The effect of journal guidelines on the reporting of antibody validation

5 Joyce M Hoek[1,2], Wytske M Hepkema [2], Willem Halffman [2]

7 [1] Department of Psychology, University of Groningen, Groningen, The Netherlands
8 [2] Institute for Science in Society, Radboud University Nijmegen, Nijmegen, The Netherlands

10 Corresponding Author:
11 Joyce M Hoek [1,2]
12 Grote Kruisstraat 2/1 Heymans Building, Groningen, Province of Groningen, 9712 TS, The
13 Netherlands
14 Email address: j.m.hoek@rug.nl

Abstract

17 **Background.** Despite the widespread use of antibodies as a research tool, problems with
18 specificity, lot-to-lot consistency and sensitivity commonly occur and may be important
19 contributing factors to the 'replication crisis' in biomedical research. This makes the validation
20 of antibodies and accurate reporting of this validation in the scientific literature extremely
21 important. Therefore, some journals now require authors to comply with antibody reporting
22 guidelines.

24 **Methods.** We used a quasi-experimental approach to assess the effectiveness of such journal
25 guidelines in improving antibody reporting in the scientific literature. In a sample of 120
26 publications, we compared the reporting of antibody validation and identification information in
27 two journals with guidelines (*Nature* and the *Journal of Comparative Neurology*) with two
28 journals without guidelines (*Science* and *Neuroscience*), before and after the introduction of
29 these guidelines.

31 **Results.** Our results suggest that the implementation of antibody reporting guidelines might have
32 some influence on the reporting of antibody validation information. The percentage of validated
33 antibodies per article slightly increased from 39% to 57% in journals with guidelines, whereas
34 this percentage decreased from 23% to 14% in journals without guidelines. Furthermore, the
35 reporting of validation information of all primary antibodies increased by 23 percentage points in
36 the journals with guidelines (OR=2.80, 95%CI=0.96-INF; adjusted p=1, one-tailed), compared to
37 a decrease of 13 percentage points in journals without guidelines. Fortunately, the guidelines
38 seem to be more effective in improving the reporting of antibody identification information. The
39 reporting of identification information of all primary antibodies used in a study increased by 58

40   percentage points (OR=17.8, 95%CI=4.8-INF; adjusted *p*=0.0003, one-tailed) in journals with
41   guidelines. This percentage also slightly increased in journals without guidelines (by 18
42   percentage points), suggesting an overall increased awareness of the importance of antibody
43   identifiability. Moreover, this suggests that reporting guidelines mostly have an influence on the
44   reporting of information that is relatively easy to provide. A small increase in the reporting of
45   validation by referencing the scientific literature or the manufacturer's data also indicates this.
46

47   **Conclusion.** Combined with the results of previous studies on journal guidelines, our study
48   suggests that the effect of journal antibody guidelines on validation practices by themselves may
49   be limited, since they mostly seem to improve antibody identification instead of actual
50   experimental validation. These guidelines, therefore, may require additional measures to ensure
51   effective implementation. However, due to the explorative nature of our study and our small
52   sample size, we must remain cautious towards other factors that might have played a role in the
53   observed change in antibody reporting behaviour.
54

55   Introduction

56   Antibodies are veritable workhorses in biomedical research. Used to label specific molecules or
57   antigens (generally proteins), they allow researchers to map biomolecular processes in the cell.
58   Through techniques such as western blotting (WB), immunohistochemistry, or ELISA, the use of
59   antibodies has become widespread in biological research.
60   However, binding antibodies to antigens is intricate and full of complications. Antibodies
61   may lack specificity and their affinity for specific antigens may vary. Furthermore, their affinity
62   may vary with experimental conditions (Baker 2015b), for example when pH levels or reagents
63   denature proteins, altering protein folding and thereby the epitopes to which antibodies bind.
64   Among tens of thousands of antibodies on offer (largely from commercial suppliers), researchers
65   need to identify precisely which antibodies are most suitable for their antigens of interest, under
66   their precise experimental circumstances. To aid this selection, extensive support tools have been
67   developed, such as the Antibodypedia, or the Antibodyregistry, which support the identification
68   of antibodies with research resource IDs (RRIDs) (Bandrowski et al. 2016).
69   Nevertheless, even with the help of such resources, the practice of antibody use remains
70   complex. While researchers may identify the correct antibody for their specific research purposes
71   *in principle*, verifying that the correct antibodies are used *in practice* is a different matter.
72   Antibodies may vary from batch to batch, suppliers may not always be able to guarantee relevant
73   quality standards, or earlier mistakes may be obfuscated by relying on locally established
74   routines, such as habits, experimental skills, and techniques passed on in a laboratory. If
75   identification and validation information about antibodies is not reported accurately, the
76   possibility of experimental replication is jeopardised, and subsequent research may be built on
77   errors. In turn, this may lead to wasted research, missed opportunities for medical innovation, or
78   even patient safety risks. Taken together, the costs involved may be considerable. In fact, some
79   commentators suggest problems with antibody validation, or lack of validation information, may

80   be an important contributing factor to the 'replication crisis' in biomedical research (Freedman et
81   al. 2015).
82   For over a decade, various researchers have expressed concerns about insufficient
83   antibody validation in biomedical research (Baker 2015a; Bordeaux et al. 2010; Saper &
84   Sawchenko 2003). The challenges are considerable. One study validating over 5000 commercial
85   antibodies for the Human Protein Atlas (HPA) showed that half of these antibodies were not
86   suitable for the specific immunohistochemistry application in the HPA. The researchers
87   concluded that every application of antibodies requires application-specific validation (Berglund
88   et al. 2008).
89   Advocates of tighter validation have suggested techniques and principles to ensure
90   correct antibodies are used and to improve reporting of antibody validation information in
91   publications, although there is no universal standard as yet (GBSI 2016). At the very least, these
92   advocates suggest that researchers should report information that establishes the identity of
93   antibodies used, via reference to RRIDs or supplier identifiers, including catalogue and batch
94   number (Vasilevsky et al. 2013). While many researchers rely on the literature to establish
95   application-specific validity of antibodies, actual validation testing avoids repetition of older
96   mistakes. This testing should verify not just the identity, avoiding supply-line errors (such as
97   misidentification during personal exchange, or transport via suppliers), but also the antibody's
98   specificity for the target protein and sensitivity in the specific application. Relatively simple
99   checks include staining a western blot to check whether antibodies recognise antigens of the
100 correct molecular weight, omitting the primary antiserum, and performing pre-adsorption
101 controls (Saper & Sawchenko 2003). While these methods might be useful as a first indicator of
102 an antibody's specificity, they are not very stringent, and antibodies might still be found to be
103 nonspecific upon more thorough validation (Andersson et al. 2017; Bordeaux et al. 2010; Hewitt
104 et al. 2014; O'hurley et al. 2014).
105   The International Working Group of Antibody Validation (IWGAV) proposed a very
106 stringent validation procedure in 2016, with five 'pillars' for application-specific validation.
107 These include genetic strategies (testing the antibody in conditions when the protein is not
108 expressed), orthogonal strategies (comparing results for varying amounts of target protein
109 identified by other means), independent antibody strategies (comparing results with alternative
110 antibodies), expression of tagged proteins (using affinity tags or fluorescent proteins), and
111 immunocapture followed by mass spectrometry. The IWGAV advises researchers to carry out at
112 least one of these five methods in order to validate whether an antibody is truly specific to the
113 application at hand. Furthermore, the IWGAV advises suppliers to also use at least one of these
114 five pillars for validation, including validation for each new batch of antibodies, and to provide
115 specific information on optimal use (Uhlen et al. 2016).
116   Several initiatives encourage researchers to improve antibody validation. Apart from calls
117 to action in editorials, the development of databases, and improved validation by suppliers, some
118 journals have also stepped up to the plate. The first and most vocal initiative was taken by the
119 *Journal of Comparative Neurology* (JCN), introducing explicit requirements for antibody

120   validation in its author instructions in 2003 (Saper 2005; Saper & Sawchenko 2003). Several
121   journals followed suit and now require various levels of antibody validation information: some
122   require only identification information; others require documentation or actual experimental
123   proof of application-specific validity (Alexander Steve et al. 2018; Gore 2013; Nature 2013).
124   In essence, these journal initiatives constitute rule-based or 'regulatory' policy,
125   attempting to change researchers' behaviour through regulation. These vary from insistent but
126   voluntary guidelines, through checklists that attempt to 'nudge' authors into compliance (Babic
127   et al. 2019), or actual rules at penalty of editorial rejection. In policy sciences, the track record of
128   rule-based behaviour modification is mixed, at best. Rules require policing, which is expensive,
129   and rules tend to fail if they lack community support, for example if they are perceived as
130   meaningless or ineffectual. In matters of research integrity, rules that blatantly diverge from
131   actual practice have been shown to induce cynicism among researchers, decreasing willingness
132   to comply (Clair 2015).
133   Although journals have long been considered powerful 'gatekeepers of science' (Crane
134   1967), and journals are generally expected to maintain standards in research communities, their
135   power to do so is not absolute. Journals are also dependent on research communities for
136   submitted articles and willingness of reviewer cooperation, compounded by the hard economics
137   of publishing in which volume is often the backbone of the business model (Larivière et al.
138   2015). Hence, the possibility of journals to raise standards may be limited by what the research
139   community is willing to maintain.
140   From this perspective, it is interesting to investigate the effect of journal antibody
141   guidelines on the presence and quality of reported validation information. For these purposes, we
142   chose to compare antibody validation information in journals with guidelines and in equivalent
143   journals without guidelines, i.e. a control group. Our hypothesis was that antibody validation
144   information would improve in the journals that introduced antibody validation guidelines.
145   Before we turn to the details of our research method, we want to discuss some key
146   methodological assumptions in validation testing, which touch upon quite fundamental issues of
147   philosophy of science. First, we want to make an important distinction between antibody
148   reliability and antibody validity. Antibodies may reliably produce the same results if
149   experimental conditions are reproduced exactly, but if these conditions are based on
150   misconceptions, they could simply reliably reproduce systematic error. A striking example of
151   how the reproduction of error might lead research astray, comes from the recent evaluation of
152   antibodies against oestrogen receptor beta in breast cancer research. Andersson et al. discovered
153   that the antibodies most cited in the literature had systematically stained the wrong proteins in
154   tissue samples and that the target protein was not even present in breast cancer tissue, affecting
155   two decades of research (Andersson et al. 2017). Hence, validation would ideally constitute the
156   experimental confirmation under different conditions, providing more robust replication, rather
157   than identical reproduction, as recent philosophy of science has argued (Leonelli 2018). In the
158   case of antibodies, the principles of the 'five pillars' offer such variation in experimental
159   conditions. In contrast, strictly speaking, reference to exact reproduction of antibodies as used in

160 the literature only constitutes proof of reliability, although it is frequently presented as proof of
161 validity.
162   Second, while exact reproduction can be considered important evidence for antibody
163 reliability in specific applications, in the practice of antibody research this evidence is not
164 absolute: it always remains possible to question whether the reproduced experiment was indeed
165 identical or performed equally expertly. Lack of exact reproduction raises the possibility that the
166 repeat experiment failed to reproduce the exact original conditions and materials, or lacked some
167 quality in the performer, such as particular experimental skills (Collins 1985). In the case of
168 antibodies, this is particularly pertinent in the case of polyclonal antibody batches that have run
169 out, for example as source animal populations are terminated. Moreover, whether an application
170 of antibodies constitutes an exact reproduction may be difficult to establish based on concise
171 information provided in databases or even publications. Lack of reproduction therefore does not
172 automatically disqualify the reliability of the original experiment and exact reproduction may be
173 hard to assess in practice.
174   In scoring the quality of validation information, we should therefore pay attention to
175 information that establishes the identity of antibodies, to evidence for antibodies' reliability in
176 the specific intended purposes, and to evidence that experimentally confirms the antibodies'
177 adequacy to identify the intended antigens with precision. While only the latter constitutes proper
178 validation, information about identity and reliability is a precondition of such higher levels of
179 validation.
180

Materials & Methods

Research strategy

183 In this exploratory study, we used a quasi-experimental study design to assess whether journal
184 antibody validation guidelines can improve antibody validation reporting (see figure 1). We
185 selected two journals that had adopted such guidelines and two journals with a comparable
186 impact factor and research area, but without guidelines. We took random samples of 15 articles
187 from each journal from a one-year timeframe before and after the two journals implemented
188 antibody reporting guidelines. We used the study performed by Han et al. (2017) as an indication
189 of the sample size and did not perform a power calculation beforehand. Since the first selected
190 journal with guidelines created awareness of the antibody problem in August 2003, the 'before
191 guidelines' samples consisted of articles that were published between August 2002 and August
192 2003. The 'after guidelines' samples consisted of articles that were published between June 2017
193 and June 2018, because the second journal with guidelines implemented its most stringent
194 guidelines in June 2017. We then used an antibody validation coding table to assess the antibody
195 validation information provided by each of the articles. We evaluated a total of 120 articles.
196

197 Journal selection

198 Intervention group

199 The *Journal of Comparative Neurology (JCN)* and *Nature* were selected as the two journals with
200 antibody validation guidelines. With regard to antibodies, *JCN* requires authors to add an
201 'Antibody Characterization' section to the methods section of a paper.[1] This section should
202 include complete identification information of the used antibodies and proof of their validity.
203 Authors can report their own validation or describe and cite evidence from the literature. The
204 guidelines further specify that just a casual reference to a supplier's data sheet does not count as
205 reliable evidence of specificity (information from Author Guidelines section of website *Journal*
206 *Comparative Neurology,* retrieved on 04-05-2018). When authors want to submit a manuscript to
207 *JCN,* they need to check a box that states that their publication provides full (validation)
208 information on all antibodies used (article submission process tested on 09-05-2018).
209   The journal *Nature* has been asking authors in the life sciences to fill out a reporting
210 checklist since May 2013. Since June 2017, this checklist has been replaced by a reporting
211 summary to be published as supplementary information alongside the article. While the previous
212 checklist and current reporting summary are not uniquely intended for antibodies, these
213 documents do ask for exhaustive identification and validation information on antibodies.
214 Furthermore, filling out this reporting summary is a requirement for publication in *Nature*. The
215 form needs to be uploaded at submission to prevent manuscript processing delays (article
216 submission process tested on 09-05-2018). A more elaborate explanation of these journals'
217 guidelines can be found in supplemental article S1.

218 Control group

219 The InCites Journal Citations Reports function of Web of Science (WoS) was used to find
220 journals comparable to *JCN* and *Nature*. Comparable journals were chosen by selecting the first
221 journal with a similar impact factor from the Cited Journal Data. Journals in this network are
222 considered similar because they fit in the same research area and are part of the same citation
223 network.
224   In the case of *JCN,* the most similar journal is *Neuroscience* with a similar impact factor
225 and research category of neurosciences. The journal *Neuroscience* does not have antibody
226 validation guidelines. No specific information about antibodies can be found in the author
227 guidelines for article submission on *Neuroscience's* website. The website does provide
228 information on resource materials in general. The source of all materials used should be provided
229 and the location of each supplier should be provided on first use in the text (information from
230 Guide for Authors section on website *Neuroscience,* retrieved 04-05-2018). However, at
231 submission, authors are not asked to provide information on the materials they used or on
232 antibodies specifically (article submission process tested on 09-05-2018).

---

[1] Editorial policies at *JCN* have recently changed. The Antibody Characterization section in force at the time of this study is no longer a requirement.

233   A comparable journal to *Nature* is *Science*, as they are both multidisciplinary journals
234   with similar impact factors. Using the same method as before, *Science* is also the first journal
235   with a comparable impact factor in the list of cited journals of *Nature*. While *Science* does state
236   that antibodies should be validated, this instruction is cursory and hard to find. It is written
237   between brackets on the 'Editorial Policies' page of the Science website, rather than in author
238   instructions. There is no further elaboration on what validation should consist of. Furthermore,
239   the manuscript submission process of *Science* never asks authors for validation information on
240   materials that are used in a study and never mentions antibodies. Hence, *Science* can be regarded
241   as having very marginal, to no guidelines on antibody reporting when compared to *Nature*
242   (information from Information for Authors/Editorial Policies section on website *Science,*
243   retrieved 04-05-2018; article submission process tested on 09-05-2018).

244   Article selection

245   WoS was used to search for articles published between 2002 and 2003 or between 2017 and
246   2018 in each of the four journals. These lists of articles were then randomly sorted by assigning
247   them a random number and sorting these numbers from smallest to largest. The first 15 articles
248   in these lists in the time span of August 2002 – August 2003 and June 2017 – June 2018
249   (according to the WoS publication date) that contained antibodies as a resource material were
250   then selected for assessment, selecting a total of 120 articles. Using antibodies as a resource
251   material means that they are used as a tool in a biological method such as western blotting or
252   immunohistochemistry. Antibodies can also themselves be the subject of scientific research. In
253   this case the article was only included in the sample if antibodies were additionally used as a
254   resource material. To determine if an antibody was used in a study, first the title of the article
255   was used to determine if the article was published in a field that could use antibodies (e.g. life
256   sciences and not physics). Next, the full text of the article was visually screened and screened for
257   the word 'anti' to see if antibodies as a resource material were mentioned in the article. If this
258   word did not occur in the full text in relation to antibodies, it was assumed that no antibodies had
259   been used. In case of doubt, supplementary information was used to determine if an article
260   should be included in the sample.

261   Coding

262   Coding protocol

263   All articles and their supplementary information were then scanned for antibody (validation)
264   information, and this information was collected in a coding table. The coding table was carefully
265   constructed by evaluating the literature for different antibody validation methods, by consulting
266   researchers working with antibodies, and by checking the table on a set of testing articles.
267   Information was collected per article, not per antibody. If multiple antibodies were validated by
268   multiple methods of validation, all methods were noted for this paper. In case of doubt about the
269   answer to a question, the decision was always made to rule in favour of the article. For example,

270   the highest number of validated antibodies would be written down in case of doubt about how
271   many antibodies were validated.
272   The coding table consisted mostly of 'yes' and 'no' questions such as: 'are all primary
273   antibodies validated?' or 'is validation carried out by positive control?' When a particular
274   question was not relevant for the evaluated article (for example, the questions about type of
275   validation when no antibodies were validated), this question was evaluated as 'not applicable'.
276   When it was impossible to answer a question for a certain article, this question was coded
277   'unclear'. The answer 'unclear' was classified as missing in data analysis.
278   We wanted to know the percentage of articles that report validation or information on all
279   primary antibodies from the sample of all articles that use antibodies. Therefore, for the
280   questions 'are all primary antibodies validated?' and 'is the basic information of all primary
281   antibodies complete?', the answer 'not applicable' was coded as 'no' to calculate the percentage
282   of articles that contain this information in all articles with antibodies. For all other questions
283   about validation type, the answer 'not applicable' was classified as missing. We made this choice
284   because we wanted to know the prevalence of these validation types in articles with validated
285   antibodies, instead of their prevalence in all articles using antibodies.
286   A full overview of the coding protocol, the questions in the coding table and the different
287   antibody validation methods is provided in supplemental article S2.

288   Implicit and explicit validation

289   Because we assessed antibody validity information by reading the publication, our coding made
290   no distinction between explicit or purposeful validation and implicit or 'accidental' validation, as
291   the intention to validate cannot always be inferred from the text. Whether explicitly presented as
292   validation or not, in both cases the antibody (and thus the article it was used in) was coded as
293   validated. However, for the interpretation of our findings, it is important to understand this
294   distinction between implicit and explicit validation.
295   Explicit validation means that the article provides information with the explicit purpose
296   to attest antibody validity. In this case, experiments are presented with the purpose of validating
297   the antibody, and/or the words antibody validation are explicitly mentioned. An example from
298   one of the evaluated articles would be the use of western blot with molecular weight markers to
299   validate the antibody (see figure 2).
300   Implicit validation here means that from reading the article, claims about the validity of
301   the antibody can be made, but no experiments have specifically been reported to establish the
302   validity of the antibody. The words 'antibody validation' are not mentioned. An example from
303   one of the evaluated articles would be the use of RNA interference as an antibody validation
304   method (see figure 3).

305   Methodological reliability

306   All 120 articles were evaluated by one of us (Joyce Hoek). An independent rater (Wytske
307   Hepkema) then assessed a randomly selected sample of these articles to determine the interrater

308 reliability. This sample consisted of 2 articles for each journal and time frame, resulting in 16
309 articles or 13% of the total sample. Since Wytske Hepkema only analysed a small percentage of
310 the total sample, we only used her answers to estimate the interrater reliability, and we used the
311 answers of Joyce Hoek for data analysis.
312   The percentage of agreement between both raters (Joyce Hoek and Wytske Hepkema)
313 was calculated using the agree function of the irr package (Gamer et al. 2012) in R. For nominal
314 variables, Cohen's kappa was calculated using the cohen.kappa function of the psych package
315 (Revelle & Revelle 2015) in R. For continuous variables, the intra-class correlation coefficient
316 (ICC) was calculated using the icc function of the irr package in R. ICC estimates were
317 calculated based on a single-measures, consistency, two-way mixed-effects model. Missing
318 values were not taken into account in the calculation of kappa or the ICC. Cohen's kappa was
319 indeterminate in some cases because one or both raters answered 'yes' or 'no' for all evaluated
320 articles. This is indicated by IND in the table. Table 1 shows the percentage agreement and
321 Cohen's kappa or intra-class correlation coefficient.
322   The agreement between raters varied considerably between the different questions, from
323 a kappa of 1 to a worse than chance kappa of -0.20. Antibody validation information is often not
324 explicitly reported, as opposed to antibody identification information, which also had a higher
325 interrater agreement. The low agreement on antibody validation questions thus illustrates the
326 current problem of antibody validation reporting: it is very difficult for a reader to assess the
327 validity of antibodies used in paper and therefore assess the validity of the performed
328 experiments. A more standardized way of reporting could improve this. The large disagreement
329 indicates that it is difficult to establish the validity of antibodies from reading the publication
330 when antibody validation is not explicitly mentioned. The disagreement between the two raters
331 when it comes to the categories 'validation carried out by the authors of the paper', 'negative and
332 positive control' and 'other types of validation' illustrates this. These types of validation are
333 often consequential to the methods used in an experiment. For example, antibodies are used to
334 validate the silencing of a gene, and, at the same time, the silencing of this gene can be seen as a
335 method to validate the antibody. This is a type of circular reasoning that only holds if either the
336 antibody or the silencing has proven to be valid by other methods. Therefore, it is difficult to
337 judge if the experiment validates the antibody or not. Because of the low agreement on these
338 questions, they were not taken into further consideration during data analysis.
339   On the other hand, the more explicit types of validation are easier to recognize and to
340 judge by readers of the paper. This is illustrated by the high percentage of agreement between
341 both raters and kappa value on the categories 'validation by reference to supplier, the literature,
342 or a database', 'molecular weight similar to target', 'spatial localization', 'pre-adsorption' and
343 'secondary antibody without primary'.

344 Data analysis

345 We compared the reporting of antibody validation information between the samples of journals
346 with and without guidelines and between the samples of August 2002 – August 2003 and June

347   2017 – June 2018. We expected the proportion of articles reporting validation to be higher in
348   2017 than in 2003 in journals that implemented guidelines. Furthermore, the proportion of
349   validation was expected to be larger in 2017 in the samples of journals that implemented
350   guidelines compared to journals that did not implement guidelines. We measured the difference
351   in validation reporting by comparing the percentage of validated antibodies per article and the
352   proportion of articles reporting any validation information. We also decided to look at the
353   proportion of articles reporting complete identification or validation information on all primary
354   antibodies as a more robust measure of validation rigour.
355   For the comparisons of proportions, one-tailed *p*-values were calculated, since these
356   hypotheses are directional. When comparing the samples of journals without guidelines between
357   2003 and 2017, no increase in proportion was expected. Likewise, the samples of similar
358   journals with and without guidelines taken at baseline in 2003 were not expected to be different.
359   For this reason, two-tailed *p*-values were calculated in these cases. Because of the small sample
360   size, we used Fisher's exact test (fisher.test in R) to compare the two proportions (instead of the
361   chi-square or z-test for two proportions). This function uses the conditional Maximum
362   Likelihood Estimate as an estimate of the odds ratio. For one-tailed tests, the odds ratio goes to
363   infinity. This is indicated by INF in the text. To account for multiple comparisons, we used the
364   Holm-Bonferroni correction (p.adjust in R) to adjust all calculated *p*-values. Data analysis was
365   carried out using R version 3.3.3.
366   In line with recent debate about the value of *p*-values (Wasserstein et al. 2019) and
367   because of the exploratory nature of our study, we decided to not only look at statistically
368   significant (*p*<0.05) results, but also take into account meaningful but non-significant
369   differences. To us, a meaningful difference is a substantial increase in proportion of validated
370   articles, accompanied by a large odds ratio. We think these differences are interesting and might
371   be good targets to further investigate if this study were to be replicated.
372   We provide our data and R code in supplemental data S1 and S2 and supplemental article
373   S3. Our data can also be found at https://doi.org/10.17026/dans-xhk-74m4
374
375   **Results**

376   Percentage of validated antibodies per article

377   For each evaluated article, the percentage of validated antibodies per article was calculated from
378   the total number of antibodies used and the number of validated antibodies. An overview of these
379   numbers can be found in supplemental table S1.
380   Figure 4 provides a descriptive overview of the percentages of validated antibodies in
381   each evaluated article as well as the mean percentage per journal. This figure shows that the
382   mean percentage of antibodies that have been validated per article for each journal slightly
383   increased in both journals with guidelines (*JCN* and *Nature*) between 2003 and 2017. Combined,
384   these journals increased from an average of 39% to 57% of validated antibodies per article. This

385   percentage decreased in the journals that did not introduce reporting guidelines. These journals
386   went from an average of 23% to 14% of validated antibodies per article.
387   Furthermore, there is a difference in reporting between similar journals with and without
388   guidelines. This difference in percentage was already present before *JCN* and *Nature* introduced
389   guidelines, but it increased after the introduction of these guidelines. The difference in average
390   percentage of validated antibodies between *JCN* and *Neuroscience* was 20 percentage points
391   before the introduction of guidelines and increased to 49 percentage points after the introduction
392   of guidelines by *JCN*. Between *Nature* and *Science,* the difference in average percentage was 12
393   percentage points, which increased to a difference of 38 percentage points after the introduction
394   of guidelines by *Nature*.

395   Validation of at least one antibody

396   Next, we made comparisons about antibody validation at the article level. Instead of looking at
397   the percentage of validated antibodies per article, we now compared the percentage of articles
398   that reported the validation of at least one antibody (see supplemental table S2 and S3). The
399   reporting of validation improved slightly, but insignificantly, in the journals that had introduced
400   guidelines. In addition, after the introduction of guidelines, 90% of the articles that use
401   antibodies in the journal with guidelines reported validation of at least one antibody compared to
402   53% of articles in journals without guidelines (OR=7.6, 95%CI=2.11-INF; adjusted *p*=.09, one-
403   tailed). However, since it might be quite simple to implicitly validate at least one antibody per
404   article, we also chose to compare the percentage of articles reporting antibody (validation)
405   information for all primary antibodies that they used as a more robust measure of antibody
406   reporting quality.

407   Validation of all primary antibodies

408   First, we compared the percentage of articles that report validation information on all the primary
409   antibodies used per study. This percentage increased between 2003 and 2017 in both journals
410   (*JCN* + *Nature*) that implemented reporting guidelines (see figure 5 and supplemental table S4
411   and S5). On average, this percentage increased by 23 percentage points in the journals that had
412   implemented guidelines (OR=2.80, 95%CI 0.96-INF; adjusted *p*=1, one-tailed). In contrast, the
413   percentage of articles with validation information on all primary antibodies decreased in both
414   journals without guidelines (*Science* + *Neuroscience*). Furthermore, there was no meaningful
415   increase in the total sample (18% to 22%, OR=1.28, 95%CI=0.48-3.52; adjusted *p*=1, two-
416   tailed).
417   At the journal level, there was an increase of 24 percentage points (OR=2.61,
418   95%CI=0.60-INF, adjusted *p*=1, one-tailed) in articles that reported validation information on all
419   primary antibodies in *JCN* and an increase of 22 percentage points in *Nature* (OR=5.3,
420   95%CI=0.59-INF; adjusted *p*=1, one-tailed).

421   By comparing the journals with guidelines to the journals without guidelines, we can see
422   that the difference in reporting between these similar journals in our sample substantially
423   increased after the introduction of guidelines. After June 2017, 46% of the articles in journals
424   that did implement guidelines (*JCN + Nature*) reported validation of all primary antibodies after
425   the introduction of guidelines, compared to 0% in journals that did not implement guidelines
426   (*Neuroscience + Science*), a difference of 46 percentage points (OR=INF, 95%CI=6.5-INF;
427   adjusted *p*=.0007, one-tailed). While a difference in percentage was already present before the
428   introduction of guidelines in 2003, this difference was much smaller (10 percentage points,
429   OR=1.96, 95%CI=0.43-10.3; adjusted *p*=1, two-tailed).
430   At the journal level, there was a large difference of 64 percentage points (OR=INF,
431   95%CI=5.1-INF, adjusted *p*=.01, one-tailed) in validation reporting between the journals *JCN*
432   and *Neuroscience* after June 2017. Between *Nature* and *Science* there was a difference in
433   reporting of 29 percentage points (OR=INF, 95%CI=1.09-INF; adjusted *p*=1, one-tailed) after
434   *Nature* had introduced guidelines, a difference that was not present before.

435   Identification of all primary antibodies

436   In addition to looking at antibody validation, we also evaluated how well antibody identification
437   information was reported for all primary antibodies used in a study. Antibodies are considered
438   identifiable if enough information is provided in the publication for readers to be able to track
439   down the antibody and obtain it themselves. In order to do so, either an RRID or the name,
440   supplier and catalogue number of a commercial antibody needs to be reported. For non-
441   commercial antibodies either an RRID or the host-animal and immunogen used needs to be
442   reported. Figure 6 and supplemental table S6 and S7 show the percentage of articles that reports
443   identification information on all primary antibodies that were used in the article.
444   The percentage of articles that reported identification information on all primary
445   antibodies increased considerably between 2003 and 2017 in journals that implemented antibody
446   guidelines. Overall, the percentage of articles that reported identification information on all
447   primary antibodies increased with 58 percentage points from 10% to 68% (OR=17.8
448   95%CI=4.8-INF; adjusted *p*=.0003, one-tailed) in journals with guidelines (*JCN + Nature*). Of
449   the separate journals, *JCN* showed the most improvement in reporting. After the implementation
450   of guidelines, its percentage increased from 20% to 93% of articles that use antibodies
451   (OR=45.2, 95%CI=5.5-INF; adjusted *p*=.003, one-tailed). In the journals *Neuroscience* and
452   *Science* this percentage also increased, but the increase was much smaller than in the journals
453   with guidelines. Overall, the percentage of articles reporting identification information on all
454   primary antibodies increased with 18 percentage points from 10% to 28% (OR=3.36,
455   95%CI=0.70-22.1; adjusted *p*=1, two-tailed) in journals without guidelines (*Neuroscience +
456   Science*). Apart from that, the overall reporting of identification information improved over time.
457   For all journals combined this percentage increased by 37 percentage points (OR=7.9,
458   95%CI=2.81-26.3; adjusted *p*=.0005, two-tailed). This suggests that other factors such as an
459   increase in awareness, the availability of identifying information, or a changing view on the

460 importance of antibody reporting by the scientific community might also play a role in how well
461 antibody identification information is reported.

462 Change in type of validation

463 If an article reported validation of at least one antibody, the types of validation that were carried
464 out in the article were further specified. An article can contain multiple types of antibody
465 validation, since one antibody can be validated in multiple ways, or multiple antibodies can be
466 validated by different methods. First, a distinction was made between validation by the authors
467 of the paper themselves and validation by means of a reference to the literature, to the
468 information of the supplier, or to validation information in a database. Validation by the authors
469 implies an experimental check in the laboratory and involves material verification, while the
470 other forms rely on documentation. Reference to the supplier implies a reliance on supplier
471 information about validity, reference to the literature implies a documentation of similar use of
472 these antibodies by other researchers, which is also possible via reference to a database.
473
474 Example of validation by the authors of an evaluated paper:
475
476   -   **Independent antibody strategies of five pillars.** *"Our immunostaining in the mouse*
477   *embryo cerebellum produced a similar labeling pattern as the other FoxP2 antibody"*
478   (Vibulyaseck et al. 2017).
479
480 Examples of validation by reference to a third party from the evaluated articles:
481
482   -   **Reference to supplier.** *"The rabbit polyclonal anti-FoxP2 antibody (AP5753b, Abgent,*
483   *San Diego, California, CA, USA) produced a single major band of 80 kDa in the mouse*
484   *heart tissue lysates in the manufacturer's western blot analysis."* (Vibulyaseck et al.
485   2017).
486   -   **Reference to the literature.** "*The goat polyclonal anti-EphA4 antibody (AF641, R&D*
487   *Systems, Minneapolis, MN, USA) recognized a single band of 110 kDa in HEK293 cell*
488   *lysate transfected with the EphA4 gene by western blot (Hashimoto et al., 2012). Its*
489   *immunohistochemical reactivity has been eliminated by preabsorption with the*
490   *recombinant EphA4 protein (Rosas et al., 2011)"* (Vibulyaseck et al. 2017).
491
492 Figure 7 and supplemental tables S8 and S9 show the changes in the prevalence of these types of
493 validation before and after the introduction of guidelines in articles that contain at least one
494 validated antibody. The percentage of articles that use validation by the authors seems extremely
495 high. It went from 87% to 85% in journals without guidelines and from 94% to 94% in journals
496 with guidelines between 2003 and 2017. However, these percentages might be misleading. These
497 high percentages were often caused by a small number of validated antibodies per article, which
498 is explained by our methodology of evaluating validation per article instead of per antibody.

499   Moreover, this method of validation is often implicitly carried out as a consequence of the
500   experimental method being used. Since the reliability of the rating of this category is
501   questionable ($\kappa = -0.20$), we did not perform further statistical analysis on this category.
502   Both the use of the validation methods 'reference to the literature' and 'reference to the
503   antibody supplier' increased in journals with guidelines (*JCN + Nature*) after the introduction of
504   these guidelines. These methods increased by 24 percentage points (OR=2.62, 95%CI=0.86-INF;
505   adjusted *p*=1, one-tailed) and 27 percentage points (OR=INF, 95%CI=1.96-INF; adjusted *p*=0.4,
506   one-tailed) respectively.
507   Before guidelines were installed, there was a small difference in the prevalence of these
508   validation types between journals with and without guidelines. This difference increased after the
509   introduction of guidelines. Before the introduction of guidelines, there was a difference of 13
510   percentage points (OR=1.99, 95%CI=0.36-14.4; adjusted *p*=1, two-tailed) in validation by
511   reference to the literature between journals with and without guidelines. After the introduction of
512   guidelines, this difference increased to 43 percentage points (OR=8.3, 95%CI=1.8-INF; adjusted
513   *p*=0.3, one-tailed). Similarly, there was no difference in the use of validation by reference to the
514   antibody supplier before the introduction of guidelines. However, after guidelines had been
515   installed, there was a difference of 27 percentage points (OR=INF, 95%CI=1.34-INF; adjusted
516   *p*=1, one-tailed) between journals with and without guidelines. It is notable that validation by
517   reference to information provided in an online antibody database is not used at all.

518   ## Change in type of validation carried out by authors of paper

519   If an antibody was experimentally validated by the article's authors, we further specified what
520   type of validation was carried out. Figure 8 and supplemental table S8 and S9 show an overview
521   of the change in reported validation types. Figure 8 contains methods that are mostly used to
522   validate antibodies explicitly. Other types of (often more implicit) validation, such as positive
523   and negative control, were also registered but not taken into account for comparisons because of
524   low interrater reliability.
525
526   Examples of validation from the evaluated articles:
527
528   -   **Pre-adsorption.** *"In control experiments, primary antibody was pre-adsorbed with rat*
529   *ANP (Bachem-Peninsula Laboratories; 1 µM overnight at 4 °C) prior to incubation with*
530   *tissue sections to assess the degree of non-specific staining. Under these conditions, low*
531   *levels of background staining were observed in cortex (see Results)"* (Wiggins et al.
532   2003).
533   -   **Spatial localization.** *"(...) a polyclonal antisera against rat ANP were used in an*
534   *attempt to visualize ANP-like-IR in the cerebral cortex and to determine the effect of CSD*
535   *on its level and cellular distribution. Consistent with previous reports, ANP-like-IR was*
536   *consistently detected in subcortical regions, with a high density of nerve-fibre staining*
537   *found throughout areas such as the bed nucleus of the stria terminalis (Fig. 4A, E), the*

538		*preoptic hypothalamus, areas of the amygdala and in the paraventricular thalamic*
539		*nucleus (see Kawata et al 1985, Skofitsch et al 1985)"* (Wiggins et al. 2003).
540	-	**Secondary antibody without primary.** *"Additional control incubations with primary*
541		*antibody omitted were routinely included and resulted in a lack of specific staining"*
542		(Wiggins et al. 2003).
543	-	**Molecular weight in western blot (and KO).** *"The rat monoclonal anti-Pcdh10*
544		*antibody (MABT20 clone 5G10, Millipore, Billerica, MA, USA) produced a single band*
545		*of 137 kDa in the P0 mouse cerebellum and whole brain in our western blot. The band*
546		*disappeared completely in samples from Pcdh10-KO mice (OL-KO homozygotes)"*
547		(Vibulyaseck et al. 2017).

549 Before the introduction of guidelines, all types of validation were used more often in journals
550 without guidelines. However, after the introduction of guidelines, the use of these validation
551 methods decreased in these journals without guidelines while there was a slight, but
552 nonsignificant, increase of their use in the journals with guidelines. This means that the
553 differences in the types of validation that were mostly used increased between journals with
554 guidelines and similar journals without guidelines. After the introduction of guidelines, there was
555 a difference of 10 percentage points in the use of pre-adsorption (OR=2.87, 95%CI=0.33-INF;
556 adjusted p=1, one-tailed), 23 percentage points in the use of spatial localization (OR=5.9,
557 95%CI=0.80-INF; adjusted p-value=1, one-tailed), 32 percentage points in the use of a
558 secondary antibody without a primary one (OR=8.6, 95%CI=1.22-INF; adjusted p=1, one-
559 tailed), and 10 percentage points in the use of molecular weight in western blot (OR=1.73,
560 95%CI=0.39-INF; adjusted p=1, one-tailed) between similar journals.

561   Interestingly, there is a difference in how often these validation types are used in the two
562 journals that introduced guidelines. In the journal *Nature*, the methods pre-adsorption and spatial
563 localization were not used at all before and after the introduction of guidelines. The small
564 increase in reporting of spatial localization is thus solely caused by its use in the journal *JCN*.

Discussion

567 These results suggest that journal guidelines might have some positive influences on antibody
568 validation reporting. For example, this is suggested by the large difference in the percentage of
569 articles reporting validation information of all primary antibodies between journals with and
570 without guidelines after the introduction of these guidelines. However, it is notable that in most
571 of our analyses a difference in reporting was already present between the four journals before the
572 introduction of guidelines. This could suggest that the journal itself (e.g. its editorial attention to
573 validation, or earlier publications about validation), regardless of the presence of reporting
574 guidelines, already has an influence on antibody validation reporting. Another explanation might
575 be that there is a difference in validation practices and journal selection between research areas
576 or groups.

577   Moreover, our results provide some insight in how journal guidelines might work. This
578   can be illustrated by comparing the increase in reporting of validation information with
579   identification information of all primary antibodies. On average, the reporting of identification
580   information improved by 58 percentage points, while the reporting of validation information
581   only improved by 23 percentage points in journals that introduced guidelines. This suggest that
582   the introduction of antibody reporting guidelines particularly affected aspects of antibody
583   reporting that are relatively easy to change. Adding identification information for each antibody
584   used in your publication, such as catalogue numbers, is easier than providing evidence of actual
585   validation. From an editorial perspective, the provision of identity information such as catalogue
586   numbers is also easier to check. At least for some of the articles, researchers seem to have opted
587   for the easier solutions to comply with antibody guidelines by providing readily available
588   information.
589   A similar indication for solutions that choose the 'path of least resistance' to comply with
590   guidelines, can be found in the use of third-party validation information. Once again, this
591   information is relatively easy to include in a paper with a simple reference and requires no extra
592   experimental work. Our results show that validation by reference to the literature increased by 24
593   percentage points in journals that introduced guidelines, while this percentage decreased in
594   journals without guidelines. Likewise, validation by reference to the antibody supplier's
595   information increased by 27 percentage points in journals with guidelines, while this method was
596   never used in journals without guidelines. These results are not significant, but the odds ratios
597   seem to suggest (especially in the case of reference to supplier's information) that these
598   validation practices merit further investigation. More research assessing how researchers regard
599   journal guidelines, how these guidelines affect their practices, and a replication of this study with
600   a larger sample size is recommended.
601   Because the guidelines mostly seem to influence aspects of antibody reporting that are
602   relatively easy to provide, we concluded that the effects of journal guidelines on antibody
603   reporting as limited. This limited effect of journal guidelines is consistent with findings in
604   similar research on journals' measures to improve reporting about research resources. A recent
605   evaluation of the 2013 *Nature* life science checklist shows that it improved the reporting of some
606   items, but also that approximately half of the compliant articles did not actually use the tools
607   recommended by the checklist (Han et al. 2017). Another evaluation of the *Nature* checklist
608   showed improvements in transparency of reported information, but much more modest
609   improvements in experimental design. Furthermore, the study found that the checklist did not
610   improve antibody reporting (NPQIPCollaborativegroup 2019). However, with regard to antibody
611   identifiability, our results are more optimistic than previous studies. Vasilevsky et al. found no
612   relationship between reporting guidelines and identifiability of resource materials in biomedical
613   research. Identifiability of resources was actually higher in journals with no or loose guidelines
614   than in journals with strict guidelines. Antibody identifiability in the *Journal of Comparative*
615   *Neurology* was only slightly higher than average across all journals (Vasilevsky et al. 2013).

616   While our results point towards some improvements, it is clear that journal guidelines are not a
617   definitive solution to fix validation and misidentification problems.
618   Much seems to depend on how journal guidelines are implemented. Mere symbolic
619   support for guidelines in editorial instructions may not be enough to improve reporting practices.
620   Baker et al. found that the endorsement of the ARRIVE (Animal Research: Reporting of In Vivo
621   Experiments) guidelines by journals had little to no impact the reporting of animal studies in
622   these journals (Baker et al. 2014). Hair et al. found that even requiring authors to complete an
623   ARRIVE checklist after submission of their manuscript did not improve compliance (Hair et al.
624   2019). In another study of journal guidelines regarding RRIDs, compliance was the lowest when
625   the journals only provided author instructions. Compliance became higher when authors received
626   mailed instructions, and it became very high when editorial staff assisted authors with their
627   RRIDs (Bandrowski et al. 2016). These results suggest that journal guidelines may have a
628   beneficial effect, but that they require additional measures to make them effective.
629   Another interesting observation from our study is that, in contrast to the reporting of
630   antibody validation, the reporting of information on the identity of all primary antibodies
631   increased in all four journals between 2003 and 2017. Moreover, this percentage increased
632   significantly by 37 percentage points in the total sample of all four journals combined. Although
633   the improvement was larger in journals that introduced guidelines, it is likely that this
634   improvement is at least partly caused by a growing awareness of antibody problems. That this
635   growing awareness only seems to have an effect on improving antibody identification reporting
636   and not on antibody validation reporting might be due to the current attention for reproducibility
637   in science. Although antibody validity is important for the overall validity of research, detailed
638   reporting on which antibodies were used is important to ensure reproducibility. However, as we
639   have argued, reproducible research is not necessarily valid research. We would therefore urge
640   not to forget the validity question in the replication crisis debate. Furthermore, it would be
641   interesting to study researchers' motivation for this change in reporting behaviour to see whether
642   and how it relates to the reproducibility crisis.

643   Limitations

644   Our study has several noteworthy limitations. As an exploratory study, our study was not pre-
645   registered, which we would recommend for possible future confirmatory replications of our
646   work. Furthermore, the labour-intensive nature of scrutinising publications for validation
647   information means our sample and interrater testing are limited. Together with a conservative
648   statistics test (Fisher's exact), this might have increased our Type II error. Having a low power
649   can have several consequences for the reliability of results, such as an overestimation of the
650   effect size. Readers of our study should be cautious of this limitation while interpreting our
651   results.
652   In addition, coding of articles inevitably implies some level of arbitrary convention and
653   interpretation, for example as we decided to code on the level of articles and not individual
654   antibodies. No distinction was made between implicit and explicit validation, giving the

655   researchers the benefit of the doubt with respect to their validation efforts, which may
656   overestimate the beneficial effect of guidelines and how often certain validation methods are
657   used. The room for interpretation about whether specific forms of experimental validation were
658   actually used, is also illustrated by the lower agreement between raters in this respect. This again
659   means that our results should be regarded with caution and a replication of this study with a
660   larger sample size and multiple assessors is recommended. Because of these limitations, we tried
661   to be as open as possible about the decisions we made during the coding process, by illustrating
662   these choices with examples from the evaluated articles in this manuscript, and by providing an
663   elaborate explanation of our choices in the coding table.
664   Last, our pseudo-experimental research design has some limitations. Because of the
665   substantial time period from which we took our samples (2003 and 2017), the time between the
666   intervention of implementing the guidelines and the measurement of changed behaviour is
667   stretched. We chose this approach because of the implementation of the guidelines by the two
668   journals at different times. However, another way to approach this could be by sampling over
669   time, for example by using interrupted time series analysis. Moreover, the quasi-experimental
670   design makes the effect of journal guideline introduction hard to isolate from wider influences,
671   such as a growing awareness of the urgency of antibody validation, or the growing knowledge
672   stored in databases such as the Antibodyregistry. Fortunately, our results do show signs of an
673   overall growing awareness of antibody problems, since the reporting of antibody identification
674   information significantly improved in the entire sample.
675

Conclusions

677   This study suggests that the implementation of antibody reporting guidelines by a journal might
678   have some positive influences on the reporting of antibody validation information, as we
679   hypothesized. The percentage of validated antibodies per article, the percentage of articles which
680   reported validation and identification of all primary antibodies, and different types of validation
681   all increased in journals with guidelines after the introduction of these guidelines. However,
682   improvements were particularly visible for forms of antibody information that are relatively easy
683   to provide, such as providing complete identification information. Strictly speaking, such
684   information supports the reliability of antibodies, and it is conceivable that securely identified
685   and previously validated antibodies can be used reliably in identical applications. Nevertheless,
686   improvements of more robust experimental validation were modest, at best. Combined with the
687   results of previous studies on journal guidelines, this suggests that the effect of journal antibody
688   guidelines by themselves may be limited and may require additional measures to ensure effective
689   implementation.
690   In light of the quasi-experimental and exploratory nature of our study, we need to be
691   careful with drawing these conclusions. Other factors, such as a general shift in research
692   practices, a change in publication behavior, a preference for publishing in a journal that adheres
693   to the values of a researcher, or changed editorial practices might also play a role. More research
694   would therefore be needed to study how researchers make decisions with regard to antibody

695 validation and how they adjust their research practices in response to changing journal
696 guidelines.
697

Acknowledgements

699 We would like to thank Jeffrey Stuart, Nicole Vasilevsky, Malcolm MacLeod and one
700 anonymous reviewer for their constructive feedback on our manuscript; Ger Pruijn, Serge
701 Horbach, Freek Oude Maatman, Merle-Marie Pittelkow, Jasmine Muradchanian and Ymkje
702 Anna de Vries for comments; and the Research Quality Team for inspiring discussions.

703

704

705

References

707 Abu-Remaileh M, Wyant GA, Kim C, Laqtom NN, Abbasi M, Chan SH, Freinkman E, and
708   Sabatini DM. 2017. Lysosomal metabolomics reveals V-ATPase-and mTOR-dependent
709   regulation of amino acid efflux from lysosomes. *Science* 358:807.
710 Alexander Steve PH, Roberts Richard E, Broughton Brad RS, Sobey Christopher G, George
711   Christopher H, Stanford SC, Cirino G, Docherty James R, Giembycz Mark A, Hoyer D,
712   Insel Paul A, Izzo Angelo A, Ji Y, MacEwan David J, Mangum J, Wonnacott S, and
713   Ahluwalia A. 2018. Goals and practicalities of immunoblotting and
714   immunohistochemistry: A guide for submission to the British Journal of Pharmacology.
715   *British Journal of Pharmacology* 175:407-411. DOI:10.1111/bph.14112
716 Andersson S, Sundberg M, Pristovsek N, Ibrahim A, Jonsson P, Katona B, Clausson C-M, Zieba
717   A, Ramström M, Söderberg O, Williams C, and Asplund A. 2017. Insufficient antibody
718   validation challenges oestrogen receptor beta research. *Nature Communications* 8:15840.
719   DOI: 10.1038/ncomms15840
720 Babic Z, Capes-Davis A, Martone ME, Bairoch A, Ozyurt IB, Gillespie TH, and Bandrowski
721   AE. 2019. Incidences of problematic cell lines are lower in papers that use RRIDs to
722   identify cell lines. *eLife* 8:e41676. DOI:10.7554/eLife.41676
723 Baker D, Lidster K, Sottomayor A, and Amor S. 2014. Two Years Later: Journals Are Not Yet
724   Enforcing the ARRIVE Guidelines on Reporting Standards for Pre-Clinical Animal
725   Studies. *Plos Biology* 12:e1001756. DOI: 10.1371/journal.pbio.1001756
726 Baker M. 2015a. Antibody anarchy: A call to order. *Nature* 527:545. DOI: 10.1038/527545a
727 Baker M. 2015b. Blame it on the antibodies. *Nature* 521:274.
728 Bandrowski A, Brush M, Grethe JS, Haendel MA, Kennedy DN, Hill S, Hof PR, Martone ME,
729   Pols M, Tan SC, Washington N, Zudilova-Seinstra E, Vasilevsky N, and Initiative RRI.
730   2016. The Resource Identification Initiative: a cultural shift in publishing. *Brain and
731   Behavior* 6:e00417. DOI:10.1002/brb3.417
732 Berglund L, Björling E, Oksvold P, Fagerberg L, Asplund A, Szigyarto CA-K, Persson A,
733   Ottosson J, Wernérus H, and Nilsson P. 2008. A genecentric Human Protein Atlas for
734   expression profiles based on antibodies. *Molecular & cellular proteomics* 7:2019-2027.
735 Bordeaux J, Welsh AW, Agarwal S, Killiam E, Baquero MT, Hanna JA, Anagnostou VK, and
736   Rimm DL. 2010. Antibody validation. *Biotechniques* 48:197.

737 Clair JA. 2015. Procedural Injustice in the System of Peer Review and Scientific Misconduct.
738   *Academy of Management Learning & Education* 14:159-172. DOI:
739   10.5465/amle.2013.0243
740 Collins H, M. 1985. *Changing order: Replication and induction in scientific practice*: University
741   of Chicago Press.
742 Crane D. 1967. The Gatekeepers of Science: Some Factors Affecting the Selection of Articles for
743   Scientific Journals. *The American Sociologist* 2:195-201.
744 Freedman LP, Cockburn IM, and Simcoe TS. 2015. The economics of reproducibility in
745   preclinical research. *Plos Biology* 13:e1002165.
746 Gamer M, Lemon J, Gamer MM, Robinson A, and Kendall's W. 2012. Package 'irr'. *Various
747   coefficients of interrater reliability and agreement*.
748 GBSI. 2016. GBSI workshop report: Antibody Validation: Standards, Policies, and Practices.
749   GBSI. Available at https://www.gbsi.org/publication/asilomar-antibody-workshop-report/
750 Gore AC. 2013. antibody validation requirements for articles published in endocrinology. Oxford
751   University Press.
752 Hair K, Macleod MR, Sena ES, Sena ES, Hair K, Macleod MR, Howells D, Bath P, Irvine C,
753   MacCallum C, Morrison G, Clark A, Alvino G, Dohm M, Liao J, Sena C, Moreland R,
754   Cramond F, Currie GL, Bahor Z, Grill P, Bannach-Brown A, Marcu D-C, Antar S,
755   Blazek K, Konold T, Dingwall M, Hohendorf V, Hosh M, Gerlei KZ, Wever KE, Jones
756   V, Quinn TJ, Karp NA, Freymann J, Shek A, Gregorc T, Rinaldi A, Jheeta P, Nazzal A,
757   Henshall DE, Storey J, Baginskaite J, de Oliveira CL, Laban K, Charbonney E, Lynn SA,
758   Cascella M, Wheater E, Baker D, Cheyne R, Christopher E, Roncon P, De-Souza EA,
759   Warda M, Corke S, Ammar Z, O'Connor L, Devonshire IM, McCann SK, Gray LJ,
760   Tanriver-Ayder E, and on behalf of the IC. 2019. A randomised controlled trial of an
761   Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *Research
762   integrity and peer review* 4:12. DOI: 10.1186/s41073-019-0069-3
763 Han S, Olonisakin TF, Pribis JP, Zupetic J, Yoon JH, Holleran KM, Jeong K, Shaikh N, Rubio
764   DM, and Lee JS. 2017. A checklist is associated with increased quality of reporting
765   preclinical biomedical research: A systematic review. *PLoS one* 12:e0183591. DOI:
766   10.1371/journal.pone.0183591
767 Hewitt SM, Baskin DG, Frevert CW, Stahl WL, and Rosa-Molinar E. 2014. Controls for
768   Immunohistochemistry: The Histochemical Society's Standards of Practice for Validation
769   of Immunohistochemical Assays. *Journal of Histochemistry and Cytochemistry* 62:693-
770   697. DOI: 10.1369/0022155414545224
771 Larivière V, Haustein S, and Mongeon P. 2015. The Oligopoly of Academic Publishers in the
772   Digital Era. *PLoS one* 10:e0127502. DOI: 10.1371/journal.pone.0127502
773 Leonelli S. 2018. Rethinking Reproducibility as a Criterion for Research Quality. Including a
774   Symposium on Mary Morgan: Curiosity, Imagination, and Surprise: Emerald Publishing
775   Limited. p 129-146.
776 Nature. 2013. Reducing our irreproducibility. *Nature* 496:398-398.
777 NPQIPCollaborativegroup. 2019. Did a change in Nature journals' editorial policy for life
778   sciences research improve reporting? *BMJ Open Science* 3:e000035. DOI:
779   10.1136/bmjos-2017-000035
780 O'hurley G, Sjöstedt E, Rahman A, Li B, Kampf C, Pontén F, Gallagher WM, and Lindskog C.
781   2014. Garbage in, garbage out: a critical evaluation of strategies used for validation of
782   immunohistochemical biomarkers. *Molecular oncology* 8:783-798.

783   Revelle W, and Revelle MW. 2015. Package 'psych'. *The Comprehensive R Archive Network*.

784   Rosner E, Rohmann KN, Bass AH, and Chagnaud BP. 2018. Inhibitory and modulatory inputs to
785   the vocal central pattern generator of a teleost fish. *Journal of Comparative Neurology*
786   526:1368-1388. DOI: 10.1002/cne.24411

787   Saper CB. 2005. An open letter to our readers on the use of antibodies. *The Journal of*
788   *Comparative Neurology* 493:477-478. DOI: 10.1002/cne.20839

789   Saper CB, and Sawchenko PE. 2003. Magic peptides, magic antibodies: guidelines for
790   appropriate controls for immunohistochemistry. *Journal of Comparative Neurology*
791   465:161-163.

792   Uhlen M, Bandrowski A, Carr S, Edwards A, Ellenberg J, Lundberg E, Rimm DL, Rodriguez H,
793   Hiltke T, Snyder M, and Yamamoto T. 2016. A proposal for validation of antibodies.
794   *Nature Methods* 13:823. DOI: 10.1038/nmeth.3995

795   Vasilevsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, LaRocca GM, and Haendel MA.
796   2013. On the reproducibility of science: unique identification of research resources in the
797   biomedical literature. *PeerJ* 1:e148.

798   Vibulyaseck S, Fujita H, Luo Y, Tran AK, Oh-Nishi A, Ono Y, Hirano S, and Sugihara I. 2017.
799   Spatial rearrangement of Purkinje cell subsets forms the transverse and longitudinal
800   compartmentalization in the mouse embryonic cerebellum. *Journal of Comparative*
801   *Neurology* 525:2971-2990. DOI: 10.1002/cne.24250

802   Wasserstein RL, Schirm AL, and Lazar NA. 2019. Moving to a world beyond "p< 0.05". Taylor
803   & Francis.

804   Wiggins AK, Shen PJ, and Gundlach AL. 2003. Atrial natriuretic peptide expression is increased
805   in rat cerebral cortex following spreading depression: possible contribution to sd-induced
806   neuroprotection. *Neuroscience* 118:715-726. DOI: 10.1016/S0306-4522(03)00006-X

807

Interrater reliability of article coding process

Question	Percentage agreement	Cohen's kappa	95% CI	N
Basic information all primary antibodies complete?	100%	1.00	1.00-1.00	16
Any validation information present (at least one antibody)?	75%	0.50	0.13-0.87	16
All primary antibodies validated?	94%	IND	/	16
If validated:				
Reference to validation by antibody supplier?	100%	IND	/	8
Reference to validation in the literature?	88%	0.75	0.31-1.00	8
Reference to validation information in database?	100%	IND	/	8
Antibody validation carried out by authors of article?	63%	-0.20	-0.49-0.088	8
If carried out by authors, which method?				
Molecular weight similar to target (in WB)?	83%	0.57	-0.12-1.00	6
Spatial localization?	100%	1.00	1.00-1.00	6
Pre-adsorption/blocking peptide?	100%	1.00	1.00-1.00	6
Using secondary antibody without primary?	100%	IND	/	6
One of the five pillars?	100%	1.00	1.00-1.00	6
Positive control?	67%	0.33	-0.23-0.90	6
Negative control?	67%	IND	/	6
Other validation method?	50%	0.00	-0.75-0.75	6

Question		ICC	95% CI	N
How many antibodies were used?		0.995	0.986-0.998	15
How many antibodies were validated?		0.684	0.263-0.886	14
Percentage of validated antibodies?		0.663	0.226-0.878	14

Figure 1

Overview of the research strategy.

Samples of 15 articles were taken from journals that did (1 and 2) and that did not (3 and 4) introduce antibody reporting guidelines. These samples were taken from a timeframe before (August 2002 - August 2003) and after (June 2017 - June 2018) the introduction of guidelines by journal 1 and 2. An increase in antibody validation reporting was expected between the samples connected with a black arrow, while no difference was expected between the samples connected with a dashed grey arrow.
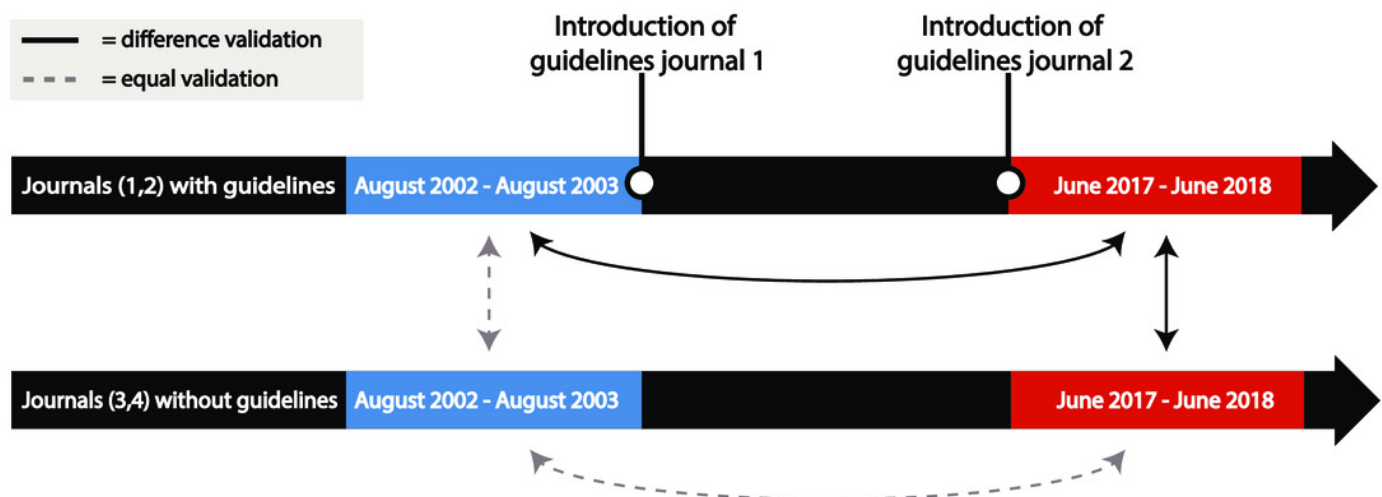
Figure 2

Example of explicit antibody validation.

To test the specificity of the antibody, a comparison was made between the western blots provided by the supplier of the antibody and western blots carried out by the authors of the paper. According to the supplier, two bands should be observed: one between 25 and 37 kDa and one between 50 and 75 kDa. According to the authors, the additional band at >250 kDa is likely caused by six connexin protein subunits that form a connexin hemi-channel. The authors explicitly mention antibody validation as the purpose of this experiment: *"We tested the antibody's specificity using a western blot of Gulf toadfish whole brain homogenates."* Figure from Rosner et al. 2018.

250 kDa –

75 kDa –
50 kDa –

37 kDa –

25 kDa –

Figure 3

Example of implicit antibody validation.

Atg7 defective cells were generated by inhibiting gene expression of the Atg7 gene with RNA interference. The cells in which Atg7 was silenced show no staining with anti-ATG7, while wild type cells do. This type of validation was marked as validation by negative control (RNA interference) and by genetic strategies of the five pillars. Figure from Abu-Remaileh ea 2017. Lysosomal metabolomics reveals V-ATPase-and mTOR-dependent regulation of amino acid efflux from lysosomes. *Science* 358:807. Reprinted with permission from AAAS.
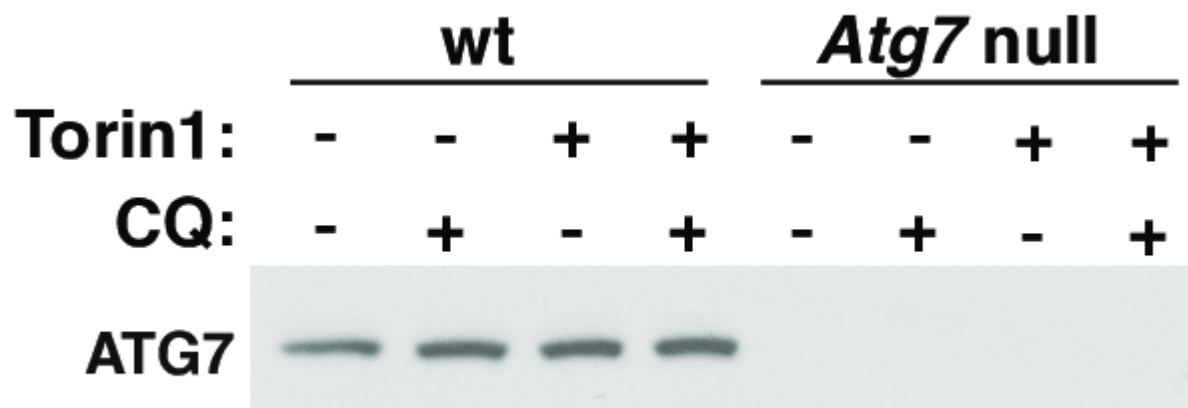
Figure 4

Change in percentage of validated antibodies per article before (2003) and after (2017)
two of the journals (*JCN* and *Nature*) implemented antibody reporting guidelines.

Grey dots represent the percentage of validated antibodies per article, red bars indicate the
mean percentage of validated antibodies per article for each journal. Sample size: *JCN* 2003
n=15, 2017 n=14; *Neuroscience* 2003 n=14, 2017 n=15; *Nature* 2003 n=15, 2017 n=14;
Science 2003 n=15, 2017 n=15.
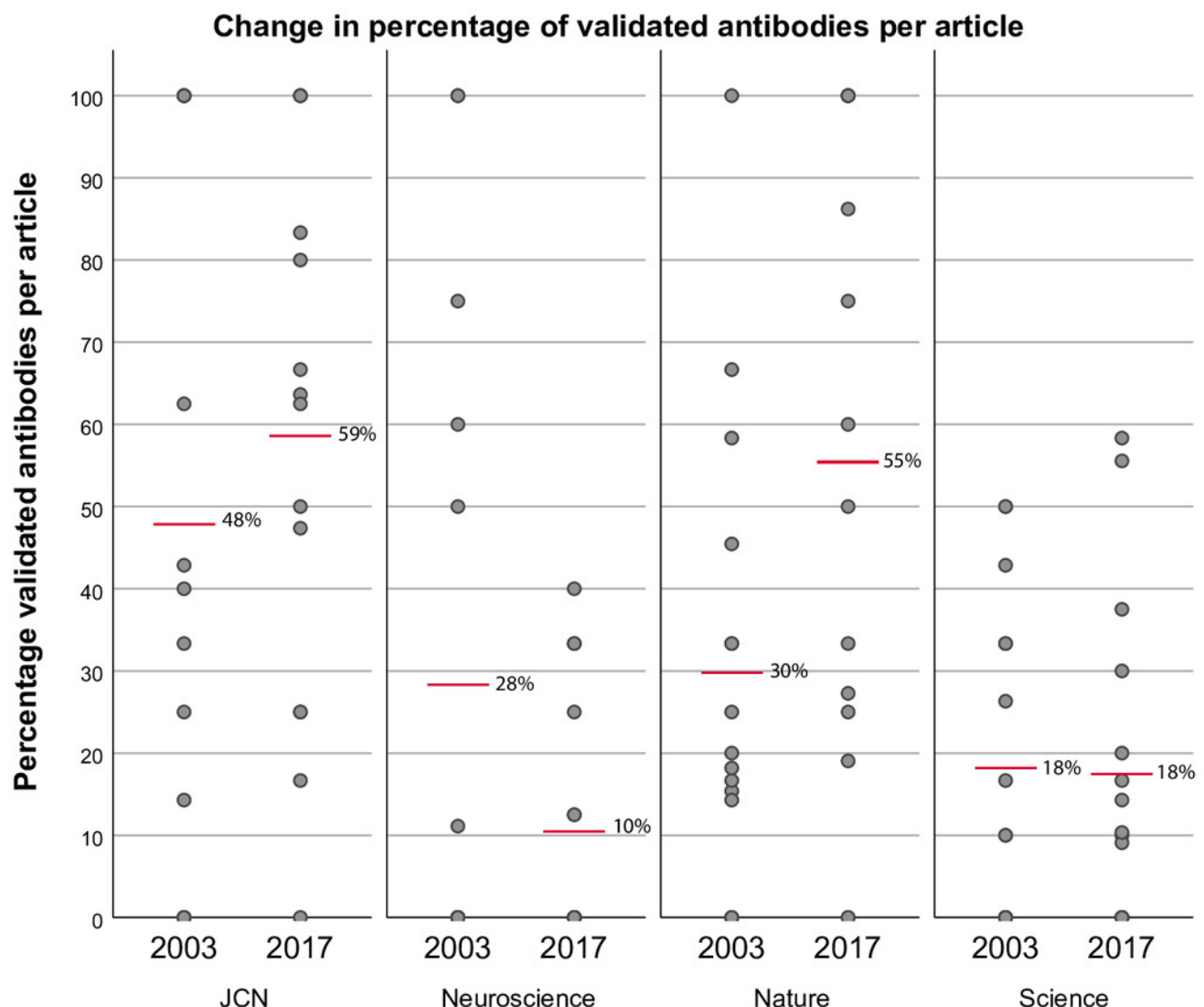
Figure 5

Change in percentage of articles that reported validation information on all primary antibodies used in the study before (2003) and after (2017) two of the journals (*JCN* and *Nature*) implemented antibody reporting guidelines.

Sample size: with guidelines 2003 n=30, 2017 n=28; without guidelines 2003 n=30, 2017 n=30, total 2003 n=60, 2017 n=58, *JCN* 2003 n=15, 2017 n=14; *Neuroscience* 2003 n=15, 2017 n=15; *Nature* 2003 n=15, 2017 n=14; *Science* 2003 n=15, 2017 n=15. *P*-values (one-tailed) were determined with Fisher's exact test and adjusted using the Holm-Bonferroni correction.
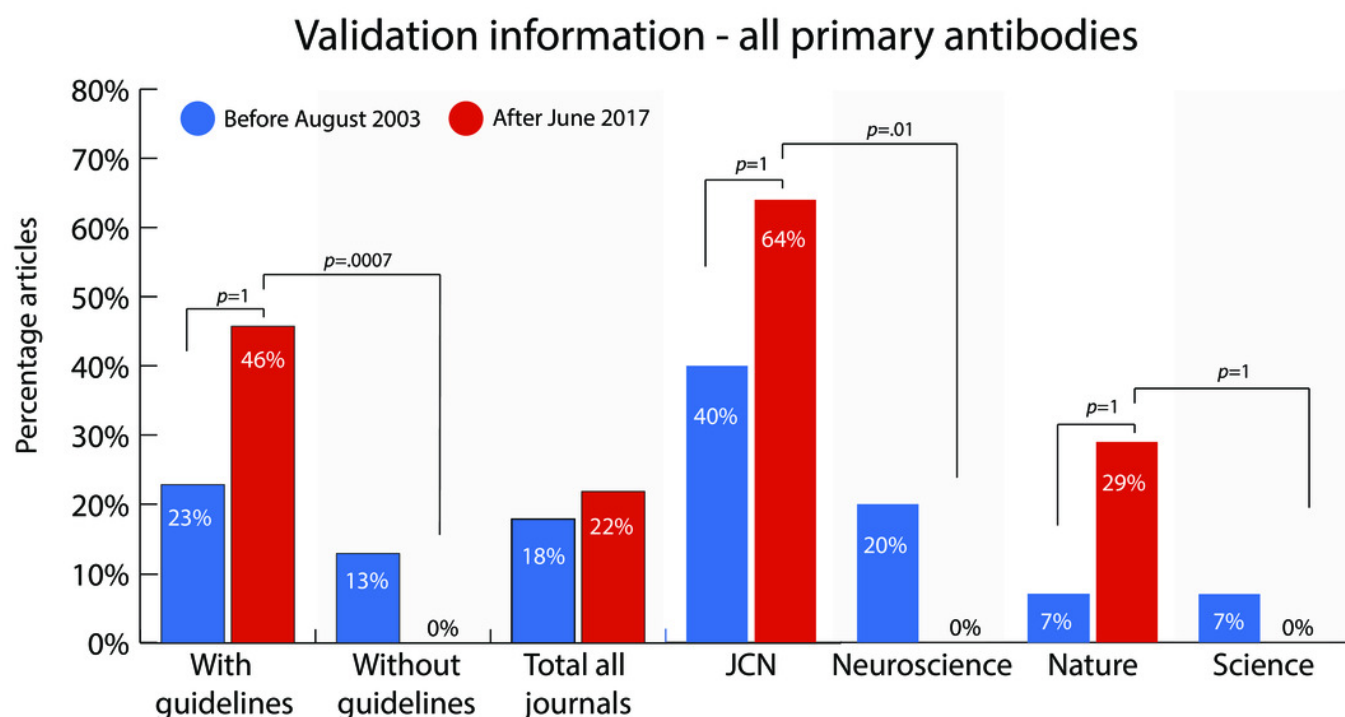
Figure 6

Change in percentage of articles that reported identification information on all primary antibodies used in the study before (2003) and after (2017) two of the journals (*JCN* and *Nature*) implemented antibody reporting guidelines.

Sample size: with guidelines 2003 n=30, 2017 n=28; without guidelines 2003 n=30, 2017 n=29; total all journals 2003 n=60, 2017 n=57; *JCN* 2003 n=15, 2017 n=15; *Neuroscience* 2003 n=15, 2017 n=15; *Nature* 2003 n=15, 2017 n=13; *Science* 2003 n=15, 2017 n=14. *P*-values were determined with Fisher's exact test and adjusted using the Holm-Bonferroni correction. [1] = two-tailed *p*-values, other *p*-values are one-tailed.
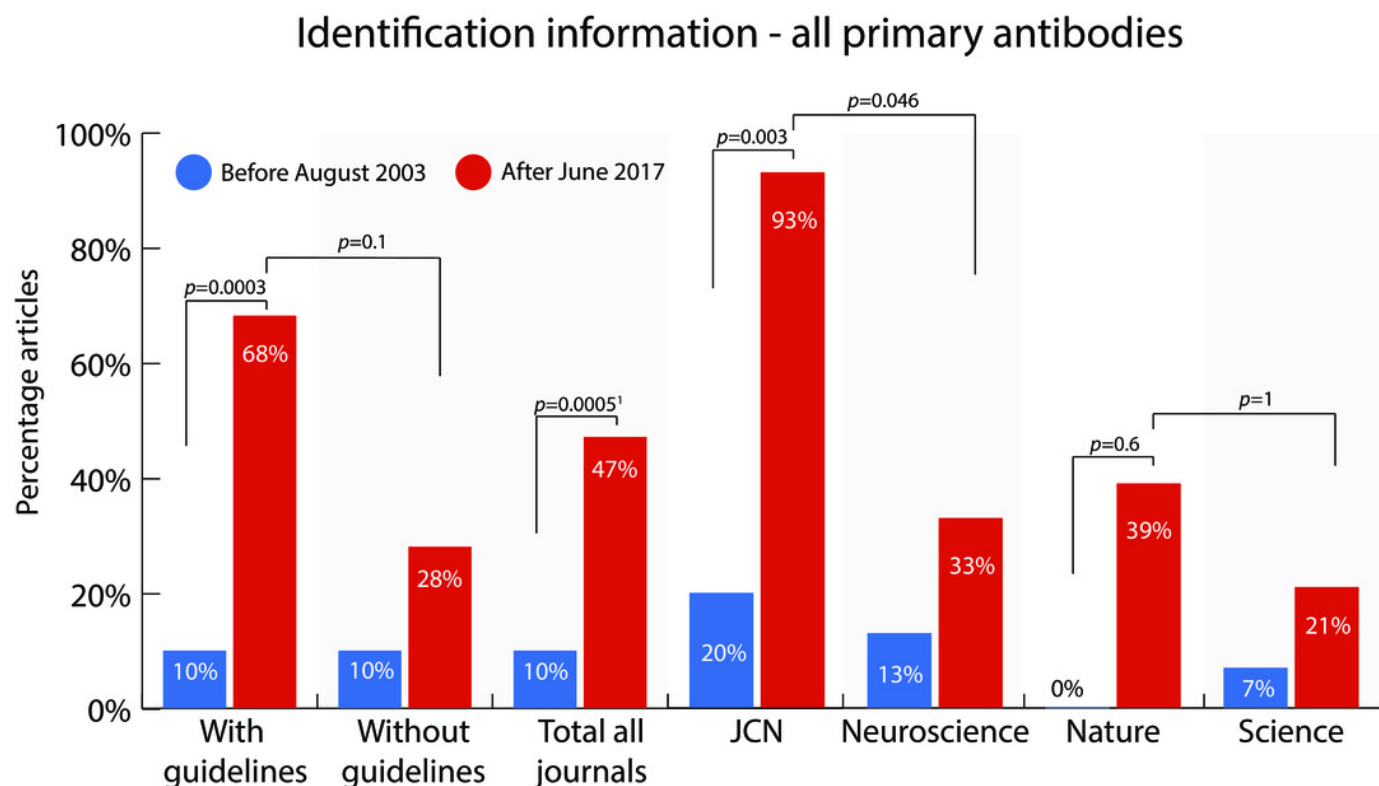
Figure 7

Change in percentage of validated articles reporting different types of validation before (2003) and after (2017) the journals with guidelines implemented antibody reporting guidelines.

Sample size (identical for all types of validation unless stated otherwise): with guidelines 2003 n=23, 2017 n=27; without guidelines 2003 n=16, 2017 n=16; exceptions: reference validation literature with guidelines 2003 n=22; reference validation supplier with guidelines 2017 n=26. P-values (one-tailed) were determined with Fisher's exact test and adjusted using the Holm-Bonferroni correction.
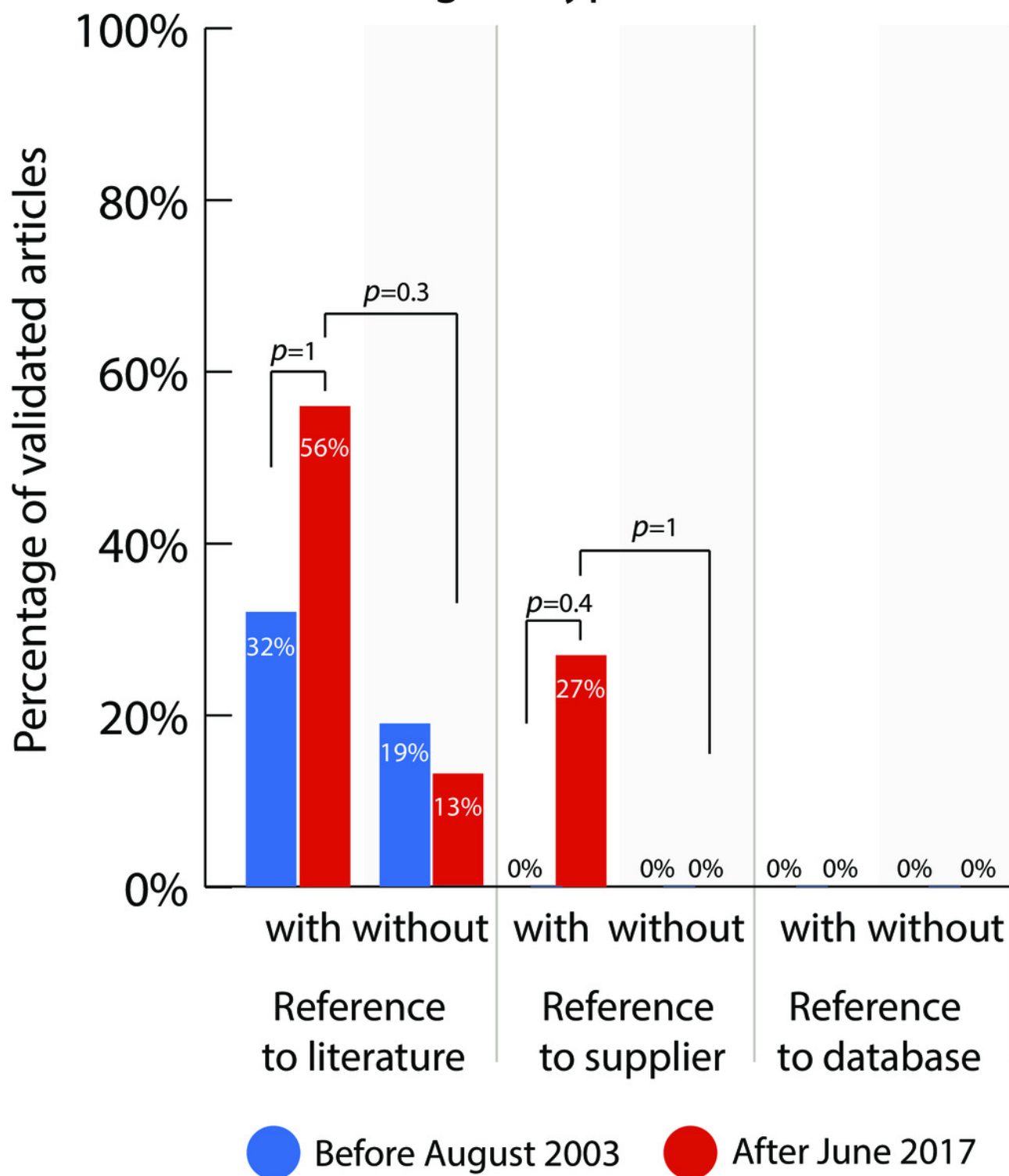
Figure 8

Change in percentage of validated articles reporting different types of validation before (2003) and after (2017) the journals with guidelines implemented antibody reporting guidelines.

Sample size (identical for all types of validation): with guidelines 2003 n=20, 2017 n=23; without guidelines 2003 n=15, 2017 n=15. *P*-values were determined with Fisher's exact test and adjusted using the Holm-Bonferroni correction. All adjusted p-values are 1.

Change in type of validation