

再現可能な心理学研究入門¹

国里愛彦²

Introduction to reproducible psychological research

Yoshihiko Kunisato²

Abstract：本論文では、心理学における再現性の危機について概観した上で、再現可能性とは何かその定義について論じた。Claerbout-Donoho-Pengによる reproducibility（再解析による再生可能性）と replicability（追試による再現可能性）の区別から始まって、Goodman et al. (2016)の3つの再現可能性（方法・結果・推論の再現可能性）について紹介した。再現可能性を高める取り組みは、近年始まったものも多く、個々の取り組みが再現可能性におけるどういう位置づけになるのかわかりにくい。そこで、Goodman et al. (2016)の3つの再現可能性ごとに、再現可能性を低める要因と高めるための方策について整理することで、再現可能性の全体像を捉えられるように解説した。最後に、方法の再現可能性（再解析による再生可能性）を高める取り組みとして、共有に向けたデータ管理、解析のパッケージ化、解析環境のコンテナ化について具体的に解説した。再現可能性を高めるための方策の中には心理学教育のなかで扱われてこなかった内容もあり、本論文が再現可能な心理学研究実践の一助となることを期待する。

Keywords：再現性の危機、再解析による再生可能性、追試による再現可能性、3つの再現可能性、共有を前提としたデータ管理、解析のパッケージ化、解析環境のコンテナ化

1. 心理学における再現性の危機

一般的に、再現性とは、ある現象が成立する条件を整えることができれば、その現象が繰り返し生じることを指す。心理学の知見の多くは再現性を前提としており、同じ対象母集団に対して、同じような操作を行えば、同じような結果が得られることが期待されている。このような再現性を前提とすることで、心理学の知見を用いて、ある現象の説明をしたり、予測したりができる。例えば、心理学の応用領域である臨床心理学では、心理検査の結果からその後の症状の変化などを予測したり、特定の対象母集団に対してどの心理療法が効果的であるかなどを判断したりする。心理検査は、特定の対象母集団において、どういうことが予想できるのかという観察研究の知見に再現性があることが前提になっている。心理療法の効果は、特定の対象母集団に対して、どういう心理療法を行うと、どういう症状改善が生じるのかという介入研究の知見に再現性があることが前提になっている。公認心理師法が成立し、心理学の知見が医療・保健、教育、産業、福祉、司法領域とさまざまな領域で活用されることが公的に期待されるようになってきている。こ

のような社会的な状況において、心理学研究における再現性の確保の必要性は非常に高いといえる。

心理学に対する社会的な責任が重くなる一方で、2010年代以降、心理学を含む複数の研究領域において、研究の再現性の問題が指摘されてきている。例えば、化学、物理学、地球科学、生物学、医学などのさまざまな研究領域の1500名の研究者を対象とした調査では、「再現可能性の危機があるか？」の質問に対して、9割が「ある（重大な危機52%、軽微な危機38%）」と回答した（Baker, 2016）。さらに、70%の研究者がほかの研究者の実験の再現に失敗したと報告しており、地球科学、医学、物理学、生物学、化学の順番で失敗したという報告の確率が高くなっていった（Baker, 2016）。研究領域によって再現性の高低に違いが認められるものの、科学全般において、再現性の危機が指摘されている。心理学も同様に再現性の危機が指摘されている。例えば、再現性の危機に関連して、100件の心理学研究を対象に大規模な追試プロジェクトが実施されている（Open Science Collaboration, 2015）。100件の内、オリジナルの研究の97件において統計的有意性が示されていたが、追試研究では35件の研究でしか統計的有意性が示されなかった（Open Science Collaboration, 2015）。心理学では36%の研究しか結果が再現できなかったという結果は、心理学者に衝撃を与えるものであった（心理学の再現性の危機にかかわる詳しい経緯については、池田・平石（2016）を参照）。これらの指摘から、心理学においても再現性を高

受稿日2019年11月13日 受理日2019年11月20日

1 本論文は、JSPS 科研費（16H05957, 19K03211, 18KT0021）と専修大学特別研究員（特例）制度の支援を受けて執筆された。

2 専修大学人間科学部心理学科（Department of Psychology, School of Human Sciences, Senshu University）

める取り組みへの関心が高まってきている。

現時点において、心理学の再現性の危機に対しては、そのセンセーショナルな内容に対する反応は落ち着いて、具体的にどのように再現可能性を高めていくのか建設的な議論や実践が進みつつある。本論文では、再現可能性について定義した上で、再現可能性を高めるための方法について論じる。

2. 再現可能性とは

日本語では、reproducibility と replicability は、ともに再現可能性と訳すことがあり、両者が区別されておらず、混乱の元となっている。この混乱は、日本語に限定した問題ではなく、英語圏においても生じている³。元々、Jon Claerbout によって最初に reproducible research が提案された (Claerbout & Karrenbach, 1992)。これは地球物理学での計算において、研究終了後に、ほかの研究者がその結果を再現できるようにするという取り組みであった。その後、統計学者の David Donoho が reproducible research においては、データとコードの公開が必要である点を明確化した (Buckheit & Donoho, 1995)。つまり、reproducible research とは、既に実施された研究を発表する際に、そのデータとコードを共有することで、ほかの研究者がその研究の結果を繰り返し得ることができるように研究することである。さらに、Roger D. Peng が、reproducibility と新規にデータ取得して同様の結果が得られることを指す replication を分けた (Peng, Dominici, & Zeger, 2006)。Claerbout-Donoho-Peng の定義は、その後の研究者に大きな影響を与えたが、一方で reproducibility と replication に対して彼らとは正反対の定義を採用する研究者や学会も存在する (Barba, 2018)。このように、再現可能性については、日本語訳だけでなく、その概念自体も未だに議論がある。しかし、Claerbout-Donoho-Peng の定義を採用する研究者は多く、National Science Foundation やアメリカ統計学会などの団体でも採用されていることから (Barba, 2018)、本論文でも Claerbout-Donoho-Peng の定義を議論の前提とする。

Claerbout-Donoho-Peng の定義では、既存データに同じ方法を用いて同じ結果が得られることを reproduce とする。つまり、reproduce は「同じデータ+同じ方法=同じ結果」が満たされることである (Barba, 2018)。データは既にあることから、ここでの同じ方法とは、元

の研究と同じ解析方法を指している。この文脈での reproduce は、新たに何かを生み出すというよりは、元の研究と全く同じデータと解析で結果を再生できることが求められている。そこで、本論文では、reproduce を「再解析による再生性」と訳すことにする。単に「再生性」でも良いように思うが、誤解をできるだけ小さくするために、「再解析による再生性」とした。「再解析による再生性」は「同じデータ+同じ方法=同じ結果」が満たされた状態だが、reproducibility は、その状態を可能にするための研究者の実践を含む。再解析による再生性を可能にするには、研究者はデータとコードを共有する必要がある。本論文では、reproducibility を「再解析による再生可能性」と訳すことにする。

次に、Claerbout-Donoho-Peng の定義では、新規のデータに同じもしくは異なる方法を用いて同じ結果が得られることを replicate とする。つまり、replicate の場合は、「新たなデータ+同じ方法=同じ結果」と「新たなデータ+違う方法=同じ結果」の2種類がある。ここでの方法は、解析だけに限定されず、データの取得方法も含む。そのため、「新たなデータ+同じ方法=同じ結果」は直接的追試、「新たなデータ+違う方法=同じ結果」は概念的追試に対応する。既に追試という言葉を使っているように、replication は追試を意味する。しかし、追試の結果同じ結果が得られるというニュアンスを含めるため、本論文では、「追試による再現性」と訳すことにする。「追試による再現性」は、新規のデータ取得によって同じ結果が得られた状態だが、replicability は、その状態を可能にするための研究者の実践を含む。追試による再現性を可能にするためには、研究者はデータ取得にかかる方法を詳細に報告する必要がある。本論文では、replicability を「追試による再現可能性」と訳すことにする⁴。

Claerbout-Donoho-Peng の定義とその混乱を踏まえて、Goodman, Fanelli, & Ioannidis (2016) は、再現可能性を、「方法の再現可能性 (methods reproducibility)」、 「結果の再現可能性 (results reproducibility)」、 「推論の再現可能性 (inferential reproducibility)」の3つに分けて定義した⁵。「方法の再現可能性」は、元の研究と同じデータに対して、同じツールや方法を使うこ

3 再現性の用語については、Barba (2018) が詳しい。

4 本論文では、議論をすすめるためにも reproducibility と replicability の翻訳案を提案した。これをたたき台として、さらなる用語と概念の洗練化が進むことを期待している。

5 Claerbout-Donoho-Peng では reproducibility を再生可能性と訳したが、Goodman et al. (2016) の文脈では reproducibility は再生可能性に限定されないので、すべて再現可能性と訳した。

とで同じ結果を得ることができることである。「結果の再現可能性」は、元の研究と同じ方法で収集された新規のデータから元の研究と同じ結果を得ることができることである。「推論の再現可能性」は、独立した追試や再解析の結果から質的に同じ結論を導くことができることである。

Claerbout-Donoho-Peng の定義との対応で考えると、「方法の再現可能性」は「再解析による再生可能性」に対応し、「結果の再現可能性」は「追試による再現可能性」に対応する (Goodman et al., 2016)⁶。Claerbout-Donoho-Peng の定義との違いは、Goodman et al. (2016) では、「結果の再現可能性」に関して、何をもって元の研究の結果と新規にデータ取得した結果を同じとするのかについて詳細に論じている点と「推論の再現性」を追加した点である。結果の再現可能性において、追試を行って元と同じ結果が得られるか検討する場合、測定誤差がかなり小さい研究領域であれば、結果が同じかどうかの判断がしやすい。しかし、心理学を含む人を対象とした研究では測定誤差が大きく、単一の追試によって、その結果の再現性を問うことには限界がある。個々の研究では統計的に有意な差はなくとも、メタ分析においては、有意な差が検出できることもある。そのため、*p* 値に基づく 2 値判断ではなく、複数の研究の効果をメタ分析などで統合して考える必要がある。このよ

うに、Goodman et al. (2016) は、人を対象とした研究の結果の再現可能性において留意すべき点を明確化した。「推論の再現可能性」は、Goodman et al. (2016) によって新たに提案された再現可能性だが、その問題性は古くから知られてきている。研究論文では、得られた結果から何らかの結論を導くが、同じ結果であっても研究者によって導かれる結論が異なることは多い。考察や結論は結果から一意に導けるものではなく、研究が行われた文脈によって結論は異なり、後述する統計の不適切な使用が行われている場合は、その結果を割り引いて考える必要もある。Goodman et al. (2016) の 3 つの再現可能性は、方法や結果だけでなく、推論の再現可能性を含めることで、より包括的かつ再現性の種類に応じた議論ができる点が優れる。そこで、本論文の以降では、3 つの再現可能性に分けて論じる。

3. 再現可能性を低める要因と高めるための方策

Goodman et al. (2016) の 3 つの再現可能性の分類に従って、それぞれの再現可能性を低める要因と高めるための方策について論じる (図 1)。

3.1 方法の再現可能性

方法の再現可能性は、「同じデータ + 同じ方法 = 同じ

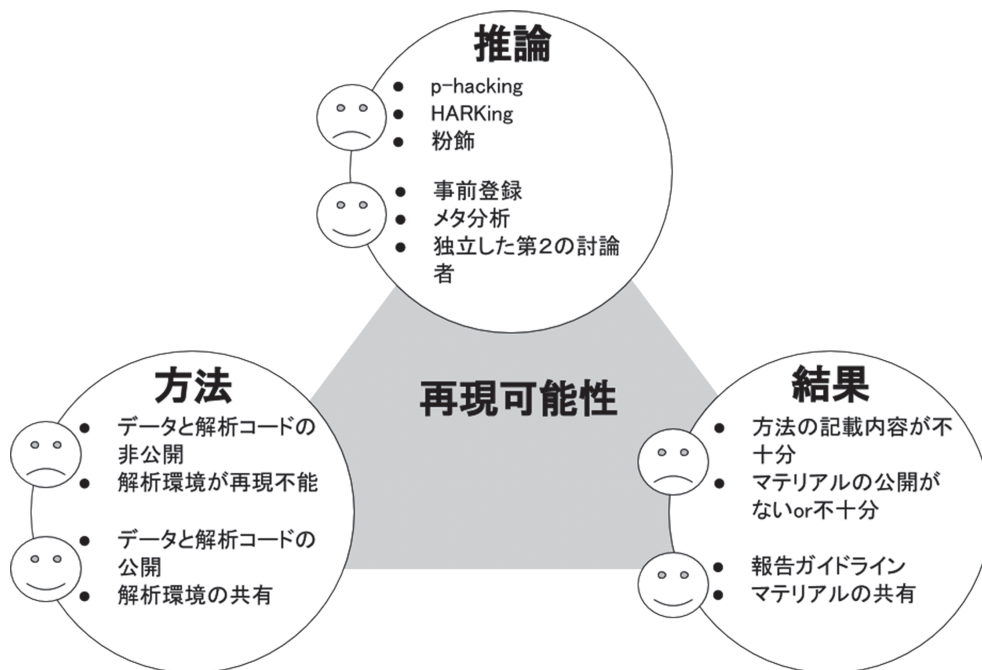


図 1 再現可能性を低める要因と高めるための方策

6 本論文では、基本的には Goodman et al. (2016) の 3 つの再現性の定義を用いるが、分かりやすさから、再解析による再生可能性や追試による再現可能性も適宜用いる。

結果」を確認できることなので、研究で使ったデータと解析コードが利用可能であることが前提条件になる。これまで心理学において、データや解析コードの共有は必須ではなかったが、徐々に学術雑誌がデータの公開を推奨するようになり、変化が生じつつある。それは、医学研究においても同様になる。2000年から2014年にかけて発表された生物医学論文の内、441本をランダム抽出して調べた研究では、生データを直接利用可能な研究は、実証データを用いた268本の研究中0本であった(Iqbal, Wallach, Khoury, Schully, & Ioannidis, 2016)。論文上にデータへのリンクが記載されているが、機能していないものもあった⁷。その後、2015年から2017年にかけて発表された生物医学論文から149本をランダム抽出して調べた結果、18.3%の研究においてデータの入手に関して記載があった(Wallach, Boyack, & Ioannidis, 2018)。このように、徐々にデータ共有は進んできている。生物医学研究に限らず、Cognition誌やPublic Library of Science (PLoS)の発行する雑誌などのように、データ共有を要求する雑誌も増えてきている。論文発表とともにデータや解析コードの公開を行う流れは今後も加速することが予想される。

再現可能性に関する指摘や雑誌の掲載指針の変化に伴って、データや解析コードの共有が進んできているが、公開されているデータや解析コードから論文で報告されている結果を再現できないという問題も指摘されつつある。例えば、データやコードの共有に関する指針が出た以降のScience誌に掲載された論文を対象とした調査では、データと解析コードの提供があったのは全体の44%であり、そのうち結果を再現できたのは26%であった(Stodden, Seiler, & Ma, 2018)。また、データ共有が求められているCognition誌に発表されたデータに対して行った調査において、再解析が可能なデータを公開している35論文の内、22論文は再現できたが(その内11論文は再現するのに著者の助けが必要だった)、13論文は何かしらの指標で結果が再現できなかった(Hardwicke et al., 2018)。このように、データと解析コードの共有がなされていても、それが活用できないものだったり、再現不可能なものだったりすることがある。

データや解析コードの共有がなされているのに、結果が再現できない理由としては、(1) データや解析コー

ドが公開を前提として作成されておらず第三者では扱えないものになっている、(2) ソフトウェアの活発な開発により、ソフトウェアなどのバージョンが論文の結果が作られた環境と異なっているなどがある。これまでデータや解析コードは研究者個人が管理するものであり、その研究者が分かればよいというものだった。そのため、第3者にも理解できる形式でデータや解析コードを書くような教育は受けてないことが多い⁸。また、データや解析コードの共有において、コンセンサスの得られた形式もまだない。次に、オープンソースソフトウェアの発展により、急激な速度で解析に用いるソフトウェアは発展しており⁹、それによって再解析による再生可能性が低くなることもある(Epskamp, 2019)。Epskamp (2019)は、その要因として、 β 版の存在とソフトのアップデートと依存関係の問題を指摘している。すでに広く使用されているソフトであってもまだ β 版のものもあり、今後大幅なアップデートがなされることがある。アップデートによっては、 β 版のときに書いた解析コードがアップデート後に正常に動作しなくなることがある。 β 版でなくてもソフトのアップデートによって、かつての解析コードが正常に動作しなくなることがある。特に、RやPythonでは単一のパッケージのみで動作するのではなく、いくつかのパッケージに依存して動作することが多い。その場合、依存しているパッケージのどこかに大幅なアップデートがあると、正常に動作しなくなる可能性がある。また、ソフトウェアを超えて、OSが違うことで動作しないこともあり、解析に使用するソフトウェアだけでなく、ソフトウェアが動作する環境も再生可能性においては考慮する必要がある。データと解析コードの共有方法や解析ソフトや解析環境の共有方法については、第4節にて詳細に論じる。

3.2 結果の再現可能性

結果の再現可能性は、「新たなデータ+同じ方法(直接的追試) = 同じ結果」と「新たなデータ+違う方法(概念的追試) = 同じ結果」を確認できることなので、追試が可能くらい詳細に方法が記載されている必要がある。もちろん論文の本文に記載できる内容には限度があるので、一部はサプリメントに記載したり、別途公開

7 リンク先にデータもコードもないことはままある。雑誌の要求に答えるために記載しているもののリンク先が機能しなかったり、著者に問い合わせても全く返事がなかったりすることはよく経験する。

8 少なくとも著者は、このような教育を受けたことはない。

9 このような活発な開発の様子は、Rでの解析でよく用いられるtidyverseパッケージ(<https://github.com/tidyverse/tidyverse>)や認知課題の作成で用いられるPsychoPy(<https://github.com/psychopy/psychopy>)のGithubページを見ると観察できる。

したり、著者問い合わせによって対応するなどが行われる。社会・心理学的介入に関する臨床試験における報告の質は低いことが指摘されており、方法において報告が必要とされる内容を記載している論文は半分以下との報告もある (Grant, Mayo-Wilson, Melendez-Torres, & Montgomery, 2013)。方法の記載が不十分だと結果の再現可能性を低める可能性がある。特に、臨床試験などにおいてはどのような介入を行ったのか詳細に記載したプロトコルや割付などの手続きについても詳細に記載されている必要がある。これは、心理学における観察研究や実験室実験においても同様で、現象の再現において必要な手続きが報告されていないことは、結果の再現可能性を低める。また、心理学において質問項目は論文上でも確認できるが、認知課題などは論文上で記載はされているものの実際に用いた認知課題のプログラムや刺激がそのまま公開されていることは多くない。文章で説明できることには限界もあるため、論文の記載内容から想定される課題と実際に用いた課題が異なることがある¹⁰。このような認知課題などの研究の材料も詳細に報告もしくは論文とは別に公開しておかないと、オリジナルの研究と追試の研究では異なる方法を行っている可能性もあり、結果の再現可能性を損なうことになる。

結果の再現可能性については、適切に追試ができるように論文上で方法について報告する必要がある。臨床研究などでは、最低限必要とされる研究報告ガイドラインについて整備されてきている。例えば、無作為化比較試験については CONSORT 声明 (Moher et al., 2010)、観察研究については STROBE 声明 (Von Elm et al., 2007)、診断精度研究については STARD 声明 (Bossuyt et al., 2015)、尺度研究については COSMIN チェックリスト (Mokkink et al., 2010; Prinsen et al., 2018)、メタ分析については PRISMA 声明 (Liberati et al., 2009)、症例研究については CARE ガイドライン (Riley et al., 2017) などがある。日本国内でも医学系学会を中心に海外で使用されている研究報告ガイドラインの解説がなされてきており、『行動療法研究 (2019年より、認知行動療法研究)』では、「行動療法研究における研究報告に関するガイドライン」と題する特集が組まれた (奥村・原井・谷・佐藤, 2014)。行動療法研究の特集において、

非薬物療法の介入研究 (奥村, 2014b)、観察研究 (竹林, 2014)、尺度研究 (土屋, 2015; 稲田, 2015)、メタ分析 (国里, 2015)、症例研究 (谷, 2015) について、日本語による解説がなされている。臨床研究だけでなく心理学研究においても、これらの報告ガイドラインは有用なため、結果の再現可能性を高めるために適宜活用することが望ましい。また、認知課題のプログラムや刺激などの材料は、Center for Open Science の Open Science Framework (<https://osf.io/>) や Github (<https://github.com/>) などのサービスを通して公開することで、その後の追試などで活用できるようにできる。研究報告ガイドラインに従って研究で用いた方法について過不足なく報告し、紙面に載せられない詳細なプロトコルや認知課題などの材料は OSF や Github を活用することで、結果の再現可能性を高めることができる。

3.3 推論の再現可能性

推論の再現可能性は、独立した追試や再解析の結果から質的に同じ結論を導くことができることであるが、同じ結果であっても研究者によって導かれる結論が異なることは多い。方法や結果に比べて、考察は結果から一意に導かれるものではなく、そこには研究者の解釈がはいる。この解釈には、研究の文脈やストーリーが入ってくるため、結果を解釈する者によって異なることがある。このように推論の再現可能性は、ほかの再現可能性よりも担保するのが難しいものだが、それをさらに難しくする要因として、(1) 統計の不適切な使用、(2) 統計学的に有意になった結果を元にした研究仮説の設定、(3) 粉飾などがある¹¹。

まず、統計の不適切な使用として、p-hacking がある。実証的な研究の多くが統計的帰無仮説検定を採用しており、 p 値が有意水準を下回るかどうかにかかわらず焦点が当てられてしまいがちになる。そのため、有意水準未満の p 値を得るためにハッキングを行う p-hacking が行われてしまう現状がある。p-hacking の具体例としては、統計学的に有意になった結果のみを報告しほかの測定した変数や条件を報告しない、参加者を少しずつ足していった統計学的に有意になったところで止める、さまざまな組み合わせで共変量を投入した分析を行って有意になった結果のみ報告するなどがある (池田・平石, 2016)。

10 実際には、著者の研究室で追試研究を計画した際に、論文の記載内容では矛盾があり認知課題が用意できなかった。そこで、著者に問い合わせた結果、論文の記載内容とは少々異なる手続きが用いられていたことが明らかになったことがある (拙取・国里, 印刷中)。その経過は Appendix として公開し、用いたウェブ実験用コードも公開した (<https://osf.io/acqb7/>)。

11 p-hacking, HARKing, 粉飾については、本論文では詳細に扱うことはできないため、池田・平石 (2016) や奥村 (2017) を参照のこと。

このような p-hacking を行ってしまうと、本質的には有意ではない差に対しても有意と判断してしまう第一種の過誤が高くなってしまふ。次に、これまで心理学者は暗黙のうちに、結果が一貫している、ストーリーとしてわかりやすい、新奇性があるなどの審美的判断基準によって論文を評価してきた。そのため、これらの審美的判断基準に沿うために、データを解析した結果をもとに序論や仮説を書くという HARKing という行為が行われてきた（池田・平石, 2016）。これは、仮説検定を偽装した行為であると同時に、検定を行ったが有意ではなかった結果をなかったことにして論文を構成することになるので、p-hacking と同様の第一種の過誤の問題が生じる。最後に、粉飾（spin）とは、臨床試験において否定的な結果が得られた場合に、結果の解釈を歪めた結論に読者を誘導する論文の書き方になる（奥村, 2017）。これは、結果から結論を導くのは一意ではないことを巧みに利用した方法で、p-hacking や HARKing などが行いにくい臨床試験において否定的な結果が得られたときに、考察やアブストラクトの結論部分で肯定的な結果のような印象を植えつける方法になる。具体的には、主要評価項目において統計的有意性が認められない場合に、群間差ではなく群内差に注目して効果を主張する、複数の副次的な評価項目のなかで有意であった部分に注目して効果を主張する、サブグループ解析を実施して有意な結果の得られた部分に注目して効果を主張する、有意な差がないことを劣ってはいないもしくは同等な効果と評価して有効性を主張するなどがある（奥村, 2014a）。論文の批判的吟味の訓練を受けた読者であれば、粉飾に気がつくかもしれないが、それらの訓練を受けていない実務家がアブストラクトと結論のみを読んだ場合は誤った情報が伝わる可能性がある。p-hacking, HARKing, 粉飾は、結果から結論を導くのを難しくしてしまうため、推論の再現可能性を低める行為になる。

推論の再現可能性を高めるためには、（1）p-hacking や HARKing ができないように事前登録をする、（2）1つの結果から結論づけるのではなくメタ分析などを用いて結論を導く、（3）独立した第2の討論者を用意するなどがある。推論の再現可能性を高める方法として、まずは p-hacking や HARKing ができないように、事前に研究仮説や方法を登録した上で、研究を実施することが挙げられる（Nosek & Ebersole, 2018）。臨床試験などにおいては心理学に先行して臨床試験登録が義務づけられるようになり、臨床試験前に研究方法や介入プロトコルに関する査読を受けることも行われてきて

いる。仮説や評価項目や解析方法を事前に登録することで、データ取得後に結果に応じて仮説を変更することができなくなるので、HARKing ができなくなる。また、事前に決めてない方法で解析したり参加者数を変更したりできなくなるので、p-hacking もできなくなる。事前審査付き事前登録制度を採用する雑誌も出てきており、データ取得前の事前登録段階で査読を受けて、それをパスした場合は、結果が仮説を支持してもしなくても掲載される（池田・平石, 2016）。次に、1つの研究知見から考察するには限界があることから、メタ分析を活用する方法がある。系統的に文献を収集して結果を統合するメタ分析によって、バイアスをできるだけ小さくした結論を導くことができ、推論の再現可能性を高めることができる（Gilmore, Diaz, Wyble, & Yarkoni, 2017）。臨床研究のガイドラインにおいては、考察においてメタ分析を行って、これまでの知見を統合した結論を出すことが推奨されている（Avidan, Ioannidis, & Mashour, 2019）。しかし、この推奨については、多くの医学雑誌で要求されていないことや系統的レビューにかかる労力などから、あまり採用されていない現状がある。そこで、系統的レビューを用いずに推論のバイアスを小さくし、粉飾などを防ぐ方法として、独立した第2の討論者（independent second discussant）が提案されている（Avidan et al., 2019）。この独立した第2の討論者とは、その論文に関わった著者グループの考察に加えて、その研究領域の独立した専門家が研究の結果から考察を書いて、1つの論文内に研究を実施した研究者の考察と独立した第2の研究者の考察を掲載することである。独立した第2の討論者にはデータと解析コードや方法の詳細が開示され、それを元に考察を執筆するため、粉飾などを避けることができる。独立した第2の討論者のオーサiershipをどうするのか、領域の専門家かつ著者らとは独立に研究を評価できる討論者をどのように用意のかなど運用上の問題はあがるが、今後議論が進むなかで推論の再現可能性を高める方法として一般に活用されるようになるかもしれない¹²。人を対象としたノイズの多い実証データからどのように結論を形成するのにかかわる推論の再現可能性は、研究知見の蓄積と応用において

12 例えば、データ取得前に方法まで記載した論文を事前審査し、データ取得後に結果以降を記載した論文は事前審査時の方法通りに実施されていれば基本的に採択とする（事前審査付き事前登録制度）。そして、その事前審査にかかわった査読者が研究者とは別に考察を書いて、2つの考察を載せた論文を掲載するという運用もありえるかもしれない。事後的に査読者は匿名ではなくなるが、査読が金銭的・学術的に無報酬なものではなくなり、明示的に知見の蓄積に貢献できることになる。

重要なテーマになる。今後も多様な議論が展開される可能性があるため、議論の推移を見守る必要があるだろう。

4. 再解析の再生可能性を高めるデータや解析コードの共有方法

3つの再現可能性について問題点と解決方法を整理してきた。以下では最も実現可能な方法の再現可能性（再解析による再生可能性）を高めるための具体的な方法について論じる。方法の再現可能性を高める上では、研究で用いたデータと解析コードの共有が必須であるが、これらはただ公開すればよいわけではない。実際、データやコードの共有を推進している雑誌の方針に従って、データやコードの共有がなされているように見えて、全く機能していないことも多い。心理学の教育において、データや解析コードの共有方法について学ぶことはなく、手探りで行った結果、再解析による再生可能性が低くなってしまっている可能性がある。

著者が遭遇した具体的な例をあげると、(1) 解析コードは公開されておらず、Word ファイルに貼りつけられたデータだけが公開されており、再解析する場合にかなり手間がかかる、(2) 論文上にデータへのリンクがあるがリンク先がプライベートなリポジトリになっておりアクセスが不可能になっている、(3) 解析コードは公開されているが説明する README もなく、わかりやすいファイル名もつけられておらず、数十あるファイルを1つずつ開いてコードを確認しなければ正常に動作させられない、(4) シミュレーションのプログラムが公開されているが既に使われなくなった OS でしか動作しないので利用できない、(5) データも解析コードも入手できたが OS やソフトのバージョンが違うためかエラーが出る、もしくは結果が微妙に異なるケースなどがあった。(1) と (2) は再解析を意識したデータの準備や公開ができていないこと、(3) は解析に使ったデータやコードのパッケージ化ができていないこと、(4) と (5) は解析やシミュレーションを行った環境の説明や共有ができていないことが問題になる。そこで、再解析による再生可能性を高めるために、本論文では、共有に向けたデータ管理、解析のパッケージ化、解析環境のコンテナ化に分けて解説する。

4.1. 共有を前提としたデータ管理

再解析の再生可能性を高めるためには、処理前の生データから最終的に論文に報告する結果まで追跡可能な

ようにした上で公開する必要がある。Wilson et al., (2017) は、科学的計算におけるデータ管理の推奨実践として以下の7点を挙げている。

1. 生データを保存する
2. 生データは1つ以上の場所にバックアップをとる
3. 第三者に提供しやすい形式にデータを整える
4. データを解析しやすいように整える
5. データ処理に用いたすべてのステップを記録する
6. 複数のデータセットを使用することを想定して、ユニークな識別子を使用する
7. ほかの研究者がアクセスし引用できるようにデジタルオブジェクト識別子 (Digital Object Identifier: DOI) の付与ができるリポジトリにデータを登録する

まず、最初に得られた生データを1つ以上の場所に保管する。多くの生データはそのままで解析に使えないので、整形したデータを生データに上書きしたくなるが、その過程にミスが入り込む余地がある。そのため、最初に得られた生データは上書きなどの操作をせずにそのまま保管し、最初に得られた生データから最終的な結果まで追跡可能にする。なお、どのように保管したとしてもデータは破損する可能性があるため、生データは複数の場所にバックアップをすることも忘れないようにする。

次に、生データを第三者に配布しやすい形式や解析しやすい形式に整え、その過程をすべて記録する。データのファイル形式は、データの構造に依存するが、心理学の場合は表形式のデータが多いので CSV 形式が適していることが多い。変数名やファイル名も統一した基準のもとで設定する¹³。そして、データは解析がしやすい形式に整えていく。R においては、データサイエンティストの Hadley Wickham が提唱する整然データ (Wickham & Grolemund, 2016) の形式に整えるとよい。整然データは、1つのセルには個々の値が入り、1つの列には個々の変数が入り、1つの行には個々の観測が入るという原則の元で整えられたデータである¹⁴。整然データ

13 ここでは深入りしないが、変数名やファイル名は、単語をアンダースコアでつなぐスネークケース (例: kunisato_yoshihiko)、一番最初の単語以外の単語の最初の文字を大文字にするキャメルケース (kunisatoYoshihiko) などさまざまな書き方がある。言語によっても好まれる書き方は異なるので、使用言語のスタイルガイドを参考に選択すると良い。なお、R の場合は、Hadley Wickham 氏の The tidyverse style guide (<https://style.tidyverse.org/>) が分かりやすいかと思われる。

になっていれば、人間も読みやすく、容易に解析に利用することができる。これらの生データから解析可能なデータへの整形過程もその後の解析もデータに対して行った処理をすべて残すようにする。行った操作がスクリプトに残るので、GUIベースの統計ソフトよりもRなどのCUIベースのソフトはこのような場合に適している。また、Rの場合はR Markdownを用いると、データに対して行った操作だけでなく、なぜそれを行ったかなどのコードに関連したメモなどを容易に残すことができる¹⁵。なお、心理学においても脳波データと質問紙などのように生データが分かれているデータセットを用いる場合がある。その場合に、それぞれのデータセットの対応づけができるように、ユニークな識別子を使う。データセットによってユニークな識別子の書き方が異なると対応づけが難しくなるため、データセット間で共通の名前を使う。

最後に、ほかの研究者がアクセスし引用できるように、公開するデータは個人のサイトではなくfigshare (<https://figshare.com/>) や Zenodo (<https://zenodo.org/>) などのような比較的永続性のあるリポジトリで公開する (Gewin, 2016)。これらのリポジトリでは、DOIが付与されるので、恒久的に引用が可能になる。公開にあたっては、データや解析コードだけでなく、それらの説明文章を含めて、後述する解析のパッケージ化したものを公開する。このようにして公開されたデータや解析コードは、ほかの研究者が利用しやすいものになり、データの再利用を含めた研究の進展、研究の引用の増加などが期待される。なお、データ公開については、データ取得前に研究参加者にその同意をとる必要がある。そ

の際に、多くの心理学研究の場合は、匿名性が確保されることを前提にしていることが多い。そのため、氏名などの個人情報はそもそも取得していないと思うが、取得されたデータの組み合わせから個人を特定されないか気をつける必要がある。例えば、ある大学の心理学の授業の受講生に協力を依頼して調査を行った場合、そのなかで1名だけ64歳の受講生がおり、その受講生のデータも公開データに含まれたとする。そのデータがどこで取得されたかは明記しなくとも著者の所属などから自明に思われる場合は、公開データの年齢データから、その64歳の受講生のデータを特定することができる。このように、個人情報が無い場合であっても、変数の組み合わせや推測を可能にする情報などから個人が特定できる場合は、当該変数を削除したり、層化するなどの工夫をする必要がある。オープンデータが普及してきてはいるが、安易に公開するのではなく、個人情報が保護されているかは慎重に確認する必要がある。

4.2. 解析のパッケージ化

解析可能なデータであっても、その解析コードをそのまま共有しても、第三者が利用可能なものになるとは限らない。データが複数ある場合には、それぞれどういうデータか説明する文章が必要になる。また、解析コードもそれがメインの解析のコードなのか、メインの解析を行うにあたって作成した関数のコードなのか分からないと第三者は利用できない。例えば、図2の左のような状態のフォルダを共有されても困るわけである。図2の左のフォルダには、データが3つあるがそれぞれがどういうデータかファイル名だけでは分からない。おそらく

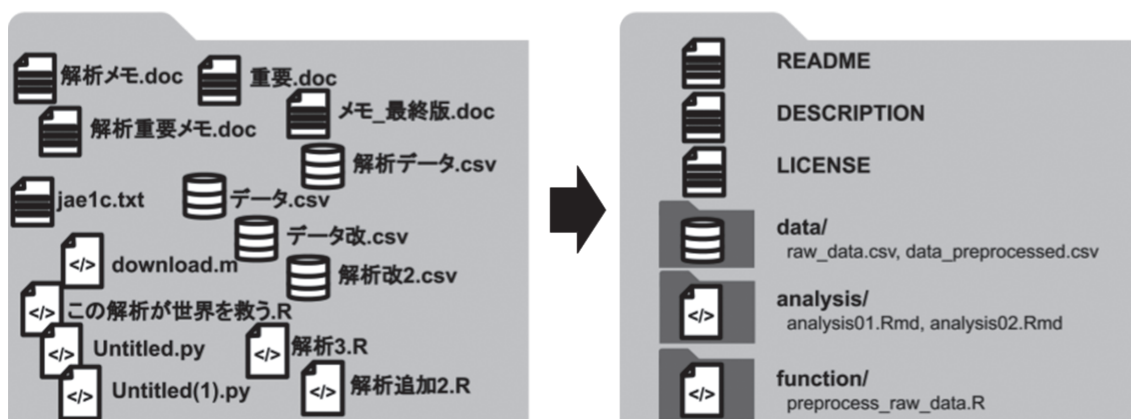


図2 解析のパッケージ化の例

14 整然データの定義にも微妙なブレがある。詳しくは、西原 (2017) を参照のこと。

15 R Markdown については、高橋 (2018) を参照のこと。

「データ改.csv」は「データ.csv」に何か手を加えたものに思えるが、確信は持てない。また、解析コードも R コードが3つ（1つは恥ずかしいし、無意味な名前がついている）、Matlab コードらしきものが1つ、たまたま生成されてしまったような Python コードが2つある。解析コードの名前から、どの解析コードがメインの解析のためのコードかわからない。これではどうしていいかわからないので、解析やデータに関するドキュメントらしきものを探さることになるが、「重要.doc」や「メモ_最終版.doc」などがあってどれが一番重要かわからない。もちろん、すべてのファイルを確認すればよいが、ファイル間に不一致がある場合は、著者に直接問い合わせが必要になる可能性が高い。

どのように整理すればわかりやすいのかについては、あまりコンセンサスはなく、各研究者によって創意工夫がなされているような現状がある。そこで、近年は、再生可能なデータ、解析コード、関連文書のまとめかたに関する枠組みとして、research compendium (<https://research-compendium.science/>) という概念が提案されてきている¹⁶。research compendium は、(1) ファイルを研究者コミュニティにおいて広く流布された方法でまとめる、(2) データ、方法、出力を明確に分けた上で、それらの関係について明示する、(3) 解析が実施された解析環境を明確にするという原則に基づいている。この原則に基づいて、Marwick, Boettiger, & Mullen (2018) は、R パッケージ開発で用いられる枠組みを research compendium として採用することを提案している。R パッケージ開発に限定する必要はないが、R パッケージ開発では合理的な配置でファイルを構成してパッケージを作成することになるので、どのような形式かあまり悩む必要がなく、決められた型に自身のデータや解析コードなどを配置すれば良いので利便性が高い (Marwick et al., 2018)。research compendium に従うと、データや解析コードを単に共有するのではなく、それらをパッケージ化して配布することになる。現状においては、各研究者が工夫してデータと解析コードの公開をしているが、パッケージ化の方法が決まっている方が、研究者個人の負担も小さく済むのと、その後のメタ分析などのデータ統合においても同じファイル構造を採

用しているデータは統合が容易になることが予想できる。

Marwick et al. (2018) に従って整理すると、図2の右のようなフォルダ構成になる。プロジェクトのフォルダを構成するサブフォルダやファイルは以下のとおりになる。

- ・ README 第三者がプロジェクトの全体像を理解できるように説明をするファイル。どの data に対して、どの analysis を用いるのかなどを記載する。
- ・ DESCRIPTION プロジェクトのメタデータや依存関係を記載するファイル。
- ・ LICENCE 公開したファイルは誰がどのように使えるものなのか説明するファイル。
- ・ data/ プロジェクトで使用したデータを配置するフォルダ。
- ・ analysis/ プロジェクトで使用した解析コードを配置するフォルダ。解析コードが複数ある場合は、どういう順番で実行するのか分かるようなファイル名にしておく。
- ・ function/ プロジェクトで使用した関数を配置するフォルダ¹⁷

図2の右のフォルダのようにデータと解析コードを配置すれば、第三者は README を読んで、analysis フォルダの analysis01.Rmd から順番に再解析ができる。関心やその後の利用に併せて、DESCRIPTION, LICENCE, function フォルダなどを確認することもできる。具体例として、著者の研究室で取り組んだ追試研究のデータと解析コードの共有リポジトリ (https://github.com/ykunisato/somatori_kunisato_2019_replication_study) を挙げる。リンク先の Github リポジトリの README を読むと、どのようにすれば解析結果を再現できるのか分かるようになっている。また、DOI も取得しているので、このリポジトリ自体の引用も可能になる。今後、research compendium がデータや解析コードの共有における標準になるかは未知数ではあるが、独自の方法ではなく、ある程度共有可能な枠組みで解析のパッケージ化は必要になってくると思われる。

4.3. 解析環境のコンテナ化

適切に管理されたデータを用いて、解析のパッケージ化ができていれば、再解析の再現可能性は担保されそう

16 research compendium をそのまま直訳すると研究概要や研究要約となり、既にある用語と重なってしまう。そのため、research compendium を解析梱包や研究大要とすることも考えたが、意味が取りにくくなるため、日本語訳をせず、同様な概念として解析のパッケージ化を用いた。

17 Marwick et al. (2018) では、R/となっているが、研究で使用する言語は R とは限らない。汎用性をもたせるために、ここでは function/ としている。

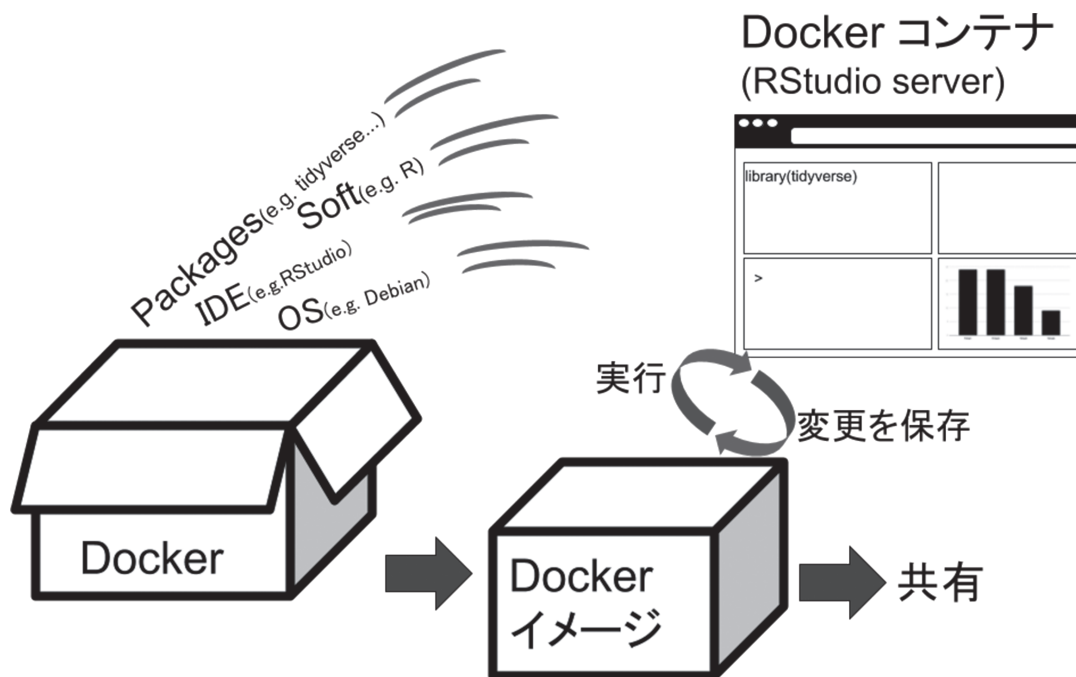


図3 解析環境のコンテナ化の例

に思われるが、解析ソフトウェアやOSなどの解析環境が違くと解析コードが適切に動作しなかったり、結果が再現しなくなったりする。再解析をする場合に、使用する解析ソフトウェアやそれが依存しているソフトウェアすべてのバージョン、OSの種類とバージョンなどを元の環境と揃えることができればよいが、現実的にはかなり難しい¹⁸。そこで、Dockerなどのコンテナ技術を使って、最小限の構成で解析ソフトウェアが動作するようなコンテナを用意して共有することで再現可能性を担保する方法が提案されてきている。Dockerとは、Docker社 (<https://www.docker.com/>) が提供するコンテナ型仮想化技術になる。コンテナとは、ソフトウェアが実行するのに必要なOSやライブラリなどを用意した仮想環境であり、Dockerを実行しているコンピュータ上でコンテナとして隔離されて動作させることができる。Dockerは、Dockerfileと呼ばれるコンテナ内に入れる内容について記載したレシピを元にしてイメージを作成し（ビルドと呼ぶ）、それをもとにコンテナを実行する（図3）。Dockerfileイメージは、Dockerfileの設定によっては、ビルドする時点によって異なる可能性があるが（例えば、最新版のソフトウェアをインストールする設定だとビルドするタイミングでバージョンが変わる）、一度ビルドしたイメージについては、コンテナ内

で完結して動作し、バージョンなども含めていつでも解析環境が再現可能になる。なお、コンテナに行った変更（新たにRパッケージを追加するなど）はイメージに反映されないので、コンテナで解析を行ってすべての作業が終了した段階でイメージとして保存して共有すると解析で使った環境をそのまま共有できる（図3）。このように、コンテナ技術は解析環境の再現可能性を担保する上では有用である。また、Dockerなどのコンテナ型仮想化技術は、従来の仮想マシンよりも容量が小さく済むため、解析環境を共有したい場合に有用である。現在、バイオインフォマティクスの領域ではDockerの活用が進んできており（Kulkarni et al., 2018）、心理学でも活用が提案されつつある（Clyburne-Sherin, Fei, Green, & Ocean, 2018; Epskamp, 2019）。

RのためのDockerコンテナとしてRockerが開発されている（Boettiger & Eddelbuettel, 2017）。Rockerではさまざまなコンテナが用意されており、ベースとなるRのはいったコンテナ（r-ver）、よく使用される統合開発環境（Integrated Development Environment: IDE）のRstudioも入ったコンテナ（rstudio）、さらにtidyverseパッケージを追加したコンテナ（tidyverse）などがある。このように、すでにあるコンテナにIDEやパッケージを追加するような形でDockerfileを書くことができる。Rockerを使用していて、足りないパッケージがあれば、オリジナルのDockerfileにそれらを書き足せばよい。著者もRockerのverseコンテナ

¹⁸ Rでは、checkpoint, renv, packratなどのRパッケージを使って、解析に使用したRパッケージや依存パッケージのバージョンを管理できる。

(java, tex などドキュメント作成に必要なものが追加されたコンテナ) をベースにして、心理学研究に必要なパッケージや Stan (Carpenter et al., 2017) も利用できるようにした Docker イメージやネットワークメタ分析、心理学的ネットワークモデリング、認知モデリングに特化した Docker イメージなどを作成している。なお、Dockerfile を元にビルドしたイメージを配布する場として dockerhub (<https://hub.docker.com/>) がある。Rocker も著者の Docker イメージも dockerhub で配布しており (<https://hub.docker.com/u/ykunisato>), 自分で作成しなくても利用することができる。

Docker の利点として、コンテナの性質を活かした解析環境の再生可能性を挙げたが、同時に解析環境の構築にかかる負担を小さくすることができる点も利点として挙げることができる。例えば、Stan や JAGS などはローカル PC の OS や設定によっては何らかの工夫をしないとインストールができないことがある。パソコン操作に不慣れなユーザーはパッケージの使い方の前の段階で挫折してしまうことになる。また、多数のパッケージを使う場合に、新たに R をインストールするたびにそれらのパッケージのインストールもしなくてはならなくなる。Docker を使えば、数ステップでいつも使う解析環境を用意することができる。この利便性は、ローカル PC だけでなく、Amazon Web Services や Google Cloud Platform を用いたクラウドコンピューティングを行う場合にも使うことができ、Docker を使ってクラウド上に素早く解析環境を用意することができる。R などのワークショップにおいて、最初に参加者の PC に R やさまざまなパッケージをインストールするのに多くの時間をかけてしまうことがあるが、Docker を使ってクラウド上に解析環境を用意しておけば、スムーズに R 自体の学習に入ることができる。Docker の利点は多いが、いくつかターミナルでのコマンド操作が必要になる点や Dockerfile の作成には Linux の知識が必要になる点など敷居が高く見える欠点もある。それに対して、GUI によって Docker と同様の取り組みがブラウザ上で可能な Code Ocean (<https://codeocean.com/>) のようなサービスも提案されてきている (Clyburne-Sherin et al., 2018)。解析環境のコンテナ化については、心理学における実践が始まったばかりにはなるが有用なサービスも出てきており、今後も定期的なフォローアップが必要と思われる。

5. おわりに

心理学研究において、研究環境の整備は、それぞれの研究者に委ねられてきた。そのことは、研究室の特色にもなっていたのかもしれない。しかし、研究実施にかかる環境の整備も高度化し、社会的な要請から再現可能な研究実践も求められるようになってきている。事前登録、オープンサイエンス、解析のパッケージ化、解析環境のコンテナ化と対応すべき技術は目まぐるしく現れている。これに対して個々の研究者の自助努力のみで対応するのは難しいかもしれない。今後は、心理学研究における研究環境の基盤整備にかかる研究の実施やそのような知識・技術を共有する場や機会がますます重要になってくるだろう。

謝辞

解析の再生可能性における Docker の意義や活用法について、前田和寛氏 (LINE Fukuoka 株式会社) からご教授いただきました。ここに感謝の意を表します。本論文の動機の核の部分は、前田氏から頂いた数々のご指摘によって徐々に形成されたものになります。

引用文献

- Avidan, M. S., Ioannidis, J. P. A., & Mashour, G. A. (2019). Independent discussion sections for improving inferential reproducibility in published research. *British Journal of Anaesthesia*, 122 (4), 413-420.
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533 (7604), 452-454.
- Barba, L. A. (2018). Terminologies for Reproducible Research. *arXiv*. Retrieved from <http://arxiv.org/abs/1802.03311>
- Boettiger, C., & Eddelbuettel, D. (2017). An Introduction to Rocker: Docker Containers for R. *The R Journal*, 9 (2), 527-536.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., ... STARD Group. (2015). STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*, 351, h5527.
- Buckheit, J. B., & Donoho, D. L. (1995). WaveLab and Reproducible Research. In A. Antoniadis & G. Oppenheim (Eds.), *Wavelets and Statistics* (pp. 55-81). New York, NY: Springer New York.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76 (1). Retrieved from <https://www.osti.gov/biblio/>

- 1430202
- Claerbout, J. F., & Karrenbach, M. (1992). Electronic documents give reproducible research a new meaning. *SEG Technical Program Expanded Abstracts 1992* (pp. 601-604). Society of Exploration Geophysicists.
- Clyburne-Sherin, A., Fei, X., Green, S., & Ocean, C. (in press). Computational Reproducibility via Containers in Psychology. *Meta-psychology*.
- Epskamp, S. (2019). Reproducibility and Replicability in a Fast-Paced Methodological World. *Advances in Methods and Practices in Psychological Science, 2* (2), 145-155.
- Gewin, V. (2016). Data sharing: An open mind on open data. *Nature, 529* (7584), 117-119.
- Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences, 1396* (1), 5-18.
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine, 8* (341), 341ps12.
- Grant, S. P., Mayo-Wilson, E., Melendez-Torres, G. J., & Montgomery, P. (2013). Reporting quality of social and psychological intervention trials: a systematic review of reporting guidelines and trial publications. *PLoS One, 8* (5), e65442.
- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsson, G., Banks, G. C., Kidwell, M. C., ... Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science, 5* (8), 180448.
- 池田 功毅・平石 界 (2016). 心理学における再現可能性危機: 問題の構造と解決策. *心理学評論, 59* (1), 3-14.
- 稲田 尚子 (2015). 尺度翻訳に関する基本指針. *行動療法研究, 41* (2), 117-125.
- Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., & Ioannidis, J. P. A. (2016). Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biology, 14* (1), e1002333.
- Kulkarni, N., Alessandri, L., Panero, R., Arigoni, M., Olivero, M., Ferrero, G., ... Calogero, R. A. (2018). Reproducible bioinformatics project: a community for reproducible bioinformatics analysis pipelines. *BMC Bioinformatics, 19*, 349.
- 国里 愛彦 (2015). 系統的展望とメタアナリシスの必須事項. *行動療法研究, 41* (1), 3-12.
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., ... Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology, 62* (10), e1-e34.
- Marwick, B., Boettiger, C., & Mullen, L. (2018). Packaging Data Analytical Work Reproducibly Using R (and Friends). *The American Statistician, 72* (1), 80-88.
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., ... Altman, D. G. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ, 340*, c332.
- Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., ... de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Medical Research Methodology, 10*, 22.
- 西原 史暁 (2017). 整然データとは何か. *情報の科学と技術, 67* (9), 448-453.
- Nosek, B. A., & Ebersole, C. R. (2018). The preregistration revolution. Proceedings of the. Retrieved from <http://www.pnas.org/content/early/2018/03/08/1708274114>. shosh
- 奥村 泰之 (2014a). アクセプトされる失敗した臨床試験の粉飾法: 無駄のない研究推進のためのピアレビュー研究. *Monthly IHEP, 301*, 23-28.
- 奥村 泰之 (2014b). 非薬物療法の介入研究の必須事項. *行動療法研究, 40* (3), 155-165.
- 奥村 泰之 (2017). 粉飾された臨床試験の判別法: 臨床試験のすべての関係者へ. *臨床評価, 45* (1), 25-34.
- 奥村 泰之・原井 宏明・谷 晋二・佐藤 寛 (2014). 研究報告の質向上に向けて. *行動療法研究, 40* (3), 151-154.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349* (6251), aac4716.
- Peng, R. D., Dominici, F., & Zeger, S. L. (2006). Reproducible epidemiologic research. *American Journal of Epidemiology, 163* (9), 783-789.
- Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 27* (5), 1147-1157.
- Riley, D. S., Barber, M. S., Kienle, G. S., Aronson, J. K., von Schoen-Angerer, T., Tugwell, P., ... Gagnier, J. J. (2017). CARE guidelines for case reports: explanation and elaboration document. *Journal of Clinical Epidemiology, 89*, 218-235.
- 柚取 恵太・国里 愛彦 (印刷中). アンヘドニア (anhedonia) と遅延割引: Lempert & Pizzagalli (2010) の追試. *心理学評論*
- Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*

- es of the United States of America*, 115 (11), 2584-2589.
- 高橋 康介 (2018). Wonderful R 3: 再現可能性のすゝめ: RStudioによるデータ解析とレポート作成 共立出版
- 竹林 由武 (2014). 観察研究の必須事項 行動療法研究, 40 (3), 167-175.
- 谷 晋三 (2015). 症例研究の必須事項 行動療法研究, 41 (1), 13-18.
- 土屋 政雄 (2015). 尺度研究の必須事項 行動療法研究, 41 (2), 107-116.
- Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Annals of Internal Medicine*, 147 (8), 573-577.
- Wallach, J. D., Boyack, K. W., & Ioannidis, J. P. A. (2018). Reproducible research practices, transparency, and open access data in the biomedical literature, 2015-2017. *PLoS Biology*, 16 (11), e2006930.
- Wickham, H., & Grommund, G. (2016). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc.
- Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLoS Computational Biology*, 13 (6), e1005510.