Methodology for Modelling Changes in Health Services Delivery and Policy

Vignette prepared for the MRC-NIHR Methodology Research Programme Advisory Group

Søren Rud Kristensen, Rachel Meacock, Matt Sutton
Manchester Centre for Health Economics, University of Manchester

16th February 2015

Summary
The modelling and evaluation of changes to health services delivery and policy pose more methodological challenges than for specific healthcare technologies. They are generally introduced in a non-experimental manner, they have widespread consequences both for outcomes and costs, and they are subject to equilibrating and offsetting mechanisms. In this vignette, we briefly describe these challenges and, using examples of papers that have sought to address them, we identify the key methodological challenges. We conclude with five priorities for unmet methodological need: better integration with theory; measuring spillovers; modelling long-term consequences; extended frameworks for cost-effectiveness; and implementation-evaluation couplets.
Background

Changes in health services delivery and health policy are often aimed at an organisational rather than at a clinician and/or patient level. While the methods for modelling and evaluating specific health technologies targeted at clinician and/or patient level are well-established, the methods for evaluating changes at an organisational level are less well-developed.

Changes at an organisational level are often introduced in a non-experimental fashion and real experiments in this area are expensive and rare (Haynes et al, 2012). The MRC have developed guidance on non-experimental evaluation of population health interventions (Craig et al, n.d.), but there are additional complexities in evaluating the impacts of changes to health services delivery and health policy that require further consideration.

At organisational level, there is a wide range of consequences that are of concern, including measures of access, quality, safety, efficiency and equity of health care. Some interventions are targeted at particular outcomes but most have a generic focus and potentially wide consequences across a range of process and outcome measures. The scope of these changes both in terms of organisations and outcomes usually precludes primary quantitative data collection, leading to a reliance on administrative data.

The inter-related nature of inputs, processes and outcomes in healthcare organisations, and the interactions between them, mean that changes to one component of a health service will have multiple consequences for the system as a whole. Relatively little is known about change mechanisms, i.e. how the inputs at an aggregate level (e.g. staff, capital and equipment) are related to the outputs (i.e. the production process), and the reactions of the various actors and agencies in the system to each other.

The fact that organisations and teams perform multiple tasks, only a subset of which can be measured, adds considerable complexity. As Holmstrom and Milgrom (1991, p.50) emphasise: “[g]iven a highly incomplete set of performance measures and a highly complex set of potential responses from the agent, how can the agent be motivated to act in the social interest?“ In services with multiple objectives such as health care, the question of whether these objectives are substitutes or complements has added pertinence (Kaarboe and Siciliani, 2011). These wider effects are also known as spillovers and have been shown to be substantial in several healthcare settings including interactions between (Baicker et al, 2013) and within organisations (Sutton et al, 2010).

A further complication arises because organisations and the processes they perform are dynamic. Thus, reactions to changes in health services delivery and health policy, and their wider consequences, are likely to have different long-term effects to their short-term effects. Our recent evaluation of the longer-term effects of a hospital pay-for-performance (P4P) programme illustrated all of these complexities (Kristensen et al, 2014).

Finally, these complexities apply both to costs and benefits. The cost-effectiveness of changes to health service delivery and health policy are particularly poorly understood, as we noted in our recent review of studies of the cost-effectiveness of P4P programmes (Meacock et al, 2014).
Aim

This vignette has been prepared for the MRC-NIHR Methodology Research Programme Advisory Group to review methods for modelling the consequences of changes in health services delivery and policy. Our focus is on economic analysis of the costs and outcomes of such interventions. This vignette describes the range of challenges faced by researchers seeking to model and evaluate changes in health services delivery and policy and concludes with a summary of areas of unmet methodological need.

Modelling the counterfactual

A key challenge when evaluating non-experimental intervention is the modelling of the counterfactual, i.e. what is likely to have happened to the intervention group had the intervention not been introduced? (Craig et al, n.d.) For this, it is critical to understand how the intervention is implemented, how treated units are assigned, the production process and the expected profile of effects over time.

Modelling changes to healthcare delivery and health policy requires clear understanding of the process of participation or assignment. There is a particular evaluation problem if selection into treatment is correlated with the expected outcomes (in economics vocabulary, selection is ‘endogenous’). There is an extensive suite of instrumental variables (IV) approaches to deal with such selection. These IV approaches require exogenous variation in participation, such as regional variation, time, distance or administrative rules. The challenge for researchers is to identify exogenous sources of variation in health service delivery and policy that affect participation strongly and only affect the outcome through the effect on participation.

An understanding of the production process is also required, as this determines the scope of the possible impacts. For example, in our evaluation of the longer-term effects of the Advancing Quality P4P programme (Kristensen et al, 2014), we identified that the healthcare teams targeted by the quality indicators treated both incentivised and unincentivised patients. As a consequence, it was plausible that the improvements in quality generated by the programme would affect all of the patients treated by these teams, not just those who were directly covered by the programme.

Another component to specifying the counterfactual is understanding how the change takes effect over time. There may be anticipation of a forthcoming change in delivery or policy. In our evaluation of Best Practice Tariffs (Allen et al, forthcoming) for example, we found rapid responses to these new incentives following their announcement but prior to their formal introduction. A recent paper by Malani and Reif (forthcoming) demonstrates the importance of attributing changes prior to introduction to the effect of a programme. The effects of the change may occur instantly or gradually. In our Best Practice Tariffs study (Allen et al, forthcoming), we used differential spline regression to show that the payment regime change was better captured by a change in time trend rather than an instantaneous change in level. Finally, the effects of an initiative to change healthcare delivery may be permanent or temporary. This issue has been explored in the context of the Quality and Outcomes Framework by Kontopantelis et al (2014). Their study indicated that performance did not fall when incentives were removed in contrast to a similar experiment in the US (Lester et al, 2010).
Overall, our assessment is that there is a plethora of approaches available for evaluating delivery and policy programmes that are introduced in a non-experimental manner, but more guidance is needed on when different methods should be used in this context.

**Attribution**

Attributing any effects detected by analysis to the specific change in health services delivery or policy under examination becomes increasingly difficult in the long-run, as competing or supporting policy changes occur alongside the primary change of interest. For example, the Advancing Quality P4P programme was absorbed into the national Commissioning for Quality and Innovation (CQUIN) initiative after 18 months, adding further complexities to our long-term evaluation of the programme (Kristensen et al, 2014).

**Mechanisms**

As Propper (2012, p.38) concluded in a review of the evidence on competition: "[t]he mechanisms by which these improvements have occurred are not well understood or researched. Difference-in-difference designs are essentially black box analyses and do not shed light on exactly how changes in incentives get translated into actions." This criticism of ‘black box’ analytical techniques, which relate a change in healthcare delivery or health policy to an outcome without tracking the intervening process, is highly germane. The inability of most evaluators to pinpoint the mechanism(s) through which change occurs undermines the general acceptance of the findings and limits their generalisability.

In this area, a considerable amount could be gained by drawing to a greater extent on existing methodologies. For example, economists and operational researchers have developed sophisticated methods for measuring various forms of efficiency and for modelling the production process (e.g. Data Envelopment Analysis, Stochastic Frontier Analysis). Despite their relevance to delivery and policy interventions, which seek to affect the production process, these are rarely used in programme evaluation studies.

**Wider consequences**

Changes to health service delivery and policy can have consequences beyond their intended effects. A policy-induced change in service delivery in one area may lead to an unintended diversion of effort away from other areas of care not covered by the policy. Alternatively, changes meant to improve service delivery in one area of care may unintentionally improve care in other areas of service. The key distinction here is between tasks that are substitutes and complements.

These complexities have importance for modelling and evaluation of changes to health service delivery and policy in two ways. First, they prompt the importance of looking beyond the intended outcomes to assess the full impact of a policy change. Second, their
possible existence highlights the need for careful consideration of possible spillover effects when choosing control groups in a non-experimental design study.

Wider consequences can also be generated by equilibrating and offsetting mechanisms within health care organisations and health systems. There is surprisingly little analysis that considers such equilibrating/offsetting effects. One exception is the study by Martin and Smith (2003), which simultaneously estimated the effects of waiting times for elective surgery in discouraging demand and encouraging supply. Their finding that demand was relatively unresponsive to waiting times suggested that reduction of waiting times through supply-side targets would not be offset substantially by increases in demand.

**Cost-effectiveness**

Whilst the effects of changes to health services delivery and policy on short-term health outcomes are frequently studied, the costs of such changes are often neglected (Maynard, 2012). Compared to the rigorous frameworks applied to the economic evaluation of single and multiple healthcare technologies (NICE, 2013), the methods for assessing the cost-effectiveness of changes to health services delivery and policy as a whole are underdeveloped and underutilised. Although administrative data sets can provide information on patient-level resource use, higher organisational-level cost data is often lacking.

The issue of opportunity cost is of particular pertinence, as the large-scale nature of health service delivery changes means that the possible health gains foregone through not providing alternative treatments are likely to be substantial and widespread throughout the whole system. Recent work by Claxton et al. (2013) provides useful estimates of the average cost of producing a quality-adjusted life year (QALY) within the NHS, and methods to calculate the magnitude of health likely to be displaced by any reductions in funding available for existing services as a result of decisions to fund new interventions which impose additional costs upon the NHS.

Quantification of the benefits of changes to health services delivery in terms of QALYs is also lacking, with effects commonly assessed only in terms of short-term outcomes such as 30-day mortality or readmission rates (Meacock et al, 2014). The impact of policies on health is, however, likely to span the lifetime of the patients affected. Methods to extrapolate these lifetime benefits from changes in short-term outcomes are therefore needed. We have begun to develop such methods, proposing a technique to calculate a discounted and quality-adjusted life expectancy (DANQALE) tariff which can be attached to changes in 30-day mortality using administrative data (Meacock et al, 2014).

**Need for theory**

While our vignette has concentrated primarily on empirical issues, the complexities we have highlighted underline the importance of a theoretical underpinning for understanding and modelling interventions that change relationships between and activities within organisations. There is a very substantial literature in economics (and of course other disciplines) that provides a theoretical understanding of organisations,
much of it directly applicable to health care organisations. These theories are too rarely applied and tested in real-world settings.

In recent work (Kristensen et al, 2014b), for example, we noted that the highly developed theoretical literature on setting optimal prices in P4P contracts had not been applied in all the recent P4P policies in healthcare. Derivation of a relatively simple theoretical model showed that price-setters needed to take account of four factors when setting prices: the benefit of the incentivised care process; the providers’ level of altruism; the opportunity cost of public funds; and the cost of the incentivised care process. We know of no examples where these have been taken into account.

There are also potential gains to be made by introducing methods developed in other areas of social sciences for evaluating interventions. Programme theories and logic models are examples of carefully developed methodologies for assessing the effect of policy interventions that can enrich evaluations of health services delivery and policy. For example, the logic model requires a precise statement of the purpose of an intervention and careful consideration of how the intervention is expected to work, including: the inputs the intervention relies on; the activities these inputs affect; the outputs and outcomes affected by the change in the short, medium and long term; and consideration of the specific context in which the intervention takes place. As such, the logic model can serve as a tool for sharpening the thinking required to identify the spillover effects highlighted above, by making assumptions and possible empirical tests about the assignment and production process explicit. Mixed methods research can offer insight on where to look for quantitative evidence of mechanisms and wider consequences.

The predictions from theoretical models can offer clarity in complex situations and it is often through the development of formal theory that the principal assumptions that underpin modelling and evaluation methodologies can be identified. There would be considerable value in bringing more of the vast theoretical literature to bear on the problem of how to model and evaluate changes in health service delivery and policy.

**Areas of unmet methodological need**

In this vignette we have identified a range of challenges for modelling and evaluating changes to health care delivery and health policy. We conclude by listing the key areas in which we believe there is unmet methodological need.

1. We noted the value of theoretical underpinning for empirical analyses of health care organisations. There is a vast theoretical literature on the behaviour of organisations and the individuals and teams that work within them. Yet, much of this literature has not been applied to healthcare. Bringing more of these structural models of organisations and systems to bear on healthcare problems would be beneficial. With the vast amount of data available in the health care context, this has the potential to bring about new contributions to the literature in these fields.

2. Those involved in delivering healthcare perform a multitude of tasks and there are complex reactions between the different actors in the healthcare system. The few studies that have measured spillovers in healthcare suggest that they may be substantial. Development of a methodology for identifying and measuring spillovers would be a
considerable contribution to furthering the science of modelling changes in healthcare delivery and health policy.

3. Too many evaluations of changes to healthcare delivery and health policy are focused on consequences in the short-term. Organisations and the processes they perform are dynamic and take time to adjust. There are also likely to be equilibrating and offsetting mechanisms. There is a need to emphasise the requirement for system-level analyses and longer evaluation periods, but the methodology for measuring and modelling longer-term consequences is worthy of further development.

4. There is widespread recognition of the need to include economic evaluation in the assessment of specific healthcare technologies. By comparison, studies of the cost-effectiveness of healthcare delivery and health policy are rare. The toolkit for economic evaluation in this space needs further development. There are examples of recent (Walker et al, 2010; Meacock et al, 2014) and ongoing development of frameworks for evaluating organisation-level changes, but further work is required.

5. We noted that changes in healthcare delivery and health policy are rarely introduced in a randomised manner. The manner in which such programmes are implemented requires a different approach to evaluation. There is a very wide range of non-experimental evaluation methods available. The challenge here is to ensure that the right evaluation method is identified for each mode of implementation. There may be value in the MRC issuing guidance on types of implementation and their corresponding evaluation method. This might encourage planners and policymakers to engage in feasible means of implementation that facilitate reliable evaluation, without requiring full randomisation.
References


