Institution: University of Birmingham

Unit of Assessment: UoA 10, Mathematical Sciences

Title of case study: Statistics research underpins methodological standards and design tools for efficient and informative trial of healthcare interventions

Period when the underpinning research was undertaken: 2011–2020

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:
Prof. Karla Hemming	Professor of Biostatistics	2009–present (Institute of Applied Health Research)
Mr Alan Girling	Reader in Medical Statistics	1987–2003 (Mathematics); 2004–present (Institute of Applied Health Research; Emeritus 2016–present)

Period when the claimed impact occurred: 2014–2020

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact

Through changing the design methodologies for a type of cutting-edge clinical trial, our statistical research has impacted **health and wellbeing**. Specifically, our improvements to stepped-wedge randomised control trials have resulted in:

- **A new technology has been adopted by trial designers** impacting an estimated 230 trials;
- **Improved reporting guidelines for clinical trial leaders** through the adoption of our methodology by *CONSORT*, the gold standard international guidelines for clinical trials, and also **mandated by publishers**, impacting on at least 277 trials in this impact period;
- **Decisions by government health services informed by research**, as the WHO and national health agencies use the technology to assess/design trials;
- **Reduced costs of healthcare trials** and improved efficiency;
- **Trial participants, estimated at over 2M, have benefited from inclusion in more efficiently designed studies** in this REF period.

2. Underpinning research

The gold standard to assess the safety and effectiveness of clinical and healthcare interventions is the randomised control trial. However, it is often necessary to randomise clusters of people (e.g. hospitals, villages) rather than individuals, either to avoid contamination, or for logistical/economic reasons. As such, cluster randomised trials (CRT) are used extensively in healthcare and the social sciences. In many cases, it is unethical, or impossible, to withhold clinical intervention from certain (control) clusters; for example, if a treatment were proving highly effective. The stepped-wedge CRT (SW-CRT) solves this problem by staggering the intervention roll-out over time. In recent years, SW-CRTs have been growing rapidly in popularity. However, prior to the research it remained unclear how to analyse incomplete or staggered trials, trials with nested clusters (e.g. hospitals within regions), or how to a priori determine the most efficient trial design for a given situation.

Prof. Karla Hemming and Mr Alan Girling of the University of Birmingham have made several fundamental theoretical contributions to the understanding of SW-CRTs to underpin planning, design and reporting. In the first stages of the research, Hemming and Girling extended the sample size calculations for CRTs, allowing investigators to improve the statistical power of trials when cluster numbers were fixed by modifying numbers of individuals per cluster [1], leading to a published industry-standard software (Stata) implementation [4]. Hemming and Girling went on to develop a unifying framework for CRTs, including SW-CRTs, which provides several important advances embodied in the following key findings:

KF1. Power and sample size calculations can now be extended to 'incomplete' trials, where there is missing data or an implementation period to interventions [3].
KF2. Parallel, staggered CRTs can be analysed as particular cases of incomplete SW-CRTs [3].
KF3. Power calculations are given for 'nested' cases in which there is clustering within clustering (for instance, wards within hospitals) [2].
KF4. The effect of the intra-cluster correlation (ICC) can be analysed in different settings. In particular, our research showed that SW-CRTs are significantly more robust to changes in ICC than CRTs, and in some designs, power can increase with ICC rather than decrease as previously assumed [3].
KF5. The efficiency of a parallel CRT and that of a SW-CRT can now be compared for a given situation (taking into account ICC) [2, 3].
KF6. To extend previous trial efficiency results through the framework of linear mixed effects modelling, enabling assumptions regarding fixed correlation structures to be taken into account — focusing on efficiency, which is of vital importance to funders [5]. The result was the development of an algorithm for optimising the times at which the intervention is introduced.

As a result of KF1–KF6, it is now possible to determine — for a given situation, research question and resources — the most efficient trial design: in other words, to minimise the resources required to establish whether a particular intervention has beneficial outcomes for the target population. Furthermore, it is now possible to use SW-CRT in situations where information is incomplete.

3. References to the research

[1] Hemming K, Girling AJ, Sitch AJ, Marsh J, Lilford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Med Res Methodol*. 2011 Jun 30; 11:102. DOI: 10.1186/1471-2288-11-102. Erratum in: *BMC Med Res Methodol*. 2017 Jan 19;17 (1):8. PubMed PMID: 21718530; PubMed Central PMCID: PMC3149598.

[2] Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ* 2015; 350:h391. DOI: 10.1136/bmj.h391

[3] Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. *Stat Med.* 2015 Jan 30;34(2):181-96. DOI: 10.1002/sim.6325. Epub 2014 Oct 24. PubMed PMID: 25346484; PubMed Central PMCID: PMC4286109.

[4] Hemming K, Girling AJ. A menu-driven facility for power and detectable-difference calculations in stepped-wedge cluster-randomized trials. *Stata Journal*. 2014. 14(2): 363-380. DOI: 10.1177/1536867X1401400208

[5] Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Stat Med*. 2016 Jun 15;35(13):2149-66. DOI: 10.1002/sim.6850. Epub 2016 Jan 7. PubMed PMID: 26748662; PubMed Central PMCID: PMC4949721.

[6] Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M, Kristunas C, Goldstein CE, Campbell MK, Girling A, Eldridge

S, Campbell MJ, Lilford RJ, Weijer C, Forbes AB, Grimshaw JM. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*. 2018 Nov 9;363:k1614. DOI: 10.1136/bmj.k1614. PubMed PMID: 30413417; PubMed Central PMCID: PMC6225589.

4. Details of the impact

Our impact is on the design, practice and reporting of healthcare trials, benefitting trial leaders, funders and publishers through **new professional standards, technology and improved resource efficiency**, and ultimately benefiting patients who participate in trials.

1. The development of a new technology that has been adopted in clinical trials

We have **produced a software tool for the optimal design of CRTs**, including SW-CRTs, which draws on all our key findings **[KF1–KF6]**. **The Shiny CRT calculator was used in an estimated 230 clinical trials this impact period**. Since October 2017, the Shiny CRT calculator has been available in beta. The calculator enables those running clinical trials to **assess the pros and cons of different trial designs before committing to a study**. This is attested to by the Director of the Duke Global Health Institute, USA, who states, "I recently used the app when designing a complex stepped wedge trial in Kenya and was able to much better communicate the implications of changing different inputs" [S3].

The calculator fills a critical need for trial leaders, as explained by a World Health Organization Medical Officer: "The capacity of trialists to estimate sample sizes for different scenarios without the help of a statistician for every calculation was limited. The Shiny CRT calculator has come to fill that gap. I have used the tool at the design stage of a complex trial in maternal health and found it extremely helpful to easily calculate sample sizes or power estimations for different trial designs […]" [S4].

The calculator regularly receives over 100 hours of use per month [S6]. Reference to the research 2 has 545 citations up to October 2020, of which 43% report the results of trials using our underpinning research or the Shiny CRT calculator directly. A conservative estimate of the number of trials directly applying our key findings and technology is thus **230 trials** in this impact period [S8].

2. Improved healthcare trial practice through changed guidelines

We have informed the CONSORT guidelines which have changed the design of over 270 clinical trials in this impact period.

In 2018, the CONSORT reporting guideline for the SW-CRT [6] was introduced [KF6], of which Hemming was lead author. The growing reach of these guidelines is evidenced by the fact that there are 277 registered 'stepped-wedge' studies on NIH (USA) website clinicaltrials.gov [S6]. **These guidelines are endorsed and enforced** by the main medical journals and editorial groups who require the use of CONSORT as a prerequisite for publication of clinical trial results, as evidenced by instructional statements for authors. For example, the *Lancet* requires that "Cluster-randomised trials must be reported according to CONSORT extended guidelines" [S9] and *Nature* states that "Reports that do not conform to the CONSORT guidelines may need to be revised before peer review" [S9].

3. WHO and French Health Services have been informed by our research

The calculator is used by assessors to evaluate the soundness and efficacy of research proposals, leading to **more efficient use of resources and better research outcomes**. As the Head of the Biometrical Department, University Hospital of Tours, France, explains [S1]: "I am regularly involved in the French Ministry of Health selection process of clinical research projects, either as a panel member or as a reviewer. In both situations, anytime the research project is a cluster randomised trial, I use the Shiny app developed by Pr Karla Hemming to check the sample size calculation. [The app] covers a field which is not covered by classical sample size calculation softwares." Similarly, the calculator's impact on health service decisions is attested

by a World Health Organization Medical Officer: “The tool facilitated an efficient brainstorming meeting in which a group of us discussed the implications of different cluster trial designs on the study feasibility, and helped us to decide the most convenient approach” [S4].

4. The costs of healthcare trials have been reduced as a result of research-led changes in practice

The capabilities of the calculator allow clinical trial leaders to explore options and conceive the most efficient trial, **saving money, minimising the numbers of subjects exposed to experimental treatments and maximising the efficacy of their trials**. This is attested to by the Director of the International Program Evaluation Unit, Hospital for Sick Children, Toronto, who states: “[The calculator] has been the basis for the design of a series of trials that have been run in Kenya and Pakistan and of a series of other studies that are currently being implemented by our teams in Pakistan, Bangladesh, Kenya, Mali and Senegal. Our team has been able to save considerable time and resources by adopting the tool and we are much less dependent on external statistical support since the shiny app has been developed” [S2]. The efficiency afforded by the calculator is further emphasised by a biostatistician, involved with clinical trials, at the Washington School of Public Health: “[The app] not only allows [users] to plan their trials to cover many eventualities, it also facilitates the design of highly efficient trials which improves use of resources (i.e. more efficient trials cost less)” [S5].

5. Patients have benefited through receiving better and more ethical trial design

It is estimated that since 2017, **over 2M patients in clinical trials have benefitted from improved trial designs** [S8]. Benefits to trial participants can occur in two ways. Better designed safety trials **expose fewer subjects to potentially dangerous interventions**. On the other hand, in efficacy trials, the SW design allows for a **greater number of patients to receive a safe intervention** of known (albeit as yet unquantified) benefit because the trial design involves all patients eventually receiving the intervention. The first effect is attested to by the Washington biostatistician: “[The app] means that fewer patients are exposed to treatments of unknown effectiveness” [S5]. And the authors of **S7.9** describe the second effect: “The stepped-wedge design **allows all hospitals to eventually receive the intervention**, which is important when the intervention is deemed likely to be beneficial. This design **helps avoid potential ethical concerns** about sites being randomized to the control arm of a trial throughout the entire study period.” We note that SW designs can now be applied, thanks to our research [KF1–KF6], in situations where it was not previously possible.

Example trials in which our CONSORT guidelines or software have directly affected patients include the evaluation of interventions to 1) Improve obstetrics via bespoke midwifery training [S7.1], 2) Reduce childhood obesity [S7.2–5], 3) Reduce malnutrition in rural China [S7.6], 4) Avoid unnecessary hospital admissions [S7.7], 5) Reduce childhood malaria incidence in Senegal [S7.8], 6) Improve clinical outcomes for myocardial infarction patients in Kerala [S7.9] and 7) Assess the benefits of a ‘telehealth’ app in US veterans [S7.10]. **In each case, the stepped wedge design, facilitated by our results, allowed all patients to receive the beneficial intervention**.

5. Sources to corroborate the impact

[S1] Testimonial: Head of the Biometrical Department, University Hospital of Tours.

[S2] Testimonial: Director of the International Program Evaluation Unit, Hospital for Sick Children, Toronto (30.9.2019).

[S3] Testimonial: Director of the Duke Global Health Institute.

[S4] Testimonial: Medical Officer, World Health Organization (30.9.2019).

[S5] Testimonial: Biostatistician, Washington School of Public Health (12.9.2019).

[S6] Results of searches for keywords 'stepped-wedge' at clinicaltrials.gov and at scholar.google.com and usage statistics for the Shiny CRT Calculator from shinyapps.io (18.10.2020).

[S7] Selected references using Shiny Calculator, CONSORT guidelines, or other underpinning research (full details available):

S7.1. Kenyon, et al. doi: 10.1186/s13063-017-2106-1.

S7.2. Adab, et al. doi: 10.1136/bmj.k211.

S7.3. Adab, et al. doi: 10.3310/hta22080.

S7.4. Breheny, et al. doi: 10.1186/s12889-017-5019-8.

S7.5. Li, et al. doi: 10.1136/bmjopen-2017-018415.

S7.6. Lin, et al. doi: 10.1186/s13063-015-0897-5.

S7.7. Foot, et al. doi: 10.1136/bmjopen-2016-015301.

S7.8. Cissé, et al. doi: 10.1371/journal.pmed.1002175.

S7.9. Huffman, et al. doi.org/10.1016/j.ahj.2016.10.026.

S7.10. Done, et al. doi: 10.1136/bmjopen-2018-022218.

[S8] Quantitative estimates used to support impact claims.

[S9] Statements from *Nature* and *The Lancet* regarding the CONSORT guidelines.