

original article

Full disclosure: doing behavioural science necessitates sharing

Gjalt-Jorn Y. Peters*Open University of the
Netherlands***Charles Abraham***University of Exeter***Rik Crutzen***Maastricht University*

The way we currently organise and report research retards behavioural science. This brief article provides an explanation of how we are holding back scientific progress, how this situation developed, and how current practice is justified. We also recommend practical, low-cost solutions, which would

facilitate scientific advances.

Reasons for full disclosure: how we hold back scientific progress

When we complete a study, we tend to publish only the results. Researchers rarely publish the questionnaires, computer tasks, intervention manuals, study protocols and other materials that were used to generate these results. Similarly, readers do not usually have access to complete data sets, all statistical analyses undertaken or the commands (e.g. 'syntax' in SPSS) needed to replicate these analyses. In addition, researchers do not always secure these supplemental materials sufficiently, and they frequently get lost over time. This has three consequences each of which retard the establishment of a science of behaviour.

1. It is not possible to critically scrutinize what researchers do not divulge in an article so the empirical data that reported results represent remain unknown and it is difficult to judge whether analyses and interpretation are optimal or even correct;

2. Accurate replication of research is compromised and often impossible;

3. Data syntheses (e.g. meta-analyses) are impeded, as they frequently have to exclude studies because they do not have access to the original data and procedures, and so may be based on incorrect assumptions.

A brief explanation of why each of these consequences is undesirable follows.

Maximum scrutiny

There are two reasons to desire maximum scrutiny of our research. First, we can make mistakes in analyses and interpretation (Bakker & Wicherts, 2011; Glantz, 1980; Wicherts, Bakker, & Molenaar, 2011) for example, when using an analytical method for the first time. This is especially likely when analyses are undertaken by those still learning analytic techniques (including both qualitative and statistical analyses). Since reviewers might also be limited in their competence, researchers' errors frequently end up in published papers (Glantz, 1980; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). Allowing further scrutiny beyond the pre-publication peer review process can teach us what we are doing wrong, thereby making us better researchers and reviewers (see also Walther & van den Bosch, 2012). Second, as pointed out by Simmons, Nelson, and Simonsohn (2011), researchers make a lot of choices that are not disclosed in research reports. These choices frequently favour significant p-values (Masicampo & Lalande, 2012). This is likely a consequence of the Significant Outcome Bias in

our literature (Peters, Ruiter, & Kok, 2012): non-significant findings are rarely submitted, and if submitted, rarely published. Greater disclosure of materials and analyses would allow detection of these biases.

Accurate replication

As was recently argued (Abraham, 2012a), scientific progress requires replications. Successful replications strengthen our evidence that a given theory or model holds, whereas unsuccessful replications can falsify previously held theories or models (see for example Milton & Wiseman, 1999 or Ritchie, Wiseman, & French, 2012). Without the ability to replicate others' work, we are not accumulating knowledge, but rather false positives: once an article makes a claim, we thwart the possibility to replicate the study and thereby challenge the claim, especially if required materials are unavailable. Publication bias plays a role here (some journals have an official policy to not publish replications; French, 2012), and this is a problem that needs to be addressed in its own right. Nonetheless, at least publishing all relevant materials would permit *accurate* replications. At present, debates regularly arise as to whether or not failures to replicate should be attributed to a mistaken interpretation of empirical reality or differences in materials or study procedures; witness, for example recent replication debates in social psychology (Doyen, Klein, Pichon, & Cleeremans, 2012; Ritchie et al., 2012). This inability to replicate and acknowledge replication failures renders behavioural scientists unable to use data to choose between better and worse models of reality and better and worse approaches to intervening in the real world (van Dongen, Abraham, Ruiter, & Veldhuizen, in press).

Data syntheses

When trying to synthesize the findings of multiple studies, either quantitatively or

qualitatively, it is crucial to understand the initial data, otherwise syntheses may result in counting apples and oranges as bananas to the detriment of behavioural science. For example, in the case of meta-analysis it is necessary to convert study results into effect size measures that use the same metric. For simple designs where means in two groups are compared, Cohen's d is the obvious choice; when relating two continuous variables, Pearson's r is often used. Published studies frequently do not report effect sizes; and it is quite common that studies have to be excluded from meta-analyses because it is not possible to compute the required statistics on the basis of the reported results. Even when willing (which not all researchers are, see Wicherts et al., 2011), authors are often unable to send meta-analysts their datasets. Publishing dataset files, as well as the commands for the analyses that were used (e.g. syntax files in SPSS, scripts in R), along with the articles would considerably enhance the quality of evidence syntheses.

Given these clear advantages, one may wonder why we have resisted full disclosure. Surprisingly, there are very good reasons why this convention, to only publish results and not data, analyses and materials, developed.

How non-disclosure developed and current views

On 6 March 1665, the first purely scientific journal was published (Oldenburg, 1665), to enable (more or less) efficient communication between researchers. Since then, many such journals have been published. Journals initially faced serious limitations: production and distribution of physical journals was expensive, so contributions had to be brief and appendices were rare. Because of this, the address of at least one of the authors was always included, to enable researchers to request supplementary materials such as questionnaires that were used,

more detailed study protocols, and more recently, computer programs and algorithms.

Since establishment of the internet, most journals are published online (some exclusively), archiving articles and supplementary materials on servers. These materials can be accessed from anywhere, and the costs of storing material on a server are negligible¹. Online-only journals are therefore relieved of the constraint of limiting the number of pages in an article, and all journals are relieved of the constraint of limiting the supplements. Therefore, it has been possible for a number of years to publish not only the report describing your methods and results, but also everything used in your study. However, despite the disappearance of practical and technological constraints, reporting practices in behavioural science remain largely unchanged. This could be remedied by the editors of leading journals; publishing only articles representing full online disclosure would change practice rapidly.

Because these constraints have been lifted, currently, there have been a number of pleas for full(er) disclosure. On the one hand, there were pleas for Open Access publication of articles reporting research funded with public money (Ghosh, 2012), but there have also been efforts to promote full disclosure of data and materials. For example, recently a consensus statement was developed by a group of health psychology researchers and journal editors urging editors to adopt a full disclosure policy in relation to behaviour change intervention development and design. The Workgroup for Intervention Development and Evaluation Research (WIDER) made four recommendations which can be summarized as: (1) provide detailed intervention descriptions, including (2) descriptions of control groups including usual care, (3) describe the intervention development process in detail and in relation to postulated change processes

and (4) provide intervention delivery manuals that enable accurate replication (see e.g., Abraham, 2012b). Yet, despite the minimal costs, too few researchers routinely publish all their materials, data, analyses scripts and output, listing a variety of reasons.

Reasons for non-disclosure

In informal discussions, people have listed a variety of reasons for their reluctance to publish everything. The main ones are discussed below.

I may want to use my data again

The APA (American Psychological Association) requires that "psychologists do not withhold the data on which their conclusions are based" (APA, 2010, p. 12). However, when a researcher wants to publish several articles about one dataset it can be risky to publish the dataset before the work is completed. After all, somebody else might beat you to it—quickly publishing ideas arising from your own data. Luckily, there is a straightforward solution that negates the concern: do not publish the entire

¹For example, a convenience sample of the Portable Document Format (PDF) files of 20 publications from 2012 shows that the average article has 21.50 pages (median = 13.50, sd = 26.86), is 594.80 kilobytes (median = 366.50, sd = 576.80), and that the average number of kilobytes for one page is 41.62 (median = 27.07, sd = 41.60). When saved as plain textfiles, a datafile with around 20 variables and 500 participants will be around 860 kilobytes if the variables are saved with 15 significant digits. Since most articles report far less data, it is safe to say that on average, one article plus supplemental materials can be easily stored in around 5 megabytes. As an example of how additional materials can be published, the datafile, R commands and output for these analyses are available online at <http://sciencerep.org/1>. Hosting prices vary, but plans exist where 1 gigabyte costs around €10 per year. This means that it is possible to host 1000 articles for around 50 euro per year.

dataset, but only those variables you described in your article. All statistical programs allow you to specify which variables should be 'kept' or 'dropped' when saving. This approach has two risks. First, it enables cherry picking of the data: researchers can omit variables that behave inconsistently with their hypotheses. This risk is addressed by the requirement to also publish the used study protocol, materials and questionnaires: these provide reviewers and other researchers with an overview of all studied variables. The second risk is that researchers conduct many trials, only publishing the data of those with results that fit their hypotheses, but not of the preceding trials with less fortunate results (the 'pre-replications', or prepublications). This risk can be addressed by requiring that not only all data and materials relating to the published study are provided, but also all data and materials relating to previous prepublications.

I am helping the competition if they have my materials

Sometimes, researchers are 'racing' each other to get results published first. At first glance, it seems as if you help the competition by publishing the materials (study protocol, computer tasks, questionnaires), because they no longer have to develop their materials themselves, which of course considerably speeds up their progress. However, materials would only be published when the article itself is published; so any race has already been won—or lost. This also resolves the potential problem where a researcher might be worried that a peer reviewer is a competitor and might abuse their 'preview access'. Of course, ideally, reviewers do have access to the materials, data, analysis scripts and output, as this would enhance the quality of the reviewing process. A solution could be to let reviewers sign an agreement to not use resources of reviewed articles until publication.

I want to sell my materials commercially

Researchers sometimes want to use their research to make money, for example by selling questionnaires or intervention manuals. Of course, when research is funded by public money, the results belong in the public domain, as taxpayers pay for the development of scientific materials and should therefore have access to the fruits of their initial investment. However, when research is conducted by commercial companies then they own their results and findings. We suggest that a clear line is drawn between scientific journals which share results and data for the advancement of science and so follow a full disclosure policy and commercial journals which do not require full disclosure because researchers publishing in them are seeking financial gain from work they own.

I don't want others to earn money with my hard work

Publishing your materials, protocols, data, analysis scripts, and output so that it is openly accessible, does not necessarily mean that everybody can use these resources however they please. A very simple way of determining which rights are provided is by using the Creative Commons licences. This not-for-profit organisation offers six licences, allowing researchers to determine whether they want others to be able to use a resource commercially or not, whether they want to be credited when the resource is used, and whether 'derivative works', works drawing on the original resource, have to be published under the same license ('share-alike'). Thus, it is easy to prohibit commercial use of research resources, securing citations, and 'paying it forward': making sure that others also share their work.

I am afraid errors are pointed out in my work

If a dataset and analysis protocol (preferably even the script file, e.g. an SPSS syntax file) is published, this makes a researcher vulnerable to others identifying errors in their methods and analyses. Few people like being told they were wrong. On the other hand, even less researchers would argue that science should be hampered to avoid researchers' or journals' embarrassment. A full disclosure behavioural science would facilitate collective acknowledgement that identification of mistakes in methodology and analyses are crucial to increasing accuracy in interpretation of data, thereby ensuring that mistakes do not retard scientific development; witness recent work in physics on faster than light neutrinos (Reich, 2011, 2012). Scientific data is not flawless. Falsification is central to scientific progress—the assumption that we will make mistakes is a basic assumption of our work. This is already reflected in letters to editors which allow researchers to react on papers (e.g. James & Smyth, 2012), and authors the possibility of reacting to such reactions (e.g. Crutzen, 2012).

My ethical committee requires me to use an informed consent where participants explicitly only provide permission for use of their data for my particular study, precluding re-analysis.

If a researcher is explicitly disallowed from publishing data because of a contract or because permission from an ethical committee is conditional upon non-publication of the data, little can be done. However, such arrangements are very rare; most legislators in fact encourage publication or research data. For example, the code of conduct for using personal data in scientific research that was developed by the Dutch universities based on the relevant Dutch legislation explicitly states "[...] in scientific research, the use of a previously created datafile

is allowed, also if the file was created for another reason, unless the file contains identifying data [...]" (VSNU, 2005). Thus, the practice of removing all identifying information from a datafile, which is already common practice for most researchers, suffices to enable publication and re-use of the datafile in most situations. Note that in any case, restrictions regarding publication of datafiles do not extend to publication of materials, study protocols, analysis scripts, and output files.

Guidelines for a Full Disclosure Science

Publishers already facilitate publication of supplementary materials, so all that is required is a change in authors' motivation which could be brought about by changes in editorial policy. A number of guidelines to optimize the benefits from publishing such supplemental materials follow, phrased as journal policy suggestions.

1. Require researchers to supply:
 - a. Everything necessary for replication (e.g., questionnaires, source code of computer tasks, or at least compiled tasks, detailed protocols, manual etc.);
 - b. For quantitative research, a datafile containing all variables involved in the analyses that are reported, and for qualitative research, the coding tree, and ideally, the sources with their codes;
 - c. A document detailing the analyses, which, together with the datafile, must enable accurate reproduction of the reported results (ideally, a script with the commands used, such as a syntax file for SPSS);
 - d. A record of all data collected relevant to the reported analyses e.g., datasets from preliminary datasets not included in the analyses;

e. The output used by the researcher on which the reported results were based (although this output should be replicable perfectly using the datafile and the analysis script, not all researchers use scripts for their analyses; in addition, such replication requires access to the same software, and researchers often use different proprietary software packages such as SPSS, SAS and STATA).

2. Require that these files be supplied in non-proprietary formats (this is important because opening proprietary formats require the purchase of specific software, which other researchers may not have). This means that:

a. Resources that cannot be provided in the preferred format, are provided in one of the default non-proprietary formats, such as plain text, Open Document Format (ODF), Portable Document Format (PDF), or Hyper Text Markup Language files (HTML), or, for images, Portable Network Graphics (PNG) or Scalable Vector Graphics (SVG);

b. Questionnaires and computer tasks are preferably provided in a format that can be imported into free non-proprietary software, for example LimeSurvey for questionnaires (LimeSurvey Project Team/Carsten Schmitz, 2012) and OpenSesame for computer tasks (Mathôt, Schreij, & Theeuwes, 2012). Stimuli and intervention materials are provided in the default formats (see 2a);

c. Quantitative datafiles are provided in generic data formats such as Tab or Comma Separated Values, and qualitative data in one of the default formats (see 2a);

d. Files with commands for statistical programs (e.g. SPSS, SAS, R) are provided in their original format, because these generally already are in plain text format. When a program is used that

does not store the analysis script in plain text, the researcher can copy-paste to one of the default formats (see 2a). Note that of course, using free non-proprietary programs such as R (R Development Core Team, 2012) is preferred, as all other researchers have access to this software;

e. The output can be provided in one of the default formats (see 2a).

These guidelines entail minimal efforts (and virtually no costs) from both journal editors and authors. However, the benefits are substantial: our evidence will become more accurate through correction of errors, replication, and much higher quality meta-analyses, and in addition, developing oneself as researcher will be much easier, which will benefit ourselves and our students once we involve these supplementary materials in our courses. Finally, it is likely that the mere requirement of publication of materials, data and analyses will already have a beneficial effect on the quality of our evidence base, as it has been shown that willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results (Wicherts et al., 2011). It is up to all of us to change and share to accelerate science, starting with our own sub-discipline; health psychology. Reactions from the editors of *Psychology & Health* and *Health Psychology Review* are more than welcome! ■

Acknowledgements

We would like to thank Gerjo Kok and Emily de Vet for feedback on earlier drafts of this text, and Catherine Bolman for advice on ethical aspects of data sharing.

References

- Abraham, C. (2012a). Creating change that counts: Evidence-led co-creation as a pathway to impact. *Psychology & Health*, 27(S1), 1.

- Abraham, C. (2012b). Designing and evaluating interventions to change health-related behaviour patterns. *Randomized Clinical Trials of Nonpharmacologic Treatments* (pp. 357–368). London: Chapman & Hall/CRC.
- American Psychological Association. (2010). Ethical principles of psychologists and code of conduct. *The American psychologist* (Vol. 57). Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20642307>
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43, 666–78. doi:10.3758/s13428-011-0089-5
- Crutzen, R. (2012). Author response: associations and mixed findings. *Mental Health and Substance Use*, 5, 271.
- Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral Priming: It's All in the Mind, but Whose Mind? (J. Lauwereyns, Ed.) *PLoS ONE*, 7, e29081. doi:10.1371/journal.pone.0029081
- French, C. (2012, March 15). Precognition studies and the curse of the failed replications. *The Guardian*. London. Retrieved from <http://www.guardian.co.uk/science/2012/mar/15/precognition-studies-curse-failed-replications>
- Ghosh, P. (2012). Report calls on government to back open access science. *BBC News: Science & Environment*. Retrieved November 7, 2012, from <http://www.bbc.co.uk/news/science-environment-18497083>
- Glantz, S. A. (1980). Biostatistics: how to detect, correct and prevent errors in the medical literature. *Circulation*, 61, 1–7. doi:10.1161/01.CIR.61.1.1
- James, P., & Smyth, B. P. (2012). Commentary on Crutzen et al. (2012), Negative associations between primary school children's perception of being allowed to drink at home and alcohol use, *Mental Health and Substance Use*, 5, 64–69.
- LimeSurvey Project Team/Carsten Schmitz. (2012). *LimeSurvey: An Open Source survey tool*. Hamburg, Germany. Retrieved from <http://limesurvey.org>
- Masicampo, E. J., & Lalande, D. R. (2012). A peculiar prevalence of p values just below .05. *Quarterly journal of experimental psychology*, 37–41. doi:10.1080/17470218.2012.711335
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior research methods*, 44, 314–24. doi:10.3758/s13428-011-0168-7
- Milton, J., & Wiseman, R. (1999). Does Psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin*, 125, 387–91.
- Oldenburg, H. (1665). Epistle Dedicatory. *Philosophical Transactions of the Royal Society of London*, 1(1-22). doi:10.1098/rstl.1665.0001
- Peters, G.-J. Y., Ruiter, R. A. C., & Kok, G. (2012). Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory. *Health Psychology Review*. doi:10.1080/17437199.2012.703527
- R Development Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from <http://www.r-project.org/>
- Reich, E. S. (2011). Faster-than-light neutrinos face time trial. *Nature.com news*. Retrieved October 12, 2012, from <http://www.nature.com/news/2011/111005/full/news.2011.575.html>
- Reich, E. S. (2012). Leaders of faster-than-light neutrino team resign. *Nature.com news*. Retrieved October 12, 2012, from <http://blogs.nature.com/news/2012/03/leaders-of-faster-than-light-neutrino-team-resign.html>
- Ritchie, S. J., Wiseman, R., & French, C. C. (2012). Failing the future: three unsuccessful attempts to replicate Bem's "retroactive facilitation of recall" effect. *PloS one*, 7, e33423. doi:10.1371/journal.pone.0033423
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–66. doi:10.1177/0956797611417632
- VSNU. (2005). *Gedragcode voor het gebruik van persoonsgegevens in wetenschappelijk onderzoek* [Code of conduct for the use of personal data in scientific research]. Den Haag, Netherlands.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, 100, 426–32. doi:10.1037/a0022790

Walther, A., & van den Bosch, J. J. F. (2012). FOSE: a framework for open science evaluation. *Frontiers in Computational Neuroscience*, 6, 1–8. doi:10.3389/fncom.2012.00032

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

van Dongen, A., Abraham, C., Ruiter, R. A., & Veldhuizen, I. (in press). Does questionnaire distribution promote blood donation? An investigation of question-behavior effects. *Annals of Behavioural Medicine*.



Gjalt-Jorn Peters

is Assistant Professor at the Faculty of Psychology, Open University, the Netherlands

gjalt-jorn@behaviorchange.eu



Charles Abraham

is Professor of Behaviour Change at the Peninsula College of Medicine & Dentistry, Universities of Exeter & Plymouth, UK.



Rik Crutzen

is a post-doctoral researcher at the Department of Health Promotion, Maastricht University /CAPHRI, the Netherlands