

特集再現可能性に向き合う心理学

データ・マテリアル・ 分析スクリプトのオープン化が拓く 心理学の未来

武藤拓之 むとう ひろゆき

京都大学人と社会の未来研究院(認知心理学)

<https://mutopsy.net/> h.muto@zm.commuifa.jp

経験科学にとってデータは神様である。データを取得するための実験や調査を構成するマテリアル、データから価値のある情報を引き出すための分析スクリプトもまた、実証研究にとって欠かせない道具である。本稿ではオープンサイエンスの視点に立って、データ・マテリアル・分析スクリプトのオープン化が心理学の信用性向上にどのように貢献するのかを具体的に論じた上で、乗り越えるべき障壁についても考える。

広義のオープンデータと 科学的知見の信用性

近年、科学研究をより開かれたものにすることを目指すオープンサイエンスの機運が高まりつつある。オープンサイエンスは論文のオープンアク

セス化や市民科学などを含む広範な理念であるが¹、本稿ではその要素の1つである広義のオープンデータに着目する。広義のオープンデータは文字通り、研究に関連した(広義の)データを第三者が利用できるように公開する営みのことである。わざわざ「広義の」と付けたのは、いわゆる観測データだけでなく、実験・調査で用いたマテリアルや分析スクリプトの公開もこの営みに含まれるからである。論旨を明確にするために、本稿におけるデータ・マテリアル・分析スクリプトの意味を右の欄に整理したので参照されたい。

広義のオープンデータは、科学的知見の信用性(credibility)を担保するためのお手軽で強力な方法である。その理由は次節で詳しく論じるが、その本質は、研究で得られた知見の確からしさを未来永劫誰でも自由に検証できるようにするという点にある。データ・マテリアル・分析スクリプトがいずれも十分利用できない「閉じた」研究を想像

The Future of Psychology Opened by Open Data, Open Materials, and Open Analysis Scripts

Hiroyuki Muto

してみよう。そのような研究は検証も反証も容易ではないため、苦労を重ねて追試をするか、易きに流れて盲信するかのいずれかの選択を余儀なくされるだろう。また、広義のオープンデータには、実施に要するコストが低いという利点もある。研究者がやるべきことは、電子ファイルをオンライン上のリポジトリにただアップロードするだけである。もちろん工夫しようと思えばいくらでもコストを割くことはできるが、最小限でよければほとんど手間はかからない。さらに、研究分野の慣習や現実的な制約にもとづいてどこまで公開するかを柔軟に決定することもできるし、論文の投稿前でも投稿後でもいつでも実践することができる（データを観測する前に可能な限り厳格に実施することが求められる事前登録と比較するとわかりやすい）。信用性を高めたいが、何から手を出したらよいかわからない、

データ——主に実験や調査を通じて観測される、量的または質的な値の集まりのこと。心理学では実験における参加者の反応や質問紙に対する回答などが含まれる。データは大雑把に、実験・調査で直接観測される生データと、そこから必要な情報を抽出し要約した集計データに分けることができる。オープンサイエンスの観点からは、情報量の多い生データを公開するのが理想的であるが、倫理的制約などによりそれができない場合には情報を減らすことも視野に入れる必要がある。

マテリアル——実験や調査などのデータ取得の手続きを構成するあらゆる具体的な要素のこと。何をマテリアルに含めるかは研究分野によって異なるが、本稿では実験で提示する刺激や実験を制御するプログラム、調査項目などの電子的に共有可能なものを念頭に置く。

分析スクリプト——データ分析(前処理を含む)のプロトコルを正確に記した電子ファイルのこと。具体的には、RやSASのような統計ソフトウェアで実行可能なコードがこれにあたる。分析スクリプトを残しにくいGUIベースの統計ソフトウェアはオープンサイエンス的にはやや不向きであるが、プロトコルを正確に共有する手段があるのであれば問題はない。

と悩む研究者にお勧めの方法である。

信用性の4つの評価次元から見たオープン化の役割

それでは、広義のオープンデータがどのように心理学の信用性向上に貢献するかを具体的に見ていこう。本節では、科学的知見の信用性を定量的に検証するためにLeBel²が導入した枠組みにもとづいて、信用性の4つの評価次元に対するデータ・マテリアル・分析スクリプトのオープン化の役割を論じる。その評価次元とは、方法とデータの透明性・分析の再生可能性・分析の頑健性・効果の再現可能性の4つである。これらの次元は互いに独立ではなく相互依存の関係にあり、基本的には前の次元を満たすことが後の次元を評価するための前提条件となるが、例外もある。本節では再生可能性・頑健性・再現可能性・一般化可能性といった用語を用いるが、これらは少し紛らわしいのでその意味を表1に整理した。以下の項を読まれる際に適宜参照されたい。

方法とデータの透明性

最初の評価次元は方法とデータの透明性 (transparency) である。透明性は、研究デザイン・データ・分析方法の詳細がどの程度開示されているかを指す概念であり、信用性の根幹を成す。

方法の透明性を高める最も素直な方法は、論文にできるだけ詳しく方法を記述することである。これに関連する取り組みとして、例えば心理学のトップジャーナルの1つである *Psychological Science* 誌は方法と結果のセクションの語数制限を2014年以降撤廃している³。透明性の高い報告を行うためのガイドラインも役に立つ。例えば、ランダム化比較試験の報告基準である CONSORT 声明⁴や、Simmons⁵が心理学者に向けて提案した 21 Word Solution、LeBel⁶が提供している PsychDisclosure.org というプラットフォームなどが参考になる。

とはいえ論文内での記述にはどうしても限界があるように思われる。筆者の経験からも、論文を

表 1—科学的知見の信用性に関する用語の整理

分析の再生可能性・分析の頑健性・効果の再現可能性・効果の一般化可能性は、検証に用いる研究デザイン(e.g., 実験操作, 課題, マテリアル)・データ・分析方法がオリジナルの研究と同じか異なるかにもとづいて定義することができる。この表は LeBel ら²⁾の Supplementary materials の Table 1 を参考に作成した。

	分析の再生可能性	分析の頑健性	効果の再現可能性	効果の一般化可能性
研究デザイン	同じ	同じ	同じ	異なる
データ	同じ	同じ	異なる	異なる
分析方法	同じ	異なる	同じ	異なる

読んでいて方法やデータの記述が不十分であると感じたことは少なくない。事前審査付きの追試研究を実施した抽取と国里⁷⁾のエピソードは、方法の不透明性が問題になることを示す具体例を提供するものである。彼らのはうつ病の症状と意思決定の関連を調べた先行研究の追試を試みる前に論文の記述の不明瞭さに気付き、その論文の第一著者に問い合わせたところ、論文に記載されていた意思決定課題の手続きが実際のものとは異なっていたことが明らかになったのである。この事例は論文の誤りに気付くことのできた幸運な例であるが、同様のエラーはあらゆる研究で起こりうるし、必ずしもこの事例のように誤りに気付ける保証はない。

そこでオープンサイエンスが力を発揮する。論文の記述の不正確さと曖昧さはデータ・マテリアル・分析スクリプトそのものを公開することで容易に補完することができる。特にマテリアルと分析スクリプトは研究の実施前に準備することができるので、研究計画とともに事前登録しておくことでより透明性を高めることができる。

オープンデータによる透明性の担保が研究不正の発覚に貢献した事例として、著名な行動経済学者である Dan Ariely らのチームが行った一連の研究とその顛末を紹介する。2012年に米国科学アカデミー紀要に掲載された Shu らの論文⁸⁾は、人に情報提供を求める際には最後に署名させるよりも最初に署名させたほうが正直な回答が得られやすいという仮説を支持する3つの実験を報告したものであった。この論文で報告された3つ目の実験は、自動車保険の顧客が保険会社に年間走行距離を申告する際の署名欄の位置を操作した

ところ、最後に署名した群よりも最初に署名した群のほうが平均的に長い走行距離を申告した、というものであった。嘘をついて走行距離を短く申告したほうが保険料は安くなるので、最初に署名することで正直な申告が増えたと解釈できる。ところが、この論文の著者を共著者に含む2020年の追試論文⁹⁾では署名の効果は再現できなかった。先ほど紹介した走行距離のデータに不自然な点があることも明らかになった。追試論文の著者らは元の論文と追試論文の両方のデータをオープンにしていたため誰でもデータを検証することが可能であった。そこで有志らがデータを精査したところ、保険会社から提供されたとされる走行距離のデータが捏造されていたことを示す決定的な証拠が見つかり¹⁰⁾、2012年の論文は撤回される運びとなった。もし元論文の著者らがデータの公開を拒んでいたら真実は闇の中であっただろう。この事例は、研究結果を適切に解釈し評価する上で透明性がいかに重要であるかを物語っている(本稿の範囲外ではあるが、署名の効果の研究は現実適用を念頭においたものであったため、本誌掲載の平石の論文¹¹⁾で議論されているエビデンスの観点からも興味深い事例である)。

分析の再生可能性

2番目の次元である分析の再生可能性(reproducibility)は、同じデータを同じ方法で分析したときに同じ結果が得られることを保証するものである(文献によってはこれを「再現可能性」や「再現性」と呼ぶものもあるが、後述の replicability とは意味が異なるので注意)。再生可能性は素朴に考えると当たり前前に達成されるように思われるかもしれないが、実際はそうでもない。例えば、Artner ら¹²⁾が2012年のアメリ

カ心理学会機関誌から無作為に抽出した46本の論文に含まれる232件の統計的な主張の再生可能性を検証したところ、元論文の結果を正しく再生できたのは232件中163件(70%)だけであり、しかもそのうち18件(全体の8%)は論文の記述と異なる分析方法を用いなければ正しく再生できなかった。さらに衝撃的なことに、元の論文で「有意」と報告されていた185件の主張のうち13件(7%)はArtnerらの再分析では有意とならなかった。このように、再生可能性は決して自明ではなく、同じデータの再分析によって元の研究の結論が覆ってしまうことすらあるのである。

再生可能性が損なわれる主な原因としては、論文の記述の不正確さ・曖昧さや分析方法の誤り(プログラムミスを含む)などが挙げられる。このようなエラーそのものをなるべく減らすように努力するのが最も素直で直接的な解決策であるが(これに関する技術的なアプローチは高橋¹³などを参照)、研究者も人間である以上、エラーを完全に排除することは難しい。その意味で、万が一エラーが発生してしまったとしても第三者がそれを検証できるようなセーフティネットがあると安心である。

そのセーフティネットこそがデータと分析スクリプトのオープン化である。オープンデータは第三者による再生可能性の検証を可能にし、オープンスクリプトは分析方法の誤りを発見しやすくしてくれる。どちらか一方だけでも有用ではあるが、データと分析スクリプトの両方をオープンにすれば、再生可能性が担保されていることを積極的に示すことができるうえ、第三者による検証の手間を軽減することもできる。

分析の頑健性

3番目の次元は分析の頑健性(robustness)である。これは、同じデータに対して異なる分析方法を適用しても結論が大きく変わらないことを保証するものである。例えば、心理学で頻繁に用いられる行動指標である反応時間(刺激が提示されてから反応を行うまでの時間)の分析でさえ、外れ値の基準や変数変換の方法、代表値の計算方法などの選択によっ

て結論が大きく変わりうることがわかっている¹⁴。また、「Many analysts, one data set」と題されたSilberzahnらの論文¹⁵では、「サッカーの審判は肌の白い選手よりも肌の黒い選手に対してレッドカードを出しやすい」という仮説を複数の研究チームが同一のデータセットを用いて独立に検証したところ、29チーム中20チーム(69%)が統計的に有意に仮説を支持する結果を報告したのに対し、残りの9チーム(31%)による分析結果は有意ではなく、推定された効果の大きさもチーム間で大きく異なっていたことが報告されている。厄介なことに、唯一の「正しい」分析方法なるものはほとんどの場合存在しないため、客観的な基準にもとづいて「最良の」分析方法を決定することは事実上不可能である。このような分析の自由度に対してどのように向き合うかが分析の頑健性を考える上で鍵となる。

分析の自由度を減らすための代表的な方法として事前登録とマルチバース分析が挙げられる。事前登録は、データを観測する前に分析方法を確定させることで自由度を減らす方法である(詳しくは本誌掲載の山田の論文¹⁶を参照されたい)。この方法は、都合のよい分析方法と結果の組み合わせを恣意的に選んで報告するチェリーピッキングを防ぐという意味では有効であるが、事前に決めた分析方法が最良の方法であるという保証はどこにもないため、分析の頑健性に対してはほとんど無力である。対照的に、考えうるすべての分析を実施することで特定の結論が得られる境界条件を画定する方法がマルチバース分析である¹⁷。文字通り「すべて」の分析方法を実行すればもはや自由度は存在しない、という理屈である。マルチバース分析は頑健性を直接的に検証することができる優れた方法ではあるが、計算コストが高いうえ、報告のために多くの紙面を必要とするため、研究者の負担が大きいのが難点である。

オープンサイエンス、特にデータのオープン化は、事前登録とマルチバース分析の限界を補う形で分析の頑健性を検証可能にしてくれる。これは、マルチバース分析を未来の誰かに「外注」するよ

うなイメージである。例えば、ある研究者が特定の分析方法を事前登録した上で研究を実施して論文を出版し、その後でその研究の重要性に気付いた別の研究者がオープンデータを使って頑健性を検証する、といった方法がとれば非常に効率的である。また、論文出版時には知られていなかった新しい(未来の)分析方法を用いて頑健性を確認することも可能になる。データだけでなく分析スクリプトもオープンにしていれば、元の分析方法との差異を明確にできるため解釈の助けになる。

オープンデータが頑健性の検証に役立つ例として、2019年に*Nature*誌に掲載されその後撤回された、Whitehouseらによる宗教認知科学の論文(通称「神論文」)が挙げられる¹⁸。Whitehouseらのチームは人類の進化史に関するオープンな大規模データベース「Seshat」を構築し、そのデータを用いた分析により、道徳を説く神が複雑な社会の出現後に誕生したと主張した。ところがその後、この研究の分析方法(欠損値の処理方法)に疑義をもったBeheimら¹⁹は、同じデータを異なる方法で分析すると、複雑な社会の出現よりも「前」に神が誕生したというまったく正反対の結果が得られることを明らかにした。この研究の場合、追試による検証はデータの収集コストの観点から困難であったと思われるが、データがオープンであったがゆえに頑健性の観点から健全に反証することができたのである。

効果の再現可能性

4番目の次元である効果の再現可能性(replicability)は、元の研究とは独立なサンプルを用いた追試において元の研究結果と本質的に同じ結果が得られることを保証するものである。追試には、元の研究とまったく同じ研究デザインと分析方法を用いて効果の再現可能性を検証する直接追試と、元の研究と概念的に同一の対象を異なるデザインで検証して効果の一般化可能性(generalizability)の検証を試みる概念的追試に大別できる。ただし、両者は厳密に区別できるものではなく、実際には連続体の一部をそのように呼んでいると考えられる

(例えば、刺激や使用言語が元の研究と異なっても「直接追試」と呼ぶことがある)。心理学における追試の動向と意義に関しては本誌掲載の三浦の論文²⁰をご参照いただくとして、ここでは広義のオープンデータが追試研究(特に、再現可能性の検証を目的とする直接追試)の効率化に果たす役割に焦点を当てて議論する。

論文の記述のみから追試研究を計画することも可能ではあるが、その記述が不正確または曖昧である可能性(柚取・国里⁷の例を思い出そう)や、そもそも分析の再生可能性と頑健性が不十分である可能性を考えると、論文の記述以上の情報が欲しくなる。マテリアルと分析スクリプトが利用できれば、それをそのまま(あるいは少し修正を加えて)追試のために利用することができる。データが利用できれば、その情報を用いてサンプルサイズ設計や細かなパラメータの調整に役立てることができる。さらに、データと分析スクリプトの両方が利用できれば、追試研究の実施前に分析の再生可能性と頑健性を確認することができる。これらに問題があることが事前にわかれば、問題を克服するための概念的追試を実施したり、あるいは追試自体をやめるといった意思決定を下すことができる。オープン化は透明性・再生可能性・頑健性の次元を介して追試研究の効率化を実現してくれるため、結果として再現可能性の検証に役立つのである。

● オープン化の障壁を乗り越えるために

前節でみた信用性の4つの評価次元に対する貢献を考えると、データ・マテリアル・分析スクリプトをオープンにしない理由はほとんどないように思われる。にもかかわらず、2014年から2017年の間に出版された心理学の論文を調査した研究²¹によると、生データのオープン化は188本中4本(2%)、マテリアルのオープン化は183本中26本(14%)、分析スクリプトのオープン化は188本中1本(1%)でしか行われていなかった。2022年現在では多少なりとも改善していると思われる(思いたい)が、まだまだ浸透しているとはとても言えない状況である。何がオープン化の障壁

となっているのだろうか。

考えられる障壁の1つとして倫理的な懸念が挙げられる。特に、臨床心理学や教育心理学のようなセンシティブなデータを扱うことの多い研究分野では大きな課題となりうる。確かに個人情報やプライバシーを含むデータをそのまま公開することはできないが、データに仮名化や匿名化を施して技術的に対処することは可能である²²。プライバシー保護はデータのもつ情報の一部を捨てることによって達成されるため(プライバシーと有用性のトレードオフ)、プライバシー保護によって分析の再生可能性や頑健性の検証が難しくなる面もあるが、それでもデータを一切公開しないよりは遥かにマシである。倫理的な障壁を乗り越えるためには、インフォームド・コンセントの取得方法を含むオープンデータに関する倫理規定と具体的なプロトコルを学界全体または研究分野ごとに整備・周知することも必要であろう。

もう1つの障壁として、研究者個人から見たリスクとインセンティブの問題が挙げられる。オープンサイエンスは科学全体の健全な発展にとっては有用であるが、研究者個人にとっては必ずしも良いことばかりではない。オープン化は研究の検証可能性、もっと言えば反証可能性を高める営みなので、研究の中身を曝け出すことは第三者による致命的な批判や「粗探し」を許容することを意味し、場合によっては論文の撤回を強いられるリスクもある。いくら健全な科学のためとはいえ、自身のキャリアパスや精神的健康のことを考えると躊躇するものも頷ける。このような葛藤がある以上、制度的にオープン化を促すことが不可欠であるように思われる。すでにいくつかの雑誌が採用している具体的な取り組みとしては、オープンデータの義務化やオープンサイエンス・バッジの付与などが挙げられる。このような取り組みが普及していけば、データ・マテリアル・分析スクリプトのオープン化が心理学の新しい「常識」になる日が訪れるかもしれない(倫理審査が実際にそうなったように)。

重要なのは、オープンサイエンスは「やる」か

「やらない」かの二択ではなくて、「どのくらいやるか」という程度問題として捉えるべきものであるということである。生データを公開できないのであれば、公開できるように処理を施したデータを公開すればよい。データを一切公開できない場合でも、マテリアルや分析スクリプトの公開を検討することはできる。障壁があるにせよ、できることから始めればよいのである。その小さな一歩を踏み出して未来の検証可能性に託すことが、心理学の、あるいは科学の信用性を高めるための大きな一歩である。

文献

- 1—三浦麻子: 心理学評論, **61**, 3(2018)
- 2—E. P. LeBel et al.: Adv. Meth. Pract. Psychol. Sci., **1**, 389 (2018)
- 3—E. Eich: Psychol. Sci., **25**, 3(2014)
- 4—K. F. Schulz et al.: BMJ, **340**, c332(2010)
- 5—J. P. Simmons et al.: <http://dx.doi.org/10.2139/ssrn.2160588> (2012)
- 6—E. P. LeBel et al.: Perspect. Psychol. Sci., **8**, 424(2013)
- 7—柚取恵太・国里愛彦: 心理学評論, **62**, 231(2019)
- 8—L. L. Shu et al.: Proc. Natl. Acad. Sci. U.S.A., **109**, 15197 (2012)
- 9—A. S. Kristal et al.: Proc. Natl. Acad. Sci. U.S.A., **117**, 7103 (2020)
- 10—U. Simonsohn et al.: <https://datacolada.org/98>(2021)
- 11—平石界: 科学, **92**(9), 806(2022)
- 12—R. Artner et al.: Psychol. Methods, **26**, 527(2021)
- 13—高橋康介: 『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(2018)
- 14—L. M. Fernández & M. A. Vadillo: R. Soc. Open Sci., **7**, 190831(2020)
- 15—R. Silberzahn et al.: Adv. Meth. Pract. Psychol. Sci., **1**, 337 (2018)
- 16—山田祐樹: 科学, **92**(9), 796(2022)
- 17—S. Steegen et al.: Perspect. Psychol. Sci., **11**, 702(2016)
- 18—H. Whitehouse et al.: Nature, **568**, 226(2019)
- 19—B. Beheim et al.: Nature, **595**, E29(2021)
- 20—三浦麻子: 科学, **92**(9), 791(2022)
- 21—T. E. Hardwicke et al.: Perspect. Psychol. Sci., **17**, 239 (2022)
- 22—佐久間淳: 『データ解析におけるプライバシー保護』. 講談社(2016)

武藤拓之 むとう ひろゆき

数理・統計モデリングや実験法などを含む心理学の科学的方法論とその実践に関心がある。具体的な研究テーマとしては、人の空間的推論や意思決定の背後にある認知過程の解明に力を入れている。著書に『たのしいベイズモデリング』(北大路書房, 分担執筆)などがある。