EDITORIAL

If it is statistically significant does it make it clinically important?

S. Beecroft
Manchester Medical School, University of Manchester
simon.beecroft@student.manchester.ac.uk

In answering any question it is first imperative to define the key terms contained therein. The question that forms the title of this essay refers to two key concepts in the application of evidence to the practice of medicine: statistical significance and clinical importance. This essay considers the link between these two notions in the context of research consisting of clinical trials evaluating interventions in healthcare.

Of the two, statistical significance appears to be the more straightforward to define, as there is currently a broad consensus on the calculation of statistical significance, although the most common definition does have its detractors. The prevailing method of identifying statistical significance in hypothesis testing is the use of the *p* value.[1] The *p* value represents the probability of obtaining results as extreme as those observed if the null hypothesis were true. In this system, statistical significance is achieved when the *p* value is below an (arbitrary) defined significance level, typically set at 0.05 in medical research. It is occasionally argued that *p* values should be incorporated into a Bayesian calculus which includes prior probability[2] in order to better represent the continually evolving nature of scientific knowledge, but this is not yet common practice.

In the medical literature the terms clinical importance,[3] clinical relevance[4] and clinical significance[5] appear to be used interchangeably, and the inconsistent terminology seems to be associated with confusion in the definition of the concept. This is unfortunate as the different terms could be better used to circumscribe some distinct concepts. *Clinical significance* could be best used in a narrow sense to describe whether an identified effect size achieves the smallest effect of clinical interest. This could free up *clinical importance* to cover a more expansive set of issues related to the application of research findings to medical practice. In this instance, clinical importance would refer to the uptake of the intervention in real-world clinical practice. Thus the redefined concept of clinical importance should

Manchester Medical Journal, 2016 © The Authors.
Published by Manchester University Press, The University of Manchester Library and the Manchester Medical School
http://dx.doi.org/10.7227/MMJ.0003

MANCHESTER 1824
Manchester University Press

encompass clinical significance in the narrow sense described above, along with processes which affect the translation of knowledge from research to practice. In addition to adding clarity, having the broader term available could help align the lexicon of medical research with a more practical and patient-centred approach.

The process of defining the key terms enables an exploration of the process of generating clinically important research as defined here. Firstly, the reported statistically significant result must truly be statistically significant; there must have been no error in generating the result. Secondly, the statistically significant result must achieve the smallest effect of clinical interest, which must be defined appropriately. Finally, the intervention must be adopted and practised effectively by clinicians and patients. Running this gauntlet successfully is far from guaranteed for any research project, and as a result there are a multitude of situations in which reported statistical significance does not equate to clinical significance or importance. The many stages involved in the translation of research into clinical practice each have their own capacity for error which could reduce or nullify the clinical importance of the findings. Error may be introduced in the identification of the research question, the design and conduct of the study, the presentation of the results, the statistical analysis of the results, and the communication and application of the research findings. The following are brief elaborations of some of the key problems encountered at each stage, and this is necessarily not an exhaustive list but should serve to highlight some important pitfalls that can sever the connection between statistical significance and clinical importance.

Research question

There are a number of ways in which research can be poorly conceived that mean no matter how significant the results they are unlikely to be clinically important. Producing redundant work or work that does not address patient concerns or investigate outcomes important to patients is a recipe for wasted time and money. Ensuring that a statistically significant result would be useful to practising clinicians at the very least means ensuring one is not duplicating work that has already been done (except in the context of replication trials). Sadly, the process of undertaking a systematic review to identify the state of existing knowledge before embarking on a new research project is not universal.[6] Work that is genuinely new may also have limited clinical applicability if it fails to focus on what matters to patients.[7,8] There have been a number of initiatives aimed at improving patient involvement in setting research agendas, but progress is slow. Both of these errors end up effectively producing research of extremely limited clinical value; unfortunately the reward structure for medical research continues to incentivise this kind of work.

Study design and conduct

In spite of decades of collective experience of designing and conducting clinical trials, the medical research community continues to produce research that is methodologically flawed. A raft of issues in the design and conduct of medical research can lead to type I error, which is to reject the null hypothesis when it is true. A brief exposition of commonly encountered problems at this stage includes insufficiently powered studies, studies that do not appreciate the prior probability of detecting an effect and studies that investigate too many outcomes.[9] These are problems that should be identified by clinicians trained in critical appraisal of literature. However, constant vigilance on behalf of those involved in front-line care is not a particularly robust system for weeding out flawed research, and much research that is unable to demonstrate the relationships it reports may well be translated into changes in clinical practice.

Statistical analysis

The phenomenon of '*p*-hacking' has received increasing scrutiny in recent years as meta-research findings have begun to suggest that there is widespread inappropriate data manipulation occurring in scientific research.[10] There are a number of ways in which data may be treated to increase the likelihood of generating a statistically significant result. These methods include using interim analyses during data collection; post hoc selection of primary outcome measures; and modifications of the treatment of outliers, subgroup analyses, and inclusion and exclusion criteria.[11] These methods may be subtle and can be difficult to identify during critical appraisal of research papers, thus leading to unchallenged claims of statistical significance that are unfounded. Unfortunately, even when such discrepancies in research are identified there are a number of barriers to correcting the scientific record,[12] not least editors' inability or unwillingness to appreciate the presence of inappropriate treatment of data in research published in their journals.[13]

Communicating research findings

If clinicians are unable to access research findings then it is impossible for the insights to be implemented in clinical practice. The issue of knowledge translation[14] has been explored by researchers in medical education, and there are a number of barriers identified to closing the implementation gap. These challenges include the polar opposite issues of restricted access to information and information overload. Many medical journals require a paid subscription or apply a per article charge which is generally prohibitively expensive for those without institutional access. Whilst

open-access journals have become more successful recently, there is still a massive wodge of medical knowledge that is stuck on the other side of a paywall for many clinicians around the world.[15] On the other hand, the sheer volume of research being produced is a problem for practising clinicians who wish to provide the best care for their patients. When the amount of time required to read all relevant journals exceeds the number of waking hours in the day[16] this is clearly an issue pertinent to the uptake of new interventions. It appears as though continuing medical education will remain a piecemeal affair for the foreseeable future; a situation which is not conducive to maximising the clinical importance of the medical research that is conducted.

It seems valid in the context of assessing the clinical importance of statistically significant results to understand their relative contribution to the sum of knowledge. The totality of evidence available via a literature review is not generated as by a pure, disinterested process. The complex interplay of a number of different motivations on behalf of researchers, institutions, commercial interests and publishers[17] means that results that do not achieve statistical significance are less likely to be published than those that do.[18] This publication bias significantly affects the context in which new research findings are assessed. The interpretation of new research does not occur in isolation; it is incorporated into existing knowledge, whether formally in the setting of systematic review and meta-analysis, or informally through amalgamation with healthcare professionals' existing decision-making heuristics. Although tools exist to identify publication bias,[19] the preponderance of statistically significant findings in medical research affects the quality of this process and limits the utility of the research base as a whole.

Application of research findings

Even with clinicians who are up to date with the literature, there can be many more barriers to implementation of interventions in clinical practice. Clinicians' own motivation and/or willingness to modify their current practice are paramount in this regard. Resistance to the adoption of new interventions is dependent on the environment, personality, identity and cognitive architecture of those expected to implement it, which define their assessment of commitment to and capacity for the intervention.[20] Once the clinician has agreed to the change of practice, other issues come to the fore. Amongst others, access to the new intervention and any necessary training required to implement it, the cost-effectiveness of the intervention compared with existing therapeutic strategies, and local governance and funding arrangements will significantly impact on ability to modify practice.[21] Once these issues have been ironed out, the stage is finally set for the clinical use of the intervention. The final hurdle is patient acceptance

of or adherence to the intervention, which can be influenced by a number of different factors. Some factors relevant to the researcher planning to investigate an intervention they wish to become clinically important are regimen complexity, economic and structural factors, patient-related factors, and pattern of healthcare delivery.[22] If the proposed intervention is overly onerous for patients or they are unable to access the service appropriately or cope with practical aspects of receiving the intervention, then its clinical importance will be manifestly reduced.

Conclusion

In conclusion, the simple answer to the question of whether statistical significance invariably means clinical importance is an emphatic no. Statistically significant results may be clinically important but the association is by no means certain; as the old proverb has it, 'There's many a slip 'twixt cup and lip.' The decoupling of statistical significance and clinical importance is important not only when attempting to ensure a robust evidence base to be drawn upon when clinicians provide care and make shared decisions with patients, but also to protect those who participate in medical research as subjects. If the clinical importance of research findings is undermined, then the risk:benefit calculus for clinical trial participants is changed and any harm suffered by participants becomes objectively less justifiable. It is therefore incumbent on those engaged in medical research and practice to be aware of these issues and to do all that is reasonably practical to mitigate them.

References

1 Sterne JAC, Smith GD. Sifting the evidence: what's wrong with significance tests? Br Med J. 2001 Jan 27;322(7280):226–31.
2 Goodman SN. Of p-values and Bayes: a modest proposal. Epidemiology. 2001 May;12(3):295–7.
3 Mon-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KBY, et al. Determination of the clinical importance of study results: a review. J Gen Intern Med. 2002 Jun;17(6):469–76.
4 Bhardwaj SS, Camacho F, Derrow A, Fleischer AB, Feldmann SR. Statistical significance and clinical relevance: the importance of power in clinical trials in dermatology. Arch Dermatol. 2004 Dec;140(12):1520–3.
5 Sedgwick P. Clinical significance versus statistical significance. Br Med J [Internet]. 2014 [cited 2016 Feb 5]; 348:g2130. doi: 10.1136/bmj.g2130.
6 Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009 Jul 4;374(9683):86–9.
7 Perel P, Miranda JJ, Ortiz Z, Casas JP. Relation between the global burden of disease and randomized clinical trials conducted in Latin

America published in the five leading medical journals. PloS ONE [Internet]. 2008 [cited 2016 Feb 5]; 3(2):e1696. doi: 10.1371/journal.pone.0001696.
8 Crowe S, Fenton M, Hall M, Cowan K, Chalmers I. Patients', clinicians' and the research communities' priorities for treatment research: there is an important mismatch. Res Involv Engagem. 2015 Jun 25;1(1):1–10.
9 Ioannidis JPA. Why most published research findings are false. PloS Med. 2005 Aug 30;2(8):696–701.
10 Gadbury GL, Allison DB. Inappropriate fiddling with statistical analyses to obtain a desirable p-value: tests to detect its presence in published literature. PloS One [Internet]. 2012 [cited 2016 Feb 5]; 7(10):e46363. doi:10.1371/journal.pone.0046363.
11 Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. PloS Biol [Internet]. 2015 [cited 2016 Feb 5]; 13(3):e1002106. doi:10.1371/journal.pbio.1002106.
12 Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: a tragedy of errors. Nature. 2016 Feb 4;530(7588):27–9.
13 Slade E, Drysdale H, Goldacre B, COMPare Team. Discrepancies between prespecified and reported outcomes. Ann Intern Med [Epub ahead of print]. 2015 [cited 2016 Feb 5]; 164(5):374. doi: 10.7326/L15-0614.
14 Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? J Contin Educ Health Prof. 2006 Winter;26(1):13–24.
15 Zachariah R, Ford N, Maher D, Bissell K, Van den Bergh R, van den Boogaard W, et al. Is operational research delivering the goods? The journey to success in low-income countries. Lancet Infect Dis. 2012 May;12(5):415–21.
16 Alper BS, Hand JA, Elliott SG, Kinkade S, Hauan MJ, Onion DK, et al. How much effort is needed to keep up with the literature relevant for primary care? J Med Libr Assoc. 2004 Oct;92(4):429–37.
17 Dickersin K. Publication bias: recognizing the problem, understanding its origins and scope, and preventing harm. In: Rothstein HR, Sutton AJ, Borenstein M, editors. Publication bias in meta-analysis: prevention, assessment and adjustments. Chichester: John Wiley & Sons; 2005. p. 11–33.
18 Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev [Internet]. 2009 [cited 2016 Feb 5]; 1:MR000006. doi: 10.1002/14651858.MR000006.pub3.
19 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA. 2006 Feb 8;295(6):676–80.
20 Laws RA, Kemp LA, Harris MF, Davies GP, Williams AM, Eames-Brown R. An exploration of how clinician attitudes and beliefs influence

the implementation of lifestyle risk factor management in primary healthcare: a grounded theory study. Implement Sci. 2009 Oct 13;4:66.
21 Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol. 2008 Jun;41(3–4):327–50.
22 Sabaté E. Adherence to long-term therapies: evidence for action. Geneva: World Health Organization; 2003.