョン, 研究計画, 効果量, サンプルサイズ, 事前登録制度

1. はじめに

本号の特集は、心理学における再現可能性の問題について真摯な議論が巻き起こる最良の機会を提供していると思われる。このような議論は再現可能性の問題だけでなく、日本の心理学研究の方法論、研究倫理、出版、さらには学会のあり方を考え、議論する機会に繋がると考えている。そこで本論では、筆者のこれまでの研究者としての経験から、特に、研究計画の立案、実施とリサーチデザインの観点から心理学研究の再現性可能性と心理学研究のあり方について考え、その「理想」について提案する。

筆者は、これまで健康心理学・社会心理学を基盤となるディシプリンとしながら、臨床医学にまたがる学際領域、特にサイコオンコロジーという分野において、調査研究や臨床研究を行い、その成果を、英文誌を中心に公表してきた。このような研究者人生の中で、厚生労働省科学研究費のプロジェクトとして、認知行動療法の1つである問

題解決療法のプログラムを日本人のがん患者向けにアレンジしたものを開発し、その有効性を前後比較研究により検証する研究に責任研究者として関わった(Hirai et al., 2012)。この研究は、医学者、看護学者との共同研究であり、いわゆる医学的臨床研究として求められるさまざまな条件を満たしながら進めなければいけない研究で、研究計画の策定と準備だけで1年以上の時間を費やすこととなった。この研究プロジェクトでの経験を振り返ることで、再現可能性や心理学研究におけるリサーチデザインの課題を提示する。

2. 心理学的臨床研究における リサーチデザイン

まず、このプロジェクトのような医学的臨床研究では、研究の「プロトコール」と呼ばれる研究計画書の作成を行わなければいけない。そのプロトコールでは、「どんな人を対象に、どんな介入を行い、何と比較し、どのようなアウトカムをどの

くらい改善するのか」, すなわち Patients, Exposure, Comparison, & Outcomes: PECO (福原, 2008) と呼ばれる研究テーマ (リサーチ・クエスチョン) を構造化するための定式に従って, それらを事前に設定しなければならなかった。加えて介入で使用する教材や介入者の訓練も含めてプログラムの開発を行いながら, 「班会議」と呼ばれる3ヶ月~6ヶ月毎に開催されるリサーチミーティングでプレゼンテーションとディスカッションを行い, 研究を実施するために必要なことについて, PECOを満たすように決めていった。ここで求められていたことは, プロトコルが完成する段階で研究の知的作業のうち8割が終わっているということであり, あらゆる可能性を考えることが研究班の研究者たちから容赦なく求められた。最終的には, 無作為化比較試験をするのか, 単純な前後比較試験をするのかという狭い意味での研究デザイン, さらにデザインにもとづくサンプルサイズを計算し, データ取得後の統計解析方法, 研究組織と役割分担, データの管理方法を含む倫理的配慮の内容についても詳細に記載することが求められた。

「班会議」でのプレゼンテーションとディスカッションは, リサーチ・クエスチョンに関する本質を突かれつつも, 予算や研究期間, 協力施設の獲得見込みという各種制約条件も考慮しながら現実的なプランニングが必要とされる, 筆者にとってはかなりハードな仕事であった。途中, 科研全体の責任者である「班長」の命令で, アメリカの心理療法の臨床研究の専門家 (David Streiner, PhD) を招いてのワークショップで「プロトコル」を (英語で) 発表することになった。この発表で, Dr. Streiner からは, われわれのプロトコルの中で, 効果量をどれくらいに設定するのか? という点について指摘を受けて, 0.5SD なのか 1SD にするのかそこをよく考えるようにというアドバイスをもらった。なぜなら 1SD の変動は誰がみても意味のある介入の効果であると考えられるからとのことであった。

筆者らの研究は, 心理的苦痛のある術後補助化学療法を終了した乳癌患者を対象とし (Patients), 問題解決療法プログラムを用いた心理的介入 (Exposure) が実行可能であり, 介入前と比べて (Comparison), HADS (Hospital anxiety and depres-

sion scale)¹⁾で測定された心理的苦痛が有意に改善する (Outcomes) かを明らかにすることを目的としていた (Hirai et al., 2012)。われわれは, 対象者と同じ母集団 (術後の乳がん患者) を対象とした観察研究を事前に行っており, そのデータに基づいて介入で得られるべき効果量とサンプルサイズの計算を行った。介入前から介入後6ヶ月までのHADS得点が11点以上であった対象者のHADS尺度の自然経過の変化量を 1.74 ± 6.82 点と推定し, この増加分に, 先述のワークショップでの Dr. Strainer のアドバイスを元に 1SD (6.82 点) を引いたもの ($6.82 - 1.74 \approx 5$), すなわち HADS 尺度の5点の減少を意味のある介入による変動であると仮定した。そこで, 想定される HADS 尺度の分散を 7.0 点, 検出すべき変化量 (効果量) を 5.0 点, タイプ I エラーの α (有意水準) を 0.05, タイプ II エラーの β を 0.20, Power (検定力) を 0.80 としてサンプルサイズの計算を行った。その結果, 必要なサンプルサイズは, $N=17$ となった。これを下限値としつつ統計解析のためのマージンを考えて最終的に 20 例を統計解析の対象とし, ドロップアウトも考慮にいれ目標症例数を $N=30$ と設定した。これを最後に記載し, プロトコルを完成させ, 実際の介入研究を始めることができた。結果としては, われわれの問題解決療法プログラムによる介入は, 心理的苦痛の高い術後乳がん患者の抑うつ・不安を, 術後6ヶ月後においても有意に低減させ ($F(2, 36)=6.45, p<0.01$), 比較的高い効果量 (0.82) を得た (Hirai et al., 2012)。

3. 効果量とサンプルサイズの事前決定と研究の頑健性

上記の経験から学んだことは, 事前に得られるべき効果量を設定し, それに基づいてサンプルサイズを計算しておくことと, PECOで定式化を行ったようなシンプルで明確なリサーチデザインの重要性, さらに研究計画段階でのレビューの有効性である。このような筆者の経験からは, 本号のテーマである心理学研究における再現可能性についても, 効果量を事前に十分に検討し, 設定す

1) 42点満点の尺度 (Zigmond & Snaith, 1993)。日本人がん患者における適応障害・大うつ病のカットオフ値が11点と設定されている (Kugaya et al., 1998)。

ることで得られた有意な結果であれば、その研究成果は、頑健であり、再現性も確保されるのではないと思われる。

しかしながら、筆者らの研究で想定した効果量はアウトカムとなった尺度の変化量のISDという大きなもので、この変化は、実際の介入の現場で見ていた感覚としては、研究参加者の様子が「目に見えて良くなる」くらいははっきりしたものであった。つまり、心理学研究において、効果量が高い結果というのは、誰の目から見ても明らかに変化したと実感できるくらいのものである。臨床研究の場合は、その実践的な必要性から、効果が明らかにあると臨床実践レベルで複数の臨床家が感じるものについて、高いコストをかけてその効果を実証的に検証するということになる。当然、そこまでやったとしても有意な結果が得られないこともある。余談にはなるが、筆者らの研究は、日本のサイコオンコロジー、臨床・健康心理学分野においてはかなり質の高い研究であると考えられるが、最終の解析対象のサンプルサイズが19名で無作為化比較試験でなかったために、サイコオンコロジー分野の国際誌である *Psycho-Oncology* 誌に投稿した際には、Editor's rejectとなり、国内の医学系の英文誌にやっと拾ってもらうことができた。このように当然の結果を導く研究をかなりの手間隙とコストをかけて行う臨床研究は、研究者としては割に合わない仕事である。

一方で、本号で取り上げられているような、再現性可能性について疑問が持たれている研究の多くは、新奇性やオリジナリティが高い心理現象を扱ったものであることが多く、事前に効果量を推定することが困難な場合が多いと考えられる。さらに、だれも発表していない結果を発表すること

がその研究者にとって最も価値のあることなので、競争者との関係も考えると、筆者が経験したように事前に比較的オープンなところで研究計画を発表し、ピアによるレビューを受けるという過程を踏むことは難しいと考えられる。

4. リサーチデザインとは

筆者は、前述のような一連の研究過程を経験する中で、プロトコル作成のために必要に迫られたということもあり、「リサーチデザイン」に関する知識を深めることとなった。その中でも「医学的研究のデザイン (Designing clinical research)」(Hulley et al., 2007) という本に示されている研究のデザイン・実施と推論の対応関係という部分が、分野を越えて心理学研究においてもその根幹となると考えている (図1)。

これによると、科学的研究者であるわれわれ心理学者は、心理学における普遍的真理を結論として得るために、研究テーマ (Research question) を設定し、前述のように研究計画に落とし込み (デザインする)、実際の研究を実施する。そこで得られた研究結果に内的妥当性 (Internal validity) がある場合、すなわち介入や実験の過程に誤りがなく研究の条件に照らし合わせて合理的であることが判断されれば、その研究結果をもとに、実際にデータを取った対象集団という研究の範囲での真理について結論を推論することができる。さらに、その結論が、一般的に適応可能であるか (外的妥当性) の議論を行った結果、当初設定した研究テーマについての結論を推論することができる。もし、設定した研究テーマに含まれる母集団 (目的母集団) を十分に代表するサンプルが得ら

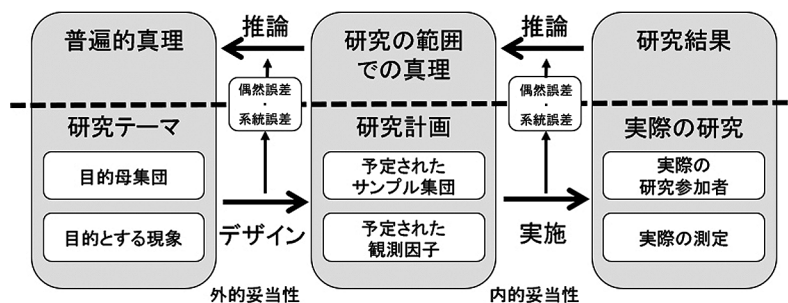


図1 研究のダイナミズム (テーマ・デザイン・実施) (Hulley et al., 2007)

れなかった、すなわち外的妥当性がなかったということになれば、得られた研究結果から普遍的真理を導くことができない。多くの臨床医学研究では、図1のように研究結果に対して、研究の外的妥当性を高めるために目的母集団を限定したりリサーチ・クエスチョン（例：心理的苦痛の高い術後乳がん患者）を設定したり、研究の内的妥当性を高めるために高い効果量を設定したりすることで普遍的真理を推論可能にする「エビデンス」を生み出そうとしている。

本号において、渡邊（2016）は、心理学におけるデータが、エビデンスとしてよりもデモンストレーションとして機能していたと述べている。これはすなわち、心理学研究者が主張したい普遍的真理について直接的な根拠となる研究結果が得られていないために、その普遍的真理そのものを主張できないが、心理学研究者の得た心理学的データが「普遍的である可能性」をデモンストレートしている、ということである。基礎的な心理学研究では、人間の心理と行動に関する普遍的な法則を導くことを究極的な目的としているため、目的母集団が「人間」として設定されるが、実際の研究の対象が、例えば大学生であったりすることも多く、外的妥当性を満たすための論拠を十分に揃えることが難しい場合が多い。さらに、人間という複雑系の反応を測定するために、心理尺度や代理指標としての生理学的指標など、測定方法への偶然誤差や未知の系統誤差の影響が大きく内的妥当性を高めることについても技術的な課題が多い。よってより適切な対象者の設定や測定方法の革新など、心理学研究の内包する方法論の問題を直ちにクリアすることは難しいと考えられる。

また、人間という複雑系を扱う心理学研究において、1つの研究結果だけを根拠として普遍的真理を示すことは難しい。それは、心理学研究は、渡邊（2016）の示すように、生体という複雑系の中でも行動という現象自体の生起確率が低い現象を多く取り扱っているからであると考えられる。複数の研究結果を積み重ねることで1つの普遍的真理を追究する必要がある。

そこで、心理学者が普遍的真理について常に見えるために必要なのは、まずは、内的妥当性を高める研究デザインと測定方法の改善に努めつつ、複数の小さい研究を積み上げることである。その

上で研究テーマから普遍的真理に対する一定の見解を得るための理論的研究とアカデミアにおける議論を深めることで、理論としての心理学の質を上げることができるのではないかと考えられる。そのためには、他の人文社会科学分野での研究の方法論やその分野の学会や雑誌での議論の仕方が参考になるのではないと思われる。

5. 心理学研究における リサーチデザインの理想

最後に、複雑かつ現象自体が確率的である心理学現象を明らかにする心理学研究におけるリサーチデザインの「理想」について考えてみたい。まず本号の池田・平石（2016）において、解決策として示されていた研究計画の事前登録制度がその1つの具体的な案となるのではないと思われる。この事前登録制度は、前述の筆者の経験で示したような事前の検定力推定とサンプルサイズ決定に基づき、「事前に想定したことが結果どうなったか」の過程を透明化し、心理学研究の成果の頑健性を担保する仕組みである。それと同時に、事前に研究計画を公表することで研究テーマのオーサiershipを事前に確定させて、その研究が仮説を支持する結果を生み出すかどうかにかかわらず、発案者となった研究者のオリジナリティや新規性を保護することができると思われる。これは、具体的には、プロトコルペーパーと呼ばれる研究計画を記した論文を投稿し、専門誌に掲載することである。この場合、プロトコルペーパーの掲載を許可する雑誌が必要になる。

さらに、研究テーマとなる心理現象について十分な理論研究を行い、理論の整合性についてさまざまな学術的観点から批判的議論を行う。この際、他の人文社会科学系の学会と同様に、理論研究のみの論文を発表し、それを雑誌上や学会で議論することでその理論の質を高めることができると思われる。同時に、理論を小さい仮説に落とし込んだ複数の実験、臨床実践、フィールド調査などでその理論で示された現象が起こるかどうかをデモンストレートするような研究を行う。これらの研究成果はすぐに普遍的真理に結びつけるような結論とせず、再び批判的な議論を行い、理論研究で示された理論の頑健性を強化したり、それ

を修正したりするために用いる。その頑健性が示された段階で、心理的現象の十分な効果量を想定した介入研究を含む実験デザインによる研究を、ピアレビューによる研究計画の立案と先に示した事前登録制度の下で行う。研究結果に対して、仮説の是非を示し、ディスカッションを行うことで、得られた結果を研究テーマとなった心理現象の普遍的真理を推論する。

最後に「理想」を実現するためには、心理学研究者のマインドセットを新たなものにすることも必要であると考え。日本の心理学アカデミアにおいて、アカデミックポストを得て、研究活動を行うためには、自分以外の研究者はライバルである。少しでもライバルよりも先に、「有意な研究成果を示したい」という研究者のマインドセットは自然なものではあるが、それが本号で取り上げられていた、*p*-HackingとQRPsに根底でつながっているように思われる。しかし「囚人のジレンマ」のようなこの状況を続けることは日本の心理学アカデミアにとって望ましいことではない。内部的競争環境を保ちつつも、アカデミア全体として、外的環境、主として研究資金の最大のステークホルダーとなっている日本国に対して、「信頼できる成果」を供給し続けて行く努力を行っていかねばいけない。そのためには、心理学研究者はたとえ研究結果が「有意」でなくても、そして研究成果を公表する雑誌のインパクトファクターが小さくとも、結果を公表することを惜しまず、研究の計画段階から研究のライバルとも直接の議論を厭わないようにするというマインドセッ

トが必要である。さらに心理学アカデミア全体に貢献することが研究者としての自らの幸福にも繋がると考えることも必要である。

文 献

- 福原俊一 (2008) リサーチ・クエスチョンの作り方 健康医療評価研究機構 (iHope).
- Hirai, K., Motooka, H., Ito, N., Wada, N., Yoshizaki, A., Shiozaki, M., Momino, K., Okuyama, T., & Akechi, T. (2012). Problem-solving therapy for psychological distress in Japanese early-stage breast cancer patients. *Japanese Journal of Clinical Oncology*, 42, 1168–1174.
- Hulley, S., Cummings, S., Browner, W., Grady, D., & Newman, T. (2007) *Designing clinical research* (3rd edition). Philadelphia, PA: Lippincott Williams & Wilkins. 木原雅子・木原正博 (訳) (2009) 医学的研究のデザイン (第3版) メディカル・サイエンス・インターナショナル.
- 池田功毅・平石 界 (2016) 心理学における再現可能危機：問題の構造と解決策 心理学評論, 59, 3–14.
- Kugaya, A., Akechi, T., Okuyama, T., Okamura, H., & Uchitomi, Y. (1998). Screening for psychological distress in Japanese cancer patients. *Japanese Journal of Clinical Oncology*, 28, 333–338.
- 渡邊芳之 (2016) 心理学のデータと再現可能性 心理学評論, 59, 98–107.
- Zigmond, A. S., & Snaith, R. P. (1993). The hospital anxiety and depression scale. 北村俊則 (訳) The hospital anxiety and depression scale (HAD尺度) 精神科診断学, 4, 371–372.

— 2016. 4. 24 受稿, 2016. 4. 30 受理 —