Institution: University of Edinburgh		
Unit of Assessment: 4		
Title of case study: G: Improving the reproducibility of preclinical research through more rigorous and robust policy and practice at publishers, funders and industry		
Period when the underpinning research was undertaken: 2003 – 2020		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
Malcolm Macleod	Chair of Neurology and Translational Neuroscience	2003 – present
Emily Sena	Stroke Association Kirby Laing Foundation Senior Lecturer	2008 – present
Period when the claimed impact occurred: August 2013 – December 2020		
Is this case study continued from a case study submitted in 2014? N		

1. Summary of the impact

Underpinning Research: Edinburgh Neuroscience research has quantified the prevalence of bias within preclinical research and shown that failure to address bias is associated with overstated or inaccurate results.

Significance and Reach of Impact: Major publishers including Nature Publishing Group have since 2013 implemented strict criteria for reporting research study methods. Major funders including the Medical Research Council and Wellcome have since 2015 included sections dedicated to experimental design and analysis in their application forms. Pharmaceutical companies are committed to improving reproducibility of preclinical research, and 18 (including Janssen, AbbVie, Roche, Novartis and Pfizer) have joined an Edinburgh-led consortium to define a shared quality management framework for industry and academia.

These elevated standards of rigour for research funding and publishing mean that research quality has risen, leading to more efficient use of taxpayers' money and improved prospects for developing effective treatments.

2. Underpinning Research

The Challenge: The reproducibility crisis and bias in preclinical research
There is a growing concern in the biomedical research community that many high-profile research findings cannot be reproduced. This concern is shared by research users such as pharmaceutical companies, who build on preclinical research to develop interventions for human disease.

An example of this phenomenon is a systematic review of the literature, performed by Edinburgh Neuroscience researchers on NXY-059, a free radical scavenger that was widely believed to have substantial neuroprotective properties in animal models of stroke [3.1]. The review was initiated because in 2006 SAINT II, a large AstraZeneca-sponsored Phase III clinical trial designed to confirm the efficacy of NXY-059, had found it to be ineffective. The results of the systematic review showed that only 3 of the underlying 9 preclinical publications reported randomisation, an important measure to reduce the risk of bias. Crucially, the studies *not* reporting randomisation gave substantially higher estimates of efficacy (53% reduction in infarct volume, compared with 20% reduction in studies that did report it) [3.1]. The researchers subsequently confirmed the association between non-reporting of risk of bias and higher treatment efficacy in a meta-meta-analysis of 13 meta-analyses of experimental stroke studies [3.2].

Thus, reproducibility of research is related to the presence of bias within experimental methods. To systematically document and quantify the prevalence and consequences of bias in preclinical research using animal models, Edinburgh Neuroscience researchers have founded and lead the international academic partnership CAMARADES. Overall,

CAMARADES work clearly shows that failure to take, or report, steps to minimise risks of bias, such as randomisation, concealment of treatment allocation, sample size calculations and blinded assessment of outcome, is associated with overstated or even inaccurate results, which are consequently difficult to reproduce.

Publication bias is pervasive in the preclinical literature and can overestimate efficacy

Using systematic reviews, Edinburgh Neuroscience researchers provided the first quantitative estimate of the prevalence and impact of study quality bias and publication bias in the literature on animal models of stroke, estimating that 1 in 7 experiments that have been conducted are never published [3.3]. Importantly, in an analysis of 4,445 studies of 160 candidate treatments for neurological disorders, the researchers found an 'excess of significance', whereby based on plausible effect sizes, 919 studies would have been expected to report a significant positive effect, but 1,719 were found to do so [3.4]. This bias towards publishing positive findings is problematic because non-publication of negative data skews the literature towards overestimating the potential efficacy of new treatments.

Inflated effects from preclinical studies may therefore lead to clinical trials that were never likely to succeed (such as SAINT II above), wasting resources and placing participants at unnecessary risk.

Steps taken to reduce risk of bias are under-reported

Having demonstrated and quantified the publication bias towards positive findings, the Edinburgh Neuroscience researchers next quantified the extent of under-reporting of measures to reduce the risk of bias: in a random sample of 2,000 biomedical publications, randomisation was reported in only 20%, blinded assessment of outcome in 3%, and sample size calculations in none. Under-reporting of these measures was found across deciles of journal impact factor, and in publications from the 5 top-ranking UK institutions according to RAE2008: only 14.4% reported randomisation, 17.3% blinding, 10.4% inclusion/exclusion criteria, and 1.4% sample size calculations [3.5].

A core set of indicators of validity is proposed

Edinburgh Neuroscience research findings were among the driving forces that prompted the US National Institute of Neurological Disorders and Stroke (NINDS) to bring together academics (including CAMARADES members), editors and representatives from funding agencies, disease advocacy groups and pharmaceutical industry, to develop recommendations for the reporting of animal experiments in both publications and grant applications. The outcome of this June 2012 meeting was a core set of indicators of validity – randomisation, blinding, sample size estimation, and data handling (e.g. pre-specifying the primary endpoint, inclusion/exclusion criteria and outliers) – that should always be reported [3.6]. The conclusions of this paper were strongly influenced by Edinburgh Neuroscience research, with 14 of the 64 references cited being from Macleod and/or Sena.

3. References to the research

[3.1] Macleod, MR, van der Worp, HB, Sena, ES, Howells, DW, Dirnagl, U & Donnan, GA 2008, 'Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality' *Stroke*, 39(10) pp. 2824-2829. doi: 10.1161/STROKEAHA.108.515957

[3.2] Crossley, NA, Sena, E, Goehler, J, Horn, J, van der Worp, B, Bath, PMW, Macleod, M & Dirnagl, U 2008, 'Empirical evidence of bias in the design of experimental stroke studies - A metaepidemiologic approach' *Stroke*, 39(3), pp.929-934 doi: 10.1161/STROKEAHA.107.498725

[3.3] Sena, ES, van der Worp, HB, Bath, PMW, Howells, DW & Macleod, MR 2010, 'Publication Bias in Reports of Animal Stroke Studies Leads to Major Overstatement of Efficacy' *PLoS Biology*, 8(3), e1000344 doi: 10.1371/journal.pbio.1000344

[3.4] Tsilidis KK, Panagiotou OA, Sena ES, Aretouli E, Evangelou E, Howells DW, Al-Shahi Salman R, Macleod MR, Ioannidis JPA 2013 Evaluation of Excess Significance Bias in Animal

Studies of Neurological Diseases. *PLOS Biology,*11: e1001609 doi: 10.1371/journal.pbio.1001609
[3.5] Macleod, MR, Lawson Mclean, A, Kyriakopoulou, A, Serghiou, S, De Wilde, A, Sherratt, N, Hirst, T, Hemblade, R, Bahor, Z, Nunes-Fonseca, C, Potluru, A, Thomson, A, Baginskitae, J, Egan, K, Vesterinen, H, Currie, GL, Churilov, L, Howells, DW & Sena, ES 2015, 'Risk of Bias in Reports of In Vivo Research: A Focus for Improvement' *PLoS Biology*, 13(10), pp. e1002273 doi: 10.1371/journal.pbio.1002273
[3.6] Landis, et al. (including Macleod, MR) 2012, 'A call for transparent reporting to optimize the predictive value of preclinical research' *Nature*, 490(7419), pp. 187-191 doi: 10.1038/nature11556

4. Details of the impact

The overarching goal of CAMARADES work is to improve the reproducibility and quality of preclinical research. This is being achieved through their research informing and improving standards and processes at publishers, funders and pharmaceutical companies, as detailed below.

Impact on publishing standards

In response to the growing concern around reproducibility of biomedical research, in April 2013 (before the REF2021 impact period), Nature Publishing Group (NPG) introduced editorial measures to improve the consistency and quality of reporting in its articles. This policy has remained in place for the entire REF2021 period and beyond. It abolished space restrictions in method sections to allow authors to describe their methods in as much detail as necessary, and provided a checklist to prompt authors to disclose technical and statistical information in their submissions; the contents of this checklist correspond to the 4 indicators of validity published in [3.6] [5.1a]. The Director of Author and Reviewer Services for NPG at the time confirms the role of Edinburgh Neuroscience research in bringing about the policy: "*Not only were the reporting criteria based on the essential indicators of validity identified in Macleod and Sena's research, but the impetus for the policy change was strongly driven by their demonstration of the poor reproducibility and under-reporting of risks of bias seen in preclinical research studies. In particular, their finding that under-reporting of steps taken to minimise bias is associated with overstated estimates of efficacy precipitated the need for a policy that requires these steps to be reported.*" [5.1b].

In 2016, Edinburgh Neuroscience researchers undertook a before-and-after study to determine whether the NPG editorial policy had resulted in improved reporting. This revealed that full compliance with the policy had reached 16%; an unprecedented improvement in the quality of reporting in NPG articles not seen in matched articles from other publishers [5.2].

Development of a minimal standards framework for all major publishers

Despite the improvements seen following the NPG editorial policy change, the before-and-after study indicated that 84% of papers were still not fully compliant. In order to improve this, in November 2018, a 9-person working group of journal editors and experts convened to develop a minimal set of reporting standards for all research in life sciences [5.3a]. Macleod is the only academic member of the working group; all others are affiliated with a publisher, including Wiley, eLife, PLoS, Springer-Nature, Cell Press/Elsevier and American Association for the Advancement of Science (which publishes the journal *Science*). This working group was expressly set up to deliver: 1) A framework setting out minimal expectations across the 4 core areas of materials, design, analysis and reporting (MDAR) 2) a checklist to serve as an implementation tool and 3) a user guide.

The MDAR checklist was tested in 13 journals in 2019, with the result that 80% of users (both editors and authors) found it useful and quick to complete (average 24 minutes). These results were presented at a meeting of the National Academies of Science, Engineering and Medicine (NASEM) in September 2019 [5.3b], with the presentation referring to papers [3.1], [3.4] and [3.6] in justifying the need for widely accepted minimal standards of reporting [5.3c]. Further

refinements to the checklist are underway. Importantly, all publishers represented on the working group have committed to rigorously enforcing the framework, once published [5.3b].

Revision of ARRIVE guidelines
Another leading effort to improve reporting standards was the ARRIVE guidelines (Kilkenny et al. 2010), coordinated by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs). Edinburgh Neuroscience research was heavily referenced in justifying why such guidelines were needed (5 out of 21 publications cited) [5.4a]. The ARRIVE guidelines were endorsed by 430 journals, but after 9 years, this had not been accompanied by substantial improvements in reporting, as demonstrated by the Edinburgh-led randomised controlled trial IICARUS [5.4b].

The results of the IICARUS trial prompted NC3Rs to convene an international expert working group (26 people, including both Macleod and Sena) to revise the ARRIVE guidelines to facilitate wider uptake [5.4c]. Launched in July 2020, the new "ARRIVE 2.0" guidelines contain re-prioritised items divided into "essential" and "recommended", with additional clarifications and illustrative examples to ensure that authors, editors and reviewers are "*better equipped to improve the rigour and transparency of the scientific process, and thus reproducibility*" [5.4d]. These revised guidelines cite 8 Edinburgh Neuroscience publications, including in explaining the reasons for the new format chosen. By December 31st 2020, the ARRIVE 2.0 website had been viewed 103,000 times [5.4e] and the paper publishing them viewed 18,700 times [5.4d].

The increased focus of publishers on quality and transparency is highlighted by the launch in 2017 of a new journal, *BMJ Open Science*, with an explicit mission to "*improve the transparency, integrity, and reproducibility of biomedical research closely aligned to medicine*." [5.5a]. The role of Edinburgh Neuroscience research in driving this focus is highlighted by the appointment of Sena as inaugural Editor-in-Chief [5.5b].

Impact on research funders
The publishers' increased attention to rigour, driven by Edinburgh Neuroscience research, has been mirrored by major funders including Wellcome, Medical Research Council (MRC), Biotechnology and Biological Sciences Research Council (BBSRC) and the United States National Institutes of Health (NIH). In 2015, NC3Rs coordinated an effort to align policies between UK funders, crediting Edinburgh Neuroscience research in providing the impetus: "*CAMARADES […] work has been instrumental in revealing issues with the quality of animal experiments, estimating the scale of the problem, and highlighting its impact on the reliability of published research. Awareness of these issues amongst the funders prompted a greater focus on the rigour of the science described in funding applications, with major UK funders such as the MRC, BBSRC, Wellcome and Cancer Research UK now requesting explicit experimental design and statistics information, providing training for panel members to assess the reliability of the research proposed, and including statisticians on funding panels*." [5.6a]. An article on the *Nature* website reporting the updated requirements also cited CAMARADES work to illustrate the need for clear guidelines [5.6b].

In the US, an updated NIH policy on reporting and rigour was announced in 2017, accompanied by a paper explaining the rationale behind the policy. This cites 9 Edinburgh publications in key sections on factors contributing to low reproducibility [5.7].

Impact on pharmaceutical industry
Edinburgh Neuroscience research findings are influencing policy and practice in the pharmaceutical industry through the European Quality in Preclinical Data (EQIPD) Consortium. EQIPD was formed in 2017 through EUR4,500,000 (GBP4,046,040; 01-21) funding for 3 years, and consists of 11 academic and 18 industry (including Pfizer, AbbVie, Janssen, Roche, and Novartis) partners from 8 countries [5.8a], with the University of Edinburgh being the coordinating institution [5.8b]. The EQIPD partners are working together to define a shared quality management framework for both industry and academia, and deliver

certified education and training in this framework to enable a smoother, faster and safer transition between preclinical and clinical testing [5.8c].

Insights from Edinburgh Neuroscience research have also led to practical changes in companies within EQIPD. An Associate Director at Janssen Pharmaceuticals stated: "*As a direct result of our engaging with this research, we have changed our internal research procedures at Janssen and put increasing emphasis on research rigor and experimental design of our preclinical studies, including more focus on randomization, blinding, upfront specification of exclusion criteria and sample size calculation, and early involvement of our biostatisticians, both as integral part of our ethics approval processes for internal projects and procedures involving animals and during experimental planning. It was the CAMARADES research […] that alerted us to the issues and informed us how best to address them*." [5.9a].

Similarly, the CEO of BioCurate, an Australian not-for-profit company aiming to generate high-quality preclinical candidates from academic research for the bio-pharmaceutical industry, stated: "*Research from the CAMARADES group was pivotal in establishing a quantitative approach to identify research reports that are at risk of bias. In addition, they have provided important insights as to how these issues can be improved*." [5.9b].

Impact on research quality

As described above, Edinburgh Neuroscience research has substantially contributed to increased rigour and robustness in the publishing and funding of academic research as well as policy and practice in the pharmaceutical industry. As a result of these improvements, the overall standard and robustness of biomedical research will rise; as an example, improvements are already seen in the reporting of experimental methods following the NPG editorial policy change [5.2].

This ongoing improvement in standards of biomedical research amounts to a more efficient and accountable use of public funds. Ultimately, the beneficiaries are patients in need of therapies, through the higher likelihood of clinical trials succeeding and producing approved therapies.

5. Sources to corroborate the impact

[5.1] NPG editorial policy change. a. Nature Announcement: Reducing Our Irreproducibility, 25th April 2013 b. Testimonial from Director of Author and Reviewer Services for Nature Publishing Group (in 2013)

[5.2] The NPQIP Collaborative group, MacLeod, M, Sena, E & Howells, DW 2019, 'Did a change in Nature journals' editorial policy for life sciences research improve reporting?' *BMJ Open Science* doi: 10.1136/bmjos-2017-000035

[5.3] Minimal standards. a. Minimal standards blog b. PLoS Blog with pilot results c. Slides presented at NASEM meeting

[5.4] ARRIVE guidelines 1.0 and 2.0 a. ARRIVE guidelines Kilkenny et al. 2010 doi: 10.1371/journal.pbio.1000412 b. IICARUS trial: Hair, K, Macleod, MR & Sena, ES 2019. doi: 10.1186/s41073-019-0069-3 c. Percie du Sert N, Hurst V, Ahluwalia A, *et al*. 2018 Revision of the ARRIVE guidelines: rationale and scope. *BMJ Open Science*. doi: 10.1136/ bmjos-2018-000002 d. New ARRIVE 2.0 guidelines doi: 10.1371/journal.pbio.3000410 e. Google Analytics for ARRIVE website to 31st December 2020

[5.5] *BMJ Open Science* a. About *BMJ Open Science* b. Dr Sena as Editor-in-Chief

[5.6] Impact on research funders a. Testimonial from Head of Experimental Design and Reporting at NC3Rs b. News article on Nature website on UK funders' new guidelines

[5.7] Hewitt et al. 2017 paper to explain new NIH policy on rigour doi:10.1093/ilar/ilx011

[5.8] EQIPD a. List of members b. UoE as coordinating institution c. EQIPD vision

[5.9] Testimonials from pharmaceutical industry a. Testimonial from Janssen Pharmaceuticals b. Testimonial from BioCurate