***W**hat is the **i**mpact of large scale implementation of **s**troke **E**arly Supported Discharge? (WISE)*

Analysis Plan for Work Package 1 (WP1)

Introduction to WISE project and role of WP1 within the overall project plan/aims

Stroke Early Supported Discharge (ESD) is a multidisciplinary team intervention that clinical trials have shown reduces length of hospital stay and reduces risk of dependency of stroke survivors. The ESD intervention comprises facilitated discharge from hospital and delivery of stroke specialist rehabilitation in patients' homes. There has been widespread implementation of ESD across England, following recommendations in UK national policy documents and clinical guidelines [1,2]. However, the type of ESD service patients receive on the ground is variable and in some regions ESD is still not offered at all [3]. The WISE study will investigate the impact of implementing ESD at scale in real world conditions and investigate which models of ESD are effective in practice.

To undertake this research, a realist evaluation approach composed of two interlinking work packages will be adopted. The aim will be to understand how and why ESD works, for whom and in what circumstances. Work package 1 will involve analysis of historical prospective national stroke audit data from hospital and community providers across the East Midlands, West Midlands, East of England and North of England. This provides for a large dataset with which to investigate a Midlands and East initiative to implement ESD and a region of England (North) identified as slower to respond. Effectiveness of ESD will be measured with evidence based metrics as defined by national clinical guidelines for stroke, and that reflect outcome measures used in the original ESD trials [4,5]. Work package 2 (WP2) will use a qualitative approach to examine further how confounding factors influence the implementation and effectiveness of ESD in practice. Case study sites predicting contrasting results for anticipated reasons will be included. Semi-structured one-to-one interviews with NHS commissioning and service managers and group interviews with provider staff will be conducted to investigate how and why ESD was implemented in the first place, perspectives on what makes an ESD service effective and cost implications of ESD models. Stroke survivors will also be interviewed to explore what matters most to them.

This WISE study will investigate whether and how current large scale implementation of ESD is achieving the outcomes clinical trial evidence suggested it would; if this is the case, findings will help facilitate further implementation of ESD nationally and internationally. In addition, this study adopts a mixed-methods approach to progress current theoretical understanding of the implementation of complex interventions. Findings will inform debate about optimal methodological approaches for evaluation of healthcare delivery.

Introduction to WP1 framework

Programme theories

To bridge the second translational gap between randomised controlled trials and clinical practice, we will adopt a realist evaluation approach [6]. Central to the realist approach is the assumption that context makes a difference to intervention outcomes. The focus is not simply to understand *if* interventions work, rather on *how* or *why* they work (or don't work). This approach is critical with regard to understanding how complex interventions, such as ESD, work in real world settings. Policymakers, commissioners and NHS providers need to understand how and why different ESD models are effective (or not) in different contexts, so that they are better equipped to make decisions about healthcare provision in their local area.

Realist evaluation involves the development of programme theories about how activities associated with interventions cause outcomes. These programme theories are then tested and refined during the evaluation process. Realist evaluators use the context-mechanism-outcome (CMO) configuration as the main structure for realist analysis [6]. Critical to the approach is the idea of mechanisms and how they relate to context. In simplistic terms, whether mechanisms lead to desired outcomes depends on the context. As a basic example relating to ESD, a CMO pattern could be that operating in a rural location (context), disrupts multidisciplinary team working (mechanism), which influences the delivery of rehabilitation in the patients' home (outcome).

We have developed programme theories relating to the implementation and effectiveness of ESD. These have been guided by conceptual frameworks of implementation science, previous research and have informed our research questions [7,8,9]. Our first programme theory hypothesises that adoption of evidence based ‘core components’ of ESD are important for the intervention to be effective in practice [7]. We believe these core components represent important ‘mechanism-outcome’ links, with regard to how effective ESD is in practice.

Core components include that ESD is delivered by *stroke specialist staff* operating as a *coordinated multidisciplinary team* (e.g. doctors, nurses, occupational therapists, physiotherapists, speech and language therapists, clinical psychologists, social workers and rehabilitation assistants). The ESD team should have *weekly* multidisciplinary team meetings and regular meetings with stroke unit hospital staff. Members of the ESD team should have *training* opportunities. Furthermore, an effective ESD intervention consists of a *coordinated, facilitated discharge* from hospital and provision of *timely rehabilitation* to *eligible* (mild to moderate) stroke patients at home within 24 hours of acute stroke unit hospital discharge. The ESD teams should plan to provide services every day of the week [7].

Further to these ESD core components, we want to better understand the extent to which contextual factors are associated with the effectiveness/contribute to the variability of ESD provision. Previous research suggest these factors include characteristics of the stroke patients accessing the service, other characteristics of the ESD team itself such as size of the team, and features of the location in which it operates [4,5,10,11]. Hence our second programme theory states that contextual factors at every level, i.e. the patient, the team and the site, need to be considered as part of an investigation of ESD effectiveness.

A third programme theory relates in particular to the geographical location in which ESD is implemented. The question of how ESD might operate in rural settings has been raised, given the fact that the majority of randomised controlled trials were conducted in urban settings [4,5,7]. Therefore, our third programme theory is the hypothesis that core components of ESD will operate differently in urban versus rural settings. Please see Figure 1 in the appendix for a visual representation of our initial programme theory configurations.

Multilevel modelling

When we believe individuals are clustered together in some way such as grouped by ESD provision, we might expect that two randomly selected individuals from the same group will tend to be more alike than two individuals selected from different groups. Such dependencies can therefore be expected to arise and we need multilevel models – also known as hierarchical linear models, mixed models, random effects models and variance components models - to analyse data with a hierarchical structure. Note that the data structure discussed here is strictly hierarchical, that is each patient belongs to a single ESD team. More generally, structures can be non-hierarchical such as cross-classification whereby patients can be nested within ESD teams and hospitals without the hospitals and ESD teams to be nested within each other. Multilevel modelling can accommodate such data structures. Furthermore, multilevel modelling can also accommodate non-normal responses which is very useful if any outcome variable cannot be treated as continuous.

If data are grouped and we have not taken account of group effects in our regression model, the independence of observations assumption will not hold thereby violating the inference obtained from a single-level multiple regression model. Specifically, the standard errors of the regression coefficients will generally be underestimated. Consequently confidence intervals will be too narrow and p-values will be too small, which may in turn lead us to infer that a predictor has a 'real' effect on the outcome when in fact the effect could be ascribed to chance. Underestimation of standard errors is particularly severe for coefficients of predictors that are measured at the group level, e.g. an indicator of whether an ESD team is mostly rural or urban. Correct standard errors will be estimated only if variation among groups is allowed for in the analysis, and multilevel modelling provides an efficient means of doing this.

Obtaining correct standard errors is just one reason for using multilevel modelling. Multilevel modelling also enables researchers to investigate the nature of between-group variability, and the effects of group-level characteristics on individual outcomes. This method allows us to explore if there is between-ESD model variability in patient outcomes and hence evaluate overall differences in ESD provision effectiveness. We can also examine if patient characteristics vary across ESD models to see if some ESD models are more effective for certain types of patients. Crucially for this study multilevel modelling allows us to explore what factors contribute to variability in patient outcomes at the patient level, e.g. age, and the ESD team level, e.g. level of rurality within the catchment area of a given team.

Where there is patient data nested within ESD or hospital team, we can test the suitability of applying a multilevel model by comparing it with single-level model. We can do this using a Likelihood Ratio Test (LRT) so long as the single-level model is nested within the multilevel model, i.e. the single-level model applies the overall sample mean to all patients and the multilevel model extends the model setup by applying the group means to their respective patients. The difference in the log-likelihood scores for each model is chi-squared distributed with the difference in number of parameters to be estimated as the degrees of freedom. The null hypothesis for the LRT is no difference between simpler and more complex models and this test can be used to compare different multilevel models as well so long as the simpler model is nested within the more complex model, i.e. the complex model contains one extra variable.

Another property of multilevel modelling is the shrinkage factor. This applies to the situation where some clusters have many patients and others have fewer. The shrinkage factor means that those groups with relatively fewer observations will be aligned more closely to the overall sample average whereas those groups who contribute many more observations will be aligned more closely to their group mean given the weight of evidence. Therefore, the groups with fewer observations will have their group mean weighted or shrunk towards the overall sample mean and this properly means that the multilevel model does not require all groups to contain the same number of observations.

If we can establish that a multilevel model framework is suitable for addressing our research questions then we can determine the amount of variation that is accruing to the differences between groups by calculating the variance partition coefficient (VPC – also known as the intra-class correlation (ICC)). The VPC measures the proportion of total variance that is due to variance among ESD/hospital teams. Total variance includes this between group-level variance plus the patient within group-level variance. This VPC measure is useful as it provides the percentage breakdown of total variance across the levels in the multilevel model indicating which level is explaining more of the variation with respect to the outcome variable.

Regarding the outcome variable itself, the multilevel model framework provides a suite of different models to suit different outcome variables in terms of their distribution. There exists a standard set of models to accommodate continuously/normally distributed outcome variables but the framework can also accommodate non-normally distributed responses such as logit for binary variables, ordered logit for ordinal variables and multinomial logit for nominal variables. This range of models means that the multilevel framework can accommodate different response type variables so that the framework need not be compromised even if a particular outcome variable cannot be treated as normal. Given its popularity among practitioners, multilevel modelling can be conducted in numerous statistical software packages such Stata, MLwiN, SPSS or R.

Aims and objectives

The overall aim of the WISE study is to identify what benefits there are for healthcare communities that have adopted ESD and to determine if the benefits of ESD suggested by randomised clinical trials (and why it was recommended in the first place), are realised in practice. This research study will investigate the impact of implementing ESD at scale in real world conditions and investigate which models of ESD are most effective in practice.

The objective of WP1 is to adopt a realist evaluation approach to address this stated aim through our three programme theories outlined above. We plan to test, modify and add to these programme theories by investigating how context influences the core components of ESD, what mechanisms are involved in delivering ESD and how this relates to what outcomes are achieved.

Research questions

To address these WP1 programme theories, we have devised a subset of research questions under one main research question as follows:

1. How effective is ESD when implemented at scale in practice?

   (a) What adopted models of ESD exist and how do these relate to evidence based recommendations?

   (b) Can realised benefits of implementing ESD be quantified by defined measures of effectiveness, i.e. reduction in length of hospital stay, responsiveness of the service, amount of rehabilitation delivered and changes in patient dependency?

   (c) What ESD model, site and patient level characteristics influence effectiveness of ESD services? Specifically:

      (i) Does the degree to which an ESD service has adopted an evidence based model relate to higher levels of effectiveness?

      (ii) Does rurality of the site in which teams are based determine variability?

      (iii) What patient characteristics explain variability in outcomes?

   (d) What are the cost-consequences of adopted ESD models?

Introduction to WP1 data

In collaboration with the Royal College of Physicians, WP1 will involve analysis of historical prospectively collected Sentinel Stroke National Audit Programme (SSNAP) data from hospital and community providers across the East Midlands, West Midlands, East of England and North of England strategic clinical networks (SCNs). This study is designed to evaluate the impact of different models of ESD operating over the four geographically defined SCNs. These data will be interrogated to compare a Midlands and East initiative to implement ESD beginning in 2012 [12] with a region of England (North of England) that has been slower to implement ESD based on SSNAP post-acute organisational audit data [3]. Sites within each of the four regions will be defined according to Clinical Commissioning Group (CCG) and Local Authority boundaries. All individual ESD teams within each site that participate in the SSNAP clinical audit will be included.

Introduction to analytical approach

A key aim of this work package is to investigate whether the degree to which an ESD service has adopted an evidence based model is related to better patient care (measured by ESD responsiveness and rehabilitation delivered) and patient dependency (Modified Rankin scale). Effects of ESD on length of hospital stay will also be investigated, however, this analysis will require a different calibration of the statistical model and will be dealt with separately.

Quasi-experimental, with respect to ESD being or not being offered on the care pathway, cross-sectional and before-and-after comparison group designs will be employed using multilevel model regression analysis. Multilevel modelling is an appropriate technique when analysing an outcome variable that is generated from a clustered/nested structure whereby the outcome under consideration is produced by patients in different hospital/ESD settings. Traditional multiple regression techniques that ignore clustering and treat patients as independent observations can lead to the standard errors of the regression coefficients being underestimated thereby overstating their statistical significance. In particular, standard errors for the coefficients of higher-level predictor variables, such as ESD service delivery core components, will be the most affected by ignoring the clustering of patients within different ESD teams.

In this WISE study, we are interested in measuring the effects that ESD teams have on stroke survivor outcomes. Measuring the effects that ESD teams have on their patients is a necessary first step to learning how ESD practices combine to generate differences between teams. Combining SSNAP clinical data at the patient level with organisational data at the ESD team level, we will attempt to measure the 'true' effects that ESD teams have on their patients by fitting two-level patients-within-teams multilevel models to patient outcomes where covariate adjustments are made for a range of patient and ESD team confounders. From these multilevel models, the ESD team-level residuals are then argued to measure the effects that these teams have on their patients having controlled for important confounding characteristics at each level. These effects are interpreted as measuring the influences ESD teams have on their patients' outcomes while they receive the ESD treatment.

The analytical approach for WP1 comprises four distinct analyses to address the research questions outlined above. These four analyses are as follows:

1. ESD models and levels of effectiveness

2. ESD impact on length of hospital stay
3. Difference-in-difference analysis
4. Cost-consequences of ESD models

With the first piece of analysis above, we are directly examining the effectiveness of ESD service provision whereby patients are nested within ESD teams. The second and third analyses focus on the impact of implementing ESD on hospital length of stay with patients nested within admitting hospitals controlling for ESD on the care pathway or not at the hospital level because the outcome variable is length of stay in hospital until discharge from acute services at the patient level. Teasing out any reduction in hospital length of stay thanks to implementing ESD would be beneficial for the commissioners of such services to learn hence why it is included within the WP1 analysis plan. Furthermore, the cost-consequence analysis is also beneficial for the commissioners to learn with respect to the financial implications pertaining to ESD provision. Therefore, the first piece of analysis directly evaluates the effectiveness of ESD whereas the subsequent three analyses indirectly evaluate its effectiveness through possible reductions in hospital length of stay and cost-savings.

Analysis and variable definitions

Multilevel modelling

Multilevel modelling is the preferred method of statistical analysis for addressing our WP1 research questions as it will allow us to explore the relationship between defined outcomes of interest with our predictor variables controlling for confounding variables likely to be associated with variation in outcomes. Specifically, multilevel modelling can appreciate the variation in outcomes as a mixture of patient variability nested within ESD service provision variability. Different variables will be included at different levels, i.e. predictor variables will refer to characteristics of the ESD service provision and confounding variables will refer to characteristics of the 'site' or geographical location of the ESD service and individual stroke patient characteristics.

Predictor variables when patients nested within ESD teams

We define predictor variables as those we are particularly interested in based on our research questions. We will investigate whether differences in levels of effectiveness of ESD teams can be explained by adoption of evidence-based core components. Detailed information about each ESD service type, collected by SSNAP in the 2015 post-acute organisational audit, will be collated by the research team. This will include information about ESD team composition, capacity, workload and organisational features, e.g. frequency of team meetings. The research team will cross-reference the organisational audit data with aggregate clinical audit data in the public domain to verify how long each ESD team has been operating for. This will provide the research team with a detailed description of the ESD model adopted by each ESD team.

ESD team organisation audit data will then be compared to ESD consensus statements and national clinical guideline recommendations to determine adoption of evidence based core components [1,2,7]. Following the findings of Fisher et al. [7] and Langhorne and Baylan [13] in terms of ESD team structure, practices and procedures for the purpose of evaluating ESD service delivery, we assume core members of the ESD team would provide the following minimum whole time equivalent (WTE) per 100 stroke patients commitment to the service; doctors 0.1, nurses 0.4, occupational therapists 1.0, physiotherapists 1.0, and speech and language therapists 0.3. We also assume that patients would have access to clinical psychologists, social workers and rehabilitation assistants as part of the ESD service. Furthermore, the ESD team should meet weekly, provide a stroke specific service more than 5 days a week commencing within 24 hours of the patient being discharged from hospital and provide training opportunities for members of the ESD team. All these assumptions are encapsulated in Table 1 below, whereby an ESD team receives a point for every assumption met or exceeded and no point if not. In order to evaluate the level of service provided by the ESD teams, we propose a scoring system based on Table 1 below that relates ESD organisational audit data directly with evidence-based core components as outlined above. We refer to this scoring system as a measure of ESD organisational strength. A big assumption made with this measure is that each component is equally weighted in terms of importance, i.e. doctors meeting their minimum recommended WTE per 100 stroke patients commitment of 0.1 gains one point as do rehabilitation assistants having training opportunities. In subsequent analysis, we will assess this assumption when evaluating these core components with respect to patient outcomes.

Based on our proposed scoring system, an ESD team can score a maximum of 17 points broken down by a maximum of 5 points for core team members meeting or exceeding recommended WTE level per 100 stroke patients, a maximum of 3 points for access to other team members, a maximum of 3 points for training opportunities, a maximum of 3 points for multidisciplinary team meetings and a maximum of 3 points for level of service provided.

	Yes	No
Core team members meeting or exceeding recommended WTE level per 100 stroke patients:		
Doctors ≥ 0.1	1	0
Nurses ≥ 0.4	1	0
Occupational therapists ≥ 1	1	0
Physiotherapists ≥ 1	1	0
Speech and language therapists ≥ 0.3	1	0
Access to other team members:		
Clinical psychologists	1	0
Social workers	1	0
Rehabilitation assistants	1	0
Training opportunities:		
Nurses	1	0
Therapists	1	0
Rehabilitation assistants	1	0
MDT meetings:		
Weekly meetings	1	0
Core team attend	1	0
ESD member attends acute meeting	1	0
Service:		

Stroke specific	1	0
Median waiting time between referral and ESD ≤ 1 day	1	0
Weekly service > 5 days	1	0

Table 1: Configuration of ESD organisational strength measure

The focus of this WISE study centres around 31 ESD services offered within 4 Strategic Clinical Networks (East Midlands, East of England, North of England and West Midlands). In terms of the ESD team core components, we have 17 individual items grouped into 5 components which can be aggregated into one total score measure. Core component scores will be calculated to reflect different aspects of the adopted ESD service model e.g. team composition (core team and others), staff training, team meetings and service specificity. We intend to analyse this core component score in three different ways:

1. Total score
2. Disaggregated by component
3. Individual item

Taking them as ordered above, we will assess the significance of each element from most aggregated to most disaggregated to evaluate which components/individual items have the greatest impact on patient outcomes. This analysis may also enable the grouping of ESD teams based on ESD service types for comparison purposes.

Therefore, we will apply this scoring system to the 31 ESD services in accordance with the aforementioned three ways we intend to analyse the scores. By way of validating the recommended consensus guidelines on WTE units per 100 stoke patients, we will also include two alternative core team scores whereby we replace the WTE recommendation with (1) patients simply having access to each core team member (one point for yes and zero point for no) and (2) the recorded amount of WTE units per 100 stroke patients for each core team member. We believe these alternative formulations may prove useful in terms of conducting a sensitivity analysis with respect to our proposed measure of organisational strength and in relation to patient outcomes. We believe this analysis can help empirically validate the current consensus regarding recommended delivery of an effective ESD service.

Predictor variables when patients nested within hospitals

When we turn our attention to the hospital length of stay and difference-in-difference analyses, the set of predictor variables will change to reflect the fact that for these analyses patients will be nested within hospitals rather than ESD teams and hence different datasets will be used. Hospital characteristics of interest are type of hospital (e.g. (non-)routinely admitting (acute)), grade/score of hospital as recorded by the SSNAP team and a measure of delayed transfers of care from hospital, derived from the Adult Social Care Outcomes Framework (ASCOF), to account for influence of provision of social care [14]. The ASCOF data reports the average daily rate of delayed transfers of care per 100,000 population aged 18 and over at Local Authority (LA) level. We can aggregate these data up to NHS Trust level using averages where multiple LAs are situated within the one NHS Trust and then we can map these rates onto the hospitals in our sample as each hospital belongs to an NHS Trust. Starting from the admitting hospital, we will control for whether or not ESD service provision is on the care pathway for stroke patients. Where ESD provision is on the care pathway, we will explore the variation (if any) in hospital length of stay with respect to the variation in ESD service provision in accordance with our previous analysis. We will also adjust for level of rurality and deprivation at the hospital level as per the ESD team analysis plus include the same patient case-mix variables at the patient level given the outcome variable for these analyses will be hospital length of stay at the patient level.

Confounding variables

At the ESD team level these confounding variables will include level of rurality [15] and deprivation [16] characteristics based on Clinical Commissioning Group (CCG) area statistics. Where an ESD team covers multiple CCGs then the weighted average level of rurality and deprivation will be calculated for those teams with the weights being derived from the NHS Quality and Outcomes Framework (QOF). The QOF records prevalence of stroke and transient ischaemic attack by CCG so we can use these data as a proxy for proportion of stroke activity in each relevant CCG rather than simply assuming the ESD teams covering multiple CCGs do so equally. The grade/score of each discharging hospital as recorded by the SSNAP team will also be taken into consideration as an indication of level of inpatient care before being referred to an ESD service. For ESD teams with multiple discharging hospitals, a weighted average SSNAP score will be calculated based on the number of patients being discharged to those ESD teams. Furthermore, a measure of the number of WTE units per member of staff will be calculated for each ESD team to provide a relative ratio of dedicated time given the number of staff in the ESD team [11]. This time to staff ratio is treated as a confounder as it does not form part of the recommended ESD service criteria outlined above but provides a way of acknowledging that the makeup of the teams can vary in terms of staff size and we are interested to learn if having more or less staff for a given number of WTE units has any bearing on types of ESD model adopted and associated levels of effectiveness. The same logic applies to the level of rurality and deprivation as we appreciate that these contextual confounders are not homogenous across the ESD teams and may impact upon service provision and patient outcomes.

In order to account for comparison between different groups of individual patients, stroke patient characteristics, collected as part of the SSNAP data set, will include age at admission, sex, pre-stroke independence, comorbidities, NIH stroke scale score on admission, type of stroke and modified Rankin score at discharge from hospital. These reflect previously validated stroke case-mix models [17,18].

Exploring effect modification

Further to the confounders outlined above being important controls when evaluating ESD service provision on patient outcomes, they may also modify the effect of ESD service provision on patient outcomes. Therefore, we can test for this by adding interaction effects between the ESD service variables and the confounders of interest while retaining the main effects in the model. Effect modification may occur even if the main effect of the confounder is found to be insignificant as the aggregate effect from the different levels of the confounding variable may cancel each other out, e.g. very rural and urban areas may produce opposite effects with respect to ESD service provision but appreciating this difference may only become apparent by interacting the different levels of rurality with the different levels of ESD service provision. Moreover, the multilevel modelling framework allows for cross-level interactions so we can explore if service provision also depends on patient case-mix confounders.

Outcome measures

An effective ESD service should be responsive (no waiting time from discharge to first contact with the patient) and deliver intensive stroke rehabilitation by a multidisciplinary team [1]. Clinical trials showed that effective ESD reduced length of hospital stay and reduced patient dependency [3,4]. Therefore, the effectiveness of ESD will be measured with evidence based metrics as defined by national clinical guidelines for stroke, and that reflect outcome measures used in the original ESD trials [4,5].

Effectiveness of ESD service models will first be measured with the following outcome measures:

(i) time to first assessment from hospital discharge,
(ii) number of minutes of therapy (occupational therapy, physiotherapy, speech and language therapy) delivered,
(iii) Modified Rankin scale (measure of patient dependency) after ESD delivered, and
(iv) length of hospital stay

Analysis 1: ESD models and levels of effectiveness

ESD services that are distinguished by their core components as well as their exposure to rurality, deprivation, SSNAP grade/score of discharging hospital and WTE commitment given team size will be compared using defined outcome measures of effectiveness.

Historical data collected by teams prospectively over a one year period (1 January 2016 to 31 December 2016) will be used. This aligns with 2015 SSNAP post-acute organisational data used to distinguish ESD model type. Effectiveness of different ESD models of provision, i.e. adoption of evidence based components, over the one year period will be measured with the following patient outcome measures: (i) time to first assessment from hospital discharge (ii) number of minutes of therapy (occupational therapy, physiotherapy, speech and language therapy) delivered (iii) Modified Rankin scale (measure of patient dependency) after ESD delivered. Adjustment will be made for patient characteristics such as age at admission, sex, pre-stroke independence, comorbidities, NIH stroke scale score on admission, type of stroke and modified Rankin score at discharge from hospital.

Analysis 2: ESD impact on length of hospital stay

A key benefit of ESD identified in the original randomised controlled trials was a reduction in length of hospital stay [4,5]. Hence we have also included this measure in our proposed analysis. A quasi-experimental cross-sectional design will be first be used in which admitting hospitals that have either directly or indirectly adopted ESD along their care pathway are compared with admitting hospitals that have not and different models of ESD service are compared. Furthermore, the hospital acute stroke units that refer stroke patients to each of the ESD teams selected for inclusion in this study will be identified (based on SSNAP clinical data). These will be categorised for comparison purposes in relation to the ESD team it refers patients to (and the adopted ESD model, as explained previously). We will also identify hospital acute stroke units within each region that do not refer to any ESD service to allow us to include non-ESD sites, as a further comparator, in the analysis.

Historical data collected prospectively over a fourteen month period will be used (1 September 2015 to 31 December 2016, dates relate to hospital rather than ESD admission). Initially, data from all stroke patients admitted to each hospital stroke unit during the 12 month time period time will be included. However, not all patients who are discharged from hospital access ESD and hence there will be a proportion of patients whose length of hospital stay in unlikely to be affected. Therefore we will also perform a further analysis on 'ESD-related' length of hospital stay for each hospital with a subgroup of patients who have similar case-mix characteristics but differ with respect to accessing (or not) ESD services upon discharge from hospital.

Analysis 3: Difference-in-difference analysis

An additional analysis using a before-and-after design will then be conducted to further investigate the magnitude of ESD impact on length of hospital stay. Difference-in-difference analysis enables investigation of differences in length of hospital stay over time and comparison between hospitals that have implemented ESD with hospitals that have not, and between different models of ESD. Hence this analysis requires use of patient-level hospital data collected before and after ESD services have begun to be implemented.

We will restrict difference-in-difference analysis to 'more recently' introduced ESD services. The post-ESD (i.e. ESD services have begun to be implemented) time period for all sites included in the analysis will be from 1 Sept 2015 to 31 Dec 2016 (in line with previous analyses). The pre-ESD (i.e. ESD services have not yet been implemented) time period will be a fourteen month period prior to the earliest start date of all ESD services included in the analysis.

Analysis 4: Cost-consequences of ESD models

Cost-consequence analyses have been recommended for complex interventions that have multiple effects as they offer a more flexible approach to presenting costs and benefits alongside each other rather than combing these into a single measure [19]. This approach also fits with the realist approach of the WISE study, in that costs (staff/travel/equipment) associated with ESD related mechanisms will be considered in light of context in which they are operating and the outcomes or benefits that are achieved.

Cost estimates will be calculated from a health service perspective, focusing on core components of ESD services (featured in previous analyses) that are associated with costs of interest to healthcare commissioners and providers [20]. Travel costs associated with delivery of rehabilitation will be estimated by defining the geographical area over which the ESD service operates, determining average distances travelled and number of patient visits made. Using patient caseload information, direct costs per patient will also be calculated. This will enable us to present associated cost estimates in relation to adopted ESD model types and the geographical context in which they are operating. The main benefits (treatment effects) will be obtained from differences in length of hospital stay and changes in modified Rankin derived from the comparison of ESD model types (described above). This final piece of analysis will overlap with WP2 as inquiring about cost of ESD provision will form part of the data collection when visiting the six selected sites.

Context-Mechanism-Outcome configurations from WP1

Using findings from these analyses conducted in WP1, CMO patterns relating to the delivery, cost and effectiveness of ESD will be identified. These will be compared to our initial programme theories and modifications will be made in light of the findings.

Addressing methodological issues

Multicollinearity

In the instance where model variables are found to be strongly related rendering the possibility of parameter estimates to be unreliable, we will conduct a sensitivity analysis on the affected variables removing those with lower relative importance with respect to our outcome variables. Moreover, site-level variables will be chosen according this study's CMO programme theories rather than control for all site-level variables in the same model thereby reducing the risk of failing to uncover any significant ESD team variability with respect to the patient-level outcome variables.

Confounding variables

Given the complexity of evaluating ESD service provision in real world settings, adequately covering all important confounders relies on the quality of the data available for WP1. The danger is that by not appropriately covering the important alternative considerations with respect to patient outcomes we may overestimate the effect of ESD service provision on patient outcomes. Therefore, we will endeavour to factor in all relevant information into our analyses and be clear when this is not possible. Moreover, WP2 may help identity further confounding issues not addressed in WP1 but these will be incorporated into our programme theories and acknowledged as further considerations when evaluating ESD service provision in real world settings.

Sample size

Sample size and power calculations are complex in multilevel modelling. This is because if subjects within a cluster are more similar than between clusters, there is a net loss of data and if the degree of similarity is high, i.e. if the ICC is large, then there may be far fewer subjects under observation from a statistical perspective. Statistical approaches must take into account the potentially clustered nature of the data collected (patients within one service potentially being more like one another than patients between services). Although we have not adopted a cluster-randomised controlled trial design, similar approaches when considering sample size and power calculations, are applicable for this study design [21,22].

In addition to conventional sample size calculation requirements, an ICC coefficient is required to represent the degree of similarity within a cluster. The design effect, which is calculated using the ICC and average cluster size, is used to estimate the extent to which the sample size should be inflated to accommodate for the homogeneity in the clustered data. Therefore, multiplying the conventional sample size by design effect gives the total sample size [22]. In this study, clusters relate to ESD teams and hospital stroke units for length of hospital stay analysis. Cluster size is determined by the number of patients associated with each ESD or hospital team during the time period involved.

Using a sample size calculation for multiple regressions, without clustering, the minimum total sample size required with a power of 80%, an alpha of 0.05 and a medium effect size of 0.25 would be 158. This figure is lower than the average cluster size of 190 based on currently published SSNAP data [23,24]. Based on previously published ICC recommendations and taking into account our outcomes of interest, we assumed an ICC of between 0.05 and 0.12 to allow for the combination of process and patient outcome measures. This resulted in a design effect of between 10.5 and 23.7. Current annual admission figures from SSNAP suggest we would have an average cluster size of 190 patients within 25 ESD teams and 640 patients within 34 hospitals thereby producing approximate total sample sizes of 4750 patients for ESD data and 21760 patients for hospital data. Deflating these total sample size numbers by the design effects produces an effective sample size range of between 201 and 455 for the ESD analysis and between 919 and 2082 for the hospital analysis. These deflated numbers represent the 'real' sample sizes after accounting for loss of data due to assumed within-cluster similarity.

References

1. Royal College of Physicians Intercollegiate Stroke Working Party. *National Clinical Guideline for Stroke.* Royal College of Physicians 2012.

2. The National Institute for Health and Care Excellence. *Stroke Rehabilitation in adults. NICE guidelines [CG162].* 2013.
3. Royal College of Physicians Intercollegiate Stroke Working Party*. Sentinel Stroke National Audit Programme (SSNAP). Post-acute organisational audit. Phase 2: Organisational audit of post-acute stroke service providers.* Royal College of Physicians 2015.
4. Langhorne P, Early Supported Discharge Trialists. Services for reducing duration of hospital care for acute stroke patients. *The Cochrane database of systematic reviews.* 2005 (2):CD000443.
5. Fearon P, Langhorne P, Early Supported Discharge Trialists. Services for reducing duration of hospital care for acute stroke patients. *The Cochrane database of systematic reviews.* 2012 (9):CD000443
6. Pawson R, Tilley N. *Realist Evaluation*. London: Sage Publications Ltd.; 1997.
7. Fisher RJ, Gaynor C, Kerr M, Langhorne P, Anderson C, Bautz-Holter E, et al. A consensus on stroke: early supported discharge. *Stroke.* 2011;42(5):1392-7.
8. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation science*. 2009;4:50.
9. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *The Journal of continuing education in the health professions.* 2006;26(1):13-24.
10. Fisher RJ, Cobley CS, Potgieter I, Moody A, Nouri F, Gaynor C, et al. Is Stroke Early Supported Discharge still effective in practice? A prospective comparative study. *Clinical rehabilitation.* 2016;30(3):268-276.
11. Ariss SM, Enderby PM, Smith T, Nancarrow SA, Bradburn MJ, Harrop D, et al. Secondary analysis and literature review of community rehabilitation and intermediate care: an information resource. *Health Services Delivery Research.* 2015;3(1).
12. NHS Midlands and East. *NHS Midlands and East Stroke Service Specification.* 2012. Available from: http://emsenate.nhs.uk/downloads/documents/cardiovascular%20disease/stroke/final-stroke-spec.pdf
13. Langhorne P, Baylan S, *Early Supported Discharge Trialists. Early supported discharge services for people with acute stroke*. Cochrane Database of Systematic Reviews 2017, Issue 7
14. Department of Health. *Measures from the Adult Social Care Outcomes Framework, England – 2014-15, Final release*. Available from: http://digital.nhs.uk/catalogue/PUB18657 [Accessed October 06 2015]
15. Office for National Statistics. *2011 Rural-Urban Classification*. Available from: www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/2011-rural-urban/index.html [Accessed January 06 2016]
16. Department for Communities and Local Government. *English indices of deprivation 2015*. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 [Accessed 30 September 2015]
17.Turner M, Barber M, Dodds H, Murphy D, Dennis M, Langhorne P, et al. Implementing a simple care bundle is associated with improved outcomes in a national cohort of patients with ischemic stroke. *Stroke.* 2015;46(4):1065-70.
18. Teale E, Forster A, Munyombwe T, Young JB. A systematic review of case-mix adjustment models for stroke. *Clinical rehabilitation.* 2012;26(9):771-786.
19. Craig LE, Wu O, Bernhardt J, Langhorne P. Approaches to economic evaluations of stroke rehabilitation. *International Journal of Stroke.* 2014; 9(1):88-100.
20.Gage H, Kaye J, Owen C, Trend P, Wade D. Evaluating rehabilitation using cost-consequence analysis: an example of Parkinson's disease. *Clinical rehabilitation.* 2006; 20(3):232-238
21. Aspinal F, Beresford B, Faria R, Kanaan M, Laver-Fawcett A, Parker G, et al. *Models Of Reablement Evaluation: a mixed methods evaluation of a complex intervention (The MORE project).* HS&DR proposal – reference number 13/01/17. 2014.
Available from: http://www.nets.nihr.ac.uk/__data/assets/pdf_file/0003/129630/PRO-13-01-17.pdf
22. Lancaster GA, Chellaswamy H, Taylor S, Lyon D, Dowrick C. Design of a clustered observational study to predict emergency admissions in the elderly: statistical reasoning in clinical practice. *Journal of Evaluation in Clinical Practice.* 2007;13(2):169-178.
23. Thomas RE, Grimshaw JM, Mollison J, McClinton S, McIntosh E, Deans H and Repper J. Cluster randomized trial of a guideline-based open access urological investigation service. *Family Practice* 2003; 20: 646–654.
24. Eccles M, Grimshaw J, Steen N, Parkin D, Purves I, McColl E et al. The design and analysis of a randomized controlled trial to evaluate computerized decision support in primary care: the COGENT study. *Family Practice* 2000; 17: 180-6.

Appendix overleaf

Figure 1: Initial programme theory configurations for WP1 relating to ESD service provision