BMJ Open

Protocol for the health economic evaluation of increasing the weekend specialist to patient ratio in hospitals in England

Samuel I Watson,[1] Yen-Fu Chen,[1] Julian F Bion,[2] Cassie P Aldridge,[2] Alan Girling,[3] Richard J Lilford,[4] on behalf of the HiSLAC Collaboration

Check for updates

[1]Warwick Medical School, University of Warwick, Coventry, UK
[2]Institute of Clinical Sciences, University of Birmingham, Birmingham, UK
[3]Institute of Applied Health Research, University of Birmingham, Birmingham, UK
[4]Division of Health and Population Sciences, Warwick Medical School, University of Warwick, Coventry, UK

Correspondence to
Dr Samuel I Watson;
s.watson.1@warwick.ac.uk

ABSTRACT

Introduction This protocol concerns the evaluation of increased specialist staffing at weekends in hospitals in England. Seven-day health services are a key policy for the UK government and other health systems trying to improve use of infrastructure and resources. A particular motivation for the 7-day policy has been the observed increase in the risk of death associated with weekend admission, which has been attributed to fewer hospital specialists being available at weekends. However, the causes of the weekend effect have not been adequately characterised; many of the excess deaths associated with the 'weekend effect' may not be preventable, and the presumed benefits of improved specialist cover might be offset by the cost of implementation.

Methods/design The Bayesian-founded method we propose will consist of four major steps. First, the development of a qualitative causal model. Specialist presence can affect multiple, interacting causal processes. One or more models will be developed from the results of an expert elicitation workshop and probabilities elicited for each model and relevant model parameters. Second, systematic review of the literature. The model from the first step will provide search limits for a review to identify relevant studies. Third, a statistical model for the effects of specialist presence on care quality and patient outcomes. Fourth, valuation of outcomes. The expected net benefits of different levels of specialist intensity will then be evaluated with respect to the posterior distributions of the parameters.

Ethics and dissemination The study was approved by the Review Subcommittee of the South West Wales REC on 11 November 2013. Informed consent was not required for accessing anonymised patient case records from which patient identifiers had been removed. The findings of this study will be published in peer-reviewed journals; the outputs from this research will also form part of the project report to the HS&DR Programme Board.

BACKGROUND

Seven-day health services are a key policy for the UK government and are of interest to other health systems trying to improve the use of infrastructure and resources in response to

Strengths and limitations of this study

► We contribute to the underdeveloped area of methodology for the analysis of complex service delivery interventions.
► We consider integration and synthesis of multiple forms of evidence from across a complex causal chain.
► This protocol presents the methods for first economic evaluation of the 7-day National Health Service in England and Wales.
► A lack of high-quality experimental data may limit any causal inferences that can be made.
► Multiple sources of uncertainty may limit conclusions even with large samples of data.

rising healthcare demands at a time of fiscal constraint. According to the organisation responsible for managing the National Health Service (NHS) in England, the aim of Seven Day Services is '*to ensure patients receive consistent high quality safe care every day of the week*'[1] by applying 10 clinical standards that must be met every day of the week in order to end '*the variation in outcomes for patients admitted to hospitals in an emergency at the weekend*'. Of the 10 standards, 4 have been prioritised, and of these, 3 are focused on increasing the input of consultants (hospital specialists): time to initial consultant review less than 14 hours following hospital admission, daily consultant review and access to consultant-directed interventions. The 7-day services policy therefore makes an explicit association between perceived worse care at weekends and the input of senior medical staff, with the implication that increasing consultant input will result in better patient outcomes.

A particular motivation[2] for the 7-day policy has been the observed increase in the risk of death associated with weekend admission, known as the 'weekend effect'.[3,4] However, the

BMJ

causes of the excess deaths associated with weekend admission are likely multifactorial, may lie outside hospitals[5] and so far no relationship has been established between the weekend effect and either quality of care in hospital or the availability of consultants.[6] Ideal consultant staffing ratios do not exist. The total cost of training a doctor to specialist level is estimated to be £510 411.[7] Therefore, if more specialists at weekends were to improve the quality of care and potentially eliminate the weekend effect, given current constraints on healthcare resources, it would be important to estimate the cost-effectiveness of this policy under different baseline assumptions. We are not aware of any studies to have conducted an economic evaluation of specialist to patient ratios to facilitate human resources decision making in the literature. Indeed, there are very few formal cost-effectiveness analyses of staffing levels at all, and these are typically limited to comparisons of 'high' versus 'low' staffing provision for specific therapies (eg, ref 8). This paucity of evidence may be due to the lack of methodological development for the evaluation of SDIs, the nature of which presents issues not usually of concern to more 'typical' health technology assessment.[9]

Evaluation of SDIs

Methods for the economic evaluation of health technologies are well developed. Typically, the results from randomised controlled trials (RCT), or other direct evidence of the effect of the intervention on patient outcomes, are incorporated into models. These models are then used to extrapolate the effect forward in time to a prespecified time horizon. The evaluation of SDIs presents an additional layer of complexity. The average effect on any one patient of SDIs is generally small, prohibiting any reasonably sized RCT, and there are often multiple patient endpoints that are relevant (figure 1). As a result, the primary endpoints of studies of SDIs are often the more frequently occurring, 'upstream' endpoints that

are themselves causes of the patient endpoints relevant to economic evaluation (eg, targeted processes in figure 1).[10]

The evaluation of SDIs requires the extrapolation of evidence across the 'causal chain' that links the intervention to patient outcomes (figure 1). Multiple forms of evidence from across this causal chain therefore need to be integrated to make quantitative inferences. This requires a model or models of the processes under consideration, the identification of the available data and a method to integrate the evidence statistically.[9 10]

The complex nature of the healthcare system highlights the importance of institutional and domain-specific knowledge first in model development and second in constructing prior densities for model parameters. Prior knowledge is incorporated using Bayesian methods. The Bayesian methods proposed here permit three other key features. First, they naturally integrate into a decision theoretic framework.[11 12] The purpose of the evaluation is to inform decision makers at the hospital or national level about whether to increase specialist intensity at the weekend. Second, they allow us to appropriately represent the uncertainty involved in such a decision. These uncertainties include biases that may be present in results from previous studies and the data collected for this study itself. Such uncertainty can be propagated through the models using Markov Chain Monte Carlo estimation. Third, the parameter estimates based on the current state of knowledge may be 'updated' with data from the High-intensity Specialist Led Acute Care (HiSLAC) project to allow both ex ante and ex post evaluations.

The method proposed here involves an econometric model to estimate treatment effects from the HiSLAC data from which the overall net benefits will be evaluated. The development of the overall model proceeds in stages and allows for uncertainty in the way in which the data are
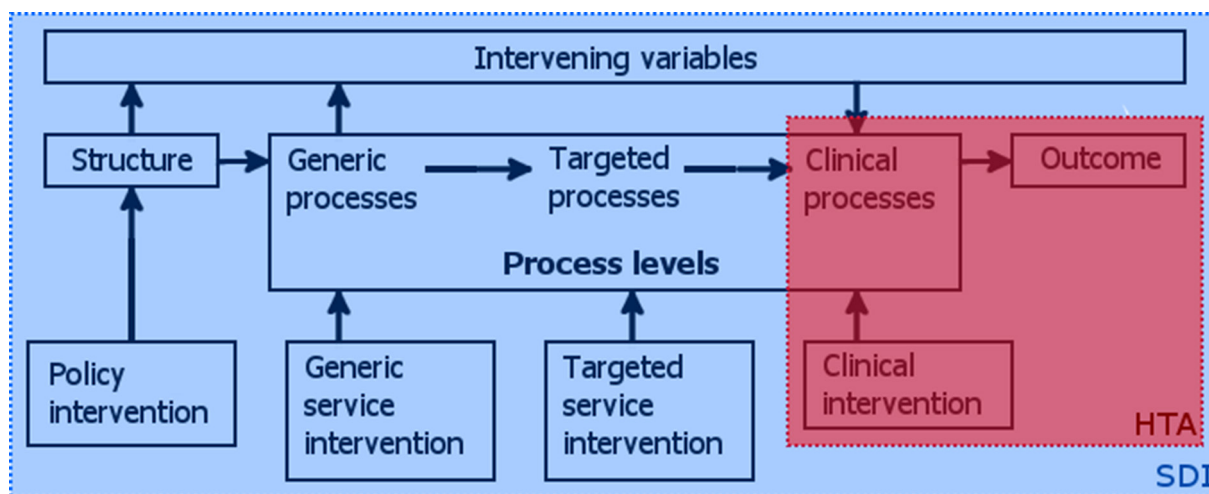

Figure 1 Modified Donabedian causal chain. Health technology assessment (HTA; red box) typically evaluates clinical interventions that directly impact on clinical processes and patient outcomes. Service delivery interventions (SDI; blue box) include policy and generic service interventions may have affects across the causal pathway. Adapted from Lilford et al[10] with permission.

combined. Results from these models will be presented alongside summary statistics from the point prevalence surveys available in other publications.[6]

Aims

The overall aim of the HiSLAC project is to determine whether increasing the intensity of specialist-led care at weekends improves outcomes for patients admitted to hospital as emergencies at weekends. The specific aims of the health economic evaluation comprise: (1) estimation of the relationship between specialist intensity and care quality; (2) valuing changes to care quality; and (3) estimating the expected net benefits (ENBs) of different levels of specialist intensity. We will do this by performing a cost–benefit analysis of a 'high intensity specialist' intervention within a decision theoretic framework. The primary objective will be to determine whether to implement a policy of increasing specialist intensity at weekends to the level observed in 'high intensity' hospitals or not. The secondary aims are: (1) to estimate the cost per quality-adjusted life years (QALY) threshold at which the decision changes and hence the cost-effectiveness of 'high intensity' specialists, (2) to determine the probability that the net benefits are positive at different cost-effectiveness thresholds, and (3) to examine the net effects of the policy in practice by valuing average changes to weekday effects. The method proposed here is intended to take into account factors that may confound the analysis of the effect of specialist intensity on patient outcomes and to identify potential limitations.

This protocol is intended to have enough flexibility such that the modelling reflects the knowledge that is gained during the process. This protocol may also serve as a framework for future evaluations of health SDIs where it is not possible or feasible to measure the effect of the intervention on the endpoints required by decision makers. We also believe that this is the first preregistered protocol for a health economic evaluation, and we hope that this sets a precedent for future analyses.

METHODS/DESIGN

This health economics evaluation forms part of the National Institute of Health Research funded HiSLAC project, a 5-year (2014–2018) research project that is currently investigating the effect of the roll out of 7-day services.[13] There are multiple other components to the HiSLAC project including an ethnographic work package and a cross-sectional analysis of specialist to patient ratios and the estimated weekend effect. A complete outline of the project is available elsewhere.[14] We are concerned here with the health economics evaluation of increased specialist staffing at weekends in hospitals in England. The analysis detailed here serves as a parallel evaluation alongside the main project. The details of the quantitative data collection for HiSLAC are provided here for completeness.

Specialist intensity will be assessed by the annual HiSLAC point prevalence survey (2014–2018). All specialists working in participating NHS Trusts in England are invited to participate in the surveys, which will collect information on specialist presence and hours provided on a (specific) Wednesday and a Sunday each year.[6] Data from Hospital Episode Statistics will also be used to determine the mean number of Wednesday and Sunday emergency admissions (those admitted via the emergency department or directly to a ward) for each hospital for each year. On this basis, it is possible to calculate specialist intensity. The natural experiment in changes to specialist intensity afforded by a shift to 7-day working will be used to evaluate the impact of specialist intensity on quality of care and patient outcomes.

Quality of care will be assessed using case record review of 4000 admissions (50% weekend) from 20 participating hospitals. Hospitals were stratified into quintiles by number of beds and then ordered by specialist intensity per 10 emergency admissions on a Sunday based on the 2014 point prevalence survey.[6] Two hospitals at the top and two at the bottom of each quintile were selected, defined for the purposes of this project to be 'high intensity' and 'low intensity' hospitals, respectively, to give 20 hospitals with a wide variation in size and specialist intensity. From each of the selected hospitals, the case records of 50 randomly selected weekend and 50 weekday admissions will be reviewed at two time points (2013/2014 and 2016/2017), before and after the adoption of a 7-day services policy by the UK government in 2015,[14] respectively. Expert review of these 4000 case records will identify errors in patient care and associated adverse events (AEs). These data will form the basis of an analysis to estimate the effect of changes to specialist intensity on clinical and economic outcomes through extrapolation from care quality assessed by preventable errors.

Development of a qualitative causal model

Causal models are common throughout statistics and econometrics to identify estimators for the treatment effects of interest. However, the processes in a complex system, such as a hospital, are not necessarily well understood. Each of the many endpoints may have many causes, of which specialist presence is only one.[10]

An expert workshop will be convened involving participants with relevant knowledge of the healthcare system, the 7-day services policy and its implications for specialist intensity, and the assessment of patient outcomes within health services research. The results of ethnographic observations of the care of emergency admissions and the role of specialists in providing such care collected by the HiSLAC project will be made available in the workshop. A facilitated discussion will explore the mechanisms by which increasing specialist intensity may affect care quality and patient outcomes. Expert participants will be asked to evaluate the probability of each potential mechanism in contributing to the effects of differing levels of specialist intensity using a visual analogue scale.
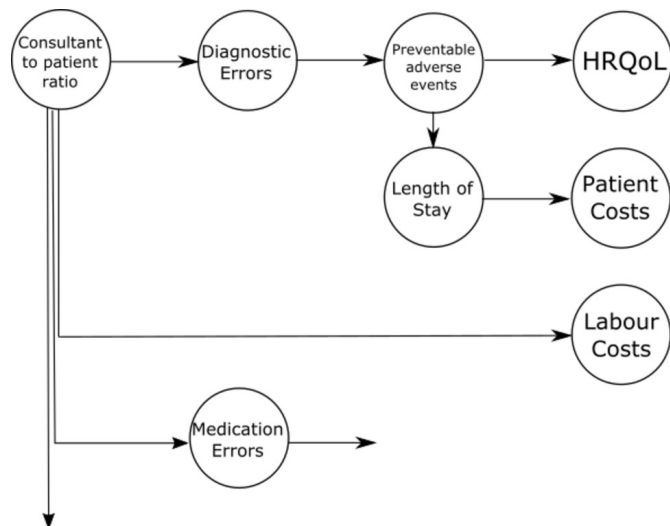
Figure 2 An example qualitative causal model for the effects of implementing high-intensity specialist acute care at the weekend. Arrows indicate conditional dependencies and are interpreted as causal relations. This is a simple model with a limited number of pathways, more may be included in the final model. HiSLAC, High-intensity Specialist Led Acute Care; LoSLAC, Low-intensity Specialist Led Acute Care; QALYs, quality-adjusted life years.

On the basis of the elicitation a qualitative causal model will be developed. We will develop models involving the causal pathways for which it is judged that there is at least at 10% probability of contributing to the effect of specialist intensity on patient outcomes, although the experts will not be made aware of this. A simple example of such a model is shown in figure 2. These models are Bayesian causal networks, which represent the joint probability distribution of the variables. Circles indicate variables, and arrows indicate a causal effect and its direction. Variables separated from one another by another variable are independent conditional on the separating variable. Latent variables, such as unobserved patient health, can also be represented on such diagrams.

Review of the literature

The model developed for the analysis reveals which data are required to make inferences about the effects of 'high intensity' specialists. Previous literature that provides data on the relationships in the models developed will be identified. Standard systematic review methods will be used. In many cases, SDIs converge on similar causal processes: many are aimed at reducing AEs that in turn impact on patient quality of life and length of stay in hospital. We are conducting a systematic review on the effect of preventable AEs on patient length of stay, costs and quality of life outcomes to support this and other similar projects.[15] The other required data are determined by the models; for example, the simple model in figure 2 reveals the need for information on the effects of diagnostic error on the risk of experiencing a preventable AE and the costs of implementing 'high intensity' specialists.

Statistical modelling and data synthesis
Expected net benefits

Let x be the specialist to patient ratio. The decision is then whether to increase the specialist to patient ratio from x_L to x_H, where these levels might be regarded as 'low intensity' and 'high intensity', respectively, for example. We consider the problem in a simple decision framework with a linear loss function: the decision is to make the change if the ENB is positive.[16] We therefore aim to estimate the ENB of increasing the specialist to patient ratio from x_L to x_H, where the policy relevant effects of specialist intensity are assumed to operate through changing care quality. For the purposes of this exposition, we will consider the model in figure 2 and concentrate on estimation of the 'treatment effect' of the specialist to patient ratio on the risk of experiencing a diagnostic error. The same modelling strategy applies to other relevant outcomes. If other branches are added to the model, for instance through avoiding prescribing error, then the same process would be followed for this branch and the resulting probabilities summed. As a secondary analysis, we will explore whether any changes to specialist staffing occur during the weekday to estimate the net effect of the policy on overall specialist staffing and care quality.

The direct costs associated with providing specialist to patient ratio x are the additional costs of labour, $c(x)$, which is increasing in x. The outcomes of providing a specialist to patient ratio are the monetary value of the preventable AEs that occur at specialist to patient ratio x, $d(x)$. The ENB of increasing the specialist to patient ratio from x_L to x_H is:

$$E(\mu(x_H,\ x_L)) = \begin{aligned} &E[(d(x_H) - d(x_L)) \\ &- (c(x_H) - c(x_L))] \end{aligned} \quad (1)$$

where $\mu(...)$ is the net benefit function and the decision would be to increase specialist intensity to 'high intensity' levels if $E(\mu(x_H, x_L)) \geq 0$.

Model parameterisation

The strategy for estimating the net benefits is outlined in figure 3. The first part of the ENB in equation (1) is the change in economically relevant outcomes due the policy change from 'low intensity' to 'high intensity' at the weekend associated with preventable AEs. Let y_k be the value of outcome k: either the health service costs or QALYs (see figure 2), then $E(d(x_H) - d(x_L)) = \sum_k [E(y_k|x_H) - E(y_K|x_L)]$, that is, the sum of the mean incremental costs and QALYs lost associated with the policy through preventable AEs. These costs and QALYs have not previously been estimated and will not be estimated directly by the HiSLAC project. However, it can be expressed in terms of the intermediary variables that mediate the effect of care quality on patient outcomes:

$$E(y_k|x_H) - E(y_k|x_L) = \begin{aligned} &[P(D|x_H) - P(D|x_L)] \\ &P(A|D)[E(y_k|A) - E(y_k|A)] \end{aligned} \quad (2)$$

where D is the event that a patient experiences a diagnostic
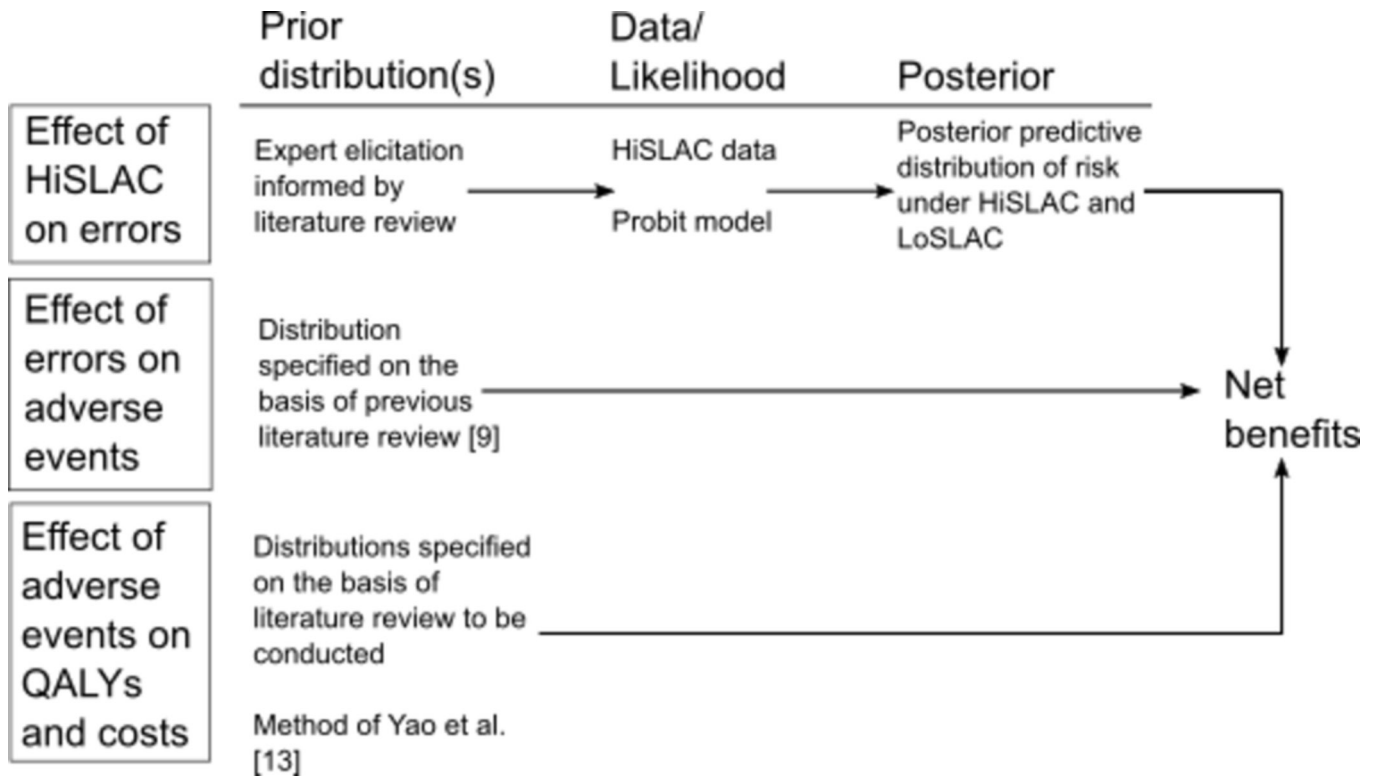
Figure 3 Summary of the data sources and specifications to be used in the analysis. HRQoL, health-related quality of life.

error, and A is the event a patient experiences an AE (and A' its complement, ie, no AE). This can be considered as valuing the effect of specialist intensity on care quality.

Estimation of the effect of specialists on errors

The first term on the right-hand side of equation (2), $P(D|x_H) - P(D|x_L)$, is the absolute difference in the risk of a patient experiencing a diagnostic error. As previously described, data will be collected from a stratified (by day of week) random sample of 100 patients, 50 weekend and 50 weekday admissions ($w = 1$ and $w = 0$ respectively), from each of 20 participating hospitals ($j = 1, ..., 20$) at two time points, (2013/2014 and 2016/2017 corresponding to $t = 1$ and $t = 2$, respectively). The specification of the model proposed here is based on first, estimating and recovering treatment effects, and second, ensuring the parameters have straightforward interpretations for the purposes of expert elicitation and the determination of appropriate prior distributions.

The model we propose is a Bayesian hierarchical model. Fundamentally, the model estimates the marginal effect of the specialist to patient ratio on the probability of experiencing the outcome, net of secular trends in the same hospitals and trends between hospitals at the weekend, in a way conceptually similar to the difference-in-differences model. Such trends may include changes in other staffing levels, which we assume will not change specifically at the weekend given the relevant guidance does not target non-specialist staffing intensity. The model is assumed to identify the causal effect of increasing

the specialist to patient ratio at the weekend based on the fact that hospitals are mandated by policy to meet certain specific specialist-specific standards but under the assumption that they will be doing so at different rates and from different starting positions given the specialist levels required to meet these standards. Hospitals with lower weekend ratios are expected to increase their ratio by more than hospitals with higher ratios. However, at the same time, there are hypothesised changes in the rates of diagnostic error or other relevant outcomes as well as possible temporal effects.

We will specify a generalised linear model for the risk of experiencing an error. For patient, treated in hospital, at time, at weekend or weekday $w = 0, 1$, respectively, we specify

$$Pr(D_{ijtw} = 1) = \begin{aligned} &F(\alpha_j + \gamma x_{jtw} + \beta_{1j}I_w \\ &+ \beta_{2j}I_T + \beta_3 I_w I_T + z'_{ijtw}\delta) \end{aligned} \quad (3)$$

where D_{ijtw} is a binary variable equal to one if the patient experienced a diagnostic error (or other outcome of interest) and zero otherwise, z_{ijtw} is a vector of patient covariates (age, sex and number of comorbidities), I_T and I_W are binary indicators for period and weekend, respectively, and x_{jtw} is the specialist to patient ratio. An informative prior density is specified for γ on the basis of an expert elicitation informed by any prior literature identified. The hospital effect is modelled as $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$, and similarly the varying slopes are modelled as $\beta_{1j} \sim N(\mu_t, \sigma_{\beta1}^2)$ and

$\beta_{2j} \sim N(\mu_w, \sigma^2_{\beta2})$. Standard deviations will be assigned half-normal priors[17] and the remaining parameters will be assigned weakly informative $N(0,5^2)$ prior distributions. Probit, logit, exponential and linear specifications will be used for the link function, F, and the best-fitting model selected (this list is not exhaustive). It is possible that the risk of different types of error, should multiple outcomes be considered, are not (conditionally) independent of one another. Therefore, a multivariate probit model will also be compared in the case of multiple outcomes. Models will be compared on the basis of the Watanabe-Akaike information criterion; further model comparisons are described later. We will estimate the posterior predictive distribution of risk of the outcome of interest for given levels of specialist intensity for the second time period, averaged over the distributions of the other covariates. From this we will estimate the (posterior predictive) absolute treatment effect.

Estimation of the effect of errors on AEs

As previously discussed, the results of the analysis of patient data need to be linked to the patient outcomes relevant to economic evaluation. Equation (2) provides the formula to do so. The second term of equation (2) is the probability of a preventable AE given a diagnostic error (or other type of error), and it is also not reported in the literature as far as we are aware. However, we can use Bayes' theorem:

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} \qquad (4)$$

A previous systematic review of AEs provides the required probabilities for the numerator on the right hand side of equation (4).[18] For example, of the studies included in that systematic review, the median reported percentage of patients experiencing an AE was 9.2%, of which a median of 43.5% were preventable; therefore, approximately 4.0% of patients experience a preventable AE ($P(A) = 0.04$). A median of 7.5% of AEs were related to diagnostic errors ($P(D|A) = 0.075$).[18] These probabilities will also be estimated from case record review data. The probability of experiencing a diagnostic error, $P(D)$, will be estimated from the case record review data.

Estimation of the effects of diagnostic errors on AEs and patient outcomes

The third term on the right hand side of equation (2), $E(y_k|A) - E(y_k|A')$, is the expected impact of an AE on outcome k: the incremental costs and the incremental QALYs lost associated with a preventable AE. As described, we are conducting a systematic review of the literature on the consequences of preventable AEs on length of stay and costs.[15] The results from the identified studies will be used to inform the specification of the distribution for the incremental effect of a preventable AEs on costs: studies from the UK will be used if possible; otherwise, studies from Organisation of Economic Cooperation and Development (OECD) countries will be used. The incremental effect of a preventable AE on length of stay will be multiplied by the average cost per bed day in the NHS from the NHS Reference Costs of the relevant year. In our mapping of the literature, no studies were identified by our systematic review that estimated the mean QALY difference.

We will extend the approach of a previous study to developing a specification for the incremental effect of an AE on QALYs.[19] AEs, following the study by Brennan *et al*, are often classified in terms of their incremental impact on patient disability and its duration. The incidence of AEs by category of health effects has been reported in a number of previous studies. These categories are typically mortality, permanent disability, moderate impairment and minimal harm. We specify a multinomial distribution for these outcomes and a Dirichlet distribution as the prior distribution for the probabilities. The Dirichlet prior will be updated with data on the outcomes of preventable AEs from previous studies.

To specify a distribution for the QALY loss to each category, information is required on the severity and duration of the effect. In the case of mortality, we will assume that the remaining life expectancy is exponentially distributed with a rate parameter equal to the reciprocal of the estimated mean life expectancy of patients who experienced a fatal AE. This is also the duration to be used for a permanent disability. For temporary disability, in the absence of further information, we specify that the duration is a distributed according to a gamma distribution. In the baseline analyses, we will select parameters such that the median duration of a temporary disability is 2 months (approximately the length of time for a broken bone to heal) and that 95% of patients will have recovered by 12 months. Moreover, for minimal harm, we specify that the duration is uniformly distributed between 1 day and 1 month. A deterministic sensitivity analysis will be conducted to examine the robustness of the results to these choices of distributions.

The health-related quality of life (HRQoL) associated with the lost years for each patient is required to estimate the QALY loss associated with mortality. To estimate the severity, we will take a weighted sample from the EQ-5D-derived HRQoL weights reported in the Health Survey for England,[20] survey years 1996, 2003–2006 and 2008, with the weights given by the age groups of those patients who died. The probability of being in a given age group conditional on having experienced a fatal preventable adverse event will be determined from the literature. We will then take the mean and SD from this weighted sample to specify a distribution for the QALY weights.

For the remaining disability categories, we will use the method of Yao *et al* and determine quality of life weights using the EQ-5D questionnaire as applied to conditions representative of each category.[19] We will use the HRQoL weight previously discussed as the baseline weights. We will also search the literature for quality of life studies

reporting on the representative conditions. The counterfactual quality of life will be as described for mortality.

To convert the estimated QALY losses to a monetary value, we will use the societal willingness to pay per QALY of £20,000 as specified by the National Institute of Health and Care Excellence. Interventions below this threshold are considered cost-effective. We will explore the decisions that would be made at a range of thresholds in a deterministic sensitivity analyses.

Elicitation of prior densities

An expert elicitation workshop will be conducted in order to elicit a prior distribution for the marginal 'treatment effect' of the specialist to patient ratio on diagnostic errors (and other types of error if appropriate). Established techniques for expert elicitation will be used[21 22]; we will follow the same protocol as a previous project.[23] The workshop involves a number of stages: (A) introduction and explanation of key concepts; (B) a training exercise to elicit beliefs about a known quantity to familiarise participants with the process; (C) group discussion of the relevant evidence; (D) first round belief elicitation; (E) a break; (F) feedback of results from the first round; and (G) second round elicitation. Data will be presented to participants in a randomised fashion to prevent anchoring bias.[24] Participants will be asked to provide quantiles of their beliefs regarding the probability of each model using the Sheffield Elicitation Framework.[25] The elicited quantiles will be transformed using the appropriate link function, and a normal distribution will be fitted to these quantiles, and they will be aggregated using linear opinion pooling. Only the results from the second round elicitation will be used.

Some researchers are uncomfortable with the notion of subjective probability. We do not offer to explore the argument into Bayesian versus frequentist interpretations of probability here. However, we do note that there is no coherent frequentist interpretation to the there being a set of causal models, each of which has a certain probability of being true; nevertheless, we contend that this is an important aspect of uncertainty that needs to be accounted for. There are often concerns about cognitive biases and heuristics during expert elicitations,[21 24] and we ensured that we followed best practice to avoid these. This included the use of training exercises, the provision of feedback and the scrambling of evidence to prevent anchoring biases. Such techniques have been previously used successfully in health services research.[19 21]

Costs

Where an increase in specialist intensity is achieved through an increase in budget, the per-patient incremental cost of HiSLAC will be determined by calculating the incremental change in the number of specialists (the incremental change in the ratio multiplied by the number of patients) and multiplying this by the cost of a whole time equivalent specialist. Specialists are paid a basic salary on a scale. Since the levels of specialist contributing to any changes in specialist intensity are unknown, we will specify a uniform distribution over the specialist pay scale for the cost of a specialist. On the basis of the specialist survey, we will use the mean specialist hours per emergency admission for 'low intensity' and 'high intensity' hospitals as the values on which to base the cost calculations.

Model checks

It is important to assess the validity of the model. If the model fails to produce a reasonable summary of the data, then it should be excluded.[26] Posterior predictive checks are a method of comparing the distribution of new data generated from the model with the observed data; if the model is a good fit to the data, then the replicated data will resemble the observed data. Such checks are based on discrepancy statistics and graphical comparisons. If the observed data lie at the extremes of the distribution from the replicated data, then the model is a poor fit. We will conduct two checks based on 100 new data sets of 4000 observations, each of which will be generated from the model of the effect of the specialist to patient ratio on diagnostic errors each with a new set of parameters generated from the posterior predictive distribution of the parameters. The two checks we will use are: first, a graphical comparison of the relationship between specialist intensity and risk of error. This is considered a check of the 'internal model' in equation (3), which should enable us to detect if there are any problems for particular values of specialist intensity. Second, we will examine the χ^2 statistic from a likelihood ratio test comparing observed and expected outcomes from observed and replicated data for the contingency table of outcomes in 'high intensity' and 'low intensity' hospitals in the two time periods.

ETHICS AND DISSEMINATION

Informed consent was not required for accessing anonymised patient case records from which patient identifiers had been removed. The findings of this study will be published in peer-reviewed journals; the outputs from this research will also form part of the project report to the HS&DR Programme Board.

REFERENCES

1. NHS England. Seven day hospital services. 2017 https://www.england.nhs.uk/ourwork/qual-clin-lead/seven-day-hospital-services/ (accessed 31 Oct 2017).
2. Prime Minister's Office. PM on plans for a seven-day NHS. 2015 https://www.gov.uk/government/speeches/pm-on-plans-for-a-seven-day-nhs (accessed 30 Oct 2015).
3. Pauls LA, Johnson-Paben R, McGready J, *et al*. The weekend effect in hospitalized patients: a meta-analysis. *J Hosp Med* 2017;12:760–6.
4. Zhou Y, Li W, Herath C, *et al*. Off-hour admission and mortality risk for 28 specific diseases: a systematic review and meta-analysis of 251 cohorts. *J Am Heart Assoc* 2016;5:1–12.
5. Anselmi L, Meacock R, Kristensen SR, *et al*. Arrival by ambulance explains variation in mortality by time of admission: retrospective study of admissions to hospital following emergency department attendance in England. *BMJ Qual Saf* 2017;26:613–21.
6. Aldridge C, Bion J, Boyal A, *et al*. Weekend specialist intensity and admission mortality in acute hospital trusts in England: a cross-sectional study. *Lancet* 2016;388:178–86.
7. Curtis L, Burns A. Unit costs of health and social care. 2016 http://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2016/ (accessed 31 Oct 2017).
8. Ganz DA, Simmons SF, Schnelle JF. Cost-effectiveness of recommended nurse staffing levels for short-stay skilled nursing facility patients. *BMC Health Serv Res* 2005;5:35.
9. Watson SI, Lilford RJ. Integrating multiple sources of evidence: a Bayesian perspective. In Raine R, Fitzpatrick R, Barratt H, Bevan G, Black N, Boaden R, et al. Challenges, solutions and future directions in the evaluation of service innovations in health care and public health. *Health Serv Deliv Res* 2016;4:1–18.
10. Lilford RJ, Chilton PJ, Hemming K, *et al*. Evaluating policy and service interventions: framework to guide selection and interpretation of study end points. *BMJ* 2010;341:c4413.
11. Goodman SN. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 1999;130:1005–13.
12. Lindley DV. The philosophy of statistics. *Stat* 2000;49:293–337.
13. Bion J. High-intensity specialist-led acute care. 2015 http://www.hislac.org (accessed 22 Jan 2015).
14. Bion J. Prospective evaluation of the impact of high-intensity specialist-led acute care (hislac) on emergency medical admissions to nhs hospitals at weekends. 2016 http://hislac.org/images/docs/HiSLAC_Protocol_V4_May_18th_2015.pdf (accessed 31 Oct 2017).
15. Watson S, Taylor C, Chen Y. A systematic review and meta-analysis to identify the health and economic consequences of adverse events at the patient-level. 2015 http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015019578 (accessed 31 Oct 2017).
16. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ* 1999;18:341–64.
17. Gelman A. Prior distribution for variance parameters in hierarchical models. *Bayesian Anal* 2006;1:515–33.
18. de Vries EN, Ramrattan MA, Smorenburg SM, *et al*. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* 2008;17:216–23.
19. Yao GL, Novielli N, Manaseki-Holland S, *et al*. Evaluation of a predevelopment service delivery intervention: an application to improve clinical handovers. *BMJ Qual Saf* 2012;21(Suppl 1):i29–38.
20. NatCen Social Research. Health survey for England. Report No: SN: 7649.
21. O'Hagan A. *Uncertain judgements: eliciting experts' probabilities*. 1st edn: John Wiley and Sons, 2006.
22. O'Hagan A. Eliciting expert beliefs in substantial practical applications. *J R Stat Soc Ser D* 1998;47:21–35.
23. Lilford RJ, Girling AJ, Sheikh A, *et al*. Protocol for evaluation of the cost-effectiveness of ePrescribing systems and candidate prototype for other related health information technologies. *BMC Health Serv Res* 2014;14:314.
24. Johnson SR, Tomlinson GA, Hawker GA, *et al*. Methods to elicit beliefs for Bayesian priors: a systematic review. *J Clin Epidemiol* 2010;63:355–69.
25. O'Hagan T. SHELF: The Sheffield Elicitation Framework. http://www.tonyohagan.co.uk/shelf/ (accessed 15 Feb 2016).
26. Gelman A, Meng X-L, Stern H. Posterior predictive assessment of model finess via realized discrepancies. *Stat Sin* 1996;6:733–807.