The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

CONSTRUCTING THE ENTIRE CARE PATHWAY OF BOWEL CANCER PATIENTS UNDERGOING EMERGENCY SURGERY TO IMPROVE THEIR QUALITY OF CARE.

SUMMARY OF THE RESEARCH

Background: Bowel cancer is a common disease with over 30,000 new cases per year in England alone. [1] This cancer has a high rate of acute presentation with around one in five patients diagnosed after an emergency hospital admission. [2] Emergency surgery to remove the tumour is usually necessary because the patient's condition is life threatening. However, it is a high-risk procedure, with one in eight patients dying within 90 days. [1]

There are guidelines recommending the care that should be given to patients undergoing emergency bowel surgery, including pre-operative documentation of the risk of surgery, having both a consultant anaesthetist and surgeon present in theatre, and transfer of all high-risk patients from theatre to critical care. [3-6] The evidence supporting these guidelines is limited. National clinical audits provide a rich source of information on bowel cancer patients, their tumour pathology, the care received, and the outcomes of care. This information is collected in several separate datasets:

- The National Bowel Cancer Audit (NBOCA) [1]
- The National Emergency Laparotomy Audit (NELA) [7]
- The Intensive Care National Audit Programme (ICNAP) [8]
- Hospital Episode Statistics dataset (HES) [9] with Office for National Statistics Mortality data (ONS) [10]

These datasets have not been analysed together to date. By linking these national datasets together the entire care pathway can be constructed so that the factors which have the greatest impact on the outcomes of care can be identified. In order to do this, methodology must be developed to minimise bias from data linkage errors, incomplete capture of patients in each audit, and confounding due to patient selection, in which the care a patient receives depends on how sick they are.

Aims and anticipated impact: The aim of this project is to identify which processes of care reduce mortality and complications in bowel cancer patients undergoing emergency surgery. In order to do this, methods will be developed to minimise bias from data linkage errors, incomplete capture of patients into each audit, and confounding due to patient selection. This will ensure that funding and efforts are directed towards the interventions with the greatest impact on patient outcomes. However, this research will have an impact far beyond the care for patients with bowel cancer. The methods that will be developed are relevant for the study of all conditions where care is being delivered by multiple providers.

Research plan: Firstly, we will develop methods for linking patients across datasets by improving the existing techniques used by NHS Digital to link national datasets in order to reduce linkage errors ("missed matches" and "false matches"). This will include developing probabilistic linkage methods (i.e. producing match weights that reflect the likelihood that two records belong to the same individual, given agreement/disagreement on a set of matching variables), and assessing the benefit of using additional information to link datasets (as well as patient identifiers), such as admitting hospital and dates and types of admissions and procedures ("indirect identifiers").

Secondly, we will evaluate a new "spine" approach to data linkage that makes use of existing linkages between the administrative dataset of all hospital admissions, Hospital Episode Statistics data (HES) and national clinical datasets. This simple approach is potentially more secure and efficient than linking each dataset in turn, because patient information does not need to be passed between organisations for each linkage (and therefore security risk is minimised). However, this spine approach will miss patients who are not identified in HES, potentially introducing selection bias. The spine approach therefore represents a trade-off between "data security" and "data quality". We will assess the impact of using the spine approach by identifying scenarios in which it is sufficiently robust.

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

Third, we will develop coding strategies to determine how to synthesise multiple sources of the same information across datasets, some of which will be conflicting. For example, surgical approach, procedure type and metastatic cancer are all recorded in NBOCA, NELA and HES, but may be recorded differently in each. This will result in a master linked dataset.

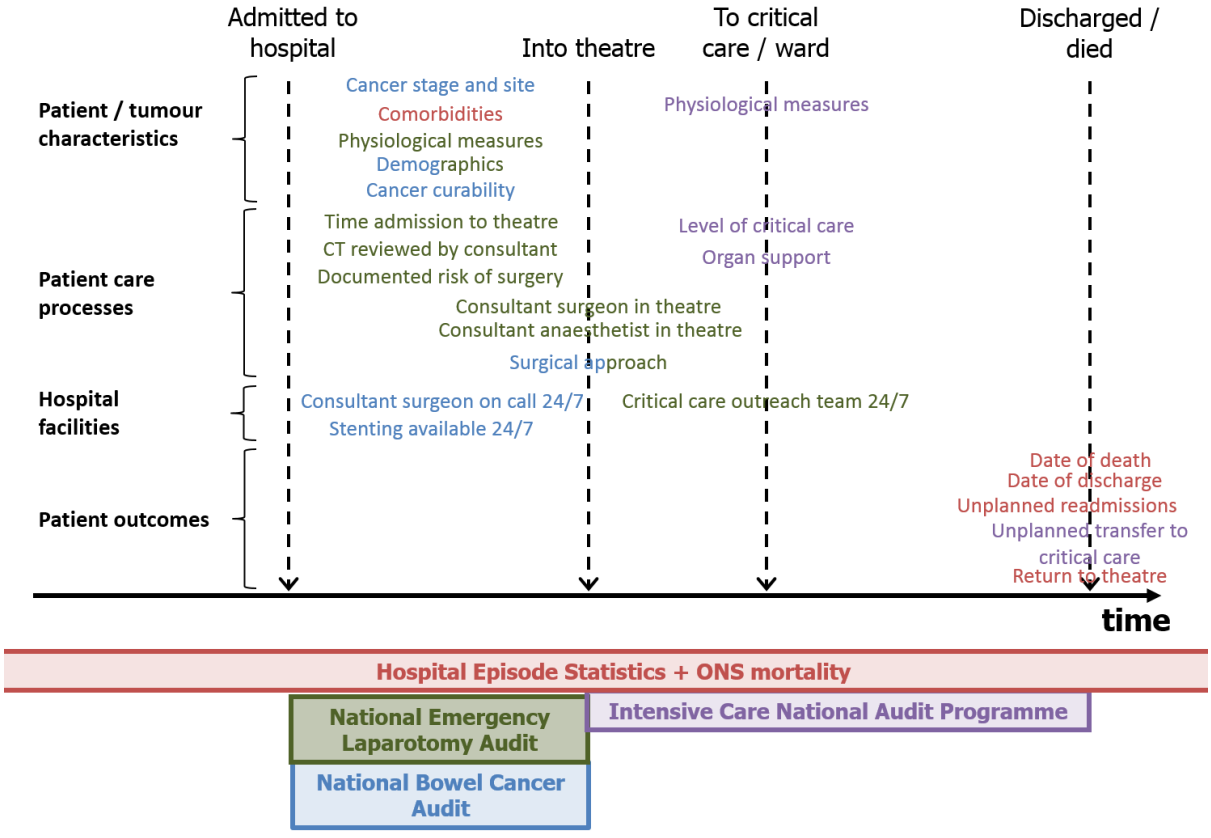
Finally, we will analyse the master linked dataset to identify the components of care which have the greatest impact on patient outcomes, using careful risk-adjustment methods to deal with confounding by patient selection. We will assess the sensitivity of our findings to the methods that we have developed by comparing the estimates obtained using the different methods of data linkage.

BACKGROUND AND RATIONALE

Bowel cancer is one of the most common cancers in the UK and one in five patients are diagnosed after an emergency admission to hospital. [1] One in eight dies after emergency surgery to remove the tumour. [1]

Guidelines recommend how care should be delivered to these patients, such as documenting the risk of surgery, getting patients into theatre within a recommended time, having a consultant surgeon and anaesthetist present in theatre, and transferring patients to critical care after theatre. [3-6,11] However, evidence that supports the guidelines is limited. Evidence could be generated using national clinical datasets that routinely collect data on patients, their tumour, care, and outcomes. However, data are collected in several separate disconnected datasets, as summarised in Figure 1.

Figure 1: Examples of some of the key data items available from each dataset at each point in the patient pathway. The colour of each data item corresponds to the colour of the dataset in the box along the bottom of the diagram.



Identifying the processes of care that have the greatest impact on outcomes could be achieved through linking these national datasets together to reconstruct the entire care pathway. Example questions that could only be answered through linking all of the national datasets are listed below, together with the dataset that collects each piece of information.

- Does a patient's delay into theatre [NELA] affect their risk of death [ONS] or surgical complications [ICNAP], HES], after taking into account how sick they are [NBOCA, NELA,

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

HES], how advanced their cancer [NBOCA], and the other processes of care they receive [NELA, NBOCA, ICNAP]?

- Does direct transfer to an appropriate level of critical care [NELA, ICNAP] affect a patient's risk of death [ONS] or surgical complications [ICNAP, HES], after taking into account how sick they are [NBOCA, NELA, HES], how advanced their cancer [NBOCA], and the other processes of care they receive [NELA, NBOCA, ICNAP]?
- Does having a consultant colorectal surgeon available at all times at a hospital [NBOCA] affect patients' risk of death [ONS] or surgical complications [ICNAP, HES], after taking into account how sick patients are [NBOCA, NELA, HES], how advanced their cancer [NBOCA], and the other processes of care they receive [NELA, NBOCA, ICNAP]?

In each example, information is needed from every dataset in order to define the relevant care processes, outcomes and risk-adjustment factors. Each analysis therefore relies on the linkage of all four datasets.

The first two questions are examples of how patient-level factors affect outcomes, whereas the third question is an example of how a hospital-level factor affects patient outcomes. Hospital-level information is available from organisational audits carried out as part of NELA and NBOCA to collect data on the services available at each hospital.

However, linking multiple datasets is not straightforward:

- Inaccurate or incomplete identifiers (e.g. NHS number and postcode) can lead to errors in linking patients. Rates of linkage errors can differ between patient sub-groups, due to differing data quality. [12,13] This can potentially introduce selection bias, if particular sub-groups of individuals are less likely than others to be included in the linked dataset, and can mean true associations are missed, or false associations found.[14-16] Linkage errors are compounded when linking multiple data sources with errors in each linkage.
- Transfer of personal information for linkage is associated with a risk of disclosure. This risk increases with the number of datasets to be linked. Differences in processes and documentation required by each organisation also have to be overcome.
- Therefore, existing methods for linkage of multiple datasets need to be further developed to ensure that an analysis of the linked dataset produces results that are accurate and reliable, and that the data linkage process is secure and efficient.

EVIDENCE EXPLAINING WHY THIS RESEARCH IS NEEDED NOW

In 2011 the National Confidential Enquiry into Patient Outcomes and Death raised concerns about the quality of care in emergency surgery in the UK. [4] Although survival of these patients is improving, a recent international comparison of emergency surgery found that outcomes were worse in England than in Australia or the US for patients undergoing emergency abdominal surgery. [17]

Over the last three years processes of care for patients undergoing emergency laparotomy in England and Wales have improved. For example, the proportion of patients with a consultant surgeon and anaesthetist present in theatre increased from 70% to 79%, the proportion with preoperative risk documented increased from 56% to 71%, and the proportion with a CT scan reported by a consultant anaesthetist has increased from 73% to 79%. [7] Over the same time-period 90-day mortality of patients having emergency surgery for bowel cancer has fallen from 14.2% to 11.9% [1].

It is currently difficult to establish which processes of care improve patient outcomes because:

1. Different aspects of care are collected in different datasets.
2. The sickest patients tend to receive the most prompt and thorough care, leading to so-called "confounding by patient selection".

Data have been collected since 2014 on the details of perioperative care in bowel cancer patients undergoing emergency surgery, in the National Emergency Laparotomy Audit. Patient and pathology information is also collected on these patients in the National Bowel Cancer Audit, and critical care

data is collected in the Intensive Care National Audit Programme. Linking these datasets together will provide, for the first time, a rich source of information on around 3,500 bowel cancer patients having emergency surgery per year [1,7].

There is wide variation in hospital adherence to guidelines, such as documented risk of surgery, consultant presence in theatre, and transfer to critical care [7] and hospitals are improving their adherence to guidelines at different rates. [7] By modelling processes and outcomes of patients over time, and across hospitals, we have a unique opportunity to disentangle the processes of care that have the greatest impact on patient outcomes. Patient-related, hospital-related and staff-related factors of care all need to be considered. The temporal relationship between processes of care and outcomes for each patient, together with modern statistical methods, will strengthen support for cause-effect associations.

AIMS AND OBJECTIVES

The aim of this project is to improve the quality of emergency care of bowel cancer patients in order to save lives and reduce complications. The project will identify the processes of care with the greatest impact on mortality and complications of surgery. The information required to do this is collected in England in several separate datasets. These will be linked in order to construct the entire patient pathway. However, methodological issues need to be overcome:

- Standard methods for linking healthcare datasets by NHS Digital rely on deterministic linkage of a set of patient identifiers e.g. NHS number, sex, date of birth and postcode. This involves a set of deterministic rules that are used to classify records as links or non-links. Such linkage methods could be improved by (i) making use of a wider set of information common to both datasets including identifiers and “indirect identifiers” such as dates and types of admissions and procedures; and (ii) by generating weights that rank record pairs in terms of the likelihood of them belonging to the same patient (probabilistic linkage).
- Most national clinical audits have already been linked to hospital admission data (HES).Using HES as the “spine” would be possible without requiring any further linkages or release of patient identifiers. However, the spine approach would miss any patients not identified in HES, potentially leading to selection bias.
- There will be multiple sources of the same information across datasets, some of which will be conflicting. For example, surgical approach, procedure type and metastatic cancer are all recorded in NBOCA, NELA and HES, but may not agree between all of the sources. Coding strategies need to be developed to overcome this.
- Causal relationships between processes and outcomes of care can be difficult to establish, especially as the care that patients receive depends on how sick they are, known as confounding by patient selection. Risk-adjustment methods need to be developed to overcome this potential bias.

This project will develop improved methods for linking multiple datasets and will assess the impact of the choice of methods on the findings of the research to identify which processes of care have the greatest impact on patient outcomes. The specific objectives are:

1. Identify which processes of care have the greatest impact on outcomes of patients with bowel cancer undergoing emergency surgery, making use of the linked datasets. In order to do this:
2. Determine the most accurate methods for linking multiple datasets in order to minimise bias due to linkage errors.
    - Evaluate probabilistic versus deterministic linkage strategies
    - Evaluate the use of indirect identifiers such as admitting hospital and dates and types of admissions and procedures, in place of and in addition to patient identifiers.
3. Assess a new approach to data linkage, which makes use of existing linkages between HES and national clinical datasets.
    - Compare this new approach, using HES as the spine, to a more inclusive approach in which every dataset is linked to every other dataset (pairwise linkage).
4. Develop coding strategies to synthesise the same information collected in multiple datasets

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

5. Develop a risk-adjustment model specific to bowel cancer patients undergoing emergency surgery, incorporating information from across the datasets about patient and tumour characteristics, and physiological measures.
6. Assess the extent to which the methods for data linkage and risk-adjustment impact on the findings.

RESEARCH PLAN / METHODS

The objectives will be tackled in four work packages (WPs), initially focussing on methods required to create a comprehensive, high quality dataset with which to investigate the effects of processes of care on outcomes in bowel cancer patients undergoing emergency surgery, and finally assessing the sensitivity of these analyses to different methods.

WP1: Improve the methods for linking multiple datasets to minimise linkage errors [Months 1-12]

The standard approach to linking national clinical audits is to use deterministic linkage on four patient identifiers (NHS number, date of birth, sex and postcode). This approach is sub-optimal because (i) it makes use of a limited amount of information to link records, and (ii) it is a deterministic method, meaning that errors or missing values in the identifiers can prevent records belonging to the same individual being linked, known as “missed matches”. Missed matches will reduce statistical power and, where linkage errors do not occur randomly, can lead to under- and over-estimation of associations. [12,13] “False matches”, in which records of different patients are incorrectly linked, tend to increase variation in the data and therefore lead to conservative estimates of associations.

Linkage accuracy can be improved by:

1. Using indirect identifiers, such as dates of admissions and procedures, as well as patient identifiers, to link records. [18-21] The use of additional information has the potential to reduce both missed matches (i.e. by providing additional information where patient identifiers are missing) and false matches (i.e. by better discriminating between records with similar identifiers). [22] In the current climate of stricter control of patient data, the use of ‘indirect’ identifiers alone, without patient identifiers, has great potential to preserve data security. It would allow analysts to link anonymised or pseudonymised datasets without the need to transfer patient identifiers between multiple organisations, which can lead to increases in the risk of disclosure of personal information as well as substantial delay.
2. Probabilistic linkage, which can substantially increase the proportion of records that can be linked. [22,23] This technique generates a match weight for each pair of records, which represents the likelihood that records belong to the same individual. Match weights are generated from the pattern of agreement between identifiers in different records, where agreement contributes positively to the match weight, and disagreement contributes a penalty. The way in which match weights are calculated takes into account the discriminatory power of identifiers (so that, for example, agreement on date of birth would produce a higher match weight than agreement on sex). It can also incorporate information on how common or rare a matching value is, for example by assigning higher weights to very old patients (who are less likely to agree on date of birth by chance) than to average aged patients (who are more likely to agree on date of birth by chance). Records are then classified as links or non-links by comparing match weights with a cut-off threshold. Work carried out by our team linking baby records to their mother’s records in HES, found that probabilistic linkage using only indirect identifiers linked 98% of babies to mothers. Probabilistic linkage can, potentially, increase the rate of false matches, depending on the threshold chosen, but probabilistic linkage methods have been shown to achieve very high linkage rates at the cost of minimal false linkages. [22-24]

We will evaluate four approaches to data linkage, the first of which is the approach currently used to link national clinical audits:

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

- Deterministic linkage on patient identifiers
- Probabilistic linkage on "proxy" identifiers, described below
- Probabilistic linkage on indirect identifiers only
- Probabilistic linkage on indirect identifiers and "proxy" identifiers

NHS Digital will carry out the standard deterministic linkage on four patient identifiers and the other linkage methods will be carried out by our team. We hold data on patient sex but not NHS number, date of birth or postcode. NHS Digital's linkage report will tell us whether there is an exact match on NHS number or not, and we will use month and year of birth as an alternative to date of birth, and lower super output area of residence as an alternative to patient postcode. We call these set of identifiers "proxy" identifiers to distinguish them from the patient identifiers used by NHS Digital to carry out data linkage. Rather than asking NHS Digital to carry out any bespoke linkage methods for this project, our approach is to replicate what is currently available to analysts of linked datasets so that the methods developed here will be widely applicable to audits, service evaluation projects and research in which national clinical datasets are linked.

Each pairwise linkage between the datasets will be evaluated separately. WP2 will explore the combinations of these pairwise linkages.

Evaluation of the linkage methods

The linkage methods will be evaluated using a published three-step process developed by members of our team: [25]

1. *Applying the linkage algorithms to a subset of gold standard data to quantify linkage error*
2. *Comparing the characteristics of linked and unlinked data from each linkage method to identify potential sources of bias*
3. *Evaluating the sensitivity of our results to the choice of linkage procedure.*

1. Applying the linkage algorithms to a subset of gold standard data to quantify linkage error

We will generate a "gold standard" linked dataset that will be used to evaluate each of the four linkage approaches described above. We will use established methods for creating a gold-standard dataset using manual review [26]: a subsample of record pairs will be drawn from the datasets to be linked, and the "true" match status of each record pair will be determined from manually inspecting all identifiers and other information available in both datasets (e.g. patient identifiers, indirect identifiers, and other clinical and geographic information). [27] The gold standard dataset will include certain matches (e.g. those agreeing on all matching variables), certain non-matches (records disagreeing on all matching variables), and records that are more difficult to classify (those with agreement on some matching variables and disagreement on others). For each record pair, we will determine the true match status, and this will be compared to the results from each of the four linkage methods described above. The linkage quality of each method will be quantified using the following metrics: [23]

- Sensitivity: The proportion of true matches that are linked
- Positive predictive value (PPV): The proportion of linked records that are true matches
- False match rate: The proportion of non-matches that are linked

2. Comparing the characteristics of linked and unlinked data from each linkage method to identify potential sources of bias

Comparisons will be made of the characteristics of linked and unlinked records of patients identified as eligible for the study. [25] This will allow us to identify any potential sources of bias, if particular subgroups of patients are more difficult to link (and therefore less likely to be included in the linked dataset used for analysis). By making comparisons across risk-adjustment variables, processes of care, hospital facilities and patient outcomes, we will be able to assess the likely impact on the analyses of the effect of care processes on patient outcomes.

3. Evaluating the sensitivity of our results to the choice of linkage procedure.

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

Strategies 1 and 2 above will provide overall measures of linkage error, and information on whether particular groups of patients are more affected by linkage error than others. Conducting a sensitivity analysis will enable us to further determine whether any linkage error is likely to introduce bias into results. In WP4b we will carry out a sensitivity analysis to evaluate the impact of the different linkage methods on the effect estimates in the analyses of interest. We will re-run each analysis using the four linkage approaches to assess how much the findings change, whether linkage bias under-estimates or over-estimates the effects of care processes and under what circumstances the impact is minimal. Although all linkage methods may have some level of linkage error, this strategy will enable us to evaluate the direction and extent of any bias in the findings. Moreover we will be able to assess the effect that different methods have on the precision of the analyses and therefore what we are able to conclude about relationships with confidence.

WP2: Assess the "spine" approach, a new approach to data linkage which makes use of existing linkages between HES and national clinical datasets [Months 10-18]

The linkage approach found to be of the highest quality in WP1 will be used as a primary strategy to carry out the linkage between each pair of datasets in this work package (other approaches will be implemented in WP4b). This should provide near-perfect linkage because, in patients undergoing major surgery in the National Bowel Cancer Audit, 93% of patients could be liked to HES using deterministic linkage on four patient identifiers [1] and probabilistic linkage methods are known to achieve even higher linkage rates at the cost of minimal false linkages. [22,23] The sensitivity and specificity of probabilistic linkage of UK electronic data on patient identifiers have both been estimated to be higher than 99%. [24] In this work package, therefore, we assume that patients captured in one or more audits but missing from another audit are missing because of incomplete capture rather than linkage errors.

Most national clinical audits are already linked to HES and therefore linkage of multiple audits using HES as the "spine" would not require any further linkages. The use of multiple linked national datasets is growing rapidly, and this new linkage strategy could provide a fundamental shift in the approach to data linkage across national audits, research projects, service evaluation projects and clinical trials for which more than two clinical datasets need to be linked. The approach would not require any further data linkages and would therefore avoid the need to transfer additional patient data between organisations. Potential benefits include reduced risk of disclosure of personal information, reduced resources required to carry out data linkage, and substantial reduction in delays to projects.

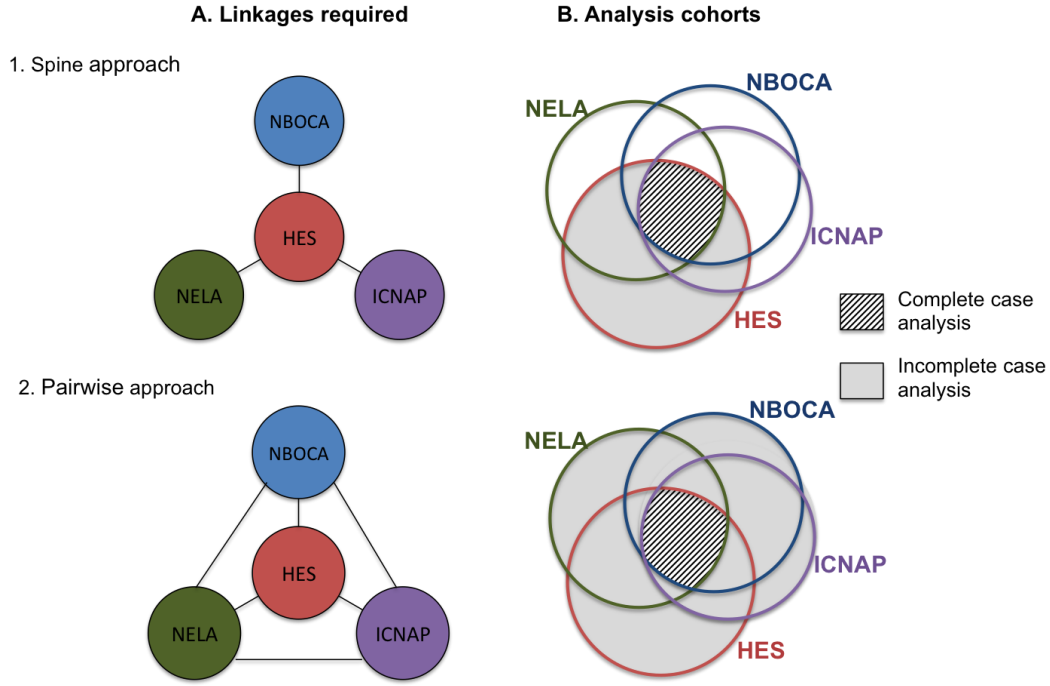
The potential risk of this new approach is that any patients not identified in HES would be missing from the analysis cohort. This would reduce statistical power and, if patients not identified in HES are not representative of all patients, could lead to bias in effect estimates. The primary use of HES is to allow hospitals to be paid for the care they deliver. Patients in this study are undergoing major surgery, therefore we expect the number missing from HES to be minimal, but due to clinical coding errors, patients may be included in HES but not recognised as eligible for the study.

Figure 2 shows two alternative approaches to linkage, a new "spine" approach and a more inclusive "pairwise" approach. The spine approach is less inclusive because eligible patients not identified in HES but captured in the other datasets cannot be included. This is because for these patients we do not know which records in NBOCA, NELA or ICNAP belong to the same patient. If we included the records in the analysis we could be including the same patient more than once. The pairwise approach can include all patients captured in at least one dataset, but the trade-off is that it requires additional linkages.

The darker shaded areas represent the analysis cohort if only those captured in every dataset are included, and we refer to this as the "complete case analysis". An alternative is to include patients who are missing from one or more datasets and consider the information from these datasets to be missing, which we call the "incomplete case analysis". The incomplete case analysis would include more patients but with less complete data, and missing data methods would be required. In reality there will be much more overlap between eligible patients in the datasets. The lack of overlap is exaggerated for demonstration.

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

Figure 2: Spine and pairwise linkage approaches



The different linkage approaches lead to three possible analysis cohorts:

- "Spine/pairwise-linkage", complete case analysis (includes only patients captured in all datasets)
- "Spine-linkage", incomplete case analysis (includes only eligible patients captured in HES)
- "Pairwise-linkage", incomplete case analysis (includes all patients eligible according to any of HES, NBOCA, NELA or ICNAP)

These three analysis cohorts will be compared in terms of statistical power and bias arising from excluding eligible patients. The third cohort will comprise all bowel cancer patients undergoing emergency surgery, except patients missing from all four datasets. The number missing from all four datasets is likely to be small as case ascertainment of the datasets has been shown to be high [1,7].

Comparing eligible patients across datasets will allow estimates of the capture into each dataset, and the capture into multiple datasets. This will allow us to quantify the overlap between the datasets in Figure 2. The key metric when assessing the spine-linkage versus pairwise-linkage approaches will be the proportion of eligible patients that each method captures. This will be assessed alongside the amount of missing information across the key data items. Secondly, the characteristics of patients captured will be compared to those not captured to assess the potential bias from subgroups of patients being less likely to be included. Thirdly, the sensitivity of the results to the choice of linkage approach will be assessed in WP4b.

WP3: Synthesise information from multiple data sources [Months 16-22].

Some information, such as metastatic cancer, surgical approach, and admission to intensive care is collected in multiple datasets. This information may be conflicting, missing, or defined differently. We will develop coding strategies to synthesise relevant information across data sources. This methodological work is not only an essential preparation for data analysis, but it will also guide strategies on how information from multiple sources can be best synthesised in other linked datasets in the future, beyond the area of bowel cancer care.

Synthesising conflicting information
First we will carry out validity checks within and between datasets, identifying inconsistencies between data items such as open surgery which is recorded as being converted from laparoscopic to

open, indication that a patient was not admitted to intensive care but with a recorded level of intensive care, etc. Next we will check the correlation of information with known associates (e.g. ASA grade with postoperative mortality, metastatic cancer with long-term survival, surgical approach with length of hospital stay etc). Together these metrics will allow us to rank the reliability of the same information across data sources so that a hierarchy of data sources can be developed for each data item. This hierarchy will be used to resolve conflicts in the linked data.

Synthesising information defined differently
The clinical members of our research team will be crucial in the development of methods for combining information from multiple sources when data items are defined differently. For example, surgical urgency is measured using different versions of the NCEPOD-classification [30] in NBOCA and NELA, and HES and NBOCA define mode of admission differently. Sometimes algorithms within a dataset are used to generate categories, for example individual surgical procedures and diagnoses are categorised into different procedure and diagnosis groups in NBOCA and NELA. Categories will be synthesised across the two datasets by comparing the algorithm used in each dataset, and, where available, by comparing with more detailed information in a third dataset, such as individual procedure codes and diagnosis codes in HES.

WP4a: Identify which processes of care have the greatest impact on outcomes of patients with bowel cancer undergoing emergency surgery [Months 20-32]

Data sources and variables of interest

At the start of the project there will be 3 years and 4 months of data available across the national datasets (patients per year having surgery / diagnosed November 2013 to March 2017), providing approximately 11,700 bowel cancer patients undergoing emergency surgery (based on 3,500 per year) [1,7] By the start of year two of the project, there will be an additional year of data, providing over 15,000 patients for the study.

Some data sources measure care directly against national recommendations for the care of emergency bowel cancer patients (NELA, NBOCA) but others do not (ICNAP). A range of care processes will be investigated, including delay into theatre, seniority of surgeon and anaesthetist in theatre, and transfer from theatre to critical care. Our clinical and patient partners will be essential in identifying patients groups and combinations of care processes (as statistical interactions) to be prioritised. The key outcomes will be postoperative mortality, longer-term survival, and complications from surgery (unplanned readmissions and returns to theatre). Input from clinicians and patients will help to identify any other important outcomes.

Examples of the types of question to be answered are:

- Does a patient's delay into theatre [NELA] affect their risk of death [ONS] or surgical complications [ICNAP], Hospital Episode Statistics (HES)], after taking into account how sick they are [NBOCA, NELA, HES], how advanced their cancer [NBOCA], and the other processes of care they receive [NELA, NBOCA, ICNAP]?
- Does direct transfer to an appropriate level of critical care [NELA, ICNAP] affect a patient's risk of death [ONS] or surgical complications [ICNAP, HES], after taking into account how sick they are [NBOCA, NELA, HES], how advanced their cancer [NBOCA], and the other processes of care they receive [NELA, NBOCA, ICNAP]?
- Does having a consultant colorectal surgeon available at all times at a hospital [NBOCA] affect patients' risk of death [ONS] or surgical complications [ICNAP, HES], after taking into account how sick patients are [NBOCA, NELA, HES], how advanced their cancer [NBOCA], and the other processes of care they receive [NELA, NBOCA, ICNAP]?

Each dataset will contribute to different elements of:

Population: Bowel cancer patients undergoing emergency surgery
Intervention: A particular process of care

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

Comparison group: Representative set of patients not given the process of care or cohort matched for important factors / risk-adjusted / propensity score methods
Outcomes: Mortality and complications of surgery

Table 1: Data sources for each element of the study PICO

	NBOCA	NELA	ICNAP	HES/ ONS
Population	●	●	●	●
Intervention	●	●	●	
Comparison group	●	●	●	●
Outcomes:				
Mortality: postoperative and long-term				●
Unplanned readmission				●
Return to theatre		●		●
Unplanned admission to critical care			●	

Statistical methods
Establishing causal relationships will require careful risk adjustment to deal with the potential bias from sicker patients being treated differently, known as confounding by patient selection. Risk adjustment methods such as multivariable regression or propensity score methods will be developed to accommodate potential confounding, missing data and clustering within hospitals simultaneously. [28,29] Clinical guidance will be crucial to understand a priori the important factors about underlying patient risk and the clinical decision making around care processes.

Multivariable regression models the relationship between processes and outcomes of care, adjusting for differences in patient and tumour characteristics, physiological status, other processes of care, and hospital and clinician factors. A well-defined risk score has been developed for all patients undergoing emergency laparotomy for any indication by members of our team. [31] A well-defined risk score has also been developed for patients across all surgical urgencies in NBOCA, again by members of our team. [32] A new risk adjustment model, specific to bowel cancer patients undergoing emergency surgery, will need to be developed, incorporating information from across the datasets about patient and tumour characteristics, and physiological measures.

As an alternative, propensity score methods will also be considered. These methods directly model the chance of a patient receiving a particular process of care based on their patient and tumour characteristics, physiological status, other processes of care, and hospital and clinician factors. Patients can then be matched on propensity scores, or stratified according to propensity score, or inverse probability weighting or risk adjustment can be used based on the propensity score. Expertise on propensity score methodology, particularly in large electronic health databases and where missing data is also a problem is a major research theme at LSHTM and we will draw on that expertise in our work. [33] We will provide an application to bowel cancer that will be accessible for clinicians and applied statisticians.

Propensity score methods can be preferable to multivariable regression when the number of risk factors is large compared to the size of the sample (unlikely to be an issue here) or when there are stark differences in characteristics between those receiving and not receiving a process of care (likely here). These methods also allow the estimation of the most relevant effect, whether it is the average effect of the care process in all patients, just in the patients who currently receive the care process, or just in the patients who do not currently receive it. [29] The choice of propensity score method will depend on the amount of overlap in propensity scores between the comparison groups. [34].

WP4b: Assess the impact the choice of methods for data linkage has on the findings [Months 26-32]
Each of the methods used in WP1, WP2 and WP4a may lead to different estimates of the impact of processes of care on outcomes. By comparing the estimates obtained using different combinations of the methods we will assess how sensitive the findings are to the choice of methods, and in what circumstances they differ.

The different methods used in WP1, WP2 and WP4a are as follows:

1. Linkage methods

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

- Deterministic linkage on patient identifiers
- Probabilistic linkage on proxy identifiers
- Probabilistic linkage on indirect identifiers only
- Probabilistic linkage on indirect identifiers and proxy identifiers

2. Spine versus linkage approaches
   - Spine / pairwise, complete case analysis (includes patients captured in all datasets)
   - Spine linkage, incomplete case analysis (includes eligible patients captured in HES)
   - Pairwise linkage, incomplete case analysis (includes all patients eligible according to any of HES, NBOCA, NELA or ICNAP)
3. Risk-adjustment methods
   - Multivariable regression
   - Propensity score methods (more than one may be appropriate and any used will be included in this sensitivity analysis)

The sensitivity analysis will include combinations of all of the above methods. The analyses in WP4a will be repeated with each combination, to explore the impact each of the methods has on the findings.

DISSEMINATION AND PROJECTED OUTPUTS

Based on our results, we will formulate recommendations on how the care of bowel cancer patients undergoing emergency surgery can be improved, as well as on the methods developed for linking and analysing multiple datasets. Our strategy to disseminate our findings will be based on a number of principles:

- We will build awareness of the project among patient groups, professional bodies and methodologists in the project's early phase through direct communications and presentations.
- We will make use of existing networks and relationships available within LSHTM to disseminate the clinical and methodological findings of the research, the RCS, the RCA and the Association of Coloproctology and the Farr Institute. Dissemination methods will include websites, presentations, press releases and social media such as Twitter and Facebook.
- We will use these networks and relationships as a two-way process: they will help to inform the research and to ensure effective dissemination of results.
- We will have clear messages that fit our wide range of audiences (e.g. patient and public, commissioners, clinicians, regulators, policy makers and methodologists).
- We will time publications around conferences and relevant events

Our findings will be disseminated as follows:

- A research report for the NIHR HS&DR programme detailing research methods, findings and conclusions of all four WPs, including recommendations for practice and an extensive summary for patients and the wider public.
- Policy advice targeting NHS England at national level and regional Clinical Senates and Academic Health Science Networks.
- Feedback to NICE, especially with regards to their clinical guideline on acutely ill adults in hospital.
- Advice to NHS England and Healthcare Quality Improvement Partnership about priorities for national audits and the methods for linking multiple audits.
- Advice to NHS Digital, National Cancer Registration and Advisory Service and other linked data providers on methods for data linkage.
- Presentations and reports targeting relevant professional bodies, including the RCS, the RCA, the Association of Coloproctology, the Health Services Research Network, the International Population Data Linkage Network, the Administrative Data Research Centre for England, and the Farr Institute (of which LSHTM is a partner).

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

- Presentations and reports to patient organisations, including the Bowel Cancer UK and Beating Bowel cancer. We will involve our PPI representatives and NBOCA and NELA patient and charity representatives to ensure that we address all relevant organisations and that the style and format of our publications is accessible.
- Research papers for peer-reviewed academic journals (e.g. BJS, BJA, IntJEpi), articles for clinical journals, and conference presentation

PLAN OF INVESTIGATION AND TIMETABLE

Tasks (time period in months from project start)

Pre-start

Recruitment of Research Fellow

NHS Digital pairwise deterministic data linkage using patient identifiers

WP1: Improve the methods for linking multiple datasets to minimise linkage errors (1-12)

Probabilistic linkage using proxy identifiers (1-3)

Probabilistic linkage using indirect identifiers (2-4)

Probabilistic linkage using proxy identifiers and indirect identifiers (3-5)

Evaluation of linkage methods (6-10)

Write publication (10-12)

WP2: Assess a new approach to data linkage which makes use of existing linkages between HES and audits (10-18)

Spine linkage and pairwise linkage (10-12)

Evaluation of size of spine and pairwise linkage (13-16)

Write publication (16-18)

WP3: Synthesise information from multiple data sources (16-22)

Synthesise conflicting information (16-18)

Synthesise differently defined information (18-20)

Write publication (20-22)

WP4a: Identify which processes of care have the greatest impact on outcomes of patients with bowel cancer undergoing emergency surgery (20-32)

Determine most relevant clinical questions regarding which processes of care impact on patient outcomes (20-22)

Develop new risk adjustment model (22-28)

Develop appropriate propensity score methods (24-30)

Write publication (30-32)

WP4b: Assess the impact the choice of methods for data linkage has on the findings (26-32)

Sensitivity analysis using different combinations of methods from WP1, WP2 and WP4a (26-30)

Write publication (30-32)

Dissemination (30-36)

Formulate recommendations and implications for bowel cancer surgery services (30-36)

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

Dissemination activities (30-36)

Final report for NIHR (32-36)

PROJECT MANAGEMENT

Dr Kate Walker will, as principal investigator, take overall responsibility for leadership and management of the study. She will chair the research project's advisory group who will convene every four months. This group will oversee the implementation of the study and comprise of all co-applicants, clinical collaborators, the public and patient representatives, and the research fellow. It will monitor the progress of completion of tasks against the project's timeline and consider remedial action if needed. The group will also discuss the implications of findings, and decide how they should be disseminated. The group's meetings will be face-to-face for London-based staff and with an option of video-conferencing facilities for those based elsewhere.

The work for this project will be carried out by a full-time research fellow who will be supervised on a daily basis by Dr Kate Walker, supported by Prof Jan Van der Meulen, Prof Linda Sharples and Dr Katie Harron. This research team will communicate on a regular basis with the public and patient representatives and the clinical collaborators to seek their input on all key issues related to research design, method development, data analysis and interpretation and reporting. The research team will meet monthly to discuss all relevant methodological, practical and logistical issues, involving other co-applicants and collaborators when necessary and appropriate.

Administrative support (10% FTE) will be available to help with arranging meetings, dealing with day to day queries and budget management.

APPROVAL BY ETHICS COMMITTEES

The project requires 4 datasets (NBOCA, NELA, ICNAP, HES/ONS) and all 6 data linkages between them. By the start time of the project we will have access to the linked datasets which will be held at the Clinical Effectiveness Unit, a collaborative partnership between the LSHTM and RCS.
CAG approvals are already in place to hold all datasets and to carry out 4 of the 6 data linkages (all except the linkage between ICNAP and HES and ICNAP and NELA). A CAG amendment has been requested to allow these two additional data linkages. The 4 data linkages which have been approved have either already been carried out (NBOCA-HES/ONS, NELA-HES/ONS) or will be linked within the next two months (NBOCA-NELA, NBOCA-ICNAP).

PATIENT AND PUBLIC INVOLVEMENT

The clinical advisory groups of the National Bowel Cancer Audit and the National Emergency Laparotomy Audit, including patient and charity representatives, helped to define the clinical questions. The NIHR North Thames CLAHRC's PPI panel gave feedback on the planned research at its workshop in March 2017. The panel indicated their unanimous support and members made clear that the importance of the research justified the access of the researchers to potentially sensitive patient data. One panel member said in support of the project "I think it is tremendous that it has something to offer both the clinical and data analysis communities" whilst another member felt strongly about the clinical topic, saying "I think it's fantastic. This research has to be done". The panel recommended that PPI representatives should be involved in the research. The planned research project's advisory group will have two PPI members who will be paid according to INVOLVE guidance. The CLAHRC's PPI panel were also asked to review the Plain English summary. Their comments led to a much clearer summary with simpler terms used, and with a clearer explanation of the implications for patients and the NHS.

The PPI panel of the NIHR North Thames CLAHRC have contributed to the design of this research proposal and we will continue to work closely with them for the duration of the project. Two PPI members will be recruited to the research project's advisory group, with the help of North Thames CLAHRC who have much experience of recruiting PPI members of research teams. These public and patient representatives will have a guiding role for all WPs. For example they will be crucial in defining

the aspects of care and outcomes of care that are most important to patients and the public. They will comment, and in some cases, contribute to project reports and research outputs.

We do not expect that any formal training will be required for our PPI members. We will provide support where and when needed to ensure that they have sufficient understanding of the research methods that we are going to use. In conjunction with London School of Hygiene & Tropical Medicine's Public Engagement Advisory Coordinator, our PPI members will be consulted on the best ways to disseminate the findings from this project to patients and the wider public.

EXPERTISE AND JUSTIFICATION OF SUPPORT REQUIRED

We will use existing national databases which we already have access to. This avoids the expense of data collection and makes the study entirely feasible. The largest proportion of the budget will therefore be staff costs.

Staff

Research team members, who will also sit on the research advisory group

Dr Kate Walker (LSHTM) is PI. She is a statistician specialising in complex methodological issues in health services research for patients with bowel cancer, dealing with linked national clinical datasets. She is the lead methodologist on the National Bowel Cancer Audit and provides methodological expertise to the National Emergency Laparotomy Audit. Together with the co-PI, she will be responsible for managing all aspects of the project, including managing the full-time research fellow. She will chair the research advisory group and the research team and will dedicate 20% of her time.

Prof Jan Van der Meulen (LSHTM) is co-PI. He brings to the project his experience in health services research and national clinical audits, particularly in the area of surgery and bowel cancer. He will provide senior oversight of the project, contributing 10% of his time. Together with KW he will ensure that the project runs to schedule and achieves all of its aims.

Dr Katie Harron (LSHTM) is a statistician with internationally recognised expertise on the methods for data linkage, including probabilistic linkage of national clinical datasets using indirect identifiers. She will lead on the development of methods for data linkage in Work Package 1, contributing 5% of her time.

Prof Linda Sharples (LSHTM) is a statistician with experience in observational and experimental studies of surgical interventions. She will contribute senior statistical expertise, in particular on multilevel modelling, risk adjustment including propensity score methods, and missing data. She will contribute 10% of her time.

One research fellow will be employed full-time on the project, planning and carrying out the analyses, and disseminating the findings, under close supervision of KW and JVM.

Other research advisory group members

Prof David Cromwell (LSHTM) is a quantitative health services researcher with experience of using linked datasets to evaluate patterns of surgery and patient outcomes. He is the senior methodologist on the National Emergency Laparotomy Audit. He will contribute to study design, methodological expertise on the National Emergency Laparotomy Audit and development of analytical methods, contributing 2.5% of his time

Dr David Harrison (ICNARC) is an expert on methods for risk adjustment and brings methodological expertise on the ICNARC Intensive Care National Audit Programme. He will provide consultancy time to the project, contributing to the methods for risk adjustment and methodological expertise on the ICNARC Case Mix Programme dataset.

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

Prof James Hill (Central Manchester University Healthcare Trust) is a consultant colorectal surgeon and clinical lead on the National Bowel Cancer Audit, with clinical trials and health services research experience. He will provide senior clinical input and will contribute 2.5% of his time.

Dr Ramani Moonesinghe (UCL) is a consultant anaesthetist and anaesthetic lead on the National Emergency Laparotomy Audit with quality improvement and health services research experience. She is Associate National Clinical Director for elective care for NHS England which will allow her to disseminate findings direct to the core of NHS leadership . She will provide senior clinical input and will contribute 2.5% of her time.

Two PPI representatives will be appointed to sit on the research project's advisory group to provide a perspective on what is most important to patients and the public.

A small amount of administration time has been requested to help with budget management, managing meetings, and dealing with day-to-day enquiries.

Non-staff

Data linkage costs

CEU secure data environment costs

Travel and subsistence costs for advisory group meetings

Project-specific photocopying and printing

Computer and STATA license

Conference attendance (fees, travel and subsistence)

Open access publishing

REFERENCES

1. National Bowel Cancer Audit Report, 2016. NHS Digital, 2016 https://www.nboca.org.uk/reports/annual-report-2016/ (Accessed 10 September 2017).

2. D Wallace, K Walker, A Kuryba, P Finan, N Scott, J van der Meulen. Identifying patients at risk of emergency admission for colorectal cancer. Br J Cancer. 2014;111: 577–580.

3. Emergency Surgery: Standards for unscheduled care, 2011. The Royal College of Surgeons of England, 2011. https://www.rcseng.ac.uk/library-and-publications/college-publications/docs/emergency-surgery-standards-for-unscheduled-care/ (Accessed 10 September 2017)

4. Knowing the Risk: A review of the peri-operative care of surgical patients. A report by the National Confidential Enquiry into Patient Outcome and Death (2011). NCEPOD, 2011. http://www.ncepod.org.uk/2011report2/downloads/POC_fullreport.pdf (Accessed 10 September 2017)

5. The Higher Risk General Surgical Patient. The Royal College of Surgeons of England, 2001. https://www.rcseng.ac.uk/library-and-publications/college-publications/docs/higher-risk-patient/ (Accessed 10 September 2017)

6. Patient Safety: A Consensus Statement. Association of Surgeons of Great Britain and Ireland, 2009. http://asgbidocuments.surgicalmembershipportal.co.uk/Consensus%20Statment/patient_safety_consensus_statement_-_april_2010_as_gone_to_press_.pdf (Accessed 10 September 2017)

7. The Second Patient Report of the National Emergency Laparotomy Audit (NELA). Royal College of Anaesthetists, 2016 http://www.nela.org.uk/reports (Accessed 10 September 2017).

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

8. Annual Quality Report 2015/16 for adult critical care. Intensive Care National Audit Programme, 2016 https://onlinereports.icnarc.org/Home (Accessed 10 September 2017).

9. NHS Digital (2017) Hospital Episode Statistics http://content.digital.nhs.uk/hes (Accessed 10 September 2017).

10. The Office for National Statistics mortality data. http://content.digital.nhs.uk/article/2677/Linked-HES-ONS-mortality-data (Accessed 10 September 2017)

11. National Institute for Health and Care Excellence, Clinical Guideline 50: Acutely ill patients in hospital, 2007 https://www.nice.org.uk/guidance/cg50 (Accessed 10 September 2017).

12. Bohensky M. Chapter 4: Bias in data linkage studies, in Methodological Developments in Data Linkage, Harron K, Dibben C and Goldstein H, Editors. 2015, Wiley: London.

13. Bohensky M, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, Brand CA. Data linkage: A powerful lresearch tool with potential problems. BMC Health Serv Res. 2010;10(1):346-352.

14. Harron K, Wade A, Gilbert R, Muller-Pebody B, Goldstein H. Evaluating bias due to data linkage error in electronic healthcare records. BMC Med Res Method. 2014; 14:36.

15. Harron K, Goldstein H, Wade A, Muller-Pebody B, Parslow R, Gilbert R. Linkage, evaluation and analysis of national administrative data: application to enhanced blood-stream infection surveillance in paediatric intensive care. PLoS One. 2013;8(12):e85278.

16. Moore, C. L., Amin, J., Gidding, H. F., & Law, M. G. (2014). A New Method for Assessing How Sensitivity and Specificity of Linkage Studies Affects Estimation. PLoS ONE, 9(7), e103690.

17. Chana P, Joy M, Casey N, Chang D, Burns EM, Arora S, Darzi AW, Faiz OD, Peden CJ. Cohort analysis of outcomes in 69 490 emergency general surgical admissions across an international benchmarking collaborative. BMJ Open 2017;7(3):e014484.

18. Harron K, Gilbert R, Cromwell DA, van der Meulen JH. Linking data for mothers and babies in de-identified electronic health data. PLoS One. 2016; 11:e0164667.

19. Méray N, Reitsma J, Ravelli A, Bonsel G. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. J Clin Epidemiol. 2007;60(9):883-91.

20. Hammill B, Hernandez A, Peterson ED, Fonarow GC, Schulman KA, Curtis L. Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. Am Heart J. 2009;157(6):995-

21. Hagger-Johnson G, Harron K, Fleming T, Gilbert R, Goldstein H, Landy R, Parslow RC. Data linkage errors in hospital administrative data when applying a pseudonymisation algorithm to paediatric intensive care records. BMJ Open. 2015;5(8):e008118

22. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. Int J Epidemiol. 2016; 45(3): 954–964.

23. Lyons RA, Jones KH, John G, Brooks CJ, Verplancke JP, Ford DV, Brown G, Leake K. The SAIL databank: linking multiple health and social care datasets. BMC Med Inform Decis Mak. 2009;9(3).

1000.

24. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. J Clin Epidemiol. 2011;64,565-572.

25. Harron K, Doidge JC, Knight HE, Gilbert RE, Goldstein H, Cromwell DA, van der Meulen JH. A guide to evaluating linkage quality for the analysis of linked data. Int J Epidemiol. 2017;1–12.4.

26. Paixão ES, Harron K, Andrade K, Teixeira MG, Fiaccone RL, Costa MDCN, Rodrigues LC. Quality of record linkage of large administrative databases in a middle income country: still births and dengue during pregnancy in Brazil. BMC Med Inform Decis Mak. 2017;17(1):108.

27. Boyd JH, Guiver T, Randall SM, Ferrante AM, Semmens JB, Anderson P, Dickinson T. A Simple Sampling Method for Estimating the Accuracy of Large Scale Record Linkage Projects. Methods of Information in Medicine. 2016; 55, 276-283.

The research project *Constructing the entire care pathway of bowel cancer patients undergoing emergency surgery to improve their quality of care* is funded by NIHR Health Services and Delivery Research Programme

28. Zakrison TL, Austin PC, McCredie VA. A systematic review of propensity score methods in the acute care surgery literature: avoiding the pitfalls and proposing a set of reporting guidelines. Eur J Trauma Emerg Surg. 2017 Mar 24. [Epub ahead of print]

29. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. Multivariate Behav Res. 2011;46(3):399-424.

30. The NCEPOD Classification of Intervention. National Confidential Enquiry into Patient Outcome and Death. http://www.ncepod.org.uk/classification.html (Accessed 10 September 2017).

31. NELA Technical Document: Development of the Risk Adjustment Model. Royal College of Anaesthetists, 2016 http://www.nela.org.uk/reports (Accessed 10 September 2017).

32. Walker K, Finan PJ, van der Meulen JH. Model for risk adjustment of postoperative mortality in patients with colorectal cancer. Br J Surg. 2015;102(3):269-80.

33. Leyrat C, Seaman SR, White IR, Douglas I, Smeeth L, Kim J, Resche-Rigon M, Carpenter JR, Williamson EK. Propensity score analysis with partially observed covariates: How should multiple imputation be used? Stat Methods Med Research, 2017, doi: 10.1177/0962280217713032. [Epub ahead of print]

34. Elze MC, Gregson J, Baber U, Williamson E, Sartori S, Mehran R, Nichols M, Stone GW, Pocock SJ. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. J Am Coll Cardiol. 2017;69(3):345-357.