

thesis (questions are put down in writing on stickers provided by the trainers).

- Trainers should announce that some issues – especially research ethics issues – do not have simple black and white answers and that the main point is to become familiar with ethical thinking and deliberation.
- Computers and cell-phones should be banned so that everyone can be involved in the debates.

Feedback from doctoral candidates about the training includes some very isolated but interesting comments:

- The dilemma exercise, which aims at creating ethical deliberation through the confrontation of moral values, can be regarded as aggressive and provoke rejection. On the contrary, it can be judged irrelevant because it is considered as being far from reality or pointless.
- Philosophical and historical references can provoke sharp or even hostile reactions, depending on each person's culture and beliefs.

Nevertheless, the overwhelming majority of feedback is positive, strengthening the pedagogical approach that has been adopted i.e., a face-to-face doctoral training in research ethics and integrity, in small classes, based on debates that are focused on the issues that are raised by the doctoral candidates themselves. Most doctoral candidates report that they have become aware of best practices concerning publications and experiments, that they have learnt to consider their work from an ethics point of view, and that they have appreciated being able to raise issues about their own theses. Clearly this costs money and the recruitment and remuneration policy of the trainers is an integral element of the (ethical) debate about this kind of training.

References:

- [1] Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene (CERNA), "Proposition de formation doctorale – Initiation à l'éthique de la recherche scientifique", 2018. [Online]. Available (in French): http://cerna-ethics-allistene.org/digitalAssets/55/55709_Cahier_CERNA_FormationDoctorale2018.pdf
- [2] Formation Doctorale "Éthique de la recherche et intégrité scientifique": une formation proposée par l'école des docteurs de Toulouse, 2017. [Online]. Available (in French): <https://hal.archives-ouvertes.fr/cel-01452867v2> (version 3 to come)

Please contact :

Catherine Tessier
ONERA/DTIS, Université de Toulouse
+33 5 62 25 29 14
catherine.tessier@onera.fr

Efficient Accumulation of Scientific Knowledge, Research Waste and Accumulation Bias

by Judith ter Schure (CWI)

An estimated 85 % of global health research investment is wasted [1]; a total of one hundred billion US dollars in the year 2009 when it was estimated. The movement to reduce this waste recommends that previous studies be taken into account when prioritising, designing and interpreting new research. Yet current practice to summarize previous studies ignores two crucial aspects: promising initial results are more likely to develop into (large) series of studies than their disappointing counterparts, and conclusive studies are more likely to trigger meta-analyses than not so noteworthy findings. Failing to account for these aspects introduces 'accumulation bias', a term coined by our Machine Learning research group to study all possible dependencies potentially involved in meta-analysis. Accumulation bias asks for new statistical methods to limit incorrect decisions from health research while avoiding research waste.

The CWI Machine Learning group in Amsterdam, The Netherlands, develops methods to allow for optional continuation in statistical testing. Thus, in contrast to standard statistical tests, the safe tests we develop retain their statistical validity if one decides, on the spot, to continue data collection and obtain a larger sample than initially planned for – for example because results so far look hopeful but not yet conclusive. Additional research into the application of these safe methods to meta-analysis was inspired by the replicability crisis in science and the movement to reduce research waste.

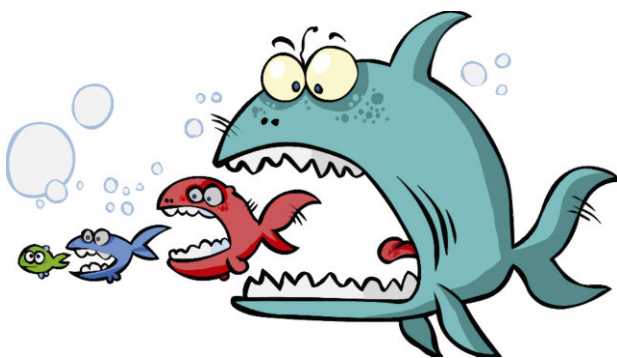
The 85 % research waste estimate is calculated by cumulatively considering waste in four successive stages of health research: (1) the choice of research questions, (2) the quality of research design and methods, (3) the adequacy of publication practices and (4) the quality of research reporting. In two of these stages, design and reporting, research waste is caused by a failure to systematically examine existing research. In terms of research design, the paper that estimated the research waste [1] stresses that new studies "should not be done unless, at the time it is initiated, the questions it proposes to address cannot be answered satisfactorily with existing evidence". Its recommendations about reporting involve that new studies should "be set in the context of systematic assessments of related studies" [1].

In 2014, a series of follow-up papers put forward by the REWARD Alliance showed that the 2009 recommendations remained just as pressing in 2014. The recommendation to always relate new studies to available research in design and reporting acquired the name evidence-based research in 2016

and has since then been promoted by the Evidence-Based Research Network.

Deciding whether existing evidence answers a research question is a difficult task that is further complicated when the accumulation of scientific studies is continuously monitored. This continuous monitoring is key to ‘living systematic reviews’, which are meta-analyses that incorporate new studies as they come available. Restricting further research when a certain boundary of evidence is crossed introduces bias, while continuous monitoring also creates multiple testing problems. As a result, reducing research waste is only feasible with statistical methods that allow for optional continuation or optional stopping.

Optional stopping is a well-studied phenomenon in statistics and machine learning, with a variety of approaches in the frequentist, Bayesian and online learning realm. These approaches are neatly combined, and much generalised, in the safe testing paradigm developed in our group. What is new to the meta-analysis setting is that dependencies arise even without continuously testing a series of studies. The very fact that a series of studies exists already introduces



Large study series are more likely when they include initial promising results within the series than when they include very disappointing ones, just like the availability of the big fish in the cartoon depends on specific smaller fish available.

Image: Bobb Klissourski (Shutterstock)

dependencies with results part of it that were at least not unacceptably disappointing to prohibit the expansion into the available series.

Meta-analysis is currently mainly considered when a study series of considerable size is available, with a median number of studies of around 15 in a typical meta-analysis [2]. Large study series are more likely when they include initial promising results within the series than when they include very disappointing ones, just like the availability of the big fish in the cartoon depends on specific smaller fish available. These dependencies introduce accumulation bias that in turn inflates false positive error rates when ignored in statistical testing.

Our research tries to determine how to deal with small-fish-dependent large fish, for various fish sizes. We intend to develop these methods for meta-analysis within the period of my PhD research (2017-2022) and involve other

researchers from evidence based medicine, the reducing waste movement, psychology’s reproducibility projects, and software projects such as JASP in implementing and communicating the results.

The recommendations of the 2009 research waste paper are increasingly being heard by chief scientific advisors, funders (such as the Dutch ZonMW) and centres for systematic reviews [3]. Now we need to implement them efficiently with the right statistics.

Links:

[L1] <http://rewardalliance.net/>

[L2] https://en.wikipedia.org/wiki/Evidence-based_research

[L3] <http://ebnetwork.org/>

References:

[1] I. Chalmers, P. Glasziou: “Avoidable waste in the production and reporting of research evidence”, *The Lancet*, 374(9683), 86-89, 2009.

[2] D. Moher, et al.: “Epidemiology and reporting characteristics of systematic reviews, *PLoS medicine*, 4(3), e78, 2007.

[3] P. Glasziou, I. Chalmers: “Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers”. *Bmj*, 363, k4645, 2018.

Please contact:

Judith ter Schure, CWI, The Netherlands

+31 20 592 4086

Judith.ter.Schure@cwi.nl