oxera
compelling economics

Agenda
Advancing economics in business

Setting the efficiency factor for NHS Trusts: has Monitor got the balance 'right'?

As sector regulator for health services in England, Monitor is responsible for designing effective incentive mechanisms for NHS management to promote the interests of patients. To that end, it is tasked with setting prices for a range of healthcare services that reflect efficient costs. As part of the 2016/17 national tariffs, a sector-wide efficiency factor of 2% has been proposed. Here we explore the methodology underpinning the efficiency factor and compare it with approaches and frameworks considered by utility regulators

The 2015/16 budget for the NHS in England is £116.4bn—equivalent to around 7% of UK gross domestic product (GDP). Faced with a combination of rising costs, increasing patient demand and a UK government commitment to run a budget surplus in 'normal' times, NHS funding has come under increasing pressure in recent years. Consequently, incentivising and delivering efficiency improvements have become key priorities for the NHS.

Under the Health and Social Care Act 2012, Monitor[1] is responsible for setting national tariffs on an annual basis for a range of healthcare services—for example, services involved in providing care to patients admitted to hospital, outpatient care, and emergency care. To drive value for money, Monitor seeks to set prices that reflect efficient costs, and incentivises NHS Trusts to reduce costs over time by finding more efficient ways of working.[2] This needs to be balanced against the need for the Trusts to maintain and improve their services in terms of safety, quality, level of integration, and access.

For 2016/17, Monitor has set national tariffs based on rolled-over prices from the previous year.[3] A sector-wide efficiency factor of 2% is proposed to be overlaid on this.[4]

This article reviews the methodology underpinning Monitor's analysis in deriving the efficiency factor,[5] contrasting it with approaches and frameworks considered by utility regulators in the UK, and makes some suggestions for the development and application of Monitor's analysis.

Contrasting Monitor with utility regulators

In a simple textbook world, healthcare providers in England would be free to set prices. Health market dynamics would be such that providers would have incentives to deliver services efficiently and effectively. Patients would identify which providers offer the best quality of care, and at the lowest price, and it is these providers who would gain at the expense of those providing poorer value for money.

However, healthcare in England does not work like this. First, most NHS care is free at the point of delivery. As such, patients cannot know whether the services offer value for money. Second, while patients may exercise some choice over elective care, they have much less choice in relation to emergency care. Given these factors, many NHS treatments are subject to a national tariff, against which providers are remunerated.

The combination of choice (exercised by elective patients seeking higher-quality treatment), and administrative action by the various regulators in health and social care,[6] is designed to produce a system whereby NHS institutions are presented with an incentive to improve their efficiency of operation. This then feeds into the tariff-setting process, which mirrors the RPI - X approach adopted by the regulated monopoly utilities.

Undertaking comparisons across peers and setting efficiency targets is common practice among utility regulators as a means of replicating some of the outcomes of a competitive market. In particular, it can create an incentive for the providers to seek out and implement efficiency gains while also passing those gains on to consumers sooner than might otherwise be the case.

In this context, utility regulators often use cost benchmarking to set cost- or price-reduction targets for organisations to achieve during the (typically multi-year) price control period, or as a basis against which to monitor the organisation's performance. Many organisations also undertake benchmarking to determine their own internal efficiency challenges for business planning purposes (or in presenting well-justified business plans to the regulator).

In the box below, we compare the economic framework commonly considered by utility regulators to determine the efficiency factor with the approach adopted by Monitor.

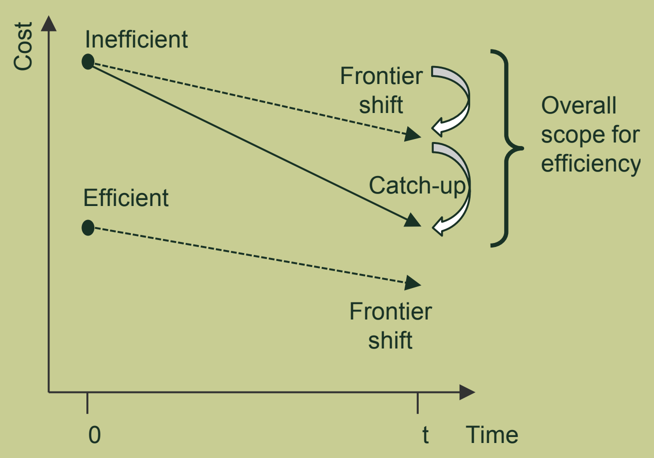
Monitor's approach to deriving the efficiency factor for 2016/17 involves comparing cost performance at the organisation level (in this case, NHS Trusts) and over time using econometric modelling. Its use of a panel data[7] framework is in line with the majority of the UK utility regulators' approaches. Ofwat (water),[8] Ofgem (energy),[9] and the Office of Rail and Road (ORR)[10] have all relied heavily on panel data modelling techniques in their most recent price control reviews.

In Monitor's case, modelling at the overall Trust (i.e. aggregate) level is a pragmatic decision. This is because the data from which the efficiency factor is derived could be subject to significant error or short-term volatility at a more disaggregated level, at least with the currently available

Comparing the approaches of utility regulators with that of Monitor

Utility regulators: catch-up and frontier-shift efficiency

The scope for efficiency gains is commonly thought of as being composed of frontier-shift and catch-up efficiencies (which utility regulators consider either explicitly or implicitly when setting efficient cost allowances). The two concepts of efficiencies are distinct and illustrated in the graph below, where the forward-looking cost profiles set for an inefficient and efficient company are depicted over time.

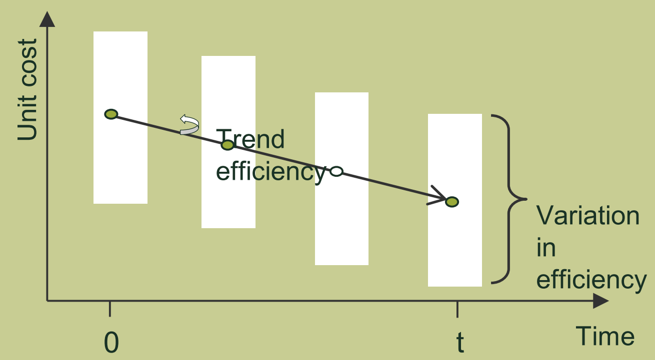


Catch-up efficiency relates to the gap between a specific company's performance and the best-practice benchmark that is set by relatively efficient (frontier) companies.

Frontier-shift relates to productivity gains that are possible for even the most efficient firm in the industry due to the adoption of new technologies, improved management practices, etc. The concept of the frontier shift is 'dynamic' given that, over time, different companies may leap-frog each other to drive the frontier forward.

Monitor: trend and variation in efficiency

Monitor uses a slightly different framework to assess the efficiency factor. This factor is likewise composed of two elements: trend and variation in efficiency, as depicted in the figure below.



Trend efficiency is the sector-wide efficiency improvement for the average Trust over time, and, by construction, could capture a combination of catch-up and frontier-shift improvements.

Variation in efficiency is similar to the concept of catch-up efficiency, and refers to the gap that less efficient Trusts have to close compared with their more efficient counterparts. Monitor assumes that this gap does not vary over time, but does vary by Trust.

Based on its analysis, Monitor has estimated an efficiency factor of between 1.5% and 2.5%, made up of a trend efficiency of 1.4% and a variation in efficiency of up to 1.1%. It notes that an efficiency factor in the region of 2% is appropriate given the scale of the financial challenge that it faces in 2016/17 and the state of the provider finances.

Source: Oxera, based on Monitor (2016), '2016/17 National Tariff Payment System: A consultation notice', February.

data.[11] In a utility regulation context, once a mechanism for collecting comparable and consistent data is established, modelling at different levels of aggregation is typically considered[12] because aggregated and disaggregated modelling have their advantages and disadvantages.[13]

In terms of benchmarking approaches, Monitor has used a panel data stochastic frontier analysis (SFA)[14] model and a random effects (RE) model.[15] Both of these have been used in a regulatory context in the UK and Europe. For example, the ORR[16] and Ofcom[17] have considered a panel SFA approach as part of the cost assessment toolkit. In Europe, SFA (alongside data envelopment analysis, DEA) appear to be the commonly used techniques.[18] Similarly, the ORR, Ofwat and Ofgem have considered a RE model in the most recent price control reviews.[19]

Finally, in contrast to the majority of other utility regulators, Monitor has incorporated service quality levels (e.g. quality of care based on patient satisfaction surveys) within cost assessment directly. Regulators such as Ofwat[20] and Ofgem[21] have recently indicated that they are seeking to integrate customer outcomes and quality of service within their cost benchmarking framework.

In summary, while the economic framework considered by Monitor differs slightly from utility regulators, the benchmarking approaches employed are broadly in line with regulatory precedents. In addition, Monitor integrates service quality levels within cost assessment, a development that other regulators are seeking to implement in future price reviews.

Reviewing Monitor's efficiency methodology

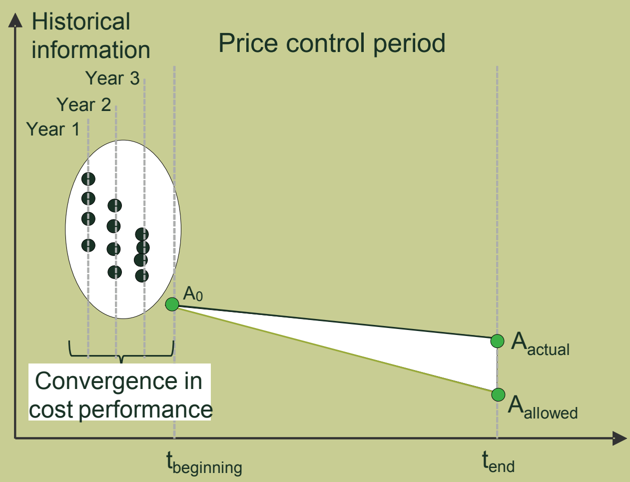
Here, we briefly discuss some areas where Monitor's methodology could be developed further, or where the application of the efficiency factor may need careful consideration.[22]

- **A uniform efficiency factor may not be appropriate for all Trusts**. Monitor's 2% efficiency factor (based on a combination of catch-up and frontier-shift efficiencies) applies to all Trusts. However, this may not be appropriate for some of the Trusts. For example, for those in the middle of the efficiency spectrum, which have historically improved their efficiency by catching up to best practice, using historical performance to inform the target could be challenging since some of the historically achieved efficiency gains may not be replicable.[23]
- **The implicit assumption that Trusts have not converged in performance may lead to upward bias in the efficiency factor**. Monitor's efficiency analysis seeks to determine the average efficiency improvements achieved by the Trusts over the historical period and applies this to the sector-average efficient prices. Monitor's econometric models assume that there is no convergence in the performance of the relatively inefficient Trusts over the period of analysis. While Monitor does mention some evidence on the lack of catch-up,[24] this assumption does not appear to be empirically tested on the data used to determine the average rate of improvement. This could result in potential upward bias in the measured efficiency factor, and the *projected* efficiency factor could be unachievable as the potential for further catch-up will be diminished. The box below illustrates this point further.

Should Monitor adjust for historical catch-up?

Assume an extreme case where complete convergence in cost performance has occurred across the sector over the historical period—i.e. all Trusts end up with the same level of unit cost at the start of the price control period.

In such a scenario, all Trusts start the price control period at the same point. They all end up *underperforming* relative to the allowed prices, as they can only achieve frontier shift, while the efficiency factor includes an element of (the historical) catch-up.



In this scenario, as the prices are set at the average, the efficiency factor derived using historical data would overestimate the scope for improvement over the future, leading to financial difficulties. In this scenario an adjustment for catch-up effects would ensure that Trusts receive a tariff consistent with that of the average efficient Trust. Other regulators have examined this issue in similar contexts.[1]

Source: Oxera analysis. [1] See Oxera (2008), 'Should DTe adjust expected productivity growth for catch-up effects when setting the X factor?', report prepared for Directie Toezicht Energie (DTe), the Dutch energy regulator, April.

- **There is relatively limited cross-checking of results with other efficiency approaches**. The modelling approach adopted by Monitor is econometric and top-down in nature, which has its limitations.[25] To ensure that the assumptions imposed on the model are not driving the result, it would be useful to cross-check the results from models using alternative assumptions, and from other modelling approaches. For example, to determine the energy networks' relative efficiencies in the RIIO price controls, Ofgem used a toolkit of approaches (including econometric benchmarking at different levels of aggregation and bottom-up assessments). In this way, the results from the different approaches can be compared and contrasted, and, based on an understanding of the approaches, some consensus could be reached to identify a robust range for the estimated inefficiencies.

On this last point, in utility regulation, according to surveys, SFA and DEA seem to rank as the most commonly used approaches.[26] In Monitor's case, DEA could be particularly useful as it can readily provide measures of frontier shift (i.e. trend efficiency) that is distinct from Trust-specific efficiency change over time (i.e. variation in efficiency).[27] DEA could also allow Trusts to readily identify their 'peers' and facilitate sharing of best practice. Similarly, more disaggregated, bottom-up or operational evidence is often used as a cross-check in a regulatory context.

What happens next?

The tariff-setting approach for the NHS is an evolving process, as is the case in network utilities. Getting the 'right' efficiency factor involves ensuring that the economic framework, model development process, and efficiency approach underlying the efficiency factor are robust, transparent, and consistent with economic best practice.

Monitor, in setting efficiency targets in its overall regulatory structure, has to strike a careful balance between incentivising quality healthcare and driving down costs. It is conscious that a non-robust application of efficiency analysis may result in unachievable targets.[28] In particular, the management of NHS Trusts need to be given realistic objectives in relation to the evolution of their costs, which will be funded via the future tariffs that Monitor sets. If future prices are set too high, difficult management decisions necessary to improve efficiency may not happen soon enough. If tariffs are set too low, management may become demotivated, as they may not be able to meet the efficiency targets set, whatever they do, with implications for financial sustainability.

In terms of the overall direction and intention of the modelling undertaken by Monitor, there are many merits to its approach in setting the efficiency factor. This article has identified some potential areas for the development and application of Monitor's analysis that it may wish to consider when determining the efficiency factor for 2017/18. Further improvements may also be possible in undertaking future assessments. The suggestions set out above could result in a more robust estimation of current and potential levels of efficiency, both across Trusts and over time. This could help secure the right balance between reducing unit costs and maintaining quality care for patients.

[1] As of 1 April 2016, Monitor is part of NHS Improvement.

[2] Monitor (2016), '2016/17 National Tariff Payment System: A consultation notice', February, p. 31, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/499594/2016-17_national_tariff_statutory_consultation.pdf.

[3] Monitor (2016), '2016/17 National Tariff Payment System: A consultation notice', February, p. 34, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/499594/2016-17_national_tariff_statutory_consultation.pdf.

[4] Other adjustments include cost uplifts, manual adjustments and inflation.

[5] In this article and Oxera's submission, we do not address some of the wider framework issues that require careful examination. For example, there is debate around how effective efficiency incentives are with a soft budget constraint (for example, government is unlikely to let hospitals fail if they do not meet targets), and whether nationally averaged tariffs (and efficiency targets) are appropriate. Another question is whether annual efficiency targets are appropriate or whether a longer timeframe could be used (e.g. to allow Trusts to implement spend-to-save measures).

[6] For example, indirectly, the actions from the Department of Health, Monitor and the Commissioners may have similar incentive properties as a result of the threat of being put into special measures and/or the management being replaced.

[7] Panel data modelling implies that data is modelled across Trusts and over time. A panel dataset increases the number of observations for modelling and can help account for Trust-specific factors that are not suitably accounted for in the modelling. Hence, it can help improve the robustness of the analysis.

[8] See Ofwat (2014), 'Final price control determination notice: policy chapter A3 – wholesale water and wastewater costs and revenues', December, p. 20, http://www.ofwat.gov.uk/wp-content/uploads/2015/10/det_pr20141212wholesale.pdf

[9] See Ofgem (2014), 'RIIO-ED1: Final determinations for the slow-track electricity distribution companies - Business plan expenditure assessment', November, https://www.ofgem.gov.uk/publications-and-updates/riio-ed1-final-determinations-slow-track-electricity-distribution-companies

[10] See Office of Rail and Road (2013), 'PR13 Efficiency Benchmarking of Network Rail using LICB', August, http://orr.gov.uk/__data/assets/pdf_file/0012/5007/pr13-efficiency-benchmarking-of-nr.pdf

[11] For example, in the collection and standardisation of cost data in the NHS.

[12] There are several examples of regulatory precedent for assessing the efficiency factor by modelling from both top-down and more detailed bottom-up perspectives. See, for example, Ofgem (2013), 'Strategy decisions for the RIIO-ED1 electricity distribution price control. Tools for cost assessment', March, https://www.ofgem.gov.uk/sites/default/files/docs/2013/02/riioed1deccostassessment_0.pdf

[13] For example, aggregated modelling may not be affected by cost categorisation/reporting issues, but requires simplifying assumptions in terms of the trade-off between activities and cost components; it may also be viewed as opaque as it typically considers high-level cost drivers. Similarly, while a disaggregated modelling approach can provide richer specifications and could capture organisation heterogeneity in a much better way than aggregated models, it can suffer from cost categorisation and reporting issues.

[14] SFA is an econometric method that estimates a cost function and the inefficiency for each observation (here, NHS Trust), while isolating the effects of uncertainty (i.e. data/modelling errors). SFA has a long academic history, having been developed in 1977. For a more detailed discussion, see Kumbhakar, S.C. and Knox Lovell, C.A. (2000), *Stochastic Frontier Analysis*, Cambridge University Press; and Kumbhakar, S., Wang, H.-J. and Horncastle, A. (2015), *Practitioner's Guide to Stochastic Frontier Analysis*, Cambridge University Press.

[15] RE is a particular form of panel data model, and, in the current context, Monitor's approach assumes that the Trust-specific effects estimated from the RE model (which do not change over time) reflect management inefficiencies.

[16] See Office of Rail and Road (2013), 'PR13 Efficiency Benchmarking of Network Rail using LICB', August, http://orr.gov.uk/__data/assets/pdf_file/0012/5007/pr13-efficiency-benchmarking-of-nr.pdf.

[17] See, for example, NERA (2008), 'The comparative efficiency of BT Openreach', a report for Ofcom, March.

[18] Specific examples include the German energy sector (where both DEA and SFA are used), see Bundesnetzagentur (2008), '2009 report by the Federal Network Agency for Electricity, Gas, Telecommunications, Post and Railway to the European Commission on the German Electricity and Gas Market', http://www.ceer.eu/portal/page/portal/EER_HOME/EER_PUBLICATIONS/NATIONAL_REPORTS/National%20reporting%202008/NR_En/E08_NR_Germany-EN_Summary.pdf; and the Finnish energy sector, where both DEA and SFA used to be considered, see Syrjänen, M., Bogetoft, P. and Agrell, P. (2006), 'Efficiency benchmarketing project B: Analogous efficiency measurement model based on Stochastic Frontier Analysis: Final Report', http://www.sumicsid.com/reg/papers/emv_d5_final.pdf, Gaia Consulting Oy. For a survey of techniques considered internationally, see Jamasb, T. and Pollitt, M. (2001), 'Benchmarking and regulation: international electricity experience', *Utilities Policy*, **9**:3, pp. 107–30.

[19] For Ofwat (2014), 'Setting price controls for 2015-20: Final price control determination notice: policy chapter A3 – wholesale water and wastewater costs and revenues', December, pp. 49–50 , http://www.ofwat.gov.uk/wp-content/uploads/2015/10/det_pr20141212wholesale.pdf. For Ofgem, see, for example, Ofgem (2014), 'RIIO-ED1: final determinations for the slow-track electricity distribution companies. Business plan expenditure assessment', November, pp. 59–60, https://www.ofgem.gov.uk/publications-and-updates/riio-ed1-final-determinations-slow-track-electricity-distribution-companies. For ORR, see, for example, Office of Rail and Road (2013), 'PR13 efficiency benchmarking of Network Rail using LICB', August, p. 4, http://orr.gov.uk/__data/assets/pdf_file/0012/5007/pr13-efficiency-benchmarking-of-nr.pdf.

[20] See, for example, Ofwat (2015), 'Water 2020: regulatory framework for wholesale markets and the 2019 price review', December, section 5.4, http://www.ofwat.gov.uk/consultation/water-2020-consultation/. One of the issues considered in Ofwat's review was the integration of its cost assessment approach with outcomes and quality of service.

[21] In the 2014/15 RIIO-GD1 Annual Report, Ofgem noted that it would work with the GDNs to develop new benchmarking tools to include outputs and consider their effectiveness in order to better reflect the RIIO regulatory model. See Ofgem (2016), 'RIIO-GD1 Annual Report 2014/15', May, https://www.ofgem.gov.uk/system/files/docs/2016/03/riio-gd1_annual_report_2014-15_final.pdf.

[22] See Oxera (2016), 'Monitor's evidence on efficiency for the 2016/17 national tariff', An independent response from Oxera Consulting LLP, March, http://www.oxera.com/getattachment/Latest-Thinking/News/March-2016/Making-the-NHS-more-efficient-Oxera-responds-to-UK/Monitor-proposal-on-efficiency-factor-independent-response-from-Oxera-(1).pdf.aspx.

[23] Monitor could consider additional incentives for the relatively efficient Trusts to incentivise innovation. For a review, see Oxera (2011), 'Encouraging efficiency in regulated sectors: Lessons from 20 years of RPI – X', Report prepared for BT, December, http://stakeholders.ofcom.org.uk/binaries/consultations/wlr-cc-2011/annexes/Encouraging_efficiency.pdf.

[24] Monitor (2016), '2016/17 National Tariff Payment System: A consultation notice', February, p. 7, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/499594/2016-17_national_tariff_statutory_consultation.pdf

[25] In particular, in standard econometric models as considered by Monitor, a functional form relating costs to cost drivers has to be assumed. This can be restrictive and potentially lead to biased estimates of the efficiency factor.

[26] See, for example, surveys in Farsi, M., Fetz, A. and Massimo F. (2007), 'Benchmarking and Regulation in the Electricity Distribution Sector', CEPE working paper 54, January, https://www.ethz.ch/content/dam/ethz/special-interest/mtec/cepe/cepe-dam/documents/research/cepe-wp/CEPE_WP54.pdf; and Jamasb, T. and Pollitt, M. (2001), 'Benchmarking and regulation: international electricity experience', Utilities Policy, **9**:3, pp. 107–30.

[27] For more on DEA, see Thanassoulis, E. (2001), *Introduction to the theory and application of Data Envelopment Analysis: A foundation text with integrated software*, Kluwer Academic Publishers, chapter 6.

[28] Monitor stated that 'it is difficult to get the efficiency factor right, and there are problems if we get it wrong. If the efficiency factor is set too high, then prices are too low. This can mean that the business of providing healthcare can become unsustainable'. Monitor (2016b), '2016/17 National Tariff Payment System: A consultation notice. Annex B5: evidence on efficiency for the 2016/17 national tariff', February, p. 3, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/499594/2016-17_national_tariff_statutory_consultation.pdf.