University of Groningen

The Role of Human Fallibility in Psychological Research

Kovacs, Marton; Hoekstra, Rink; Aczel, Balazs

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Empirical Article

The Role of Human Fallibility in Psychological Research: A Survey of Mistakes in Data Management

Marton Kovacs[1,2] , **Rink Hoekstra**[3] , and
Balazs Aczel[2]
[1]Doctoral School of Psychology, ELTE Eotvos Lorand University, Budapest, Hungary; [2]Institute of
Psychology, ELTE Eotvos Lorand University, Budapest, Hungary; and [3]GION Education/Research,
University of Groningen, Groningen, the Netherlands

Abstract

Errors are an inevitable consequence of human fallibility, and researchers are no exception. Most researchers can recall major frustrations or serious time delays due to human errors while collecting, analyzing, or reporting data. The present study is an exploration of mistakes made during the data-management process in psychological research. We surveyed 488 researchers regarding the type, frequency, seriousness, and outcome of mistakes that have occurred in their research team during the last 5 years. The majority of respondents suggested that mistakes occurred with very low or low frequency. Most respondents reported that the most frequent mistakes led to insignificant or minor consequences, such as time loss or frustration. The most serious mistakes caused insignificant or minor consequences for about a third of respondents, moderate consequences for almost half of respondents, and major or extreme consequences for about one fifth of respondents. The most frequently reported types of mistakes were ambiguous naming/defining of data, version control error, and wrong data processing/analysis. Most mistakes were reportedly due to poor project preparation or management and/or personal difficulties (physical or cognitive constraints). With these initial exploratory findings, we do not aim to provide a description representative for psychological scientists but, rather, to lay the groundwork for a systematic investigation of human fallibility in research data management and the development of solutions to reduce errors and mitigate their impact.

Everybody makes mistakes, and scientists are no exception. The research process is a highly complex affair involving a variety of self-taught, unsupervised, and ad hoc manual procedures that are vulnerable to human error. Such errors include accidentally overwriting data, analyzing the wrong data set, misapplying a randomization procedure, mislabeling experimental conditions, or copying and pasting the wrong test statistics. When errors are discovered, it is common to blame the researcher, but some errors should be expected as an inevitable consequence of human fallibility (Hardwicke et al., 2014).

The field of psychology is currently immersed in a self-reflective era during which the credibility of the literature has come under serious scrutiny (Nelson et al., 2018; Vazire, 2018). Much attention in this discussion has been paid to the impact of existing methodological and statistical practices that have been identified as threats to the validity of scientific claims and the efficiency of

Corresponding Author:
Balazs Aczel, Institute of Psychology, ELTE Eotvos Lorand University
Email: aczel.balazs@ppk.elte.hu

Table 1. Links to All External Materials Related to the Study

Name	Link
The main survey exported from Qualtrics	https://osf.io/67dfz/
Preregistration of the primary study	https://osf.io/myu3v
OSF repository of the project	https://osf.io/fg7yb/
Definitions of the groups	https://github.com/marton-balazs-kovacs/ researchers_mistake_script/tree/master/Data/ Processed/grouping/definition
Examples of data-management mistakes shown during the validation of the grouping process	https://osf.io/3sf9j/
Instructions for the raters during validation	https://osf.io/awr6s/
Link to the preprint	https://psyarxiv.com/xcykz/

knowledge accumulation (John et al., 2012; Simmons et al., 2011). The impact of basic human error, however, has received relatively sparse attention, and existing evidence is limited to specific circumstances. For example, reviewing published studies, Rosenthal (1978) found that researcher observations of participants were occasionally miscoded. In a more recent study, Nuijten et al. (2016) performed an automated assessment of thousands of psychology articles and observed at least one statistical reporting inconsistency in half of them. Finally, Hardwicke et al. (2018) attempted to directly reproduce target values reported in 35 psychology articles by repeating the original analyses. Twenty-four of these articles contained at least one value that could not be reproduced within a 10% margin of error. Although these studies highlight the role of human error in specific circumstances, what is missing is a systematic assessment of the nature, frequency, and severity of data-management mistakes in psychology. A detailed characterization of data-management mistakes may help with the identification and dissemination of solutions that are most needed to improve this aspect of psychological research.

The goal of the present survey is to start the exploration of the role of human error in the management of psychological data. *Research data management* is an umbrella term concerning all stages of a research project that have an effect on the data. This is the definition we use throughout the article. These stages typically consist of many manual procedures, which makes them especially vulnerable to human error. We aimed to survey researchers from the field of psychology and ask them to describe and rate mistakes that they encountered in their own research. Given the sparsity of research in this topic and the nonrepresentativeness of our sample, our goal was explicitly exploratory and descriptive.

Disclosures

Preregistration

This was an exploratory study, and it was not our intention to test any hypotheses. Nevertheless, we preregistered a

study protocol (https://osf.io/myu3v) outlining our rationale, methods, and analysis plan to make clear which aspects of the study were preplanned and which were developed during or after data collection.

In the preregistration, we proposed to group the collected mistakes into traditional data-management stages, but we have used an updated version of the data-management stages that we think is more nuanced (see Fig. S1 in the Supplemental Material available online). The data preprocessing procedures and the validation of the grouping process (see Method section) were not preregistered. We are not aware of any other deviations from the preregistered protocol.

Data, materials, and online resources

All data, materials, and R code for the analyses and figures can be accessed at the project's OSF page: https://osf.io/fg7yb/. A list of links to specific external materials can be found in Table 1.

Reporting

We report the rationale for our sample size, all data exclusions, all manipulations, and all measures conducted during the study.

Ethical approval

Ethical permission was provided by Eotvos Lorand University Faculty of Education and Psychology Ethical board in Hungary. We collected no identifying information from the respondents. This study was conducted in accordance with the Declaration of Helsinki.

Method

Sample

We contacted 16,412 corresponding authors of articles published between 2010 and 2018 in a journal that had *psychology* among its labels in the ScienceDirect

Table 2. List of Questions From the Survey About the Research Data-Management Mistakes

Property of the mistake	Question	Variable type
Mistakes in general	The frequency of mistakes in general	Likert-type scale
Most frequent mistake	Description of the mistake	Free text
	Cause of the mistake	Free text
	Outcome of the mistake	Multiple choice with free-text option
	Frequency of the mistake	Likert-type scale
	Seriousness of the mistake	Likert-type scale
Most serious mistake	Description of the mistake	Free text
	Cause of the mistake	Free text
	Outcome of the mistake	Multiple choice with free-text option
	Seriousness of the mistake	Likert-type scale

database. Participation was voluntary and anonymous. To encourage participation, we offered to support the Center for Open Science with $0.20 for each completed survey. For the detailed description of the email address collection method and the recruitment, see the Supplemental Material.

Materials

We developed a questionnaire (summarized in Table 2) and corresponding scales (see Tables 3 and 4) for the exploration of the mistakes made during the data-management process in psychological research (available at https://osf.io/67dfz/).

In this questionnaire, we first aimed to measure how often researchers commit data-management mistakes in general. Therefore, we asked them how frequently they believe any kind of data-management mistake happens in their research team; responses were on a 5-point Likert-type scale from *very low* to *very high* frequency (Table 3). Next, we asked the respondents to specify the most frequent mistake that happened in their research team during the last 5 years and how frequently that mistake occurs (on the same frequency scale as above), how serious they think the outcome of that mistake was (on a 5-point Likert-type scale ranging from *insignificant* to *extreme* severity; see Table 4), the cause of that mistake (free-text response), and what negative outcome occurred (select one from financial loss, erroneous conclusion, time loss, inefficiency, frustration, and other, please specify).

We also asked researchers to write down the most serious mistake that happened in their research team during the last 5 years, how serious they think the outcome of that mistake was (on the same seriousness scale as above), the cause of that mistake (free-text response), and what negative outcome occurred (select one from financial loss, erroneous conclusion, time loss, inefficiency, frustration, and other, please specify).

Finally, as background information questions, we asked respondents to specify their research field (they could choose one from the following options: social psychology, applied psychology, personality psychology, clinical psychology, developmental and educational psychology, experimental and cognitive psychology, neurophysiology and physiological psychology, methodology and statistics, or other) and the number of years they have worked in that field.

Procedure

The participants received the Qualtrics survey link in an email (available at https://osf.io/67dfz/). All questions were optional except the background-information questions. The topic of the survey was introduced by eight brief examples of research data-management mistakes (partially sourced from a pilot study; see the Supplemental Material). The completion of the survey took a median of 6 min.

Number of responses

Out of the 16,412 sent emails, 14,033 were delivered, and the remaining 2,379 bounced. All in all, 779 researchers (response rate = 5%) started our survey, out of which we excluded 19 respondents who did not accept the informed consent form and 271 respondents who did not answer any of the questions listed in Table 2. We also excluded one respondent who did not answer any of the compulsory questions regarding background information. The survey software and

Table 3. Frequency Scale for Research Data-Management Mistakes

Frequency level	Description
Very low	Occurs never or rarely
Low	Occurs in some of the projects
Moderate	Occurs in half of the projects
High	Occurs in most of the projects
Very high	Occurs in (almost) all of the projects

Table 4. Seriousness Scale for Research Data-Management Mistakes

Seriousness level	Possible consequences	Example
Insignificant	Minutes of time loss, insignificant financial loss, no effect on conclusions	Occasional typos in the variable names
Minor	Some project delay and/or money loss, short-timed frustration, no effect on conclusions	Have to rerun the analysis
Moderate	Definite time and/or money loss, mild frustration, potential effect on some conclusions	Have to record part of the whole sample again
Major	Great project delay and/or money loss, affecting some conclusions of the article, a considerable level of frustration	Have to redo the whole data collection
Extreme	Project failure, serious time and/or money loss, strongly affecting the central conclusion of the article, damaged professional reputation	Article withdrawal

personal correspondence indicated that some respondents redistributed the survey link among their colleagues. Because the respondents who answered the forwarded survey also indicated their field of research and the years they have spent on the field, we decided to keep their responses (24 respondents after exclusions). Ultimately, the data of 488 respondents remained for further analysis.

Data preprocessing

The data-preprocessing pipeline was considerably different for the investigation of the frequency and seriousness ratings of the mistakes and for the free-text responses (description of the mistakes, their causes, and their outcomes). Thus, we describe the preprocessing of the ratings and the free-text responses separately.
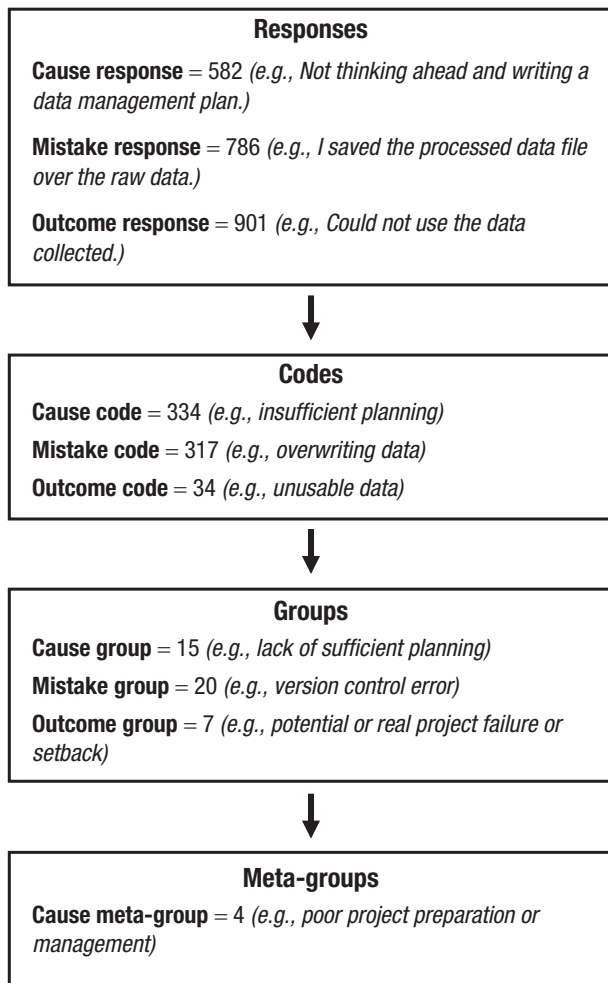
Preprocessing of the frequency and seriousness ratings. For the preprocessing of the frequency and seriousness ratings, we used the data from the remaining 488 respondents after the initial exclusions. The respondents had to provide a frequency rating and seriousness rating for the most frequent mistake and a seriousness rating for

Table 5. Number of Mistakes Descriptions Excluded for Each Exclusion Criteria

Exclusion criteria	Property of the mistake	Number of mistakes excluded for the property
Ambiguous	Most frequent	0
	Most serious	40
Irrelevant content	Most frequent	22
	Most serious	62
Out of time frame	Most frequent	8
	Most serious	5

the most serious mistake that they described (for the description of the rating scales, see Tables 3 and 4). In some cases, the respondents described more than one mistake in their free-text response, but we included only one rating per question per respondent. Only those frequency and seriousness ratings in which the corresponding description of the mistake passed the following exclusion procedures were included in our analyses. First, because describing a mistake was not compulsory, we worked with the description of 449 most frequent mistakes and the description of 404 most serious mistakes after excluding the missing responses. Second, we excluded responses in which the description of the mistake provided by the participant was ambiguous (e.g., respondent wrote "see above," and it was not clear which answer they were referring to) or irrelevant to the given question or the researcher stated that the mistake occurred before the 5-year time frame we were interested in. After this exclusion, we were left with the descriptions of 419 most frequent mistakes and 297 most serious mistakes. Table 5 contains the number of mistake descriptions that we excluded in this step for each exclusion criteria. Finally, because providing a rating for the described mistakes was also not compulsory, one seriousness rating was not reported for a description of a most serious mistake. Therefore, it is missing from the analyses. At the end of the data preprocessing, we were left with 419 frequency and seriousness ratings of the most frequent mistakes and 296 seriousness ratings of the most serious mistakes. These ratings were provided by 426 respondents. For the overall frequency of mistakes in the team question, we had 486 responses left after excluding two missing responses.

Preprocessing of the free-text responses. To analyze the free-text responses describing the research data-management mistakes, their causes, and their outcomes, we categorized them into groups according to similarity

Responses

Cause response = 582 *(e.g., Not thinking ahead and writing a data management plan.)*

Mistake response = 786 *(e.g., I saved the processed data file over the raw data.)*

Outcome response = 901 *(e.g., Could not use the data collected.)*

↓

Codes

Cause code = 334 *(e.g., insufficient planning)*

Mistake code = 317 *(e.g., overwriting data)*

Outcome code = 34 *(e.g., unusable data)*

↓

Groups

Cause group = 15 *(e.g., lack of sufficient planning)*

Mistake group = 20 *(e.g., version control error)*

Outcome group = 7 *(e.g., potential or real project failure or setback)*

↓

Meta-groups

Cause meta-group = 4 *(e.g., poor project preparation or management)*

Fig. 1. Flowchart illustrating the categorization of free-text responses into groups. The number of responses indicate their counts after both the separation of the responses and the exclusions. Here, we report only the final number of items for each level of grouping. Illustrative examples are shown as italicized text in parentheses.

by using thematic analysis (Braun & Clarke, 2006), a qualitative method that helps identify and highlight central features in texts (for a summary, see Fig. 1). The grouping process was carried out by two team members (B. Aczel and M. Kovacs), and all disagreements were resolved by discussion. Below, we describe creation of the groups in detail.

Preparing data for the grouping process. For the preprocessing of the free-text responses, we started the process with responses from 488 respondents. Respondents were asked to describe their most frequent and most serious mistakes and their causes and outcomes in a free-text response (see Table 2). For the outcomes of the mistakes, we provided a list of options with the possibility of writing a free-text response if none of the provided options were applicable. However, we applied the same data-preprocessing method to the outcomes of the mistakes as

to the descriptions of the mistakes and their causes for the sake of simplicity. The preprocessing methodology was applied separately to the descriptions of mistakes, causes, and outcomes. Answering these questions was not compulsory, so there were missing responses. Moreover, as mentioned in the Preprocessing of the Frequency and Seriousness Ratings section, we excluded the responses in which the description of the mistake was ambiguous or irrelevant to the given question or the researcher stated that the mistake occurred before the prescribed time frame (i.e., past 5 years). We applied the same exclusion criteria to the descriptions of the causes and the outcomes as well. When the respondents provided more than one description of a cause, a mistake, or an outcome in their free-text response, we treated each response separately in the grouping process. Thus, after the initial exclusions and the separation of the responses, we had 931 descriptions of causes, 835 descriptions of mistakes, and 920 descriptions of outcomes. Figure 2 shows the number of responses left after each stage of the grouping process broken down by the aspects of a research data-management mistake (cause of the mistake, the mistake itself, outcome of the mistake) and property of the mistake (most frequent mistake, most serious mistake). Furthermore, we excluded additional responses as explained in the Coding Process and the Grouping Process sections.

Creating codes. As the first step of the grouping process, we summarized each response by a short plain-text code in a systematic way. Each code highlighted a central feature of the given answer. We excluded all responses from further steps of the thematic grouping if we did not find the text to contain sufficient information regarding the given survey question. At the end of the coding process, we had 317 different codes for the descriptions of the mistakes, 334 for the causes of the mistakes, and 34 for the outcomes of the mistakes.

Creating groups. As the second step, we categorized the codes into higher level groups. A group describes the essence of a collection of codes. Each time a code did not fit any of the existing groups, we created a new group according to the given code. At this stage, we excluded those responses of which the codes could not be categorized into any of the groups because the code did not contain sufficient or relevant information. Using the codes, we identified 20 different groups of mistake types, 15 groups of causes of mistakes, and seven groups of outcomes of mistakes. Following this, we created a definition for each group by listing the codes that were assigned to that group. Finally, each free-text response inherited the group label assigned to its code. At the end of the thematic grouping process, there were 786 descriptions of mistakes, 582 causes of mistakes, and 901 outcomes of mistakes assigned to groups.

	Cause	Mistake description	Outcome
Excluding missing responses	430	449	462
	365	*404*	*380*
Excluding irrelevant, ambiguous, and out of timeframe responses	418	419	457
	302	*297*	*368*
Separation	549	529	507
	382	*306*	*413*
Exclusion in creating codes	517	514	504
	355	*291*	*411*
Exclusion in creating groups	356	506	496
	226	*280*	*405*

Fig. 2. Flowchart illustrating the number of responses broken down by aspects of the mistake and property (most frequent and most serious) after each preprocessing stage of the free-text responses. The number of responses for the most frequent mistakes are written in the upper row, and the number of responses for the most serious mistakes are written in italics in the lower row.

Creating metagroups. As the third step, we created four metagroups to decrease the number of groups for the causes of mistakes to ease comprehension and aid visualization. The creation of the metagroups was carried out through a discussion in a nonsystematic way. The four metagroups were created according to overlapping themes between the groups. Table 6 shows which cause groups were assigned to which metagroups.

Table 6. Metagroups for Mistake Causes

| Metagroup	Cause group
Poor project preparation or management	Bad or lack of planning, bad or lack of standards, bad skill management, miscommunication, failure to automate an error prone task, time-management issue
External difficulties	High task complexity, technical issues
Lack of knowledge	Lack of knowledge/experience
Personal difficulties	Carelessness, inattention, lack of control, overconfidence, physical or cognitive constraints

Results

Background information

Among the 488 respondents, the three most commonly identified psychology fields were experimental and cognitive psychology ($N = 88$), social psychology ($N = 62$), clinical psychology, and developmental and educational psychology ($N = 45$ for both), although the largest group ($N = 116$) of the respondents could not associate themselves with any of the listed research fields. The median time spent in their field was 15 years (interquartile range = 15). For a summary of the respondents' research fields and the distribution of their years spent in their field, see the Supplemental Material.

General overview of data-management mistakes

To obtain a general overview of data-management mistakes, we investigated the overall frequency of mistakes, the frequency and seriousness of the most frequent mistakes, and the seriousness of the most serious mistakes
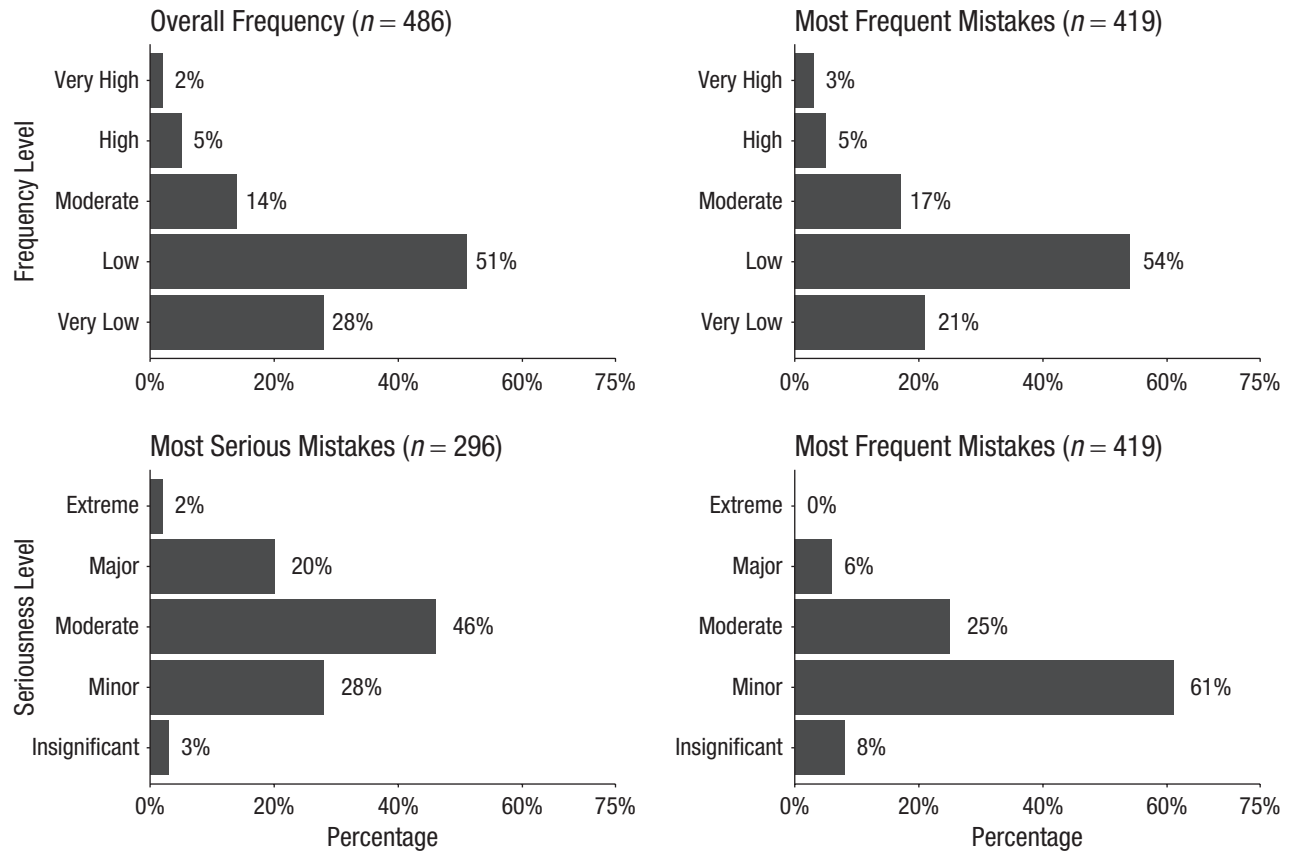
Fig. 3. Distribution of all responses presented in the General Overview of Data-Management Mistakes section. Each plot shows the percentages on the *x*-axis, and the levels of either the frequency scale (see Table 2) or the seriousness scale (see Table 3) are shown on the *y*-axis. Percentages may not sum to 100 because of rounding. For the counts behind these percentages, see Figure S4 in the Supplemental Material available online.

(for the questions, see Table 2). All the results for this section are shown in Figure 3, and the text below provides a summary of the results.

The overall frequency of mistakes. Responses suggested that the overall occurrence of mistakes was infrequent; 79% (384 of 486) of respondents reported that mistakes occurred with very low or low frequency, whereas for 21% (*n* = 102) of the remaining respondents, mistakes had moderate, high, or very high frequency.

The most frequent mistakes. When researchers were asked how frequently the most frequent mistake happened in their research team, 75% (314 of 419) of them indicated that it had low or very low frequency, whereas for the remaining 25% (*n* = 105) of the teams, the most frequent mistake had moderate, high, or very high frequency.

The most frequent mistakes reportedly led to insignificant or minor consequences (e.g., minutes of time loss, insignificant financial loss, no effect on conclusions) for 69% (289 of 419) of respondents, moderate consequences for 25% (*n* = 104) of respondents, and

major or extreme consequences for the remaining 6% (*n* = 26) of respondents.

The most serious mistakes. When asked about the most serious data-management mistake that occurred in their team during the last 5 years, 31% (93 of 296) of respondents reported that the mistake led to insignificant or minor consequences (e.g., minutes of time loss, insignificant financial loss, no effect on conclusions), 46% (*n* = 137) reported that the mistake led to moderate consequences, and the remaining 22% (*n* = 66) reported that the mistake led to major or extreme consequences.

Data-management mistake types, causes, and outcomes

Frequency of data-management mistake types. Through the grouping process, we sorted the 786 descriptions of the most frequent (*N* = 506) and most serious (*N* = 280) data-management mistakes into 20 different mistake types. To determine which type of mistakes are the most frequent, in our sample, we counted how many

Table 7. Research Data-Management Mistake Type Groups and the Number of Their Occurrences

Mistake type group	Most frequent mistakes ($N = 506$)	Most serious mistakes ($N = 280$)
Ambiguous naming/defining of data	86	16
Version control error	62	20
Wrong data processing/analysis	47	32
Data coding error	45	26
Data input error	37	21
Loss of materials/documentation/data	30	26
Programming error	35	19
Data transfer error	41	8
Data selection/merging error	21	17
Technical/infrastructural problem	17	21
Oversight in study design or measurement	14	18
Bad or poor documentation	21	7
Participant allocation error	15	8
Bad connection of data points	8	8
Wrong reporting	7	8
Wrong software or hardware settings	5	10
Data or file organization error	6	4
Deviation from the protocol	5	4
Violation in ethics	2	4
Project management error	2	3

Note: The mistake types are arranged in a descending order according to the number of times they were reported for the most frequent and most serious mistakes together.

times a mistake type was reported by respondents. For this analysis, we kept multiple responses provided by single respondents. Table 7 shows how many times a mistake type reportedly occurred for the most frequent and most serious mistakes. The three most frequently reported mistake types for the most frequent mistakes were ambiguous naming/defining of data (86 of 506), version control error ($n = 62$), and wrong data processing/analysis ($n = 47$). The three most frequently reported mistake types for the most serious mistakes were wrong data processing/analysis (32 of 280), data coding error ($n = 26$), and loss of materials/documentation/data ($n = 26$).

Mistake types and their reported causes. Figure 4 shows the data-management mistake types for the most frequent mistakes and the proportions of the metalevel grouping for their reported causes. The relationship between the mistake types and causes can also be viewed separately for the most serious mistakes (see Fig. S5 in the Supplemental Material). Cases were omitted from the analyses if the respondent described more than one mistake and more than one cause was associated with them because the mistake and its cause could not be unambiguously connected. In case of a one-to-many mapping, we assumed that the respondent wished to report several causes that led to a mistake or one cause that led to several mistakes.

The most common causes assumed by the researchers to be responsible for these most frequent mistake types were poor project preparation or management (43%) and personal difficulties (29%). For the most serious mistake types, the most common causes were the same: 39% for the poor project preparation or management and 37% for the personal difficulties.

Mistake types and their reported outcomes. Figure 5 shows the frequency of the data-management mistake types for the most frequent mistakes and the proportions of their reported outcomes. The relationship between the mistake types and reported outcomes can also be viewed separately for the most serious mistakes (see Fig. S6 in the Supplemental Material). Cases in which the respondent described more than one mistake and reported more than one negative outcome associated with those were omitted from this analysis. The most commonly reported outcomes that we could clearly associate with the mistake types for the most frequent mistakes were time loss (67%) and frustration (21%). The most common outcomes associated with the most serious mistakes were the same: 46% for time loss and 26% for frustration.

Mistakes types and data-management stages. We categorized each mistake type according to the data-management stage (or overlap of multiple stages) during
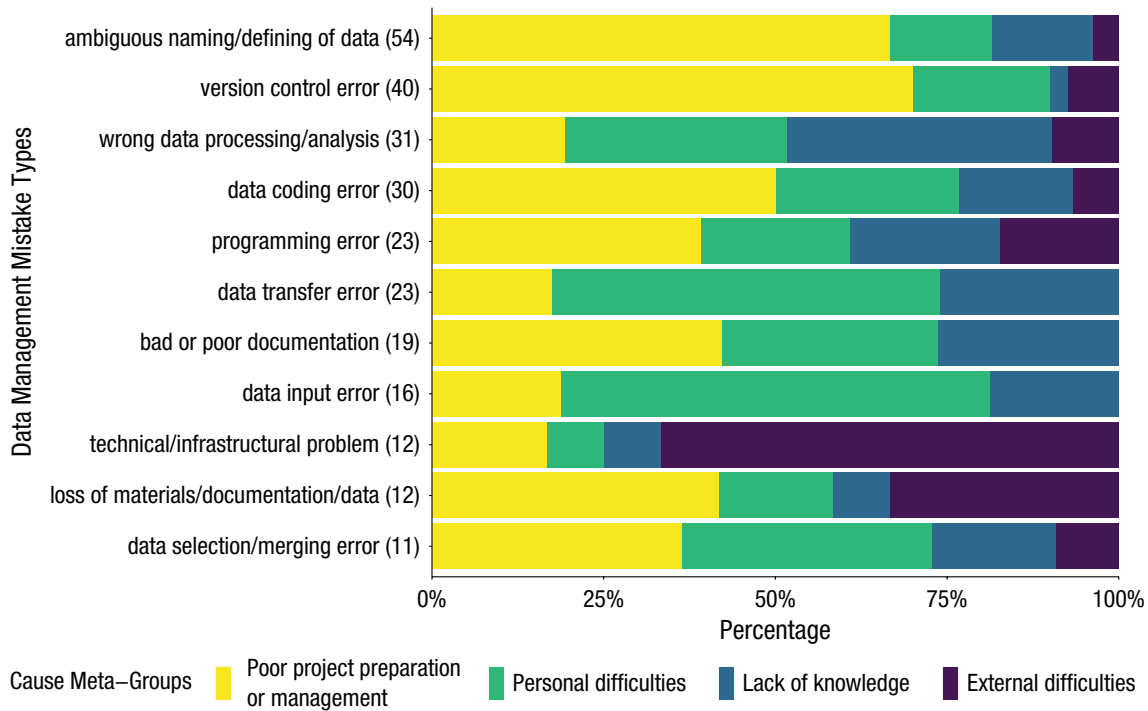
Fig. 4. The frequency of the data-management mistake types for the most frequent mistakes and the proportions for the metalevel grouping (see Table 6) of their reported causes. The mistake types are presented in decreasing order from the top to the bottom by the number of research teams who reported the specific mistake type. Mistake types with fewer than 10 occurrences are not displayed. The numbers in parentheses represent the number of times a given mistake type was reported after cases with multiple mistakes/causes were omitted.

which it was likely to have occurred (Fig. 6). Most types of mistakes belong to the overlap of the data processing/analysis, data creation/collection, and the data archiving/sharing sections. We developed the data-management model used for the present categorization. For a more detailed description of development, see the Supplemental Material.

Discussion

The results of this survey showed that data-management mistakes are ubiquitous in many labs conducting psychological research. Most respondents believed that data-management mistakes occur infrequently in their own research; one fifth of them observed mistakes in moderate, high, or very high frequency. The most serious mistakes had only minor consequences for one third of the research teams in our sample, whereas for one fifth of them, they came with major or extreme repercussions (e.g., project failure or erroneous conclusions). Naturally, this survey was not capable of detecting undiscovered or unreported mistakes, and therefore, it is plausible that our numbers underestimate the actual frequency of data-management mistakes. These exploratory findings do not aim to provide exact estimates but, rather, to help identify some common data-management mistakes and

potential causes and outcomes, which may facilitate the education about existing solutions and the development of novel mistake-mitigation strategies.

Respondents reported a variety of mistakes occurring across the research data-management pipeline. Deciding which mistakes are of highest priority to address will require consideration of their frequency and seriousness and the potential resources needed to address them. The majority of respondents reported that the most frequent mistakes, involving ambiguous naming or defining of data, version control error, and wrong data processing/analysis, can be associated with the data-processing and analysis stage. These mistakes were mostly assumed to be the result of poor project planning or management. Most frequently, the cost of these mistakes is time loss and frustration. We assume that erroneous conclusions are less frequent outcomes of these mistakes only because the reporters discovered the mistakes before publicizing their results. Hence, the proportion of conclusions that remain defective in the literature because of data-management mistakes is dependent on the efficiency of the existing checking procedures.

Most mistake types were categorized to more than a single stage because they can happen at several points of the data-management pipeline. The mistakes that were typical of most stages were found to be ambiguous
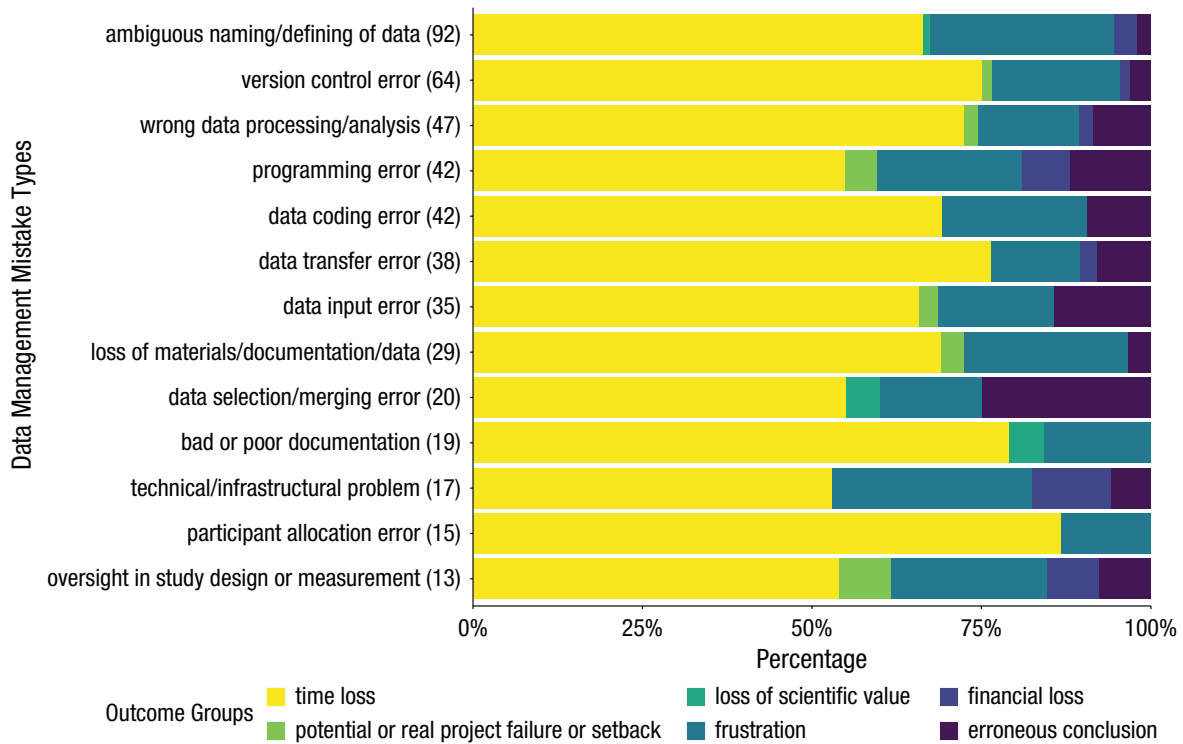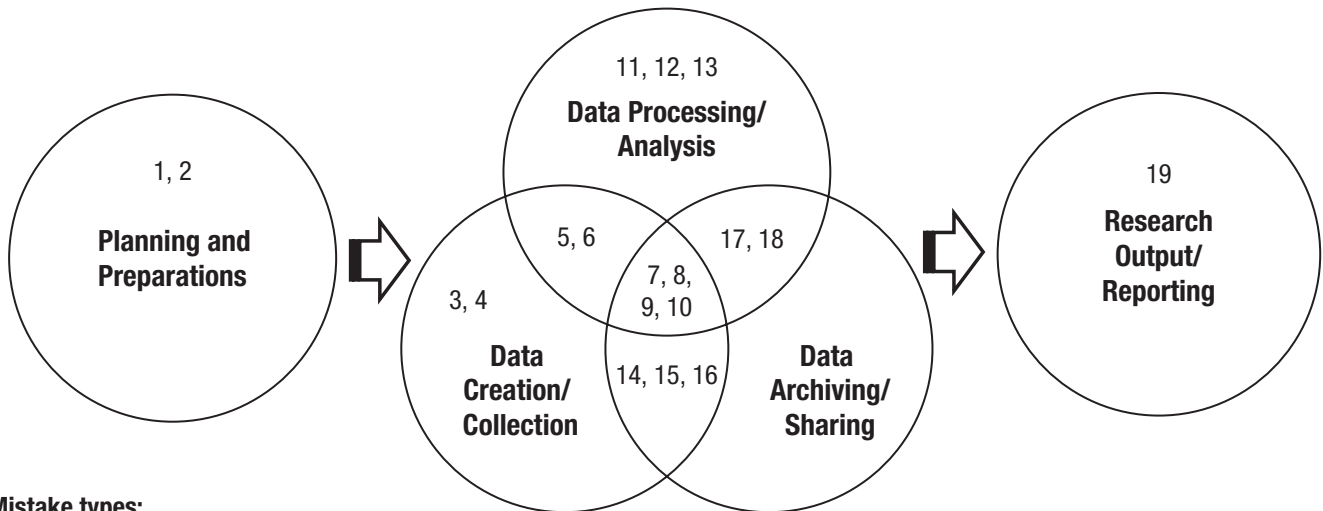
Fig. 5. The frequency of data-management mistake types for the most frequent mistakes and the proportions of the reported outcomes. The mistake types are presented in decreasing order from the top to the bottom by the number of research teams who reported the specific mistake type. Mistake types with fewer than 10 occurrences are not displayed. Numbers in parentheses represent the number of times a respondent reported the given mistake type.

naming/defining of data, data or file organization error, deviation from the protocol, and programming error.

A number of generic solutions and guidelines have been proposed to assist researchers with their data management. Using personal experience, Rouder et al. (2019) described five principles to minimize and mitigate research mistakes: (a) a lab culture focused on learning from mistakes, (b) implementing computer automation, (c) standardization, (d) coded analysis, and (e) elaborate manuscripts. Others have pointed toward the need for formal training in data management (Barone et al., 2017; Tenopir et al., 2016). Note that an increasing number of university library services provide dedicated support for data-management plans (Michener, 2015). Data librarians are specialized in providing support in managing research data (Semeler et al., 2019). Various guidelines and checklists have been developed to help researchers adopt transparent research workflows (Aczel et al., 2020; Klein et al., 2018), comprehensive reporting (e.g., the EQUATOR Network, https://www.equator-network .org/), reusability of data holdings (Wilkinson et al., 2016), and ethical and efficient research management (Bareille et al., 2017; Giesen, 2015). Dedicated software tools (e.g., R Markdown; Baumer & Udwin, 2015) are available to make data management more efficient, transparent, and less prone to error. We present a

noncomprehensive collection of existing error-mitigation tools or strategies corresponding to a number of our mistake types in Table 8. Note that the cause of the mistake can play an important role in the efficiency of the error-mitigation strategies. For example, if a person makes mistakes in data management not because of the lack of knowledge but because of some personal difficulties, then the potential solution will require more than mistake-specialized strategies.

This survey was intended to be exploratory and descriptive, and several caveats and limitations should be considered when interpreting the results. First, because the survey relied on researchers' self-report, the study will not have detected mistakes that were undiscovered, forgotten, or otherwise unreported. The findings may, therefore, highlight the existence of some pertinent data-management mistakes and perhaps their relative frequency but should not be interpreted as reliable estimates of mistake prevalence in psychological science. Second, although the number of researchers responding to the survey and passing our exclusions is adequate ($N = 488$) for exploratory purposes, the overall response rate (before exclusions) was very low (5%), which suggests that the findings are potentially strongly affected by self-selection bias. The overall direction of influence of such bias is difficult to predict because

Fig. 6. Mistake types categorized by research data-management stage. The numbers indicate the mistake types (see https://osf.io/76d24/). Mistake 20 (technical or infrastructure problems) is not part of any stage because it is an external factor but can have an effect on the efficiency of the data-management pipeline.

potential differences between respondents and nonrespondents are nontrivial (e.g., people who have made more mistakes may have been more likely to take part in the survey because it was more relevant to them or less likely to take part because reporting mistakes may have felt more embarrassing for them). Third, we gained only limited knowledge about the background of the respondents because many could not assign themselves to any of the psychological subfields offered in the survey and chose instead the "other" category. Finally, the

Table 8. Existing Error-Mitigation Strategies for the Most Frequent and/or Serious Data-Management Mistakes

Mistake type	Existing error-mitigation strategy
Ambiguous naming or defining of data	Using naming standards (e.g., Gorgolewski et al., 2016)
	Using codebooks (Arslan, 2019; Buchanan et al., 2021)
	Creating data management plans (Michener, 2015)
Version control error	Using a version control system such as Git (Blischak et al., 2016)
Wrong data processing/analysis	Copiloting (Veldkamp et al., 2014)
	Creating data-management plans (Michener, 2015)
	Using statistical code language (Python, R)
Loss of materials/documentation/data	Using a version control system such as Git (Blischak et al., 2016)
	Sharing information in online repositories (Klein et al., 2018)
Programming error	Copiloting (Veldkamp et al., 2014)
	Use software tests and code commenting (Michener, 2015)
Oversight in study design and measurement	Clear project structuring (Rybicki, 2019)
	Registered report format (Chambers, 2013)
Poor documentation	Transparent research workflow (Klein et al., 2018)
	Preregistration (Nosek et al., 2019)
Data or file organization error	Using data-specification standards (e.g., Gorgolewski et al., 2016) or file-organization standards (e.g., *The DRESS Protocol*, n.d.)

survey yielded a large quantity of partly qualitative data, and it was necessary to rely on our own subjective assessment to generate a meaningful summary. We attempted to improve objectivity by having at least two team members dual code all responses, but some subjectivity was required, nonetheless.

Psychological science is currently undergoing a period of heightened concern about the credibility and validity of its research practices and results (Vazire, 2018). Metaresearch efforts have focused on documenting major threats to credibility, such as fraud, questionable research practices, and low transparency (Hardwicke et al., 2020), but have paid relatively sparse attention to the role of basic human error. The present study has highlighted some pertinent mistakes that can percolate into the research pipeline that may reduce efficiency and potentially undermine the validity of scientific claims. Future work may look to build on these findings and develop a systematic exploration of human fallibility in research data management. Repeating our methodology on a representative sample could provide valuable information in this regard and identify the weaknesses of research efficiency. We suggest three major research questions for the continuation of this endeavor: (a) What practices do researchers use to improve efficiency and quality control in data management? (b) What prevents researchers from using existing solutions? and (c) What is needed to increase adoption of these solutions?

Transparency

ORCID iDs

Marton Kovacs [iD] https://orcid.org/0000-0002-8142-8492
Rink Hoekstra [iD] https://orcid.org/0000-0002-1588-7527
Balazs Aczel [iD] https://orcid.org/0000-0001-9364-4988

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/25152459211045930

References

Aczel, B., Szaszi, B., Sarafoglou, A., Kekecs, Z., Kucharský, Š., Benjamin, D., Chambers, C. D., Fisher, A., Gelman, A., Gernsbacher, M. A., Ioannidis, J. P., Johnson, E., Jonas, K., Kousta, S., Lilienfeld, S. O., Lindsay, D. S., Morey, C. C., Munafò, M., Newell, B. R., . . . Wagenmakers, E.-J. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, *4*(1), 4–6. https://doi.org/10.1038/s41562-019-0772-6

Arslan, R. C. (2019). How to automatically document data with the codebook package to facilitate data reuse. *Advances in Methods and Practices in Psychological Science*, *2*(2), 169–187. https://doi.org/10.1177/2515245919838783

Bareille, R., Baudouin-Massot, B., Carreno, M. P., Fournier, S., Lebret, N., Remy-Jouet, I., & Giesen, E. (2017). Preventive actions to avoid questionable research practices. Use of EERM (Ethical and Efficient Research Management) during Arrival and Departure of a co-worker. *International Journal of Metrology and Quality Engineering*, *8*, 10. https://doi.org/10.1051/ijmqe/2016029

Barone, L., Williams, J., & Micklos, D. (2017). Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators. *PLOS Computational Biology*, *13*(10), Article e1005755. https://doi.org/10.1371/journal.pcbi.1005755

Baumer, B., & Udwin, D. (2015). R markdown. *Wiley Interdisciplinary Reviews: Computational Statistics*, *7*(3), 167–177. https://doi.org/10.1002/wics.1348

Blischak, J. D., Davenport, E. R., & Wilson, G. (2016). A quick introduction to version control with Git and GitHub. *PLOS Computational Biology*, *12*(1), Article e1004668. https://doi.org/10.1371/journal.pcbi.1004668

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Buchanan, E. M., Crain, S. E., Cunningham, A. L., Johnson, H. R., Stash, H., Papadatou-Pastou, M., Isager, P. M., Carlsson, R., & Aczel, B. (2021). Getting started creating data dictionaries: How to create a shareable data set. *Advances in Methods and Practices in Psychological Science*, *4*(1), https://doi.org/10.1177/2515245920928007

Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, *49*(3), 609–610. http://doi.org/10.1016/j.cortex.2012.12.016

The DRESS Protocol. (n.d.). https://www.projecttier.org/tier-protocol/dress-protocol/

Giesen, E. (2015). Ethical and efficient research management: A new challenge for an old problem. *International Journal of Metrology and Quality Engineering*, *6*(4), 406. https://doi.org/10.1051/ijmqe/2015028

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., . . . Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.44

Hardwicke, T. E., Jameel, L., Jones, M., Walczak, E. J., & Weinberg, L. M. (2014). Only human: Scientists, systems, and suspect statistics. *Opticon1826*, *16*(25), 1–12.

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., & Henry Tessler, M. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448.

Hardwicke, T. E., Serghiou, S., Janiaud, P., Danchev, V., Crüwell, S., Goodman, S. N., & Ioannidis, J. P. A. (2020). Calibrating the scientific ecosystem through meta-research. *Annual Review of Statistics and Its Application*, *7*, 11–37. https://doi.org/10.1146/annurev-statistics-031219-041104

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Johnson, H. R., Stash, H., Papadatou-Pastou, M., Isager, P. M., Carlsson, R., & Aczel, B. (n.d.). *Getting started creating data dictionaries: How to create a shareable dataset Erin M. Buchanan 12abc Sarah E. Crain 1abc Arielle Cunningham 1abc*.

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., Ijzerman, H., Nilsonne, G., Vanpaemel, W., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1), Article 20. https://doi.org/10.1525/collabra.158

Michener, W. K. (2015). Ten simple rules for creating a good data management plan. *PLOS Computational Biology*, *11*(10), Article e1004525. https://doi.org/10.1371/journal.pcbi.1004525

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, *69*(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., Veer, A. E. van't, & Vazire, S. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

Nuijten, M. B., Hartgerink, C. H., Van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205–1226.

Rosenthal, R. (1978). How often are our numbers wrong? *American Psychologist*, *33*(11), 1005–1008.

Rouder, J. N., Haaf, J. M., & Snyder, H. K. (2019). Minimizing mistakes in psychological science. *Advances in Methods and Practices in Psychological Science*, *2*(1), 3–11. https://doi.org/10.1177/2515245918801915

Rybicki, J. (2019). Best practices in structuring data science projects. In Z. Wilimowska, L. Borzemski, & J. Świątek (Eds.), *Information systems architecture and technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology – ISAT 2018* (pp. 348–357). Springer International Publishing.

Semeler, A. R., Pinto, A. L., & Rozados, H. B. F. (2019). Data science in data librarianship: Core competencies of a data librarian. *Journal of Librarianship and Information Science*, *51*(3), 771–780. https://doi.org/10.1177/0961000617742465

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Tenopir, C., Allard, S., Sinha, P., Pollock, D., Newman, J., Dalton, E., Frame, M., & Baird, L. (2016). Data management education from the perspective of science educators. *International Journal of Digital Curation*, *11*(1). https://doi.org/10.2218/ijdc.v11i1.389

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411–417.

Veldkamp, C. L., Nuijten, M. B., Dominguez-Alvarez, L., van Assen, M. A., & Wicherts, J. M. (2014). Statistical reporting errors and collaboration on statistical analyses in psychological science. *PLOS ONE*, *9*(12), Article e114876. https://doi.org/10.1371/journal.pone.0114876

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18