Essays

Discovering the value of a "failed" trial

Helen Speyer
Mental Health Centre Copenhagen, Denmark; helene.speyer@regionh.dk

DOI: 10.20316/ESE.2018.44.18004

Abstract
Increased focus on replicability in science has led to legislation and regulation to minimize the file drawer phenomenon. An alternative approach could be to encourage authors to write papers with impact rather than papers in high impact journals. Based on personal experience, this essay suggests a systematic framework developed to facilitate the extraction of valuable knowledge from a "failed" trial. First, "negative" results should be differentiated into inconclusive, neutral, negative and statistically significant but clinically irrelevant. Second, to avoid cherry-picking of references, systematic search should be performed when the results are integrated in current research. Third, acknowledging that the tested hypothesis might be wrong can initiate de-implementation in clinical practice and suggest that further research should look for an alternative approach.

Keywords: clinical trial, negative results, publishing, replication

Discovering the value of a "failed" trial

Editors have often been held responsible for the current replication crisis in science as they are assumed to prefer papers reporting positive results, which will lead to a higher rate of type 1 errors than the normally accepted risk of 5%. This publication bias could also be explained by the authors themselves, as they do not submit their negative findings. Authors state several specific reasons for not submitting negative findings; lack of enthusiasm, personal conflicts of interest or low expectations of acceptance.[1] If papers were published based on quality rather than direction of results, the replication crisis could be counteracted and science might progress more rapidly. Legislation and regulation, such as preregistration, might increase the rate of submissions of negative findings but will not affect the quality of the papers. Awareness of how negative results could be presented in an influential way might nurture the intrinsic motivation of authors. As a consequence, papers might be more read-worthy and increase the contribution to the scientific field. A discussion is required on how we can move focus from striving to get into high impact journals to writing papers with high impact.

In this essay, I will describe my personal experience with publishing negative findings. During the process, I developed a systematic approach to answer important questions based on my negative results. Hopefully, this framework can inspire editors to guide authors in writing up their negative findings in high quality papers, so they can counter the effects of publication bias.

The CHANGE trial was an ambitious clinical trial aiming to reduce premature mortality in patients with schizophrenia by offering affiliation to a lifestyle coach.[2] After three years of work on the trial, giving more than 20 enthusiastic talks, recruiting 428 patients and spending about 30,000 euro, we finally had the results. A few weeks earlier, a book had been published about the methods used by the coaches in the trial and *The Lancet Psychiatry* had invited us to submit the results. The results were 75 *p*-values exceeding 0.05, which is statistically very unlikely, as one in twenty should be positive just by coincidence when the null hypothesis is true.[3] At that point, I realised that finding my name in *Lancet Psychiatry* as well as invitations to exciting conferences would not be anything but a dream.

While digesting the disappointment, I kept alternating between four principal ways to deal with my negative findings. My first impulse was to store the file safely in the electronic drawer on the desktop of my computer. I had no difficulties arguing that my time was spent much better writing on something else, which would benefit not only me, but also the tax payers paying my salary. However, after the acute frustration had settled down, I decided that I would not contribute to publication bias this time. So, I started writing.

My second strategy was to devalue the whole trial, state that the results were inconclusive, and that if I received another very large grant I could conduct a new trial, address its shortcomings, and show that lifestyle coaching indeed works. However, I honestly could not come up with any revolutionary ideas that would markedly improve the design. Even today, three years later, I do think it was a well-designed and well conducted trial.

The third opportunity was to reframe the boring results to somewhat more uplifting (positive) findings. There are several ways to do that; change statistical model, use a different outcome, remove outliers or do subgroup analyses, just to mention a few. This process is colloquially known as HARKING (hypothesis after results are known), p-hacking or data massaging.[4] The key point is to keep on going until one gets a significant *p*-value, which will eventually come by chance. Alternatively, one can bend the interpretation, so p=0.09 would be "trend significant" and 0.1 would be "highly suggestive" and thus publish "positive" findings…

After a few weeks as a data massage therapist, I looked myself in the mirror and gave up. Even after sacrificing my identity as a rigorous researcher, no patterns of effect came through. I faced the dreadful reality: the most likely

explanation for the negative results was simply that the intervention did not work. Our great idea was not that great after all. In fact, it was probably wrong.

Framework for dealing with negative results

My mission from that point was to pull as much value as possible out of the trial. I wanted to write a paper truly worth reading and I wanted our results to contribute to the body of knowledge. I started reflecting on what questions we could actually answer, and came up with two: Is this research area a dead end? Should we adjust interventions we have implemented in practice? These two questions, namely *implications for practice and research,* ought to be addressed with the same enthusiasm as in trials with positive results. Three steps seemed necessary to approach a sound conclusion systematically.

1. Differentiate negative results

Negative results are often used as a common classification of results that could not reject the null-hypothesis. There are four types: i) inconclusive results are those that stem from underpowered trials, also called type II errors;[5] ii) neutral results stem from a trial with sufficient power to state that there is no effect; iii) true negative results are results that significantly affected the outcome of interest, but in the opposite direction to that which was hypothesized; iv) results that are statistically significant but clinically irrelevant, if the effect size is trivial.

Trials with inconclusive results due to small study size have limited value. One could ask whether underpowered trials should be conducted at all, if it could have been known in advance that they were underpowered. In our case, the study had sufficient power, demonstrated by a-priori power calculations, to rule out type II errors for primary and secondary outcomes. Bearing this in mind, as well as the high internal and external validity, we believe that our results were truly neutral.

2. Integrate results into current research

No matter how well conducted the study is, its results might still be due to random errors. Therefore, integration into the existing literature is crucial. Most authors use a few paragraphs to mention similar studies, sometimes selecting studies with the same conclusion(s). This is called confirmation bias (the urge to have confirmation that the world behaves like you expected it to behave). To avoid the temptation of cherry picking, we chose to do a systematic literature search and synthesize the results in a meta-analysis, including a funnel plot to address publication bias.

In the case of our research, lifestyle interventions for schizophrenia are already considered evidence based and recommended in the NICE guidelines.[6] However, our literature search showed that lifestyle interventions targeting people with severe mental illness do not affect metabolic outcomes such as blood pressure, glucose or cholesterol. Regarding weight, there was a statistically significant reduction of approximately 2kg, which was not sustained at 2-12 months follow up. We judged this small effect to be clinically irrelevant and concluded that our research findings were in line with the literature. The meta-analysis has not been published yet.

Looking at similar interventions in the general population, our conclusion was confirmed.[7] Individualised lifestyle interventions have a limited potential to reduce metabolic risk factors. Surprisingly, some studies even noted a slightly elevated risk of severe adverse events like death and stroke in the intervention groups[8,9] Denmark. Adults aged 30–60 years were randomly drawn from a population and were randomised to intervention group (n = 11,483). In fact, sociological and philosophical research has for years pointed at potential harmful effects of lifestyle interventions on general health, such as stress and stigma.[10]

3. Discuss implications for research and practice

The big question is whether now we have enough research data from this area to close the case. Based on the available literature, the scientific community will be able to judge and economic resources and academic interest might be allocated to interventions that might work.

To qualify the next step in research, we need to specify "what did not work?" In the inductive phase of a clinical trial, a hypothesis is developed based on what is already known on the topic. This process, and the exact hypothesis describing why we think X will affect Y, is rarely provided in detail in psychosocial research. In CHANGE and similar trials, the underlying hypothesis is based on the assumption that *people can and will adopt a healthier lifestyle, provided they have sufficient knowledge and support*. Or in other words, lifestyle pattern is a matter of willpower and freedom of choice. This assumption is likely to be wrong, as lifestyle is mainly a function of social determinants.[11] Acknowledging that the theoretical assumptions of a trial are wrong might be most frightening and humiliating for the author but could contribute to a paradigm shift in their beliefs.

The obvious implication from the results of our trial for clinical practice is that we should not offer this kind of intervention, as there is no beneficial effect. Systematic and mandatory screening for unhealthy lifestyle has already been implemented in Denmark in people with schizophrenia. The question is whether this type of screening is ethically correct. Screening makes sense only if an intervention is available that will improve patient outcome. As this seems not to be the case for lifestyle counselling for patients with schizophrenia, resources could better be invested in interventions with proven efficacy.

The systematic approach to identify the value of negative findings, nurture academic enthusiasm and facilitated the writing of a paper that was published in a leading psychiatric journal.[2]

Not all readers appreciate the promotion of results that do not fit public opinion. Both editors and reviewers might be allegiance and confirmation biased. Lifestyle promotion is a movement with devoted adherents; to state that it is probably ineffective is provocative. In fact, a reviewer claimed we were secretly promoting the pharma industry. Editors should be aware of these emotional pitfalls.

In the current system, editors are the gate-keepers of the publication process. They have the right and duty to publish well written papers. The quality of papers with initially disappointing results should be valued as much as those that confirm an a priori hypothesis.

The solution to publication bias is not to have journals dedicated to negative results or rules that push editors to accept negative findings. Instead, we should raise the quality of papers with negative results, so they are equally interesting to publish. Editors should encourage and guide authors to improve the quality of papers, so decision makers have a comprehensive picture of knowledge pertaining to their field.

References

1 Song F, Loke Y, Hooper L. Why are medical and health-related studies not being published? A systematic review of reasons given by investigators. *PLoS One*. 2014;9(10):e110418.

2 Speyer H, Christian Brix Nørgaard H, Birk M, Karlsen M, Storch Jakobsen A, Pedersen K, et al. The CHANGE trial : no superiority of lifestyle coaching plus care coordination plus treatment as usual compared to treatment as usual alone in reducing risk of cardiovascular disease in adults with schizophrenia spectrum disorders and abdominal obesity. *World Psychiatry*. 2016;15(June):155–65.

3 Gasparikova-Krasnec M, Ging N. Experimental Replication and Professional Cooperation. *American Psychologist*. 1987;42:266.

4 Wicherts JM, Veldkamp CLS, Augusteijn HEM, Bakker M, van Aert RCM, van Assen MALM. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*. 2016;7:1832.

5 Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, *et al*. Power failure: why small sample size undermines the reliability of neuroscience. Vol. 14, Nature reviews. *Neuroscience*; 2013. p. 365–76.

6 Kuipers E, Yesufu-Udechuku A, Taylor C, Kendall T. Management of psychosis and schizophrenia in adults: summary of updated NICE guidance. *BMJ*. 2014;348:g1173.

7 Jorgensen T, Jacobsen RK, Toft U, Aadahl M, Glumer C, Pisinger C. Effect of screening and lifestyle counselling on incidence of ischaemic heart disease in general population: Inter99 randomised trial. *BMJ*. 2014;348:g3617.

8 Bender AM, Jørgensen T, Pisinger C. Do high participation rates improve effects of population-based general health checks? *Preventative Medicine* (Baltim). 2017;100:269–74.

9 Skaaby T, Jorgensen T, Linneberg A. Effects of invitation to participate in health surveys on the incidence of cardiovascular disease: a randomized general population study. International Journal of Epidemiology. 2017;46(2):603–11.

10 Holtug N, Kongsholm N, Lægaard S, Nielsen MEJ. Etik i forebyggelse og sundhedsfremme [Internet]. Sundhedsstyrelsen. 2009. 59 p. Available from: http://www.sst.dk/Udgivelser/2009/Etik i forebyggelse og sundhedsfremme.aspx

11 Marmot M. Economic and social determinants of disease. *Bulletin of the World Health Organization*. 2001;79(10):988–9.