POPULATING SERVICE DELIVERY MODELS USING OBSERVATIONAL DATA: CASE STUDY ON ENDOSCOPY PROVISION FOR ACUTE UPPER GASTROINTESTINAL BLEEDING

REPORT FOR THE GUIDELINES TECHNICAL SUPPORT UNIT

23rd July 2013

Sarah Davis[1], Mónica Hernández[1], Allan Wailoo[1], Mark Strong[1]

[1]School of Health and Related Research, University of Sheffield

ScHARR, University of Sheffield, Regent Court, 30 Regent Street
Sheffield, S1 4DA

Acknowledgements

The authors gratefully acknowledge the contributions of Kel Palmer, David Patch, Stephen Atkinson and Mark Donnelly, who provided clinical expertise. The views, and any errors or omissions, expressed in this document are of the authors only.

EXECUTIVE SUMMARY

Introduction

Acute upper gastrointestinal bleeding (AUGIB) is a common medical emergency. In patients presenting with AUGIB, endoscopy provides diagnostic information and also facilitates the delivery of a range of endoscopic therapies. The optimal timing of endoscopy varies according to the severity of bleeding with some patients requiring very urgent endoscopy and others who can safely wait several days. Endoscopy provision varies across the UK with the main differences being in the provision of daily lists versus weekday lists and the provision of a formal out-of-hours on-call rota versus informal out-of-hours endoscopy provision.

The 2012 National Institute for Health and Care Excellence (NICE) Clinical Guideline on, 'Acute upper gastrointestinal bleeding: Management', made recommendations on the optimal timing of endoscopy and gave some guidance on how services might be arranged to provide timely endoscopy. One of the difficulties in making evidence based recommendations on service delivery issues is that there is usually a lack of randomised controlled studies comparing the alternative service delivery models. In this situation it becomes necessary to rely on observational data to determine the impact of alternative service delivery models on the outcomes of interest. In the 2012 NICE Clinical Guideline, the recommendations on endoscopy provision were informed by a statistical analysis of an observational dataset which was used to estimate the relationship between endoscopy timing and mortality and length of stay. This was then used to populate an economic model comparing alternative service delivery models for endoscopy provision.

Our aim was to use the clinical question on the optimal timing of endoscopy in AUGIB from the 2012 NICE guideline as a case study to examine the methodological challenges of addressing service delivery questions within clinical guidelines. In particular we aimed to;

1. consider the data required to populate a cost-effectiveness model comparing different service delivery models for endoscopy in AUGIB
2. consider the observational data that might be collected to inform that model within the timeframe of a typical guideline, with particular focus in this case on the data required to estimate the relationship between timing of endoscopy and the outcomes of mortality and length of stay

3. consider how that new observational data set might be analysed to give an unbiased estimate of the relationship between timing of endoscopy and mortality and length of stay

Main findings

The following points summarise the main findings which were considered to be generalisable to other service delivery questions within clinical guidelines.

When identifying the patient outcomes to be included within the economic model it is important to determine the degree of precision needed to reflect important variations in costs and QALYs benefits between the strategies. For example, in the case of AUGIB it is important to consider whether discharge time needs to be measured to the nearest minute or the nearest day in order to capture the true opportunity cost of delayed discharge.

It is also important that the model captures all the important differences in costs and QALYs between the service delivery models being compared whether they fall on the target population or the wider population. In this case, broadening access to endoscopy may also reduce time to endoscopy in patients requiring endoscopy for reasons other than AUGIB.

When estimating the cost-effectiveness of alternative service delivery models it is important to recognise that different service delivery models may be optimal in different regions. There is therefore a need to consider local variations within the model. This may include variation in;

- the numbers and clinical characteristics of patients presenting in different regions
- the organisational environment such as on-site access to other relevant services
- the feasibility of collaborative arrangements which may be dependent on the distance to nearby hospitals.

When collecting additional data to inform a cost-effectiveness model, it is important to be careful to define the included population in such a way as to minimise the inclusion of patients unaffected by the choice between different service delivery models. In this example, when estimating the relationship between time to endoscopy and outcomes, it would be useful to identify and exclude patients admitted following an outpatient endoscopy.

Observational data can be used to measure the effect of different service delivery models on patient and organisational outcomes either directly by comparing outcomes according to the service delivery arrangement at that site or indirectly through an intermediate variable such as time to endoscopy. The latter is a more flexible approach that may allow a variety of existing and non-existing service delivery models to be explored. However, a lack of variability in current service provision may make it difficult to use current practice as a natural experiment. In this case, there is a lack of variability in actual access to urgent out-of-hours (OOH) endoscopy for unstable patients despite there being no formal OOH provision in around half of hospitals included in the 2007 audit. This makes it unlikely that any future collection and analysis of observational data would be able to provide a completely unbiased estimate of the true survival benefit attributable to providing urgent OOH endoscopy to unstable patients.

Several statistical techniques exist to deal with confounding when estimating causal effects from an observational data set. The three main techniques discussed here are regression, propensity scores matching and instrumental variables. All of the techniques rely on careful identification of potential confounding factors and the relationships between these factors and either treatment allocation or outcomes. When addressing service delivery questions, it is important to consider both patient level and organisational level confounders. Therefore when designing observational studies to answer service delivery questions it is important to seek expert advice from those with an in-depth knowledge of the clinical area and how current service provision operates, to ensure that what is being estimated using the various statistical techniques is clinically appropriate.

Given that none of the statistical approaches described above can guarantee to remove all confounding influences, it would be wise to compare the results achieved using several different techniques. It would be particularly useful to compare the results achieved through an instrumental variables approach with those achieved through either regression or propensity score matching as these make different assumptions regarding the presence of unmeasured confounders.

If potential instrumental variables are only strong predictors of treatment allocation within a subset of the population it may be possible to use different instruments for different groups within the population to provide an estimate of the ‘local’ average treatment effect. It is local

in the sense that it is identified through those for which the instrumental variable induces a change in the treatment. In this case, the two strongest candidates for instrumental variables are the provision of daily lists in patients presenting at the weekend and provision of a formal OOH endoscopy rota for high risk patients presenting OOH, but neither of these instruments are likely to be strong predictors of time to endoscopy in the whole AUGIB population.

There is a balance to be struck when collecting and analysing observational data with regards to the number of confounding factors that should be included in the data collection exercise and the feasibility of conducting the study. Limiting the data collection to routine data that can be obtained retrospectively will increase the chances of achieving high completion rates. However, this may limit the ability of the analysis to capture subtle differences in outcomes between different service delivery models, such as a failure to capture the resource implications of differences between morning and evening discharges if the analysis is restricted to collecting date rather than time of discharge.

The feasibility of further data collection to inform guideline development should be considered from the scoping stage. In some cases additional data collection may not be feasible within the usual guideline development time-frame and in such cases a judgement would need to be made between the value of the additional data and the cost of extending the timescale for guideline development.

Conclusions specific to the example of endoscopy provision in AUGIB

Whilst we anticipate that the lack of variability in time to endoscopy for unstable patients would make it difficult to estimate the survival benefit of timely endoscopy in this group, we are optimistic that a causal inference could be made regarding the benefits of timely endoscopy for stable patients. The estimates which informed the economic evaluation within the published NICE guideline may be improved upon if more complete data can be obtained on relevant confounding factors and data is collected on the provision of daily lists for use as an instrumental variable.

CONTENTS

TABLES AND FIGURES

ABBREVIATIONS AND DEFINITIONS

A&E	Accident and Emergency
AUGIB	Acute upper gastrointestinal bleeding
BP	Blood pressure
BSG/NHSBT	British Society of Gastroenterology / NHS Blood and Transplant
CG141	NICE Clinical Guideline, 'Acute upper gastrointestinal bleeding: Management
GP	General Practitioner
NICE	National Institute for Health and Care Excellence
NHS	National Health Service
OOH	Out-of-hours
QALY	Quality adjusted life year

1. INTRODUCTION

1.1. BACKGROUND

Acute upper gastrointestinal bleeding (AUGIB) is a common medical emergency with an incidence of 103/100,000 per annum[1]. The principle diagnostic test for patients with AUGIB is endoscopy. Endoscopy not only provides diagnostic and prognostic information, it also facilitates the delivery of a range of endoscopic therapies[2].

The optimal timing of endoscopy varies according to the severity of bleeding. Patients with active haematemesis / melaena may benefit from urgent diagnostic and therapeutic endoscopy done as soon as possible after resuscitation and in these cases endoscopic therapy may be life saving. Conversely those with relatively trivial bleeding may safely wait several days for endoscopy.

Endoscopy provision varies across the UK with only 52% of hospitals reporting access to a formal out-of-hours (OOH) endoscopy rota in 2007[3]. The 2012 National Institute for Health and Care Excellence (NICE) Clinical Guideline, 'Acute upper gastrointestinal bleeding: Management,' (CG141) recommends that endoscopy should be offered to unstable patients with severe AUGIB immediately after resuscitation, and within 24 hours in all other patients with AUGIB[2]. It also recommends that units with more than 330 cases of AUGIB per year should offer daily endoscopy lists. The latter recommendation was informed by a cost-effectiveness analysis which assessed the costs and benefits of various service delivery models for providing endoscopy and found that daily lists were likely to be cost-effective provided that this facilitated endoscopy within 24 hours and provided the unit expected at least 330 cases of AUGIB per annum.

In the cost-effectiveness analysis which informed CG141, the costs of implementing alternative service models to achieve endoscopy within specific time frames were offset in the model against savings generated by early discharge and reduced length of stay. The model also estimated the impact of the different service models on quality adjusted life years gained (QALYs) over a 28 day period by estimating the impact of prompt endoscopy on mortality rates and the impact of early discharge on health utility. The different service delivery models compared in the guideline economic analysis were assumed to facilitate endoscopy within a specific time period (e.g daily lists facilitate endoscopy within 24 hours), meaning that the cost-effectiveness results were then driven by the relationships between time to endoscopy and time to discharge and death. These relationships were informed by a

statistical analysis of observational data from an audit conducted by the British Society of Gastroenterology / National Health Service Blood and Transplant (BSG/NHSBT) in 2007 [3]. This analysis was recognized by the guideline authors as having 'potentially serious limitations due to the possibility of confounding factors which were not controlled'. Despite this limitation, it is recognized that some estimate of the cost-effectiveness of different service delivery models is required and therefore some estimates of the relationships between endoscopy timing and economic outcomes such as mortality and length of stay are necessary and these are likely to be based on analysis of observational data.

The NICE guideline does not make prescriptive recommendations regarding how smaller units seeing fewer than 330 cases per annum should arrange their services to deliver endoscopy within 24 hours, although it does discuss the potential for smaller units to offer safe transfer to larger sites or peripatetic endoscopy to achieve timely endoscopy in collaboration with other units. One example of such a collaborative service, which was already in place at the time the NICE Guideline, is reported by Shokouhi *et al*[4]. The East and North Hertfordshire National Health Service (NHS) Trust operates a collaborative weekend service across two sites. Patients from the Queen Elisabeth II Hospital are transferred, after being medically stabilized, to the Lister Hospital which offers weekend endoscopy lists. Unstable patients requiring urgent endoscopy outside of the weekend list receive OOH endoscopy without transfer between sites. Shokouhi *et al.* claim that their collaborative service has multiple benefits but the cost-effectiveness of the service is not explicitly addressed. An economic model which explicitly examines the costs and benefits of such collaborative services would be useful in determining whether they are a cost-effective alternative to either daily lists or weekday lists in units seeing fewer patients.

Our aim was to use the clinical question on the optimal timing of endoscopy in AUGIB from the 2012 NICE guideline as a case study to examine the methodological challenges of addressing service delivery questions within clinical guidelines. In particular we aimed to;

1. consider the data required to populate a cost-effectiveness model comparing different service delivery models for endoscopy in AUGIB
2. consider the observational data that might be collected to inform that model within the timeframe of a typical guideline, with particular focus in this case on the data required

to estimate the relationship between timing of endoscopy and the outcomes of mortality and length of stay

3. consider how that new observational data set might be analysed to give an unbiased estimate of the relationship between timing of endoscopy and mortality and length of stay

2. MODELLING THE COST-EFFECTIVENESS OF ENDOSCOPY SERVICES

2.1. ALTERNATIVE SERVICE DELIVERY MODELS

There is currently variation across hospital sites in the provision of endoscopy for AUGIB outside of the normal working week (8am to 5pm, Monday to Friday). After discussion with clinical experts we identified the following components of endoscopy services as being potential options for inclusion in an economic evaluation.

- Daily versus weekday endoscopy lists
- Formal on-call endoscopy rota versus informal OOH endoscopy for unstable patients
- Protected AUGIB emergency slots in daily endoscopy lists
- Discharge of low risk patients for outpatient endoscopy

As stated earlier, in the 2007 audit it was found that only 52% of hospitals had a consultant led OOH endoscopy rota. However, even in those hospitals without an OOH rota, 13% of patients received an OOH endoscopy, which suggest that a substantial number of endoscopies are being done through ad-hoc or informal OOH arrangements based on the good-will of the staff involved[3]. These informal OOH arrangements are understandable given the life-threatening nature of AUGIB in some patients and the lack of universal formal OOH arrangements. However, this ad-hoc provision makes it difficult to assess the benefit of formal OOH provision against a comparator of no OOH provision which is the official policy in those sites without formal OOH provision. A comparison against informal OOH provision is possible, but may be difficult to evaluate due to a lack of information on the monetary or opportunity cost of such arrangements.

The audit found that 58% of hospitals had a dedicated endoscopy slot within their weekday lists for patients with AUGIB who did not require urgent OOH endoscopy[3]. These slots allow

patients admitted overnight to receive prompt endoscopy without disruption to the routine endoscopy list but they are clearly not a universal feature of endoscopy provision.

The audit found that 21% of patients were discharged with an outpatient referral for endoscopy[5] suggesting that some hospitals are able to book outpatient endoscopy appointments within a short enough time frame to allow low risk patients to be discharged without endoscopy. However, our clinical advisors did not think that all hospitals had access to urgent outpatient endoscopy, resulting in some low risk patients waiting for inpatient endoscopy.

Some hospitals operate weekend endoscopy lists[4;6], but no data is available on this from the 2007 audit as questions on weekend lists were not included in their survey of organisational aspects of care[3]. In the 2007 audit, endoscopies occurring over the weekend were categorised as OOH, but it would be important in the economic model to distinguish between endoscopies occurring at the weekend through weekend lists and those occurring through either formal or informal on-call provisions.

As described earlier, some hospitals operate collaboratively to provide services across a number of sites[4] and a variety of collaborative options may be practical depending on the local situation. These may include transfer between sites for weekend endoscopy, or peripatetic endoscopy in which the on-call team travel to the patient. Ideally any economic model developed to assess endoscopy service delivery models should be flexible enough to consider collaborative services across multiple sites in addition to single site services. However, it was considered that any collaborative arrangement is likely to be focused on providing one of the service options above, with the additional dimension of considering whether it is more cost-effective to do so collaboratively or as a single site.

2.2. HEALTH ECONOMIC OUTCOMES

In any cost-effectiveness model, the primary outcomes are the mean incremental costs and QALYs between the strategies being compared. In this case, the costs are likely to be largely determined by the staffing costs for the endoscopy service and the length of hospital stay. The cost per patient of providing weekend lists or a formal OOH rota is likely to depend on the number of cases using that service per annum and therefore may vary according to local

circumstances. For the economic model used to inform CG141, data were available on the variation in presentations across sites and the timing of presentations[2]. This was combined with information on unit costs for consultants and nurses working unsocial hours to calculate the additional costs of providing weekend lists or extended access to endoscopy through on-call rotas[2]. Determining costs for on-call provision may require some assumptions to be made regarding the time taken to complete an on-call endoscopy and any impact on activity the following day. This may be particularly difficult when assessing the opportunity cost of informal OOH provision. When modelling a collaborative service it would be necessary to capture all the relevant costs across all the collaborating sites including costs for patient transfer or travel time for staff working across multiple sites. If the service delivery model being evaluated included a policy of providing outpatient endoscopy referrals for low risk patients, the cost of outpatient endoscopy would need to be included within the analysis.

The actual treatment received for AUGIB following endoscopy is unlikely to vary significantly between different endoscopy strategies, but some differences in costs may be introduced if one strategy results in more patients dying prior to endoscopy or leaving against medical advice prior to endoscopy so these outcomes would need to be captured in the model, unless their impact is negligible. If the AUGIB treatment is merely delayed due to endoscopy delay, then the long-term health outcomes are unlikely to differ significantly according to endoscopy service provision. However, if endoscopy delay results in a longer length of stay, then any additional time spent in hospital is likely to be at a lower quality of life than time spent at home after discharge, so there may be some QALY gain associated with early discharge. If endoscopic therapy provides symptomatic benefit to the patient there may also be a difference in health utility pre- versus post-endoscopy. There may also be some survival benefit attributable to prompt endoscopy in unstable patients, although analysis of the 2007 audit was unable to demonstrate a statistically significant mortality benefit associated with the provision of a formal OOH on-call rota to allow prompt endoscopy in such patients. The risk-adjusted post-endoscopy in-hospital mortality was 21% higher for patients in hospitals without formal OOH endoscopy rotas (1.21, 95% CI 0.96 to 1.51) but this was of borderline statistical significance ($p=0.10$)[3]. The lack of statistical significance may simply have been due to small numbers. However, it may also have been due to OOH endoscopy being provided through ad-hoc arrangements despite a lack of formal provision, as this would reduce the variability in time to endoscopy between hospitals with and without formal OOH endoscopy provision. Based on the analysis conducted to inform CG141, the main trade offs

are likely to be between the cost of endoscopy provision and the 'hotel' costs of extended hospital stay[2]. The key inputs to the economic evaluation are therefore likely to be;

- Cost of endoscopy service
- Length of hospital stay (pre and post endoscopy)
- Inpatient costs per day (pre and post-endoscopy)
- Inpatient health utility pre and post-endoscopy
- Health utility post discharge
- Inpatient mortality (pre and post endoscopy)
- Post-discharge mortality

There may also be significant benefits to other non-AUGIB patients associated with broadening access to endoscopy. For example, early endoscopy may prevent aspiration pneumonia in patients with dysphagia due to oesophageal stricture and may reduce time to surgery in patients with severe colitis. Increasing the formal OOH provision for urgent AUGIB cases may also reduce the disruption associated with having to slot urgent AUGIB cases into routine endoscopy lists. Conversely the provision of a formal OOH endoscopy rota may reduce the availability of those staff involved to participate in other on-call rotas. It may not be feasible to include these broader implications explicitly in the model, but the relative size and direction of any cost or QALY effects for other patient groups should be considered before they are excluded from the model.

2.3. MODEL STRUCTURE

A model evaluating alternative endoscopy services should be able to capture key features of the care pathway that may influence either time to endoscopy or patient outcomes. The NICE Care Pathway for AUGIB was used as a starting point for a discussion with our clinical advisors regarding the factors influencing time to endoscopy and the factors influencing time to discharge or mortality risk. From this, the care pathway shown in Figure 1 was developed.

Various patient related factors were identified and these are shown in the 'Patient attributes' box in Figure 1. The most obvious patient related factors are those factors already incorporated in one of the recognised risk scoring tools such as the Blatchford score[7], which is useful in identifying patients requiring intervention, and the Rockall score[8], which is useful in predicting those patients at higher risk of death. These factors included in these risk scores

tools are detailed in Table 1. Some additional patient related factors were also identified, such as route of referral with patients referred via accident and emergency (A&E) being more likely to have preliminary risk assessment and blood tests results available than those referred directly by their general practitioner (GP). Time since last meal was also identified as a potential factor which could delay endoscopy in some patients.

It is recognised that patients who experience AUGIB during an admission for another reason are at increased risk of death[5]. To some extent this will reflect differences in comorbidities and other patient attributes between those admitted for AUGIB and those who experience AUGIB following admission for another reason. However, there was some concern among our clinical advisors that there may be a delay in recognising AUGIB and initiating management if patients experience a bleed in an inpatient environment where AUGIB is uncommon. This delay may in itself contribute to the increased risk of mortality. Therefore AUGIB occurring in established inpatients would be an important factor to capture as it could determine both time to endoscopy and mortality.

The influence of time of presentation on time to endoscopy will be dependent on endoscopy service provision, but time of presentation may also be a proxy for other relevant factors which might influence length of stay or mortality. For example, it has been found that patients presenting at the weekend with AUGIB are more likely to be haemodynamically unstable and more likely to present with haematemesis[9]. When an AUGIB is experienced by an established in-patient, care should be taken to accurately record the time of presentation with AUGIB as this needs to be distinguished from the time of admission.

The NICE Care Pathway for AUGIB assumes that all patients progress from presentation to endoscopy and then to treatment, with the exception of low risk patients with a pre-endoscopy Blatchford score of zero who may be considered for early discharge. However, it was noted that some patients included in the 2007 audit did not follow the standard care pathway and were discharged without endoscopy. For example 125/6750 patients included in the 2007 audit were 'specifically categorised for no active treatment or investigation when they first presented with AUGIB'[10] and would therefore be unlikely to receive an endoscopy regardless of whether it was available. There were also a number (96/6750) of patients in the audit who took their own discharge against medical advice without endoscopy[10]. Another route to discharge without endoscopy for low risk patients is via referral for outpatient

endoscopy. It was acknowledged that these categories of patients who may be discharged without endoscopy should be properly accounted for in any future data collection, and within the cost-effectiveness model. Failure to do so may overestimate the benefits of providing broader access to endoscopy as benefits may be estimated in patients who are unlikely to receive endoscopy in practice. For example, in the analysis conducted to inform CG141, 99.5% of patients presenting received endoscopy under the continuous access strategy, but only 67.7% received endoscopy under the weekday service. Some of this difference may be explained by the fact that endoscopy rates for the weekday service were based on the audit data itself, whilst endoscopy rates for the continuous access service were based on assumptions which may not have taken into account those patients who present with AUGIB who would be unlikely to receive endoscopy under any service delivery model.

Table 1 Factors included in the Blatchford and Rockall risk scores

Risk score	Factors included
Blatchford [7]	• Systolic blood pressure • Pulse • Haemglobin • Blood urea • Presentation with melaena • Presentation with syncope • Comorbidites (hepatic disease, cardiac failure)
Rockall [8]	• Age • Shock (based on pulse and systolic blood pressure) • Comorbidity (Chronic failure, ischaemic heart disease, any major comorbidity, renal failure, liver disease, metastatic cancer) • Diagnosis* (Mallory-Weiss tear, no lesion identified, gastrointestinal malignancy) • Stigmata of recent bleeding* (blood in the upper gastrointestinal tract, adherent clot, visible or spurting vessel) *only known post endoscopy

Several post endoscopy factors were identified as being potentially important determinants of time to discharge such as AUGIB treatments (other than endoscopic therapy), re-bleeding events, non-AUGIB healthcare needs (e.g treatment for comorbidities) and social care needs. It was also noted that some parts of the care pathway make take longer depending on whether the activity occurs during normal working hours or OOH. For example, a decision regarding whether someone is fit to be discharged may be delayed over the weekend until an appropriate staff member is available to assess the patient. Discharge may also be delayed until appropriate discharge medications can be provided by pharmacy services. This has been denoted on Figure 1 by the term, 'seven day working'. Any cost-effectiveness model would need to account for variations in the time taken to complete certain activities over the working week.

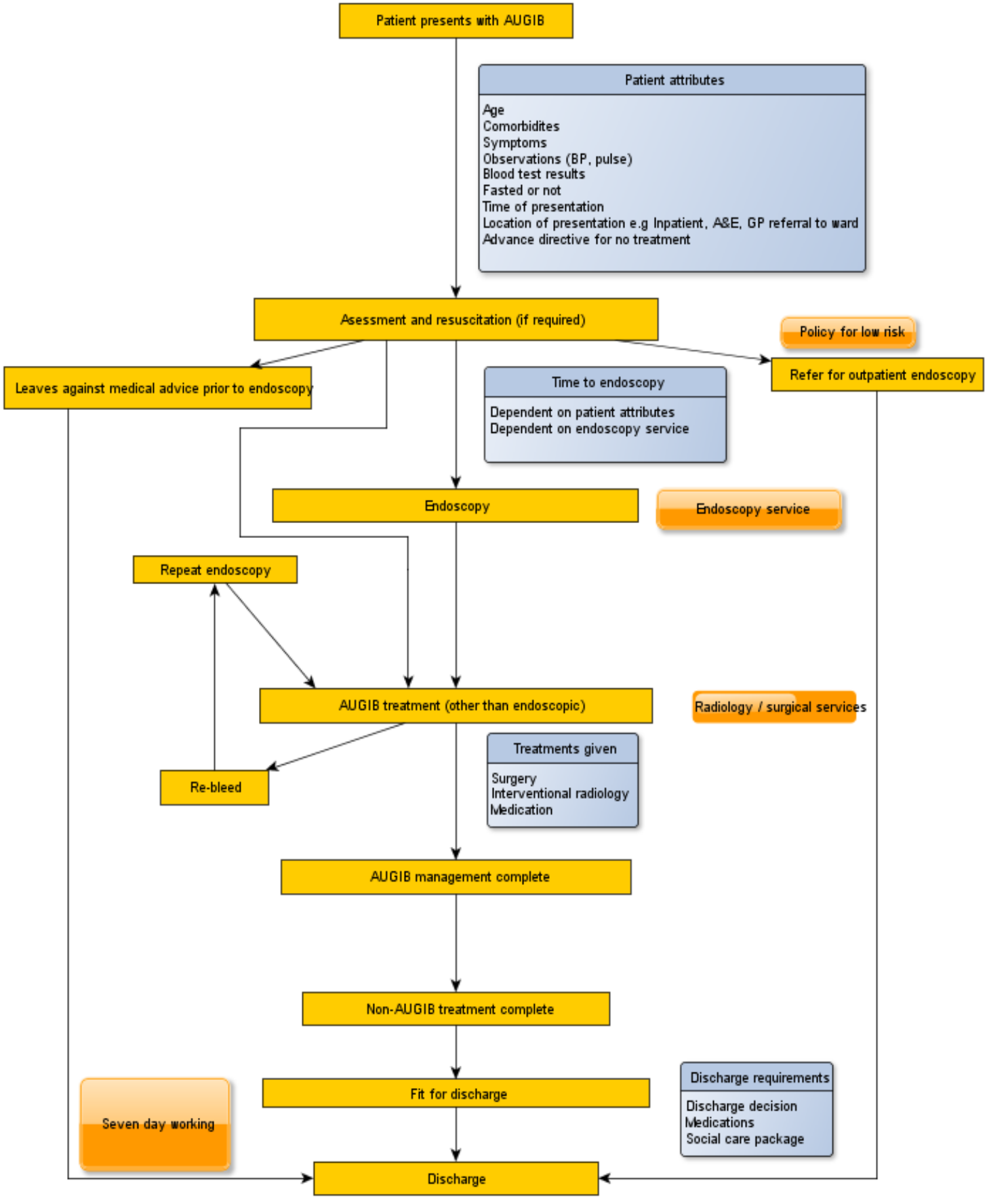
Ideally, one would want to model the individual components of the care pathway to allow a more detailed analysis of the impact of different endoscopy service models. For example, if it were possible to distinguish between the duration of AUGIB treatment and the overall length of hospital stay, it may be possible to measure the impact of reducing time to endoscopy even if the overall length of hospital stay is unaffected due to it being largely determined by social care needs and treatment durations for comorbidities. However, this is unlikely to be possible if the analysis is restricted to routine data as this is unlikely to distinguish the duration of stay for different diagnoses or comorbidities. Therefore, the model may be restricted to using overall length of hospital stay.

It is also important to consider how exactly each of the outcomes needs to be measured to be used meaningfully within the model. For example, the economic model which informed CG141, used a one hour Markov time cycle. Such a detailed analysis was considered necessary to assess the impact of their 'continuous access' strategy which was assumed to provide access to endoscopy within 4 hours[2]. However the 2007 BSG/NHSBT audit on which this model was based recorded the date of discharge and death but not the time of discharge and death meaning that times had to be imputed using assumptions. This imputed data on time of discharge and death was then used to calculate the hourly probability of discharge and death such that the costs attributable to length of stay were allowed to vary in the model according to the number of hours spent as an inpatient. Collecting data on the precise time of discharge is unlikely to be feasible based on the data routinely recorded in patient notes. Whilst some hospital management systems record time of discharge, this is not always easy

to link with individual patient notes. In reality, discharging a patient one hour earlier in the day may not represent a true cost saving or productivity gain to the NHS as the bed may not be filled by another patient until the following day. It may be better for the model to attribute costs for inpatient stays according to whole or half day units rather than hourly. In which case, some assumptions could be used to infer the time of discharge based on the date of discharge and the time of endoscopy. For example it could be assumed in the model that a whole day of hospital cost is accrued for patients discharged on the same day as their endoscopy and for each day prior to the day of endoscopy, whilst a half day of cost is accrued for patients discharged on any day other than the day of endoscopy. This would allow for there to be some cost saving attributable to rapid endoscopy allowing discharge on the same day of endoscopy.

Ideally the model should be able to capture variations in demand or resources between hospitals or regions allowing the decision analysis to be tailored to local needs. For example, the numbers presenting per annum or the case mix of patients presenting may vary across hospitals or regions. The most cost-effective solution for a large acute hospital in a densely populated area may be 24/7 endoscopy provided alongside a 24/7 interventional radiology service. Conversely in an area with several small hospitals, serving a less densely populated area, the most cost-effective solution may involve a collaborative on-call service covering several hospitals. Therefore a model which is capable of assessing service delivery options across single or multiple sites according to the needs of patients presenting in a particular region would be useful.

Figure 1 AUGIB care pathway showing factors influencing time to endoscopy, length of stay and mortality



3. ESTIMATING THE IMPACT OF ALTERNATIVE ENDOSCOPY SERVICE DELIVERY MODELS ON COSTS AND PATIENT OUTCOMES

A cost-effectiveness model compares the costs and health outcomes (survival and quality of life) of alternative health care interventions or policies. It does so by combining data from many different sources to estimate the overall impact of alternative policies over the full time period where they may be expected to differ. To estimate the cost-effectiveness of a particular intervention or policy some information is needed on the impact of that intervention or policy on patient outcomes. When evaluating the cost-effectiveness of drug therapies, inter alia, it is usual to base the economic model on data from randomised controlled trials which provide an unbiased estimate of the relative effect of the different treatments by ensuring that the patient groups receiving each treatment are similar. However it is not usually feasible to conduct a randomised controlled trial when evaluating alternative service delivery models which are much more complex than an individual drug or procedure and can only be implemented at a hospital level. In the absence of trial data we would need to try to infer the impact of different service delivery models on patient outcomes from an observational study.

3.1. DIRECT EVALUATION OF POLICIES CURRENTLY IN PLACE

Given that that there is currently some variability in service provision across the NHS, one option would be to try to measure the variation in outcomes according to different types of service provision directly. An example of this is the analysis conducted by Hearnshaw *et al*. which used data from the 2007 BSG/NHSBT audit to compare patient outcomes, such as in-hospital mortality, according to whether or not the hospital had a formal OOH endoscopy on-call rota[3]. This direct approach has several limitations. Firstly, the policies to be compared would have to be defined and then the service provided in different hospitals would need to be categorised into one of those policies. If the categories are too broad, then important variation between the different hospital policies may be missed, but if too many different policies are specified, then there may not be sufficient hospitals in each category to make a worthwhile comparison. The approach taken by Hearnshaw *et al*. was to separate hospitals into two categories according to whether or not there was a formal OOH rota[3]. However, they also collected other information regarding OOH endoscopy provision such as the availability of endoscopy nurses which could have been used to specify more categories of OOH

provision. Secondly there may be interactions between different aspects of endoscopy policy. For example, the impact of not having a formal OOH on-call rota is likely to be smaller in hospitals with weekend lists as routine endoscopy services are available for a greater proportion of the week. So it may not be possible to consider OOH on-call service provision separately from other aspects of service provision. This could be addressed by carefully categorising the provision across hospitals using consistent definitions of terms such as OOH but this categorisation should take into account any possible interactions between different aspects of endoscopy policy. Another disadvantage of measuring the impact of differing policies directly is that it doesn't allow for any change in hospital policy during the period of data collection. Nor does it allow any policy to be explored which isn't currently implemented within the NHS.

3.2. INDIRECT EVALUATION OF POLICIES VIA AN INTERMEDIATE OUTCOME

An alternative approach is to measure the impact of time to endoscopy on patient outcomes and to estimate the impact of different policies on time to endoscopy. This was the approach taken in the economic model which informed CG141. This is a more flexible approach, as once the relationship between time to endoscopy and outcomes is known, then a variety of existing and non-existing policies for providing endoscopy can be explored. In such a model, it would also be possible to examine whether the optimal policy differs according to differing local circumstances and even whether it would be robust to any anticipated changes in health care needs. One disadvantage of this indirect approach is that it would then be necessary to somehow estimate the relationship between different endoscopy policies and time to endoscopy. This could be based on actual data from hospitals employing the particular service delivery models being compared. For example, in the model which informed CG141, time to endoscopy in the absence of OOH provision was based on audit data from hospitals reporting no OOH service. In the absence of data on specific policies, assumptions could be made regarding their impact on time to endoscopy. For example, in the analysis which informed CG141, it was assumed that daily endoscopy lists would allow endoscopy within 24 hours even though data were not collected on the provision of daily lists in the 2007 audit.

3.3. THE PROBLEM OF CONFOUNDING IN OBSERVATIONAL DATA SETS

The main difficulty in using either approach is adjusting for the fact that the patients being compared, or the care they receive, may differ in ways other than the one we want to use to explain variation in the outcomes. So for example, in the first option described above in which different endoscopy services are compared directly, a difference in case mix between the hospitals with and without weekend endoscopy lists may lead to the relationship between weekend endoscopy and time to discharge being either over or under estimated. Or in the second option, where time to endoscopy is used as an intermediate outcome, if patients with signs of shock have a shorter time to endoscopy and shock predicts mortality then this may wrongly suggest that earlier endoscopy results in a higher mortality. This is known as confounding and it is possible to adjust for confounding factors provided they are identified and measured within the observational data set being used to determine the relationship of interest.

3.4. IDENTIFYING POTENTIAL CONFOUNDING FACTORS

We asked our clinical advisors to identify factors which could influence time to endoscopy, time to discharge and time to death. We also asked them to say whether these are routinely recorded in patient notes and could therefore be retrospectively collected. Figure 2 shows the factors identified and their expected relationships. They can be broadly categorised as patient factors and organisational factors. A few of the factors identified could be dependent both on the patient presenting and the organisation to which they present. For example, the specific treatment given may be dependent on patient factors, such as the diagnosis of the cause of the bleeding, but the treatment given may vary according to whether a particular intervention is available at that hospital. Another example is that severity of bleeding may be a weaker predictor of the likelihood of receiving an OOH endoscopy in hospitals which have weekend lists compared to those with only weekday lists.

Time of presentation is clearly a very important factor that is likely to interact with other confounding factors. As already discussed, patients presenting at different times may have different clinical characteristics[9]. The impact of time of presentation on time to endoscopy will depend on the particular endoscopy service at the hospital to which the patient presents, which in turn will determine whether the endoscopy is provided within a routine list or via an on-call service. Receiving an endoscopy via the on-call service might be a confounding factor

in its own right if the on-call team has a different skill mix to the team providing routine endoscopy. For example, access to endoscopy nurses and anesthetists may vary according to whether or not the procedure is conducted during a routine list or via the on-call service. Time of presentation might also affect the treatment given depending on whether all treatments are available 24/7 at that particular hospital. For example, interventional radiology might not be available 24/7 at all hospital sites.

The patient factors at presentation (age, shock, blood pressure (BP), pulse, comorbidities, melaena, syncope, haemogobin, blood urea, time of presentation, route of presentation) and endoscopy findings were all considered to be potential confounding factors and easily accessible from patient notes, with the exception being time since last meal which is not always recorded. Although this may not be a problem if we are only concerned with time since last meal in those patients where it has been an important factor in delaying endoscopy, as it is likely to be recorded in such cases.

In terms of organizational factors, it should be possible to record whether a particular hospital had a formal OOH rota, whether they offered weekend lists, and whether they had protected AUGIB slots within their endoscopy lists, as such questions regarding endoscopy provision were included in the 2007 BSG/NHSBT audit. However, ascertaining whether a particular patient received their endoscopy via a routine list or via the on-call service would have to be inferred from information on the time of endoscopy and details regarding local service provision. It was noted in particular that it would be hard to ascertain if an individual patient used a protected AUGIB emergency slot within a list. It was also noted that access to urgent outpatient endoscopy, allowing low risk patients to be discharged with an outpatient referral, might be harder to ascertain but could be determined if the right question was targeted at the right individual.

We discussed with our clinical advisors whether any factors that might influence the quality of the endoscopy such as availability of endoscopy nurses and anesthetists and the seniority or expertise of the endoscopist should be recorded. It was agreed that it might be difficult to measure and record whether a particular endoscopist was competent in particular therapeutic procedures, but it might be possible to record factors such as whether the endoscopy was performed by a consultant or another staff member. Factors such as the availability of

anaesthetists and endoscopy nurses were considered to have some impact on the quality of care, but the impact would be small and therefore these might not be worthwhile recording.

Intermediate patient outcomes such as re-bleeds and the subsequent need for a repeat endoscopy were identified as being likely to have a significant impact on hospital length of stay. As re-bleeds occur after endoscopy, it is possible that the likelihood of experiencing a re-bleed will be influenced by the time to endoscopy. However it is also possible that the likelihood of re-bleeding is a proxy for some other unmeasured factor that varies between patients prior to endoscopy. It would therefore be important to measure this intermediate patient outcome as it may be able to indicate whether important confounders have been left out when conducting an analysis using either regression or propensity score matching to adjust for confounding (see chapter 4 for a discussion of these statistical methods).

Transfusion of blood products may occur prior to endoscopy and there is some evidence from the 2007 BSG/NHSBT audit to suggest that transfusion may increase the risk of re-bleeding and mortality, although this may ultimately be due to differences in confounding factors between those receiving early transfusion and those receiving no transfusion[11]. Given this uncertainty it would be wise to include transfusion as a potential confounding factor in any future analysis.

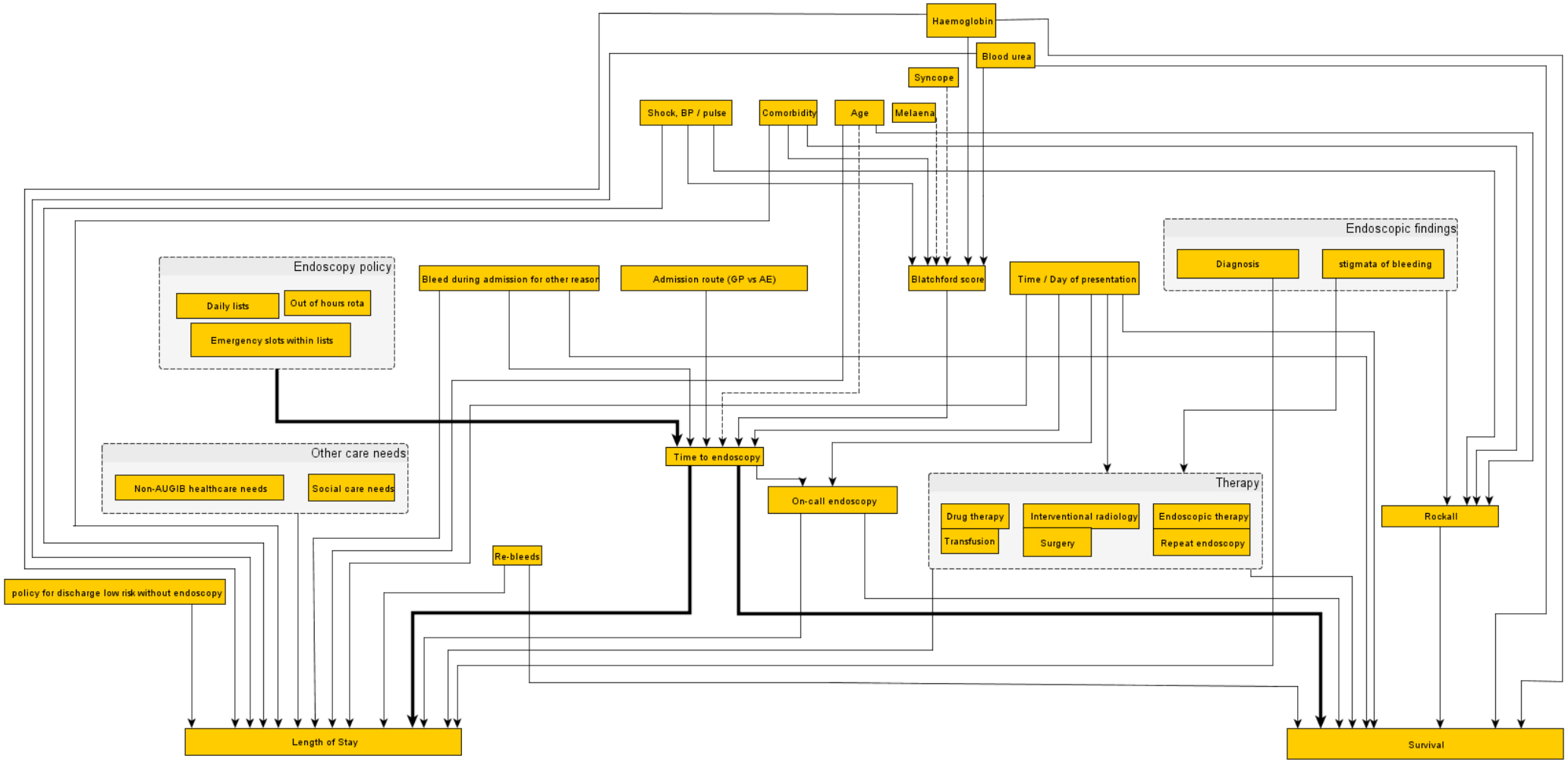
Other treatments for AUGIB such as surgery, interventional radiology and drug therapy, were recognized as having a potential impact on either time to discharge or mortality. These factors are only likely to act as confounding factors if there is some correlation between the treatments offered to an individual and their time to endoscopy. It was discussed whether factors such as OOH access to interventional radiology might be predictive of outcomes and may correlate with OOH access to endoscopy, but this factor was felt to be relevant to too few patients to be a significant predictor. Access to emergency surgery and blood products were not considered to vary significantly between hospitals.

Treatment for comorbidities and delays in discharge related to social care needs were identified as significant determinants of length of stay which might make it difficult to detect the impact of more efficient management of AUGIB. We discussed whether it might be possible to record when the patient's AUGIB management was completed or when they were

‘medically fit’ for discharge but waiting for social care needs to be addressed. However, it was felt that this would be difficult to record consistently using routinely collected data.

As length of stay may be significantly affected by whether hospitals have efficient procedures for weekend discharge, we discussed whether it was worthwhile trying to measure some factors which might influence this such as weekend access to pharmacy. However, this was considered to be too difficult to capture.

Figure 2 Potential confounding factors and the interaction between these and the relationship of interest (shown in bold lines)



4. STATISTICAL METHODS FOR ADJUSTING FOR CONFOUNDING FACTORS

The three main techniques discussed here are regression, propensity scores matching and instrumental variables. Both regression and propensity score matching work well as long as selection into treatment is due to measured covariates and there are no unobserved factors influencing treatment selection. One of the differences between the two is that for a regression we need to make some functional form assumptions which in some cases might be overly restrictive. Another more important difference is that they estimate slightly different parameters with regression estimating the average treatment effect for the population and propensity score matching estimating the average treatment effect for the treated group. Additional assumptions are needed to estimate the average treatment effect for the population when using a propensity score approach. If there are unobserved factors that influence treatment selection, then none of the above work and an instrumental variable needs to be used, although this also requires a functional form to be specified, which again may impose restrictive assumptions on the data.

4.1. REGRESSION

Regression can be used to estimate the difference in outcomes, in this case time to discharge and time to death, according to treatment allocation. In our example, treatment allocation could be the actual service provision e.g weekday versus weekend lists, or it could a categorisation of the time to endoscopy, e.g more or less than 24 hours. Regression can be used to adjust for known confounders. In regression it is assumed that the relationship between the outcome and the potential confounders is linear in parameters and that the slope of the regression line does not vary according to treatment allocation. So in our example, we would have to assume that the impact of haemoglobin on length of stay is consistent between patients receiving endoscopy within 24 hours and those receiving endoscopy after 24 hours, once other confounding factors which may differ between these groups have been adjusted for. It is possible to extend the model to deal with situations where the slopes are not equal between the groups with different treatment allocations by including interactions between baseline variables and treatment effects. However, if the treatment interacts with any of the

baseline measures, and those baseline-by-treatment interactions are not included in the model, then this will result in a biased estimate of the treatment effect.

It is also assumed that there are no unmeasured confounders. The chance of this is increased by including as many potential confounding factors as possible within the regression. Inclusion should be guided by knowledge of previous research in the area and theoretical considerations. Factors thought to influence treatment selection or outcome should not be excluded from the regression because they do not meet traditional tests of statistical significance. Confounding factors measured after treatment allocation may be problematic if these are potentially influenced by the treatment allocation. For example, in this case it would be necessary to consider whether factors such as endoscopy findings, or the need for particular AUGIB treatments might be influenced by time to endoscopy. When addressing service delivery questions such as this, it is important to consider both patient level and organisational level confounders for inclusion in the regression equation.

Another difficulty with regression is that it works better when the characteristics of those allocated to different treatments are similar in terms of the confounding factors. This is unlikely to be true if those confounding factors are strong predictors of treatment allocation. In our case, this becomes problematic if those patients having early endoscopy are very different from those having late endoscopy. This might be less of a problem if the hospital's endoscopy service provision is used as the treatment allocation variable rather than the time to endoscopy experienced by individual patients, although this results in a less flexible model framework as described earlier.

Cox regression models are commonly used to deal with confounding factors when the outcome of interest is the time to an event, in this case either time to discharge or time to death. The proportional hazards assumption for Cox regression requires that the relative hazard associated with a particular treatment or confounding factor is constant over time even if the hazard itself is varying. This methodology was rejected in the analysis conducted to inform the economic model for CG141 [2]. The reasons given were that that this method would assume:

- No interactions between factors we are adjusting for (i.e. age had the same effect in people with liver disease as people without)

- The effect of each factor within a Rockall score was the same regardless of:
  - o Follow up time
  - o Time to endoscopy

Instead a stratified analysis was conducted with the Rockall score used as the stratifying factor. This was chosen as it was considered to;

- represent a combination of valid risk factors influencing time to endoscopy
- be a validated predictor of mortality
- be a variable with good completion rates within the data set.

Given that CCG141 recommends that the Blatchford score is used for risk assessment prior to endoscopy[2], stratification by Blatchford score rather than Rockall score might be more appropriate in any future observational study as this is more likely to predict time to endoscopy.

4.2. PROPENSITY SCORE MATCHING

Propensity score matching allows a single score to be used to match patients who received different treatment allocations, but who are similar in other known confounders. The propensity score is the probability of receiving a particular treatment given the observed covariates (confounding factors). The covariates included in the propensity score should be those thought to predict treatment allocation. Often these are the same covariates as those that you would put in the regression but this does not need to be case. Patients who received different treatments can be matched based on their propensity score and the treatment effect can be estimated within the matched cohort. This method also requires the groups who received different treatment allocations to have substantial overlap in terms of their propensity scores. If there is poor overlap in propensity scores between the patients receiving different treatments, then this may lead to a substantial number of patients not being matched with an equivalent patient receiving a different treatment. This reduces the generalizability of the relationship estimated from the matched group. One benefit of propensity score matching over regression is that the overlap of the two groups is an explicit outcome of the matching process. In addition to being used to match patients, propensity scores can also be used in selection models (based on control functions) to adjust regressions according to the probability of a patient receiving a particular treatment. One difference between propensity

score matching and selection models is that matching assumes that, conditional on the covariates, the unobservables are independent of the treatment whereas in selection models there is a model linking the unobservables to the treatment selection.

When selecting covariates to include in the propensity score, again, it is generally better to err on the side of including more rather than less potential confounding variables[12]. Omitting important variables can lead to bias in the estimates. However, only variables that are unaffected by the treatment should be included, that is variables that are either fixed over time or measured before the treatment took place. Confounding variables shouldn't be excluded from the propensity scores just because they aren't statistically significant predictors of treatment allocation, but including factors which are not causally associated with the outcome may unnecessarily reduce the statistical power.

Propensity score matching was originally devised to deal with situations where there are two possible treatment allocations. However, it can be extended to more than two groups and to continuous treatments[13;14]. It may therefore be possible to use propensity score matching with time to endoscopy treated as a continuous variable. This may be more efficient than treating time as an ordered categorical variable by splitting it up in to a large number of discrete time intervals as this may make it more difficult to specify an appropriate functional form.

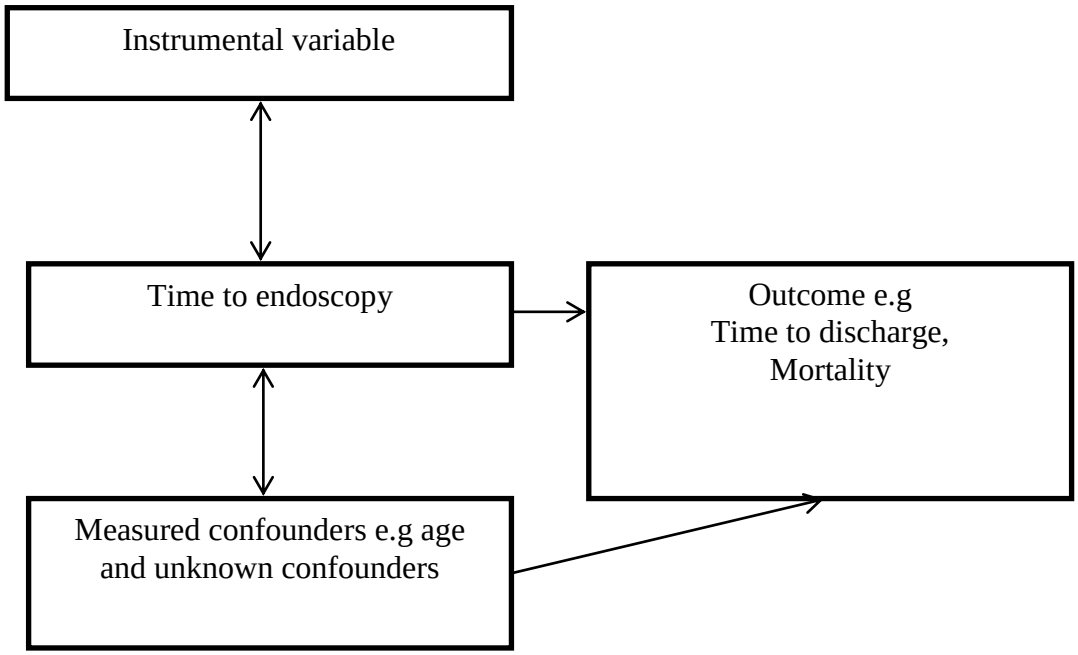
4.3. INSTRUMENTAL VARIABLES

An instrumental variable is one which mimics treatment allocation by being a good, but not necessarily a perfect, predictor of the treatment received by an individual within the observational dataset. As illustrated in Figure 3, a valid instrumental variable needs to be related to the outcome of interest through its affect on treatment allocation alone and not via any other route. In particular, it cannot be related to any measured or unmeasured confounders which affect both treatment allocation and the treatment outcome. Instrumental variables are better suited than propensity score matching to situations where the treatment allocation can be specified as a continuous variable such as time to endoscopy. The main difficulty in using an instrumental variable approach is in identifying an appropriate instrument.

In the case of AUGIB, an instrumental variable is required which is a good predictor of time to endoscopy, but which isn't related to any potential confounding factors that might be related to both time to endoscopy and the outcome e.g mortality or time to discharge. Figure 2 shows the following factors as being predictive of time to endoscopy, but not predictive of either length of stay or mortality except through time to endoscopy

- Endoscopy policy
    - Daily lists versus weekday lists
    - Formal versus informal OOH rota
    - Dedicated AUGIB emergency slots in lists
- Admission route (A&E versus GP)

Figure 3 An illustration of an instrumental variable analysis



The 2007 BSG/NHSBT audit collected data on whether participating hospitals had an OOH consultant on-call rota for endoscopy. One might expect the provision of an OOH on-call rota to be a good instrumental variable as it shouldn't be related to any patient level variables, such as age or comorbidity, which may predict outcomes. One might also expect it to be a strong predictor of time to endoscopy in the group of patients presenting OOH. However, the

findings of the BSG/NHSBT audit suggest that this relationship is complicated by informal OOH endoscopy provision in those hospitals without a formal OOH on-call rota. In the 56% of sites without such a rota, 13% of patients underwent endoscopy OOH. The rate was 20% in hospitals with an OOH rota (Hearnshaw 2010). This difference in the rate of OOH endoscopy was statistically significant ($p=0.001$) even after adjusting for hospital clustering effects and pre-endoscopy Rockall Score using binary regression. There was also a significant difference ($p<0.05$) in the time to first endoscopy between sites with and without a formal endoscopy rota. This suggests that the provision of a formal OOH rota may be somewhat predictive of time to endoscopy. However, given that endoscopy is being provided based on clinical judgement in those hospitals without an OOH on-call rota, it is probable that the more severe cases will be more likely to receive OOH endoscopy, reducing variability in time to endoscopy for the patient group most likely to benefit from the provision of an OOH rota. This lack of variability in time to endoscopy in the most severe cases may make it difficult to demonstrate any mortality benefit attributable to the provision of a formal OOH rota.

A better candidate for an instrumental variable may be the provision of daily lists as this might be a predictor of time to endoscopy in a subset of patients who present at or just before the weekend. The benefit of daily lists would be mainly seen in medium to low risk patients presenting at the weekend, who wouldn't otherwise receive endoscopy until the following Monday. Those patients requiring very urgent endoscopy would probably receive an endoscopy during the weekend in the absence of weekend lists through either formal on-call endoscopy provision or through informal OOH endoscopy provision. Therefore weekend lists may only be a good predictor of time to endoscopy in a subset of patients.

Hospitals were also asked about whether dedicated AUGIB endoscopy slots were available. These may affect time to endoscopy independently of other factors, however, this may not be a strong predictor if hospitals without dedicated slots arrange their lists such that patients with AUGIB are generally seen early in the day even when there are no dedicated slots. As such, this was not considered to be a particularly strong candidate for an instrumental variable.

Admission route might be a good instrumental variable in that it was felt by the clinical advisors to be unrelated to time to discharge or time to death and to potentially result in a delay of 4 to 6 hours. However, this delay wouldn't be vey informative when trying to decide

between daily or weekday lists. Furthermore, before selecting this as an instrumental variable, it would be worth examining the data to see whether patients presenting to A&E had more severe symptoms than those presenting to their GP.

The other factors identified in Figure 2 as being predictive of time to endoscopy were not considered to be suitable candidates for an instrumental variable. For example, weekend versus weekday presentation is likely to predict time to endoscopy in those hospitals which do not have weekend lists. However analysis of data from the same BSG/NHSBT audit[9] found that patients presenting at the weekend were more likely to present with shock and hematemesis and less likely to present with melaena. This suggests that presentation at the weekend would not be a good instrumental variable as these factors are likely to predict outcomes as well as time to endoscopy. Patients who bleed whilst admitted as an inpatient for another reason are recognised to be at higher risk of death than those admitted for AUGIB[5] and therefore this wouldn't be a good instrument variable. Clearly those patient factors included in either the Blatchford or Rockall risk scores are likely to be related to outcomes and therefore would not make a good instrumental variable.

The two strongest candidates for instrumental variables are the provision of daily lists and provision of a formal OOH endoscopy rota. However, neither of these instruments are likely to be strong predictors of time to endoscopy in the whole AUGIB population. Weekend lists may be a good instrument for medium risk patients presenting at the weekend and a formal OOH endoscopy rota may be a good instrument in high risk patients presenting OOHs. Therefore, it might be useful to pre-specify subgroups within the whole population who are likely to be affected differently by different aspects of endoscopy policy and to look for instrumental variables that are strong predictors of time to endoscopy within each subpopulation. Using different instrumental variables in different groups would provide an estimate of the 'local' average treatment effect. It is local in the sense that it is identified through those for which the instrumental variable induces a change in the treatment.

The AUGIB example discussed here illustrates the fact that the main difficulty in applying an instrumental variable approach which is identifying a good instrumental variable. None of the factors discussed above were considered to be a good variable in the whole population presenting with AUGIB. This reflects the fact that the population presenting with AUGIB is diverse, with different service provision policies affecting patients differently according to

their time of presentation and the severity of their bleed, although this may not be so problematic if the population can be divided prospectively into homogeneous subgroups.

4.4. OVERALL STATISTICAL APPROACH

Given that none of the approaches described above can guarantee to remove all confounding influences, it would be wise to compare the results achieved using several different techniques. It would be particularly useful to compare the results achieved through an instrumental variables approach with those achieved through either regression or propensity score matching as these make different assumptions regarding the presence of unmeasured confounders.

5. FEASIBILITY CONSIDERATIONS

5.1. SAMPLE SIZE AND STATISTICAL POWER

The potential for detecting the impact of OOH endoscopy provision on outcomes such as length of stay and mortality is hindered by the fact that the impact of OOH provision on time to endoscopy will vary according to the time of presentation of the AUGIB with many patient's care being unaffected by OOH provision. For example, in the 2007 BSG/NHSBT audit, 59% of patients presented OOH (i.e. outside of 8am to 5pm, Monday to Friday) with 20% presenting between midnight and 8am[3]. Whilst OOH endoscopy provision might have some impact on those presenting during normal working hours, the inclusion of this population in the analysis is likely to dilute the observed effect of OOH endoscopy provision on outcomes. As a result, a conservative measure would be to power any statistical analysis assuming no effect in those presenting during normal working hours.

A further problem with sample size arises when considering the subgroup of AUGIB patients who require very urgent endoscopy. The impact of endoscopy delay on mortality is likely to be greatest in this group, but detecting this mortality effect could be difficult if the number of these very urgent patients within the data set is small as suggested by our clinical advisors who estimated that 5 to 10% of patients required very urgent endoscopy.

The need to adjust for clustering effects at the hospital level may further reduce statistical power and attempts should be made to recruit as many hospital sites as possible. The 2007 BSG/NHSBT audit recruited 84% (217/257) of all acute admitting units in the UK although data from only 208 sites were included in the final analysis [5].

5.2. VARIABILITY IN CURRENT PROVISION

Any cost-effectiveness model examining alternative service delivery models for endoscopy should include an estimate of costs and outcomes for patient undergoing very urgent endoscopy OOH, whether that is through a formal OOH rota or through informal ad-hoc OOH cover, as the number of patients using such services might be affected by other aspects of service provision such as weekend lists. As discussed earlier, it may be difficult to assess the benefit of providing a formal OOH endoscopy on-call rota to allow urgent endoscopy in unstable patients if these patients currently receive urgent endoscopy despite a lack of formal provision. Even if no difference is found between time to endoscopy or outcomes between patients receiving very urgent endoscopy through informal versus formal OOH provision, it should still be possible to estimate the costs of these two options and care should be taken not to undervalue the opportunity cost of informal OOH cover. It may be beneficial to collect data on the actual resource use or opportunity cost associated with providing informal OOH cover to populate a future economic model, although this could be based on a smaller sample of hospital sites than the audit examining time to endoscopy.

5.3. COMPLETION RATES

The 2007 BSG/NHSBT audit invited 257 hospitals to participate and received data on 8939 potential cases of AUGIB from 217 hospitals. Of these 1090 cases (12%) were excluded because the submitted data were incomplete[5]. Reasons given for incomplete data were, 'time pressure, loss of case records or late realisation that entry criteria had not been met[5].' Any future data collection should take this completion rate into account, allowing for this within the calculation of sample size and taking steps to make data collection as simple as possible. Patients presenting to hospitals which operate within a collaborative AUGIB service may also be transferred between sites for endoscopy. This may result in the time to endoscopy being censored unless the hospital records can be linked across sites.

5.4. TIME SCALES FOR DATA COLLECTION

NICE's interim methods guide for developing service guidance discusses the process for identifying evidence including accessing existing audit data and making a call for evidence from stakeholders[15]. However, it does not consider the feasibility of commissioning an observational study to collect data required to populate the economic model during the guideline development period. The 2007 BSG/NHSBT audit collected data during a 2 month period in May / June 2007. Trusts were actively recruited from 6 months before the audit period and the results were reported in December 2007. Therefore this audit took a full year to conduct, without even taking into account the time spent planning the audit. Given that the average NICE Clinical Guideline has around 15 months of development time from the final scope being signed-off to the draft guideline being submitted for consultation, it might be possible to collect audit data on a similarly short audit period provided the analysis can be planned relatively quickly after the scope is finalised.

The feasibility of further data collection to inform guideline development should be considered from the scoping stage. In some cases additional data collection may not be feasible within the usual guideline development time frame and in such cases a judgement would need to be made between the value of the additional data and the cost of extending the timescale for guideline development.

5.5. TECHNICAL EXPERTISE REQUIRED

It has already been recognised within NICE's interim methods guideline for developing service guidance that operational research (OR) methods are likely to be the most appropriate way to assess cost effectiveness and that OR experts should be consulted for advice on the suitability of methods for certain areas [15]. We would agree with this and would also suggest that where additional data is to be collected and analysed to inform the economic evaluation, this should be done with advice from statisticians / econometricians with expertise in the application of the techniques discussed in section 4.

6. CONCLUSIONS

Several issues have been identified from examining this particular example which might be applicable to other service delivery questions.

A lack of variability in current service provision may make it difficult to use current practice as a natural experiment. In this case, there is a lack of variability in actual access to urgent OOH endoscopy for unstable patients despite there being no formal OOH provision in around half of hospitals included in the 2007 audit. This makes it unlikely that any future collection and analysis of observational data would be able to provide a completely unbiased estimate of the true survival benefit attributable to providing urgent OOH endoscopy to unstable patients.

There is a need to consider local variations both in terms of the numbers and clinical characteristics of patients presenting in different areas, and in terms of the organisational environment such as on-site access to other relevant services, and the feasibility of collaborative arrangements which may be dependent on the distance to nearby hospitals.

There is also a balance to be struck when collecting and analysing observational data with regards to the number of confounding factors that should be included in the data collection exercise and the feasibility of conducting the study. Limiting the data collection to routine data that can be obtained retrospectively will increase the chances of achieving high completion rates. However, this may limit the ability of the analysis to capture subtle differences in outcomes between different service delivery models, such as a failure to capture the resource implications of differences between morning and evening discharges if the analysis is restricted to collecting date rather than time of discharge.

It is also important to be careful to define the included population in such a way as to minimise the inclusion of patients unaffected by the choice between different service delivery models. In this example, it would have been useful to identify and exclude patients admitted following an outpatient endoscopy prior to analysing the data.

Several statistical techniques exist to deal with confounding when estimating causal effects from an observational data set. All of the techniques rely on careful identification of potential confounding factors and the relationships between these factors and either treatment allocation or outcomes.

Whilst we anticipate that the lack of variability in time to endoscopy for unstable patients would make it difficult to estimate the survival benefit of timely endoscopy in this group, we

are optimistic that a causal inference could be made regarding the benefits of timely endoscopy for stable patients. The estimates which informed the economic evaluation within the published NICE guideline may be improved upon if more complete data can be obtained on relevant confounding factors and data is collected on the provision of daily lists for use as an instrumental variable.

7. REFERENCES

(1) Rockall TA, Logan RFA, Devlin HB, Northfield TC. Incidence of and Mortality from Acute Upper Gastrointestinal Hemorrhage in the United-Kingdom. *British Medical Journal* 1995; 311(6999):222-226.

(2) National Clinical Guideline Centre. Acute Upper Gastrointestinal Bleeding: Management. 2012. London, The National Clinical Guideline Centre at The Royal College of Physicians.

(3) Hearnshaw SA, Logan RF, Lowe D, Travis SP, Murphy MF, Palmer KR. Use of endoscopy for management of acute upper gastrointestinal bleeding in the UK: results of a nationwide audit. *Gut* 2010; 59(8):1022-1029.

(4) Shokouhi BN, Khan M, Carter MJ, Khan NQ, Mills P, Morris D et al. The setting up and running of a cross-county out-of-hours gastrointestinal bleed service: a possible blueprint for the future. *Frontline Gastroenterology* 2012.

(5) Hearnshaw SA, Logan RF, Lowe D, Travis SP, Murphy MF, Palmer KR. Acute upper gastrointestinal bleeding in the UK: patient characteristics, diagnoses and outcomes in the 2007 UK audit. *Gut* 2011; 60(10):1327-1335.

(6) Davies AHG, Shaq S, Brind AM, Bowling TE, Green JRB. Availability of fully staffed GI endoscopy lists at the weekend for inpatients: does it make a difference? *Clinical Medicine* 2003; 3(2):189-190.

(7) Blatchford O, Murray WR, Blatchford M. A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. *Lancet* 2000; 356(9238):1318-1321.

(8) Rockall TA, Logan RFA, Devlin HB, Northfield TC. Risk assessment after acute upper gastrointestinal haemorrhage. *Gut* 1996; 38(3):316-321.

(9) Jairath V, Kahan BC, Logan RF, Hearnshaw SA, Travis SP, Murphy MF et al. Mortality from acute upper gastrointestinal bleeding in the United kingdom: does it display a "weekend effect"? *Am J Gastroenterol* 2011; 106(9):1621-1628.

(10) Dalton D, Grant-Casey J, Hearnshaw S, Lowe D, Travis S, Rockall T et al. UK Comparative Audit of Upper Gastrointestinal Bleeding and the Use of Blood. 2007. London, British Society of Gastroenterology.

(11) Hearnshaw SA, Logan RF, Palmer KR, Card TR, Travis SP, Murphy MF. Outcomes following early red blood cell transfusion in acute upper gastrointestinal bleeding. *Aliment Pharmacol Ther* 2010; 32(2):215-224.

(12) Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *American Journal of Epidemiology* 2006; 163(12):1149-1156.

(13) Hirano K, Imbens GW. The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives* 2004;73-84.

(14) Imai K, Van Dyk DA. Causal inference with general treatment regimes. *Journal of the American Statistical Association* 2004; 99(467).

(15) National Institute for Health and Clinical Excellence. Interim methods guide for developing service guidance. 2013. London, National Institute for Health and Clinical Excellence.