Intelligence xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Intelligence

Publish (your data) or (let the data) perish! Why not publish your data too?

Jelte M. Wicherts *, Marjan Bakker

University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 19 December 2011
Received in revised form 4 January 2012
Accepted 11 January 2012
Available online xxxx

Keywords:
Data sharing
Science policy
Data archiving
GPA

ABSTRACT

The authors argue that upon publication of a paper, the data should be made available through online archives or repositories. Reasons for not sharing data are discussed and contrasted with advantages of sharing, which include abiding by the scientific principle of openness, keeping the data for posterity, increasing one's impact, facilitation of secondary analyses and collaborations, prevention and correction of errors, and meeting funding agencies' increasingly stringent stipulations concerning the dissemination of data. Practicing what they preach, the authors include data as an online appendix to this editorial. These data are from a cohort of psychology freshmen who completed Raven's Advanced Progressive Matrices, tests of Numerical Ability, Number Series, Hidden Figures, Vocabulary, Verbal Analogies, and Logical Reasoning, two Big Five personality inventories, and scales for social desirability and impression management. Student's sex and grade point average (GPA) are also included. Data could be used to study predictive validity of cognitive ability tests, Extraversion, Neuroticism, Conscientiousness, Openness to Experience, Agreeableness, and the general factor of personality, as well as sex differences, differential prediction, and relations between personality and intelligence.

1. Introduction

Congratulations! You have just received the good news from the editor, popped open the bubbly to celebrate your latest scientific contribution, and – after having sobered up – are now finalizing your manuscript for publication in this journal. You have written an interesting paper, and are eager to share your empirical findings with the world. So why not go all the way, and publish your raw data too?

2. The cons of data sharing

Of course, there may be insurmountable reasons not to share the data. You may not own the data to begin with. Or sharing may be hampered by ethical considerations relating to privacy. The data may be part of a longitudinal study that will be featured in other articles, and are thus part of an ongoing project.

Perhaps you are hesitant to share your data for other reasons. You probably invested considerable time and resources in collecting and analyzing the data, and may therefore consider yourself the rightful owner (even if, say, the taxpayer helped finance the data). You may be afraid that some nitpicking statistician may one day discover an error in one of the statistical analyses that you reported in your paper or that a researcher with opposing views will disagree with your analyses. You may feel that you will be able to publish another paper with the data, provided you had the time. You may consider it too much trouble to document the data properly, which clearly takes time and effort. But such impediments to data sharing may be petty when compared to the benefits of sharing.

3. The pros of data sharing

You may not realize that there are many advantages to publishing the data following publication. First, the required

* Corresponding author at: Department of Psychology, Psychological Methods, University of Amsterdam, Weesperplein 4, 1018 XA, Amsterdam, The Netherlands. Tel.: +31 205257067; fax: +31 206390026.
E-mail address: J.M.Wicherts@uva.nl (J.M. Wicherts).

0160-2896/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2012.01.004

data archiving, i.e., making them suitable for dissemination, ensures that your data are not (ultimately) lost. One of us once requested data from a close colleague, who responded: "Sure I will send you those data, but it's like seven computers ago, and so please allow me some time to hunt them down". Apparently, archiving is not every psychologist's cup of tea; in a survey among 192 Dutch research psychologists, two thirds indicated to not archive their data (Voorbrood, 2010). Those who indicated to do so often confused proper archiving of data with the simple act of saving the file on their current computer. When Wicherts et al. requested data from 141 psychological researchers back in 2005 (Wicherts, Borsboom, Kats, & Molenaar, 2006), the relevant papers had been published less than 12 months earlier and only 27% of the researchers were willing and able to share (some of) their data. The low response rate was mostly due to the disappearance of the data and researchers' failure to document the data file in an accessible format. Only a small number of researchers flatly refused. Although we do expect *your* data handling practices to be relatively good in light of the relative complexity of intelligence test data, it would still be good to outsource archiving so as to save the data for posterity.

Second, sharing the data is consistent with the universal scientific norms of openness and rigor. We recently found that psychologists who fail to share their data after publication commit considerably more errors in the reporting of statistical results and report findings that are more likely to be debatable (Wicherts, Bakker, & Molenaar, 2011). Openness with respect to data is good for you and scientific progress, while concealment may facilitate all sorts of trouble, including faulty analyses, overly positive reporting of results, and even misconduct (Wicherts, 2011).

Third, publishing one's data has been shown to increase citation scores of the papers which first feature the data (Piwowar, Day, & Fridsma, 2007). So sharing is good for the publishing journal and it is good for the citation score on your resume. Notwithstanding our critique of Lynn and Vanhanen's (2006) work on national IQ (e.g., Wicherts, Dolan, Carlson, & van der Maas, 2010), their work was followed up by many researchers simply because Lynn and Vanhanen published their data in an accessible form.

Fourth, sharing the data encourages more research because it enables secondary (novel) analyses, for which you may currently lack time, knowledge, expertise, or even interest. Advances in both statistical techniques and knowledge of our research topic will certainly open up novel possibilities with your data in the future. We are sure that many of your colleagues, who use your published data for novel analyses, will have the courtesy to ask you to join the novel research project (as happened in, e.g., Dolan et al., 2006). Besides, having first-hand knowledge of the data, plus possible relevant background information, your help may be essential in those future studies. Gifts are often returned. In addition, meta-analysts often find a considerable number of published studies to be useless because primary researchers failed to report the effect size or correlation of interest, i.e., crucial information that is lost once the data are unavailable (Ortego & Botella, 2010). Proper data archiving is essential for the quality of meta-analyses (Cooper & Patall, 2009).

Fifth, sharing the data facilitates subsequent reanalyses, which may have diverse beneficial results. On the one hand, they may result in the correction of potential errors in the analyses and in the reporting of results (which are quite common; Bakker & Wicherts, 2011; Wicherts et al., 2011) and may even help prevent such errors in the first place (Wicherts, 2011). On the other hand, reanalyses may emphasize the robustness of your original substantive results. Of course, laying oneself open to criticism and subsequent correction is good scientific practice, but psychologically not necessarily palatable. However, if trivial errors are found, these are very unlikely to be publishable and will receive little attention (although ultimately one is well served by knowledge of such errors). If serious errors are found, you should politely thank the person who found them and learn from the experience (which is, after all, the essence of scientific progress).

Sixth, many funding agencies including the National Institutes of Health (NIH), National Science Foundation (NSF), the Wellcome Trust, the Medical Research Council (MRC), the Deutsche Forschungsgemeinschaft (DFG), and the Netherlands Organization for Research (NWO) in our country have stipulated that grantees should either write a data sharing plan as part of proposals or make the data publically available upon completion of the project. In fact, you are already obliged keep your data in accessible form for at least five years and to share your data for verification purposes upon request (American Psychological Association, 2010). Obviously: you need not worry about such stipulations, if you routinely submit your data along with your paper.

4. Putting the money where our mouth is

Data can be archived in several online data repositories. For instance, data could be put on the website of the International Society for Intelligence Research (http://www.isironline.org/) by emailing Timothy Bates. Together with IT-specialists and librarians at the University of Amsterdam, we are currently developing a repository that will eventually include much of the data from over 40 years of our freshman-testing program. Moreover, Elsevier's *ScienceDirect* offers the possibility to archive data as an online appendix (e.g., Hassal & Sherrat, 2011; Johnson & Bouchard, 2011). To illustrate the possibilities offered by Elsevier, we attach a data set to the current editorial that is potentially interesting for future use. The file includes data from our freshman-testing program called "Testweek" (Busato, Prins, Elshout, & Hamaker, 2000; Smits, Dolan, Vorst, Wicherts, & Timmerman, 2011; Wicherts & Vorst, 2010) in which 537 students (age: $M = 21.0$, $SD = 4.3$) took the Advanced Progressive Matrices (Raven, Court, & Raven, 1996), a test of Arithmetic, a Number Series test, a Hidden Figures Test, a test of Vocabulary, a test of Verbal Analogies, and a Logical Reasoning test (Elshout, 1976). Also included are data from a Dutch big five personality inventory (Elshout & Akkerman, 1975), the NEO-PI-R (Hoekstra, Ormel, & Fruyt, 1996), scales of social desirability and impression management (based on work by Paulhus, 1984; Wicherts, 2002), sex of the participants, and grade point averages of the freshmen's first trimester that may act as outcome variable. We had ambitious plans to use these data in studies of sex differences, differential prediction, advanced item response models, the general factor of personality (de Vries, 2011; Van der Linden, Te Nijenhuis, & Bakker,

Please cite this article as: Wicherts, J.M., & Bakker, M., Publish (your data) or (let the data) perish! Why not publish your data too?, *Intelligence* (2012), doi:10.1016/j.intell.2012.01.004

2010), the link between IQ and personality, etc., but we simply lacked the time. We therefore gladly make these data available for the consideration of the scientific community.

The data are anonymized and may be used for any purpose. Our only request is a proper citation to this editorial to make our point. The data are described in Table 1 and are given in an Excel file. The PDF in the supplement provides additional information on univariate distributions, some data-handling, number of items, time limits, and scale reliabilities. Fig. 1 gives the path diagram with standardized loadings of a reasonably well-fitting factor model of the seven cognitive tests; χ^2–$SB = 16.95$, $DF = 10$, $p = .076$, $RMSEA = .040$, $SRMR = 0.028$, $CFI = .992$. This model was fitted in LISREL 8.80 to the data of 429 participants who completed all seven subtests and who were native speakers of Dutch. The Satorra Bentler-scaling was inspired by the non-normality in some of the variables (caveat lector!). Perhaps the most interesting feature of sharing data is that anyone may disagree with the primary authors' analytic choices and run alternative models, use other selections of the data, possibly leading to different conclusions. To be sure: there is simply nothing more scientific than a debate about the raw data. Subscale-level data (for the NEO-PI-R), item-level data, and test materials (in Dutch) are available upon request.

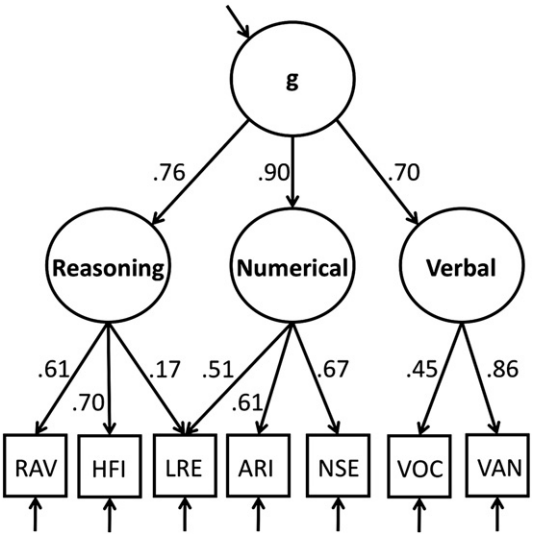

Fig. 1. Proposed factor model of the seven cognitive tests in the data set ($N = 429$) with standardized loadings.

5. Conclusion

By not archiving our data, much valuable data get lost every year. For instance, you may still have a 5¼ inch floppy disk with valuable data, but nowhere to insert it. We should stop this waste and start archiving in an accessible and robust way. This particularly applies to the types of wealthy multivariate data sets that we as intelligence researchers often collect. Perhaps the best example of secondary analyses in our field is Carroll's (1993) treatise on the structure of individual differences in cognitive abilities in which he factor analyzed data from over 460 studies by using their correlation matrices. This shows that it would be wise to always publish the full covariance matrix to accompany any factor analytic application. But the raw data could be essential. For instance, Molenaar and colleagues (Molenaar, Dolan, & Maas, 2011; Molenaar, Dolan, Wicherts, & van der Maas, 2010) recently developed novel methods to study Spearman's Law of Diminishing Returns and these methods can only be applied with the raw data. Our understanding of this phenomenon as well as many others rests greatly on the possibilities to replicate these results on the basis of “old” data and in fact

Table 1
Description of variables in the data file.

Variable	Description	*N*	Min	Max	Mean	SD
Sex	Sex of participant	534	1(=M)	2(=F)	1.72	.451
Language	Native Dutch speaker?	523	1(=Y)	2(=N)	1.10	.305
ZGPA	Z of 1st tri. Grade Point Average	314	−3.06	2.25	0.00	1.00
Rav_until	APM completed until item no.	521	14	36	27.24	4.13
Rav_score	APM no. correct	521	3.00	31.00	21.17	4.33
E5pft	Extraversion – 5PFT	517	30	90	64.00	10.81
A5pft	Agreeableness – 5PFT	517	25	96	72.17	9.27
C5pft	Conscientiousness – 5PFT	517	33	90	61.34	9.74
N5pft	Neuroticism – 5PFT	517	19	83	45.89	12.19
O5pft	Openness – 5PFT	517	39	96	66.40	9.32
Lretot	Logical reasoning total score	513	−14	40	12.62	9.01
Nsetot	Number series total score	527	−23	23	6.68	7.46
Voctot	Vocabulary total score	523	−28	21	−8.82	8.57
Hfitot	Hidden figures total score	512	−26	29	7.96	10.24
Vantot	Verbal analogies total score	519	−30	30	5.52	11.74
Aritot	Arithmetic total score	521	−52	48	12.74	7.86
Im	Impression Management	515	10	39	23.86	4.56
Sd	Self Deception	515	8	39	19.03	6.19
nNEO	Neuroticism – NEO-PI-R	500	79	210	135.67	21.73
eNeo	Extraversion – NEO-PI-R	500	112	212	167.37	17.04
oNEO	Openness – NEO-PI-R	500	127	220	172.47	16.33
aNEO	Agreeableness – NEO-PI-R	500	93	212	164.88	16.83
cNEO	Conscientiousness – NEO-PI-R	500	87	212	153.97	18.60

Note: Valid *N* (listwise) = 280.

Please cite this article as: Wicherts, J.M., & Bakker, M., Publish (your data) or (let the data) perish! Why not publish your data too?, *Intelligence* (2012), doi:10.1016/j.intell.2012.01.004

Roberto Colom kindly shared the Spanish WAIS-III standardization data precisely for that purpose (see also Dolan et al., 2006).

The call for more openness with respect to research data is rapidly increasing in many scientific fields (Nature, 2009), and, as intelligence researchers we should not lag behind. We owe much of our knowledge of intelligence from statistical analyses that have evolved over the years and are subject to further future refinement. Our ideas may chance, which may lead us to revisit our data. Not so long ago both editors and publishers, in view of journal space, desired journal articles to be written as densely as possible. But times have changed. With the onset of online publishing, we can publish sizeable files as online appendices to journal articles. Publishing the data in this way (or what other way the future may bring) has numerous advantages for both individual researchers and the field of intelligence research. Sharing data should become the rule rather than the exception; we owe it to our research participants and to science.

Acknowledgments

The preparation of this article was supported by Grant 400-08-215 from the Netherlands Organization for Research (NWO) and a SURFshare-grant from the Surf Foundation. We thank Floor van Alphen and Harrie Vorst for their help in preparing the data.

Appendix A. Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.intell.2012.01.004.

References

American Psychological Association (2010). *Publication Manual of the American Psychological Association* (Sixth Edition). Washington, DC: American Psychological Association.

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*, 666–678. doi:10.3758/s13428-011-0089-5.

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, *29*, 1057–1068. doi:10.1016/S0191-8869(99)00253-6.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press.

Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods*, *14*, 165–176. doi:10.1037/a0015565.

de Vries, R. E. (2011). No evidence for a General Factor of Personality in the HEXACO Personality Inventory. *Personality and Individual Differences*, *45*, 229–232. doi:10.1016/j.jrp.2010.12.002.

Dolan, C. V., Colom, R., Abad, F. J., Wicherts, J. M., Hessen, D. J., & van de Sluis, S. (2006). Multi-group covariance and mean structure modeling of the relationship between the WAIS-III common factors and sex educational attainment in Spain. *Intelligence*, *34*, 193–210. doi:10.1016/j.intell.2005.09.003.

Elshout, J. J. (1976). *Karakteristieke moeilijkheden in het denken. [Characteristic difficulties in thinking].* doctoral dissertation, University of Amsterdam, Amsterdam, The Netherlands.

Elshout, J. J., & Akkerman, A. (1975). *Vijf Persoonlijkheids-factoren Test 5PFT, handleiding [Five personality factors test 5PFT, manual].* Nijmegen, The Netherlands: Berkhout Nijmegen B.V.

Hassal, C., & Sherrat, T. (2011). Statistical inference and spatial patterns in correlates of IQ. *Intelligence*, *39*, 303–310. doi:10.1016/j.intell.2011.05.001.

Hoekstra, H. A., Ormel, H., & Fruyt, F. D. (1996). *NEO-PI-R Personality Inventory Dutch Manual.* Lisse, The Netherlands: Swets & Zeitlinger.

Johnson, W., & Bouchard, T. J. (2011). The MISTRA data: Forty-two mental ability tests in three batteries. *Intelligence*, *39*, 82–88. doi:10.1016/j.intell.2011.02.010.

Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality.* Augusta, GA: Washington Summit Publishers.

Molenaar, D., Dolan, C. V., & Maas, H. L. J. (2011). Modeling ability differentiation in the second-order factor model. *Structural Equation Modeling*, *18*, 578–594. doi:10.1080/10705511.2011.607095.

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. L. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, *38*, 611–624. doi:10.1016/j.intell.2010.09.002.

Nature (2009). Data's shameful neglect. *Nature*, *461*, 145. doi:10.1038/461145a.

Ortego, C., & Botella, J. (2010). The hard but necessary task of gathering order-one effect size indices in meta-analysis. *Psicológica*, *31*, 291–315.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, *46*, 598–609.

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, *2*, e308. doi:10.1371/journal.pone.0000308.

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford, UK: Oxford Psychologists Press.

Smits, I. A. M., Dolan, C. V., Vorst, H. C. M., Wicherts, J. M., & Timmerman, M. E. (2011). Cohort differences in big five personality factors over a period of 25 years. *Journal of Personality and Social Psychology*, *100*, 1124–1138. doi:10.1037/a0022874.

Van der Linden, D., Te Nijenhuis, J., & Bakker, A. B. (2010). The General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, *44*, 315–327. doi:10.1016/j.jrp.2010.03.003.

Voorbrood, C. (2010). *Archivering, beschikbaarstelling en hergebruik van onderzoeksdata in de psychologie [Archiving, sharing, and reusing of psychological research data].* The Hague, The Netherlands: Data Archiving and Networked Services.

Wicherts, J. M. (2002). *Meetequivalentie tussen pen & papier en computerafname van vragenlijsten: Alexithymie, sociale wenselijkheid en angst [Measurement equivalence of paper-and-pencil and computerized administration of questionnaires: Alexithymia, social desirability, and anxiety].* Unpublished Master's Thesis, University of Amsterdam.

Wicherts, J. M. (2011). Psychology must learn a lesson from fraud case. *Nature*, *480*, 7. doi:10.1038/480007a.

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.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726–728. doi:10.1037/0003-066X.61.7.726.

Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. J. (2010). Raven's tests performance of Africans: Average performance, psychometric properties, and the Flynn Effect. *Learning and Individual Differences*, *20*, 135–151. doi:10.1016/j.lindif.2009.12.001.

Wicherts, J. M., & Vorst, H. C. M. (2010). The relation between specialty choice of psychology students and their interests, personality, and cognitive abilities. *Learning and Individual Differences*, *20*, 494–500. doi:10.1016/j.lindif.2010.01.004.

Please cite this article as: Wicherts, J.M., & Bakker, M., Publish (your data) or (let the data) perish! Why not publish your data too?, *Intelligence* (2012), doi:10.1016/j.intell.2012.01.004