Project Deliverable Report

Deliverable D2 – Report on the likely cost of the various prototype interventions based on a model of the likely costs

Work Package 6

Task Report on the likely cost of the various prototype interventions based on a model of the likely costs

Date of delivery Contractual: 01-11-2009 Actual: 30-10-2009

Code name Version: 1.0 Draft □ Final X

Type of deliverable Report

Security (distribution level) Public

Contributors UMCU, FAD, BHAM, CMJ/NCQA, ASL 10 FIRENZE, RUNMC, OUNL, KI

Authors (Partner) Prof. Richard J. Lilford
Mr. Peter J. Chilton
Dr. Karla Hemming
Dr. Celia A. Brown
Mr. Alan J. Girling
Prof. Paul Barach
Miss Wendy Fox-Kirk

Contact Person Professor Richard J. Lilford

WP/Task responsible WP 6

EC Project Officer Dr. J. Paehler

Abstract (for dissemination) A report providing a model for the assessment of cost effectiveness of organisational interventions such as handover in healthcare settings. The proposed model is an extended Donabedian Chain. Upstream interventions have diffuse effects creating problems with statistical power if downstream end points are measured. The difficulty of measuring the effect of upstream interventions such as that for handover is discussed and the conclusion reached that if an intervention is relatively inexpensive then worthwhile (cost effective) effects will not be detectable at the level of the patient. The importance of assessing the value of information is also discussed with Bayesian techniques suggested as the key methodology to address this issue.

Keywords list Handover, cost-effectiveness model, intervention, Bayesian techniques
Executive Summary

Poor continuity of clinical care (with multiple provider involvement), either at a patient’s referral to a hospital by a primary care specialist or at a patient’s discharge from the hospital (further jointly referred to as ‘handover’), is a critical aspect of a patient’s care. Incomplete handovers to a secondary/tertiary care unit or discharge from hospital can lead to adverse events for patients that may ultimately lead to either life threatening situations during treatment/surgery or avoidable treatment and/or re-hospitalizations after the patient’s discharge.

Avedis Donabedian famously described a causal chain linking structure, process and outcome. We propose a more fine-grained subdivision of process into: generic processes, targeted processes and clinical processes. The deconstructed chain provides a tool to conceptualise the knock-on effects of interventions at one level on downstream end points. The purpose of this deliverable is to consider the implications of these relationships for selection of the ‘primary’ end points that determine the sample size to evaluate the potential handover practices that this project is developing and evaluating.

We found that:

1) Targeted service interventions have a greater effect on clinical processes than on contingent outcomes. Process is frequently a more cost-effective primary end point than outcome.

2) The further to the left an intervention is applied in the causal chain, the more diffuse are its potential effects on clinical processes. However, many of these processes converge on a small number of patient outcomes. Thus generic interventions, in contrast to targeted service interventions, are more likely to yield detectable changes in outcomes than in clinical processes.

3) The magnitude of effects that generic interventions have on outcomes might nevertheless be small, but such interventions may be cost-effective if their net excess costs are also small. Studies powered to detect small changes in patient level end points are not value for money when large changes in end points are neither plausible nor necessary to justify the use of the intervention. In these circumstances studies should be powered on surrogate upstream outcomes.

4) Bayesian methods are useful in both determining the “Value of Information” for studies of different sizes and in the analysis of data from different sources and of different types.

Developing and justifying the implementation of effective handover service interventions, either systemic, educational or regulatory, requires generating economic evidence to inform policy and practice about the costs of ineffective handovers and the estimate of the economic impact of the tools to improve the patient handover process.
Table of Contents

1. INTRODUCTION .................................................................................................................. 3
   1.1 AN EXTENDED DONABEDIAN CHAIN ........................................................................... 3
   1.2 CLASSIFICATION OF SERVICE DELIVERY AND POLICY RESEARCH ......................... 4
   1.3 FOCUS OF THIS REPORT ............................................................................................. 4

2. MEASURING THE EFFECTIVENESS OF TARGETED SERVICE INTERVENTIONS ........... 5
   2.1 DOWNSTREAM END POINTS ...................................................................................... 5

3. MEASURING THE EFFECTIVENESS OF GENERIC PROCESS AND STRUCTURAL
   INTERVENTIONS .............................................................................................................. 7
   3.1 DOWNSTREAM END POINTS ...................................................................................... 7
   3.2 NARROW VS. DIFFUSE EFFECTS .............................................................................. 7
   3.3 SELECTING END POINTS ........................................................................................... 7

4. COST-EFFECTIVENESS OF STUDIES OF POLICY AND GENERIC SERVICE INTERVENTIONS
   ........................................................................................................................................... 8
   4.1 WORTHWHILE EFFECTS MAY NOT BE DETECTABLE BY MEASUREMENTS OF EITHER
       CLINICAL PROCESS OR OUTCOME ............................................................................. 8
   4.2 WHAT CAN BE DONE WHEN WORTHWHILE EFFECTS CAN NOT BE DETECTED AT THE
       PATIENT LEVEL? .......................................................................................................... 9
   4.3 VALUE OF INFORMATION .......................................................................................... 10

5. APPLICATION OF THE MODEL TO HANDOVER .......................................................... 11

6. DISCUSSION ...................................................................................................................... 13
   6.1 MULTI-COMPONENT INTERVENTIONS ....................................................................... 13
   6.2 SPILLOVER EFFECTS .................................................................................................. 13
   6.3 BAYESIAN METHODS ................................................................................................ 13
   6.4 REPRESENTATIONAL NATURE OF THE MODEL PRESENTED IN THIS REPORT ............. 14

7. CONCLUSION AND RECOMMENDATIONS ................................................................... 14

REFERENCES .......................................................................................................................... 16

APPENDIX .............................................................................................................................. 22

Table 1. Total sample sizes for the study of change in compliance of a clinical process (stroke
algorithm) vs. outcomes contingent on compliance (death or permanent disability) .............. 22
1. Introduction

Explanatory Note on Variation to Deliverable 2

In Annex 1 – Description of Work for the grant agreement for this collaborative project, Deliverable 2 (D2) is described as follows:

D2 – Report on the likely cost of the various prototype interventions, based on a model of the likely costs.

This report addresses D2 in part rather than in whole and concentrates on the development of a model to derive the likely costs and cost effectiveness of handover interventions. The original desire to apply the model to actual prototype handover interventions will be achieved once the data from the Handover project are available to inform the design of new handover prototype interventions. The model described within this report sets out the key issues including major methodological difficulties in achieving a satisfactory assessment of the cost effectiveness of upstream interventions with a diffuse effect in complex organisational (socio-technical) environments such as the interface between the hospital and the community. Full cost effectiveness analyses for new handover interventions arising from this project will be provided in the next deliverable due from work package 6 (D7) in month 25.

There is a broad consensus that treatments should be evaluated by means of clinical trials, using patient outcomes as end points. However, methods for evaluation of service delivery (management) and policy interventions are contested. In this report we consider one aspect of study design for these complex interventions – the selection of the primary end point. By primary end point we mean the end point that will be given the most weight in the interpretation of study results and around which the calculations of sample size will be framed. Our analysis begins with a consideration of the types of end points that may be used.

1.1 An extended Donabedian chain

Avedis Donabedian conceptualised a causal chain linking structure, process and outcome (1). This model provides an idealised framework for the design and interpretation of studies to evaluate the effectiveness of service delivery or policy interventions (2-4). We explore the implications of a more fine grained model in which the process level is divided into three further categories or sub-levels (5). The clinical processes sub-level is the category closest to the patient (encompassing treatments such as drugs, devices, procedures, “talking therapy”, complementary therapy and so on) (6). Further to the left in the causal chain we identify a targeted processes sub-level, comprising processes targeted specifically at improving particular clinical processes (examples include training...
in the use of a device and a clinical rule built into a computer system). Further left still we distinguish a generic processes sub-level (for example, a novel handover intervention). Structure lies at the extreme left of the chain. We will use ‘structure’ to describe factors such as the reimbursement method and provision of major facilities, which are controlled by policy makers rather than managers or clinicians working within organisations. A more fine grained classification would be possible (a point to which we return later) but the levels described are sufficient for the purposes of the analysis that follows. The effects of interventions impact at all levels downstream of an intervention (2).

1.2 Classification of Service Delivery and Policy Research

An intervention may be applied at any level in the extended chain with the exception of the extreme right (outcome) level. Service delivery and policy interventions may therefore be heuristically classified into four groups, according to the level at which they are applied:

1. Clinical interventions
2. Targeted service interventions
3. Generic service interventions
4. Structural (policy) interventions

The study of clinical interventions (treatments) is typically called Health Technology Assessment (HTA), while the study of the effects of interventions at the other end of the chain – the structural level – falls under policy research. However, evaluations of interventions at the two middle levels (targeted and generic process interventions) tend to be lumped together as part of management research. Service delivery and Organisational (SDO) research or health services research. Part of the message of this report is that these portmanteau rubrics cover very different types of research from a methodological point of view, while the interventions at the generic process and structural levels have many features in common.

1.3 Focus of this report

It is widely accepted that service delivery or policy interventions (such as handover) should undergo careful ‘pre-implementation’ evaluation in order to maximise prospects for success and minimise foreseeable risk (1;7). When an intervention is rolled out in practice, a comprehensive evaluation may have a number of components (2;8-11):

1) Base-line observations at all levels in the extended causal chain so that context can be described.
2) Monitoring changes at the level of the intervention to observe fidelity of uptake or show how it was modified in practice (3).
3) Observations of (any) downstream changes to reflect the effects of the intervention (3).
4) Both quantitative and qualitative observations across the causal chain to better understand not only what happened but why – that is to promote theoretical understanding (12).
In this report we focus on but one aspect of a comprehensive evaluation framework of a novel handover intervention, by examining the quantitative relationships between the various downstream levels in the extended version of Donabedian’s chain (component 3 above). It is not our purpose to critique different study designs – observational vs. experimental, single site vs. multi-site etc – and this issue will be considered only insofar as it may affect the above quantitative relationships. The merits and limitations of the various designs have been considered in detail elsewhere (13).

2. Measuring the Effectiveness of Targeted Service Interventions

2.1 Downstream end points

Here we consider “near patient” service delivery interventions, such as implementing specific management interventions targeted at improving staff compliance with a particular clinical process. For example, immediate thrombolysis is an effective treatment for a patient with thrombotic stroke (14), but it is seldom used (15). A targeted service intervention might involve writing and disseminating a local protocol for stroke care along with measures to ensure 24 hour availability of CT scans and thrombolytic medicines.

There are two downstream levels at which effects resulting from targeted service interventions may be observed – clinical processes (e.g. proportion of eligible patients who receive timely thrombolysis), and outcome (e.g. proportion of patients who recover from stroke). Ideally, both process and outcome would be measured. However, selecting a sample of sufficient size to measure plausible changes in both clinical processes and outcomes risks wasteful redundancy given an established link between the clinical process and the corresponding outcome. If the premise that the treatment is effective does not hold, then the logical response is to first establish the validity of the process by means of a clinical trial – itself a special form of targeted service intervention (16). Given a clinical process of established effectiveness, the most cost-effective type of ‘primary’ end point (clinical process or outcome) should be selected. The costs of a study are then related to:

1) Sample size (number of participating centres and the number of the patients sampled in each centre), and
2) The cost of making each observation.

Changes in clinical outcome rates (such as mortality or infection rates) can never be bigger than changes in the clinical error rates on which they are contingent and are usually much smaller. The required sample sizes increase in a non-linear fashion as the postulated effect size decreases. Larger, often much larger, samples are therefore required to detect changes in outcome rates than are needed to detect changes in the corresponding clinical process. The sample sizes needed, for example to demonstrate improvements in the clinical process and outcomes of stroke care are compared by type of study in Table 1 (See appendix). To show differences in outcomes the sample size required might be some
two orders of magnitude larger than those needed to demonstrate changes in the clinical processes on which they are contingent. Since the costs of a study are a function of sample size, the consideration of sample size favours use of clinical process over outcome as the ‘primary’ end point in studies of targeted service interventions when the validity of the process as a proxy is not in dispute.

However, the study costs are also a function of the cost of making each observation. Clinical processes are generally more difficult and expensive to measure than outcomes. Many important outcomes such as mortality and infection rates are collected routinely by health service surveillance systems. However, such systems seldom carry the numerator (process failure) and denominator (opportunity for failure) data from which rates of process failure/error should be constructed (20). It is necessary to harvest this information from case notes, bespoke data collection pro formas or direct non-participant observations (21). The cost of making reliable measurements of clinical process failures may be considerable, depending on the process concerned (22).

There are thus competing forces at work when evaluating a targeted service intervention; the generally higher cost per case of measuring clinical processes is in tension with the greater number of cases that must be sampled to measure outcomes with commensurate precision. The greater the size (in absolute terms) of the hypothesised effect on outcome and the more expensive the process for collection of salient clinical processes, the stronger becomes the argument to rely on outcome measures. For example, an influential study to assess the impact of targeted processes to reduce infection associated with central venous lines was based on the measurement of blood stream infections – the outcome of interest (23). The direct observation of clinical processes would have been very expensive and substantial effects on infection rates were expected (and observed) (23). However, as the signal (change in outcome due to intervention) diminishes in relation to the noise (changes in outcome due to uncontrolled sources of variation), a study based on process measurement will become more cost-effective. Such is the case in Landrigan’s study where the effects of fatigue on the quality of care delivered by medical interns in the Intensive Care Unit (ICU) were ascertained by means of direct observation (24).
3. Measuring the Effectiveness of Generic Process and Structural Interventions

3.1 Downstream end points

Generic process interventions, such as an improved handover process, have the potential to affect targeted processes, clinical processes and outcomes. They may impact on clinical processes through targeted service processes or indirectly through intervening variables (such as morale, sickness absence, ‘culture’, knowledge) (25), so there are four downstream 'levels' available as evidence of their effects. Moving yet further to the left, structural interventions (such as building a new hospital, increasing reimbursement rates, or acquiring ‘teaching’ status) can exert effects through a yet greater number of downstream levels and intervening variables; five in all.

3.2 Narrow vs. diffuse effects

The further to the left an intervention is applied in the chain, the greater the number of downstream processes that may be affected. For example a targeted service intervention, such as a forced function to prevent misconnection of oxygen delivery pipes, may be expected to affect only one clinical process – gas delivery, i.e. a narrow or coupled effect. However, a generic process intervention (such as a system of “appraisal” for all staff) or a structural intervention (such as augmenting resources to improve the nurse to patient ratio) has the potential to impact on a myriad of clinical processes across an institution – a diffuse effect. However, these clinical processes converge on certain discrete clinical outcomes. For example, mortality rates and quality of life scores are the final common pathway for a large numbers of discrete clinical processes ranging from stroke care to cancer management.

3.3 Selecting end points

With hundreds or thousands of downstream clinical processes, each of which may be affected to some degree, it may be quixotic to seek effectiveness of an intervention with diffuse effects by making observations of clinical processes. For example, Donchin et al. estimated that patients in intensive care units experience a mean of 178 clinical processes per day (26). In an organisation such as a hospital there would be many thousands of actions that could be affected by a change such as the ratio of physicians to patients. The effect on each clinical process might be so small that large, perhaps impracticaly and unaffordably enormous, samples would be required to avoid high probabilities (or the near certainty) of false null results. It would be logistically taxing to enumerate compliance with all (or even a meaningful proportion) of the clinical processes that might be affected downstream. An alternative might be a ‘holistic’ assessment of process that assesses the quality of care implicitly. For example, by asking experts to review clinical case notes. However, such methods have low reliability (22) and are expensive, since they depend on large amounts of clinical expert time. Given limited resources, it will
often make more sense to determine the sample size on the basis of outcomes on which large numbers of processes converge and which often can be measured at low cost. Patient outcomes also encapsulate the net effect of generic interventions on many individual processes, some of which may be affected in a negative direction. That is to say, the intervention may have various positive and negative effects across processes and these are consolidated among a limited number of outcomes.

A word of caution must be expressed at this point. Although multiple clinical processes converge on a limited number of outcomes, it does not follow that the effect on even these end points will be of sufficient magnitude to be observed in studies of affordable and manageable size – a possibility that will now be considered.

4. Cost-effectiveness of studies of policy and generic service interventions

4.1 Worthwhile effects may not be detectable by measurements of either clinical process or outcome

The sensitivity of outcomes to various structural and generic process interventions is likely to be highly variable. For example, a 50% increase in the nurse to patient ratio may achieve a large improvement in outcomes when compared with instituting a system of staff appraisal. The results of specimen calculations on the sample sizes required to detect different magnitudes of true improvement in end points are shown under different study designs. It can readily be seen that, irrespective of study type, large samples are needed to demonstrate an absolute change in an end point of less than five percentage points at an 80% statistical power threshold. When a change in end point is less than 2 percentage points (for a baseline of 10%), around 20,000 observations are required for a controlled before and after study design. However, effects of small magnitude are not necessarily nugatory. Other things being equal, the more expensive an intervention, the greater the size of effect needed to justify its use. The National Institute for Health and Clinical Excellence (NICE), the organisation responsible for policy-level resource allocation decisions in England and Wales, uses a heuristic maximum of approximately £30,000 (~€34,500; US$50,000) for a healthy life year (27). The maximum cost below which interventions of different effect sizes could be judged cost-effective by NICE is illustrated by a hypothetical case in which a patient stands to gain five years of healthy life if the intervention is successful. These calculations show that there is a sizable gap between the size of effect necessary for an inexpensive intervention to be deemed cost-effective and the size of effect that can be detected reliably by studies of manageable and affordable size. For example, assuming a baseline of 10% mortality, an intervention whose cost, when averaged out over a cluster, amounts to £300 (~€350; $500) per patient would have to save only 0.2 lives (of 5 years mean duration) per 100 patients admitted to an institution to be cost-effective (ignoring discounting and assuming lives saved in perfect health):
Cost per life saved = \frac{\text{Cost per patient} \times \text{Number of patients}}{\text{Number of lives saved} \times \text{Mean duration of a saved life}}

\begin{align*}
\text{Cost per life saved} &= \frac{300 \times 100}{0.2 \times 5} \times \frac{30,000}{1} = 30,000
\end{align*}

This shows that even a simple before and after study would require over 700,000 patients to detect an effect of this magnitude. Nationwide studies or international comparisons based on routinely collected data might provide such samples, but tend to leave cause and effect explanations for any observed difference unresolved. This is especially so when effect sizes are small, since even small biases drive proportionately greater inaccuracy under these circumstances (28).

The costs of studies needed to demonstrate an effect of a given magnitude are greater for studies with contemporaneous controls than for studies without such controls. A trade-off exists between accuracy and precision given the assumption that studies with contemporaneous controls are less prone to bias (13). The uncomfortable conclusion reached is that valid studies of sufficient precision to observe worthwhile effects may be unaffordable or logistically impossible to conduct under some circumstances. Inexpensive generic service innovations of low perceived risk, such as an explicit employee sickness absence policy or improving access to continuing education, might be cost-effective at a level of effectiveness that cannot be detected through patient level observations, i.e. observations of either outcomes or clinical processes.

4.2 What can be done when worthwhile effects can not be detected at the patient level?

The above ‘inconvenient truth’ raises the question of what should be done in such cases, given that decisions about policy and service delivery interventions still have to be made – interventions have to be implemented or abandoned; if implemented they may or may not be evaluated. If evaluated the sample size may be determined on the basis of patient level end points or points further upstream. We take our clue from Walter Charleton who said that, “The ‘reasonable man’ will not require demonstrations or proofs that ‘exclude all dubiosity, and compel assent,’ but will accept moral and physical proofs that are the best that may be gained.” (29). In circumstances where cost-effective interventions are unlikely to produce measurable effects on patient level end points, we advocate making a judgement on effectiveness informed by observations further upstream in the extended chain, along with information from other studies (2;3;5;13;25). A framework for evaluation, such as that proposed in the introduction, includes consideration of previous studies of similar interventions, the theoretical basis for the intervention (7), the results of (any) pre-implementation evaluations (alpha-testing) (5), observations of the ‘fidelity’ of uptake of the intervention (25), effects on intervening variables (25) and the triangulation of quantitative and qualitative observations at all levels in the chain (2;3). Depending on cost, observations may still be made at the patient level (see below) but these are given less prominence in two respects. Firstly, an upstream end point is selected as the basis for sample size calculations. Secondly, patient level observations are not expected to reach
conventional levels of statistical significance and are interpreted accordingly – a point to which we return in the context of Bayesian statistical analysis in the discussion. By way of example, consider a package of educational interventions designed to improve knowledge and attitudes relating to patient safety, with a cost of £10 (~€12; $17) per patient when averaged over an institution. The intervention would be cost-effective even if it saved one life for every 1,000 patients treated. As pointed out above, it might be asking too much, irrespective of study design, to expect to observe improvements in clinical processes or outcomes from such an intervention. The sample size would be enormous. However, the finding that staff attended the educational events, reported positively on the experience and demonstrated improved scores on a validated patient safety culture tool may provide sufficient encouragement to continue the intervention (30).

4.3 Value of Information

It may, of course, be very difficult to decide whether or not it would be cost-effective to ‘power’ a study on patient level end points rather than to rely on upstream end points. Furthermore, there may be value in measuring patient level end points, even if the sample size is based on upstream end points. This is for two reasons. Firstly, the effect on patient level end points might be bigger than expected and a significant effect may materialise even though this was not anticipated ex ante. Secondly, effects on patient level end points can be interpreted in a Bayesian paradigm, even when they are not statistically significant in the frequentist paradigm. Deciding whether or not to ‘power’ a study on patient level end points and if not, whether (and what) patient level end points to measure, is a matter for informed judgement.

These various decisions can be supported by formal Value of Information (VoI) methods to determine the form of evaluation that represents the most cost-effective use of resources. Such methods are well established in HTA (31-34) and the framework presented here suggests that they may have a role (perhaps an even greater role) in the design of service delivery and policy interventions. VoI turns on estimates of the costs of various studies and also the estimates of the probabilities of effects within the plausible range. The latter calls for a Bayesian approach, a point to which we return below.
5. Application of the Model to Handover

Handover is at the generic end of the spectrum and its effects are diffuse. This results in many potential process outputs:-

- prescribing
- reading test results
- arranging follow on tests
- communication with social services
- arranging follow on appointments
- giving advice/instructions to the patient

Each of these has a large set of sub-process e.g. for prescribing the range of potential drug errors. This confirms the diffuse nature of the effects of any handover intervention.

From this analysis, it therefore makes sense to evaluate outcomes rather than process. However, we learn from the above discussion that evaluation of outcomes might be resistant to the effect of the intervention but small effects are worthwhile when the intervention is not expensive.

In figure 1, we posit a handover intervention which costs €10,000 with opportunity costs (staff time off for training etc.) at €1,000,000 and travel at €3,000. We work through the consequences of such a scenario for a study design. Note that often a wrong assumption is made concerning the calculation of costs of a new intervention, in that, the cost of the intervention itself is perceived to be the cost, this is erroneous. To get a ‘true’ cost of any new intervention it is necessary to also consider the opportunity costs. These refer to the costs associated with taking staff away from their usual tasks to either take up training or to engage in the new tasks associated with the intervention. Costs in figure 1 are only estimates of both intervention costs and opportunity costs associated with the implementation of a new handover intervention.

This proposed model suggests a cost of €10,000 for 1 QALY. This is clearly highly cost effective. However, to prove this a study would need to measure a change in death rate of 1 in 1000. This is clearly an impossible task.
**Figure 1. A Worked Example To Show The Potential Cost-Effectiveness Of An Intervention To Improve The Safety Of Handovers, Considering Only Mortality**

**Plausible effect: Mortality**

10% of patients die within one year of discharge

↓

5% die from preventable reasons

↓

Therefore, 0.5% of all deaths are preventable

↓

Improving handover can give a 20% reduction in deaths

↓

Therefore, 0.1% reduced deaths

↓

With 20,000 discharges per year, this corresponds with 20 lives saved.

**Intervention costs (including opportunity costs)**

<table>
<thead>
<tr>
<th>Components</th>
<th>Cost</th>
<th>Opportunity Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Handover Intervention</td>
<td>€10,000</td>
<td></td>
</tr>
<tr>
<td>200 nurses (€50,000)</td>
<td></td>
<td>€500,000</td>
</tr>
<tr>
<td>100 doctors (€100,000)</td>
<td></td>
<td>€500,000</td>
</tr>
<tr>
<td>Travel</td>
<td>€3,000</td>
<td></td>
</tr>
</tbody>
</table>

~€1,000,000 for a hospital with 20,000 discharges = €50 per discharge

€50 gives 20 lives saved, with an average of 5 years.

Willingness to pay for a full QALY is €50,000

Therefore we get 100 life-years for €1,000,000.

Therefore 1 life-year for €10,000, which is below the threshold (i.e. Full QALY = €50,000)
6. Discussion

6.1 Multi-Component Interventions

Multi-component service interventions may contain both generic and targeted elements (35). An evaluation in this case should consist of observations relevant to both generic elements (e.g. measurements of effects on intervening variables and perhaps outcomes), and also relevant to the specific components (where targeted clinical processes are relevant). Such a situation would arise, for example, when a hospital adopts ‘lean service re-design’ both as a remedy for specific perceived shortcomings and as a motivating principle designed to prompt clinical services to innovate under their own initiative.

6.2 Spillover effects

The arguments presented here are based on a causal chain operating from left to right. However, there may be circumstances where bi-directional flow is plausible – a specific targeted intervention may produce upstream (feedback) effects. These in turn could bring about downstream changes (feed-forward) in a related activity within the same domain. For example, introduction of clinical guidelines for asthma care in general practice may sensitis clinicians to the use of guidelines in general and thereby produce improvements in diabetes care (36). This phenomenon is sometimes referred to as the ‘halo effect’ (37), although such spillover effects can be harmful as well as beneficial; for example, incentives to follow diabetes care guidelines may deflect attention away from other worthy goals. The corollary of potential spillover effects when targeted specific interventions are implemented is that the end points observed may need to be widened to take account of plausible positive and negative effects in related practices. How wide the search should be and where the effects become so diffuse that the intervention has more in common with the generic rather than the targeted genre is a matter for informed judgement.

6.3 Bayesian methods

Formal VoI requires estimates of effect size captured in the form of a Bayesian ‘prior’ (31-34). Such a prior can be used both for sample size calculations and also for later Bayesian analysis of study data where the ‘prior’ is updated by empirical information once this becomes available (2,11;38-42). The Bayesian approach can make good use of changes in end points that fall short of conventional levels of statistical significance, as previously pointed out in the context of rare diseases (43). Lastly, Bayesian methods provide an explicit method to combine (triangulate) information from outside a particular study along with multiple end points obtained within that study (38). The Bayesian philosophy does not, however, remove the need to specify sample size and hence define a ‘primary end point’.
6.4 Representational nature of the model presented in this report

This analysis shows that an extended version of Donabedian’s chain can help in the selection of primary end points that determine sample size, and hence costs and feasibility of evaluative studies of health policy and service delivery interventions. The model presented in this report represents processes of considerable complexity. For example, Donabedian’s process level could have been divided into more than three sub-levels; the underlying construct is, in all likelihood, a continuum. The nub of the argument is that the further to the left an intervention is applied, the greater the number of ‘downstream’ end points that might be affected until a point is reached where there are too many to capture at the clinical processes level. We cannot determine where this point lies in a formulaic way. Like many aspects of study design and interpretation, it calls for judgement. Even finely structured models such as Bayesian analysis and VoI provide an aid to human judgement and are not a substitute for it. Models are “simplified views of the world that help us think about complex issues, but are not true representations of the complexity itself” (44). It is in this spirit that the model presented here is offered.

7. Conclusion and Recommendations

If the assumptions in the proposed model and the worked example outlined in figure 1. are realistic (to be tested in later work) then with a relatively low cost of an intervention there is little value in running a full evaluation of impact on patient outcomes due to the proposed small and difficult to detect effects. Thus, impact would need to be assessed through a series of interviews.

Outcomes due to the proposed intervention may be small and difficult to detect. Thus, impact would need to be assessed using a different approach.

Firstly, there is a need to test the assumptions by populating the model with better data.

a) Examining previous literature
b) Conducting interviews with experts, staff and patients

These estimates of potential effects, significance and costs could be captured within a Bayesian framework.

Secondly, we will conduct a VoI study in which we will calculate the cost effectiveness of conducting a study to measure end points at the level of the patient. If as we expect, this shows that such a study would either be prohibitively expensive (more than the
intervention cost) or logistically impossible (not enough hospitals in Europe to test), then we have to be satisfied with a less robust study based upon upstream observations. Such a study could be conducted within a Bayesian paradigm bearing in mind that upstream effects are a necessary but not sufficient requirement to show changes down at the patient level.
References


(31) Claxton K. The irrelevance of inference: a decision-making approach to the
stochastic evaluation of health care technologies. J Health Econ 1999; 18(3):341-
364. [PMID: 10537899].

Modeling payback from research into the efficacy of left-ventricular assist devices
[PMID: 17493314].

(33) Girling AJ, Lilford RJ, Braunholtz DA, Gillett WR. Sample-size calculations for
trials that inform individual treatment decisions: a 'true-choice' approach. Clin

19605884].

(35) Shirley PJ. The Safer Patients Initiative: the UK experience of attempting to

England evidence based guidelines development project: methods of guideline

(37) Francis J, Perlin JB. Improving performance through knowledge translation in the
[PMID: 16557507].

(38) Lilford RJ, Braunholtz D. Reconciling the Quantitative and Qualitative Traditions:


## D2. Report on Model of Cost Effectiveness

### Appendix

*Table 1. Total sample sizes for the study of change in compliance of a clinical process (stroke algorithm) vs. outcomes contingent on compliance (death or permanent disability)*

<table>
<thead>
<tr>
<th>Study type</th>
<th>Percentage point change in uptake of algorithm</th>
<th>Sample size:</th>
<th>Number of sites</th>
<th>Contingent percentage point change in bad outcome</th>
<th>Sample size:</th>
<th>Number of sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single centre before &amp; after design (or RCT of individuals)</td>
<td>5</td>
<td>1,372</td>
<td>1</td>
<td>1</td>
<td>49,282</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>398</td>
<td>1</td>
<td>2</td>
<td>12,078</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>124</td>
<td>1</td>
<td>4</td>
<td>2,894</td>
<td>1</td>
</tr>
<tr>
<td>Multi site cross sectional study, cluster*, k=50</td>
<td>5</td>
<td>5,500</td>
<td>110</td>
<td>1</td>
<td>121,800</td>
<td>2,436</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>784</td>
<td>32</td>
<td>2</td>
<td>29,900</td>
<td>598</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>243</td>
<td>10</td>
<td>4</td>
<td>7,200</td>
<td>144</td>
</tr>
<tr>
<td>Controlled before and after, study on 20 sites*</td>
<td>5</td>
<td>4,992</td>
<td>20</td>
<td>1</td>
<td>199,020</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1,352</td>
<td>20</td>
<td>2</td>
<td>49,260</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>380</td>
<td>20</td>
<td>4</td>
<td>12,056</td>
<td>20</td>
</tr>
</tbody>
</table>
k = number of individuals per site.
ICC = Intra-class correlation
* An ICC of 0.06 is assumed for process outcomes, and 0.03 for patient outcomes (18).
+ Assumes rate in control group is unchanged pre and post comparison times and that ICC is negligible (19).
Illustration: Evaluating a percentage point change from 10% to 20% in use of stroke algorithm would require a total sample size of 398 individuals (divided equally between 2 groups); while a similar study spread over multiple sites, with 50 individuals per site, would require 784 individuals (which equates to 32 sites). The number of individuals required to evaluate the improvements in outcome from 20% to 18% are 12,078 individuals, or in a multiple site study, 29,900 individuals spread over 598 sites.