STUDY PROTOCOL

Effectiveness of pre-operative anaemia screening and increased Tranexamic acid dose on outcomes following unilateral primary, elective total hip or knee replacement: a statistical analysis plan for an interrupted time series and regression discontinuity study [version 1; peer review: 2 approved]

Ashley B. Scrimshire ^{1,2}, Caroline Fairhurst ¹, Catriona McDaid ¹, David J. Torgerson ¹

¹Department of Health Sciences, University of York, York, UK

²Northumbria Healthcare NHS Trust, Wansbeck, UK

V1 First published: 01 Apr 2020, 9:224
<https://doi.org/10.12688/f1000research.22962.1>
 Latest published: 13 May 2021, 9:224
<https://doi.org/10.12688/f1000research.22962.2>

Abstract

Perioperative blood transfusion is associated with poorer postoperative outcomes following hip and knee replacement surgery. Evidence for the effectiveness of some measures aimed at reducing blood transfusions in this setting are limited and often rely on weak pre-post study designs. Quasi-experimental study designs such as interrupted time series (ITS) and regression discontinuity design (RDD) address many of the weaknesses of the pre-post study design. In addition, *a priori* publication of statistical analysis plans for such studies increases their transparency and likely validity, as readers are able to distinguish between pre-planned and exploratory analyses. As such, this article, written prospective of any analysis, provides the statistical analysis plan for an ITS and RDD study based on a data set of 20,772 primary elective hip and knee replacement patients in a single English NHS Trust. The primary aim is to evaluate the impact of a preoperative anaemia optimisation service on perioperative blood transfusion (within 7 days of surgery) using both ITS and RDD methods. A secondary aim is to evaluate the impact of a policy of increased tranexamic acid dose given at the time of surgery, using ITS methods.

Open Peer Review

Approval Status

	1	2
version 2 (revision) 13 May 2021		
version 1 01 Apr 2020	 view	 view
1. David Reeves , University of Manchester, Manchester, UK 2. Al Ozonoff , Boston Children's Hospital, Boston, USA Harvard Medical School, Boston, USA		
Any reports and responses or comments on the article can be found at the end of the article.		

Keywords

Regression discontinuity, interrupted time series, quasi-experimental, anaemia, orthopaedics

Corresponding author: Ashley B. Scrimshire (ashley.scrimshire@nhs.net)

Author roles: **Scrimshire AB:** Conceptualization, Methodology, Writing – Original Draft Preparation, Writing – Review & Editing; **Fairhurst C:** Writing – Review & Editing; **McDaid C:** Writing – Review & Editing; **Torgerson DJ:** Conceptualization, Writing – Review & Editing

Competing interests: No specific funding was received for this study. However, all authors are involved in a clinical trial which is partially funded by Vifor Pharma who market an IV iron preparation. This manuscript has no direct relevance, but the trial topic was related to anaemia management. The lead author (AS) is currently undertaking a PhD which is partially funded by the same trial grant. Vifor Pharma (or any other funding body) have had no input, involvement or influence in this manuscript, and have not seen the manuscript prior to submission. York Trials Unit acknowledge funding from the British Orthopaedic Association.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2020 Scrimshire AB *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Scrimshire AB, Fairhurst C, McDaid C and Torgerson DJ. **Effectiveness of pre-operative anaemia screening and increased Tranexamic acid dose on outcomes following unilateral primary, elective total hip or knee replacement: a statistical analysis plan for an interrupted time series and regression discontinuity study [version 1; peer review: 2 approved]** F1000Research 2020, 9:224 <https://doi.org/10.12688/f1000research.22962.1>

First published: 01 Apr 2020, 9:224 <https://doi.org/10.12688/f1000research.22962.1>

Introduction

Peri-operative red blood cell (RBC) transfusion is associated with poorer post-operative outcomes across surgical disciplines, including elective total hip (THR) and knee replacement (TKR) surgery¹⁻³. Multi-modal patient blood management (PBM) programmes aim to reduce blood transfusions and the associated complications. Two core elements of PBM are peri-operative tranexamic acid (TXA) and pre-operative anaemia optimisation. However, debate exists around optimal TXA dose and there is a lack of high quality randomised controlled trials (RCT) into preoperative anaemia optimisation, with much of the evidence coming from pre-post design observational studies. The pre-post study design is common in the medical literature and causal associations are often inferred from them. However, they are subject to several flaws, including being unable to separate temporal changes from intervention effect and not accounting for regression to the mean⁴. This frequently leads to over-estimation of a treatment effect and has been described as the weakest observational study method^{4,5}.

Although RCTs are considered the gold standard for evaluating changes in healthcare, they are not always feasible and the results may not always be generalisable to real world populations^{4,6,7}. A recent study comparing characteristics of patients recruited to peri-operative medicine RCTs and national registry data, shows significant differences in age, sex and ethnicity exist, potentially limiting the generalisability of RCT results⁸. In addition, an RCT into preoperative anaemia optimisation may prove challenging as this practice is already recommended in multiple guidelines, as part of wider PBM programmes⁹⁻¹¹. Where an RCT is not feasible quasi-experimental study designs such as interrupted time series (ITS) and regression discontinuity (RDD), can provide more robust evidence as they eliminate some of the threats to internal validity seen in pre-post studies.

The prospective publication of statistical analysis plans (SAP) for observational studies increases their transparency and likely their validity, as readers are able to distinguish between pre-planned and exploratory analyses¹². This paper, written prospective of any analysis being performed, provides the SAP for a quasi-experimental study using ITS and RDD methods on a large dataset of elective THR and TKR patients from an English NHS Trust.

The primary aim of this study is to evaluate the clinical effectiveness of introducing a preoperative anaemia optimisation programme, using iron, for patients undergoing primary elective THR or TKR surgery. A secondary aim is to evaluate the clinical effectiveness of introducing a policy of increased intravenous TXA dose on induction of anaesthesia (15mg/kg maximum 1.2g increased to 30mg/kg max. 2.5g). Both interventions take place in the presence of a well-established, multi-modal enhanced recovery programme, detailed elsewhere¹³.

Although similar in design, ITS and RDD examine data from a different perspective. ITS is concerned with population level changes over time, whilst RDD uses patient level data to focus on effects on outcomes around intervention thresholds. These two analyses will provide complimentary results on the effectiveness

of introducing a preoperative anaemia optimisation programme and an increased TXA dose of 30 mg/kg in an NHS Trust¹⁴.

Statistical Analysis Plan

Data source

Over time the orthopaedic department at Northumbria Healthcare NHS Foundation Trust (NHCT) has introduced a range of interventions aimed at improving post-operative outcomes for patients undergoing elective lower limb arthroplasty. These have been well documented in a series of pre-post cohort studies^{13,15,16}. Details of the interventions and a timeline are given in [Table 1](#) and [Figure 1](#).

As part of on-going service evaluation, a large dataset of 20,772 patients who have undergone primary elective THR or TKR at NHCT has been compiled. This includes data from hospital electronic record systems, such as the Patient Administration System and Blood Transfusion database, and a prospectively maintained database for the pre-operative anaemia screening service. Data includes patient demographics, comorbidities, pre-operative anaemia screening results (i.e. haemoglobin concentration, Hb), anaemia treatment given, operative details, post-operative complications, blood transfusions and length of hospital stay (LoS). The full dataset covers a time period from January 2008 to March 2019.

Ethical approval was not required as this is a retrospective study of routinely collected data. Local Caldicott guardian approval was given for use of this data. Data flow will be presented in a STROBE diagram in the resulting publication¹⁷. Population characteristics (age, gender, comorbidities, type of surgery) and descriptive statistics will be presented in tables for the cohorts being studied. Analyses will be performed using R and RStudio (version R-3.6.2 for mac, R Core Team 2013, <http://www.R-project.org/>) on an intention to treat basis. Results will be presented in terms of absolute and relative values with 95% confidence intervals where appropriate. Results will be considered statistically significant if the p-value ≤ 0.05 .

Outcomes

The primary outcome is the proportion of patients receiving perioperative allogenic RBC transfusion (within 7 days of surgery). Secondary outcomes are the quantity of blood transfused (RBC units), LoS (in days), critical care admission rate (within 30 days) and emergency readmission rate (within 30 days)^{1,2,16}.

Interrupted Time Series

ITS using segmented regression has several strengths over the pre-post study design. In particular, it controls for secular trends over time, provides powerful, easy to understand visual outputs and may improve generalisability to the wider population^{6,7,18}. For this study, data are available to evaluate both policies described above in an ITS analysis.

The two interventions in this study were introduced at specific, well defined time points, allowing for clear separation of pre- and post-intervention periods ([Figure 1](#)). An early step in ITS analysis is to generate summary statistics for each time period and undertake simple pre-post comparisons¹⁹. This will be performed

Table 1. Summary of interventions introduced in Northumbria Healthcare NHS Foundation Trust aimed at improving arthroplasty patient outcomes.

Author, year	Intervention	Control cohort	Intervention cohort	Statistically significant outcomes reported by authors (p<0.05)
Malviya, 2011 ¹³	Multimodal Enhanced Recovery Programme*	Jan 2004 – April 2008 (n=1500)	May 2008 – Nov 2009 (n=3000)	Reduced mortality, transfusions and LoS
Morrison, 2017 ¹⁵	Increased dose of IV Tranexamic acid (30mg/kg max 2.5g)	May 2008 – July 2011 (n=2637)	Feb 2012** – Jan 2013 (n=1814)	Reduced transfusions
Pujol-Nicholas, 2017 ¹⁶	Introduction of pre-operative anaemia screening and optimisation pathway	Feb 2012 – Jan 2013 (n=1814)	Feb 2013 – May 2014 (n=1622)	Reduced transfusion, readmissions, critical care admissions, LoS and costs.

* Includes introduction of IV Tranexamic acid, 15mg/kg max 1.2g, at induction of anaesthesia

** Policy introduced in August 2011. This study allowed a 6-month implementation period to ensure the change in practice had been adopted

LoS = length of hospital stay

Figure 1. Timeline of patient blood management interventions introduced at Northumbria NHS Foundation Trust. ERP = Enhanced Recovery Programme IV = Intravenous, TXA = Tranexamic acid, ITS = interrupted time series, RDD = regression discontinuity design.

in this study and later compared to the results from ITS and RDD analyses.

Data description. ITS is said to work best with short-term outcomes that change quickly after implementation of an intervention or after a clearly defined lag period⁷. This study is examining short-term outcomes, however, some delays to observed changes in outcomes are expected. The orthopaedic department has previously reported that a 6-month lag period was required to fully adopt the increased TXA dose policy¹⁵. This same lag will therefore be incorporated into the ITS analysis. Regarding the

introduction of the preoperative anaemia optimisation programme, staff running the anaemia service report that this started promptly on 01/02/2013, after detailed planning, and uptake was rapid. However, a lag to observed changes in outcomes will be inevitable due to surgical waiting list times. Comparing screening and surgery dates for the first 10 anaemic and 10 non-anaemic patients from the cohort shows all but one had their surgery within 6 months of screening. Therefore, a 6-month implementation period is also considered appropriate following introduction of the anaemia optimisation service (Figure 1). Lag periods will be accounted for by excluding this data from

analysis²⁰. ITS requires sequential measures of the outcome, at regular intervals, before and after the intervention time points^{19,20}. In keeping with many ITS studies, individual level outcome data will be converted to, and presented as, proportions or means at monthly intervals and a segmented-regression analysis performed²⁰. ITS plots will be generated and visually inspected to determine if linear or non-linear regression modelling is appropriate. A minimum of 8 data points pre- and post-intervention are desirable^{19,20}. It is expected the shortest time frame being analysed in this study will include 12 months/data points, thus surpassing this requirement.

Addressing threats to validity. Time varying confounders are the main threat to validity of ITS studies²⁰. These are specific to each ITS study and are carefully considered later in this SAP. However, the most robust way to account for time varying confounders, even those that are unknown, is to model against a control group. This could either be a different population not exposed to the intervention, or if individual level data are available, by splitting the data into two groups, one group targeted by an intervention and another who are not. In this study, data for a different group of patients is not available for either intervention. The TXA policy is targeted at all THR or TKR patients so this data cannot be split. However, the anaemia optimisation policy targets a specific group of anaemic patients so the population can be split into two groups to increase the robustness of this analysis. As such the two interventions will be modelled separately, the TXA intervention without a control group, the anaemia screening intervention with a control group.

Time varying confounders specific to the primary outcome of this study may include other PBM interventions, those relevant will be discussed in turn. A restrictive blood transfusion policy was introduced Trust-wide in 2007 and has been unchanged since^{15,16}. A multimodal enhanced recovery programme (ERP), including IV TXA on induction of anaesthesia, was introduced in May 2008. In keeping with other similar policy changes in this unit a 6-month implementation period for the ERP is considered appropriate. To account for this, data from 1st January 2008 to 31st October 2008 will be excluded from this analysis. In addition, patient warming has been introduced locally²¹ but a Cochrane review shows this does not affect surgical transfusion rates, so will not be considered any further²². Intra-operative cell salvage has never been routinely used locally for the procedures being studied. To the best of our knowledge, no other relevant cointerventions have been introduced during the study period. Any unaccounted for, gradual changes in practice, would be detected in the pre-intervention slope of the TXA analysis and in the control group for the anaemia analysis.

Other considerations include changes in data coding, validity and reliability over time. The data for this study is considered reliable as it comes from a number of NHS Trust electronic databases detailed earlier in this paper. There have been no material changes to data collection methods or outcome reporting over the study period.

Changes in the population over time can also affect ITS reliability, however there have been no known substantial changes in the

population served by NHCT over the study period. This study includes a continually enrolled population, so is not subject to population attrition over time. Although no changes to diagnostic criteria for ischaemic heart disease (IHD) are known to have occurred during the study period, this comorbidity is specifically mentioned in the NHCT transfusion policy and lowers the threshold for considering transfusion. For completeness, rates of IHD will be plotted against time and visually inspected for any patterns, particularly around the time of the interventions. If required IHD will be included in the ITS modelling.

Developing the model. Autocorrelation, including seasonality, will be tested using the Durbin-Watson test including a lag of up to 12 time points (to account for seasonality), visual inspection of residual plots for patterns over time, and interpretation of autocorrelation and partial autocorrelation function plots. If identified, autoregressive and/or moving average relationships will be included in the final ITS model^{19,23}. Data will be inspected for wild data points and where identified, explanations will be sought, and exclusion considered.

Sensitivity analysis. An optimal model will be developed and described for these ITS analyses. The impact of decisions taken during this process such as inclusion/exclusion of wild data points and autocorrelation adjustments will be tested in sensitivity analyses. Further analyses of data stratified by surgery type (THR or TKR) and/or by patient gender, will be conducted if data permits, as these may impact on outcomes.

Regression discontinuity

RDD estimates the local average treatment effect when treatment decisions are based around a cut-off value for a continuous variable²⁴. For example, giving iron (the treatment) with the intention of reducing RBC transfusion and LoS (the outcomes) to patients whose Hb (the assignment variable) falls below a pre-defined cut-off of 120g/L for females or 130 g/L for males (the threshold). RDD makes use of this threshold and assumes that individuals who lie just above it belong to the same population as those who lie just below it, and assignment to treatment or not is considered random²⁵.

The main strength of RDD lies in its ability to achieve a balance of unobserved factors in patients that fall, by chance, either side of the threshold value, much like an RCT²⁶. The local nature of the effect examined in RDD can also be used in optimising threshold levels. In this case we may be able to examine if a threshold Hb of 120 or 130g/L may be more appropriate for females, as is being suggested in some studies^{11,26–28}. As the TXA policy affects all patients it is only possible to conduct an RDD analysis for the anaemia screening programme, using data since the inception of this programme (1st February 2013, [Figure 1](#)).

Data description. In this study the continuous assignment variable will be preoperative Hb concentration. The outcome assessment, for primary and secondary outcomes (listed above), are observed universally for patients who receive treatment with iron or not. Details of how treatment is assigned has been previously reported, and is shown in [Figure 2](#)¹⁶. Notably the treatment thresholds are

Figure 2. Northumbria Healthcare NHS Foundation Trust anaemia pathway demonstrating haemoglobin threshold values used to determine treatment¹⁶.

different for males (Hb 130g/L) and females (120g/L), so data will be split by gender for analysis. Also, some patients are referred back to their GP for further investigations and surgery is deferred (Hb <105g/L (female) or <115g/L (male), ferritin <12 or >100ng/mL). As such patients referred back to their GP will be excluded from this analysis as allocation to iron treatment or not is unclear¹⁶.

Addressing threats to validity. Manipulation of treatment status by patients through the assignment variable (Hb) is highly unlikely. However, it is possible the reporting of the assignment variable could be manipulated by clinicians, though there is a protocol which healthcare professionals are required to strictly adhere to. To assess the statistical integrity of the data assignment variable (Hb) data will be plotted on a histogram and visually inspected for bunching around the threshold values.

To test if groups either side of the threshold are comparable, plots of Hb against average number of comorbidities per patient and age will be generated. If no differences are seen just either side of the threshold, this will support the assumption that assignment around the threshold is random and establish there has been no manipulation of treatment status, similar to an RCT²⁶. If differences are seen, and it is possible the more anaemic patients (lower Hb) have more comorbidities and/or are older, then this will be factored into optimal bandwidth size selection, detailed below. Similar to ITS, RDD is sensitive to co-interventions introduced around the threshold Hb value. There are no known co-interventions introduced locally for this patient cohort around the threshold Hb values.

Developing the model. Hb data will be presented in bins, the size of which will depend on the data available, but will be either

1, 2 or 5 g/L each. Outcome data for this study is in the form of discrete variables (transfused/readmitted/critical care admission; yes/no, number of inpatient days). As such outcome data will be converted to a probability (i.e. risk of transfusion) or average (i.e. mean number of days) for each bin. 95% confidence intervals will be plotted alongside the probabilities or averages where applicable.

A plot of assignment variable (Hb) against treatment status will be created to confirm if a sharp or fuzzy discontinuity design is most appropriate²⁶. Separate scatter plots of primary and secondary outcomes against Hb will be created. These will be inspected visually for a jump at the threshold value, indicative of a treatment effect, and to determine if linear or non-linear modelling is most appropriate. Data driven methods will be used to determine optimal bandwidth sizes either side of the threshold value.

Sensitivity analysis. The robustness of the resulting effect size estimate will be tested by constructing multiple models of varying bandwidth size. Data will be inspected for wild data points and consideration given to exclusion. Sensitivity analysis with and without wild data points will be performed. Further analyses with data stratified by surgery type (THR or TKR) will be conducted where data permits.

Comparing ITS and RDD

Both ITS and RDD have significant advantages over the typical pre-post analysis often seen in medical literature. When an intervention is introduced rapidly and short-term outcomes are frequently assessed, ITS can be considered a sub-type of RDD in which the assignment variable is time and the cut-off occurs when the policy is introduced²⁶.

It is unusual to have a dataset amenable to both types of analysis, however they provide different perspectives. Whilst both designs share the strength of not being bound by the selective inclusion criteria of an RCT, thus potentially improving generalisability, they also have their limitations.

In the case of RDD, in order to ensure groups either side of the threshold are similar the focus is on an effect close to the threshold value. (i.e. patients with Hb 119 or 121g/L are likely very similar, but patients with Hb 90 or 140 are likely different in other, unmeasured parameters). This limits the generalisability of findings to values that lie far from the threshold. In the case of ITS the results can be impacted by several factors such as autocorrelation and unmeasured confounders, which we have

attempted to address in the analysis design. Also, the findings from ITS can only indicate an associative not a causal relationship between intervention and outcomes. Whereas RDD has the potential to demonstrate causation.

Dissemination

Publication of study results will be sought in a high impact journal.

Study status

Study data has been collected and analysis pending awaiting publication of this statistical analysis plan.

Data availability

No data is associated with this article.

References

- Fowler AJ, Ahmad T, Phull MK, *et al.*: **Meta-analysis of the association between preoperative anaemia and mortality after surgery.** *Br J Surg.* 2015; 102(11): 1314–24.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Viola J, Gomez MM, Restrepo C, *et al.*: **Preoperative anemia increases postoperative complications and mortality following total joint arthroplasty.** *J Arthroplasty.* 2015; 30(5): 846–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Musallam KM, Tamim HM, Richards T, *et al.*: **Preoperative anaemia and postoperative outcomes in non-cardiac surgery: a retrospective cohort study.** *Lancet.* 2011; 378(9800): 1396–407.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Torgerson DJ, Torgerson CJ: **Designing Randomised Trials in Health, Education and the Social Sciences.** Palgrave Macmillan. 2008.
[Publisher Full Text](#)
- Cook TD, Campbell DT: **Quasi-Experimentation: Design and Analysis Issues for Field Settings.** Boston: Houghton Mifflin. 1979.
[Reference Source](#)
- Kontopantelis E, Doran T, Springate DA, *et al.*: **Regression based quasi-experimental approach when randomisation is not an option: Interrupted time series analysis.** *BMJ.* 2015; 350: h2750.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Penfold RB, Zhang F: **Use of interrupted time series analysis in evaluating health care quality improvements.** *Acad Pediatr.* 2013; 13(6 Suppl): S38–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Lindsay WA, Murphy MM, Almghairi DS, *et al.*: **Age, sex, race and ethnicity representativeness of randomised controlled trials in peri-operative medicine.** *Anaesthesia.* 2020; 1–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
- NICE Guideline NG24 Methods, evidence and recommendations. *Natl Inst Heal Care Excell.* 2015.
[Reference Source](#)
- Mueller M, Van Remoortel H, Meybohm P, *et al.*: **Patient Blood Management: Recommendations from the 2018 Frankfurt Consensus Conference.** *JAMA.* 2019; 321(10): 983–97.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Muñoz M, Acheson AG, Auerbach M, *et al.*: **International consensus statement on the peri-operative management of anaemia and iron deficiency.** *Anaesthesia.* 2017; 72(2): 233–47.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hiemstra B, Keus F, Wetterslev J, *et al.*: **DEBATE-statistical analysis plans for observational studies.** *BMC Med Res Methodol.* 2019; 19(1): 233.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Malviya A, Martin K, Harper J, *et al.*: **Enhanced recovery program for hip and knee replacement reduces death rate.** *Acta Orthop.* 2011; 82(5): 577–81.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Portela MC, Pronovost PJ, Woodcock T, *et al.*: **How to study improvement interventions: a brief overview of possible study types.** *BMJ Qual Saf.* 2015; 24(5): 325–36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Morrison RJ, Tsang B, Fishley W, *et al.*: **Dose optimisation of intravenous tranexamic acid for elective hip and knee arthroplasty: The effectiveness of a single pre-operative dose.** *Bone Joint Res.* 2017; 6(8): 499–505.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Pujol-Nicolas A, Morrison R, Casson C, *et al.*: **Preoperative screening and intervention for mild anemia with low iron stores in elective hip and knee arthroplasty.** *Transfusion.* 2017; 57(12): 3049–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
- von Elm E, Altman DG, Egger M, *et al.*: **Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies.** *BMJ.* 2007; 335(7624): 806–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wagner AK, Soumerai SB, Zhang F, *et al.*: **Segmented regression analysis of interrupted time series studies in medication use research.** *J Clin Pharm Ther.* 2002; 27(4): 299–309.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bernal JL, Cummins S, Gasparrini A: **Interrupted time series regression for the evaluation of public health interventions: a tutorial.** *Int J Epidemiol.* 2017; 46(1): 348–55.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jandoc R, Burden AM, Mamdani M, *et al.*: **Interrupted time series analysis in drug utilization research is increasing: systematic review and recommendations.** *J Clin Epidemiol.* 2015; 68(8): 950–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Althoff FC, Neb H, Herrmann E, *et al.*: **Multimodal Patient Blood Management Program Based on a Three-pillar Strategy: A Systematic Review and Meta-analysis.** *Ann Surg.* 2019; 269(5): 794–804.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Madrid E, Urrutia G, Roqué i Figuls M, *et al.*: **Active body surface warming systems for preventing complications caused by inadvertent perioperative hypothermia in adults.** *Cochrane Database Syst Rev.* 2016; (4): CD009016.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Bhaskaran K, Gasparrini A, Hajat S, *et al.*: **Time series regression studies in environmental epidemiology.** *Int J Epidemiol.* 2013; 42(4): 1187–95.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Venkataramani AS, Bor J, Jena AB: **Regression discontinuity designs in healthcare research.** *BMJ.* 2016; 352: i1216.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- O’Keeffe AG, Geneletti S, Baio G, *et al.*: **Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data.** *BMJ.* 2014; 349: g5293.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Moscoe E, Bor J, Bärnighausen T, *et al.*: **Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice.** *J Clin Epidemiol.* 2015; 68(2): 122–33.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Munting KE, Klein AA: **Optimisation of pre-operative anaemia in patients before elective major surgery - why, who, when and how?** *Anaesthesia.* 2019; 74 Suppl 1: 49–57.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chaplin DD, Cook TD, Zurovac J, *et al.*: **The Internal and External Validity of the Regression Discontinuity Design: A Meta-Analysis of 15 Within-Study Comparisons.** *J Policy Anal Manag.* 2018; 37(2): 403–29.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 02 November 2020

<https://doi.org/10.5256/f1000research.25351.r72020>

© 2020 Ozonoff A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

AI Ozonoff

¹ Precision Vaccines Program, Division of Infectious Diseases, Boston Children's Hospital, Boston, MA, USA

² Department of Pediatrics, Harvard Medical School, Boston, MA, USA

The authors provide a prospective statistical analysis plan for a forthcoming study of orthopedic surgical cases from the UK National Health Service (NHS). The study uses quasi-experimental Interrupted Time Series (ITS) and Regression Discontinuity (RD) designs. The statistical analysis plan (SAP) is clear and well-written although there are some aspects that could be clarified. It is understood that many details of the model development will be determined upon examination of the data and thus a prospective statistical analysis plan should not be over-specified. However, there are some important elements of the modeling process that could be further explained in terms of what available methods might be considered.

Specific comments follow:

1. Data source. There could be more added to the section on data sources. The data set includes N=20,772 patients who have undergone primary elective THR or TKR surgeries at NHCT. Are there any surgeries excluded? A brief 1-2 sentences to state explicitly and formally the inclusion and exclusion criteria would be useful.
2. Interrupted Time Series. The exclusion of data from the six month period following each intervention seems overly conservative. Since the data are available for screening versus surgery time, the intervention can be modeled not as a binary (0/1) indicator but rather as a continuous implementation variable ranging from 0 to 1, estimated by the monthly proportion of surgeries for which patients received screening. Thus the effect of the intervention is modeled as a weighted average during the six month period following intervention which is a more efficient use of data and should provide a more precise estimate of the intervention effect.
3. The discussion of threats to ITS validity does not give much credit to the possibility that the patient population may change over the course of the study evaluation. While the authors

note that there are no known changes in the overall population served by NHCT, it seems more plausible that there are shifts in demographics or other characteristics of the population receiving THR or TKR. Simple examinations of sex, age, and other clinical factors over the 11+ years of the study period seem warranted if only to verify that there are no major changes in the study population.

4. The discussion of threats to RDD validity could be sharpened. Most of the methods described involve visual inspection of graphical parameters with little formal testing planned. Comparability of groups on either side of the threshold might test formally the hypothesis of difference between groups as would be done for an RCT. It is not explained how bandwidth selection would address the threat to validity posed by incomparability, especially if an observed difference might be explained by manipulation of the assignment variable.
5. Describing model development, the bins for Hg will be chosen from options of 1, 2, or 5 g/L each with no explanation of what considerations are important nor how the data will drive the decision. Similarly, there should be detail provided on which 'data driven methods' will inform bandwidth selection.
6. There is no discussion of how to determine the functional form of the regression. What alternatives are considered if the relationship between Hg and outcome does not appear linear. There is a mention that non-linear models are considered without much insight into what methods are available in this case.

Minor edits/corrections:

1. Outcomes p3. Rates might be better specified with the appropriate denominator e.g. 30-day critical care readmission rate (per 1000 surgeries).
2. Data description p5. Typo 'determin' => 'determine'.
3. Regression discontinuity p5. Typo 'populationas' => 'population as'.
4. Addressing threats to validity p6. Typo 'adhered to' => 'adhere to'.
5. Final phrase 'Whereas RDD...' is a fragment => combine with the previous sentence.

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistics, epidemiology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 26 Mar 2021

Ashley Scrimshire, University of York, UK, York, UK

Thank you for taking the time to provide considered feedback on our article and for engaging in further discussion on your comments, this is very much appreciated. Your comments have been incorporated into the revised manuscript and a summary of our responses is given below.

Comment: Data source. There could be more added to the section on data sources. The data set includes N=20,772 patients who have undergone primary elective THR or TKR surgeries at NHCT. Are there any surgeries excluded? A brief 1-2 sentences to state explicitly and formally the inclusion and exclusion criteria would be useful.

Response: Thank you for your comment. This has been clarified in the text and Table 2 now presents the eligible procedure codes.

Comment: Interrupted Time Series. The exclusion of data from the six month period following each intervention seems overly conservative. Since the data are available for screening versus surgery time, the intervention can be modeled not as a binary (0/1) indicator but rather as a continuous implementation variable ranging from 0 to 1, estimated by the monthly proportion of surgeries for which patients received screening. Thus the effect of the intervention is modeled as a weighted average during the six month period following intervention which is a more efficient use of data and should provide a more precise estimate of the intervention effect.

Response: This is a very interesting point and thank you for engaging in further discussion on this. The paper has been updated. We now plan to include a sensitivity analysis modelling the intervention as a continuous implementation variable as suggested by the reviewer.

Comment: The discussion of threats to ITS validity does not give much credit to the possibility that the patient population may change over the course of the study evaluation. While the authors note that there are no known changes in the overall population served by NHCT, it seems more plausible that there are shifts in demographics or other characteristics of the population receiving THR or TKR. Simple examinations of sex, age, and other clinical factors over the 11+ years of the study period seem warranted if only to verify that there are no major changes in the study population.

Response: Thank you for highlighting this oversight, we agree with your comments. We will include comparison of key patient demographics and characteristics in the pre- and post-intervention groups. This has been included in the text.

Comment: The discussion of threats to RDD validity could be sharpened. Most of the methods described involve visual inspection of graphical parameters with little formal testing planned. Comparability of groups on either side of the threshold might test formally the hypothesis of difference between groups as would be done for an RCT.

Response: Thank you for your advice on these points. The manuscript has been updated to make this more robust. In particular we plan to generate tables and undertake statistical tests comparing non-outcome characteristics for groups either side of the threshold.

Comment: It is not explained how bandwidth selection would address the threat to validity posed by incomparability, especially if an observed difference might be explained by manipulation of the assignment variable.

Response: As detailed in the updated manuscript, manipulation of the assignment variable by patients is not considered likely in this scenario. However this will be explored within the data.

We have now clarified our analysis plans in the manuscript. As such bandwidth selection is only relevant to our planned non-parametric sensitivity analysis, rather than the primary parametric analyses which will use all data. Sensitivity analyses in which the model incorporates predicted factors that may influence outcome such as age, comorbidities, will be undertaken. In addition, variables that are identified as being unbalanced between the two groups (i.e. as a result of possible manipulation of the assignment variable) will be included as covariates in further sensitivity analyses.

Comment: Describing model development, the bins for Hg will be chosen from options of 1, 2, or 5 g/L each with no explanation of what considerations are important nor how the data will drive the decision. Similarly, there should be detail provided on which 'data driven methods' will inform bandwidth selection.

Response: Agreed, we were not clear on our approach to this. The text has been updated. We intend to first plot data using a range of bin sizes and visually inspect these to rule out ones that are clearly too wide or too narrow. We will go on to conduct F-tests (using 2k dummies and interactions) to identify bin widths that over smooth the data. From the remaining choices we will pick the widest bin size that is not rejected by either F-test. As for bandwidth selection, this is only relevant to our planned nonparametric sensitivity analysis. Here we intend to use the cross-validation method to inform bandwidth selection.

Comment: There is no discussion of how to determine the functional form of the regression. What alternatives are considered if the relationship between Hg and outcome does not appear linear. There is a mention that non-linear models are considered without much insight into what methods are available in this case.

Response: Agreed, we had not been clear on this, the text has now been updated. Our intentions are that after bin size has been selected plots will first be inspected visually. The F-Test approach will then be used to determine the functional form of the regression. Starting with a simple linear model and adding a higher order term until the F-test is no longer statistically significant. Robustness checks for this model in which the outer most 1,5 and 10% of data points are dropped will be conducted.

Comment: Outcomes p3. Rates might be better specified with the appropriate denominator e.g. 30-day critical care readmission rate (per 1000 surgeries).

Response: Agreed, this has now been updated in the text

Comment: Typos/grammar:

Data description p5. Typo 'determin' => 'determine'.

Regression discontinuity p5. Typo 'populationas' => 'population as'.

Addressing threats to validity p6. Typo 'adhered to' => 'adhere to'.

Final phrase 'Whereas RDD...' is a fragment => combine with the previous sentence

Response: These have been corrected, thank you for highlighting.

Competing Interests: None

Reviewer Report 08 October 2020

<https://doi.org/10.5256/f1000research.25351.r70571>

© 2020 Reeves D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

David Reeves

NIHR School for Primary Care Research, Centre for Primary Care and Health Services Research, Manchester Academic Health Science Centre (MAHSC), University of Manchester, Manchester, UK

It is very welcome to see a statistical analysis plan for an observational study submitted for publication, as this is relatively rare. It is also quite brave for the authors to do so: unlike an RCT, the form that the analysis of an observational dataset takes is largely dictated by the data available, and *a priori* plans that seem good in theory frequently need a major overhaul in the light of the actual data. Publishing a plan in advance risks making oneself a hostage to fortune if one's good ideas subsequently turn out to be not so feasible in practice. However, what we do not know in the present instance, is how much of this plan is genuinely *a priori*, and how much based on data exploration and analysis already undertaken, though from the precise sample sizes and details presented I suspect quite a bit. Nonetheless, publication is still very worthwhile since the paper provides a level of detail probably not possible in a paper presenting the actual findings of

the analysis, given the word-length restrictions of most publications.

The authors present a mostly well-written and well-thought-out proposal that uses statistical methods of interrupted time-series and regression discontinuity analysis to evaluate the impact of changes in Hospital Trust policy around care for patients admitted for elective lower limb arthroplasty, on outcomes for those patients. The methods proposed are sound and it is good to see them being applied in this context. My comments below are largely concerned with improving the clarity around specifics of the analyses, issues around seasonal effects, and addressing autocorrelation.

Table 1: Use of the terms “control cohort” and “intervention cohort” here is a little confusing, as the term “controls” is also used later under Addressing Threats to Validity, where it is applied to a sub-group of anaemic patients during the intervention period – i.e. a different control cohort. I would have preferred Table 1 to use terms such as “pre-intervention” and “post-intervention” to avoid confusion.

If I understand Table 1 and Figure 1 correctly, there will be 3 ITS analyses, although the paper could be clearer about this in the text. Moreover, the majority of the intervention cohort for the first ITS (TXA started) will also be part of the control cohort for the second ITS (increased TXA) – since the date ranges overlap – and the intervention cohort for ITS 2 will be identical to the control group for ITS 3 (pre-op anaemia optimisation). If this is the case (or even if it is not) the authors need to clarify the situation here. Overlapping cohorts mean that the analyses will not be independent and may have implications for interpretation of the findings.

For clarity I would like to have seen Figure 1 indicate the control cohort periods for each ITS, as well as the implementation and intervention periods. It took me some time work out how Table 1 related to Figure 1 in terms of the time-periods involved.

Outcomes will be analysed in the form of monthly means or proportions. One issue here, which is not mentioned in the paper, is that the sample size will vary considerably over time. For the first ITS (TXA started) the control period covers approx 50 months and the sample size is 1500, implying a mean sample size of 30 patients per month – very small when the outcome is a proportion; whereas the intervention period is about 20 months with a total sample of 3000, indicating 150 per month. Thus outcome means/proportions will be far more variable over the control period. I haven't checked, but the same may apply to the other ITS analyses. Ideally data-point variability should be taken into account in the analysis, and is something that the authors should at least mention and discuss the implications of, in the paper.

Each ITS analysis will incorporate a 6-month implementation period between the pre and post periods, for which data will be dropped from the model. One concern here is the potential for an annual cycle in the data values. I cannot say if elective lower limb arthroplasty is subject to seasonal variation, but certainly hospital admissions for many other conditions are. The risk here is that 6 months can represent the time between the lowest and highest points in an annual cycle. Thus it is conceivable that at the end of the pre period, the cycle will be at it's lowest point, but at the subsequent start of the post period, it will be at the top (or vice-versa). Particular care will need to be taken to evaluate whether any change in level or trend at this point can be explained by the presence of an annual cycle. The authors acknowledge the potential for seasonality in their discussion of autocorrelation (see below). However, I would like to see a specific sensitivity

analysis designed to assess robustness against the threat of an annual cycle, regardless of the outcome of any tests for autocorrelation, given the use of a 6-month lag.

Tests for autocorrelation, using the Durbin-Watson, are planned, using a lag of up to 12 time-points. However, these tests are likely to have very low power, given the numbers of data-points and the measurement error around the individual values (which at times will be very wide). To interpret a non-significant test as implying an absence of autocorrelation would be highly questionable. The data series will almost inevitably in reality possess autocorrelation, even if undetected by the DW, and in my view it would be better to conduct analysis under the assumption that autocorrelation is present. As I have suggested above, a sensitivity test against an annual cycle should be conducted regardless.

Is the rationale for, and objectives of, the study clearly described?

Yes

Is the study design appropriate for the research question?

Yes

Are sufficient details of the methods provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Not applicable

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistics, Health research.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 26 Mar 2021

Ashley Scrimshire, University of York, UK, York, UK

Thank you for taking the time to provide considered and insightful feedback on our article. Your comments have been addressed in the revised manuscript. A summary of responses is given below.

Comment: Table 1: Use of the terms “control cohort” and “intervention cohort” here is a little confusing, as the term “controls” is also used later under Addressing Threats to Validity, where it is applied to a sub-group of anaemic patients during the intervention period – i.e. a different control cohort. I would have preferred Table 1 to use terms such as “pre-intervention” and “post-intervention” to avoid confusion.

Response: Agreed, this was unclear. Table 1 has now been updated and the terms “pre-intervention” and “post-intervention” have replaced “control cohort” and “intervention cohort” to avoid confusion. This table also now outlines previously published, pre-post

design cohort studies from this unit and does not outline the time periods for this analysis. A new Table 3 in the paper clearly outlines the time periods included in this analysis.

Comment: If I understand Table 1 and Figure 1 correctly, there will be 3 ITS analyses, although the paper could be clearer about this in the text. Moreover, the majority of the intervention cohort for the first ITS (TXA started) will also be part of the control cohort for the second ITS (increased TXA) – since the date ranges overlap – and the intervention cohort for ITS 2 will be identical to the control group for ITS 3 (pre-op anaemia optimisation). If this is the case (or even if it is not) the authors need to clarify the situation here. Overlapping cohorts mean that the analyses will not be independent and may have implications for interpretation of the findings.

Response: We agree that these figures and accompanying explanations could be clearer. The text has been updated to clarify that there will be two primary ITS analyses, plus secondary and sensitivity analyses. Figure 1 has been updated to clearly demarcate the pre- and post-intervention periods for each analysis. A new Table 3 also outlines the planned analyses and the time periods included in each.

Comment: Outcomes will be analysed in the form of monthly means or proportions. One issue here, which is not mentioned in the paper, is that the sample size will vary considerably over time. For the first ITS (TXA started) the control period covers approx 50 months and the sample size is 1500, implying a mean sample size of 30 patients per month – very small when the outcome is a proportion; whereas the intervention period is about 20 months with a total sample of 3000, indicating 150 per month. Thus outcome means/proportions will be far more variable over the control period. I haven't checked, but the same may apply to the other ITS analyses. Ideally data-point variability should be taken into account in the analysis, and is something that the authors should at least mention and discuss the implications of, in the paper.

Response: Thank you for your comment. Data-point variability is expected, although not to the degree in the reviewer comment. Hopefully this is clearer now the time periods that are included in this study have been clarified in response to previous comments. Data-point variability has now been discussed in the amended text. The primary analysis will include all THR/TKR procedures in the dataset. Here the expected counts per month are 100 or more, so proportions will be used.

For the secondary analyses, the data will be split into anaemic and non-anaemic sub-groups. Here, it is expected around 20-30% of patients per month will be anaemic, so the counts are expected to drop. In this instance analyses using proportions and counts will be undertaken.

Comment: Each ITS analysis will incorporate a 6-month implementation period between the pre and post periods, for which data will be dropped from the model. One concern here is the potential for an annual cycle in the data values. I cannot say if elective lower limb arthroplasty is subject to seasonal variation, but certainly hospital admissions for many other conditions are. The risk here is that 6 months can

represent the time between the lowest and highest points in an annual cycle. Thus it is conceivable that at the end of the pre period, the cycle will be at it's lowest point, but at the subsequent start of the post period, it will be at the top (or vice-versa). Particular care will need to be taken to evaluate whether any change in level or trend at this point can be explained by the presence of an annual cycle. The authors acknowledge the potential for seasonality in their discussion of autocorrelation (see below). However, I would like to see a specific sensitivity analysis designed to assess robustness against the threat of an annual cycle, regardless of the outcome of any tests for autocorrelation, given the use of a 6-month lag.

Tests for autocorrelation, using the Durbin-Watson, are planned, using a lag of up to 12 time-points. However, these tests are likely to have very low power, given the numbers of data-points and the measurement error around the individual values (which at times will be very wide). To interpret a non-significant test as implying an absence of autocorrelation would be highly questionable. The data series will almost inevitably in reality possess autocorrelation, even if undetected by the DW, and in my view it would be better to conduct analysis under the assumption that autocorrelation is present. As I have suggested above, a sensitivity test against an annual cycle should be conducted regardless.

Response: Thank you for your advice on this. The paper has been amended and the analyses will assume autocorrelation is present and a sensitivity analysis assuming seasonality has also been incorporated.

Competing Interests: None

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research