

Shining More Light on Dark Data

Sandra Petty, Hugo Stephenson, and Sarah Hadley

In this article, the team at the NY-based, nonprofit Center for Biomedical Research Transparency (CBMRT)¹ discusses the conditions which generate dark data and how providing a mechanism for publishing high quality negative and inconclusive results alongside “positive” ones is helping to shine more light on these valuable biomedical data.

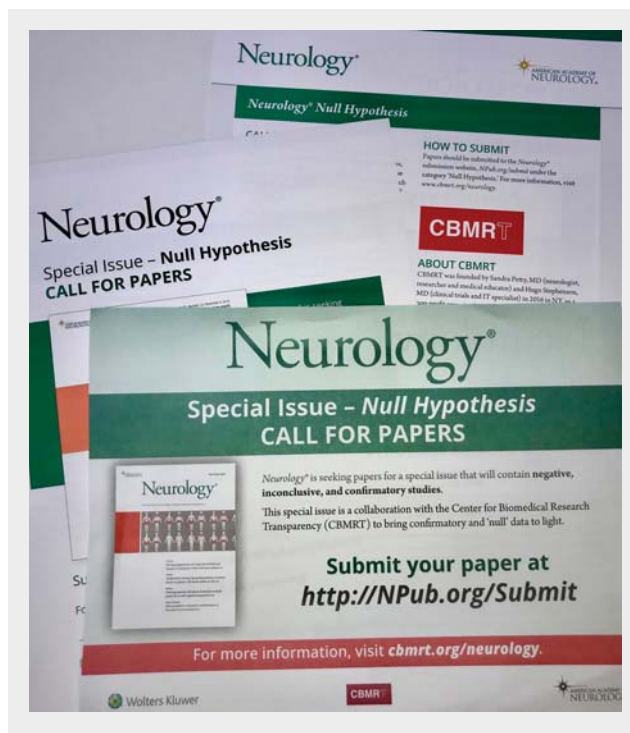
What is Dark Data?

Publication bias is a well-known issue among scientists and clinicians. Journals often like to publish positive, headline-catching results; it’s good for business. It is estimated that for clinical trials alone, positive results are almost twice as likely to be published as negative or inconclusive results.² This incentivizes scientists to put their negative or inconclusive findings—from nonetheless well-designed and executed studies—in the bottom drawer, leading to an incomplete picture of research across many scientific fields. These unpublished negative and inconclusive data exist as dark data, hidden in lab books around the world, undiscoverable to future researchers, and useless to clinicians who might value this knowledge when making treatment decisions (e.g., “Drug X worked in three out of three published trials, but what about the three unpublished ones?”). Per neurologist and CBMRT co-founder Dr Sandra Petty:

“As a physician, this issue is concerning; it is no less concerning for patients. To quote one of my astounded patients: “Don’t you know all this already?!”

Evidence suggests that over half of clinical trial results remain unpublished 30 months after trial completion (and one-third remained unpublished 51 months [median] post trial completion).³ This figure is likely to be significantly higher for biomedical research in the laboratory, which is harder to track with limited preclinical research registries and information.

PROF SANDRA PETTY is co-founder and CEO of Center for Biomedical Research Transparency (CBMRT). DR HUGO STEPHENSON is co-founder of CBMRT. SARAH HADLEY is Deputy Director of CBMRT.



Dark data also represents significant research waste, which an issue now very much in focus among funders and the scientific community as scientists may actually duplicate research that has already been completed but never published. By some estimates upwards of 80% of medical research funding is wasted, which equates to around \$160 billion in global medical research spend per annum.⁴ This figure includes wastage not just through non-publication of research, but also through unclear, incomplete, or inaccurate published results and poor study design. Put simply, researchers conduct many experiments and trials as a result of research funding they receive. Researchers often select the research with the best, usually positive results, in which to invest their time to write up and submit for publication. But most of the experiments and their results are never written up, let alone submitted for publication or made discoverable for future researchers. This creates an

CONTINUED

environment where future grant recipients have no ability to learn from prior work that has ultimately been funded by donors and taxpayers.

Why Does Dark Data Exist?

The causes of dark data are multifactorial⁵ and span the spectrum of research and reporting activity.

At one extreme, in a highly competitive research environment, there exists a perception amongst researchers that drawing attention to efforts that have been unsuccessful in demonstrating an expected outcome can work against their career goals and chances of future funding. In a data-driven world, changing this perception goes to the heart of research culture, and involves recognizing and celebrating those who have pursued well-planned and designed avenues of research, even if those results are not “positive.”

At the other extreme, competition for space in top-tier peer reviewed journals has meant that null hypothesis manuscripts have faced a high bar for acceptance and compete against papers with positive results, which could be seen to have higher commercial value in terms of attracting citation, subscriptions, and reprints. This results in repeated experiences of manuscript rejection. Many journal editors, believe, despite evidence to the contrary, that null hypothesis articles are less likely to be cited in future papers, with citation being used as a crude indicator for research relevance and impact. In addition, the “novelty” of a study can be a consideration for journals. That is, in making publication decisions, journals often assess whether the research is “new, true, and does anyone care.” Negative, inconclusive, and confirmatory results may not meet journals’ expectations for novel and unique research. Since the analysis, writing, and manuscript drafting processes are time consuming for time-poor researchers, many choose to focus their efforts on research that they perceive has a greater chance of publication success. Changing this perception requires close interaction with major journals, their editorial teams, and establishment of dedicated space for well-designed studies that result in negative and inconclusive outcomes.

What to Do About Dark Data?

The emergence of the modern open science movement almost two decades ago has spurred a near-continuous development of innovative tools and initiatives that form today’s open science infrastructure. Undoubtedly, these developments have helped bring the issue of dark data to light:

- **Open access mega journals** such as *BMJ Open* and *Medicine* are helping get more dark data published by giving less consideration to novelty, and greater acceptance of negative results and confirmatory studies
- **Open data initiatives** including open source software and workflow tools and data sharing initiatives, of which there are over 300 in biomedicine alone. These include Figshare, YODA, the Genomic Data Commons, and FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles which promote access and utilization of existing electronic data, algorithms, and analytical tools. These initiatives help to make dark data more discoverable. Therefore, even if a study has not resulted in publication the underlying data are now easily sharable.
- **Preprint servers** where researchers can upload complete scientific manuscripts to a public server. Almost 2,400 biology preprints are being added to public servers such as bioRxiv and PeerJ each month, and the recent launch of MedRxiv has extended the service into medical, clinical, and related health sciences. Preprint servers provide an opportunity for researchers to share their preliminary results in the interests of both drawing early attention to their work and of adding to a knowledge set in a more timely manner. Well executed negative, inconclusive and confirmatory studies receive equal representation alongside positive results.
- The **Declaration on Research Assessment (DORA)** is a set of recommendations designed to improve the ways in which the outputs of scholarly research are evaluated. The Declaration currently has over 12,800 individual signatories and 872 scientific organization signatories. By encouraging a shift away from publication metrics towards making assessments based on scientific content, publication bias is downplayed and reporting of otherwise dark data incentivized.
- The **Consolidated Standards of Reporting Trials (CONSORT)** is an evidence-based, 25-item checklist endorsed by 585 journals for reporting randomized trials and is designed to improve completeness and transparency in trial reporting. Placing greater emphasis on reporting underlying methodology serves to level the playing field between high quality positive, negative, and inconclusive results.
- **Funder evaluation tools.** As funders focus more on the outcomes of their medical research expenditure, they will increasingly rely on platforms such as Digital Science’s “Dimensions” which leverage machine learning and NLP technologies to build connections between clinical trials, publications, policies, and patents data and in turn track research impact through customized metrics. At a minimum, dark data resulting from research grants will be more readily identifiable.

CONTINUED

It would seem however that this impressive open science infrastructure may be a necessary but not completely sufficient set of resources to achieve research transparency to the degree that dark data is brought to light. Our view is that a continuing shift in research culture across the biomedical research ecosystem is also needed to achieve a permanent state of transparency. We envisage an environment where researchers are enabled to utilize more of these resources, and where research output incentives and funding trends are redefined.

Culture change comes about through a combination of different drivers such as technological changes and invention, network and infrastructure creation, leadership, exchange, and education, and does not require significant investment. As noted by the Royal Society as part of its Research Culture Program (which focuses in particular on research integrity):

“Enhancing research culture doesn’t require major effort and resources. Organizations across the UK and globally have made changes linked to integrity that have improved their research culture. These range from simple approaches such as using informal communication channels to nurture a supportive environment, discussing successes and “failures”, to embedding research integrity into the heart of institutional culture, requiring research leaders and senior administrators to lead by example.”⁶

The Center for Biomedical Research Transparency (CBMRT) is another non-profit organization focused on enhancing research culture by facilitating transparent reporting of biomedical research. CBMRT’s goal is to ensure that all biomedical results, including negative and inconclusive results (dark data), are discoverable and accessible in the interests of patient safety and research efficiency.

To achieve this, CBMRT works with major medical societies and their existing, highly respected journals to call for papers with null or inconclusive data and publish as a special edition called Null Hypothesis. This initiative is directly changing research culture by reducing the probability of manuscript rejection, and celebrating researchers who write their dark data with publication in journals of impact for their peers. As noted by one Null Hypothesis author, Dr Kevin Messacar:

“I applaud the efforts of CBMRT in combatting publication bias. Considerable effort was put into gathering the retrospective data from the clinical experience of off-label fluoxetine use for AFM with great uncertainty whether anyone would publish it without positive findings. The study was conducted with equipoise given the ultimate goal of figuring

out whether this novel use of the drug as an antiviral was having any clinical impact. We were so pleased that, despite the negative findings, Neurology gave it fair consideration and chose to feature it in the null hypothesis edition. If we don’t publish what doesn’t work, it will take us much longer to get to what actually works.”⁷

CBMRT’s first Null Hypothesis partnership, launched with the American Academy of Neurology (AAN) and its flagship journal *Neurology* has been a great success, resulting in a thirty-fold increase in inflow of papers documenting negative and inconclusive findings, and significantly raised awareness of such data and its value across the international community of neurologists. In April 2019, CBMRT and *Neurology* produced and circulated a full edition of *Neurology* dedicated to papers with negative and inconclusive findings, with the articles achieving above average levels of citation and even attention in the lay press. Null Hypothesis articles go through the same peer review process as all other *Neurology* submissions and are made freely available online ahead of print. As a result of this success, CBMRT is formalizing a long-term partnership with AAN and *Neurology* for future editions of *Neurology* Null Hypothesis, and working with major societies to replicate the model in other therapeutic areas including cardiology, oncology, and infectious disease.

Negative results journals have been attempted in the past (most notably the *Journal of Negative Results in Biomedicine* from Springer/BioMed Central) with somewhat limited success. The key to the success of Null Hypothesis is that it is the product of collaboration: medical societies and their journals contribute the publishing infrastructure and CBMRT leverages its Global Ambassador Network of over 1,000 biomedical professionals to generate a steady flow of journal submissions. Furthermore, Null Hypothesis is a model that is easily replicable across therapeutic areas, creating a commonly-branded and identifiable movement that puts an infrastructure for dark data publication firmly in the research mainstream.

The Null Hypothesis initiative runs alongside CBMRT’s US-European Biomedical Transparency Summit Series. The annual, free summits engage and connect a diverse group of stakeholders across the spectrum of biomedical research activity and drive the culture change required to increase transparency. Outstanding speakers across the United States and Europe are invited from government, industry, academia, and the not-for-profit sector. Summit participants are similarly diverse; CBMRT focuses in particular on ensuring that early career researchers and patient-centered research organizations are well-represented. The Summits

CONTINUED

cover a wide range of transparency topics including policy developments, evolution of the publishing model, data sharing innovations, and research methodology.

There are several other successful initiatives focused specifically on driving culture change towards greater transparency in biomedical research. There are awards which signal the importance of publishing data where the results do not confirm the expected outcome or original hypothesis, such as the ECNP Preclinical Network Data Prize for published “negative” scientific results, and the Symbiont Awards which recognize exemplars in data sharing practice. The AllTrials–BMJ “Unreported Clinical Trials of the Week” campaign draws attention to the need for greater transparency on clinical trial methods and results by shining a spotlight on clinical trials that haven’t published results. And the ReproducibiliTea journal club initiative is now running in 27 countries, bringing young university researchers together across disciplines to discuss diverse issues, papers, and ideas about improving science.

As clinicians and scientists, we are in so many ways indebted to the quality of research that has gone before us to gain understanding of diseases and therapies, to inspire and inform our own research study design, and most importantly to inform and optimize treatment outcomes for our patients. However, unless we achieve balanced and transparent reporting through the revelation of dark data we risk an incomplete understanding of the state of our field, of our treatments, and of the scientific evidence-based

knowledge we share with research participants and patients. The infrastructure exists; the task remains to capitalize on this by continuing the positive shift in research culture across the biomedical ecosystem.

References and Links

1. <http://www.cbmr.org/>
2. 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* 2009;(1):MR000006. <https://doi.org/10.1002/14651858.MR000006.pub3>.
3. Ross J, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov *BMJ* 2012;344:d7292. <https://doi.org/10.1136/bmj.d7292>.
4. Glasziou P, Chalmers I. Paul Glasziou and Iain Chalmers: is 85% of health research really “wasted”? *BMJ Opinion* January 14, 2016. [accessed December 11, 2019]. <https://blogs.bmj.com/bmj/2016/01/14/paul-glasziou-and-iain-chalmers-is-85-of-health-research-really-wasted/>.
5. Johnson RT and Dickersin K. Publication bias against negative results from clinical trials: three of the seven deadly sins. *Nat Clin Pract Neurol* 2007;3(11):590–591. <https://doi.org/10.1038/ncpneuro0618>.
6. Chaplin K, Price D. 7 ways to promote better research culture. World Economic Forum Annual Meeting of the New Champions, September 18, 2018. <https://www.weforum.org/agenda/2018/09/7-ways-to-promote-better-research-culture/>.
7. Messacar K, Sillau S, Hopkins SE, Otten C, Wilson-Murphy M, Wong B, Santoro JD, Treister A, Bains HK, Torres A, et al. Safety, tolerability, and efficacy of fluoxetine as an antiviral for acute flaccid myelitis. *Neurology* 2019;92(18):e2118–e2126. <https://doi.org/10.1212/WNL.0000000000006670>.

VISIT US AT THE CSE ANNUAL MEETING

YOUR SCIENCE MATTERS

Publishing Tools & Services Built for
Society Publishers

COMPOSITION // EDITORIAL SERVICES // PRINTING // ONLINE PUBLISHING // MARKETING SERVICES



allenpress.com // sales@allenpress.com // (800) 627-0326