Protocol: Emergency Surgery Or noT (the ESORT study)

1. Summary of Research

1.1. Summary and objectives of the study

The aim of this 24-month study is to evaluate the effectiveness and cost-effectiveness of emergency surgery for common acute conditions such as appendicitis, that present as emergency admissions to NHS hospitals. The 'intervention' strategy is emergency surgery within the index hospital episode, and the 'comparator' strategy is non-operative care including: 'medical management', a 'non-surgical' procedure, and the possibility of subsequent planned (elective) surgery.

The specific objectives are to evaluate the:

1. effectiveness of emergency surgery versus non-operative care for common acute conditions presenting as emergency admissions across broad ICD-10 categories.
2. relative cost-effectiveness of emergency surgery versus non-operative care across broad ICD-10 categories.
3. clinical and cost-effectiveness of operative versus non-operative care for specific patient subgroups, including diagnostic subcategories and patient characteristics.

1.2. Overview of methods

The study will apply an instrumental variable (IV) design that can fully address confounding and provide accurate estimates of treatment effectiveness even when there are unmeasured differences between the comparison groups, for example in patient case-mix. This study will extend the IV design developed by a co-applicant (Keele) that used surgeon's tendency to operate (TTO) to evaluate emergency surgery versus non-operative strategies in the United States (US)[1]. We will apply this IV design to Hospital Episode Statistics (HES) database extracts from 2009-2020. Our preparatory work using an extract of HES data (2014-5), found that the essential assumptions behind the IV design were realistic, in that the TTO, predicted the receipt of emergency surgery, and was unlikely to directly affect the outcomes.

We have chosen to include five acute conditions with well-defined intervention and comparator strategies where there is clinical uncertainty, and wide variation in the receipt of emergency surgery across the NHS. Focusing on these conditions will help ensure this research can inform future clinical guidelines and service provision. We will define final inclusion criteria for each condition as part of the proposed research, and in consultation with a multidisciplinary clinical panel.

The main outcomes are: mortality at 30 days and one year after the index emergency admission, hospital re-admission, and days alive and out of hospital up to 30 days after the index admission. The final choice of outcomes will be informed by a PPI 'study design' workshop.

The cost-effectiveness analyses (CEA) will use resource use data and mortality data from HES and the Office for National Statistics (ONS), and health-related quality of life (HRQoL) estimates from the literature. We will report the relative effectiveness and cost-effectiveness of emergency surgery for each acute condition. We will apply a local IV approach[2], to report results for pre-specified subgroups of policy relevance including patient demographics, comorbidities and route of admission. The study will provide future commissioners of acute services and of research, with evidence on those patient subgroups for which emergency surgery is relatively cost-effective, those for whom non-operative strategies are more worthwhile and those where further research is warranted.

This project is funded by the NIHR Health Services Research and Development Programme.

2. Project Overview

2.1. Background and rationale

For patients with some acute conditions who present as emergency hospital admissions there is clinical uncertainty as to whether or not they should have emergency surgery. Within the initial emergency hospital admission some patients have emergency surgery (operative care), whereas other patients with similar diagnoses only receive non-operative care which can include medical management or non-surgical procedures (for example, interventional radiology), with some of these patients having surgery deferred to the elective (planned) setting. Complications and readmission rates are higher for emergency than for elective surgery[3]. Within the emergency general surgery (EGS) specialty, some patients with acute conditions have improved health following emergency surgery and others from non-operative care. However, for many patients the relative benefits, risks and costs of emergency surgery versus non-operative care are unknown. Research is required to evaluate the relative effectiveness and cost-effectiveness of emergency surgery versus non-operative strategies for patients with common acute conditions presenting as emergency hospital admissions to NHS hospitals in England.

In England, although there are approximately 4,000 NHS consultant general surgeons who spend on average 50% of their time on EGS, there is insufficient capacity to provide emergency surgery 24/7[4]. In 2018, there were 697,314 EGS admissions to NHS trusts in England, of which 305,507 (43.8%) did not receive an operative procedure. The Getting it Right First Time (GIRFT) report for EGS, found wide variation across NHS trusts in care quality and outcomes after emergency surgery[4], which may reflect local logistical and resource constraints, but also clinical uncertainty[5]. For common acute conditions, such as diverticulitis, there are well-developed non-operative strategies and little evidence that emergency surgery leads to better outcomes. Research in the US has reported higher 30-day mortality following emergency surgery versus non-operative care for diverticulitis and some other acute conditions[1].

2.1.1. Brief literature review

There are few RCTs of emergency surgery versus non-operative strategies, amid ethical concerns about randomisation [6-11]. RCTs contrasting emergency surgery and non-operative care for patients with acute conditions have included selective patient samples, with insufficient sample sizes and follow-up duration[6-11]. These RCTs and ensuing meta-analyses have provided equivocal results for some acute conditions such as appendicitis; some have concluded that emergency surgery reduces mortality, some that it increases mortality, and some that outcomes are similar to non-operative strategies[12-22]. An ongoing RCT is comparing operative and non-operative strategies for patients with appendicitis presenting as emergency admissions to US hospitals[23], but the patients included may differ from those presenting in routine practice in the UK.

In observational studies comparing operative to non-operative care, the major concern is confounding by indication; patients who receive operative care may be sicker. Traditional risk adjustment methods are unable to fully allow for prognostic differences between the patients receiving alternative strategies because some case-mix measures, such as patient frailty, are unmeasured. As these unmeasured variables may predict both treatment receipt and outcome, these studies are liable to provide biased estimates of the effectiveness of operative care[24]. An IV design can provide accurate estimates of treatment effectiveness even when there are unmeasured differences between the comparison groups[25,26]. A recent study by Keele et al (2018) developed an IV design to address confounding when evaluating emergency surgery in the US. This study found that for some conditions, including diverticulitis, emergency surgery led to higher 30-day mortality than non-operative care[1]. However, these results may not apply to the NHS, where thresholds for operative care may be different. Also, this study did not investigate the relative cost-effectiveness of emergency surgery, nor was the study designed to report the results by clinical subgroup - evidence on both these aspects are required to inform practice policies in the NHS.

2.1.2 Why this research is needed now

The President of the Royal College of Surgeons (RCS) and the Secretary of State for Health and Social Care have endorsed the GIRFT recommendations for reducing variations in the use of emergency surgery. Recent clinical guidelines and reports highlight the lack of evidence evaluating emergency surgery[4,5,27]. The National Emergency Laparotomy Audit (NELA) recommended that the quickest improvements to the quality of EGS could be made by systematic use of protocols and care pathways, including appropriate prioritisation for emergency surgery[27]. Patient representatives, surgeons, provider managers and commissioners agree that evidence on which patients benefit from emergency surgery is urgently required.

2.1.3 Building on existing work

This study will extend the IV design developed by a co-applicant (Keele) to evaluate emergency surgery versus non-operative strategies in the US[1]. This IV design follows research in pharmaco-epidemiology that used clinician preference as an instrument for treatment receipt[26]. This study design assumes that the surgeon's preference for emergency surgery predicts how likely the patient is to receive surgery, but does not have an independent effect on the outcome (mortality). The Keele et al study undertook an extensive assessment of such assumptions in the context of US administrative data. This study found that these assumptions were plausible in this US setting where the data have some similar features to the HES data proposed for use in this application.

Our study will estimate the effectiveness and cost-effectiveness of emergency surgery for acute conditions presenting as emergency admissions to NHS hospitals in England. Our study will access large-scale routine data on emergency hospital admissions to NHS trusts in England from the HES database. In extending the IV design to the NHS in England, we will draw on our previous experience of using HES data for evaluating the cost-effectiveness of alternative forms of elective surgery[28], and for auditing EGS provision[27]. We will carefully assess the plausibility of the underlying IV assumptions for each of the acute conditions concerned, and present results according to policy-relevant subgroups.

2.2. Research plan/methods

2.2.1. Methods

Overview

The study will estimate the effectiveness and cost-effectiveness of emergency surgery versus non-operative strategies for common acute conditions that present as emergency admissions to NHS hospitals in England. We will define the target population from emergency admissions within a single data source (HES) to ensure consistent definitions, for example of inclusion criteria, across the patient cohort. The target population will include five acute conditions that present as emergency hospital admissions: appendicitis, cholecystitis, diverticulitis, acute symptomatic hernia, intestinal obstruction. The 'intervention' strategy is defined as emergency surgery within the index hospital episode, and the 'comparator' strategy as non-operative care or 'watchful waiting' within that episode which could include: 'medical management', a 'non-surgical' procedure, and then possibly subsequent planned (elective) surgery.

For each acute condition, the specific definitions of the study population, intervention and comparator will be refined at an early stage of the planned research. The initial definitions of the populations (P), interventions (I) and comparators (C) will be developed with the full HES dataset (phase 1). These initial definitions will be presented to a clinical panel, and then refined through structured discussion. The refined PIC definitions will be applied to the HES dataset, presented to a virtual clinical panel, and finalised (phase 2).

The study will access linked data for emergency hospital admissions from HES and the ONS (data request submitted April 2018; approved 20 September 2018; revised request submitted May 2020; expected data receipt, Sept 2020. We will collate information on diagnosis (ICD-10), case mix, surgical procedures received, resource use and outcomes, including mortality.

Instrumental variable design

The critical challenge for any non-randomised study is that there are liable to be potential confounders that are not observed. While we could apply traditional risk adjustment approaches, the major problem is that HES does not contain detailed measures of patient physiology, or prognostic measures such as frailty that are likely to influence the choice of operative versus non-operative care, and outcome. Hence, risk adjustment methods would be likely to provide biased estimates of the effectiveness of operative versus non-operative care. We therefore propose an IV study design[1,25,26]. A valid IV design can provide accurate estimates of treatment effectiveness even when there are unmeasured differences between the comparison groups[25,26]. An IV encourages receipt of the treatment, in this case emergency surgery, but does not have a direct effect on the outcome, for example 30-day mortality, except through treatment receipt .

Our study will use an IV developed and validated by Keele et al (2018) for evaluating emergency surgery in the US. The precedent study covered 50 acute conditions including the five that we will consider in the proposed study. The IV for the receipt of surgery is the TTO defined as the propensity of an individual (e.g. surgeon) or unit (e.g. surgical team, hospital or trust) to use operative versus non-operative management[1]. Keele et al (2018) found that TTO was a valid IV in that it varied widely across surgeons and was not correlated with surgical skill or experience[1]. We will use TTO as an IV to estimate the relative effectiveness and cost-effectiveness of emergency surgery versus comparator strategies in the NHS while minimising bias from unobserved confounding.

2.2.2. Objectives

Objective 1

We will identify cohorts of emergency admissions to NHS hospitals between 1 April 2009 and 30 June 2020 (subject to data availability). Our preparatory research has considered five of the 50 diagnostic categories from the precedent study[1]. These five acute conditions were chosen because: they are common reasons for emergency admissions in the NHS with clearly defined non-operative strategies, and there is clinical uncertainty as to whether emergency surgery is effective and cost-effective compared to alternative strategies (see Table 1).

The final definitions of inclusion criteria for each population, interventions, and comparators will be informed by a clinical panel that will be convened during the initial phase of the study (months 0-6). The clinical panel will have three purposes: first to finalise the choice of target population according to ICD-10 diagnostic codes; second to define the list of ‘intervention’ and ‘comparator’ strategies from Office of Population Censuses and Surveys (OPCS) procedure codes; and third to check the plausibility of the main IV assumptions (see IV section below). As well as obtaining the most plausible definitions and assumptions for the main analysis, we will also obtain alternative but reasonable definitions to inform accompanying sensitivity analyses (see analysis section below).

Members of the clinical panel are required to have: involvement in the clinical management or commissioning of acute services for patients with the conditions listed above, and an understanding of the common operative and non-operative alternatives for these patients. The members of the clinical panel will be drawn from clinicians working within EGS including surgeons and nurses, but also other specialities including geriatric medicine, gastroenterology, radiology, acute and emergency medicine. The panel will participate in a face-to-face workshop (Design Workshop 1) to finalise the key definitions and assumptions that will be used in the study. We will also convene a Patient and Public design workshop (Design Workshop 2) that will complement the clinical panel by helping contextualise the choice of populations, interventions and comparators, and advise on the

selection of outcomes from HES and literature-based HRQoL estimates (see also objective 2, and PPI section for full details).

Preparatory work to date has utilised emergency admission data for 2014-15 for a sample of half the acute NHS trusts in England. We used ICD-10 diagnosis codes to identify cohorts of emergency admissions, and an initial list of OPCS procedure codes to define relevant surgical procedures for these acute conditions. Table 1 lists the broad acute conditions, intervention and comparator strategies. The TTO estimates from the pilot data highlight that for each of these five acute conditions, there is wide variation in the proportion of patients who have emergency surgery across NHS trust hospitals.

Population

In the proposed study, the populations of interest will be defined by adult patients with an emergency admission to an NHS hospital in England between 1 April 2009 and 30 June 2020 for which one of the five eligible acute conditions is the primary reason for admission. Eligibility will be based on a final list of ICD-10 codes that specify that the acute condition is the 'main' diagnosis, and according to the findings of the first design workshop. For each patient, this admission then forms their *index admission*.

The exemplar, provisional 3-character ICD-10 codes for defining populations comprise: acute and unspecified appendicitis (K35, K37); cholelithiasis (K80); inguinal hernia (K40); diverticular disease of intestine (K57); vascular disorders of the intestine (K55). The final target populations for each condition will be defined with the expert clinical panel, taking place on 29 March 2020, which will determine the specific complications for inclusion related to each pathology.

Our initial analyses will also investigate pathways to diagnosis and to surgical or non-surgical care. This will help the research team and clinical panel refine the criteria for inclusion and exclusion, and for intervention or comparator. We will also examine inconsistencies between diagnoses and procedures and ensure the final definition of acute conditions ensure that for the cohort included they form the main reason for admission. This will involve: examination of diagnosis codes (specific codes and position in all episodes within an emergency admission); source of emergency admission (A&E, GP, inter-hospital transfer, etc); other potential diagnoses, e.g., other abdominal pain (R10); and multiple admissions for the same patient.

Exclusion criteria for defining the *index admission* will include: age less than 18, a previous emergency admission for the condition within the year prior to the index admission, referrals from tertiary referral centres (see IV assumptions below), and any additional exclusion criteria deemed appropriate by the clinical panel, for example minimum surgical volume thresholds for a particular hospital.

Table 1: Exemplar acute conditions, interventions and comparator strategies, and trust-level tendency to operate (TTO), from pilot HES data (2014-5)

Acute condition[1]	Population[2] (estimated cohort 09-18)	Intervention[3]	Comparator[3]	NHS Trust TTO[4] median (IQR)	F statistic[5]
Appendicitis	333,000	Appendectomy	Supportive care; Antibiotic treatment	89.3% (84.9%, 92.8%)	114
Cholecystitis	422,000	Cholecystectomy	Delayed cholecystectomy	17.3% (11.2%, 25.0%)	65
Diverticulitis	224,000	Resection; Colectomy; Laparoscopic lavage	Antibiotic therapy; intravenous fluid therapy Percutaneous	9.0% (6.6%, 10.9%)	678

			drainage of abscess		
Acute symptomatic hernia[6]	70,000	Repair of hernia	Watchful waiting; Delayed repair of hernia	46.7% (39.4%, 58.7%)	148
Intestinal obstruction (small or large bowel)	235,000	Adhesiolysis; Colonic stenting; Bowel resection	Watchful waiting with intravenous fluid	29.9% (25.5%, 34.3%)	931

[1]Acute condition refers to pathology and ensuing complications. The clinical panel will define specific complications related to pathology for the final inclusion criteria [2]Population defined by ICD-10 codes. [3]Intervention and elective surgery comparators are defined by OPCS codes. [4]Percentage of the eligible population with an intervention within index HES episode. [5]From Cragg-Donald weak identification test. [6]Includes inguinal hernia, femoral and incisional.

Intervention

For each acute condition, surgical intervention is defined as an operative procedure corresponding to one of a pre-determined set of OPCS Classification of Interventions and Procedures codes within the index emergency admission. Table 1 includes examples of typical procedures or treatments for each acute condition. Code lists will be refined during the initial phase of the research (see later section for further details). Final criteria for defining a relevant surgical intervention will recognise the timing of surgery with respect to hospital admission, and its relationship with other surgical and non-surgical procedures within the admission. These final criteria will stipulate what delay to surgery is required for it to meet the criteria for a 'comparator' rather than 'an intervention strategy'. Alternative specifications will be considered in sensitivity analyses.

Our preparatory work has identified an extensive list of provisional OPCS procedure codes that indicates the patient has had a surgical intervention. Exemplar procedures include: H01.2 Emergency excision of abnormal appendix (appendicitis); J18.3 Total cholecystectomy (cholecystitis); T20.2 Primary repair of inguinal hernia using insert of prosthetic material (inguinal hernia); H33.5 Rectosigmoidectomy and closure of rectal stump and exteriorisation of bowel (diverticular disease); T41.3 Freeing of adhesions of peritoneum (intestinal obstruction).

Comparator strategy

For each acute condition, the comparator is defined as an eligible admission whose care does not include an operative procedure considered to constitute the intervention. This includes patients receiving non-surgical treatment or diagnostic evaluation as identified by OPCS codes, and those with no procedure recorded during their index admission. The comparator strategies include delayed surgery, defined as having a relevant surgical procedure but after the specified time-frame for the intervention.

Case mix and potential confounders

Information from HES on patient characteristics will be available at the patient-level and will comprise socio-demographic characteristics (age, sex, ethnicity and decile group of the index of multiple deprivation). Comorbidities will be derived using HES records from all admissions for each patient in the year prior to and including the index admission using the RCS Charlson score[29]. The definition of co-morbidities will use both information on past medical history according to chronic conditions, but also according to reasons for previous admissions. Potential higher-level confounders will include surgical volume for relevant emergency and elective procedures for each acute condition. These surgical volumes will be calculated for each financial year from 2008-09 to partial 2020-21 using HES records for all relevant patients. We will report these measures of volume at the levels of the consultant surgeon, hospital (including surgical team), and NHS trust.

Outcomes

The main outcomes will be mortality at 30 days, 90 days and one year after the index emergency admission, readmissions, and days alive and out of hospital prior to 30 days. For each index emergency admission, we will access information on mortality, re-interventions and readmissions (including emergency and elective surgery) from HES and ONS data. Further surgery within 30 days and one year will be defined according to OPCS intervention codes. Readmission will be defined as any hospital admission within one year of the index admission, subdivided according to emergency versus elective, including for planned (re)intervention. The definition of mortality at 30 days, 90 days and one year following the index admission will be according to indicators provided by NHS Digital and will use ONS date of death. We will also have access to date of death through ONS linkage, and will calculate the number of days alive and out of hospital before day 30. We will convene a Patient and Public workshop (Design Workshop 2) to discuss the study outcomes (see PPI section for full details).

Missing and miscoded data

The main potential issues are missing values for case mix variables and failure to record diagnoses, procedures or outcomes. Missing or unreported ethnicity data in eligible admissions (c10%), and deprivation (c1%) will be minimised by using ethnicity and deprivation data from patients' other linked episodes. HES data for age and sex are almost complete. We will recognise the potential impact of improvements in the coding of diagnoses in HES between 2008-2020 by including a covariate for financial year in the analysis models. We will examine the use of diagnosis/procedure codes over time periods to identify other potential issues that need to be considered in defining PICO criteria, for example expanding eligibility to include secondary as well as primary diagnoses. For any missing covariate data, we will make plain any assumptions that we do make, drawing on previous work from ourselves and our collaborators. For example, for baseline covariates with missing values we will examine whether this pattern of missingness differed across the intervention and comparator groups. We will use the approach recommended by Rosenbaum (2010) to making minimal, but transparent assumptions about the missing data while balancing patterns of missingness across intervention and control groups[30].

Instrumental variable

Our proposed IV design follows research in pharmaco-epidemiology[26] that uses provider preference as an instrument for treatment receipt. In this study, the IV will be the TTO, defined for each acute condition, as the proportion of emergency admissions during which an eligible surgical procedure is undertaken. Hospitals and surgeons will be identified from the HES provider/site of treatment codes, and the consultant/specialty codes as in precedent research by the applicants[1,30]. In the primary analysis, TTO will be defined at the hospital-level, to recognise that in the NHS, multidisciplinary team input informs the decision to operate. We will also report sensitivity analyses with TTO defined at the trust-level, as in the pilot data, and at the consultant-level, for comparison with the precedent study[1]. We will calculate TTO by financial year (2008-09 to partial 2020-21) using HES records for all relevant patients.

Sample size

We have used a HES data extract which is for 50% of NHS acute hospital trusts for a single financial year (2014-15). We find that this pilot data would provide at least 99% power to detect an absolute risk difference of 10% for all conditions except diverticulitis (57% power). The proposed study will have access to HES data for all NHS trusts for five years, and so we envisage that for each of the five acute conditions, the study will have at least 90% power to detect differences as small as 5% at the 5% level of statistical significance.

Analysis

Checking IV assumptions

For a variable to be a valid instrument for treatment receipt it has to i) predict the receipt of treatment; ii) be independent of baseline covariates; and iii) only affect the outcome indirectly through the treatment[25]. It is critical to carefully assess these assumptions in an IV analysis. Keele et al. completed a full assessment as to whether TTO met the essential conditions for being an IV in the US context[1]. We will carefully assess whether TTO meets the criteria for an IV for each of the five acute conditions described for patients presenting as emergency admissions to NHS hospitals in England.

In the HES pilot data, we found that the trust-level TTO was strongly associated with receipt of emergency surgery; for each condition the F-statistics exceeded the critical value of 10, and therefore assumption i) was satisfied (see Table 1) [31]. We will repeat these tests of instrument strength for the full HES data for each level of the TTO. For assumption ii) we found that the trust-level TTO balanced the observed covariates, and we will explore this further in the full HES data[32,33]. Although we cannot assess empirically whether TTO is independent of unmeasured confounders, in the emergency setting this assumption is likely to be justified for hospitals, since the vast majority of patients will attend their local hospital without consideration of the hospital-level TTO. This assumption will be bolstered by the exclusion of patients referred to tertiary referral centres for whom this assumption is less plausible. These patients can be identified from HES data on admission source, and distance to provider. For assumption iii) it seems unlikely that after adjusting for the volume of EGS and the outcomes of previous admissions, that the TTO would have a direct effect on the outcomes. For example, it is unlikely that just because a surgical team prefers to operate, that the patients' outcome would be better (or worse) unless that patient actually underwent an operation. We will use falsification tests to probe the plausibility of assumptions (ii) and (iii). Specifically, we will identify subgroups where all patients receive emergency surgery, and test whether the TTO has an effect on outcomes in these subgroups[34,35].

Most IV applications also assume that there are 'no defiers' – patients treated contrary to the physicians' preference[33,36]. Here, this implies that there are no patients who receive surgery simply because they present in a high TTO hospital and vice versa. Following the approach of Swanson et al. (2015) [37] for checking the assumption of 'no defiers', we will survey the clinical panel about their preferences and choice of strategy for a set of hypothetical patients with the above acute conditions using vignette patient profiles. This will allow us to assess the magnitude and direction of any bias due to the presence of no defiers.

Estimation and sensitivity analyses

We will use the generalized effect ratio[38], that provides the same point estimates as two-stage least squares, but also correct confidence intervals[39]. For survival outcomes, we will apply IV methods for Cox survival models[40,41].

We will undertake three extensive sets of sensitivity analyses. First, we will draw from the views of the clinical panel to test whether the conclusions are robust to alternative but plausible selection criteria for the subpopulations, interventions and comparators of interest (see earlier section). Second, we will assess the robustness of the major IV assumptions by reporting nonparametric bounds for the IV estimate[42], and investigating the extent to which results are robust to the association of an unobserved confounder with both the IV and the outcome[25,43]. We will investigate whether our results are robust to alternative statistical models by estimating results based on logistic regression models using the plug-in principle for estimation[44].

Output

We will report the relative effectiveness of emergency surgery versus comparator strategies for each of the five acute conditions, for patients presenting as emergency hospital admissions to NHS hospitals in England.

Objective 2

Design

The CEA will take a hospital and personal social services perspective (PSS)[45]. In the base case, the study will report results over a one-year time horizon and incorporate patient-level resource use (HES) and mortality data (ONS), and estimates from the literature for HRQoL[46], community care costs[46], and unit costs[47,48].

Resource use and unit costs

The study will measure resource use items that are the major drivers of relative cost and cost-effectiveness for patients with relevant acute conditions[49,50]. These resource use categories are: the receipt of operative and other procedures, the duration of hospital stay for the index admission, subsequent readmissions or re-interventions, and transfer to continuing care, in particular, care homes.

The study will use patient-level resource use data from the HES extract for each eligible index emergency admission. We will collate data on the procedures received according to OPCS procedure codes, and the overall length of stay (LoS) in hospital. We will access data on the number of critical care bed-days, with the level of care defined by the number of organs supported, according to the Adult Critical Care data linked to HES[51]. We will extract HES data on subsequent admissions to recognise that patients may transfer to another hospital for example for rehabilitation. We will use information on discharge destination, including transfer home or to a care home.

For the eligible patients, we will use our HES extract to identify all other admissions to hospitals in England up to one year after the index admission. We will include all these hospital readmissions, including those where the OPCS code indicates that the patient received an operative procedure either as delayed surgery (comparator group), or as a re-intervention (intervention or comparator group).

The unit costs of critical care bed-days according to number of organs supported, and of days in general wards will be taken from the NHS payment by results database[47], and the unit costs of care home days from the PSSRU unit cost database[48]. For operative and non-operative procedures, unit costs that are directly relevant to the UK will be taken from our literature review[52-54]. Resource use measures will be combined with unit costs to report total costs per patient up to one year.

Outcomes

We will use OPCS data on each patients' vital status and date of death within one-year of the index admission to calculate the number of life years. We will undertake a full literature review of HRQoL for each of the five acute conditions. The review will select those studies, which meet the following criteria: they have measured HRQoL following emergency admission for patients with the diagnoses listed in Table 1, with a recommended, generic HRQoL instrument for patients in the United Kingdom, or a country with similar health state preferences.

We undertook a full literature review of HRQoL following emergency admissions for each of the five conditions. The main finding was that for some but not all of these conditions appropriate HRQoL were available from the literature. For example, for patients that present with an emergency appendicitis, appropriate measure of HRQoL (EuroQoL, EQ-5D) for the year after the index emergency admission will be available from the CODA trial[23]. For, other conditions our initial review suggests that appropriate HRQoL may not be available directly from the literature.

We therefore also propose a complimentary appropriate to estimating HRQoL for those conditions without appropriate literature-based utility values. We will use a published 'EQ-5D 'utility calculator' that can provide appropriate EQ-5D (HRQoL) scores for patients in the UK according to the

diagnostic categories (ICD10), age, and gender of the individual patients included in our study from the HES dataset[55]. While this approach has been extensively used in previous cost-utilities studies, we will carefully check the plausibility of the HRQoL estimates for the conditions in our study. This 'sense check' will take two forms: first, for those conditions such as cholelithiasis where there are already published HRQoL values we have cross-compared the values across the two alternative sources, and found that they are similar[52,56]. Second, we will present the HRQoL from the EQ-5D calculator to the clinical and PPI panels, as a further check of the 'face validity' of the estimates.

Finally, QALYs will be calculated using the 'area under the curve' approach by combining the most appropriate HRQoL utility score (base case) with the data available from HES/ONS on each individual patient's survival. To examine the robustness of the findings to the choice of HRQoL, we will undertake sensitivity analysis by using alternative but also plausible estimates of HRQoL from the literature, or taking the approach adopted by Meacock et al[57] for estimating QALY for health service evaluations that use HES data when no relevant HRQoL data are available from the literature.

(see separate PPI section).

We will then report quality-adjusted life years (QALYs) up to one year by combining survival time with QoL estimates using the 'area under the curve' approach[58].

Analysis

We will estimate the relative costs and cost-effectiveness of emergency surgery (operative) versus comparator strategies for each acute condition according to broad ICD-10 diagnostic categories. We will report incremental costs per life year and per QALY gained. The main cost-effectiveness metric will be the incremental net monetary benefit, which will be calculated by valuing QALYs by the NICE recommended thresholds of £20,000 and £30,000 per QALY, and then subtracting the incremental cost[45]. In the base case analysis, we will use 3 stage least-squares (3SLS) regression approaches that we have developed for IV designs that allow for the correlation between the costs and QALYs while estimating cost-effectiveness[59].

We will undertake extensive sensitivity analyses to test whether conclusions are robust to key assumptions. We will consider the impact of taking a hospital versus PSS perspective, to using different sources for the HRQoL and unit cost data, to alternative assumptions about the distribution of the cost data (e.g. gamma or log normal distribution) and to taking a longer time horizon (lifetime vs one year). To estimate lifetime cost-effectiveness, we will access use all HES-ONS linked mortality and resource use data subsequent to the index admission (up to March 2018), which will provide up to 10 years follow-up data. The long-term modelling will extrapolate from these observed data by fitting alternative parametric survival curves (e.g. Weibull, exponential, lognormal, log logistic and Gompertz). We will choose the extrapolation approach that gives the most plausible predictions of long-term survival[60].

Outputs

We will report the relative cost-effectiveness of emergency surgery versus comparator strategies for these five common acute conditions.

Objective 3

Design

We will estimate the relative effectiveness and cost-effectiveness of emergency surgery, according to patient subgroup. We will recognise that effectiveness and cost-effectiveness may differ according to observed and unobserved patient- and service-level characteristics. We will predict counterfactual outcomes following both emergency surgery (operative care) and the comparator for each patient using the local-IV (LIV) approach developed by Basu[2,61,62]. We will generate individual-level

estimates of effectiveness and cost-effectiveness. We will aggregate these individual-level estimates to report relative effectiveness and cost-effectiveness according to relevant subgroups defined by risk factors such as refined diagnostic group, age, ethnicity, index of multiple deprivation, co-morbidities, past medical history, admission route. We will report results according to each risk factor alone and then in combination.

The LIV approach will estimate person-centered average treatment (PeT) effects allowing for confounding, but will also recognise that the effects of emergency surgery are likely to be heterogeneous according to the patients' characteristics[2,61-63]. We will identify "marginal" patients for whom the physician is in equipoise about the decision for surgery according to both those prognostic characteristics that are measured in HES (e.g. age, comorbidity, diagnosis), and those that are unmeasured (e.g. illness severity). For these "marginal patients" a small change (or nudge) in the TTO can determine whether they receive surgery, but will not change the distribution of their risk factors. That is, at the time of emergency admission, Patients A and B may be at similar overall risk of death, but Patient A may be admitted to a hospital with a slightly higher TTO, and is more likely to have an operative procedure than Patient B.

By comparing outcomes for two groups of patients defined according to small differences in the TTO (the IV), but with similar risk profiles, we will therefore provide estimates of the causal effect of emergency surgery for marginal patients. By repeating this contrast across different levels of the TTO, we will estimate the required treatment effects for sets of marginal patients with different characteristics. The person-level effects of emergency surgery will then be estimated by averaging the effects for those marginal patients who share the same observed characteristics. The LIV approach will exploit information about the choice of surgery for each individual according to their observed characteristics, and this relationship between the choice and the observed risk factors will be informative about the level of each patients unobserved characteristics.

Analysis

First, we will estimate each patient's propensity for emergency surgery according to their observed characteristics and the hospital TTO using a probit model. Second, we will estimate the relationship of the observed patient characteristics and their propensity for emergency surgery with each outcome. The effects of emergency surgery on binary outcomes will be estimated using probit models, count outcomes with Poisson models, and continuous outcomes with generalised linear models. From these regression models we will estimate the effect of a change in the propensity for emergency surgery on each outcome, that is the causal effects of surgery for marginal patients providing PeT effects estimates for each patient.

The resultant PeT effects can be interpreted for each patient as the difference in their predicted outcomes with versus without emergency surgery. These person-level treatment effects will be aggregated to report the effectiveness and cost-effectiveness of emergency surgery for each of the five conditions, and for each pre-specified subgroup of interest, for example to age, gender, ethnicity, number of co-morbidities and route of hospital admission (accident and emergency, GP or elsewhere in hospital). The effectiveness of emergency surgery versus the comparator will be reported as mean (95% CI) differences in mortality, readmissions, costs, QALYs and incremental net monetary benefits.

Finally, we will develop predictive models (logistic regression) to examine which risk factors alone or in combination, predict which subgroups emergency surgery is relatively effective or cost-effective for based on the patients' PeT effects. We will define the clinical benefit of interest as a 10% difference in one-year mortality, and the metric of cost-effectiveness as a positive incremental net benefit. All standard errors will be calculated with non-parametric bootstrapping, and will account for clustering of individuals within hospitals.

To assess whether findings are robust to the choice of LIV model, we will consider alternative statistical models including: different functional forms; the inclusion versus exclusion of higher order

terms for continuous baseline measures such as age, and interaction terms between all the covariates.

Outputs

We will report the relative effectiveness and cost-effectiveness according to policy-relevant patient subgroups including, age, gender, past medical history and route of hospital admission. We will identify subgroups where there may be cost savings from increased uptake of either operative or non-operative strategies. We will itemise those resource items where reduction in use (e.g. consumables required for operative procedures) may lead to short-term cost savings.

3. Project Outputs

We will report the relative effectiveness and cost-effectiveness according to policy-relevant patient subgroups including, age, gender, past medical history and route of hospital admission. We will identify subgroups where there may be cost savings from increased uptake of either operative or non-operative strategies. We will itemise those resource items where reduction in use (e.g. consumables required for operative procedures) may lead to short-term cost savings.

4. Dissemination, projected outputs, knowledge mobilisation

This work is driven by a significant gap in our understanding of the risks, benefits and costs of emergency surgery and it is important that our recommendations for clinical practice are fed into NHS guidelines and can be used to shape practice. The study has been designed to maximise research impact. The specific research questions directly tackle those priorities that have been designated by authoritative sources such as the GIRFT report, clinical opinion leaders, service commissioners, leaders of surgical networks and PPI representatives. These views have shaped the objectives and the choice of conditions, operative and non-operative strategies. Direct communication of knowledge to key clinical organisations, and where appropriate, input into clinical guideline development, will be ensured by team members (Royal Colleges and Association of Surgeons – Cromwell, Hinchliffe; NHS England – Moonesinghe). This will include contributing to future RCS initiatives[64], and working with NHS Rightcare to modify decision aids aimed at supporting shared decision making[65]. We will also draw on members of our advisory group who have roles in guideline development, and advising the Department of Health, and who will help ensure we exploit opportunities for maximising the impact of this research.

The research will provide recommendations to commissioners and providers of surgical services on those services where disinvestment is warranted, those where additional investment is required, and those where additional evidence, for example from new RCTs, would be of greatest value.

The research findings will be presented at high profile national and international clinical conferences (surgical, perioperative). While the study has used HES data from England, we will ensure that both the empirical and methodological insights have wide relevance. We will work with our international network of surgeons, and with our international advisory panel (see below), to consider carefully the direct implications of the findings to other countries. The methods will be discussed at academic meetings, in particular the Health Services Research network, and Health Economist study group meetings in the UK, the European Causal inference network, and the American Health economists study group meeting (USA).

The design and translation workshops will draw on the views of patient representatives, and will consider the translation of the findings into accessible and informative summaries appropriate to the audiences of interest, and effective modes of dissemination via special interest groups (see below for further details). The findings will be presented to select groups of professional stakeholders (for example NICE guidelines committees, NHS England), commissioners and managers of surgical services and those setting future research priorities. We will work closely with Dr Rachel Kelz and colleagues at the University of Pennsylvania to ensure that this work is disseminated throughout

international surgical networks, in particular through presentations at the American College of Surgeons Clinical Congress. This congress will provide an excellent opportunity to translate the implications of the findings, and the wider application of the methods to key surgical opinion leaders.

We will develop a study website that will be an important repository of information about the study methods and findings, for both lay and professional audiences. It will include information for parties interested in applying to join the PPI panel and a means by which to do so, and will clearly display the HES privacy notice. The website will reside within the main LSHTM website (e.g. esort.lshtm.ac.uk) and will be developed by a study team member with expertise in developing similar websites (RS), then maintained by Beth Silver (research manager).

How will you inform and engage patients, NHS and wider population about your work? (see also separate PPI subsection)

Engagement with patients and the public, as well as NHS clinicians, is a key activity in this research, as we aim to produce findings which are clinically useful, reflect stakeholder priorities, and are informed by different areas of expertise. We have embedded these aims throughout the study, including a translation workshop with PPI representatives (month 17) where an effective information and engagement strategy will be discussed and finalised. PPI representatives will be asked to contribute to the development of materials reporting our key messages so that they are informative and clear, speak to patient interest and reach different audiences. Their views will be sought on which are the most appropriate organisations and special interest groups to support dissemination of our findings and translation of results into practice. This work is driven by a significant gap in our understanding of the risks and benefits of emergency surgery and it is important that our recommendations for clinical practice are fed into NHS guidelines and can be used to shape practice. In order to contribute to further understanding of PPI it is important that our methods are reported to the research community. To do so we will draw upon the GRIPP2 reporting mechanism for patient and public involvement in health and social care research[66].

We will work with the LSHTM media department, and our lay representatives to ensure the findings are accessible to the broader public. The study website will be an important repository of information about the study methods and findings, for both lay and professional audiences. A full and complete account of the research will be made available by open access as a publication in the NIHR HS&DR Journal. Research papers will be published in peer-reviewed journals.

5. Project/research timetable

This 24-month project has had partially prepared HES data from the outset (anticipated receipt of full HES data: September 2020). The design workshop / Clinical Panel in month 6 will help finalise the target populations, outcome measures and analysis plans. We will submit the final study protocol for university ethics review (month 6) and publish analysis plans in open access journals (months 6-9). We will extract additional parameters required for the CEA from the literature (objective 2) and analyse the data to provide the requisite estimates to meet the study objectives (months 9-18). We will prepare results for conference presentations, national and translation workshops, papers for peer-review (months 18-22) and the final draft project report (months 21-24). We will hold biweekly team meetings, and advisory group meetings (months 9 and 16). The costs requested include those for an experienced research manager to co-ordinate this complex project.

Timetable

Grant Start date: October 2019

Months 0-6	Data cleaning, testing of provisional coding algorithms Extraction of relevant HES index admissions according to ICD-10 and OPCS codes Preparation of summary data for clinical panel, calculation of TTO

Months 3-6	Draft study protocol University ethics application and approval
Month 6	Interim Report to NIHR Receive and start work on new HES data Detailed literature review of QoL data
Month 7	*Clinical panel* to refine selection of sub conditions and check IV assumptions, finalisation of subconditions (Design workshop 1)
Month 9	*Advisory Group Meeting* *Design workshop with PPI (Design workshop 2)*
Months 6-9	Application of final coding algorithms to full data Submission of paper to Health Economics Study Group on IV design Final study protocols, study design publications in open access journals
Months 9-12	IV estimation of effectiveness of emergency surgery, main analysis Collation of literature review on unit costs
Month 12	*Interim report to NIHR*
Months 13-17	Undertake cost-effectiveness analysis (objective 2) Draft paper I on effectiveness of emergency surgery
Month 16	*Advisory Group Meeting (Preliminary results)*
Month 17	Translation workshop with PPI (study results)
Months 17-19	Undertake estimation of person-level treatment effectiveness (objective 3) Draft paper II on cost-effectiveness of emergency surgery
Month 20-22	Submission of paper I Draft paper III on person-level estimation of effectiveness and cost-effectiveness
Months 19-21	Finalisation and Submission of paper II Finalisation and Submission of paper III
Months 18-23	Presentations at international clinical and health economics conferences
Months 22-23	Draft final report Production of materials to special interest groups
Month 24	*Submission of final report to NIHR*

6. Research governance

The project involves the use of routinely allocated anonymous administrative data, and discussions with PPI and clinical representatives. The project will not require approval by NHS ethics committees, but will require approval by the local university ethics committee

7. Data Storage and Transfer

Patient data will be provided by NHS Digital under a Data Sharing Agreement (DSA). The full application for HES admitted patient data and critical care data and derived ONS mortality data was submitted in May 2020 with data expected to arrive in September 2020. The data custodian as specified in the DSA is LSHTM. Data will be transferred by secure protocol in agreement with NHS Digital (e.g. secure electronic file transfer (SEFT)) and will be held on LSHTM's secure server in accordance with the DSA and LSHTM's contract with NHS Digital. Access will be restricted to study personnel (Richard Grieve, Andrew Hutchings, Silvia Moler Zapata, and to be appointed) as notified to IT Services and in accordance with LSHTM's NHS Digital Data Security and Protection Toolkit mandates for accessing and using the secure server.

8. Project Management

Richard Grieve will take overall responsibility for project delivery (20% WTE); he will guide the team, ensure close collaboration between the methodological and clinical inputs, monitor progress against timelines and lead the study management group (Richard Grieve, Andrew Hutchings, Assistant

Professor (TBC)). The study management group will meet bi-weekly in person and will report to the study advisory group.

The study advisory panel will be chaired by Iain Anderson, Consultant Surgeon, Salford Royal Hospital NHS Foundation Trust, Manchester, president-elect Association of Surgeons of Great Britain and Ireland. The study advisory group will also include Dr Rachel Kelz who is PI of the precedent study in the US, Professor James Carpenter and internationally recognised biostatistician with expertise in missing data, and Professor Sir Nick Black an eminent health services research with specific interests in outcome measurement in emergency surgery. The meetings of the study advisory group will be at months 5 and 11. These timings have been chosen to ensure key strategic input, and so that the advisory group can offer an overall assessment of the project's progress to help ensure timely delivery of each project component.

Andrew Hutchings (50% WTE) will be responsible for the preparation of the HES data (all objectives), the Assistant Professor (TBC, 10% WTE) will design and undertake the main IV analysis (objectives 1-3). Luke Keele will advise on both study design and the IV analysis. The PhD student in health economics (Silvia Moler Zapata) will undertake the literature review on HRQoL and unit costs supervised by RG (objective 2). Stephen O'Neill (5% WTE) will help the Assistant Professor with the person-level IV analysis (objective 3). Professor Anirban Basu, University of Washington, who developed the person-level IV approach, has agreed to a consultancy role on the project and will advise on the implementation of the local-IV method (objective 3).

Ramani Moonesinghe, Robert Hinchliffe (each 2.5% WTE) and Geoff Bellingan (1%) will oversee the clinical input to the project, and ensure the design and interpretation focus on implications for service change. Dr. Rachel Kelz will serve as the international clinical advisor to the project.

David Cromwell (2.5%) WTE) will bring a national surgical perspective, and links to surgical networks. Claire Snowdon (7.5% WTE) will lead the PPI input including the design and translation workshops. Paul Chandler (2.5% WTE) will be actively involved throughout the project to provide oversight from the public perspective.

The project will be supported by Beth Silver, (20% WTE), who will co-ordinate activity across the three institutions, help organise the clinical and translation workshops, manage the budget, schedule, plan timely input from collaborators, update materials for the project website, and help ensure timely delivery of the peer-reviewed publications and final draft report.

8. Public and Patient Involvement

The application has benefited from the input of Paul Chandler (PC), a Patient Research Ambassador Initiative member, and Stephen Harkins (SH) who has experienced emergency surgery for a common condition. The PPI representatives felt that the design and interpretation of the quantitative study should be informed by experiences of patients presenting as emergency admissions. PC and SH therefore helped plan two workshops with public and patient representatives, surgeons, and research commissioners. The design workshop will inform the selection of acute conditions and outcome measures, and the translation workshop will focus on interpretation and communication of results. PC and SH have encouraged the team to elicit views of i) surgical patients, ii) non-operative care patients, and iii) expectations of the healthy public via the workshops. The proposed workshops have PPI support from North Thames Collaboration for Leadership in Applied Health Research (communications officer, Stephen Towndrow).

References

1. Keele, L.J., et al., *An Instrumental Variables Design for the Effect of Emergency General Surgery.* Epidemiological Methods (In Press), 2018.

2. Basu, A., *Estimating Person-Centered Treatment (Pet) Effects Using Instrumental Variables: An Application to Evaluating Prostate Cancer Treatments.* J Appl Econ (Chichester Engl), 2014. **29**(4): p. 671-691.
3. Havens, J.M., et al., *The excess morbidity and mortality of emergency general surgery.* Journal of Trauma and Acute Care Surgery, 2015. **78**(2): p. 306-311.
4. Abercrombie, J. *Getting it Right First Time (GiRFT) report (2017).* General Surgery August 2017 [cited 2018 22 Apr]; Available from: http://gettingitrightfirsttime.co.uk/national-general-surgery-report-published-2/.
5. Watson, R., et al., *Emergency general surgery: challenges and opportunities*. 2016, Nuffield Trust.
6. Siddiqui, T., et al., *Early versus delayed laparoscopic cholecystectomy for acute cholecystitis: a meta-analysis of randomized clinical trials.* The American Journal of Surgery, 2008. **195**(1): p. 40-47.
7. Collaborators, E., et al., *Methodological overview of systematic reviews to establish the evidence base for emergency general surgery.* British Journal of Surgery, 2017. **104**(5): p. 513-524.
8. Roulin, D., et al., *Early Versus Delayed Cholecystectomy for Acute Cholecystitis, Are the 72 hours Still the Rule?* Annals of surgery, 2016. **264**(5): p. 717-722.
9. Gutt, C.N., et al., *Acute cholecystitis: early versus delayed cholecystectomy, a multicenter randomized trial (ACDC study, NCT00447304).* Annals of surgery, 2013. **258**(3): p. 385-393.
10. Gargya, V., S. Smolarek, and T.N. Walsh, *Concerns About Acute Cholecystitis: Early Versus Delayed Cholecystectomy—A Multicenter Randomized Trial.* Annals of surgery, 2015. **262**(2): p. e63-e64.
11. Zogg, C.K., et al., *The Need to Consider Longer-term Outcomes of Care.* Annals of surgery, 2017. **266**(1): p. 66-75.
12. Crofts, T.J., et al., *A randomized trial of nonoperative treatment for perforated peptic ulcer.* New England Journal of Medicine, 1989. **320**(15): p. 970-973.
13. Loozen, C.S., et al., *Conservative treatment of acute cholecystitis: a systematic review and pooled analysis.* Surgical endoscopy, 2017. **31**(2): p. 504-515.
14. Tan, C., B. Dasari, and K. Gardiner, *Systematic review and meta-analysis of randomized clinical trials of self-expanding metallic stents as a bridge to surgery versus emergency surgery for malignant left-sided large bowel obstruction.* British Journal of Surgery, 2012. **99**(4): p. 469-476.
15. Harnoss, J.C., et al., *Antibiotics versus surgical therapy for uncomplicated appendicitis: systematic review and meta-analysis of controlled trials (PROSPERO 2015: CRD42015016882).* Annals of surgery, 2017. **265**(5): p. 889-900.
16. Johansson, M., et al., *Management of acute cholecystitis in the laparoscopic era: results of a prospective, randomized clinical trial.* Journal of gastrointestinal surgery, 2003. **7**(5): p. 642-645.
17. Hansson, J., et al., *Randomized clinical trial of antibiotic therapy versus appendicectomy as primary treatment of acute appendicitis in unselected patients.* British Journal of Surgery: Incorporating European Journal of Surgery and Swiss Surgery, 2009. **96**(5): p. 473-481.
18. Fitzgibbons, R.J., et al., *Watchful waiting vs repair of inguinal hernia in minimally symptomatic men: a randomized clinical trial.* Jama, 2006. **295**(3): p. 285-292.
19. Gurusamy, K., et al., *Meta-analysis of randomized controlled trials on the safety and effectiveness of early versus delayed laparoscopic cholecystectomy for acute cholecystitis.* British journal of surgery, 2010. **97**(2): p. 141-150.
20. Boerma, D., et al., *Wait-and-see policy or laparoscopic cholecystectomy after endoscopic sphincterotomy for bile-duct stones: a randomised trial.* The Lancet, 2002. **360**(9335): p. 761-765.
21. Styrud, J., et al., *Appendectomy versus antibiotic treatment in acute appendicitis. a prospective multicenter randomized controlled trial.* World journal of surgery, 2006. **30**(6): p. 1033.
22. Salminen, P., et al., *Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: the APPAC randomized clinical trial.* Jama, 2015. **313**(23): p. 2340-2348.
23. Davidson, G.H., et al., *Comparison of Outcomes of antibiotic Drugs and Appendectomy (CODA) trial: a protocol for the pragmatic randomised study of appendicitis treatment.* BMJ Open, 2017. **7**(11).
24. Cao, A.M., G.D. Eslick, and M.R. Cox, *Early laparoscopic cholecystectomy is superior to delayed acute cholecystitis: a meta-analysis of case–control studies.* Surgical endoscopy, 2016. **30**(3): p. 1172-1182.
25. Baiocchi, M., J. Cheng, and D.S. Small, *Instrumental variable methods for causal inference.* Stat Med, 2014. **33**(13): p. 2297-2340.

26. Brookhart, M.A. and S. Schneeweiss, *Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results.* The international journal of biostatistics, 2007. **3**(1).
27. National Emergency Laparotomy Audit Project team, *Third Patient Report of the National Emergency Laparotomy Audit*. 2017: London
28. Pennington, M., et al., *Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis.* Bmj, 2013. **346**: p. f1026.
29. Armitage, J.N., J.H. van der Meulen, and G. Royal College of Surgeons Co-morbidity Consensus, *Identifying co-morbidity in surgical patients using administrative data with the Royal College of Surgeons Charlson Score.* Br J Surg, 2010. **97**(5): p. 772-81.
30. Varagunam, M., A. Hutchings, and N. Black, *Relationship between patient-reported outcomes of elective surgery and hospital and consultant volume.* Med Care, 2015. **53**(4): p. 310-6.
31. Stock, J. and M. Yogo, *Testing for Weak Instruments in Linear IV Regression*, in *Identification and Inference for Econometric Models*, D. Andrews, Editor. 2005, Cambridge University Press: New York. p. 80-108.
32. Baiocchi, M., et al., *Near/far matching: a study design approach to instrumental variables.* Health Serv Outcomes Res Methodol, 2012. **12**(4): p. 237-253.
33. Swanson, S.A. and M.A. Hernán, *Commentary: how to report instrumental variable analyses (suggestions welcome).* Epidemiology, 2013. **24**(3): p. 370-374.
34. Glymour, M.M., E.J. Tchetgen Tchetgen, and J.M. Robins, *Credible Mendelian randomization studies: approaches for evaluating the instrumental variable assumptions.* American journal of epidemiology, 2012. **175**(4): p. 332-339.
35. Yang, F., et al., *Dissonant conclusions when testing the validity of an instrumental variable.* The American Statistician, 2014. **68**(4): p. 253-263.
36. Angrist, J. and J.-S. Pischke, *Mostly harmless econometrics: an empiricists guide*. 2009, Princeton: Princeton University Press.
37. Swanson, S.A., et al., *Definition and evaluation of the monotonicity condition for preference-based instruments.* Epidemiology (Cambridge, Mass.), 2015. **26**(3): p. 414.
38. Baiocchi, M., et al., *Building a stronger instrument in an observational study of perinatal care for premature infants.* J Am Stat Assoc, 2010. **105**(492): p. 1285-1296.
39. Kang, H., L. Peck, and L. Keele, *Inference for instrumental variables: a randomization inference approach.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 2016.
40. Tchetgen, E.J.T., et al., *Instrumental variable estimation in a survival context.* Epidemiology (Cambridge, Mass.), 2015. **26**(3): p. 402.
41. MacKenzie, T.A., et al., *Using instrumental variables to estimate a Cox's proportional hazards regression subject to additive confounding.* Health Services and Outcomes Research Methodology, 2014. **14**(1-2): p. 54-68.
42. Balke, A. and J. Pearl, *Bounds on treatment effects from studies with imperfect compliance.* Journal of the American Statistical Association, 1997. **92**(439): p. 1171-1176.
43. Small, D.S. and P.R. Rosenbaum, *War and wages: the strength of instrumental variables and their sensitivity to unobserved biases.* Journal of the American Statistical Association, 2008. **103**(483): p. 924-933.
44. Terza, J.V., A. Basu, and P.J. Rathouz, *Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling.* Journal of health economics, 2008. **27**(3): p. 531-543.
45. National Institute for Health and Care Excellence, *Guide to the methods of tehcnology appraisal 2013*. 2013: London
46. Ulug, P., et al., *Strategy of endovascular versus open repair for patients with clinical diagnosis of ruptured abdominal aortic aneurysm: the IMPROVE RCT.* Health technology assessment (Winchester, England), 2018. **22**(31): p. 1-122.
47. Department of Health, *Payment by Results database*. 2016, Department of Health: London.
48. Curtis, L., *Unit costs of health and social care.* 2016, Personal Social Services Research Unit, University of Kent

49. Braithwaite, B., et al., *Endovascular strategy or open repair for ruptured abdominal aortic aneurysm: one-year outcomes from the IMPROVE randomized trial.* European heart journal, 2015. **36**(31): p. 2061-2069.
50. Campbell, H., et al., *Costs and quality of life associated with acute upper gastrointestinal bleeding in the UK: cohort analysis of patients in a cluster randomised trial.* BMJ open, 2015. **5**(4): p. e007230.
51. NHS Digital. *Hospital Episode Statistic Dictionary* [cited 2018 1 Sept]; Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary.
52. Wilson, E., et al., *Cost–utility and value-of-information analysis of early versus delayed laparoscopic cholecystectomy for acute cholecystitis.* British Journal of Surgery, 2010. **97**(2): p. 210-219.
53. Brazzelli, M., et al., *Clinical effectiveness and cost-effectiveness of cholecystectomy compared with observation/conservative management for preventing recurrent symptoms and complications in adults presenting with uncomplicated symptomatic gallstones or cholecystitis: a systematic review and economic evaluation.* Health Technology Assessment, 2014.
54. Macafee, D.A.L., et al., *Prospective randomized trial using cost–utility analysis of early versus delayed laparoscopic cholecystectomy for acute gallbladder disease.* BJS, 2009. **96**(9): p. 1031-1040.
55. Sullivan, P.W., et al., *Catalogue of EQ-5D scores for the United Kingdom.* Medical Decision Making, 2011. **31**(6): p. 800-804.
56. Sutton, A.J., et al., *Cost-effectiveness of emergency versus delayed laparoscopic cholecystectomy for acute gallbladder pathology.* Br J Surg, 2017. **104**(1): p. 98-107.
57. Meacock, R., S.R. Kristensen, and M. Sutton, *The cost-effectiveness of using financial incentives to improve provider quality: a framework and application.* Health Econ, 2014. **23**(1): p. 1-13.
58. Manca, A., N. Hawkins, and M.J. Sculpher, *Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility.* Health economics, 2005. **14**(5): p. 487-496.
59. Diaz-Ordaz, K., A. Franchini, and R. Grieve, *Methods for estimating complier average causal effects for cost-effectiveness analysis.* Journal of the Royal Statistical Society: Series A (Statistics in Society), 2018. **181**(1): p. 277-297.
60. Latimer, N.R., *Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide.* Medical Decision Making, 2013. **33**(6): p. 743-754.
61. Basu, A. and J.L. Gore, *Are elderly patients with clinically localized prostate cancer overtreated? Exploring heterogeneity in survival effects.* Med Care, 2015. **53**(1): p. 79.
62. Heckman, J.J. and E.J. Vytlacil, *Local instrumental variables and latent variable models for identifying and bounding treatment effects.* Proceedings of the national Academy of Sciences, 1999. **96**(8): p. 4730-4734.
63. Heckman, J.J. and E. Vytlacil, *Local instrumental variables*, in *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, C. Hsiao, K. Morimue, and J. Powell, Editors. 2001, Cambridge University Press. p. 1-46.
64. Royal College of Surgeons. *Commission on the future of surgery*. 2018 [cited 2018 22 Apr]; Available from: https://www.rcseng.ac.uk/standards-and-research/future-of-surgery/.
65. NHS England. *Shared Decision Making Programme*. Right Care [cited 2018 22 Apr]; Available from: https://www.england.nhs.uk/rightcare/shared-decision-making/.
66. Staniszewska, S., et al., *GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research.* BMJ, 2017. **358**: p. j3453.