Received: 2 November 2018 | Revised: 18 December 2018 | Accepted: 6 January 2019

DOI: 10.1002/gepi.22191

RESEARCH ARTICLE

WILEY Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

Presidential address: Six open questions to genetic epidemiologists

Inke R. König

Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany

Correspondence
Inke R. König, Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Lübeck 23562, Germany.
Email: Inke.Koenig@imbs.uni-luebeck.de

Abstract

Given the rapid pace with which genomics and other -omics disciplines are evolving, it is sometimes necessary to shift down a gear to consider more general scientific questions. In this line, in my presidential address I formulate six questions for genetic epidemiologists to ponder on. These cover the areas of reproducibility, statistical significance, chance findings, precision medicine and related fields such as bioinformatics and data science. Possible hints at responses are presented to foster our further discussion of these topics.

KEYWORDS

data science, precision medicine, *p* values, reproducibility, significance

Dear colleagues and friends, I have spent a long time thinking about what to talk about today. From all the presidential addresses in the past few years, I learnt that this presentation should be on a high scientific level but must not be boring. It should give useful guidance and set standards for our research in genetic epidemiology, but should at the same time be entertaining and inspiring. Ideally, I would have liked to present you relevant messages or recommendations concerning our discipline, but nothing came to my mind. Therefore, I decided to turn the tables and formulate six questions to you instead. I am convinced that we, as a society of genetic epidemiologists, should address these questions in the near future.

1 | THE REPRODUCIBILITY CRISIS: *P* VALUES AND SIGNIFICANCE THRESHOLDS

In 2016, Nature published the results of a survey on 1,576 scientists from different disciplines (Baker, 2016). It was reported that 52% of the scientists believed that there is a significant reproducibility crisis, 38% believed in a slight crisis, and only 10% said that there was no crisis or did not answer the question. Focusing on those who are usually our collaborators, namely scientists from medicine or biology, 65% and 75%, respectively, reported to have failed to reproduce someone else's experiment, and 55% and 60% had even failed to reproduce an experiment of their own. Asked about how this situation could be improved, the majority of scientists agreed with the statements of using a "more robust experimental design" or "better statistics."

The call for improving the use of statistical methods was echoed in a number of publications that specifically commented on the problematic use of *p* values and significance thresholds (Nuzzo, 2014) and their common misinterpretations. Based on that, the American Statistical Association (ASA) took a drastic step by giving explicit recommendations in their "statement on statistical significance and *p*-values" (Wasserstein & Lazar, 2016). This covers the three issues of misinterpretation, overtrust, and misuses of *p* values.

Among the six principles they laid out, I would like to highlight one, namely "Scientific conclusions and business or policy decisions should not be based on only whether a *p*-value passes a specific threshold." As the supplementary

This is an open access article under the terms of the Creative Commons Attribution NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2019 The Authors. *Genetic Epidemiology* Published by Wiley Periodicals, Inc.

material, Greenland et al. (2016) gave details on common misinterpretations of *p* values and related concepts, citing Fisher (1956) on this specific point with "No scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas." A number of comments on the recommendations by the ASA were published subsequently (e.g., Wellek, 2017, and accompanying comments), emphasizing the problematic use of the fixed threshold of 0.05 as significance level.

We, therefore, see that a strict reliance on a fixed threshold is found to be problematic. Given this general background of the discussion in biostatistics, let me briefly summarize the use of significance levels in genetic epidemiology.

In classical linkage studies, we traditionally reject our null hypothesis of no linkage if the logarithm of the odds (LOD) score is greater than 3. This threshold was derived by Morton in the context of the sequential probability ratio test (Morton, 1955), that is, a sequential study design not used for this purpose any longer. Later on, the same threshold was further substantiated when the focus shifted to genome-wide linkage studies. Lander and Kruglyak (1995) published their notable guidelines for statistical significance, taking the multiple testing along the genome into account. Assuming that an infinitely dense marker map is used, they showed that a *LOD* score of 3.3 matches a global significance level of 5%. For a typical marker map, however, it was shown that this threshold again reduced to a critical *LOD* score of 3. This paper elicited some discussions. Specifically, Witte, Elston, and Schork (1996, see also Elston, 1997) disputed the use of thresholds at all by claiming "There is no longer much use for cutpoints, except to provide investigators, referees and editors an opportunity for mechanistic explanation of results under the false guise of objectivity." and "We would prefer that one simply present *P* values and educate the reader in how to interpret them, or else present posterior probabilities." However, this did not diminish the use of the *LOD* score threshold for claiming statistical significance in the following decades.

Instead, fixed thresholds were defined with a similar stringency for genome-wide association studies (GWAs). Here, we typically reject our null hypothesis of no association if the $p < 5 \times 10^{-8}$. This threshold goes back to Risch and Merikangas (1996) who assumed the presence of 100,000 genes with an average of five diallelic genetic variants in each. Further assuming that both alleles are tested separately, this led to about one million tests to be performed for a GWA, and a Bonferroni correction for multiple testing on this number to achieve a global significance level of 5% resulted in a local significance level of 5×10^{-8}. Once GWAs were performed in practice, a number of publications derived global significance levels taking the number of single-nucleotide polymorphisms and linkage disequilibrium correctly into account (The Wellcome Trust Case Control Consortium, 2007; Dudbridge & Gusnanto, 2008; Duggal, Gillanders, Holmes, & Bailey-Wilson, 2008; Gao, Starmer, & Martin, 2008; Pe'er, Yelensky, Altshuler, & Daly, 2008), thus leading to slightly less drastic significance levels in the range of 5×10^{-7} (The Wellcome Trust Case Control Consortium, 2007) to 7.2×10^{-8} (Dudbridge & Gusnanto, 2008). However, these were not widely adopted in practice, but a review on GWAs identified two current practices. First, claims for genome-wide significance are mostly based on the local significance level of 5×10^{-8} (Jannot, Ehret, & Perneger, 2015); and second, whenever this strict threshold is not reached, another, more liberal, threshold is used to select genetic variants to take forward for further scrutiny. This is usually defined ad hoc and without sound statistical justification. This general strategy has recently been extended to other situations with similar multiple testing challenges such as genome-wide interaction studies (Becker, Herold, Meesters, Mattheisen, & Baur, 2011), sequencing-based studies (Pulit, de With, & de Bakker, 2017) or epigenome-wide studies (Saffari et al., 2018). It should be noted that these publications provide statistically sound correction methods for the respective multiple testing problems, helping to enforce the reproducibility of the results (see Question #3). On the downside, fixed thresholds are again suggested for binary decisions. Therefore, I would like to pose my

Question #1: How do we deal with p values and significance thresholds in genetic epidemiology?

Instead of providing answers to the questions, Table 1 shows possible cheat sheets that might assist in drafting responses. In this case, we can refer to a number of current recommendations from the literature (Ioannidis, 2018). For example, Benjamin et al. (2018) recommended that the usual threshold for statistical significance be generally lowered from 5% to 0.5%, which has been dismissed by others (e.g., Trafimow et al., 2018). In other publications, it was suggested not to use *p* values anymore but rely on alternative methods instead. Interestingly, there is a Bayesian line of argumentation to justify the *LOD* score threshold of 3, showing that under a number of assumptions, this threshold yields a false-positive frequency of about 5% (see also Khoury, Beaty, & Cohen, 1993). Also, a number of approaches have been proposed for alternative methods in genetic epidemiology more recently (e.g., Strug, 2018). Another recommendation is not to rely on thresholds as such but to use and interpret *p* values differently as "graded measures of the strength of evidence against H_0"

TABLE 1 Six open questions to genetic epidemiologists

No.	Question	Cheat sheet
1	How do we deal with *p* values and significance thresholds in genetic epidemiology?	• Lower the threshold from 0.05 to 0.005 • Get rid of *p* values and use alternative methods • Get rid of thresholds, but interpret *p* values as "graded measures of the strength of evidence against H_0" • Rely on accumulation of evidence
2	How can we ensure methods reproducibility?	• Provide (toy) free data • Recognize simulation as science • Acknowledge contributions, e.g., through citing shared code • Enforce collaborations
3	What does results reproducibility mean to us?	• Distinguish terms taking genetic heterogeneity into account • Acquire more knowledge of population differences
4	How can we guard against the possible flood of chance findings?	• Work against publication bias • Introduce standard disclosure such as "We report how we determined our sample size, all data exclusions (if any), all manipulations and all measures in the study." • Preregister all studies, at least those not merely exploring data
5	What is the role of genetic epidemiology in precision medicine?	• Fill tracks with meaningful methods • Take responsibility for your methods and results • Think about how clinically useful your results are • Communicate your methods and results
6	Are we all bioinformaticians/computational biologists/data scientists? Do we differ from other disciplines? If so, what makes us special?	• Know your special skills (e.g., study designs, family-based studies, population-based research, evaluation of clinical utility) • Seize opportunities to learn from others since genetics develops quickly

(Fisher, 1956) or as "measure of surprise: the smaller it is, the more surprising the results are if H_0 is true" (Amrhein, Korner-Nievergelt, & Roth, 2017). Finally, many researchers have emphasized the need to accumulate evidence across studies. This suggestion is appealing but leads back to our starting point of the reproducibility crisis.

2 | THE REPRODUCIBILITY CRISIS: METHODS REPRODUCIBILITY

To think about the accumulation of evidence across studies, it is worthwhile thinking about what we mean by "reproducibility." Useful definitions of this have been given by Goodman, Fanelli, and Ioannidis (2016) distinguishing between methods reproducibility, results reproducibility, and inferential reproducibility. Of these, methods reproducibility means that experimental and computational procedures are exactly implemented as before, with the same results. To achieve this, recommendations have been given to journals proposing to check the reproducibility as part of the publication process; and to funding agencies stressing the importance to enforce respective programs (Stodden et al., 2016). More important in our context, recommendations for researchers include the sharing of data, code, and workflows, in open-trusted repositories (Stodden et al., 2016). Focusing on these, a number of obstacles are obvious. For example, a recent study on Wellcome trust researchers (van Den Eynden et al., 2016) revealed five hurdles for the sharing of data that (a) data may be disclosive and contain confidential information; (b) third party rights and participant permission need to be respected; (c) data may be misused or misinterpreted; (d) sharing data might jeopardize future publication opportunities; and (e) time and effort are required to prepare and deposit data. Additional barriers to share programming code included (a) the time, funding, and skills required to prepare code for sharing; and (b) the challenge to make the code sustainable (van Den Eynden et al., 2016).

Facing these obstacles leads to my

Question #2: How can we in genetic epidemiology ensure methods reproducibility given the obvious hurdles? How much effort do we want to invest to ensure methods reproducibility?

On the cheat sheet (Table 1), it may be suggested to always provide free data, even if these are just artificial data sets, to allow others to use and check the code on. Producing meaningful artificial data that is similar enough to real-life scenarios is extremely difficult, and simulation of data should be recognized as a science in itself (Chen et al., 2015; König et al., 2016). If data and or code provided by others is used, acknowledging that adequately is important. Finally, encouraging collaborations from the beginning of projects might facilitate a reciprocal check of the methods reproducibility.

3 | THE REPRODUCIBILITY CRISIS: RESULTS REPRODUCIBILITY

Further subdomains of reproducibility as defined by Goodman et al. (2016) include results reproducibility, meaning corroborating results in a new study with the same experimental methods, and inferential reproducibility, stating that knowledge claims are strengthened through replication or reanalysis.

But what, exactly, does it mean to corroborate results in a new study in genetic epidemiology? For the replication of genetic associations, standards have been established in 2007 by a dedicated working group that are well adapted in the field (Chanock et al., 2007). Among the criteria listed, it is stated that a "similar population" is assumed for replication. Given the nature of genetic data, this raises the question of the definition of a "similar population."

In this context it may be helpful to distinguish between two scenarios (Igl, König, & Ziegler, 2009; König, 2011): In the first scenario, the second sample that is assessed to reproduce the results is drawn from the same population as the original sample, and there is only random variation between the samples. If the results can be corroborated, we may term this "replication." In the second scenario, the populations differ systematically, for example, through ethnic differences. If in this case, we fail to reproduce the original findings, this does not necessarily mean that replication failed, because discordant results can be due to the differences in the underlying populations. If, however, we are able to reproduce the results, the associations are more generalizable, and we may term this "validation." Keeping this in mind may be helpful, but in specific situations, it may not be obvious whether the populations of original and second samples are similar enough to expect replication.

When underlying genetic associations are used to construct polygenic risk scores, a number of recent examples show that the distribution of the scores widely differ between populations from different continents. This has been demonstrated, for instance, for polygenic risk scores for schizophrenia (Curtis, 2018), type 2 diabetes, and coronary heart disease (Reisberg, Iljasenko, Läll, Fischer, & Vilo, 2017). As a consequence, applying genetic risk cut-offs derived from one population to individuals from another population is virtually useless, even taking different prevalence rates into account. These differences between continents are rather obvious; to what degree this is also true for populations that are seemingly more similar because they are from the same continent, remains to be investigated.

Therefore, given that specification of a similar population may not be straightforward, let us pose the following

Question #3: What does results reproducibility mean to us?

On the cheat sheet, it might be helpful to always be precise in what to expect taking genetic stratification and heterogeneity into account. This emphasizes that we probably need to understand possible population differences even in seemingly similar populations much better than we do now.

4 | THE REPRODUCIBILITY CRISIS: CHANCE FINDINGS

We see that in some instances, we may not be able to reproduce a finding of a genetic association because our populations were not sufficiently similar. In other instances, reproducibility may be compromised by chance findings, and a number of specific problems have been described that lead to false-positive results (Forstmeier, Wagenmakers, & Parker, 2017). These include first the seeking for novelty, meaning that there is a preference for new and unexpected results. Another problem has been coined "hypothesizing after the results are known" or "HARKing" as opposed to defining hypotheses before performing the actual study in a transparent way. Related to that are the various flavors of multiple testing that include using multiple analysis strategies, modeling, and testing. Obviously, these problems are aggravated if study methods and results are reported only selectively.

In genetic epidemiology, multiple testing has long been an issue. The challenge of false-positive findings in candidate gene association studies has been described frequently (Ioannidis, Ntzani, Trikalinos, & Contopoulos-Ioannidis, 2001). However, it has been acknowledged that the problem of false-positive genetic association results has been remedied by the introduction of GWAs, where the multiplicity is well-controlled and sample sizes are usually large by now (Ioannidis, 2018).

A similar challenge is faced in classical epidemiological association studies, where a multitude of nongenetic risk factors are being investigated for association with disease-relevant phenotypes. For this problem, it has been claimed that the risk of false-positive classical association results may be reduced by using Mendelian randomization studies, given that these can evaluate the causality of the investigated risk factors.

Basically, in Mendelian randomization studies (Davey Smith & Ebrahim, 2003; Grover, Del Greco, Stein, & Ziegler, 2017; Ziegler, Mwambi, & König, 2015), one or a number of genetic variants are identified that can act as instrumental variables for the risk factor in that they are associated with the risk factor but have no direct effect on the disease. Since these are genetic variables, they have two distinct advantages: First, the alleles of the variants are distributed randomly at conception, so that, in effect, the predisposition for the risk factor is distributed randomly, thus approximating an experimental setting. Second, since genetic variants are stable from conception on, they always precede other possibly confounding factors. This allows any relationship between these genetic variants and disease to be interpreted as evidence for causality of the risk factor on the phenotype, if the following assumptions are fulfilled: (a) The genetic variants are associated with the risk factor; (b) There is no association between the genetic variants on the one hand and confounders of the relationship between risk factor and the disease on the other; (c) Conditioning on the risk factor and possible confounders, there is no direct association between the genetic variants and the disease. In summary, this study design is an extremely interesting approach to investigate causality. However, it does not protect against the production of false-positive results, because, again, any number of risk factors can be investigated as long there is genetic data that can be used as an instrumental variable. If we want to control the multiplicity of hypotheses, we would have to define the entire space of risk factors, analogously to considering the entire genome in a GWA. Instead, current practice tends to do the opposite. For example, preliminary results of a systematic review on Mendelian randomization studies for neurodegenerative disorders show that more than 80% of the publications only report the investigation of one single risk factor (Grover, Del Greco M., & König, 2018), resembling the pick of single candidate genes in the pre-GWAs era.

This leads me to my

Question #4: How can we guard against the possible flood of chance findings?

It should be noted that this certainly does not only pertain to Mendelian randomization studies with multiple risk factors, but is a more general issue based on the problems mentioned above.

The first item on the cheat sheet (Table 1) is a hard nut to crack. To work against the publication bias, researchers need to try to publish null findings, and reviewers and editors should acknowledge the scientific soundness, not necessarily the novelty of data. An interesting suggestion was to introduce a standard disclosure to all publications along the lines of "We report how we determined our sample size, all data exclusions (if any), all manipulations and all measures in the study." (Simonsohn, Nelson, & Simmons, 2014). More strictly, it has been recommended to preregister all studies, at least for those that want to go beyond merely exploring the data. If it has to be stated in advance which data are assessed and how the data is analyzed, which is standard in clinical trials, cherry-picking and selective reporting of seemingly interesting data will be more difficult. In that vein, it has already been shown that preregistration leads to the report of more null findings, thus can be an effective measure against publication bias (Allen & Mehler, 2018; Warren, 2018). Further advantages of this, but also the challenges, are succinctly described in the recent literature (Nosek, Ebersole, DeHaven, & Mellor, 2018).

5 | HOT TOPICS: PRECISION MEDICINE

For the two remaining questions, I would like us to look forward, where it is important to consider where to put genetic epidemiology in the landscape of newly evolving topics, the first of these being precision medicine.

The general idea of precision medicine is that large-scale data is assessed in patients and nondiseased subjects on different levels of assessment including genomics and other -omics sources (see Figure 1; König, Fuchs, Hansen, von Mutius, & Kopp, 2017). This wealth of data that may fulfill the criteria of big data is then processed on different tracks with different dedicated purposes. First, the data may be mined to detect previously unknown disease subtypes. Second, diagnostic and prognostic models may be developed and validated, and third, treatment response may be predicted. Precision medicine in this definition is a process, because validated categories and models are fed back to the phenotyping stage to define the subsequent assessment of subjects. In order for precision medicine to be successful, validated models need to be disseminated and communicated, providing accessible and easy-to-use algorithms for clinical practice.

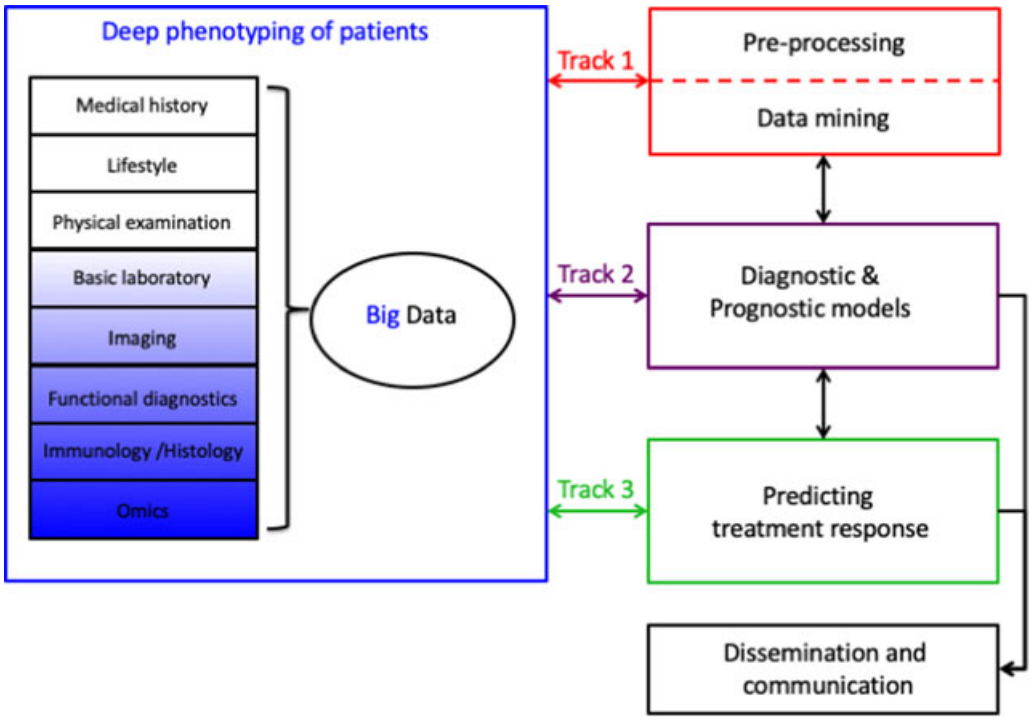

FIGURE 1 Process of precision medicine. In the *deep phenotyping* stage, information on patients is gathered on different levels. The color shading indicates that the more voluminous and complex the data set becomes, the more likely it is to meet the presupposition for precision medicine and Big Data. Data is then forwarded for further analysis to tracks 1–3. In *Track 1*, data is preprocessed including variable selection and mined for unknown structure. In *Track 2*, variables from the previous stages may be used to develop and validate diagnostic and prognostic models. Clinical relevance of these models may be investigated in studies showing the effect of the implementation of the models or by forwarding the models to track 3. In *Track 3*, specific models are developed and validated that aim at predicting treatment response partly building on previously developed models. Results from tracks 1–3 are fed back to the deep phenotyping stage to define subsequent assessment of patients. Models from Tracks 2 and 3 need to be disseminated and communicated providing accessible and easy-to-use algorithms for clinical practice. Reproduced with permission of the ©ERS 2018: *European Respiratory Journal*, 2016, *48*, 664–673. DOI: 10.1183/13993003.00436-2016

Given this definition of precision medicine leads to my

Questions #5: What is the role of genetic epidemiology in precision medicine?

Currently, genetic epidemiology already participates in the process of precision medicine by providing methods and models (López de maturana, Pineda, Brand, Van Steen, & Malats, 2016; Milani, Leitsalu, & Metspalu, 2015; Thomas, 2017) mostly for the development and validation for diagnostic and prognostic models using genomic data (Table 1). However, it might not always be obvious that by providing methods along these tracks, we also participate in finally providing algorithms for clinical practice, and responsibility for this should be acknowledged. Thinking about how useful results may be for patients and nondiseased individuals may also be necessary. Finally, precision medicine is only successful if models are communicated not only to medical co-operation partners but also to patients, and methodological researchers may contribute to that by making approaches and results as understandable and transparent as possible.

6 | HOT TOPICS: BIOINFORMATICS, COMPUTATIONAL BIOLOGY, AND DATA SCIENCE

Finally, a number of novel disciplines have been emerging over the past years that are more or less closely related to our discipline of genetic epidemiology. For instance, computational biology has been defined as "The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems" (Huerta, Haseltine, Liu, Downing, & Seto, 2000), which might in many practical applications show overlap with genetic epidemiological work. Similarly, the same report defines bioinformatics as "Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data" (Huerta et al., 2000; see also Luscombe, Greenbaum, & Gerstein, 2001). At last, data science "seeks to contextualize and establish best

practices for harnessing the recent and continuing advances in computing infrastructure, data processing, and analysis. (...) These infrastructures are necessary to access, analyze, and rapidly manage cross-scale big data in biomedical research..." (Zhu & Zheng, 2018).

Obviously, the borders between these disciplines and our field are blurred. Thus, without going into details, this again puts us into the position to sharpen our scope (Ziegler & König, 2014), leading to my

Question #6: Are we all bioinformaticians /computational biologists /genetic data scientists now? Or do we differ from the other disciplines? If so, what makes us special?

And I really do not have good hints to answer this, but I think that it is good to know our special skills. Thankfully, the comments and tweets after my presentation reminded me of "Genetic Epidemiology with a capital "E"" (Thomas, 2000, 2012), and being proficient in study designs, population-based research, family-based methods, and evaluation of individual risk as well as clinical utility, just to name a few, is likely to be part of the answer. Looking beyond our specific current scope, we should keep in mind how quickly our field is developing, and so seize the opportunity to learn from others.

ACKNOWLEDGMENTS

I am grateful to my colleagues at the Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, for their valuable feedback and constructive critique of my presentation.

ORCID

Inke R. König http://orcid.org/0000-0003-0504-6465

REFERENCES

Allen, C, Mehler, DMA. 2018. Open science challenges, benefits and tips in early career and beyond. Preprint at psyarxiv.com/3czyt

Amrhein, V., Korner-Nievergelt, F., & Roth, T. (2017). The earth is flat ($p > 0.05$): Significance thresholds and the crisis of unreplicable research. *PeerJ*, *5*, e3544.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*, 452–454.

Becker, T., Herold, C., Meesters, C., Mattheisen, M., & Baur, M. P. (2011). Significance levels in genome-wide interaction analysis (GWIA). *Annals of Human Genetics*, *75*, 29–35.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., ... Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*, 6–10.

Chanock, S. J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D. J., Thomas, G., ... Collins, F. S. (2007). Replicating genotype-phenotype associations. *Nature*, *447*, 655–660.

Chen, H. S., Hutter, C. M., Mechanic, L. E., Amos, C. I., Bafna, V., Hauser, E. R., ... Feuer, E. J. (2015). Genetic simulation tools for post-genome wide association studies of complex diseases. *Genetic Epidemiology*, *39*, 11–19.

Curtis, D. (2018). Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia. *Psychiatric Genetics*, *28*, 85–89.

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology*, *32*, 1–22.

Dudbridge, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology*, *32*, 227–34.

Duggal, P., Gillanders, E. M., Holmes, T. N., & Bailey-Wilson, J. E. (2008). Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics*, *9*, 516.

Elston, R. C. (1997). 1996 William Allan Award Address. Algorithms and inferences: The challenge of multifactorial diseases. *American Journal of Human Genetics*, *60*, 255–262.

van Den Eynden, V, Knight, G, Vlad, A, Radler, B, Tenopir, C, Leon, D, ... Corti, L. 2016. Towards open research: Practices, experiences, barriers and opportunities. Wellcome Trust, Retrieved from researchonline.lshtm.ac.uk/3332696/

Fisher, R. (1956). Statistical methods and scientific inference, Edinburgh: Oliver and Boyd.

Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings—A practical guide. *Biological Reviews of the Cambridge Philosophical Society*, *92*, 1941–68.

Gao, X., Starmer, J., & Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology*, *32*, 361–369.

Goodman, S. N., Fanelli, D., & Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Science Translational Medicine*, *8*, 341ps12.

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*, 337–50.

Grover, S., Del Greco M. F., Stein, C. M., & Ziegler, A. (2017). Mendelian randomization. *Methods in Molecular Biology*, *1666*, 581–628.

Grover, S., Del Greco M. F., & König, I. R. (2018). Evaluating the current state of Mendelian Randomization studies: A protocol for a systematic review on methodological and clinical aspects using neurodegenerative disorders as outcome. *Systematic Reviews*, *7*, 145.

Huerta, M., Haseltine, F., Liu, Y., Downing, G., & Seto, B. (2000). NIH working definition of bioinformatics and computational biology. *Biomedical Information Science & Technology Initiative and National Institutes of Health*. Retrieved from. 2digitstechcom.ipage.com/uploads/2/9/0/1/2901227/compubiodef.pdf

Igl, B.-W., König, I. R., & Ziegler, A. (2009). What do we mean by "replication" and "validation" in genome-wide association studies? *Human Heredity*, *67*, 66–68.

Ioannidis, J. P. A. (2018). The proposal to lower *p* value thresholds to .005. *Journal of the American Medical Association*, *319*, 1429–1430.

Ioannidis, J. P. A., Ntzani, E. E., Trikalinos, T. A., & Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nature Genetics*, *29*, 306–309.

Jannot, A. S., Ehret, G., & Perneger, T. (2015). $P < 5 \times 10^{-8}$ has emerged as a standard of statistical significance for genome-wide association studies. *Journal of Clinical Epidemiology*, *68*, 460–465.

Khoury, M. J., Beaty, T. H., & Cohen, B. H. (1993). *Fundamentals of genetic epidemiology*. New York, NY: Oxford University Press.

Konig, I. R. (2011). Validation in genetic association studies. *Briefings in Bioinformatics*, *12*, 253–258.

König, I. R., Auerbach, J., Gola, D., Held, E., Holzinger, E. R., Legault, M.-A., … Yang, H.-C. (2016). Machine learning and data mining in complex genomic data—A review on the lessons learned in Genetic Analysis Workshop 19. *BMC Genetics*, *17*, 1.

König, I. R., Fuchs, O., Hansen, G., von Mutius, E., & Kopp, M. V. (2017). What is precision medicine? *European Respiratory Journal*, *50*, 1700391.

Lander, E., & Kruglyak, L. (1995). Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. *Nature Genetics*, *11*, 241–247.

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, *40*, 346–358.

López de maturana, E., Pineda, S., Brand, A., Van Steen, K., & Malats, N. (2016). Toward the integration of Omics data in epidemiological studies: Still a "long and winding road". *Genetic Epidemiology*, *40*, 558–69.

Milani, L., Leitsalu, L., & Metspalu, A. (2015). An epidemiological perspective of personalized medizine: The Estonian experience. *Journal of Internal Medicine*, *277*, 188–200.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American Journal of Human Genetics*, *7*, 277–318.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of National Academy of Science of the United States of America*, *115*, 2600–2606.

Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, *506*, 150–152.

Pe'er, I., Yelensky, R., Altshuler, D., & Daly, M. J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genetic Epidemiology*, *32*, 381–385.

Pulit, S. L., de With, S. A. J., & de Bakker, P. I. W. (2017). Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genetic Epidemiology*, *41*, 145–151.

Reisberg, S., Iljasenko, T., Läll, K., Fischer, K., & Vilo, J. (2017). Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLOS One*, *12*, e0179238.

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*, 1516–1517.

Saffari, A., Silver, M. J., Zavattari, P., Moi, L., Columbano, A., Meaburn, E. L., & Dudbridge, F. (2018). Estimation of a significance threshold for epigenome-wide association studies. *Genetic Epidemiology*, *42*, 20–33.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology. General*, *143*, 534–47.

Stodden, V., McNutt, M., Bailey, D. H., Deelman, E., Gil, Y., Hanson, B., … Taufer, M. (2016). Enhancing reproducibility for computational methods. *Science*, *354*, 1240–1241.

Strug, L. J. (2018). The evidential statistical paradigm in genetics. *Genetic Epidemiology*, *42*, 590–607.

The Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*, 661–678.

Thomas, D. C. (2000). Genetic epidemiology with a capital "E". *Genetic Epidemiology*, *19*, 289–300.

Thomas, D. C. (2012). Genetic epidemiology with a capital E: Where will we be in another 10 years? *Genetic Epidemiology*, *36*, 179–82.

Thomas, D. C. (2017). What does "Precision Medicine" have to say about prevention? *Epidemiology*, *28*, 479–483.

Trafimow, D., Amrhein, V., Areshenkoff, C. N., Barrera-Causil, C. J., Beh, E. J., Bilgiç, Y. K., … Marmolejo-Ramos, F. (2018). Manipulating the alpha level cannot cure significance testing. *Frontiers in Psychology*, *9*, 699.

Warren, M. (2018). First analysis of "pre-registered" studies shows sharp rise in null findings. *Nature*. Retrieved from. https://www.nature.com/articles/d41586-018-07118-1

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *American Statistician*, *70*, 129–33.

Wellek, S. (2017). A critical evaluation of the current "p-value controversy". *Biometrical Journal*, *59*, 854–872.

Witte, J. S., Elston, R. C., & Schork, N. J. (1996). Genetic dissection of complex traits. *Nature Genetics*, *12*, 355–356.

Zhu, L., & Zheng, W. J. (2018). Informatics, data science, and artificial intelligence. *Journal of the American Medical Association*, *320*, 1103–1104.

Ziegler, A., & König, I. R. (2014). Celebrating the 30th anniversary of genetic epidemiology: How to define our scope? *Genetic Epidemiology*, *38*, 379–380.

Ziegler, A., Mwambi, H., & König, I. R. (2015). Mendelian randomization versus path models: Making causal inferences in genetic epidemiology. *Human Heredity*, *79*, 194–204.

How to cite this article: König IR. Presidential address: Six open questions to genetic epidemiologists. *Genet. Epidemiol.* 2019;43: 242–249. https://doi.org/10.1002/gepi.22191