Studia Psychologica, Vol. 65, No. 1, 2023, 56-70
https://doi.org/10.31577/sp.2023.01.866

Statistical Misconceptions, Awareness, and Attitudes towards Open Science Practices in Slovak Psychology Researchers

Jakub Rajčáni[1], Lenka Vargová[2], Matúš Adamkovič[2,3,4], Pavol Kačmár[5]
[1] Department of Psychology, Faculty of Arts, Comenius University, Gondova 2, 81102, Bratislava, Slovak Republic
[2] Institute of Psychology, Faculty of Arts, University of Prešov, 17. Novembra 1, 08001, Prešov, Slovak Republic
[3] Institute of Social Sciences, Centre of Social and Psychological Sciences SAS, Šancová 56, 81105 Bratislava, Slovak Republic
[4] Faculty of Humanities and Social Sciences, University of Jyväskylä, Keskussairaalantie 2, FI-40014 Jyväskylä, Finland
[5] Department of Psychology, Faculty of Arts, Pavol Jozef Šafárik University in Košice, Šrobárova 2, 04059, Košice, Slovak Republic

In the years following the reproducibility crisis in behavioral sciences, increased attention of the scientific community has been dedicated to the correct application of statistical inference and promotion of open science practices. In the present survey, we contacted psychology researchers, lecturers, and doctoral students from all universities in Slovakia and the Slovak Academy of Sciences via email. Together we received answers from 65 participants. Questions in the survey covered the most common misconceptions about statistical hypothesis testing, as well as awareness, attitudes, and barriers related to the adherence to open science practices. We found a high prevalence of statistical misconceptions, namely related to the interpretation of *p*-values and interpretation of null results. At the same time, participants indicated mostly positive attitudes towards open science practices, such as data sharing and preregistration, and were highly interested in further training. These results provide an insight into the needs of the Slovak psychology research community. This is an important step in the further dissemination of open science practices and the prevention of common statistical and methodological errors.

Key words: survey, statistical misconceptions, reproducibility, replicability, open science practices

Introduction

In recent years, reproducibility and robustness of research findings have been widely discussed across disciplines (e.g., Camerer et al., 2018; Klein et al., 2018; Open Science Collaboration, 2015). The term reproducibility

Correspondence concerning this article should be addressed to Jakub Rajčáni, Department of Psychology, Faculty of Arts, Comenius University, Gondova 2, 81102, Bratislava, Slovak Republic.
E-mail: jakub.rajcani@uniba.sk

Supporting material to this paper is available at https://osf.io/nw5a4/

Received April 1, 2022

crisis[1] (Baker, 2016) has been used to describe the state of low replicability of empirical research in psychology (Open Science Collaboration, 2015; Klein et al., 2018), as well as in other disciplines, such as economics (Camerer et al., 2016), cancer research (Begley & Elis, 2012; Errington et al., 2021), or neuroscience (Carp et al., 2012).

Throughout the years, several factors contributing to low replicability rates have been identified. These factors include methodological errors, low statistical power, publication bias, and questionable research practices (Ioannidis, 2005). Questionable research practices involve, for instance, undisclosed flexibility during data analysis to obtain a significant result (p-hacking; Simmons et al., 2011; John et al., 2012) or hypothesizing after the results are known (HARKing; Kerr, 1998). Surveys have shown that these practices are disturbingly prevalent among researchers (Agnoli et al., 2017; John et al., 2012; but see also Fiedler & Schwarz, 2016), while simulations show that they substantially inflate error rates in research (Ioannidis, 2005).

While the high prevalence of questionable research practices is often discussed in the context of research misconduct, another side of the problem points toward a widespread misunderstanding or misinterpretation of statistical inference. Researchers are not immune to common cognitive biases and simplifications, especially when dealing with unintuitive concepts like probability, which are essential in hypothesis testing (Nuzzo, 2016). Lack of understanding of statistical inference could be one of the key factors contributing to the use of questionable practices in data analysis. As statistical misconceptions are fairly common (Haller & Kraus, 2002), they cause not only conceptual but also practical problems (Wetzels et al., 2011), and they complicate our understanding of psychological phenomena (Wagenmakers et al., 2011).

Because Null Hypothesis Significance Testing (NHST) is still the most prevalent inferential approach in psychology and related disciplines (Amrhein et al., 2019; Cumming et al., 2007), common misinterpretations of NHST concepts pose a serious problem. They often result from the practice of NHST, which is an unambiguous amalgam of Fisher's and Neyman-Pearson's approaches (Dienes, 2008; Perezgonzalez, 2015), leading to confusion in the interpretation of *p*-value or omission of crucial steps in research planning (e.g., power analysis for sample size determination). Previous research has shown that misunderstandings and oversimplifications of NHST are prevalent not only among students but also among senior researchers and even methodology teachers (Haller & Kraus, 2002). Most common NHST misconceptions concern the erroneous assumptions that the *p*-value represents the probability of the null hypothesis being true ("inverse probability fallacy"); that *p*-value is a measure of effect ("effect size fallacy"), or the probability the result will replicate in further studies ("replication fallacy", Nickerson, 2000; Badenes-Ribera et al., 2016; Goodman, 2008; Greenland et al., 2016; Ropovik, 2017). Another pervasive misconception is related to the interpretation of negative results, for example, mistaking the absence of evidence for the evidence of absence of an effect (Altman & Bland, 1995; Alderson, 2004).

In the present debate on statistical misconceptions, many authors have emphasized the importance of moving beyond *p*-values (Amrhein et al., 2019), calling for augmentation or even replacement of NHST by reporting effect sizes (e.g., Kelley & Preacher, 2012), con-

[1] Although some authors differentiate the terms replicability and reproducibility (for further discussion see e.g., Nosek et al., 2022), for the sake of simplicity we will use the term reproducibility crisis in this manuscript in a more general sense in accordance with Baker (2016).

fidence intervals (e.g., Cumming, 2013), or Bayesian approach to hypothesis testing (e.g., Dienes & Mclatchie, 2018; Wagenmakers et al., 2018). However, a solution does not have to be a replacement of NHST by other methods, but by properly understood and applied NHST (Lakens, 2021). Identification of statistical misconceptions and consequent training is still a present-day problem.

In recent years, many advancements in research reproducibility have taken place. These include the promotion of open science practices (e.g., Munafo et al., 2017; Nosek et al., 2012), which bring more transparency into the researchers' decisions and data analytic processes. These cover practices such as preregistration of research hypotheses that enable clear distinction between what was and what was not predicted (Nosek et al., 2018); and sharing of datasets, materials, and analytical code, which enable others to re-run the analysis or reuse the data (Murray-Rust, 2008; Nosek et al., 2012). Nevertheless, evaluating the interest of the research community and perceived obstacles is very important in the progress toward open science.

To help mitigate the statistical misconceptions and facilitate the use of open science practices in a local context, we present data from a survey on analytic practices of Slovak psychology researchers. The aim of the present study is twofold. First, to identify the most common errors in understanding and interpretation of statistical concepts and estimate their prevalence among Slovak psychology researchers a decade into the reproducibility crisis. Second, we aim to examine researchers' attitudes toward open science practices, perceived barriers, and demand for further training in this topic. Having this information is crucial as it could help target the weak spots and tailor more effective training and education (e.g., workshops, webinars, seminars, or courses) for both academics and students in the Slovak context and language. Furthermore, data from a survey like this are important to compare with findings from other countries and cultural backgrounds.

Methods

Participants and Procedure

We have reached Slovak researchers, university lecturers, and professors, as well as doctoral students in the field of psychology, whose professional email contacts were available on their institution websites in spring 2020 (N = 347). The data was collected in two phases. In the first wave, data were collected between May 11 and May 30, 2020, via Microsoft Forms. Due to a technical error, the data collection was terminated prematurely with the loss of some data. Data collection continued from January 13 to May 21, 2021, via Qualtrics software. In the second run, participants were asked to indicate if they participated in this survey about a year ago. None of the participants answered yes to this item[2].

The study protocol followed the Declaration of Helsinki and was approved by the Ethics Committee of the Faculty of Arts, Comenius University, Bratislava, Slovakia with the number EK05/2020.

Altogether, 65 academics participated in the survey (17.4% of the contacted sample). Data from 14 and 51 participants were available from the first and the second wave of data collection, respectively. As we aimed to include participants who completed at least 20% of the survey, two participants (specifically, had the progress of 2%) were excluded from the analyses. Only 40 participants (11.5% of the contacted sample) have completed the whole survey. For each part of the

[2] Note that there are some minor differences between items in the first and the second data collection. These are reflected upon in the limitations.

survey, we report as much data as possible given the completion rate. In sum, we were able to use data from 52 to 63 participants for the items on demographic characteristics, work experiences, preferences, and perceived barriers regarding open science practices, data from 47 participants for the items concerning statistical/methodological knowledge, and data from 40 participants for two tasks that involved interpretation of statistical results. Naturally, the sample size limits the precision of the estimates presented below. For more detailed characteristics of the participants, see Table 1.

Survey

The online survey was presented in the Slovak language. It consisted of three parts, the first covered the demographic characteristics

Table 1 *Characteristics of the research sample*

Variable	*N* or Mean	Percentage or *SD*
Gender		
Female	40	63.5%
Male	21	33.3%
Other/NA	2	3.2%
Position		
Academic	46	73%
Ph.D. student	17	27%
Years in academia	10.4	8.6
Teaching methodology or statistics courses	23	43.4
Total number of papers authored		
0-10	23	39.7%
11-25	12	20.7%
26-50	7	12.1%
51-100	9	15.5%
100+	7	12.1%
Data analysis performed		
Alone	31	54.4%
In collaboration	18	31.6%
Someone other from the team	5	8.8%
External statistician	1	1.8%
Other	2	3.5%
Statistical software used (more than one option)		
SPSS	50	87.7%
PSPP	5	8.8%
Excel (or MS Office)	16	28.1%
R	9	15.8%
JASP	11	19.3%
Jamovi	12	21.1%
Other	4	7.0%

of the sample, including the questions on the number of authored studies, experience with supervising students, teaching methodological or statistical courses, etc. (for an outline of the questions and results, see Table 1 and the supplementary material). The second part covered questions on attitudes toward replication studies and open science practices. In this part, the participants were asked to rate their interest in methodology and statistics, rate their trust in published findings, and subjectively compare their understanding and quality of statistics and methodology in their papers with their peers (all on a scale from 0 to 100, with a higher number indicating higher quality). This was followed by several items focused on experience and attitudes towards topics like preregistration, data sharing, the distinction between exploratory and confirmatory research, and meta-analysis. Besides the participants' awareness of these issues, we were interested in the obstacles the participants perceive in implementing open science practices in their research. In the third part, items concerning common misinterpretations in statistical inference were surveyed in a quiz-like manner. These covered topics of NHST (e.g., understanding of *p*-value), effect sizes, confidence intervals, and basics of Bayesian statistics. Some of the questions were based on previous surveys (Haller & Kraus, 2002), on misconceptions discussed in the published literature (e.g., Gelman & Stern, 2006; Greenland et al., 2016), and on the materials from the course "Improving your statistical inferences" by Lakens (2016). The survey questions in Slovak are available in the supplementary materials.

Statistical Analysis

Considering the purpose of the present study, descriptive statistics were used. To account for the sample size limitations, we computed 95% CIs around the prevalence estimates via an online calculator (see Kohn & Senyak, 2022). All analyses were performed in R. Items were analyzed one by one, however, answers to questions about NHST, effect sizes, confidence intervals, and basics of Bayesian statistics were also analyzed by calculating an overall performance score. Please note that in some instances, it was not possible to precisely determine if a participant skipped the item or did not know the correct answers. In such cases, we used the best guess. This, however, has only a minor effect on the results and does not change them substantially. Data, analytical script, and more detailed results are available in the supplementary material (https://osf.io/nw5a4/).

Results

In general, on a scale from 0 to 100 (50 defined as subjectively perceived average in the population of researchers), the participants have indicated relatively high interest in methodology (M = 72.77, SD = 22.07), and above average satisfaction with their statistical practices (M = 63.63, SD = 21.97). They perceived both their statistical (M = 60.56, SD = 22.59) and methodological practices (M = 64.88, SD = 19.79) as slightly better compared to other researchers. The participants' trust in published literature was moderate to high (M = (69.96, SD = 18.1).

Statistical Knowledge and Misconceptions

We asked participants to indicate whether they use NHST, confidence intervals (CIs), effect sizes, and Bayesian statistics. Seventy-nine percent (95%CI [64%, 89%]) of the participants use NHST, 15% (95%CI [6%, 28%]) stated they know it, but do not use it, while only 6% (95%CI [01%, 18%]) reported they do not use nor know this approach. Forty-eight percent (95%CI [32%, 63%]) indicated using confidence intervals, 25%, (95%CI [13%, 40%]) stated they know CIs, but do not use them, while 27% (95%CI [15%, 43%]) do not use nor know this approach. For the topic of effect sizes, 67%, (95%CI [51%, 81%]) indicated they usually report an effect size measure, 16%, (95%CI [7%, 31%]) do not report effect sizes but know about them, and 16% (95%CI [7%, 31%]) do not report nor know the con-

Table 2 *Prevalence of misconceptions about p-values (N = 47)*

Misconception	Yes/Agree [95% CI]	*No/Disagree* [95% CI]	Does not know/does not want to answer [95% CI]
a/ *P*-value represents the probability of the null hypothesis being true.	35% [21%, 50%]	*52%* [37%, 67%]	13% [5%, 26%]
b/ If *p* = 0.01 in one study and *p* = 0.09 in a second study, does it mean that a larger effect was observed in the first study compared to the second study?	22% [11%, 36%]	*70%* [54%, 82%]	8% [2%, 21%]
c/ *P*-value represents the probability that our result is the result of chance.	41% [27%, 57%]	*43%* [29%, 59%]	15% [6%, 29%]
d/ If we obtain a statistically non-significant result (*p* > 0.05), we accept the null hypothesis (i.e., we conclude that the examined effect does not exist).	53% [38%, 68%]	*33%* [20%, 49%]	13% [5%, 27%]
e/ If you test the same hypothesis in two populations (e.g., male and female), with *p* < 0.05 in one sample and *p* > 0.05 in the other sample, does it mean that there is a significant difference between the groups?	40% [26%, 56%]	*44%* [30%, 60%]	16% [6%, 29%]
f/ If there is no examined effect in the population (i.e. the null hypothesis is true):			
there is a greater chance that we will get a high *p*-value than a low one			49% [34%, 64%]
there is a greater chance that we will get a low *p*-value than a high one			9% [2%, 21%]
all p-values are equally likely			*20%* [10%, 35%]
I don't know/I don't want to answer			22% [11%, 37%]

Note. Correct answer to questions *a* - *e* is No/Disagree; correct answer to question *f* is „all *p*-values are equally likely". The correct answers are highlighted in italic.

cept of effect sizes. The usage of Bayesian statistics was reported by 7% (95%CI [1%, 19%]) of the participants, 33% (95%CI [19%, 49%]) do not use it but know it, while 60% (95%CI [44%, 75%]) do not know nor use this approach.

In the quiz-like part of the survey, we tested the participants' knowledge of statistical concepts and the prevalence of statistical misconceptions by asking them 14 questions, six related to *p*-values, three to confidence intervals, two to effect sizes, one to statistical power, and two to Bayesian statistics. The number of correct answers (out of 14) ranged from 0 to 13, with M = 5.71 and SD = 3.45 (about 39% of correct answers). More specifically, the mean score for the *p*-value questions was 2.55 (SD = 1.77) (out of 6), the mean score for the confidence interval questions was 0.94 (SD = 0.92) (out of 3), the mean score for the effect size questions was 0.84 (SD = 0.72) (out of 2), the mean score for the statistical power question was 0.45 (SD = 0.50) (out of 1), and the mean score for the items on Bayesian statistics was 0.70 (SD = 0.78) (out of 2). Table 2 summarizes answers to questions concerning misconceptions about *p*-values. For even more detailed results from the items related to CIs, effect sizes, power, and Bayesian statistics see the supplementary materials.

Sample size justification. We also asked the participants how they determine the required sample size in their studies (with the possibility of selecting multiple answers). Only 35% (95%CI [23%, 48%]) of the participants stated they calculate a priori power analysis to determine sample size, with 27% (95%CI [17%, 40%]) stating they compute a priori power analysis with a predetermined smallest effect size of interest (SESOI; Lakens et al., 2018), and 19% (95%CI [10%, 31%]) with the effect size estimate based on similar studies (there is a substantial overlap between both answers). Other methods were the usage of a conventional rule (rule of thumb) to set the minimum sample size in 16% (95%CI [8%, 27%]) and the determination of the sample size based on financial and time constraints in 35% (95%CI [23%, 48%]). About 8% (95%CI [3%, 18%]) do not consider sample size before data collection at all.

Applied tasks. Finally, in the applied tasks, the participants were shown two results of *t*-tests that included corresponding *t*-statistics, degrees of freedom, *p*-values, 95% confidence intervals, Cohen's *d*'s, and BFs_{10}. The first scenario was accompanied by 8 statements and the second scenario was accompanied by 7 statements. In both cases, 3 of the statements were true. In the first scenario, 35% (95%CI [21%, 52%]) of participants marked one correct response, 15% (95%CI [6%, 30%]) two correct responses and 2% (95%CI [0%, 13%]) of the sample correctly selected all three options. Similar results were observed for the second scenario with 35% (95%CI [21%, 52%]) of the participants marking one correct answer, 17% (95%CI [7%, 33%]) two correct answers, and 5% (95%CI [0%, 17%]) having all the three answers correct.

Attitudes, Barriers, and Demands

In the second part of the survey, we asked about attitudes, perceived barriers, and demand for further training in statistical and methodological practices. The main topics covered here were replication studies, preregistration of hypotheses, differentiation between exploratory and confirmatory research, data sharing, and meta-analysis.

Replication. The need for replication was perceived as highly relevant (M = 83.39, SD = 15.7). Almost 35% (95%CI [23%, 48%]) have experience with conducting a replication study (10 out of these 22 participants report multiple experiences), while 47.6%

(95%CI [35%, 61%]) have never conducted a replication. The majority of participants, 86.5%, (95%CI [74%, 94%]), plan to or consider replicating their studies in the future. The most common barriers to conduct a replication were that researchers prefer to focus on new problems 36.5% (95%CI [24%, 51%]), followed by a perceived difficulty to publish replications, 28.8% (95%CI [17%, 43%]), a lack of support from the affiliated institutions, 17.3%, (95%CI [8%, 30%]), and the opinion that putting effort into replications will slow down their careers 3.8% (95%CI [0%, 13%]).

Preregistration. Sixty-three percent (95%CI [49%, 76%]) of participants reported that they know what preregistration is. In the second question, participants indicated the perceived importance of preregistration (M = 68.55, SD = 27.64) using a scale from 0-not important at all to 100-absolutely important. About 14% (95%CI [7%, 25%]) of the participants have already preregistered a study and almost 40% (95%CI [28%, 53%]) plan to do so in the future. Besides not knowing about preregistration at all, we identified the following barriers preventing preregistering a study: a lack of information about preregistration, 25%, (95%CI [14%, 39%]); lack of guidelines and standards on how to prepare preregistration, 17.3%, (95%CI [8%, 30%]); fear that it will complicate the publication process, 9.6% (95%CI [3%, 21%]); difficulty to formulate hypotheses a priori, 9.6%, (95%CI [3%, 21%]); and considering the problem of preregistrations to be unimportant, 8%, (95%CI [2%, 19%]).

Exploratory and confirmatory research. A related issue is a clear differentiation between exploratory and confirmatory research practices. On a scale where 0 = completely exploratory and 100 = completely confirmatory, the participants indicated that their research is somewhere in the middle (M = 54.3, SD = 24.5), while they consider the differentiation between the two to be an important issue (M = 80.1, SD = 18.64). When asked about distinguishing exploratory from confirmatory parts in their work, 73% (95%CI [60%, 83%]) of the participants stated they generally differentiate between the two, 21% (95%CI [11%, 33%]) sometimes differentiate between the two, and 6% (95%CI [1%, 16%]) reported that they do not discriminate exploratory from confirmatory research at all.

Data sharing. In general, the participants support the importance of data sharing (M = 75.21, SD = 22.06). On the other hand, less than 30% (95%CI [17%, 43%]) have experience with sharing data with their research (about one-third of them share data regularly). About 64% (95%CI [49%, 76%]) would like to try it in the future, whereas 8% (95%CI [2%, 19%]), refrain from or were against data sharing. The main barriers preventing data sharing were the fear that data will be misused, 34% (95%CI [21%, 49%]), lack of technical background and knowledge of how to share data, 21.3% (95%CI [11%, 36%]), fear of exposing mistakes, 17% (95%CI [8%, 31%]), giving a possible advantage to competing researchers 10.6% (95%CI [4%, 23%]), and a lack of support from collaborators 6.4% (95%CI [1%, 17%]).

Meta-analysis. Despite the indisputable importance of evidence synthesis, the topic of meta-analysis is highly underrepresented in the Slovak context. As such, we asked the participants about their experience with meta-analyses. Only 6% of the participants (95%CI [1%, 16%]) indicated they co-authored a meta-analysis. The perceived importance of meta-analyses in scientific publishing on a scale from 0 to 100 was M = 84.12 (SD = 15.45), while the participants' self-evaluation of understanding the outputs of a meta-analysis was M = 63.85 (SD = 22.79).

Additional training. The vast majority, 96.1% (95%CI [87%, 99%]) of the participants expressed their interest in additional training in open science practices, statistics, and quantitative methodology. More specifically, we identified the greatest demand for training in the following areas: meta-analysis 67.3% (95%CI [53%, 80%]), preregistration 61.5% (95%CI [47%, 75%]), reporting of results 61.5% (95%CI [47%, 75%]), questionable research practices and how to avoid them, 57.7% (95%CI [43%, 71%]), sample size determination 55.8% (95%CI [41%, 70%]), data sharing 53.8% (95%CI [40%, 68%]), doing replications 50% (95%CI [36%, 64%]), sharing of reproducible analytic code 30.8% (95%CI [19%, 45%]), and equivalence testing, 9.6% (95%CI [3%, 21%]).

Discussion

The present paper provides the results of a survey on the frequency of statistical misconceptions among psychology researchers and Ph.D. students in Slovakia. The survey also monitors their attitudes toward open science. Besides mapping general attitudes to open science practices, we identify the barriers preventing people from adhering to open science practices and discuss potential solutions.

Overall, the 18% response rate results in a relatively small sample, as the population of psychology researchers, lecturers, and postgraduate students in Slovakia is limited. However, the achieved response rate is comparable to other similar surveys, for example, a survey of open science practices done in Germany (Abele-Brehm et al., 2019), or an international survey on attitudes towards data sharing (Houtkoop et al., 2018).

Statistical Misconceptions

Our estimates of the frequency of statistical misconceptions among Slovak psychology researchers could be put into perspective by a comparison with other surveys, such as the one from Germany (Haller & Kraus, 2002), Spain (Badenes-Ribera et al., 2015), or Italy and Chile (Badenes-Ribera et al., 2016). The "inverse probability fallacy" which was described as the most frequent in previous surveys tested via the statement "*p*-value represents the probability of the null hypothesis being true" was correctly rejected by 52% of our sample. This is between the estimates from Spain (42%, Badenes-Ribera et al., 2015) and estimates from Germany (74%; Haller & Kraus, 2002), Italy, and Chile (76.2%; Badenes-Ribera et al., 2016). However, the frequency of the rejection of the misconceptions relies on the way the statement is formulated, as our estimate dropped to 43% when the question was asked in a different way ("is *p*-value the probability that our finding occurred due to chance?"). Correct rejection of all multiple statements with "inverse probability fallacy" is generally lower (6.2%; Badenes-Ribera et al., 2015; 38.4% Badenes-Ribera, et al., 2016).

Seventy percent of our sample correctly rejected the erroneous interpretation of *p*-value as an indicator of effect size ("effect size fallacy") which is slightly less than in previous surveys (86.8% Badenes-Ribera et al., 2015; 87.8%, Badenes-Ribera et al., 2016), the rate of this error is, however, also dependent on the exact wording of the question. We have also tested a misconception pointed out by Gelman & Stern (2006), that if there is $p < 0.05$ in one group but not in the other, researchers tend to interpret it as a difference between groups. This erroneous assumption was correctly rejected by 44% of the sample, while 40% agreed with it, showing that this is one of the most prevalent errors.

We found the lowest percentage of correct answers (33%) with the statement that if $p > 0.05$, the null should be accepted (meaning there is no difference). This idea seems intuitive at first glance; however, it is often incorrect as $p > 0.05$ could also indicate low statistical power for detecting the (desired) effect. The majority of the surveyed researchers, therefore, confused the absence of evidence (e.g., low statistical power) with the evidence of the absence of an effect (Altman & Bland, 1995; Alderson, 2004). This misconception is very problematic in studies with low statistical power, as it leads to inflation of false-negative results and subsequent interpretations. Aczel et al. (2018) have recently shown that in 2015 abstracts of three prominent psychology journals, 72% of non-significant results were misinterpreted as evidence of the absence of an effect. Similarly, according to Amrhein et al. (2019), this misconception is shared by roughly 51% of researchers based on published literature, which is very close to our estimate in Slovak researchers (53%).

The misconceptions discussed above are a problem as they influence researchers' behaviors and therefore the current state of published scientific literature. One of the commonly proposed solutions is moving beyond NHST and *p*-values (Amrhein et al., 2019). Besides the high frequency of misconceptions, another argument for this is its support of dichotomous categorical thinking, which is usually inconsistent with quantitative science (Cumming, 2013). Currently, the American Psychological Association has called for reporting effect sizes and confidence intervals in addition to *p*-values (Cumming et al., 2012), while others emphasize the benefits of using Bayesian statistics (Wagenmakers et al., 2018) or likelihood inference (Dienes, 2008). Problems in current employment of NHST however could be solved within the NHST framework. Popularization of methods such as equivalence tests (Lakens et al., 2018) and support of properly understood NHST (Lakens, 2021) can also help amend the misconceptions and systematic errors described here.

Open Science Practices

Overall, the participants consider reproducibility and openness in science as important, which reflects the intensity of the ongoing debates on this topic. At the same time, despite the reproducibility crisis, participants still considered the published literature to be trustworthy, which goes in line with the findings of Baker (2016). Replications were deemed to be important, while at the same time only a minority of the sample has done a replication study. The main barriers for replications were related to the incentivization of novel research and experienced lack of interest and support from journals and institutions.

A slight majority of the sample have basic knowledge about preregistrations and it was perceived as an important practice overall. Besides lack of awareness of the topic in general, participants perceived they lack knowledge about preregistration standards (e.g., widely accepted preregistration templates). Such standards are, however, available (see Bowman et al., 2016; Preregistration Task Force, 2020; the recent norms by international psychological societies), and, as such, should be brought to researchers' attention during relevant workshops and dissemination activities.

Obstacles preventing people from data sharing mostly lie in fears of data misuse, theft, or exposing themselves if errors in the author's work are found (see also Houtkoop et al., 2018). In our survey, most participants selected data misuse as a barrier. One possibility is a misuse of personal information about participants, which is prevented if the authors adhere to ethical standards and legal obligations. The second problem is data scooping (Houtkoop et al., 2018), that is, when someone publishes a paper based on a dataset first, without crediting the original author. Although this fear is common, the chance of data being stolen is probably highly overestimated (Gewin, 2016). It is also not possible if the data are published under the name of the original author using a public copyright license (e.g., Creative Commons, 2022), or shared via a repository with a digital identifier (DOI). Thus, the authors claim their authorship and, at the same time, enable others to freely use the data.

Another issue is the question of data ownership. One of the participants stated that "The data is [my] intellectual property. I accept sharing them only either to verify their authenticity or to sell them." Although this could apply to certain kinds of privately funded research, for most of the research, this argument is not defensible. If the research is financed by taxpayers, the data belongs to the public and should not be deemed as the exclusive property of a researcher/research team (e.g., Chambers, 2019). In addition, according to the Ethical principles of psychologists and code of conduct (American Psychological Association, 2016), after results are published, the researcher should not withhold the data, from which the conclusions presented in the study were derived. Additionally, fear of exposing one's errors is, too, only a hardly justifiable obstacle (Nosek et al., 2012). According to Witcherts et al. (2011), the reluctance of data sharing is positively correlated with the extent of actual mistakes, namely mistakes that influence the result of a tested hypothesis.

Finally, lack of information on data management was also one of the most frequently mentioned barriers. This could, however, be improved relatively easily by additional training. These days, training, tools, and support for data management planning are offered by institutions (e.g., specialized centers at universities) and online platforms (e.g., https://dmponline.dcc.ac.uk). Moreover, guidelines on how to make data findable, accessible, interoperable, and usable have been established (known as FAIR principles; see David et al., 2020; Wilkinson et al., 2016).

One of the most important findings from our survey is that 96% of participants expressed their interest in additional training in open science practices and methodology of behavioral research. In particular, the majority of researchers indicated a further interest in topics such as meta-analysis and data synthesis, preregistration, reporting of results, questionable research practices, sample size determination, and data sharing. Further materials on methodology, statistics, and open science practices are becoming available in the Slovak language as well. For example, the Slovak reproducibility network (https://www.slovakrn.org/) offers various regular workshops and materials that are freely available online, with more to come in the following years.

Study Limitations

This survey has several limitations. Firstly, due to the small sample size, the precision of the estimates is limited. This was because the survey was done in a small country with a very limited population of psychology researchers, lecturers, and postgraduate students. A similar problem is the self-selection of participants that could have resulted in a sample of researchers who are more interested in statistics, methodology, and open practices. This could have positively inflated the results (e.g., higher awareness of the topics, better scores on the test, fewer statistical misunderstandings) and the real situation could be worse to an unknown extent. Another possible limitation is that participants could have searched for answers to some of the questions online as this has not been controlled for (e.g., by using software that prevents opening a browser), and none of the participants were asked about it either (e.g., whether they searched for the answers online). We recommend paying attention to this aspect in future research.

Secondly, due to a technical error, the original questionnaire was irreversibly deleted resulting in loss of a part of the responses. Because of that, the data collection was stopped and restarted several months later. We had reconstructed the original questionnaire based on our notes and continued the data collection using a different software. This has resulted in small inconsistencies between the two versions of the survey (e.g., the items "informed_consent" and "methodology_statistics_interest" were added. There is also the possibility of omission of some options in the multiple-choice questions). However, we tried to recreate the original questionnaire and reconcile all possible differences to minimize inconsistencies. It is also possible that some participants could have taken the survey twice. However, since we controlled for duplex responding (participants were asked to indicate if they participated in this survey before) and got no positive answers, we assume that none did.

Conclusions

The aim of this study was to describe the extent of statistical and methodological misconceptions in Slovak psychology researchers as well as their opinions and attitudes toward open science practices. Misunderstandings about NHST and *p*-value, which have been mentioned in the literature for years, are still common among Slovak researchers. On the other hand, the participants showed largely positive attitudes towards open science practices and expressed interest in future training. The barriers preventing Slovak psychology researchers from greater adherence to open science practices could be partly attributed to lack of information. These present results imply that there is a space and potential for improvement and that researchers' demand for training in methodology and open science practices should be facilitated, especially in the current digital age.

Acknowledgements

This study was supported by project KEGA-079UK-4/2021 and project PRIMUS/20/HUM/009.

Autors' ORCID
Jakub Rajčáni
https://orcid.org/0000-0001-6407-9916
Lenka Vargová
https://orcid.org/0000-0002-8758-0997
Matúš Adamkovič
https://orcid.org/0000-0002-9648-9108
Pavol Kačmár
https://orcid.org/0000-0003-0076-1945

References

Abele-Brehm, A. E., Gollwitzer, M., Steinberg, U., & Schönbrodt, F. D. (2019). Attitudes toward open science and public data sharing. *Social Psychology*. https://doi.org/10.1027/1864-9335/a000384

Aczel, B., Palfi, B., Szollosi, A., Kovacs, M., Szaszi, B., Szecsi, P., ... & Wagenmakers, E. J. (2018). Quantifying support for the null hypothesis in psychology: An empirical investigation. *Advances in Methods and Practices in Psychological Science, 1*(3), 357–366. https://doi.org/10.1177/2515245918773742

Agnoli, F., Wicherts, J. M., Veldkamp, C. L. S., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLOS ONE, 12*(3), e0172792. https://doi.org/10.1371/journal.pone.0172792

Alderson, P. (2004). Absence of evidence is not evidence of absence. *BMJ, 328*(7438), 476–477. https://doi.org/10.1136/bmj.328.7438.476

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *BMJ, 311*(7003), 485. https://doi.org/10.1136/bmj.311.7003.485

American Psychological Association. (2016). Revision of Ethical Standard 3.04 of the" Ethical Principles of Psychologists and Code of Conduct"(2002, as amended 2010). *The American Psychologist, 71*(9), 900. https://doi.org/10.1037/amp0000102

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*, 305–307. https://doi.org/10.1038/d41586-019-00857-9

Badenes-Ribera, L., Frías-Navarro, D., Monterde-i-Bort, H., & Pascual-Soler, M. (2015). Interpretation of the p value: A national survey study in academic psychologists from Spain. *Psicothema*, 290–295. https://doi.org/10.7334/psicothema2014.283

Badenes-Ribera, L., Frias-Navarro, D., Iotti, B., Bonilla-Campos, A., & Longobardi, C. (2016). Misconceptions of the p-value among Chilean and Italian Academic Psychologists. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01247

Baker, M. (2016). Reproducibility crisis. *Nature, 533*(26), 353–66. https://doi.org/10.1038/533452a

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature, 483*(7391), 531–533. https://doi.org/10.1038/483531a

Bowman, S. D., DeHaven, A. C., Errington, T. M., Hardwicke, T. E., Mellor, D. T., Nosek, B. A., & Soderberg, C. K. (2016). OSF Prereg Template. Retrieved from https://osf.io/preprints/metaarxiv/epgjd

Carp, J. (2012). The secret lives of experiments: Methods reporting in the fMRI literature. *Neuroimage, 63*(1), 289–300. https://doi.org/10.1016/j.neuroimage.2012.07.004

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 1*. https://doi.org/10.1038/s41562-018-0399-z

Chambers, C. (2019). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice* (2nd ed.). Princeton University Press

Creative Commons (2022, June 30th). https://creativecommons.org/about/cclicenses/

Cumming, G. (2013). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis (1st ed.). Routledge. https://doi.org/10.4324/9780203807002

Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American psychological association publication manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology, 64*(3), 138–146. https://doi.org/10.1111/j.1742-9536.2011.00037.x

Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., Lo, J., McMenamin, N., & Wilson, S. (2007). Statistical reform in psychology: Is anything changing? *Psychological Science, 18*(3), 230–232. https://doi.org/10.1111/j.1467-9280.2007.01881.x

David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., … & Alliance – SHAring Reward and Credit (SHARC) Interest Group, T. R. D. (2020). FAIRness literacy: The Achilles' heel of applying FAIR principles. *Data Science Journal, 19*(1), 32. http://doi.org/10.5334/dsj-2020-032

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Palgrave Macmillan.

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review, 25*(1), 207–218. https://doi.org/10.3758/s13423-017-1266-z

Errington, T. M., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Challenges for assessing replicability in preclinical cancer biology. *ELife, 10*. https://doi.org/10.7554/elife.67995

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. Social *Psychological and Personality Science, 7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Haller, H., & Kraus, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7*(1), 1–20.

Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science, 1*(1), 70-85. https://doi.org/10.1177/2515245917751886

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician, 60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Gewin, V. (2016). Data sharing: An open mind on open data. *Nature, 529*(7584), 117–119. https://doi.org/10.1038/nj7584-117a

Goodman, S. (2008). A dirty dozen: Twelve p-value misconceptions. *Seminars in Hematology, 45*(3), 135–140. https://doi.org/10.1053/j.seminhematol.2008.04.003

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

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

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*(2), 137–152. https://doi.org/10.1037/a0028086

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review, 2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … & Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–490. https://doi.org/10.1177/2515245918810225

Kohn, M. A., & Senyak, J. *Sample Size Calculators*. UCSF CTSI. 22 June 2022. Available at https://sample-size.net/confidence-interval-proportion/ [Accessed 25 June 2022].

Lakens, D. (2016). *Improving your statistical inferences (MOOC)*. https://www.coursera.org/learn/statistical-inferences

Lakens, D. (2021). The practical alternative to the p value is the correctly used p value. *Perspectives on Psychological Science, 16*(3), 639–648. https://doi.org/10.1177/1745691620958012

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, N. P. du, Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), s41562-016-0021–016. https://doi.org/10.1038/s41562-016-0021

Murray-Rust, P. (2008). Open data in science. *Nature Precedings*, 1-1. https://doi.org/10.1038/npre.2008.1526.1

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*(2), 241. https://doi.org/10.1037/1082-989X.5.2.241

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. https://doi.org/10.1073/pnas.1708274114

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Nuzzo, R. (2015). Fooling ourselves. *Nature, 526*(7572), 182. https://doi.org/10.1038/526182a

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Perezgonzalez, J. D. (2015). Fisher, Neyman-Pearson or NHST? A tutorial for teaching data testing. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00223

Preregistration Task Force. (2020). Preregistration Standards for Psychology – the Psychological Research Preregistration-Quantitative (aka PRP-QUANT) Template. ZPID (Leibniz Institute for Psychology). http://dx.doi.org/10.23668/psycharchives.4584

Ropovik, I. (2017). O význame p-hodnôt: Reflexia na silnejúcu kritiku testov významnosti. *Československá Psychologie, 61*(5), 502–516.

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

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review, 25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100*(3), 426–432. https://doi.org/10.1037/a0022790

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6*(3), 291–298. https://doi.org/10.1177/1745691611406923

Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PloS one, 6*(11), e26828. https://doi.org/10.1371/journal.pone.0026828

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. 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*(1), 160018. https://doi.org/10.1038/sdata.2016.18