**The Brazilian Reproducibility Initiative:
a systematic assessment of Brazilian biomedical science**

Olavo B. Amaral, Kleber Neves, Ana P. Wasilewska-Sampaio, Clarissa. F. D. Carneiro

Institute of Medical Biochemistry Leopoldo de Meis, Federal University of Rio de Janeiro, Brazil.

Corresponding author:

Olavo B. Amaral, M.D., PhD.

Instituto de Bioquímica Médica Leopoldo de Meis

Av. Carlos Chagas Filho 373, E-38

Cidade Universitária

Rio de Janeiro, RJ, Brazil

CEP 21941-902

Phone: +55-21-39386762

E-mail: olavo@bioqmed.ufrj.br

Abstract

With concerns over research reproducibility on the rise, systematic replications of published science have become an important tool to estimate the replicability of findings in specific areas. Nevertheless, such initiatives are still uncommon in biomedical science, and have never been performed at a national level. The Brazilian Reproducibility Initiative is a multicenter, systematic effort to assess the reproducibility of the country's biomedical research by replicating between 60 and 100 experiments from Brazilian life sciences articles. The project will focus on a set of common laboratory methods, performing each experiment in multiple institutions across the country, with the reproducibility of published findings analyzed in the light of interlaboratory variability. The results, due in 2021, will allow us not only to estimate the reproducibility of Brazilian biomedical science, but also to investigate if there are aspects of the published literature that can be used to predict it.

Introduction

Over recent years, the debate on research reproducibility has received increasing attention from the biomedical research community, as well as from the public at large. Both in the scientific literature (Collins & Tabak, 2014) and in the lay media (Economist, 2013; Harris, 2017), the narrative that science faces a reproducibility crisis has become prominent (Baker, 2016). Suspicions about low rates of results reproducibility or replicability – the ability to reproduce the results of an experiment or article in a direct replication (Goodman et al., 2016) – in specific areas of research were initially raised by theoretical models (Ioannidis, 2005a), systematic reviews of the existing literature (Ioannidis, 2005b) and alarm calls by the pharmaceutical industry (Begley & Ellis, 2012; Prinz et al., 2011). Although these studies are few and restricted to specific fields of science, their numbers support the idea that published findings cannot be accepted at face value and that systematic assessments of reproducibility are warranted.

Interestingly, the largest initiatives to replicate published findings in specific areas of science have been carried out independently of governmental funding agencies, based on bottom-up collaborations and mostly supported by private funders. The Reproducibility Project: Psychology (RPP), which systematically reproduced 100 articles in psychology (Open Science Collaboration, 2015), was followed by similar initiatives in the fields of experimental economics (Camerer et al., 2016), philosophy (Cova et al., 2018) and social sciences (Camerer et al., 2018), with replication rates ranging between 36 and 78%. Undergoing projects include the Reproducibility Project: Cancer Biology (RP:CB; Errington et al., 2014), and the Prostate Cancer Foundation-Movember Foundation Reproducibility Initiative (Tan et al., 2015) which are replicating high-impact studies in basic cancer research through commercial services in the Science Exchange platform.

Although such projects are a welcome development, they are almost invariably limited to specific research topics or communities. Moreover, apart from the initiatives in cancer biology, most have arisen in fields in which experiments are relatively inexpensive and straightforward to perform – thus, in most areas of biomedical science, this issue remains unstudied. Moreover, as science funding and evaluation are mainly carried out at a national or local level by agencies and institutions, reproducibility data remains separated from the stakeholders that assess research quality. In this sense, one wonders whether systematic evaluation of reproducibility shouldn't be more directly tied to the levels at which research funding is distributed. Nevertheless, local or national assessments of research reproducibility are virtually unheard of.

In a time of pleas to preserve science investment in many countries, we believe that the scientific community should be able to support them with data on reproducibility. For this reason, we have obtained funding from the Serrapilheira Institute, a recently created Brazilian nonprofit institution, in order to systematically assess the replicability of the country's biomedical research. Our aim is to replicate between 60 and 100 experiments within life sciences articles from Brazilian institutions, focusing on common methods and performing each experiment in multiple sites within a network of collaborating laboratories. This will allow us not only to estimate the reproducibility of

Brazilian biomedical science, but also to investigate what aspects of the published literature can be used to predict it.

Brazilian science in a nutshell

Scientific research in Brazil started to take an institutional form in the second half of the 20th century, despite the earlier existence of important organizations such as the Brazilian Academy of Sciences (1916) and the Universities of Brazil (later the Federal University of Rio de Janeiro) (1920) and São Paulo (1934). In 1951, the federal government created the first agency dedicated to funding research and executing national science policy, the National Council for Scientific and Technological Development (CNPq). In the same year, the Coordination for Improvement of Graduate Level Personnel (CAPES) was founded. Both federal agencies remain the major funders of Brazilian science. Still, Brazil would wait another decade before graduate-level education was formalized in 1965 (Schwartzman, 2001).

As the number of researchers increased, CAPES took up on the challenge of creating a national evaluation system for graduate education programs in Brazil (Barata, 2016). In the 1990s, the criteria for evaluation began to include quantitative indicators, such as numbers of articles published. In 1998, significant changes were made with the aim of trying to establish articles in international peer-reviewed journals as the main goal, and individual research areas were left free to design their own criteria for ranking journals. In 2007, amidst the largest-ever expansion in the number of federal universities, the journal ranking system in the life sciences became based on impact factors for the previous year, and remains so to this day (CAPES, 2016).

Today, Brazil has over 200,000 PhDs, with more than 10,000 graduating every year (Centro de Gestão e Estudos Estratégicos, 2016). Although the evaluation system is seen as an achievement, it is subject to much criticism, revolving around the centralizing power of CAPES (Hostins, 2006) and the excessive focus on quantitative metrics (Pinto & Andrade, 1999). Many analysts criticize the country's research as largely composed of "salami science", growing in absolute numbers but

lacking in impact, originality and influence (Righetti, 2013). Interestingly, research reproducibility has been a secondary concern in these criticisms, and awareness of the issue has begun to rise only recently.

With the economic and political crisis afflicting the country since 2014, science funding has suffered a sequence of severe cuts. As the Ministry for Science and Technology was merged with that of Communications, a recent constitutional amendment essentially froze science funding at 2016 levels for 20 years (Angelo, 2016). The federal budget for the Ministry suffered a 44% cut in 2017 and reached levels corresponding to roughly a third of those invested a decade earlier (Floresti, 2017), leading scientific societies to position themselves in defense of research funding (Brazilian Society for Scientific Progress, 2018). Concurrently, CAPES has initiated discussions on how to reform its evaluation system (Academia Brasileira de Ciências, 2018). At this delicate moment, in which a new federal government is about to take office, an empirical assessment of the country's scientific output seems warranted to inform such debates.

The Brazilian Reproducibility Initiative: aims and scope

The Brazilian Reproducibility Initiative was started in early 2018 as a systematic effort to evaluate the reproducibility of Brazilian biomedical science. Openly inspired by multicenter efforts such as RPP (Open Science Collaboration, 2015), RP:CB (Errington et al., 2014) and the Many Labs projects (Ebersole et al., 2016; Klein et al., 2014; Klein et al., 2018), our goal is to replicate between 60 and 100 experiments from published Brazilian articles in the life sciences, focusing on common methods and performing each experiment in multiple sites within a network of collaborating laboratories. The project's coordinating team at the Federal University of Rio de Janeiro is responsible for the selection of methods and experiments, as well as for the recruitment and management of collaborating labs. Experiments are set to begin in mid-2019, in order for the project to achieve its final results by 2021.

Any project with the ambition of estimating the reproducibility of a country's science is inevitably limited in scope by the expertise of the participating teams. However, we will aim for the most representative sample that can be achieved without compromising feasibility, through the use of the strategies described below. Nevertheless, representativeness will be limited by the selected techniques and biological models, as well as by our inclusion and exclusion criteria – which include the cost and commercial availability of materials and the expertise of the replicating labs.

Focus on individual experiments

Our first choice was to base our sample on experiments rather than articles. As studies in basic biomedical science usually involve many experiments with different methods revolving around a hypothesis, trying to reproduce a whole study, or even its main findings, can be cumbersome for a large-scale initiative. Partly because of this, the RP:CB, which originally planned to reproduce the main findings of 50 studies, has been downsized to 18 studies (Kaiser, 2018). Moreover, some of the replications in the RP:CB reproduced some of the original experiments but not others; thus, the published Replication Studies in the RP:CB do not say if a particular study has been successful or not, and the project uses multiple measures to assess reproducibility.

Contrary to studies, experiments have well defined effect sizes, and although different criteria can be used for what constitutes a successful replication (Goodman et al., 2016; Open Science Collaboration, 2015), they can be defined objectively, allowing a quantitative assessment of reproducibility. Naturally, there is a downside in that replication of a single experiment is usually not enough to confirm or refute the conclusions of an article (Camerer et al., 2018). However, if one's focus is not on the studies themselves, but rather on evaluating reproducibility on a larger scale, we believe that experiments represent a more manageable unit than studies.

Selection of methods

No replication initiative, no matter how large, can aim to reproduce every kind of experiment. Thus, our next choice was to limit our scope to common methodologies that are widely

available in the country, in order to ensure that we will have a large enough network of potential collaborators.

To provide a list of candidate methods, we started by performing an initial review of a sample of articles in Web of Science life sciences journals published in 2017, filtering for papers which (a) had all authors affiliated with a Brazilian institution, (b) presented experimental results on a biological model and (c) did not use clinical or ecological samples. One hundred randomly selected articles had data extracted concerning the models, experimental interventions and methods used to analyze outcomes, with the main results shown in Figure 1A and B – for more details on this step of the project, see <https://osf.io/f2a6y/>.

Based on this initial review, we restricted our scope to experiments using rodents and cell lines, which were by far the most prevalent models (present in 77% and 16% of articles, respectively). After a first round of automated full-text assessment of 5000 Brazilian articles between 1998 and 2017, we selected 10 commonly used techniques (Figure 1C) as candidates for replication experiments. An open call for collaborating labs within the country was then set up, and labs were allowed to register through an online form for performing experiments with one or more of these techniques and models during a 3-month period. After this period, we used this input (as well as other criteria such as cost analysis) to select 5 methods for the replication effort: MTT assay, reverse transcriptase polymerase chain reaction (RT-PCR), elevated plus maze, western blot and immunohisto/cytochemistry. We are starting the project with the first three methods, while inclusion of the latter two will be confirmed after a more detailed cost analysis based on the fully developed protocols.

Figure 1. (A) Most frequent biological models used in main experiments within a sample of 100 Brazilian life sciences articles. (B) Most frequent methods used for quantitative outcome detection in these experiments. ‘Cell count’, ‘enzyme activity’ and ‘blood tests’ include various experiments for which methodologies vary and/or are not described fully in articles. Nociception tests, although frequent, were not considered for replication due to animal welfare considerations. (C) Flowchart describing the first full-text screening round to identify articles in our candidate techniques, which led us to select our final set of 5 methods.

We are currently selecting articles using these techniques by full-text screening of a random sample of life sciences articles from the past 20 years in which most of the authors, including the corresponding one, are based in a Brazilian institution. From each of these articles, we select the first experiment using the technique of interest, defined as a quantitative comparison of a single outcome between two experimental groups. Although the final outcome of the experiment should be assessed using the method of interest, other laboratory techniques are likely to be involved in the model and experimental procedures that precede this step.

We will restrict our sample to experiments that (a) represent one of the main findings of the article, defined by mention of its results in the abstract, (b) present significant differences between groups, in order to allow us to perform sample size calculations, (c) use commercially available materials, (d) have all experimental procedures falling within the expertise of at least 3 laboratories in our network and (e) have an estimated cost below 0.5% of the project's total budget. For each included technique, 20 experiments will be selected, with the biological model and other features of the experiment left open to variation in order to maximize representativeness. A more detailed protocol for this step is available at <https://osf.io/57f8s/>.

After experiments are selected, we will record each study's methods description in standardized description forms, which will be used to define replication protocols. These experiments will then be assigned to 3 laboratories each by the coordinating team, which will confirm that they have the necessary expertise in order to perform it.

Multicenter replication

A central tenet of our project is that replication should be performed in multiple laboratories. As discussed in other replication projects (Errington et al., 2014; Gilbert et al., 2016; Open Science Collaboration, 2015) a single failed replication is not enough to refute the original finding, as there are many reasons that can explain discrepancies between results (Goodman et al., 2016). While some of them – such as misconduct or bias in performing or analyzing the original experiment – are problematic, others – such as unrecognized methodological differences or chance – are not necessarily as alarming. Reproducibility estimates based on single replications cannot distinguish between these causes, and can thus be misleading in terms of their diagnoses (Jamieson, 2018).

This problem is made worse by the fact that data on interlaboratory variability for most methods is scarce: even though simulations demonstrate that multicenter replications are an efficient way to improve reproducibility (Voelkl et al., 2018), they are exceedingly rare in most fields of basic biomedical science. Isolated attempts at investigating this issue in specific fields have shown that, even when different labs try to follow the same protocol, unrecognized methodological

variables can still lead to a large amount of variation (Crabbe et al., 1999; Hines et al., 2014; Massonnet et al., 2010). Thus, it might be unrealistic to expect that reproducing a published experiment – for which protocol details will probably be lacking (Hair et al., 2018; Kilkenny et al., 2009) – will yield similar results in a different laboratory.

In our view, the best way to differentiate irreproducibility due to bias or error from that induced by methodological variables alone is to perform replications at multiple sites. In this way, an estimate of interlaboratory variation can be obtained for every experiment, allowing one to analyze whether the original result falls within the expected variation range. Multicenter approaches have been used successfully in the area of psychology (Ebersole et al., 2016; Klein et al., 2014; Klein et al., 2018), showing that some results are robust across populations, while others do not reproduce well in any of the replication sites.

Our plan for the Brazilian Reproducibility Initiative is to perform each individual replication in at least 3 different laboratories; this, however, opens up questions about how much standardization is desirable. Although one should follow the original protocol in a direct replication, there are myriad steps that will not be well described. And while some might seem like glaring omissions, such as the absence of species, sex and age information in animal studies (Kilkenny et al., 2009), others might simply be overlooked variables – how often does one describe the exact duration and intensity of sample agitation (Hines et al., 2014), for example? When conditions are not specified, one is left with two choices. One of them is to standardize steps as much as possible, building a single, detailed replication protocol for all labs. However, this will reduce interlaboratory variation to an artificially low level, making the original experiment likely to fall outside the effect range observed in the replications.

To avoid this, we will take a more naturalistic approach. Although details included in the original article will be followed explicitly in order for the replication to be as direct as possible, steps which are not described will be left open for each replication team to fill based on their best judgment. Replication teams will be required to record those choices in detailed methods

description forms, but it is possible – and desirable – for them to vary according to each laboratory’s experience. Methodological discrepancies in this case should approach those observed between research groups working independently, providing a realistic estimate of interlaboratory variation for the assessment of published findings. This approach will also allow us to explore the impact of methodological variation on the experimental results – a topic perhaps as important as reproducibility itself – as a secondary outcome of the project.

A central issue in other replication projects has been engagement with the original authors in order to revise protocols (Camerer et al., 2018; Errington et al., 2014; Open Science Collaboration, 2015). While we feel this is a worthy endeavor, the rate of response to calls for sharing protocols, data or code is erratic (Hardwicke & Ioannidis, 2018; Stodden et al., 2018; Wicherts et al., 2011). Moreover, having access to unreported information is likely to overestimate the reproducibility of a finding based on published information, leading results to deviate from a ‘naturalistic’ estimate of reproducibility (Coyne, 2016). Thus, although we will contact the original authors for protocol details when these are available, in order to assess methodological variation between published studies and replications, this information will not be made available to the replication teams. They will receive only the protocol description from the published article, with no mention of its results or origin, in order to minimize bias. While we cannot be sure that this form of blinding will be effective, as experiments could be recognizable by scientists working in the same field, replicating labs will be stimulated not to seek this information.

Lastly, although non-described protocol steps will be left open to variation, methodological issues that are consensually recognized to reduce error and bias will be enforced. Thus, bias control measures such as blinding of researchers to experimental groups will be used whenever possible, and sample sizes will be calculated to provide each experiment with a power of 95% to detect the original difference – as in other surveys, we are setting our power estimates at a greater than usual rate due to the recognition that the original results are likely to be inflated by publication bias (Camerer et al., 2018; Cova et al., 2018, Klein et al., 2018). Moreover, if additional positive and/or

negative controls are judged to be necessary to interpret outcomes, they will also be added to the experiment.

To ensure that these steps are followed – as well as to adjudicate on any necessary protocol adaptations, such as substitutions in equipment or materials – each individual protocol will be reviewed after completion in a round-robin approach (Silberzahn et al., 2018) by (a) the project's coordinating team and (b) an independent laboratory working with the same technique that is not directly involved in the replication. Each of the three protocol versions of every experiment will be sent to a different reviewing lab, in order to minimize the risk of overstandardization. Suggestions and criticisms to the protocol will be sent back to the replicating team, and experiments will only start after both labs and the coordinating team reach consensus that the protocol (a) does not deviate excessively from the published one and can be considered a direct replication and (b) includes all necessary bias control measures and controls to ensure the validity of the results.

Evaluating replications

As previous projects have shown, there are many ways to define a successful replication, all of which have caveats (Camerer et al., 2018, 2016; Goodman et al., 2016; Open Science Collaboration, 2015). Reproducibility of the general conclusions on the existence of an effect (e.g. two results finding a statistically significant difference in the same direction) might not be accompanied by reproducibility of the effect size; conversely, studies with effect sizes that are similar to each other might have different outcomes in significance tests (Simonsohn, 2015). Moreover, if non-replication occurs, it is hard to judge whether the original study or the replication is closer to the true result. Although one can argue that, if replications are conducted in an unbiased manner and have higher statistical power, they are more likely to be accurate, the possibility of undetected methodological differences preclude one from attributing non-replication to failures in the original studies (Errington et al., 2014).

Multisite replication is a useful way to circumvent some of these controversies, as if the variation between unbiased replications in different labs is known, it is possible to determine whether the original result is within this variability range. Thus, the primary outcome of our analysis will be the percentage of original studies with effect sizes falling within the 95% prediction interval of a meta-analysis of the 3 replications. Nevertheless, we acknowledge that this definition also has caveats: if interlaboratory variability is high, prediction intervals can be wide, leading a large amount of results to be considered “reproducible”. Thus, replication estimates obtained by these methods are likely to be optimistic. On the other hand, failed replications will be more likely to reflect true biases, errors or deficiencies in the original experiments (Patil et al., 2016).

An additional problem is that, given our naturalistic approach to reproducibility, incomplete reporting in the original study might increase interlaboratory variation and artificially improve our primary outcome. With this in mind, we will include other ways to define reproducibility as secondary outcomes, such as the statistical significance of the pooled replication studies, the significance of the effect in a meta-analysis including the original result and replication attempts, and a statistical comparison between the pooled effect sizes of the replications and the original result. We will also examine thoroughness of methodological reporting as an independent outcome, in order to evaluate the possibility of bias caused by incomplete reporting.

Moreover, we will explore correlations between results and differences in particular steps of each technique; nevertheless, we cannot know in advance whether methodological variability will be sufficient to draw conclusions on these issues. As each experiment will be performed in only 3 labs, while there are myriad steps to each technique, it is unlikely that we will be able to pinpoint specific sources of variation between results of individual experiments. Nevertheless, by quantifying the variation across protocols for the whole experiment, as well as for large sections of it (model, experimental intervention, outcome detection), we can try to observe whether the degree of variation correlates with variability in results. Such analyses, however, will only be planned once protocols are completed, so as to have a better idea of the range of variability across them.

Finally, we will try to identify factors in the original studies that can predict reproducibility, as such proxies could be highly useful to guide the evaluation of published science. These will include features shown to predict reproducibility in previous work, such as effect sizes, significance levels and subjective assessment by prediction markets (Dreber et al., 2015; Camerer et al., 2016, 2018; Open Science Collaboration, 2015); the pool of researchers used for the latter, however, will be different from those performing replications, so as not to compromise blinding with respect to study source and results. Other factors to be investigated include (a) the presence of bias control measures in the original study, such as blinding and sample size calculations, (b) the number of citations and impact factor of the journal, (c) the experience of the study's principal investigator, (d) the Brazilian region of origin, (e) the technique used, (f) the type of biological model and (g) the area of research. As our sample of experiments will be obtained randomly, we cannot ensure that there will be enough variability in all factors to explore them meaningfully. Nevertheless, we should be able to analyze some variables that have not been well explored in previous replication attempts, such as 'impact' defined by citations and publication venues, as most previous studies have focused on particular subsets of journals (Camerer et al., 2018; Open Science Collaboration, 2015) or impact tiers (Errington et al., 2014; Ioannidis, 2005b).

A question that cannot be answered directly by our study design is whether any correlations found in our sample of articles can be extrapolated to different methods in Brazilian biomedical science, as well as to other regions of the world. For some factors, including the reproducibility estimates themselves and their correlation with local variables, extrapolations to the international scenario are clearly not warranted. On the other hand, relationships between reproducibility and methodological variables, as well as with article features, can plausibly apply to other countries, although this can only be known for sure by performing studies in other regions.

All of our analyses will be preregistered at the Open Science Framework in advance of data collection. All our datasets will be made public and updated progressively as replications are performed – a process planned to go on until 2021. As an additional measure to promote

transparency and engage the Brazilian scientific community in the project, we are posting our methods description forms for public consultation and review (see <http://reprodutibilidade.bio.br/public-consultation>), and will do so for the analysis plan as well.

Potential Challenges

A multicenter project involving the replication of experiments in multiple laboratories across a country of continental proportions is bound to meet challenges. The first of them is that the project is fully dependent on the interest of Brazilian laboratories to participate. Nevertheless, the response to our first call for collaborators exceeded our expectations, reaching a total of 71 laboratories in 43 institutions across 19 Brazilian states.. The project received coverage by the Brazilian media (Ciscati, 2018; Neves & Amaral, 2018; Pesquisa FAPESP, 2018) and achieved good visibility in social networks, contributing to this widespread response. While we cannot be sure that all laboratories will remain in the project until its conclusion, it seems very likely that we will have the means to perform our full set of replications, particularly as laboratories will be funded for their participation.

Concerns also arise from the perception that replicating other scientists' work indicates mistrust of the original results, a problem that is potentiated by the conflation of the reproducibility debate with that on research misconduct (Jamieson, 2018). Thus, from the start, we are taking steps to ensure that the project is viewed as we conceive it: a first-person initiative of the Brazilian scientific community to evaluate its own practices. We will also be impersonal in our choice of results to replicate, working with random samples and performing our analysis at the level of experiments; thus, even if a finding is not deemed reproducible, this will not necessarily invalidate an article's conclusions or call a researcher into question.

An additional challenge is to ensure that participating labs have sufficient expertise with a methodology or model to provide accurate results. Ensuring that the original protocol is indeed being followed is likely to require steps such as cell line/animal strain authentication and positive

controls for experimental validation. Nevertheless, we prefer this naturalistic approach to the alternative of providing each laboratory with animals or samples from a single source, which would inevitably underestimate variability. Moreover, while making sure that a lab is capable of performing a given experiment adequately is a challenge we cannot address perfectly, this is a problem of science as a whole – and if our project can build expertise on how to perform minimal certification of academic laboratories, this could be useful for other purposes as well.

A final challenge will be to put the results into perspective once they are obtained. Based on the results of previous reproducibility projects, a degree of irreproducibility is expected (Camerer et al., 2016, 2018; Cova et al., 2018; Open Science Collaboration, 2015) and may raise concerns about Brazilian science, as there will be no estimates from other countries for comparison. Nevertheless, our view is that, no matter the results, they are bound to put Brazil at the vanguard of the reproducibility debate, if only because we will likely be the first country to produce such an estimate.

Conclusions

With the rise in awareness over reproducibility issues, systematic replication initiatives have begun to develop in various research fields (Camerer et al., 2016, 2018; Cova et al., 2018; Errington et al., 2014; Open Science Collaboration, 2015; Tan et al., 2015). Our study offers a different perspective on the concept, covering different research areas in the life sciences with focus in a particular country.

This kind of initiative inevitably causes controversy both on the validity of the effort (Coyne, 2016; Nature Medicine, 2016) and on the interpretation of the results (Baker & Dolgin, 2017; Gilbert et al., 2016; Patil et al., 2016). Nevertheless, multicenter replication efforts are as much about the process as about the data. Thus, if we attain enough visibility within the Brazilian scientific community, a large part of our mission – fostering the debate on reproducibility and how to evaluate it – will have been achieved. Moreover, it is healthy for scientists to be reminded that

self-correction and confirmation are a part of science, and that published findings are passive of independent replication. There is still much work to be done in order for replication results to be incorporated into research assessment (Ioannidis, 2014; Munafò et al., 2017), but this kind of reminder by itself might conceivably be enough to initiate cultural and behavioral change.

Finally, for those involved as collaborators, one of the main returns will be the experience of tackling a large scientific question collectively in a transparent and rigorous way. We believe that large-scale efforts can help to lead an overly competitive culture back to the Mertonian ideal of communality, and hope to engage not only collaborators but the Brazilian scientific community at large through data sharing, public consultations and social media (centered in our website at <http://reprodutibilidade.bio.br/home>). The life sciences community in Brazil is large enough to need this kind of challenge, but perhaps still small enough to answer cohesively. We thus hope that the Brazilian Reproducibility Initiative, through its process as much as through its results, can have a positive impact on the scientific culture of our country for years to come.

Acknowledgements

The Brazilian Reproducibility Initiative is funded by a grant from the Serrapilheira Institute. K.N. and A.P.W.S. are funded by this grant. C.F.D.C. is funded by a scholarship from CNPq.

The authors are indebted to Peter Rodgers, Timothy Errington and Richard Klein, whose constructive criticism from a round of review at *eLife* have been incorporated in this version of the manuscript.

References

- Academia Brasileira de Ciências. (2018). Considerações sobre o processo de avaliação da pós-graduação da Capes. Rio de Janeiro, RJ, Brazil. Retrieved from: http://www.abc.org.br/IMG/pdf/documento_pg_da_abc_22032018_fim.pdf
- Angelo, C. (2016). Brazil's scientists battle to escape 20-year funding freeze. *Nature* **539**:480. doi:

10.1038/nature.2016.21014

Baker, M. (2016). Is there a reproducibility crisis? *Nature* **533**:452–454. doi: 10.1038/533452a

Baker, M., & Dolgin, E. (2017). Reproducibility project yields muddy results. *Nature* **541**:269–270. doi: 10.1038/541269a

Barata, R. C. B. (2016). Dez coisas que você deveria saber sobre o Qualis. *Revista Brasileira de Pós-Graduação* **13**:13-40. doi: 10.21713/2358-2332.2016.v13.947

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature* **483**: 531–3. doi: 10.1038/483531a

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M. & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science* **351**:1433–6. doi: 10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E.-J. & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour* **2**:637-444. doi: 10.1038/s41562-018-0399-z

CAPES (2016). Considerações sobre Qualis Periódicos. Retrieved from: http://capes.gov.br/images/documentos/Qualis_periodicos_2016/Consider%C3%A7%C3%B5es_qualis_Biol%C3%B3gicas_II.pdf

Centro de Gestão e Estudos Estratégicos (2016). *Mestres e Doutores 2015 - Estudos da demografia da base técnico-científica brasileira*. CGEE, Brasília, DF, Brazil. ISBN: 978-85-5569-114-0.

Ciscati, R. (2018). Projeto vai replicar experimentos de cientistas brasileiros para checar sua eficiência. *Jornal O Globo*. Published on April 22nd 2018. Retrieved from: <https://oglobo.globo.com/sociedade/ciencia/projeto-vai-replicar-experimentos-de-cientistas->

- Collins, F. S., & Tabak, L. A. (2014). NIH plans to enhance reproducibility. *Nature* **505**:612–613.
doi: 10.1038/505612a
- Cova, F., Strickland, B., Abatista, A., Allard, A., Andow, J., Attie, M., Beebe, J., Berniūnas, R., Boudesseul, J., Colombo, M., Cushman, F., Diaz, R., van Dongen, N.D.N., Dranseika, V., Earp, B.D., Torres, A.G., Hannikainen, I., Hernández-Conde, J.V., Hu, W., Jaquet, F. Khalifa, K., Kim, H., Kneer, M., Knobe, J., Kurthy, M., Lantian, A., Liao, S., Machery, E., Moerenhout, T., Mott, C., Phelan, M., Phillips, J., Rambharose, N., Reuter, K., Romero, F., Sousa, P., Sprenger, J., Thalabard, E., Tobia, K., Viciana, H., Wilkenfeld, D. & Zhou,X. (2018). Estimating the Reproducibility of Experimental Philosophy. *Review of Philosophy and Psychology*. doi: 10.1007/s13164-018-0400-9
- Coyne, J. C. (2016). Replication initiatives will not salvage the trustworthiness of psychology. *BMC Psychology* **4**:28. doi: 10.1186/s40359-016-0134-3
- Crabbe, J. C., Wahlsten, D., & Dudek, B. C. (1999). Genetics of Mouse Behavior: Interactions with Laboratory Environment. *Science* **284**:1670-2. doi: 10.1126/science.284.5420.1670
- Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B.A., & Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences of the United States of America* **50**: 15343-15347. doi: 10.1073/pnas.1516179112
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., *et al.* (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology* **67**:68–82. doi: 10.1016/J.JESP.2015.10.012
- Economist (2013). Trouble at the lab. *The Economist*. Published on October 18th 2013. Retrieved from: <https://www.economist.com/briefing/2013/10/18/trouble-at-the-lab>
- Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J. & Nosek, B. A. (2014). An open

investigation of the reproducibility of cancer biology research. *eLife* **3**:e04333. doi: 10.7554/eLife.04333

Floresti, F. (2017). A ciência brasileira vai quebrar? *Revista Galileu*. Published on September 29th 2017. Retrieved from: <https://revistagalileu.globo.com/Revista/noticia/2017/09/ciencia-brasileira-vai-quebrar.html>

Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science* **351**:1037. doi: 10.1126/science.aad7243

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine* **8**:341ps12. doi: 10.1126/scitranslmed.aaf5027

Hair, K., Macleod, M. R., Sena, E. S., & The IICARus Collaboration. (2018). A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *bioRxiv*. doi: 10.1101/370874

Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLoS ONE* **13**:e0201856. doi: 10.1371/journal.pone.0201856

Harris, R. (2017). Rigor Mortis. *Basic Books*, New York, N.Y., U.S.A. ISBN: 978-0465097906

Hines, W. C., Su, Y., Kuhn, I., Polyak, K., & Bissell, M. J. (2014). Sorting out the FACS: a devil in the details. *Cell Reports* **6**:779–81. doi: 10.1016/j.celrep.2014.02.021

Hostins, R. C. L. (2006). Os Planos Nacionais de Pós-graduação (PNPG) e suas repercussões na pós-graduação brasileira. *Perspectiva* **24**:133–160. URL: <https://periodicos.ufsc.br/index.php/perspectiva/article/view/10315/9578>

Ioannidis, J. P. A. (2005a). Why most published research findings are false. *PLoS Medicine* **2**:e124. doi: 10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2005b). Contradicted and initially stronger effects in highly cited clinical research. *JAMA* **294**:218-228. doi: 10.1001/jama.294.2.218

Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Medicine* **11**:e1001747.

doi: 10.1371/journal.pmed.1001747

- Jamieson, K. H. (2018). Crisis or self-correction: Rethinking media narratives about the well-being of science. *Proceedings of the National Academy of Sciences of the United States of America* **115**:2620–7. doi: 10.1073/pnas.1708276114
- Kaiser, J. (2018). Plan to replicate 50 high-impact cancer papers shrinks to just 18. *Science*. doi: 10.1126/science.aau9619
- Kilkenny, C., Parsons, N., Kadyszewski, E., Festing, M. F. W., Cuthill, I. C., Fry, D., Hutton, J. & Altman, D. G. (2009). Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS ONE* **4**:e7824. doi: 10.1371/journal.pone.0007824
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., *et al.* (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology* **45**:142–152. doi: 10.1027/1864-9335/a000178
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B., Adams Jr., R.B., Alper, S., *et al.* (2018). Many Labs 2 : Investigating variation in replicability across sample and setting. *PsyArXiv*. doi:10.31234/osf.io/9654g
- Massonnet, C., Vile, D., Fabre, J., Hannah, M. A., Caldana, C., Lisec, J., Beemster, G.T.S., Meyer, R.C., Messerli, G., Gronlund, J.T., Perkovic, J., Wigmore, E., May, S., Bevan, M.W., Meyer, C., Rubio-Díaz, S., Weigel, D., Micol, J.L., Buchanan-Wollaston, V., Fiorani, F., Walsh, S., Rinn, B., Gruissem, W., Hilson, P., Hennig, L., Willmitzer, L. & Granier, C. (2010). Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three *Arabidopsis* accessions cultivated in ten laboratories. *Plant Physiology* **152**:2142–57. doi: 10.1104/pp.109.148338
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E-J, Ware, J.J. & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour* **1**:0021. doi: 10.1038/s41562-016-0021
- Nature Medicine (2016). Take the long view. *Nature Medicine* **22**:1. doi: 10.1038/nm.4033

- Neves, K., & Amaral, O. B. (2018). Abrindo a caixa-preta. *Ciência Hoje*. Published on May 24th 2018. Retrieved from: <http://cienciahoje.org.br/artigo/abrindo-a-caixa-preta>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science* **349**:aac4716. doi: 10.1126/science.aac4716
- Patil, P., Peng, R. D., & Leek, J. T. (2016). What Should Researchers Expect When They Replicate Studies? A Statistical View of Replicability in Psychological Science. *Perspectives on Psychological Science* **11**:539–544. doi: 10.1177/1745691616646366
- Pesquisa FAPESP (2018). Uma rede para reproduzir experimentos. *Revista Pesquisa Fapesp*. Published on May 2018. Retrieved from: <http://revistapesquisa.fapesp.br/2018/05/17/uma-rede-para-reproduzir-experimentos>
- Pinto, A. C., & Andrade, J. B. (1999). Fator de impacto de revistas científicas: qual o significado deste parâmetro? *Química Nova* **22**:448–53. doi: 10.1590/S0100-40421999000300026
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery* **10**:712. doi: 10.1038/nrd3439-c1
- Righetti, S. (2013). Brasil cresce em produção científica, mas índice de qualidade cai. *Folha de S. Paulo*. Published on April 22nd 2013. Retrieved from: <https://www1.folha.uol.com.br/ciencia/2013/04/1266521-brasil-cresce-em-producao-cientifica-mas-indice-de-qualidade-cai.shtml>
- Sociedade Brasileira para o Progresso da Ciência. (2018). Carta Aberta ao presidente da República em defesa da Capes recebe mais de 50 assinaturas e é destaque na imprensa nacional. *Portal SBPC*. Published on August 6th 2018. Retrieved from: <http://portal.sbpnet.org.br/noticias/carta-aberta-ao-presidente-da-republica-em-defesa-da-capes-recebe-mais-de-50-assinaturas-e-e-destaque-na-imprensa-nacional>
- Schwartzman, S. (2001). Um espaço para ciência: a formação da comunidade científica no Brasil. Ministério de Ciência e Tecnologia. Brasília, DF, Brazil. Retrieved from:

<http://livroaberto.ibict.br/handle/1/757>

Silberzahn, R., Uhlmann, E. L., Martin, D., Anselmi, P., Aust, F., Awtrey, E. C., *et al.* (2018).

Many analysts, one dataset: Making transparent how variations in analytical choices affect results. *Advances in Methods and Practices in Psychological Science* **1**:337-356. doi: 10.1177/2515245917747646

Simonsohn, U (2015). Small telescopes: detectability and the evaluation of replication results.

Psychological Science **26**:559–569. doi: 10.1177/0956797614567341

Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for

computational reproducibility. *Proceedings of the National Academy of Sciences of the United States of America* **115**:2584–9. doi: 10.1073/pnas.1708290115

Tan, E. F., Perfito, N., & Lomax, J. (2015). Prostate Cancer Foundation-Movember Foundation

Reproducibility Initiative. *Open Science Framework*. URL: <https://osf.io/ih9qt/>

Voelkl, B., Vogt, L., Sena, E.S., Würbel, H (2018). Reproducibility of preclinical animal research

improves with heterogeneity of study samples. *PLoS Biol* **16**:e2003693. doi: 10.1371/journal.pbio.2003693

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to

the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE* **6**:e26828. doi: 10.1371/journal.pone.0026828