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: awaiting peer review]

Ashley B. Scrimshire, Caroline Fairhurst, Catriona McDaid, David Torgerson

1Department of Health Sciences, University of York, York, UK
2Northumbria Healthcare NHS Trust, Wansbeck, UK

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.

Keywords
Regression discontinuity, interrupted time series, quasi-experimental, anaemia, orthopaedics
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) surgery. 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 mean. This frequently leads to over-estimation of a treatment effect and has been described as the weakest observational study method.

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. 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 results. 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 programmes. 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 analyses. 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 elsewhere.

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 Trust.

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. 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 publication. 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).

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. 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 comparisons. This will be performed...
Table 1. Summary of interventions introduced in Northumbria Healthcare NHS Foundation Trust aimed at improving arthroplasty patient outcomes.

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Intervention</th>
<th>Control cohort</th>
<th>Intervention cohort</th>
<th>Statistically significant outcomes reported by authors (p&lt;0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morrison, 2017</td>
<td>Increased dose of IV Tranexamic acid (30mg/kg max 2.5g)</td>
<td>May 2008 – July 2011 (n=2637)</td>
<td>Feb 2012** – Jan 2013 (n=1814)</td>
<td>Reduced transfusions</td>
</tr>
</tbody>
</table>

* 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

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 period. 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 policy. 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
Analysis. ITS requires sequential measures of the outcome, at regular intervals, before and after the intervention time points. 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 performed. 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. 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 studies. 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. 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 locally but a Cochrane review shows this does not affect surgical transfusion rates, so will not be considered any further. 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. 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 variable. For example, giving iron (the treatment) with the intention of reducing RBC transfusion and LoS (the outcome) 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 random.

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 RCT. 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. 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).

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. Notably the treatment thresholds are...
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 unclear.

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 adhered 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 RCT. 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 appropriate. 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 introduced.
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 findigs to values that lie far from the threshold. In the case of ITS the results can be impacted by several factors such as autoccorrelation 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

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