Some considerations relating to the attribution of NHS activity to outcomes for people with long-term conditions

Julien Forder

Jose-Luis Fernandez

Ray Fitzpatrick

Contents

Exec Summary

Background

The NHS Outcomes Framework incorporates a series of outcome measures. The aim is to assess what the NHS does in terms of what really matters to patients and care recipients. This assessment can form the basis of a mechanism to drive improvement in the NHS. In this regard, the purpose of the NHS Outcomes Framework is not just to measure outcomes but also to act as an accountability and performance mechanism.

The challenge with this approach is that outcome indicators are affected by a range of factors, not just the actions and activities of the NHS or other public services. As such, it is not immediately possible to attribute any observed change in an outcome indicator to the actions of the NHS. Some part of any observed change may be due to other factors that are outside the control of the NHS. There is a need, in other words, to disentangle the effects of 'external factors' from the effects of NHS activity and actions (in the broadest sense) on outcome indicators.

The main aim of this paper is to investigate this attribution or disentangling issue and suggest some ways in which it might be addressed. The focus is on people with long-term conditions and on Domain 2 of the NHS Outcomes Framework – Enhancing quality of life for people with long-term conditions (LTCs). The overarching measure for Domain 2 is the health-related quality of life (HRQOL) of people with LTCs. The main indicator is population level EQ-5D scores as established using surveys – mainly the GP Patient Survey (GPPS).

Measuring the impact of NHS activity

A change in HRQOL over time (or a difference in HRQOL between NHS organisational units such as CCGs) can be caused by: a change in the amount of NHS resourcing over time, a change in efficiency, and a change in the range of external or 'other' factors, outside NHS control. The goal is to find a way to identify the impact of NHS activity. Going further, we might want to find a way specifically to estimate the value of the change in NHS efficiency.

For this purpose, NHS activity can have both direct effects on HRQOL – for example restorative impacts where NHS activity helps to reduce the severity of the person's long-

term condition or through management of the consequences (e.g. pain) – and indirect effects on HRQOL, with NHS activity affecting the (other) determinants of HRQOL in the population. Examples of the latter include case-finding/diagnostic-related impact and preventative impacts.

We distinguish these impact paths of NHS activity because they imply different approaches to measuring the quality/efficiency of that activity. If we are chiefly concerned with assessing the quality of activity in terms of its direct effects, then a range of factors are outside the NHS in this regard and should be controlled for as 'external factors' (for example, factors such as the number of conditions a person suffers from, or their economic position). If we want to rate activity according to the widest set of possible consequences that NHS activity might achieve, including all the possible indirect effects, then there will be fewer other factors that should be accounted for in the analysis. In this case, we would not control for the severity/number of conditions or economic position as this could be affected by preventative activity.

This paper proposes two attribution methods, which both use population level data on HRQOL:

• the *residual difference* method;

• the *full-production function* method.

The residual difference method starts with measuring changes HRQOL rates in the population (over time) and then subtracts the effects of changes in external control factors over that period. Any remaining change in HRQOL gives an indication of the change the effectiveness of NHS activity.

This method requires a first step whereby the impact of all 'external factors' is determined, such as using a statistical model of the relationship between HRQOL and external control factors, or using a re-weighting standardisation approach.

The second step is to subtract the impact of other factors from the change in observed HRQOL. After subtraction, what we are left with is an estimate of the change in the impact of NHS activity other than any effect caused by any change in need factors. The result is an indicator that shows changes associated with NHS activity that are, in theory, independent of need effects. Nonetheless, this indicator will not just show efficiency changes. It will also embody all other non-needs-related changes in NHS activity: e.g. changes in budget or commissioner preferences. Consequently, such an indicator is best used to inform changes in performance relative to the baseline year (or benchmark unit of observation (e.g. CCG).

The production function approach uses data on NHS activity directly to estimate the relationship between (changes in) NHS activity and (changes in) HRQOL rates. This method

would estimate the average gain in HRQOL for each additional £1 of NHS activity and compare this with average gain estimated for the following year (or in comparing one area with another). Ideally, a new production function would be estimated on a yearly basis to produce a value for efficiency for each year.

Another option is a hybrid of these two methods. This involves using a full production function estimate to predict the HRQOL score for the following year and comparing the predicted HRQOL score with the actual score. Any difference would be assumed to be due to improvements in NHS efficiency. This assumption is made on the basis that all need factors have been accounted for in the analysis.

Determining whether specific control factors for HRQOL are internal or external can be challenging. In particular, some factors are partially 'internal': that is, affected by both NHS activity and by external influences – for example, the number of co-morbid conditions a person reports will depend on NHS activity but also on lifestyle, genetics, household composition and so on. We need to 'decompose' these control factors into their internal and external components. One way to do this is to 'model' the HRQOL control factor independently using existing and new factor-specific external controls to develop an adjusted or 'externalised' control factor. For example, we would use models of disease prevalence for the major LTCs to come up with some estimate for how prevalence would change year on year beyond any change in relevant NHS activity.

As regards the 'controlling' process, relevant factors include:

- Prevalence rates of major LTCs (diagnosed), including number of conditions
- Economic wherewithal or deprivation
- Local (non-NHS) services e.g. social care
- Population: age and sex structure
- Household composition, marital status or equivalent
- Area characteristics e.g. urban/rural

All could be used as external controls if we are only interested in assessing the direct effects of NHS activity. But if we are also interested in the indirect effects then, arguably, only the last two factors are fully external. The other factors would have to be decomposed.

There are further specific considerations regarding how the control factors are modelled.

- Identified need and severity – Changes in the apparent prevalence of long term conditions can also arise due to changes in the likelihood of someone with a particular condition, of a particular level of severity, self-reporting or being diagnosed with that condition. This 'identified need' could be partly influenced by NHS activity as well as factors that are external to the NHS, and this would need to be decomposed.

4

- Other service effects – The impact of non-NHS services, including social care and other public health services on HRQOL, can be regarded and measured as an external need effect. Ideally, we would use indicators of these other service factors which are independent of any NHS activity effect.
- Historical need patterns – With regard to people with long-term conditions, historical need patterns are likely to be important as control factors. Ideally, data on control factors from previous years would be included in the statistical modelling.
- Prevention – To account for prevention effects using the production function approach requires data on prevention-relevant activity in the whole population. This will generally be historic data and, consequently, of limited availability. One option is to use lagged values of total NHS activity.

Limitations

In practice there are a number of difficulties with applying these approaches, many relating to the (lack of) availability of suitable control factors and the correct specification of their relationship through time.

Not all the 'other factors' can be measured in practice. Therefore, even controlling for the observable factors, we cannot be sure whether a change in outcomes/HRQOL is due to NHS activity (i.e. efficiency) or some other unobservable factor(s). For the residual difference method, any omitted needs factor will mean that changes in HRQOL are wrongly attributed to NHS activity. There is also a problem if control factors are themselves influenced by NHS activity, especially if this influence occurs relatively quickly. Omitted variables can also cause problems for the production function approach, but in this case there are statistical methods that can potentially help to limit the problem (such as instrumental variables estimation, as long as appropriate instruments can be found).

The separate issue of lagged-effects arises because cause and effect can occur at different time points. In theory, the inclusion of lagged need and activity variables in the estimation can account for these effects. In practice, these variables are likely to be missing or unobservable, especially very long-lagged effects. There are different challenges in addressing missing lagged needs variables and missing lagged NHS activity (prevention effects). Missing lagged activity variables means that prevention or enduring effects of past use of services will not be measured in the production model approach, although more immediate effects would be captured.

Another issue is mapping the range of NHS activity associated with supporting people with long-term conditions to respondents in the GPPS survey. This task presents two particular challenges: there is the problem of identifying appropriate activity and the separate problem of mapping this to particular patients. Data linkage between datasets can help. Identifying particular types of activity is more difficult. Nonetheless, for national

performance and accountability purposes, an inclusive definition of NHS activity, which included all types of primary and secondary care service use, might be sufficient.

Finally, the NHS Outcomes Framework uses the EQ-5D HRQOL measure as the overarching indicator for Domain 2. However, EQ-5D focuses on personal functioning, and some interventions for people with LTCs are provided to help them manage the consequences of their condition, given that restoration of personal functioning is not possible. When this is the case, the use of EQ-5D alone is likely to result in under-measurement of the impact of NHS activity on quality of life for people with LTCs.

There are a number of methods that can be used tackle these problems, particularly for unobservables. These include finding 'proxy' variables for the unobservable factors, using panel/longitudinal datasets, and making expert assumptions about long-term factors.

Accountability

The Mandate from the Government to NHS England is the main strategic mechanism by which the NHS is to be held to account. NHS England is required to 'make progress' with respect to each of the domains in the NHS Outcomes Framework. Overall 'progress' could be measured in various ways. There is a distinction between expecting the NHS to become more effective within the total available budget or irrespective of budget limits. This distinction is particularly important if the NHS budget is being reduced. Since the size of the NHS budget is outside the control of NHS England, it might be most relevant to judge how much progress has been made by measuring the improvement in HRQOL stemming just from an increase in efficiency. We might also wish to specify the time period over which the improvement is being judged.

The impact of NHS activity can be assessed in aggregate terms at the national level or in terms of the distribution of progress gains across the country – the latter approach capturing the regional variation in the NHS. In other words, the objective may be to improve overall HRQOL but also to ensure that progress is reasonably even across the country.

Regarding the distribution of gains, progress could be assessed in comparative terms between areas or other organisational units. The methods outlined above could be applied at, for example, CCG level to assess how the impact of NHS activity on HRQOL differed between areas. Given the factors beyond NHS control in any locality, and given the level and nature of NHS activity, this assessment would indicate which areas were above and which below expected levels of HRQOL in their populations. These results would have more limited use in helping to understand why localities were away from expected values, but would be a mechanism to facilitate or trigger further investigation of the causes.

Conclusion

Assessing the progress of the NHS in terms of the HRQOL outcomes it produces for people with LTCs goes directly to the heart of what matters to patients. The problem with final outcome measures is being able to attribute changes in HRQOL to the activity of the NHS.

Of the two attribution methods considered in this paper, the residual difference approach is the most straightforward to implement. To date, this approach has been exemplified for year-on-year national changes in HRQOL, but it could be extended to compare performance between organisational units such as CCGs.

The production function approach is arguably the most theoretically robust method and the one that provides direct efficiency comparisons. However, it is a complex method that is highly demanding on data and could not be implemented at present given the data that are available.

Regardless of the practical challenges, it is necessary to adjust for the impact of factors outside NHS control when making accountability judgements about NHS performance. In view of the different types of impacts of NHS activity, with their implications about the choice of appropriate control factors and data-requirement in each case, the use of several performance indices is suggested. Some would use more control factors and so be focused on the quality of NHS care for people with LTCs; others would use fewer controls and so could encompass a wider set of impacts, including prevention, but potentially also embody more 'noise'. The set of indices would be used to support an overall interpretation of the performance of NHS England.

1 Introduction

Government policy on health and social care has a strong emphasis on the outcomes of NHS and social care activity. An outcomes approach is directly concerned with the impact that services and support have on patients, services users, carers and so on, rather than (just) a focus on the types and intensities of services and interventions that are provided. Outcomes can be measured in a number of ways that would include the mortality, health status, wellbeing, impairment, safety and experience of patients and other people affected by services.

The NHS Outcomes Framework incorporates a series of outcome measures. The aim is to assess what the NHS does in terms of what really matters to patients and care recipients. This assessment can form the basis of a mechanism to drive improvement in the NHS. In this regard, the purpose of the NHS Outcomes Framework is not just to measure outcomes but also to act as an accountability and performance mechanism. In particular, the purpose is to:

- provide a national-level overview of how well the NHS is performing;
- allow the DH to hold the NHS to account via NHS England (formally the NHS Commissioning Board), and to allow Parliament to hold the Secretary of State to account, for improvement in outcomes; and
- act as a catalyst for driving up quality throughout the NHS by encouraging a change in culture and behaviour.

The aim of gearing performance and accountability mechanisms directly around the relevant outcome indicators is that it motivates improvement efforts to be targeted *directly* on achieving best outcomes for patients and care recipients.

The challenge with this approach is that outcome indicators are affected by a range of factors, not just the actions and activities of the NHS or other public services. In particular, it is generally not possible to immediately attribute any observed change in a generic outcome indicator to the actions of the NHS. [1] Some part of any observed change may be due to other factors that are outside the control of the NHS. There is a need, in other words, to disentangle the effects of 'external factors' from the effects of NHS activity and actions (in the broadest sense) on outcome indicators.

An alternative is to use 'process' indicators, such as waiting times, admission rates and so on. As these process indicators are within the control of the NHS, this is an approach which side-steps any significant attribution issue. The general problem, however, is one of causation: having 'good' processes does not necessarily mean good final outcomes for service users. Indeed, it is often difficult to define what 'good' looks like for process

---

[1] Although it might be possible to link condition-specific outcome measures with particular behaviour.

measures. Particular process measures can be validated against generic outcome measures in specific studies, which limits this problem, but in doing limits our overall appraisal of NHS activity because we would need a whole range of process indicators that would need to be aggregated in some way. For specific assessment (e.g. condition-specific groupings) process measures will be valuable.

The use of process indicators can also lead to an over-emphasis on the process target itself, rather than the outcome goal that underlies the target.

When considering how NHS 'activity' could produce changes in the desired outcome indicator, there are different ways in which the term activity can be interpreted for accountability purposes. Broadly speaking, the amount of any improvement in population outcomes due to NHS activity will depend on, first, the total scale of that activity, in terms of total NHS resourcing (e.g. the budget per capita), and second, how that activity is organised and delivered (e.g. how the NHS budget is used). The latter will include: clinical decisions about what services and interventions are provided; choices and incentives about the quality of service delivery by providers; and selection of providers in affecting cost-efficiency.

We might want to be clear about this distinction when considering accountability questions. Do we expect the NHS to become more effective within the total available budget or irrespective of budget limits? This distinction is particularly important if the NHS budget is being reduced. In that case, an outcome indicator could show a decrease over time, despite an increase in cost-effectiveness/productivity. From a service evaluation perspective, we want to focus on the degree of *efficiency* of NHS activity – the extent to which the NHS contributes to improving outcomes with a given level of funding – as well as analysing the final levels of outcomes enjoyed by people in need of care. There are many types of activity on which we could focus, including types of intervention, commissioning or delivery process, and target patient groups.

As regards accountability, there is also the question of whether the focus should be on national differences in NHS performance over time (in producing HRQOL gains in the population, given need), and/or comparative performance as between NHS organisational units (e.g. between CCGs). Focusing on differences between NHS units would allow some identification of performance outliers compared to the national average. These results would be a good basis for directing improvement efforts towards under-performing areas.

The main aim of this paper is to investigate the attribution or disentangling issue and suggest some ways in which it might be addressed. There are a series of challenges in this task and this paper makes some suggestions as to how they might be addressed.

The focus is on people with long-term conditions (LTCs) and on Domain 2 of the NHS OF, *Enhancing quality of life for people with long-term conditions*.

The overarching measure for Domain 2 is the *health-related quality of life* (HRQOL) of people with LTCs. The main indicator is population level EQ5D scores as established using surveys – mainly the GP patients' survey (GPPS).[2]

The Mandate from the Government to NHS England is the main strategic mechanism by which the NHS is to be held to account. NHS England is required to 'make progress' with respect to each of the domains in the NHS Outcomes Framework.

In what follows we generally refer to 'NHS' activity because we are primarily concerned with the responsibility of NHS England. However, this term should not be taken to mean services and support just provided by the NHS. Indeed, care of people with long-term conditions will increasingly span the range of conventional organisational distinctions, that is, including health and social care (and potentially other services such as housing, benefits etc.). Moreover, this wider set of activities will fall increasing under NHS England's remit as it works in partnership with local authorities and other agencies. In taking an outcomes focus, NHS England is responsible for the "quality of life for people with long-term conditions", and achieving best improvement in this respect is likely to entail a range of support and service activities. This paper is about methods to assess how much progress is made in this regard.

2   What is the impact of NHS activity?

The implied objective of NHS activity with regard to Domain 2 is to produce improvements in the HRQOL of people with LTCs. Broadly speaking, NHS activity could produce such an improvement overall by:

- An increased level of NHS activity per capita, facilitated by an increase in NHS funding (i.e. moving the NHS up the HRQOL output curve).
- An improvement in how efficiently the given NHS budget is used in any year. In this sense, we mean how much HRQOL is gained overall from each £1 of NHS budget. Efficiency changes could be driven in many ways, including innovation, better use of services, improved clinical practice and better cost control.

A change in HRQOL over time (or a difference in HRQOL between NHS organisational units such as CCGs) can be caused by: a change in the amount of NHS resourcing over time, a change in efficiency, and a change in the range of external or 'other' factors, outside NHS control, such as the characteristics of the local population (see also Table 1 below). As a shorthand, we can write:

---

[2] and also potentially the Health Survey for England (HSE)

$$\Delta HRQOL_t = \Delta \pi_t x_t + \pi_t \Delta x_t + \Delta impact\_other\_factors_t \qquad (1)$$

where x_t is 'NHS activity' and π_t is NHS efficiency (and Δ is a change).

The goal is to find a way to isolate the impact of NHS activity i.e. the amount $\Delta \pi_t x_t + \pi_t \Delta x_t$ in the equation. Indeed, going further, we might want to find a way to specifically estimate the value of $\Delta \pi_t$, the change in efficiency. Below, we outline two methods for measuring the impact of NHS activity. Both require us to 'control for', i.e. remove, any changes in HRQOL that is due to changes in these other factors.

In what follows, when we refer to 'NHS activity' this is not to imply that actual NHS activity is in any sense homogenous or can be treated in purely aggregate terms. None of the methods we outline below requires such an assumption. Indeed, we would expect specific types of NHS activity to have quite different impacts on HRQOL. So the impact of NHS activity is actually the sum of the impact on population HRQOL of a whole range of activity.

2.1.1 NHS activity

For the purposes of assessing NHS activity we might identify two ways in which NHS activity could have an impact on HRQOL of people with long-term conditions,

- First, as direct effects on HRQOL, including for example *restorative impacts* (i.e. where NHS activity helps to reduce the severity of the person's long-term condition) and *management* impacts (i.e. helping people with long-term conditions cope with the consequences of their condition: e.g. pain management). These are examples of tertiary prevention (preventing the negative impact and sequelae of established disease);
- Second, as indirect effects on HRQOL with NHS activity affecting the (other) determinants of HRQOL in the population. Examples include *case-finding/ diagnostic-related impact* (i.e. identifying people with LTCs and the severity of their conditions) and *preventative impacts* (i.e. where NHS activity reduces the chance that people develop LTCs in the first place). In this case, we are mostly referring to secondary (e.g. early diagnosis and treatment) and primary prevention (reducing risk of onset).

We distinguish these impact paths of NHS activity because they imply different approaches to measuring the quality/efficiency of that activity. In particular, if we are chiefly concerned with assessing the quality of activity in terms of its direct (restorative or condition management/tertiary prevention) effects, then a range of factors are outside the NHS in this regard and should be controlled for as 'other factors'. If we want to rate activity according to the widest set of possible consequences that NHS activity might achieve, including all the possible indirect effects (i.e. including the primary and secondary prevention), then there will be fewer other factors that should be accounted for in the analysis. In practice many specific NHS activities have a range of possible consequences, but this does not mean that

we need to assess activity against the full range. Indeed, the broader we cast this net, the greater is the loss of specificity in our assessment. In other words, as we outlined below, we would be trading having a wider scope for a potential loss in accuracy.

When we consider only the direct HRQOL consequences, it would be appropriate to regard factors such as the number of diagnosed conditions per person as outside the impact path of NHS activity. The number of diagnosed conditions a person suffers from, however, has a strong impact on health-related quality of life independently of any direct NHS activity effect. This factor is also strongly associated with the nature of the direct (restorative and management) impacts and costs of NHS activity, and so for both reasons it should be accounted for in any assessment of this NHS activity. Essentially if are only considering the direct impacts of NHS activity then not controlling for external factors like number of conditions would lead to an unfair assessment. Differences in these factors (between organisational units or time periods) would generate differences in HRQOL scores even if the quality of care in achieving direct impacts was the same. In not controlling for this factor, the HRQOL data would appear to suggest that there was a difference in quality when this was not actually the case.

Similarly, rates of economic deprivation are (non-care) determinants of HRQOL, and these factors should be controlled for if NHS activity is being assessed only in terms of its direct restorative and management impacts.

If we are also interested in the potential indirect effects of NHS activity, such as preventative and case-finding effects, then we should not control for factors like prevalence rates and economic deprivation, because changes in such factors could be part of the impact of NHS activity, so measured. Prevention impacts, for example, would affect prevalence rates, population age structure, economic deprivation and so on, which in turn would have a knock-on effect on HRQOL. The direct (restorative/management) impact on HRQOL of NHS activity might be *affected by* these external factors, but does not, by definition affect these external factors.

In also assessing case-finding/diagnostic and, especially, preventative impacts, the potential scope of NHS activity is much wider, leaving fewer potential factors outside the control of those undertaking these activities.

The timing of measured activity is also different between these activity types. Restorative and management activities tend to have short time periods between intervention and impact, whereas preventative activities might take many years to have impact.

For these reasons, it is useful to be re-write (1) as:

$$\Delta HRQOL_t \qquad\qquad\qquad\qquad\qquad (2)$$
$$= \Delta(direct)activities_t$$
$$+ \Delta internal\_control\_factors_t(\Delta(indirect)\_activities_{t\ lagged})$$
$$+ \Delta external\_control\_factors_t$$

where $internal\_control\_factors_t$ are variables that could affect HRQOL but are themselves potentially affected by NHS activities. An example is diagnosed prevalence rates of major LTCs. By contrast, $external\_control\_factors_t$ are those factors that are external to (i.e. unaffected by) NHS activity of all types e.g. the gender breakdown in the population. Both internal and external control factors would be used in relation to when assessing only direct impacts, but only the external control factors should be used when indirect (case-finding or prevention) impacts of NHS activity are also considered.

Some potentially relevant control factors are listed in the following table. Also shown are those control factors that might be regarded as external (rather than internal) with respect to the different types of NHS activity. In some cases, the relevant control factor may be partially external. For example, NHS activity could help people to return to the working population, affecting economic deprivation rates, but global financial shocks leading to recession are clearly outside the control of the NHS.

Table 1. External control factors

Control factor	*External* control (rather than internal)	
	Direct impacts (e.g. restorative/ management impacts)	Indirect impacts (e.g. Case-finding/ prevention impacts)
Prevalence rates of major LTCs (diagnosed), including number of conditions	Yes	No/Partially
Economic wherewithal or deprivation	Yes	Partially
Local (non-NHS) services e.g. social care	Yes	Partially
Population: age and sex structure	Yes	Partially
Household composition, marital status or equivalent	Yes	Yes
Area characteristics e.g. urban/rural	Yes	Yes

Some of these factors are more relevant when we are comparing between (geographically differentiated) units rather than through time. For example, area characteristics will likely change very slowly from year to year, but differ markedly from CCG to CCG.

2.2 Attribution methods

In this paper we concentrate on statistical methods that use population level data on HRQOL. Before considering these methods, it is worth noting potential issues with sampling that would affect any statistical approach. In particular, we are concerned with sampling HRQOL and other data from the GPPS. This is a large survey, which helps minimise many sampling problems. Nonetheless, it is not clear how consistently the GPPS is administered in different areas, which could create bias. If this potential bias remained largely unchanged through time, then assessing year-on-year differences in HRQOL is minimally affected. However, where different population units are surveyed at each cross section, these biases could be important. In other words, systematic differences in HRQOL could appear from year-to-year as a result of inconsistent sampling. Whatever approach is used going forward, the significance of any such inconsistencies would need to be investigated. Our initial analyses using the GPPS survey suggest, however, that very large changes in the

composition of the sample are required in order to bring about significant changes in the average EQ5D scores (See our previous QORU report, Fernandez, Hughes et al. 2013).

We consider two main statistical methods:

- the **residual difference** method, and
- the **full-production function** method

The *residual difference* method takes data on HRQOL and relevant external factors and produces an estimate of the impact of NHS activity by subtraction:

$$\Delta HRQOL_t(given\_other\_factors) \qquad\qquad (3)$$
$$= \Delta HRQOL_t - \Delta external\_control\_factors_t$$

Any remaining changes in HRQOL that are not accounted for by the impact on HRQOL of changes in other factors give an indication of the change in the effectiveness of NHS activity. For example, after collecting all data on other factors, we find the difference in these factors should account for a 0.5% change in HRQOL over a year. But we actually see a 2% improvement in HRQOL, so we can deduce that 'NHS activity' changes contributed the extra 1.5% change, or three-quarters of the improvement.

This method requires a first step whereby the impact of 'external factors' is determined. This first step can be accomplished using a regression model estimation of HRQOL on the external control factors (i.e. $HRQOL_t = external\_control\_factors_t + error_t$) or using a re-weighting standardisation approach, as outlined in our previous QORU report (Fernandez, Hughes et al. 2013). This method does not use NHS activity data. Nonetheless, we should note that, because the level and type of NHS activity is partly determined by external factors, then part of the normal effect of services does get picked up in the first-step estimation – see below for a discussion of this point.

The second step is to subtract the impact of other factors from the change in observed HRQOL. After subtraction, what we are left with is an estimate of the change in the impact of NHS activity other than any effect caused by any change in need factors. As such, the residual difference is not the change in the total effect of NHS activity, but instead the effect net of need changes. The result is an indicator that shows changes associated with NHS activity that are, in theory, independent of need effects. Nonetheless, we should be clear that this indicator will *not* just show efficiency changes. It will also embody all other non-needs-related changes in NHS activity: e.g. changes in budget or commissioner preferences.

This method requires that all relevant needs factors are captured. If this were not the case, some observed change in HRQOL could be due to a change in an unobserved need factor. This assumption is discussed below. Also, see the Appendix for more detail.

For the above reasons, the resulting indicator is quite difficult to interpret in any *absolute* sense. Rather, it should be used to inform *changes* in performance, either through time or between units of observation (e.g. CCGs). In this case, the baseline effect of NHS activity on HRQOL (and also the external effects of the relevant control factors) that works through the first-step equation acts as a benchmark when assessing future years' HRQOL scores. The benchmark embodies NHS efficiency/performance at baseline, so after accounting for changes in external control factors in the next year(s), any further change should indicate changes in performance. There are a number of statistical methods available to help in interpreting the residual effect (e.g. frontier estimation).

It is clear, in any case, that without making any adjustment for need change there is a good chance that change in NHS performance will be inaccurately measured.

Although lacking direct indicators of the relationship between services and outcomes, the residual difference approach could be used to identify possible loci of best practice, which could be further investigated using other techniques, such as in-depth analyses of local commissioning processes and levels of service provision.

The *production function* approach is a method to estimate efficiency changes ($\Delta \pi_t$) directly. A full production function would be estimated each year using regression methods, but in this case including some variable(s) for NHS activity directly in the model:

$$HRQOL_t = \pi_t x_t + external\_control\_factors_t + error_t \qquad (4)$$

Data on NHS activity (x_t) need to be combined with the HRQOL data. With this data, plus other factors data, regression analysis could be used to directly estimate the impact of NHS activity on HRQOL, π_t. Ideally, a new production function would be estimated each year to produce a value for efficiency, π_t, for each year. Performance year-on-year would be assessed by comparing estimated efficiency in the base year (i.e. π_t) with the value a year later i.e. π_{t+1}.[3]

To build on the above example, in the base year population HRQOL is 0.7 and in the following year is 0.714 i.e. a 2% increase, as above. Applying the production function method, suppose in the base year that we found that NHS activity contributed 0.25 of the 0.7 but a year later contributed 0.2605 of the 0.714 total. Therefore the results show an efficiency improvement which, as above, accounts for three-quarters of the 2% overall improvement in average HRQOL.[4]

---

[3] It would be possible to standardise for scale effects (diminishing marginal productivity) by calculating efficiency at the baseline value of x_t i.e. $\pi_{t+1}(x_t)$ compared to $\pi_t(x_t)$.

[4] This figure constitutes a 4.2% improvement in efficiency (=0.2605/0.25).

Another option is a hybrid of these two methods. It involves predicting the HRQOL score for the following year (time t + 1) using (4)[5] and comparing the predicted HRQOL score with the actual score in that following year. Any difference could be assumed to be due to improvements in NHS efficiency. This assumption is made on the basis that all need factors have been accounted for in the analysis.

See the Appendix for further details of these methods.

All of these approaches attempt to account for factors other than NHS activity that might affect HRQOL. In practice there are a number of difficulties with applying these approaches, many relating to the (lack of) availability of suitable control factors and the correct specification of their relationship through time. These limitations and potential solutions are considered in the next section. Before that, we can briefly consider the main categories of need/control factor and potential data sources that would be (ideally) used in the analysis.

2.2.1 Control factor modelling

Determining whether specific control factors for HRQOL are internal or external can be challenging. In particular, some factors are partially 'internal': that is, affected by both NHS activity and purely external influences. For example, the number of co-morbid conditions a person reports will depend on NHS activity but also on lifestyle, genetics, household composition and so on. If we include this variable as an (external) control factor, then it will 'control-out' any preventative improvements in NHS activity, for example. If it is left out, then any changes in its value due to changes in its external drivers will not be taken into account when assessing performance.

This problem can, in theory, be overcome by 'decomposing' relevant HRQOL control factors into their internal and external components. The external component would then be used as one of the external control factors in either the residual difference or production function methods. There are two approaches we might use to do this decomposition. The first is to 'model' the HRQOL control factor independently using existing and new factor-specific external controls to develop an adjusted or 'externalised' control factor. For example, we would use models of disease prevalence for the major LTCs to come up with some estimate for how prevalence would change year on year beyond any change in relevant NHS activity. The second approach is to bypass this step and simply include these factor-specific external controls directly in the set of external control factors for estimating the main HRQOL models above.[6] This latter approach obviously requires that the new control variables were available in the main HRQOL analysis dataset.

---

[5] Taking the error to be zero.
[6] This is a 'reduced form' approach.

The 'modelling' in the first approach could be undertaken in a number of ways. One option would be to use the results of existing disease specific modelling studies. Again, the aim would be to estimate how the specific control factor might change irrespective of NHS activity. A more pragmatic option would be to use expert opinion and assumptions for this purpose.

Table 1 above indicates which HRQOL control factors would ideally be modelled and decomposed. In practice, there would clearly be limits on how much of this additional modelling could be undertaken and included. A strategy might be to leave out contentious control factors in the main analysis. This strategy would be less likely to adversely affect the production function approach – see section 3.1 below.

2.2.2 Identified need and severity

Changes in the apparent prevalence of long-term condition can also arise due to changes in the likelihood of someone with a particular condition, of a particular level of severity, self-reporting or being diagnosed with that condition. This 'identified need' could be partly influenced by NHS activity – i.e. diagnosis-related activity – as well as factors that are external to the NHS. Ideally, we would want to decompose these effects, as above, either finding suitable proxy indicators or modelling a relevant control factor adjustment.

The problem with using unadjusted prevalence/severity measures as controls is particularly apparent in this case. For example, suppose the NHS improved its identification and diagnosis of people with LTCs. In turn, this would lead to better treatment of individuals and so improvements in HRQOL. By contrast, controlling population average HRQOL by the unadjusted proportion in the population with the diagnosed condition would remove some of this beneficial effect.

2.2.3 Other service effects

The impact of non-NHS services, including social care and other public health services on HRQOL, can be regarded and measured as an external need effect. As with the expressed need factors, the provision of other services might be affected by current or past levels of NHS activity. Ideally, we would use indicators of these other service factors which are independent of any NHS activity effect.

2.2.4 Historical need patterns

With regard to people with long-term conditions, historical need patterns are likely to be important as control factors. Ideally, data on control factors from previous years would be included in the statistical modelling. For example, current HRQOL may depend on how long a person has had a condition and, therefore, past prevalence.

Accounting for long-term effects requires repeated observations on the same unit of analysis, such as data about respondents in a longitudinal or follow-up survey, or characteristics data drawn from the same small area if geographical data linkage is used.

2.2.5 Prevention effects

In using the GPPS to measure population HRQOL averages, we would need to include the whole population in the analysis, including those people without reported LTCs.

To account for prevention effects using the production function approach requires data on prevention-relevant activity. This will generally be historic data and, consequently, of limited availability. One option is to use lagged values of total NHS activity. We would still need a way to link historic data with respondents in the GPPS. It would also be important to account for historic changes in the numbers of people with LTCs and the severity of the condition.

3 Potential limitations

There are a series of challenges in using these methods in practice, with some limitations being more significant for the residual difference method, and others a greater problem for the production model approach.

3.1 Un-observables

Not all the 'control factors' can be measured in practice (not, at least, in a way that allows us to gauge their effect on current HRQOL). Therefore, even controlling for the observable factors, we cannot be sure whether a change in outcomes/HRQOL is due to NHS activity (i.e. efficiency) or some other unobservable factor. This un-observables problem is particularly relevant when using survey data (as in this case) because a survey is limited in terms of what mitigating factors data can be collected.

The consequences of this problem can be significant. Take the *residual difference* method. Any omitted needs factor will mean that changes in HRQOL are wrongly attributed to NHS activity. For example, suppose there was a widespread 'flu outbreak in a given year. This outbreak would reduce population HRQOL scores for that year. Without accounting for this effect routinely, NHS productivity would have appeared to have reduced for that year-on-year comparison.

There is also a problem if needs factors are themselves influenced by NHS activity, especially if this influence occurs relatively quickly. In that case, including the affected need factor in the first stage regression will 'mask' the influence of NHS activity. An example (for sake of argument) might be a newly NHS licensed drug that lowers prevalence rates for a long-term condition (rather than helps to manage the symptoms). By including the improvement in prevalence as a control factor, we remove the beneficial effect of NHS activity (the new drug).[7] This problem arises even if we had no missing/un-observable needs factors. A

---

[7] This is endogeneity bias where the error in a regression analysis is correlated with some of the independent variables (e.g. needs factors). In this case, needs factors are a function of the omitted NHS activity variable – see Appendix.

solution is to estimate an HRQOL production function, including NHS activity. This is the hybrid method outlined above.

The omission of NHS activity or need factors that are correlated with the included needs factors would also lead to biased coefficients on the included variables. This is because the included needs factors will reflect some of the effect on HRQOL of the missing factors as well as their own effect.

Omitted needs factor problems will be relatively modest if we are primarily concerned with year-on-year changes in NHS performance, and the omitted needs factors are expected to only change gradually over time compared to the (change in) effect of NHS activity. The problem is further mitigated if the effects of omitted needs factors are not concentrated in particular localities across the country. Changes in social norms regarding smoking and exercise are examples of slowly-changing omitted factors.

Omitted variables can also cause problems for the *production function* approach, but in this case there are statistical methods that can help to limit the problem. Instrumental variables estimation is a well-established method that can be used for estimating the full production function (4), even where there are omitted/un-observable factors (Claxton, Martin et al. 2013; Forder, Malley et al. 2013, forthcoming).

This solution requires that we find appropriate instrumental variables – see Appendix. They should be correlated with NHS activity, but not (directly) with population HRQOL. Possible instruments in this case would be those that capture differences in policy regarding the types and intensities of services and support provided to patients in different NHS commissioning areas. These policy choices or local preferences on the part of NHS decision-makers will affect the scale and type of NHS activity for each patient with given needs/conditions, but the determination of those preferences is unlikely to be affected by factors that impact on the HRQOL of particular patients.

Another possibility is to use total budgets for CCGs as instruments in predicting programme budget-based activity variables. Total budgets are determined on a set of observable needs factors and likely to have much smaller correlation with un-observed needs factors.

Supply factors (e.g. provider capacity) might also be considered as instruments for NHS activity since capacity allocation decisions tend to have strong historical determinants and are not much affected by current (un-observed) needs factors.

It is nonetheless difficult in practice to find ideal instruments, and the use of poor instruments could introduce more bias than the original omitted variable bias. Furthermore, IV estimation is sensitive to model specification, and the confidence intervals on π_t could be quite large. As such, we might anticipate that only substantial changes in efficiency year on year would be detectable, although the large sample size of the GPPS helps in this regard.

3.2 Lagged-effects

Cause and effect can occur at different time points. Current NHS activity may not have a significant effect on outcomes for some time: e.g. smoking cessation programmes that reduce long-term COPD prevalence. This is a lagged-effect problem. In theory, the inclusion of lagged need and activity variables in the estimation can account for these effects. In practice, these variables are likely to be missing or un-observable, especially very long-lagged effects, and create similar problems to those outlined above.

The significance of these problems is different between the two attribution methods. It is useful to differentiate between the problem of missing lagged needs variables and missing lagged NHS activity (prevention effects).

In as far as they do not include service indicators, lagged service-effect problems are somewhat side-stepped with *residual difference* method. NHS activity variables are not used with this approach so lagged NHS prevention effects cannot be separately estimated. The problem of missing lagged *need* effects is not avoided (which is particularly relevant if we are attempting to capture preventative effects of the NHS leading to changes in the prevalence and severity of conditions).

These problems arise in a different way with the *production model* approach. Missing lagged activity variables means that prevention or enduring effects of past use of services will not be measured. Intuitively, these effects would seem to be important for people with long-term care needs. Yet it will also be very difficult to find good measures of previous activity, especially if we want to go back more than a few years, and particularly so if we want to identify specific preventative interventions. As a consequence, we will only be able, realistically, to measure the impact of *current* NHS activity on HRQOL and perhaps limited previous years to total NHS activity. Of course, if we are attempting to build up a picture through time of productivity/efficiency trends, then current impact may be sufficient for that purpose.

Not including lagged activity will also bias the effects of current NHS activity on HRQOL if current and past activity levels are (directly) correlated. In some cases, we might anticipate the existence of such correlation: for example, if an intervention spans a number of years, or it reflects a persistent policy position by the local NHS. However, in other cases, we would argue that it is needs factors (e.g. underlying condition) that are correlated through time, not activity levels *per se*. Activity levels respond to needs.

Missing lagged needs are less problematic with this approach because the consequences can be mitigated using IV estimation (as long as the instruments we use have little correlation with the missing lagged needs factors). The same advantage of using IV with missing needs factors applies as in the un-observables case outlined above. Indeed, we can

regard this as a form of un-observables problem because the lagged effects are not in themselves the issue, but rather the lack of data about relevant historical need and activity.

3.3   Associating NHS activity with patients

A challenge for the production function approach is the need to find good measures of relevant NHS activity. Ideally, we would map the range of NHS activity associated with supporting people with long-term conditions to respondents in the GP patient survey. This task presents two particular challenges: there is the problem of identifying appropriate activity and the separate problem of mapping this to particular patients.

There are a number of practical ways in which we can tackle these problems. The mapping or linking of data can be done between the GPPS and administrative datasets using NHS number or a similar identifier. For example, secondary care activity can be linked in using the HES dataset. There is also some linkage to primary care dataset, such as QOF data, and there are some questions on GP service use in the survey. Alternatively, mapping of some NHS services can be done at small area level. Some prescription data also exist.

Identifying particular types of activity is more difficult. Nonetheless, for national performance and accountability purposes, we would probably want to take an inclusive definition of NHS activity, including all types of primary and secondary care service use.

A more pragmatic approach would be to use total expenditure at CCG level. A multi-level modelling approach could cope with outcomes data at individual level in the GPPS and activity level at the CCG level. Some refinement could be achieved by using the programme budget classifications.

Lagged activity data could also be proxied using current and historical secondary care use for people in the GPPS from HES.

Some particular limitations would arise from the general lack of administrative data concerning allied health professional services, such as physiotherapy, that would appear to be particularly relevant for a long-term care population.

There are also limits to the availability of linked social care and public health service data, although some mapping could occur using a small-area approach.

3.4   Outcome measures

The NHS Outcomes Framework uses the EQ-5D HRQOL measure as the overarching indicator for Domain 2. Some recent analysis in the Quality and Outcomes Research Unit (QORU) suggests that some aspects of quality of life valued by people with long-term conditions may not be fully detected by EQ-5D. One of the issues is that EQ-5D focuses on personal functioning. However, some interventions for people with LTCs are provided to help them manage the consequences of their condition, given that restoration of personal

functioning is not possible. If this is the case, the use of EQ-5D alone is likely to result in under-measurement of the impact of NHS activity on quality of life for people with LTCs.

4   Additional methods to improve attribution

There are a number of methods that can be used to tackle the problems of un-observables and lagged-effects when the main indicator (HRQOL) is drawn from sample data.

There are also other methods for assessing changes in efficiency over time that do not involve population-level surveys of HRQOL. For example, local health organisations could be audited to determine the extent to which, where relevant, more cost-effective interventions have displaced less cost-effective existing interventions. A further consideration of these methods is outside the remit of this paper.

4.1   Un-observables

4.1.1   Proxy variables – data linkage between survey and other data sets

The most straightforward method is to find 'proxy' variables for the un-observable factors. Many of the relevant factors are not collected in the GPPS (although more are available in the Health Survey for England - HSE), but these data could be linked with other datasets. One method is to exploit the large size of the GPPS and use sub-sample average values of HRQOL for an analysis at small-area level (using the ONS geographical classification). Many routine administrative datasets and the Census can be mapped by ONS geography. Alternatively, it might be possible to collect individual respondent addresses in the GPPS and map people by postcode or to link individual respondents directly to activity datasets (such as HES) using the person's NHS number or equivalent.

It would also be important to map in some indicator of NHS activity by small area or directly to individuals: e.g. using programme budgets data by CCG and patient/condition group, or expenditure data. Cost-weighted activity data is another possibility. Ideally, activity associated with the care of people with LTCs would be used, although there are no doubt difficulties in identifying this expenditure. More aggregated data could be used at the cost of some additional 'noise' in the estimation.

4.1.2   Panel/longitudinal datasets

Panel data have repeated observations for the same individual over time. They allow estimation of differences in HRQOL when time-invariant or person-invariant factors are missing or un-observable.[8] For example, there may be clusters of environmental factors in certain locations that have an impact on HRQOL that is not picked up by included needs

---

[8] Using, for example, a difference-in-difference approach.

factors. If these effects are time-invariant (or at least change at a slow rate), then a panel data estimation can adjust for these missing factors.

Making the GPPS a follow-up survey would be an important component in constructing a suitable panel dataset. A mechanism to link respondents with NHS activity and relevant control factors would also be required.

IV analysis would greatly benefit from the availability of a panel dataset, rather than run on cross-sectional data. With a panel dataset, an IV production function could be estimated that explicitly differentiated efficiency rates for different time periods. In other words, it should be possible to estimate an efficiency trend for a number of years. For example, estimating a rolling 3- or 5-year efficiency trend might be possible.

A repeated/longitudinal dataset helps to minimise sampling error since the same respondents are surveyed.

4.1.3 Expert assumptions about long-term factors
Some of relevant factors that are not available in a GPPS-linked dataset would include long-term prevalence rates for main disease groups and other similar factors. In some cases, without any good proxies, the analysis would have to fall back on expert assumptions and 'off-model' projections about long-term trends in these rates.

Clearly, the results of an analysis of the progress of the NHS with regard to HRQOL will be sensitive to the nature of these assumptions. Some sensitivity analysis would help cast light on the extent to which the assumptions drive the results. Nonetheless, to just ignore these factors because of a lack of data is to make an implicit assumption that they do not matter and/or are unchanging through time.

5   Interpreting the results for accountability purposes

5.1   Efficiency or total activity
Regarding questions of accountability (of NHS England), the 'progress' as required by the Mandate could be identified as any improvement in the total impact of NHS activity, both in terms of efficiency and/or total resourcing: i.e. a positive value of $\Delta\pi_t x_t + \pi_t \Delta x_t$ in (1). However, to some extent, the size of the NHS budget – and hence overall expenditure – is outside the control of NHS England. For this reason, it might be better to judge how much progress has been made by measuring the change in efficiency, $\Delta\pi_t$. These alternative accountable objectives for NHS England would have quite different implications in a period where real NHS expenditure was falling.

To be explicit about the options, overall 'progress' could be measured as:

(1) An improvement of current HRQOL stemming from an increase in just current efficiency

(2) An improvement of current HRQOL stemming from an increase in total current resourcing (NHS budget) and/or efficiency

(3) An improvement of current HRQOL regardless of cause

Generally speaking, NHS England has most control with respect to the first option and least control with respect to the last.

We might also wish to be clear about the timing issue and judge progress in terms of:

(4) An improvement in the (upward) trend in HRQOL over some period as a result of a change in the trend of NHS efficiency and also, possibly, NHS budget.

5.2   National or local improvement

The impact of NHS activity can be assessed in aggregate terms at the national level: i.e. changes in the average impact or adjusted HRQOL score over time at the England level. However, with considerable regional variation in the NHS, accountability concerns could be extended to include the *distribution* of progress gains across the country. For example, if one area performed very poorly, but this was more than offset by significant improvement in another area, could NHS England be regarded as delivering on its objectives? In other words, the objective may be to improve overall HRQOL but also to ensure that progress is reasonably even across the country.

As well as being a broad accountability tool, progress measures have considerable value if they can be used to help guide improvements in performance. In this sense, progress could be assessed in comparative terms, between areas or other organisational units. The methods outlined above could be applied at, for example, CCG level to assess how the impact of NHS activity on HRQOL differed between areas. In other words, given the factors beyond NHS control in any locality, and given the level and nature of NHS activity, this assessment would indicate which areas were above and which below expected levels of HRQOL in their populations. These results would have more limited use in helping to understand *why* localities were away from expected values, but would be a mechanism to facilitate or trigger further investigation.

Both the residual difference and production function approaches would give an equation (i.e. (1) above) for predicting HRQOL as it is affected on average by needs and other control factors and, where relevant, NHS activity levels. As such, if NHS activity and needs/other factors in a given locality had the *same* impact on HRQOL as the England average, actual observed HRQOL rates would be the same as those predicted by the estimated equation. We would infer that areas showing significantly lower actual values of HRQOL were either less efficient at producing HRQOL, were affected by some unobserved need characteristics

that differed markedly from the England average or, when using the residual difference approach, had different levels of NHS activity unrelated to local needs characteristics.

6   Concluding points

Assessing the progress of the NHS in terms of the *HRQOL outcomes* it produces for people with LTCs goes directly to the heart of what matters to most of those people. The problem with final outcome measures is being able to attribute changes in HRQOL to the activity of the NHS. Only considering (un-adjusted) changes in average population HRQOL scores is not going to be informative for NHS accountability purposes, because some of the drivers of change are outside the control of the NHS.

In this paper, some tentative suggestions have been made as to statistical methods that might be used to make this attribution. Two options were proposed: a *residual difference* approach and a *full production model* approach. The former method starts with measuring changes HRQOL rates in the population (over time) and then removes the effects of changes in needs and other control factors over that period and/or across observation 'units' in a standardisation process. Any remaining change is linked to 'NHS activity' (other than changes in NHS activity driven directly by changes in need).

The latter method uses data on NHS activity directly to estimate the relationship between (changes in) NHS activity and (changes in) HRQOL rates. This method would estimate the average gain in HRQOL for each additional £1 of NHS activity and compare this with average gain estimated for the following year (or in comparing one regional with another).

A hybrid method involves using a full production function estimation to calculate an expected HRQOL score for the following year. These expected values would be compared with the actual HRQOL score in that following year, with any difference assumed to be due to improvements in NHS efficiency.

In general, the residual difference approach is the most straightforward to implement. In our previous report we suggested a standardisation process that uses needs data from the GP patient's survey. This standardisation process could be further developed using data on needs factors drawn from a number of other sources that are currently available. Thus far, the standardisation process outlined in our report has been exemplified for year-on-year national changes in HRQOL but could be used to compare performance between organisational units such as CCGs. This extension would involve generating expected HRQOL scores for each unit using the standardisation process, and then subtracting the expected from the actual HRQOL score. The (residual) difference would identify areas with better (higher) HRQOL scores in their populations than the national average, and those with worse outcomes.

The production function approach is arguably the most theoretically robust method and the one that provides direct efficiency comparisons. However, it is a complex method that is highly demanding on data and could not be implemented at present given the data that are available. A particular difficulty would be finding good measures of (appropriate) NHS activity that could be linked with the GPPS. This approach (and the other) would also benefit from the GPPS taking a longitudinal format. Conducting the assessment on a rolling timeframe would facilitate the interpretation of the data.

Table 2 below summarises the pros and cons of each method.

Table 2. Comparing attribution methods

	Residual difference	Full production model
Pros	• Straightforward to implement • Likely to be more credible • Side-steps missing lagged activity effect problems (e.g. primary/secondary prevention effects)	• More theoretically robust – can be used to estimate the marginal effectiveness of NHS activity on HRQOL • Less susceptible to un-observable effects • Allows impacts on HRQOL to be differentiated by activity type
Cons	• Can only be used to consider changes in the effectiveness of NHS activity (through time and/or between units). Need to assume that the effect of external needs factors on activity remains constant. • More susceptible to un-observables problem • Missing lagged need effects can be problematic	• Computationally demanding • High data requirements, esp: need linked activity data, linked lagged activity data, and instrumental variables • May lack credibility

The particular types of analysis will also depend on choices of whether it is 'efficiency' or some other indicator that is the basis for accountability. We would also suggest that these methods would give useful comparative performance insights when applied across different areas or specific NHS organisational units.

Another relevant decision is on whether to include indirect as well as direct impacts of NHS activity in our assessment. Clearly, including potential indirect impacts is more comprehensive in theory but it does undermine our ability to control for external factors that might affect our assessment of the direct impacts. In other words, we trade a potential loss of accuracy for a gain in comprehensiveness.

Any method and associated outcome measure(s) would need to have credibility with a range of interested parties (clinicians, managers, politicians and the public). Credibility will hinge on whether people consider the approach to be robust, valid, and reliable, and to be able to make this judgement, the method needs to be sufficiently transparent and straightforward to assess. There is likely to be a conflict in this regard because the attribution and measurement problems are significant in this case and would likely require more complex solutions. And yet complexity itself often undermines credibility. The *residual difference* method is more likely to engender credibility. Use of EQ-5D as a measure is also more contentious in relation to long-term conditions, although clearly very well-established in relation to acute and curative healthcare. The indicator could be supplanted with a more appropriate (credible) generic measure of care-related quality of life for people with long-term conditions if and when such a measure is developed.

Although there are practical challenges in using these dis-entangling or attribution methods to develop performance indices, we would nonetheless underline the importance of controlling for need and other factors outside NHS control when making accountability judgements about NHS performance. In view of the different potential impacts of NHS activity, the choice of appropriate control factors in each case and the data-related challenges, we would suggest that several performance indices are used. Some would use more control factors and so be focused on the quality of NHS care for people with LTCs – the direct impacts. Other might use fewer controls and therefore not limit assessment of potential indirect effects. The set of indices would be used to support an interpretation of the performance of NHS England.

7   References

Claxton, K., S. Martin, M. Soares, N. Rice, E. Spackman, S. Hinde, N. Devlin, P. C. Smith and M. Sculpher (2013). Methods for the Estimation of the NICE Cost Effectiveness Threshold. CHE Research Paper 81. York, University of York

Fernandez, J.-L., A. Hughes, K. Watson, J. Forder and R. Fitzpatrick (2013). Report on using the GPPS to assess trends in EQ-5D scores for people with long-term conditions. QORU Discussion Paper, University of Kent and South East Public Health Observatory (Public Health England)

Forder, J., J. Malley, A.-M. Towers and A. Netten (2013, forthcoming). "Using cost-effectiveness estimates from survey data to guide commissioning: an application to home care." Health Economics Forthcoming.

8 Appendix

8.1 Methods for attribution

The full production model is

$$y_t = \beta_0 + \pi_t(q_t, x_t)x_t(z_t, m_t) + \sum_{l=1}^{L} \pi_{t-l}(q_{t-l})x_{t-l}(z_{t-l}, m_{t-l}) + \beta_{m1}m_t$$

$$+ \sum_{l=1}^{L} \beta_{ml}m_{t-l} + \beta_{n1}n_t(q_t, x_t, m_t) + \sum_{l=1}^{L} \beta_{nl}n_{t-l}(q_{t-1}, x_{t-l}, m_{t-1})$$

(5)

where y_t is current HRQOL in the population, x_t is (a vector of) NHS activity[9], m_t are exogenous underlying control factors, n_t are internal factors that are affected by NHS activity and exogenous control factors, and z_t are external factors that influence NHS activity but do not have a direct effect on HRQOL.[10] The function π_t is the co-efficient measuring the direct impact of NHS activity on HRQOL. The term q_t reflects how efficiently NHS activity produces HRQOL. Similarly, the β_m terms are the size of the impact of external control factors on HRQOL. Finally, the β_n terms measure the size of the indirect impacts of NHS activity. The model incorporates lags of between 2 and *L* years.

The first step of a *residual difference* approach is used to calculate (and so later remove) the effects of changes in external needs factors over time. In principle, we could estimate the following function at baseline time (t^B) for this purpose:

$$y_{t^B} = \alpha_0 + \alpha_{21}m_{t^B} + \sum_{l=1}^{L} \alpha_{2l}m_{t^B-l} + \alpha_{31}n_{t^B}(x_t, m_{t^B})$$

$$+ \sum_{l=1}^{L} \alpha_{3l}n_{t^B-l}(x_{t^B-l}, m_{t^B-1}) + \epsilon_{t^B}$$

(6)

The problem is that some of the potential effect of NHS activity works through the n_t term which we do not want to remove.

This problem can be avoided if these control factors can be decomposed in the data into an external need and activity-led component, respectively: $n_t^m(m_t)$ and $n_t^x(x_t)$ where the former is not influenced by NHS activity. We can therefore re-write the current HRQOL production function (5) as:

---

[9] For convenience we refer to NHS activity as a single term, but actually this encompasses a vector effect of many different components of NHS activity.

[10] When we discuss external control factors in the main text we are referring to the subset of all exogenous factors m that are available for the analysis.

$$y_t = \beta_0 + \pi_t(q_t, x_t)x_t(z_t) + \sum_{l=1}^{L} \pi_{t-l}(q_{t-l})x_{t-l}(z_{t-l}) + \beta_{m1}m_t \tag{7}$$

$$+ \sum_{l=1}^{L} \beta_{ml}m_{t-l} + \beta_{n1}n_t^x(q_t)x_t(m_t) + \beta_{n1}n_t^m(m_t)$$

$$+ \sum_{l=1}^{L} [\beta_{nl}n_{t-l}^x(q_{t-l})x_{t-l}(m_t) + \beta_{nl}n_{t-l}^m(m_{t-l})]$$

The *change* in population HRQOL over time will be driven by changes in control factors, which work directly and also through their influence on the level of NHS activity. Also, the change in HRQOL over time will be due to changes in NHS efficiency. Taken together, HRQOL change is therefore:

$$\Delta y_t = \Delta\pi_t(\Delta q_t)x_t + \pi_t(q_t)\Delta x_t(\Delta m_t) + \pi_t(q_t)\Delta x_t(\Delta z_t) \tag{8}$$

$$+ \sum_{l=1}^{L} [\Delta\pi_{t-l}(\Delta q_{t-l})x_{t-l} + \pi_{t-l}(q_{t-l})\Delta x_{t-l}(\Delta m_{t-l})$$

$$+ \pi_{t-l}(q_{t-l})\Delta x_{t-l}(\Delta z_{t-l})] + \beta_{m1}\Delta m_t + \sum_{l=1}^{L} \beta_{ml}\Delta m_{t-l} + \beta_{n1}\Delta n_t^x(\Delta q_t)x_t$$

$$+ \beta_{n1}\Delta n_t^m(\Delta m_t) + \sum_{l=1}^{L} [\beta_{nl}\Delta n_{t-l}^x(\Delta q_{t-l})x_{t-l} + \beta_{nl}\Delta n_{t-l}^m(\Delta m_{t-l})]$$

With need factor decomposition, we instead estimate the following (reduced-form) function in step one, as follows. In this case, we used adjusted internal control factors $\hat{n}_{t^B}^m$, modelled to reflect changes in relevant exogenous factors.

$$y_{t^B} = \theta_0 + \theta_{21}m_{t^B} + \sum_{l=1}^{L} \theta_{2l}m_{t^B-l} + \theta_{31}\hat{n}_{t^B}^m(m_{t^B}) \tag{9}$$

$$+ \sum_{l=1}^{L} \theta_{3l}\hat{n}_{t^B-l}^m(m_{t^B-l}) + \epsilon_{t^B}$$

Note that the coefficients θ on the need variables in this case will embody endogeneity bias from baseline activity effects as $\epsilon_{t^B} = \epsilon_{t^B}(x_t(m_t) \dots)$. The coefficients, in other words, will not just reflect pure need effects. This is a problem if we seek to isolate absolute activity effects at any given time – because some of the effect would be controlled out via (9) – but not a problem if we are concerned with changes in the effect of NHS activity after baseline time (t^B).

In the second step, we would use this reduced-form function to predict changes in HRQOL that stemmed from changes in control factors (given NHS activity at baseline time t^B) i.e.:

$$\Delta\hat{y}_t(\Delta m) = \theta_{21}\Delta m_t + \sum_{l=1}^{L} \theta_{2l}\Delta m_{t-l} + \theta_{31}\Delta\hat{n}_t^m(\Delta m_t) + \sum_{l=1}^{L} \theta_{3l}\Delta\hat{n}_{t-l}^m(\Delta m_{t-l}) \tag{10}$$

This predicted difference would be used to approximate for control factor-led changes in (8) for time $t \geq t^B$:

$$\pi_t(q_{t^B}) \Delta x_t (\Delta m_t) + \sum_{l=1}^{L} \left[\pi_{t-l}(q_{t^B-l}) \Delta x_{t-l}(\Delta m_{t-l}) \right] \tag{11}$$

$$+ \beta_{m1} \Delta m_t + \sum_{l=1}^{L} \beta_{ml} \Delta m_{t-l} + \beta_{n1} \Delta n_t^m (\Delta m_t) + \sum_{l=1}^{L} \beta_{nl} \Delta n_{t-l}^m (\Delta m_{t-l})$$

$$\cong \Delta \hat{y}_t (\Delta m)$$

This is an approximation because (10) encompasses both the direct external need effects – working through the β terms – and the effects of external need in generating a change in activity levels for constant efficiency q_{t^B}, working through $\pi_t(q_{t^B})$. As noted, this is not a problem when looking at differences as long as we assume that change in activity resulting from pure external need effects remains constant through time (according to the relationship estimated at baseline i.e. with q_{t^B}). It would of course be possible to re-estimate (9) on a regular basis.

We can then substitute using $\Delta \hat{y}_t (\Delta m)$ into (8) to get the total change in HRQOL:

$$\Delta y_t \cong \Delta \hat{y}_t (\Delta m_t) \tag{12}$$

$$+ \Delta \pi_t (\Delta q_t) x_t + \pi_t \Delta x_t (\Delta z_t) + \sum_{l=1}^{L} \left[\Delta \pi_{t-l}(\Delta q_{t-l}) x_{t-l} + \pi_{t-l} \Delta x_{t-l}(\Delta z_{t-l}) \right]$$

$$+ \beta_{n1} \Delta n_t^x (\Delta q_t) x_t + \sum_{l=1}^{L} \beta_{nl} \Delta n_{t-l}^x (\Delta q_{t-l}) x_{t-l}$$

$$\cong \Delta \hat{y}_t (\Delta m_t) + \Delta \hat{y}_t (\Delta q_t)$$

i.e. changes in HRQOL decomposed into control factor-led changes $\Delta \hat{y}_t (\Delta m)$, changes due to efficiency differences through time i.e. due to Δq_t and Δq_{t-l}, and changes in activity levels for non-control factor related reasons, $\Delta x_t (\Delta z_t)$. It follows that if efficiency were constant through time i.e. $\Delta q_t = 0$ and $\Delta q_{t-l} = 0$ for all l, and non-need activity levels were unchanged, then the observed change would be due entirely to the effect of a change in need i.e. $\Delta y_t = \Delta \hat{y}_t (\Delta m)$. As such, the difference between $\Delta \hat{y}_t (\Delta q_t) \cong \Delta y_t - \Delta \hat{y}_t (\Delta m)$ is a measure of the change in impact of (non-need) activity through time, potentially including efficiency effects. The *size* of the change $\Delta \hat{y}_t (\Delta q_t)$ is difficult to interpret directly, but it could be used to create a performance index, for example, $\Delta \hat{y}_t (\Delta q_t)/y_0$, that would indicate performance changes through time.

The full *production model* approach would estimate a full HRQOL function at GP patient survey respondent level. It would require NHS activity, including past activity, to be determined and linked to individual respondents. An IV estimation could be conducted if any of the control factors were missing (using the z_t factors as instruments). To measure the full effects of NHS activity, we would again need to decompose the non-external control factors, *n*. The function to be estimated would be:

$$y_t = \sigma_0 + \sigma_{1t}x_t(z_t) + \sum_{l=1}^{L}\sigma_{1t-l}x_{t-l}(z_{t-l}) + \sigma_{2t}m_t + \sum_{l=1}^{L}\sigma_{2t-l}m_{t-l} \qquad (13)$$

$$+ \sigma_{3t}\hat{n}_t^m(m_t) + \sum_{l=1}^{L}\sigma_{3t-l}\hat{n}_{t-l}^m(m_{t-1}) + \eta_t$$

The parameters σ_{1t} and σ_{1t-l} would be directly estimated. Using (7), we have (marginal) efficiency as:

$$\frac{\partial y_t}{\partial x_{t-l}} = \pi_{t-l}(q_{t-l}) + \beta_{nl}n_{t-l}^x(q_{t-l}) = \sigma_{1t-l}, \qquad \forall l = 0,\dots,L \qquad (14)$$

i.e. the impact on HRQOL of each (additional) £1 spending on NHS activity is the estimated parameter σ_{1t} and lagged activity by σ_{1t-l}. The change in marginal efficiency is $\frac{\partial^2 y_t}{\partial x_{t-l}\partial t} = \frac{\partial \sigma_{1t-l}}{\partial t} \cong \Delta\sigma_{1t-l}$, $\forall l = 0,\dots,L$, or:

$$\frac{\partial^2 y_t}{\partial x_{t-l}\partial t} = \frac{\partial\left[\frac{\partial y_t}{\partial x_{t-l}}\right]}{\partial t} \cong \Delta\pi_{t-l}(\Delta q_{t-l}) + \beta_{nl}\Delta n_{t-l}^x(\Delta q_{t-l}) = \Delta\sigma_{1t-l}, \qquad \forall l = 0,\dots,L \qquad (15)$$

for each lagged period of NHS activity l, including the current period (no lags).

These estimates could also be used to calculate *total* change in HRQOL over time, being the product of the (marginal) change in efficiency and activity:

$$\Delta\hat{y}_t(\Delta q_t) \qquad (16)$$
$$\cong [\Delta\pi_t(\Delta q_t) + \beta_{n1}\Delta n_t^x(\Delta q_t)]x_t$$
$$+ \left[\sum_{l=1}^{L}\Delta\pi_{t-l}(\Delta q_{t-l}) + \sum_{l=1}^{L}\beta_{nl}\Delta n_{t-l}^x(\Delta q_{t-l})\right]x_{t-l}$$
$$\cong \Delta\sigma_{1t-l}x_t + \sum_{l=1}^{L}\Delta\sigma_{1t-l}x_{t-l}$$

This estimate should be close to the total change estimate, $\Delta\hat{y}_t(\Delta q_t)$, made using the residual difference approach in (12) when two conditions hold. First, the estimated marginal productivity (efficiency) σ_{1t-l} would have to be close to actual average productivity (i.e. that the additional HRQOL gain for each £1 spent was approximately equal to σ_{1t-l}). Second, there would have to be no changes in the level of activity for non-need reasons i.e. $\Delta x_t(\Delta z_t) = 0$ and $\Delta x_{t-l}(\Delta z_{t-l}) = 0$.

When these two conditions do not hold, these two approaches are likely to give different results. Nonetheless, both can be used to inform assessments of NHS performance. With the production function approach, we would be focusing on changes through time in marginal productivity rather than the total changes in HRQOL after controlling for need with the residual difference method.

8.2 Instrumental variables

An essential requirement of (OLS) regression is that the error term is uncorrelated with the explanatory variables. This condition is violated when we have missing control factors that partly determine the NHS activity used by a patient. A way to resolve this problem is to use an instrumental variables approach. The predicted value of NHS activity is used in the main estimation where predicted NHS activity is based on a first-stage regression model that uses factors ('instruments') which are not correlated with the main model error.