A natural experiment using Scottish clinical data to estimate the real-world effectiveness of adjuvant chemotherapy in breast cancer patients

Principle Investigator:
Dr Peter S Hall, Edinburgh Cancer Research Centre, University of Edinburgh.

Project Description

Introduction

The need for efficient production of real-world evidence: Rigorous assessments of the causal impact of healthcare interventions require unbiased research designs. Randomised controlled trials (RCT) are the gold standard for reliably measuring treatment effects and constitute the primary source of evidence to inform decision-making in health care. However, RCTs are costly, may not be always feasible for practical or ethical reasons, and may have poor external validity. In an attempt to improve the efficiency and real-world relevance of healthcare research, better exploitation of pre-existing or routinely collected healthcare data has been advocated. Observational studies, in which assignment to treatment is non-random, may be a less expensive and more pragmatic source of evidence. However, these approaches rely on strong and sometimes unverifiable assumptions and suffer from various sources of bias. The nonrandomized or "quasi-experimental" designs available to researchers differ in their ability to mirror random assignment under reasonable assumptions; their historical limitation has been that their applicability largely depends on data availability and quality. The new revolution in digital healthcare information coupled with novel analysis methods means that such methods may now offer real solutions to evidence generation.

Regression Discontinuity – the method: The quasi-experimental regression discontinuity (RD) design has been widely used in social sciences[1-4] and has been heralded as a simple to implement and transparent method for providing "real world" effects of treatments, but is underused in healthcare[5-9]. RD applies when participants are assigned to an intervention using a cut-off value (or threshold) of a *continuous* assignment variable, e.g. a risk score or test result. The treatment effect is estimated by comparing outcomes in individuals who lie just below the cut-off with those just above it; under several assumptions, any discontinuity in the outcome at the cut-off can be attributed to treatment.

The clinical need for improved evidence: The use of adjuvant chemotherapy after surgical treatment of early breast cancer is a major contributor to the reduction in mortality from breast cancer over the last three decades. A global collaboration of trialists published a definitive individual patient meta-analysis of 100,000 women with breast cancer, concluding that chemotherapy reduces the risk of dying from breast cancer by about a third[10]. However, the historical clinical trials upon which this evidence relies were performed in highly selected patient populations including few patients older than 70. In the real world we treat patients who would never have been included in those trials due to advanced age, comorbidity, frailty, socioeconomic status or even ethnicity, but such patients are increasingly being treated worldwide with toxic chemotherapy without direct supporting evidence. Recent attempts to conduct further randomised controlled trials in such patients have failed due to poor recruitment, presumably due to a perceived lack of equipoise[11]. It is therefore unknown whether these patients benefit or are harmed by chemotherapy, and alternative methods for measuring their outcomes from treatment are urgently needed. As the decision to proceed with chemotherapy is partly based on specific values of a continuous score, this clinical question appears well suited to the RD method.

A unique opportunity in Scotland: The very high quality and pre-existing linkage of Scottish healthcare and cancer registration datasets makes Scotland an ideal testbed to evaluate a new method to estimate effect sizes from routine data to resolve the above question, which has been unanswerable for many years internationally. We also need to support projects of this nature in Scotland to develop researchers with the necessary data analytic skills, a high priority for the Farr Institute. The results of this project will have

direct clinical impact on Scots with breast cancer, who currently face treatment decisions made in the face of uncertain evidence. This project will also provide feasibility data to support a larger UK-wide study of similar design, and will have immediate international impact on the ongoing fierce debate about using chemotherapy in this situation.

Aims

1. Measure the benefit from chemotherapy in a real-world early breast cancer population
2. Validate commonly used web-based benefit prediction decision tools in Scotland
3. Determine the feasibility of using the regression discontinuity design in this context

Research questions

Is chemotherapy effective in patients with characteristics out-with the eligibility criteria of historical clinical trials?

It remains unknown whether the benefit from chemotherapy observed in clinical trials is maintained in patients out-with the eligibility criteria of those trials. For example half of new diagnoses of invasive breast cancer are in women over the age of 65, but the average age of women in trials was closer to 50. Reliable alternative methods for estimating the effects (benefits and harms) of chemotherapy in these populations are urgently required to guide the selection of patients for chemotherapy. Over 500 patients with early breast cancer are treated with chemotherapy each year in Scotland, many of whom are elderly or comorbid, but these are being treated despite poor evidence. Many more elderly patients are denied chemotherapy due to concerns about non-evidence based treatment. National patient advocacy groups such as the Independent Cancer Patients Voice are calling for research to address the ageism in the availability of existing evidence. By providing an estimate of the treatment effect of chemotherapy in these common but hard to study populations, clinicians will for the first time be able to access evidence to better guide patients in making the difficult decision about whether to undergo chemotherapy.

Are Adjuvant! Online and NHS predict valid, reliable tools for aiding chemotherapy decisions in Scottish patients?

The decision to proceed with chemotherapy for an individual patient is based on the perceived magnitude of benefits and risks set in the context of clinical and patient preferences. For marginal cases the decision is generally informed by a score calculated using one of two online tools: Adjuvant! Online (www.adjuvantonline.com) or NHS Predict (www.predict.nhs.uk). These provide an estimation of the 10-year absolute reduction in expected mortality as a continuous probability. National consensus guidelines (NICE and SIGN Guidelines) recommend either of these scores as the objective basis for a recommendation from a clinician for a patient to undergo chemotherapy or not. The score is likely to provide the basis of the next Scottish Quality Performance Indicator (QPI) in this setting. Patients with a predicted absolute survival benefit <3% are generally not recommended for chemotherapy while chemotherapy is generally recommended for those with a benefit >5%. 3-5% calculated benefit represents a grey area in which patients are offered chemotherapy, but with the caveat that the benefit is marginal and they need to make a personal decision. However, given differences in presentation stage, genetics etc. in Scotland vs. England vs US, where these scores were developed, we need to establish the accuracy of these tools in Scots with breast cancer.

Can the regression discontinuity method provide an answer to question 1?

While the clinical question above appears well suited to the RD method, feasibility and validity need to be demonstrated. If the method proves to be applicable in this context then it will be possible to answer research question 1 and provide valuable information to support clinical decision making in Scotland, and further afield. The high quality historical Scottish cancer datasets provide an ideal test case for a subsequent UK-wide or international research grant application, which will be prepared with the broader aims of: (1) generating knowledge about the benefits and side effects of breast cancer therapies in

large patient cohorts in routine practice settings, and (2) developing guidelines for the appropriate use of RD methods in the light of the increasing availability of routinely collected data.

Translation

"Clinical" side: Clinicians and patients are currently making treatment decisions on the basis of indirect or absent evidence. This research will, for the first time, provide direct evidence to inform patient choice. There are several well-formed dissemination mechanisms available in Scotland, the UK and internationally. These include the Scottish Breast Cancer Clinical Trials Collaborative group, the Scottish Cancer Networks, the SIGN guideline development groups, the UK Breast Cancer Clinical Reference Group, the UK National Cancer Research Institute and the European and American Oncology Societies. We plan a submission to a high profile journal such as The Lancet. A report of the research findings will be made available and presentations of the results will be made to all stakeholder institutions.

"Methods" side: The Farr Institute, MRC Medical Bioinformatics Centres and health services research community are seeking new methods for estimating the impact of treatments and interventions using large routine datasets. This study will contribute to these goals through evaluation of the feasibility and validity of a relevant method. While RD has been applied in social sciences, there is a lack of evidence showing its transferability to the measurement of treatment effectiveness. A MRC-funded project is currently evaluating the use of RD in primary care but there is no example of application in specialised care, such as oncology. Our findings will be of interest to the Farr, health researchers, economists, statisticians and epidemiologists,both as methodologists and as applied researchers. The Scottish government is also investing heavily in using real-world data to evaluate new drugs - suitable methods are urgently required.

"Users of evidence" side: In the longer term, our findings will increase the value of routine datasets and reduce financial and methodological barriers to health technology assessment. Potential beneficiaries include HTA research funders, evaluators, policy makers. If feasibility is proven, this will allow more policy and practice decisions to be based on solid evidence, even when ethics/logistics make RCTs difficult. It will also increase the proportion of evidence to inform decisions that originates from routine practice. There is an acknowledged need for evidence about the real-world effectiveness of chemotherapy across the UK and internationally. We therefore see this Scottish study as laying the foundation of a larger study of similar design using UK-wide data including Hospital Episode Statistics and the English Cancer Registry. We plan a subsequent larger UK proposal to the Cancer Research UK Population Research Committee.

Plan and methods

The research project will be divided into three phases. The first phase will consist of preparing the data for analysis. Then, the external validity of the scoring tools will be assessed and the feasibility of applying a RD design will be determined. Finally, estimates of treatment effects from the RD design will be compared to treatment effects extracted from the literature.

Data preparation (months 1-6)

Patients with a new diagnosis of invasive breast cancer treated with curative surgical resection between 2001 and 2011 will be identified from the Scottish Cancer Registry providing a sample size greater than 24,000 patients (estimated from Scottish Cancer Networks Audit data). A data extract from the registry (SMR06) containing staging and initial treatment information is already linked to hospital discharge data (SMR01) and mortality data and will be provided by Information Services Division (ISD) and stored for analysis within the National Services Scotland Safe Haven. Approval will be sought from the Public Benefit Privacy Panel.

Information required to calculate the scores central to the implementation of the empirical strategy are shown in table A. The Scottish Cancer Registry has been shown to be global leader in the completeness and reliability of data[12,13], with few discrepancies compared to

other nations. Based on pilot data we anticipate missing data in surgically treated patients being below 5% for the majority of required variables. During this phase of the project, data-fields for a sample of the population will be cross-validated with Network Audit teams which have 100% source data verification since 2008 to provide a more robust quantification of missing data. Co-morbidity scores will be calculated from pre-diagnosis inpatient diagnoses (SMR01) using methods previously validated by ISD. In all analyses described below, the sensitivity of results to missing data will be assessed by comparing results from: 1) a complete case analysis and 2) an analysis involving multiple imputation.

Table A: Predictors required for Adjuvant! Online and NHS Predict

	Adjuvant! Online	**NHS Predict**
Age at diagnosis	Continuous	Continuous
Mode of detection	n/a	Screen-detected | Symptomatic | unknown
Tumour size	0.1-1 |1.1-2 | 2.1-3 | 3.1-5 | >5cm	Continuous (mm)
Tumour grade	1 | 2 | 3 | unknown	1 | 2 | 3 | unknown
Number of positive nodes	0 | 1-3 | 4-9 | >9	Continuous
Oestrogen receptor status	Positive | Negative	Positive | Negative
HER2 status	n/a (but can adjust for)	Positive | Negative | unknown
Ki67 status	n/a	Positive | Negative | unknown
Comorbidity	Perfect health | Average for age | minor | moderate | severe | v. severe	n/a

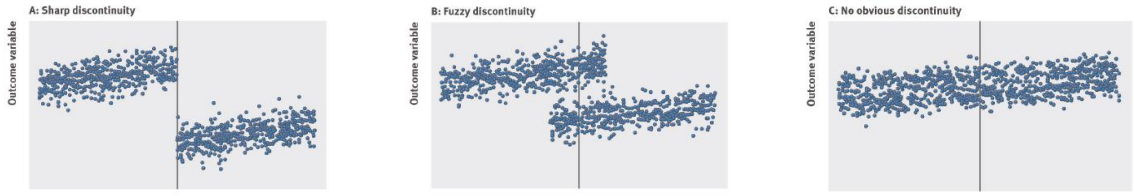
External validation of the risk score (months 7&8):

We will first assess to what extent the risk scores are valid in a Scottish population, i.e. whether the score accurately predicts overall survival and breast cancer specific survival. Following previous studies[14,15], we will compare predicted 10-year survival estimates using Adjuvant! Online and NHS Predict with observed outcomes. Discrimination and calibration of the different tools will be assessed by calculating sensitivity and specificity, area under the receiver-operator characteristic (ROC) curve (AUC), calibration curves and net-reclassification index[16-18]. Outputs of this analysis will therefore include plots of the difference between observed and predicted outcomes for overall and breast cancer specific survival and AUC and their associated c-indices.

Assessment of score-based treatment assignment mechanism (months 9-11):

A crucial step for implementing the RD method is the assessment of the treatment assignment mechanism, i.e. whether the probability of receiving the treatment changes sharply around score thresholds, and whether patients with scores close to the threshold are observationally similar[9]. We will first describe the distribution of predicted 10-year survival benefit from chemotherapy computed using NHS Predict and Adjuvant! Online and identify potential clusters around the decision thresholds (i.e. 3% and 5%). Figure 1 illustrates the three main situations encountered when analysing data anticipated to be suitable to RD. Panel A illustrates the "ideal" case, where there is strict adherence to the cut-off rule ("sharp" RD design). In such case, treatment allocation is clear around the threshold. Conversely, panel B illustrates a case where many individuals with values above the cut-off are not treated and others below the cut-off are treated ("fuzzy" RD design). Panel C reflects a case where there is no discontinuity at the cut-off and where RD would not be appropriate.

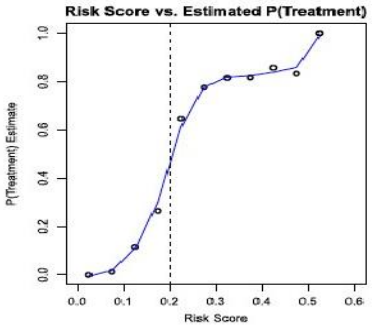
Figure 1: Simulated plots of treatment assignment variable against outcome.



Source: O'Keeffe et al. (2014)[6]

Given the weight of clinician and patient preferences on the decision to give chemotherapy, and the number of potential unobserved treatment determinants, the data for our study will likely correspond to a "fuzzy" RD design. However, this will have to be formally determined by analysing the probability of receiving chemotherapy treatment as a function of the score, and carefully investigating changes at the decision thresholds. The example figure below illustrates one possible shape for the probability to be treated around the threshold of a risk score[8]. In this particular case, we observe a sharp change in the probability to be treated around the threshold (0.2 in this case), which is what is required for identification.

Figure 2: Probability to receive treatment around the cut-off Source: Geneletti et al. (2015)[8]



Such analysis, and the implementation of RD more generally, requires sufficient sample size as the effects of interests are identified in a sub-population of patients with scores in the vicinity of the thresholds. Based on Lothian audit data, currently 28% of surgically treated patients (c. 7000 patients) fall *within* the 3-5% marginal predicted benefit category, which ensures sufficient sample size to implement the RD design and base the estimation of local average treatment effects on a reasonably large sample size.

As a last step to assess the internal validity of the RD design, we will assess covariate balance around the threshold using patient characteristics not included in the scoring process, such as comorbidity and deprivation[9]. Practically, we will perform a series of statistical tests to assess the difference in these characteristics between groups of patients based on their calculated risk score. This will be conducted for a range of "bandwidths" around the threshold (i.e. 2-3% v.s. 3-4%; 1-3% vs. 3-5%, etc.). These preliminary analyses will allow us to assess the likely biases influencing the assignment mechanism and to make a decision on the most appropriate threshold value to use. In addition to the mean estimates, we will verify whether the distribution of these variables are smooth around the threshold, as a key assumption of the RD design is that nothing else other than the probability of being treated changes sharply around the threshold.

Treatment effect estimates and assessment of bias (months 12-15):

If the validity of applying RD to these data is demonstrated, the design will be implemented in the full sample of 24,000 breast cancer patients to obtain reference treatment effect estimates. Formal criteria for progressing with this stage will depend on the feasibility of using the RD design in this setting and will be ratified by our Advisory Group. Based on the work of Moscoe et al 2015[9] they will depend on demonstration that: (i) the assignment variable is measured and reported continuously, (ii) the outcome variables are observed for all patients, independent of whether they were assigned the treatment or not and, (iii) precise information on how treatment is assigned to patients is available to determine whether the design is "sharp" or "fuzzy." Table B provides an overview of the main elements of the "fuzzy" RD design using the formal notation of the causal inference

literature. Briefly, our setting resembles that of an instrumental variable strategy where the cut-off values are associated with discrete changes in the probability to receive the treatment (i.e. impact of the score cut-off on expected outcomes weighted by the impact of the cut-off on the probability to be treated). We will focus on overall survival and breast cancer-specific survival as outcomes at 5 and 10 years and will assess the sensitivity of the results to methodological choices (i.e. the "bandwidth" around the threshold and functional form).

We will then compare RCT-based treatment effects to those obtained from RD in a trial-eligible sub-population[19]. Relevant RCT-based treatment effects will be extracted from the literature along with details on the population in which these effects were obtained. When possible, treatment effects obtained in subgroups of patients will be extracted (by age group, gender, etc.). The RD design will then be applied in a matched sample of patients from the registry data. The matching could be based on a single characteristic (e.g. restriction to patients in a particular age group) or involve multiple criteria. We will use the procedure described by Altman and Bland[20] to compare the treatment effects estimates obtained with the RD design to those extracted from the literature. This will allow quantifying the potential bias of RD as compared to RCT and assessing whether RD can be implemented routinely for the assessment of the real-world effectiveness of chemotherapy. If the estimates derived from RD in the RCT populations match within 10%, we will finally apply RD methods to subgroups of individuals for whom RCT evidence is lacking, such as older patients and those with co-morbidities or high social deprivation to provide a first empirical estimate for these groups.

Table B: formal notation and application to Adjuvant Chemotherapy (AC)

	Notation	**Application to AC**
Outcomes of interest	Y	overall survival, breast cancer-specific survival
Continuous assignment variable	X	Estimated absolute 10-year survival benefit from chemotherapy obtained from the NHS PREDICT tool and the Adjuvant! Online tool
Cut-off value (from guidelines)	x_0	X<3%: chemotherapy not recommended 3%<X<5%: chemotherapy discussed as a possible option >5%: chemotherapy recommended **x_0**: 3% or 5%
Cut-off indicator	$Z=1$ if $X \geq x_0$ and $Z=0$ if $X<x_0$	
Treatment indicator	T=1 if patient received treatment T=0 otherwise	
Fuzzy RD: Local Average Treatment Effect	$LATE=\frac{E(Y\|Z=1)-E(Y\|Z=0)}{E(T\|Z=1)-E(T\|Z=0)}$	

Key references

1. Lee, D.S. and Lemieux, T. Journal of Economic Literature, 2010. 48(2): 281-355. **2.** Cook, T.D. J of Econometrics, 2008. 142(2): 636-654. **3.** Imbens, G.W. and Lemieux, T. Journal of econometrics, 2008. 142(2): 615-635. **4.** Hahn, J., P., et al. Econometrica, 2001. 69(1): 201-209. **5.** Venkataramani, A.S., et al. BMJ 2016; 352 :i1216. **6.** O'Keeffe, A.G., et al. BMJ, 2014. 349: g5293. **7.** Bor, J., et al. Epidemiology, 2014. 25(5): 729. **8.** Geneletti, S., et al. Statistics in medicine, 2015. **9.** Moscoe, E., J., et al. Journal of clinical epidemiology, 2015. 68(2): 122-133. **10.** EBCTC. The Lancet, 2012. 379(9814): 432-444. **11.** Leonard, R., et al. British journal of cancer, 2011. 105(9): 1260-1266. **12.** Brewster, D.H., and Stockton, D.L. The Breast, 2008. 17:106-108. **13.** Brewster, D.H., et al. European Journal of Cancer, 38: 414-417. **14.** de Glas, N.A., et al. The lancet oncology, 2014. 15(7):722-729. **15.** Wishart, G., et al. European Journal of Surgical Oncology (EJSO), 2011. 37(5): 411-417. **16.** Crowson, C.S., et al. Stat Methods Med Res 2014; DOI:10.1177/0962280213497434. **17.** Harrell, F.E. Jr., et al. Stat Med. 1996;15:361–387. **18.** Heagerty, P.J. and Zheng, Y. Biometrics 2005; 61:92–105. **19.** Benson, K. and Hartz, A.J. NEJM. 2000; 342:1878-1886 **20.** Altman, D.G. and Bland, J.M. BMJ. 2003;326:219. **21.** Pauriah, M., et al. Heart. 2014;0:1–6. doi:10.1136/heartjnl-2013-304678.