evidence for
better health care

nuffieldtrust

Feeling the crunch: NHS finances to 2020

Sally Gainsbury

August 2016

About this report

As recognised by the NHS's Five Year Forward View, by 2020 the NHS will need to find savings of around £22 billion in order to balance its books. But there has been no clear articulation of how that gap is expected to be closed. The options for doing so include NHS providers becoming more efficient; NHS commissioners reducing the pace at which NHS activity is increasing each year, either through reducing demand or limiting access to care; more funding for the NHS; or some combination of these. This analysis examines different scenarios to determine exactly what it would take to close the gap.

Suggested citation

Gainsbury S (2016) *Feeling the crunch: NHS finances to 2020.* Nuffield Trust

Contents

Key points

- **The NHS in England will struggle to meet the requirement**, set by the Five Year Forward View, **to save £22 billion by 2020**.
- Our analysis examines the funding gap in detail and concludes that even if hospitals and other NHS providers made cost savings of 2 per cent a year, year after year, the **funding gap would still stand at around £6 billion** by 2020–21.
- For the most part, the task of **closing the health service's funding gap will be borne by health care providers making efficiency savings**. From 2011–12 to 2014–15 the target for annual efficiency savings was set at 4% per year. Providers have actually achieved savings closer to 2% over the last three years, resulting in an underlying provider deficit in 2015–16 of £3.7 billion.
- Continuing with **2% annual efficiencies to 2020–21 would not be sufficient to close the funding gap by itself**. The recent 'financial reset' assumes even higher efficiency savings for 2016–17 of 4%. But that would still leave providers with an underlying deficit in 2016–17 of £2.35 billion. Reducing that deficit altogether would require providers to make further efficiencies of 4% in 2017–18 and follow that with efficiencies of 3% in 2018–19. That level of recurrent, sustained efficiency saving **has never been achieved to date and would still require funds to be taken from the Sustainability and Transformation Fund** (S&TF) to balance provider deficits in the meantime.
- The S&TF can only be spent once. **If most of the funds are used to plug the deficit, there will be little money for the transformative service change** that is required to modernise and reshape NHS services for long-term financial sustainability.
- Activity is growing by an estimated 3.1% per year. Even if NHS providers manage to make the huge additional efficiencies set out above, a **sustainable balance can only be brought into the system by 2020–21 if NHS commissioners also manage to curb the rate at which NHS activity is growing** by a third.
- The NHS is relying on service change and new models of care to curb the growth in activity and treat patients more cheaply. This is highly unlikely without access to the S&TF for transformation. As such the two tasks of **huge provider efficiencies and successful commissioner investment in reducing demand growth** need to happen in a **timely and coordinated fashion**.
- Providers are in deficit in part because **the rate they have been paid for the procedures and treatments they carry out –** set by the national NHS tariff – does not cover their costs and **has been cut by an average 1.6%** in cash terms a year over the last six years.
- NHS England has agreed to increase the tariff this financial year, easing the pressure on providers slightly, but **pushing commissioners into deficit for the first time by 2018–19**.
- **If commissioners fail** in their attempts **to reduce the rate at which demand is growing**, or if additional funding cannot be secured, the **NHS will face some unpalatable decisions in order to curb the growth in activity** and bring the books into balance. These could include extending waiting times for treatment, raising the threshold at which patients become eligible for treatment, cutting some services altogether, or closing whole sites or hospitals.
- These rationing dilemmas come immediately after the EU referendum, which may have **heightened public expectations that there will be new investment** for the NHS.

Introduction

The NHS faces a £22 billion funding shortfall four-and-a-half years from now.[1] That is no longer an abstract number designed to scare the NHS into action. Its reality is hitting home already: a £3.7 billion underlying provider deficit in 2015–16; commissioners only balancing their books through one-off, non-recurrent funds; and finally the Department of Health busting its budget by £200 million despite having made over £1 billion worth of technical adjustments and switches.[2]

Yet we still know little about how the £22 billion gap will be closed. Regional Sustainability and Transformation Plans are in development – largely behind closed doors – and are not due to be completed until October this year. A new 'financial special measures' regime is now underway, but that is an attempt to stop the hole getting any deeper, rather than filling it. This analysis is an attempt to set out how the gap might be closed in theory, if things – many things – go well, and to raise some alarms about what might happen in practice if they don't.

The trouble with tariff: or how things went £3.7 billion wrong

The huge financial deficit now pervading NHS hospitals and other providers – around £3.7 billion after the smoke and mirrors of one-off accountancy adjustments have been removed – is testament to the fact that the traditional method of saving NHS cash has run out of steam.[3] That method cannot be put too simply: over the last five years it has boiled down to paying hospitals less – in cash terms – for the same procedure, for each year that passes.

The mechanism for doing that was the NHS tariff, which determines a national price for thousands of treatments and packages of care. Every year between 2010–11 and 2015–16 the tariff was ratchetted down another notch, cutting the cash amount hospitals were effectively paid for each patient by an average 1.6% a year. Factor in NHS-specific inflation, and that was a real-terms cut of 3.8%, year after year.

The rationale for that was to drive waste out of NHS providers; to force hospitals and other services to cut their unit costs in order to stay afloat by ensuring their rapidly diminishing income-per-treatment covered the expense of providing that care (see Figure 1 on the next page). It worked for a few years, but by 2013–14, the pace of provider cost-cutting started to fall behind the pace of the year-on-year cuts to the tariff. To keep up with the cuts in their income, providers needed to make a 4% cut in their operating costs every year from 2011–12.

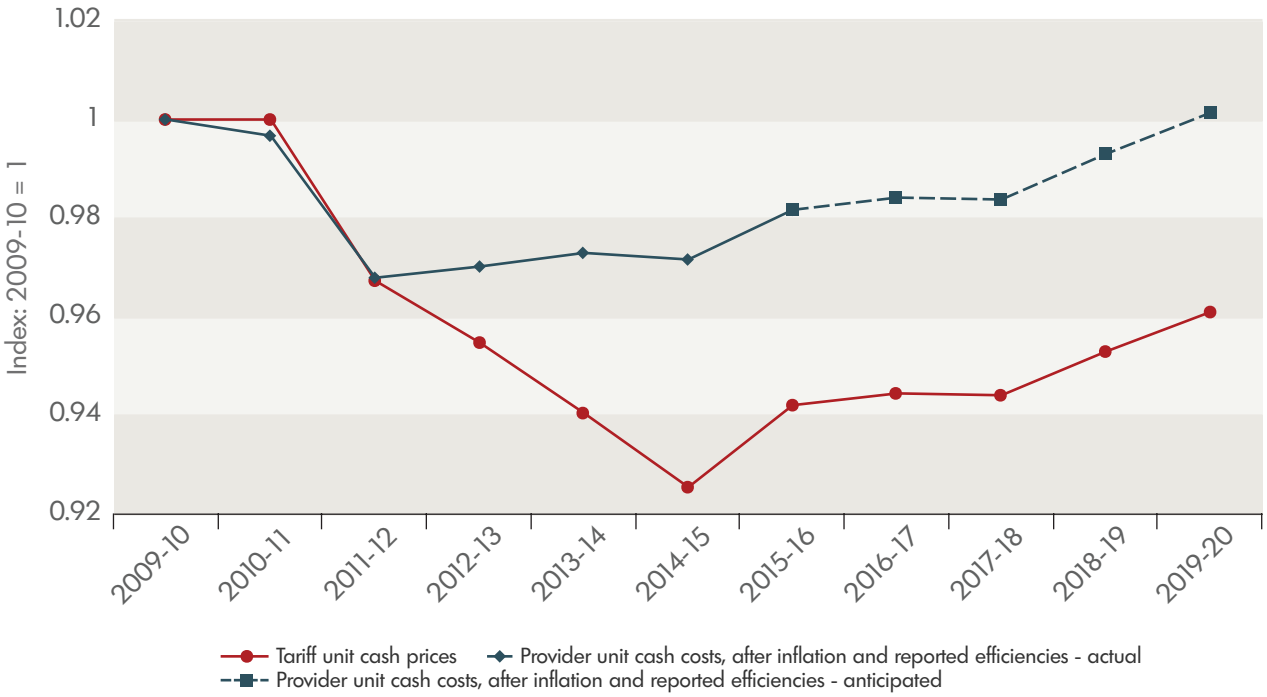
1 The £22 billion gap (£30 billion before additional funding was announced in the 2015 Spending Review) was first set out by NHS England in July 2015.

2 For the most succinct and timely analysis of the Department of Health's accounts for 2015–16, see the blog post written by the Department's former director general for finance, Richard Douglas.

3 NHS accounts report a deficit for the provider sector in 2015–16 of £2.5 billion. That figure follows a significant effort by (and pressure upon) NHS providers to reduce the reported deficit through a series of one-off accountancy adjustments.

Our £3.7 billion underlying deficit figure is drawn from our analysis of provider income and expenditure - detailed later on in this analysis. However, it is also borne out by note 2.7 in NHS Improvement's report on the fourth-quarter finances for the provider sector, which explained that the reported £2.5 billion deficit was after £1.5 billion of financial improvement of which only £300 million was termed 'operational improvement'.

Figure 1: Index of tariff unit prices versus provider unit costs



> By [2015–16], NHS providers were being paid £925 in cash for the same procedure they would have been paid £1,000 to perform in 2009–10: the equivalent of a real-terms cut of 20% to just £800.

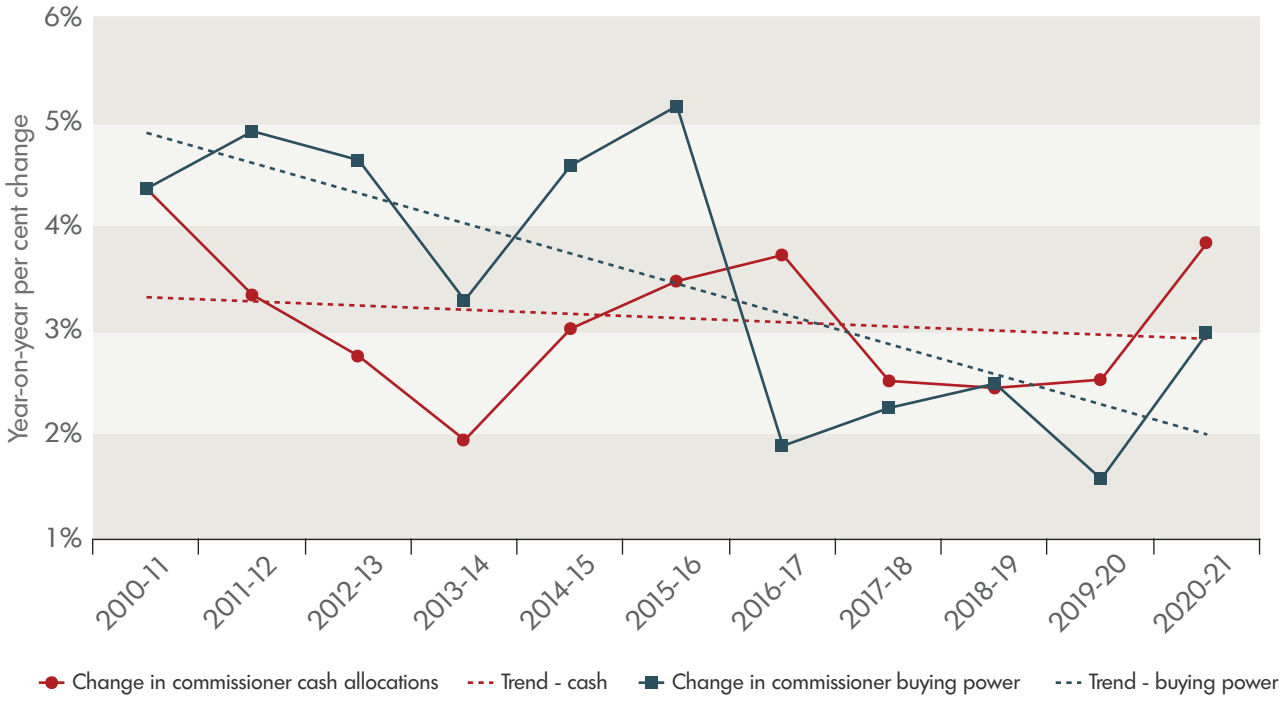
By 2013–14, however, their annual cost-cutting efforts were nearer 2%. As a result, that year, a £600 million underlying deficit emerged in NHS provider accounts.[4] The following year the underlying deficit more than doubled to £1.5 billion, and then in 2015–16 it more than doubled again. By that year, NHS providers were being paid £925 in cash for the same procedure they would have been paid £1,000 to perform in 2009–10: the equivalent of a real-terms cut of 20% to just £800.

This year, things have had to be different. Forced by the sheer size and unsustainability of the provider deficit, tariff prices have been increased, in cash terms, for the first time this decade. The increase is modest – just 1%, which is more than outstripped by the expected inflation in provider costs (determined predominantly by salary, drug and equipment costs), but it will at least mean that in order to keep pace with cuts to the tariff, providers will only need to find 2% cost cuts a year, rather than the 4% required of late. That will not leave NHS providers laughing, but it should mean their deficits will stop increasing as a proportion of their turnover, just as long as they keep up their recent trend of 2% recurrent cost cuts a year.

The more significant impact, however, is on NHS commissioners – NHS England centrally, and clinical commissioning groups (CCGs) locally. For them, the main

4 The underlying provider deficit for 2013–14 was made up of a £108 million deficit reported in the accounts, plus £509 million in revenue support to providers in financial distress. That support was provided in the form of revenue injections, which flattered the reported deficit between provider income and expenditure.

Figure 2: The declining fortunes of commissioners



measure of their spending power is not simply much how much extra cash they get from the government each year, but what direction cash prices in the tariff are moving. When tariff cash prices are cut, commissioner buying power is increased above and beyond the headline increase in their allocations. But when tariff cash prices shift upwards, the reverse is true, and commissioner buying power is diminished (see Figure 2).

So while annual increases in NHS commissioner headline cash allocations are set to remain fairly constant over this current decade – averaging 3.1% a year from 2009–10 to 2015–16 and 3% in the budgets announced from 2016–17 onwards – the cash increase in tariff prices means that 3% cash increase will translate into significantly less buying power than commissioners have been used to: an average of just 2.4% growth a year from 2016–17 onwards, down from an average of 4.5% between 2009–10 and 2015–16.[5]

That is a significant problem for commissioners, because demand for hospital and other NHS services is currently growing at 3.1% a year –considerably faster than commissioners' ability to pay for it, this year and beyond.

So while measures to curtail NHS spending in the first half of this decade centred almost exclusively on reducing the unit cost of care by squeezing ever greater efficiencies out of providers, the second half of this decade – and the quest to make £22 billion worth of savings – will see a sharpening of the incentive for commissioners to curtail spending by scaling back the *quantity* of care purchased, and possibly its *quality* too.

Unit costs up + activity up = commissioners go bust?

We can bring this incentive into sharp relief – and explore its potential implications – by calculating the future cost to commissioners and providers of NHS care if the

5 This figure excludes the Sustainability and Transformation Fund (S&TF), which is discussed in more detail later. When the S&TF is included, commissioner real-terms buying power between 2015–16 and 2020–21 would increase at an average rate of 3% a year.

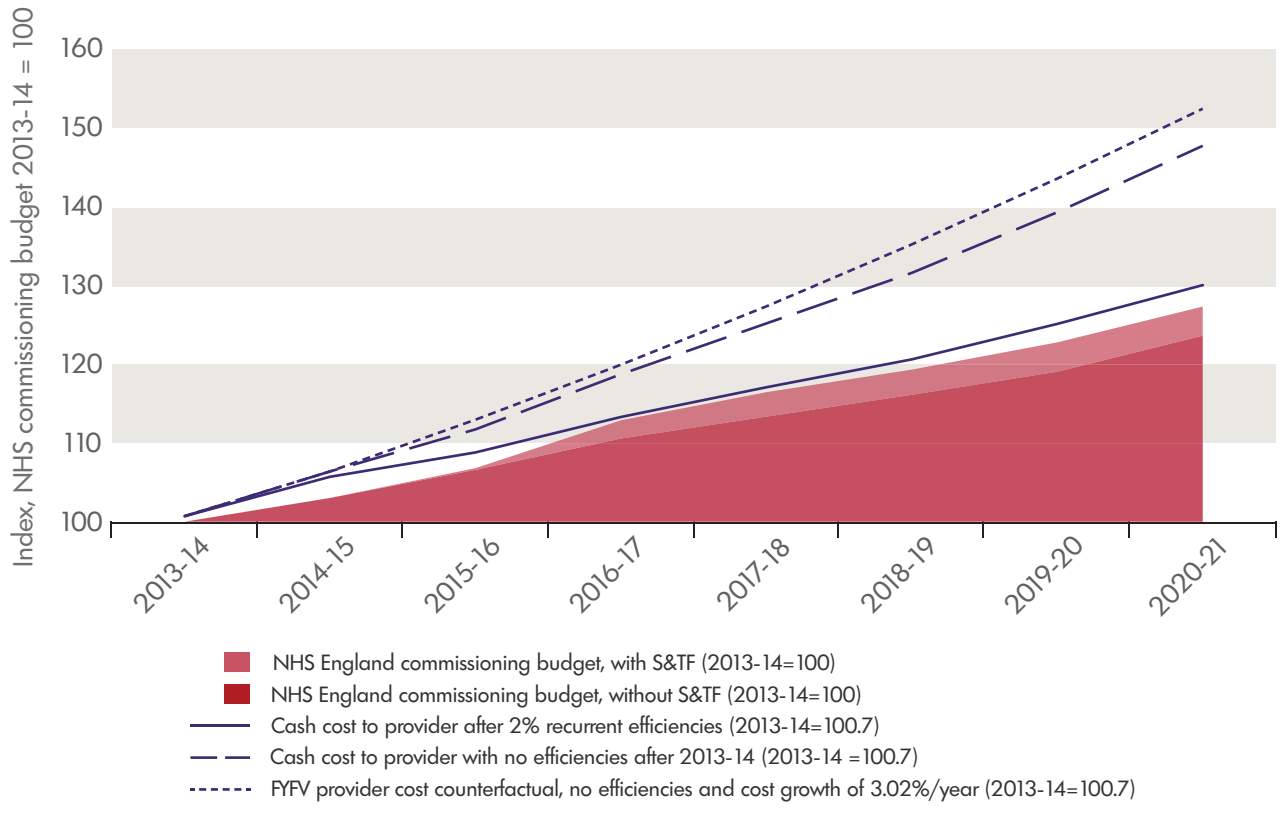
volume of care provided continues to grow at its current rate. These costs can then be compared to the NHS's commissioning budget up to 2020–21, to give a measure of their affordability, or otherwise.

NHS England's central assumption for activity growth across the NHS as a whole is a continuation of the recent trend of 3.1% a year until 2020–21; that is, NHS hospitals and other services are assumed to provide 3.1% more care each year by treating more patients with more complications through more intense and more advanced medicine.

We can calculate how much that growing quantity of care will cost to provide, and how much commissioners will be asked to pay for it, by using the guidance set out by NHS Improvement on NHS tariff prices to 2020–21. This shows how much inflation providers are expected to face, how much of that they will be expected to offset through efficiencies (as we have seen, 2% from now on) and how much prices will increase for commissioners. Although the tariff only applies directly to secondary care (over two thirds of all NHS care), it is reasonable to assume – at least for this headline analysis – that similar cost pressures (staff, drugs, equipment) will apply across all NHS services.[6]

Figure 3 shows these rising costs and sets them against the NHS commissioning budget to 2020–21, announced as part of the November 2015 Spending Review. Costs in 2013–14 are set at 100 in the index, and a 10% increase in cost is shown as 110. (A little further below we will look at these costs in cash terms, when we zoom in on the secondary care budget, where we are able to be more precise about the relationship between the volume and cost of health care).

Figure 3: Index of 3.1% annual activity growth applied to NHS-wide costs



6 While the tariff only applies directly to around 40% of secondary care, it is used as the main reference point for determining changes in costs to commissioners for activities outside the tariff.

Box 1: Modelling the cost of future activity

To model the cost of future activity we have first established the baseline cash cost of secondary care for each year from 2009–10 to 2014–15. This was derived from the NHS Reference Cost dataset, which details the cash cost to NHS hospitals and other secondary care services of providing £75 billion worth of the total £100 billion annually spent on NHS care. Provider and commissioner cash costs for projected activity beyond 2014–15 are calculated by first applying the activity growth rate from the 2014–15 cash cost to provider and then multiplying the value of activity by either the expected net change in provider costs for the year (in other words, the product of inflation as defined by the tariff and the 2% efficiency assumption) or the planned net change in tariff cash prices.

For all years except 2019–20 we have taken our measure of provider inflation from the figures published as part of the NHS tariff (or NHS Improvement's guidance on the future tariff). 2019–20 is the exception because the published figure for provider inflation that year (2%) does not include the impact of the change in the public sector pension discount rate announced as part of the 2016 Budget. We calculate that will add approximately 0.7% to secondary care provider costs in 2019–20 and so have adjusted the inflation figure for that year from 2% to 2.7% and assumed the cost will be passed onto commissioners in the form of a 0.7% increase to tariff rates. We have assessed the sensitivity of our analysis to this adjustment which is to increase the size of the reported overall deficit in the secondary care sector by 2020–21 by approximately £600 million. As the size of the projected deficit discussed in almost all our scenarios is far in excess of this sum, the thrust of this analysis is unaffected.

Finally, as the reference cost dataset covers service cost (£62 billion in 2014–15) rather than full operational cost (£75 billion), values for each year are multiplied up to reflect provider full operational costs, net of education and training costs (approximately 3.6%) as those costs are typically funded by Health Education England, rather than commissioners. These adjustments are made using the reconciliation data published by the Department of Health for the years from 2012–13 onwards.

Read more about our method for doing this in Box 1.

The shape of Figure 3 on the previous page is familiar: the purple dotted line at the top represents NHS England's 'do nothing' scenario, where unit costs simply rise in line with inflation (around 3% a year from 2015–16 onwards) because providers make no efficiencies to mitigate it. The red shaded area shows the NHS's total available resources – made up of its core budget (dark red) and then topped up by the £2 billion to £3.4 billion a year Sustainability and Transformation Fund (S&TF, in light red), which we will discuss more later. The gap of almost one fifth between the two by 2020–21 is close to £22 billion – the original 'NHS funding gap' highlighted in NHS England's Five Year Forward View (FYFV).

The bad news, though, is that it isn't £22 billion, it is £23 billion: sharp eyes may notice the provider cost lines do not start at 100, but at just under 101. That is to account for the £600 million underlying provider deficit in 2013–14, which entails that the baseline cost of NHS services in 2013–14 was £600 million (or 0.7%) more than implied by NHS England's original calculations.

Time, then, for some good news: since the calculation of that £22 billion gap three years ago, NHS England has revised down its estimate of average annual NHS provider inflation from 2015–16 onwards, from 3% to 2.5% – largely to reflect the effect of continued public sector pay restraint announced in 2015. Our purple dashed line factors that revision into the 'do nothing' cost forecast. It reduces the gap to £19 billion – the £4 billion difference making up part of the 'central savings' claimed by NHS England towards the total £22 billion/£23 billion target.[7]

7 NHS England's calculations suggest a saving from the pay cap of nearer £5 billion. However, this was calculated prior to the announcement at Budget 2016 of a change in the public sector pension discount rate, which we estimate will increase the pension cost of NHS employers by around 1% from 2019–20 onwards, reducing the central savings on the NHS pay budget from £5 billion to £4 billion.

There is further good news: things may be bad for NHS providers, but we do not (yet) live in a world where hospitals and other services do not make any year-on-year efficiency savings. The solid purple line in the middle shows provider costs after recurrent provider efficiencies of 2% a year – around the level of year-on-year efficiencies achieved in recent years, and the level of annual cost cuts providers will now need to keep making if they are to keep up with the 2% real-terms cut to tariff prices each year.

It looks like those 2% annual provider efficiencies save a lot – and they do: around £16 billion a year by 2020–21 compared to the 'do nothing' scenario.

But here is our problem: even after those 2% cost savings, year on year, the gap between the provider costs (solid purple line) and the total NHS budget (red area) gets progressively wider from 2018–19 onwards. By 2020–21 it reaches 2.1% of total available resources, indicating a £2.5 billion net deficit in the NHS budget.

And then there is the question of the S&TF. That £2 billion to £3.4 billion a year fund was announced as part of the 2015 Spending Review and is supposed to fund investment in much-needed service transformation; to secure the long-term sustainability of NHS services and to deliver the new models of care set out in the FYFV. But here we see it swallowed up just funding the annual increases in day-to-day activity; nothing is spare for investment. The dark red area at the bottom of the graph shows what the core budget would be without the S&TF: heading straight for a £6 billion NHS-wide deficit by 2020–21.

This is the backdrop against which NHS England, NHS Improvement and most particularly the Treasury are demanding two further changes from local NHS organisations. From providers, they want to see additional efficiencies, above and beyond the 2% they have achieved in recent years; and from commissioners they hope to see a full percentage-point reduction in the speed at which the volume of care provided by the NHS is growing.

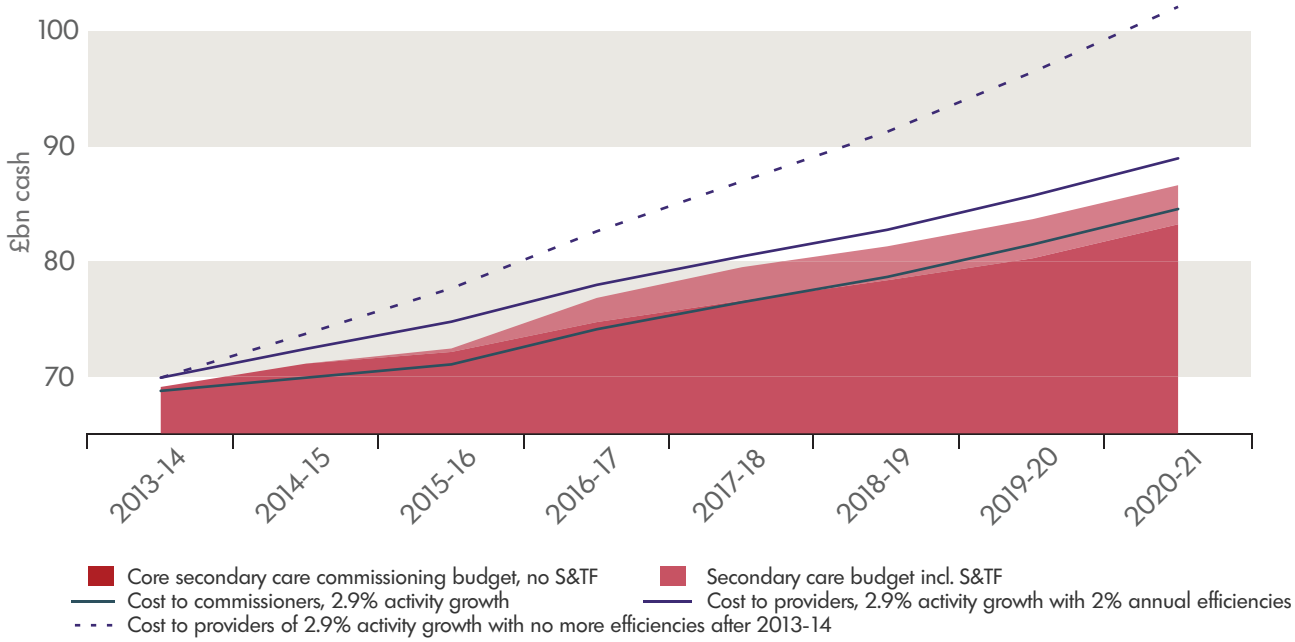
To explore the likely impact of those further efficiencies and spending reductions it is useful now to shift away from the NHS England budget as a whole to look more specifically, and in cash terms, at the secondary care sector, which makes up approximately 72% of the NHS commissioning budget, and from which the vast bulk of NHS England's estimated £15 billion of 'locally delivered' savings will come.

Zoom in on secondary care: sharing the pain of five more years of deficits?

Figure 4 presents the secondary care-only version of the £23 billion funding gap. The core secondary care budget (red area in the chart) is set at 72% of NHS England's resource budget – a figure taken from the modelling behind the NHS England's Five Year Forward View – while the light red area supplements that by adding the entire S&TF.

Read more about our workings on this in Box 2.

Figure 4: Funding gap in secondary care: 'base reform' scenario



Box 2: Modelling the cost of secondary care activity

To model the affordability or otherwise of rising secondary care activity we have first defined 'secondary care' as all hospital (general and specialist), community (including continuing health care) and mental health activity. NHS England's recent briefing paper on the modelling behind the Five Year Forward View showed how core spending in this area accounted for £69 billion (or 71%) of NHS England's total spending in 2014–15. A further £2 billion worth of NHS's England spending that year is categorised in the paper as 'commitment pressure' and as military and justice system spending.

Assuming that 71% of that spending was made in the secondary care sector brings the combined total secondary care spend to £70.2 billion, or 72% of NHS England's resource allocation. We have used that 72% figure to extrapolate an NHS secondary care commissioning budget from 2013–14 to 2020–21. To calculate the figure for the years from 2016–17 onwards (where the full NHS England budget is supplemented by the ring fenced S&TF) we have calculated first a 'core' secondary care budget as 72% of NHS England's total planned spending for each year, minus planned S&TF spending, as set out in NHS England's December 2015 board paper detailing financial allocations. Our S&TF enhanced figure for secondary care then includes the entire S&TF for that year, rather than a 72% proportion, as we believe it most likely that the vast bulk of the S&TF will be spent in secondary care. No adjustment is made to reflect NHS transfers to local authority social care spending under the Better Care Fund, or to reflect the 1% non-recurrent spending and 0.5% contingency requirements set by NHS England.

A slightly lower rate of activity increase is shown here (2.9% a year rather than 3.1%) as this echoes -NHS England's assumption for activity growth in the sector (the overall growth rate is pushed up to 3.1% by an expected increase in community prescribing activity of almost 5% a year).

Our headline 'gap' by 2020–21 here is the £15.5 billion gap between the secondary care budget (including the full S&TF) and the cost of providing activity with no additional efficiencies after 2013–14 (shown again by the dotted purple line).

The good news: provider efficiencies in 2014–15 and 2015–16 have already reduced that gap by around £3.4 billion, and if providers continue to make recurrent efficiencies of 2% a year they will close the gap by a further £9.8 billion by 2020–21. The resulting cost to providers after those efficiencies is shown by our solid purple line. That would leave a secondary care-wide deficit of around £2.3 billion in 2020–21– but only after it had absorbed the entire S&TF on funding the cost of the growing volume of day-to-day care.

Without the S&TF, the secondary care deficit after recurrent provider efficiencies of 2% a year would be £5.7 billion by 2020–21. This £5.7 billion is the intransigent end of the £15.5 billion secondary care funding gap that is now causing sleepless nights amongst commissioners and providers, and which is our focus for the rest of this analysis.

Sharing the £5.7 billion pain: commissioners hit bust by 2018–19

We can get a sense of how commissioners and providers are sharing the pain of that £5.7 billion gap by looking now at how much the growing quantity of secondary care would cost commissioners. This is shown by our teal line, which prices up the expected volume of care by the NHS tariff to show the price that would need to be paid to providers by commissioners.

At no point would those commissioner payments meet the actual cost of providing care – that much we know: even if providers make 2% cost savings a year, the amount they will get paid under the tariff will be cut by an equal amount, and so the level of their deficit will remain similar in proportion to what it is today – reaching around £4.4 billion in cash terms by 2020–21.

That is bad enough, but the more dramatic swing into the red will be for commissioners. The gradual cash increases in tariff prices, combined with the growing volume of care purchased, means that from 2018–19 onwards, the core secondary care budget will be insufficient to cover the tariff rate for the quantity of care provided.

> By 2019–20, over a third of the S&TF would be needed to subsidise commissioner costs under the tariff, while the remaining £2 billion or so would be absorbed by provider deficits. No S&TF would be available for investment in 'service transformation'.

That would mean that by 2019–20, over a third of the S&TF would be needed to subsidise commissioner costs under the tariff, while the remaining £2 billion or so

would be absorbed by provider deficits. No S&TF would be available for investment in 'service transformation'. The underlying £5.7 billion deficit against the 2020–21 core commissioning budget would be formed of a £4.4 billion deficit in provider accounts and a £1.3 billion overspend by commissioners.

Increase provider efficiency to free up investment for transformation?

We saw earlier that the £3.7 billion underlying provider deficit in 2015–16 was testament to the redundancy of relying almost exclusively on high levels of provider cost saving to curtail NHS spending. Indeed, Lord Carter's report on hospital efficiency published earlier this year looked extensively at the opportunities for further cost savings in acute hospitals to 2020 and found a rate of just 2% a year could be expected: half the rate required of (and missed by) hospitals in recent years.

Nevertheless, secondary care providers are now being asked to exceed that 2% again. The rationale is simply that without further efficiencies from providers, the entire S&TF will dwindle to naught and the NHS as a whole would be bust.

Secondary care providers have been asked to end 2016–17 with an underlying deficit no greater than £2.35 billion – down from £3.7 billion in 2015–16.[8]

Meeting that target will require providers to make an extra 2% recurrent efficiency saving, over and above the 2% already planned (and required by the tariff). Four per cent efficiency in 2016–17 is equivalent to around £3 billion in cost savings. Doing that all over again in 2017–18 with another 4% cost cuts, and following that in 2018–19 with only a slightly lower cost cut of 3% would finally bring provider costs down into line with their income under the tariff.

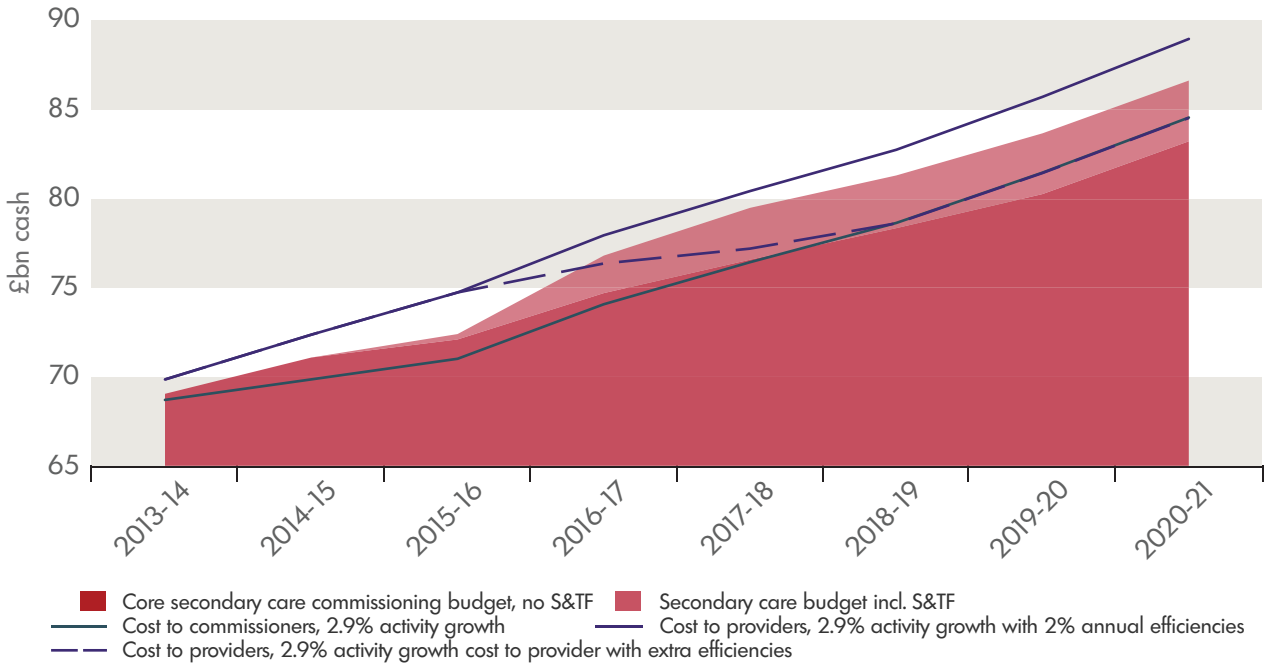
Figure 5 on the next page shows the impact these extra efficiencies would have on provider costs, with recurrent efficiency savings returning to 2% from 2019–20 onwards. As this figure shows, two years of recurrent efficiency savings at 4% (that is, twice the rate of recent trends) followed by a year at 3% would bring provider costs into line with the payments due to them under the tariff by the end of 2018–19. Over the three years, that would require providers to remove an extra £4 billion from their cost base, permanently, and in addition to the £6 billion already required by the 2% annual cut in the tariff.

It is a big ask, but it is only after that level of cost reduction that the need to spend the S&TF on filling the provider deficit would be removed, freeing up part of it for investment in transformation.

What that would not do, however, is reduce tariff costs for commissioners (teal line). That is because extra provider efficiency does nothing – at least in the short term – to reduce tariff bills for commissioners, which are determined by the volume of care purchased. By 2018–19 commissioners would still hit the point where they would be unable to pay the invoices from providers out of their core budgets. In 2019–20 a large £1.2 billion annual overspend would emerge in commissioner accounts, requiring them to use over a third of the S&TF just to fund recurrent activity growth under

8 If delivered, the £2.35 billion underlying deficit in 2016–17 would be reported in provider accounts as a deficit of £550 million, with the remaining £1.8 billion in deficit offset by extra non-recurrent income from the Sustainability and Transformation Fund.

Figure 5: Funding gap in secondary care: 'extra provider efficiency for 3 years' scenario



the tariff – that is, activity increases that get consolidated, or 'baked' into the activity baseline for each following year.

Besides a begging bowl outside the Treasury, this would leave commissioners with nowhere else to turn but to reduce the rate at which activity levels are growing.

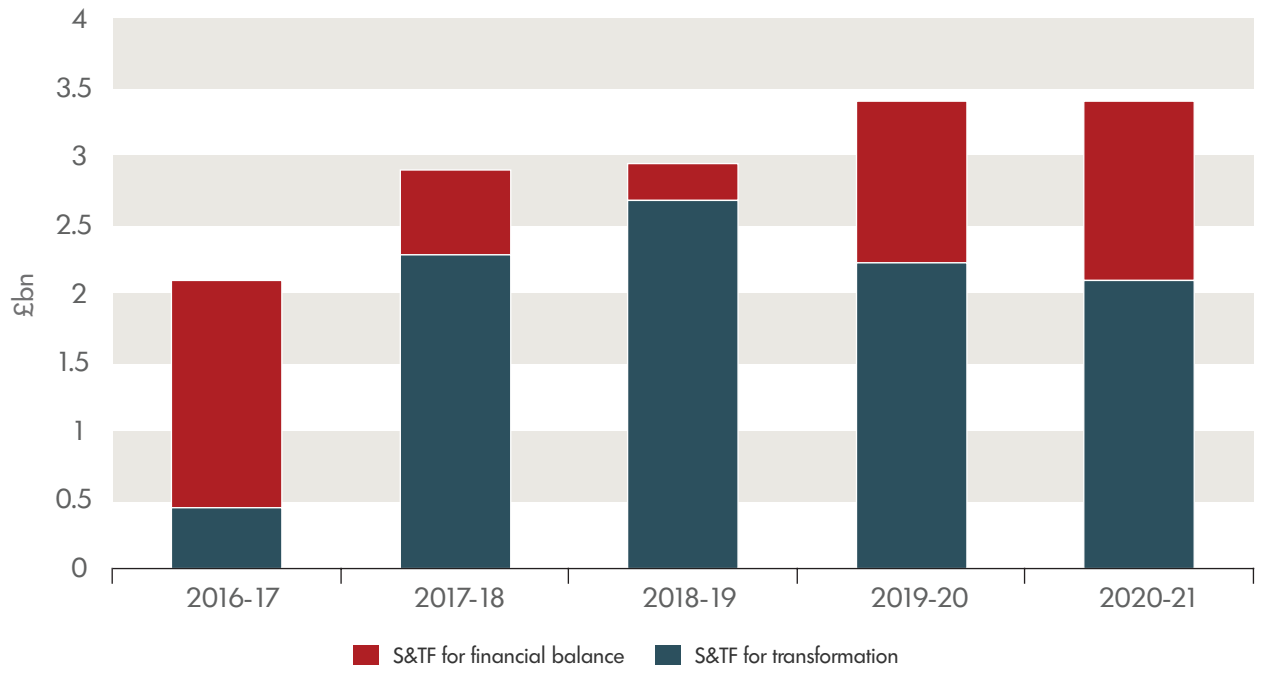
We are in the realm of theory rather than evidence-based practice here, but it is worth considering what the impact on secondary care costs would be if the modest amounts of S&TF freed up through extra provider efficiencies were successfully invested in measures to reduce the speed at which activity levels were growing.

Figure 6 on the next page shows how much S&TF would be available for genuine investment in service transformation in each of the five years to 2020–21 *if*, and only *if*, providers managed to find 4% recurrent efficiencies in 2016–17 and 2017–18 and then follow that with a year of 3% efficiencies before returning, in 2019–20, to the 2% level required to keep up with the year-on-year real-terms cuts to the tariff.

While the vast bulk of the S&TF would be eaten up by provider deficits in 2016–17, around £2.3 billion would, in theory, be available for investment in transformation in 2017–18. Somewhat paradoxically, 2018–19 (the year the NHS as a whole is set to receive the lowest cash increase in its funding) would emerge as the year with the most significant investment fund for transformation, at around £2.7 billion of the total £3 billion fund for that year. In the years after 2018–19, the continued rise in secondary care activity would again require increasing chunks of the S&TF to be used to fund the cost of that activity growth, as commissioner tariff bills exceeded their core budgets.

There is very little systematic evidence on the scale and cost-effectiveness of investment in demand management, but it is NHS England's hope that investment in early diagnosis, better primary care and improved patient self-management, for example, will serve to improve population health and reduce demand for secondary care and

Figure 6: S&TF available for transformation: 'extra provider efficiency for 3 years' scenario



therefore slow the pace of activity growth by around one percentage point: from 2.9% a year to 1.9%.

As we have seen, substantial funds will not be available for investment in that transformation until 2017–18. It therefore does not seem plausible to expect any real change in the activity growth rate until after then. Figure 7 on the next page models the impact on secondary care costs of a gradual, staged, one percentage point reduction in the activity growth rate over the three years from 2018–19 to 2020–21.

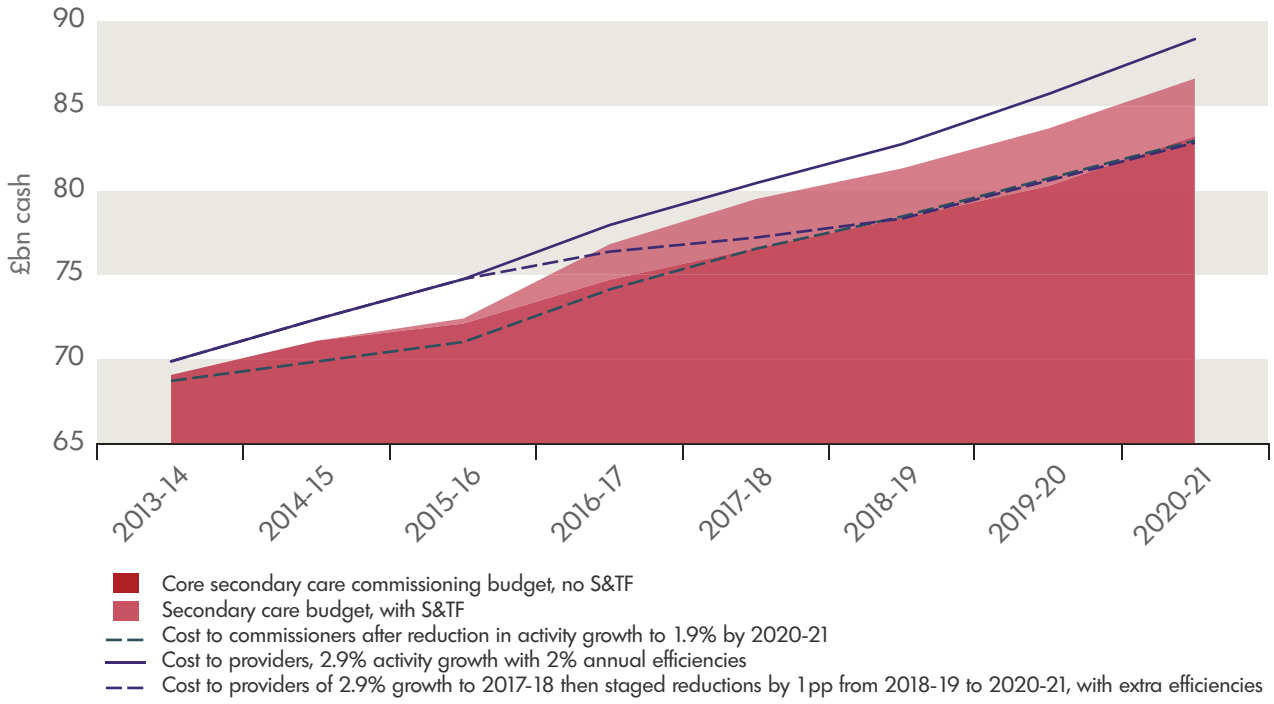
It is a big ask, but Figure 7 perhaps shows why NHS England is banking on the twin hopes of substantial extra provider efficiencies in the first instance, followed by commissioner investment to achieve a percentage point reduction in the rate at which activity is growing.

As we saw in Figures 5 and 6, extra provider efficiencies of up to 4% a year bring provider costs into line with the income they are able to command from commissioners under the tariff. Reducing the growth rate in overall activity from 2.9% to 1.9% then moves that provider cost – and the invoices providers send to commissioners – into line with what commissioners can afford, from their core budgets, meaning that by 2020–21, the lines in our graph converge and the system would finally be in balance.

Success in this strategy would imply an annual saving by 2020–21 through stemming activity growth of £1.7 billion a year and rising. That would imply a rather heroic and rapid return on any transformational investment, for which there is little or no evidence base.

And there are further reasons to be sceptical as to whether or not this can be achieved. What if, for example, those extra efficiencies were not forthcoming in the three years from 2016–17? After all, the current size of the provider deficit is testament to the inability of providers to make similarly scaled cost savings over the last three years. It is also worth noting that the current level of recurrent efficiencies (2% a year) are being

Figure 7: Funding gap in secondary care: 'one percentage point reduction' scenario



made amid annual activity growth of 2.9%, enabling providers to reduce their average unit costs in areas where extra activity can be generated at only marginal extra cost.[9] Slowing the growth rate to 1.9% might reduce the scope for further provider savings through that means.

And even if those efficiencies were forthcoming, there is still no guarantee that commissioner investments in new models of care designed to reduce the activity growth rate will succeed. Indeed, while planned initiatives such as giving dementia patients their own named clinician and bespoke care pathway, or linking care homes to hospital clinicians, may well improve the quality of care for patients, they might also serve to simply add additional activity rather than supplant and diminish growth in existing levels.

Rationing without investment

It is not a palatable scenario, but there is a risk that without either or both additional provider efficiencies and subsequent successful investment in genuinely reducing the growth in demand, commissioners will resort instead to curtailing growth through crude service rationing. They might do that directly – by raising the threshold for treatment and access to certain services – or they might attempt to do it indirectly, by shutting services and hoping that patient and clinicians' expectations for care and treatment will wane, rather than simply divert to alternative services elsewhere.

9 A simple example of this would be where a provider has already covered the fixed costs of providing a service, and so caring for additional patients only entails the marginal cost of each extra patient – i.e. the medicines and devices consumed in their care and the additional costs of providing clinical staff time.

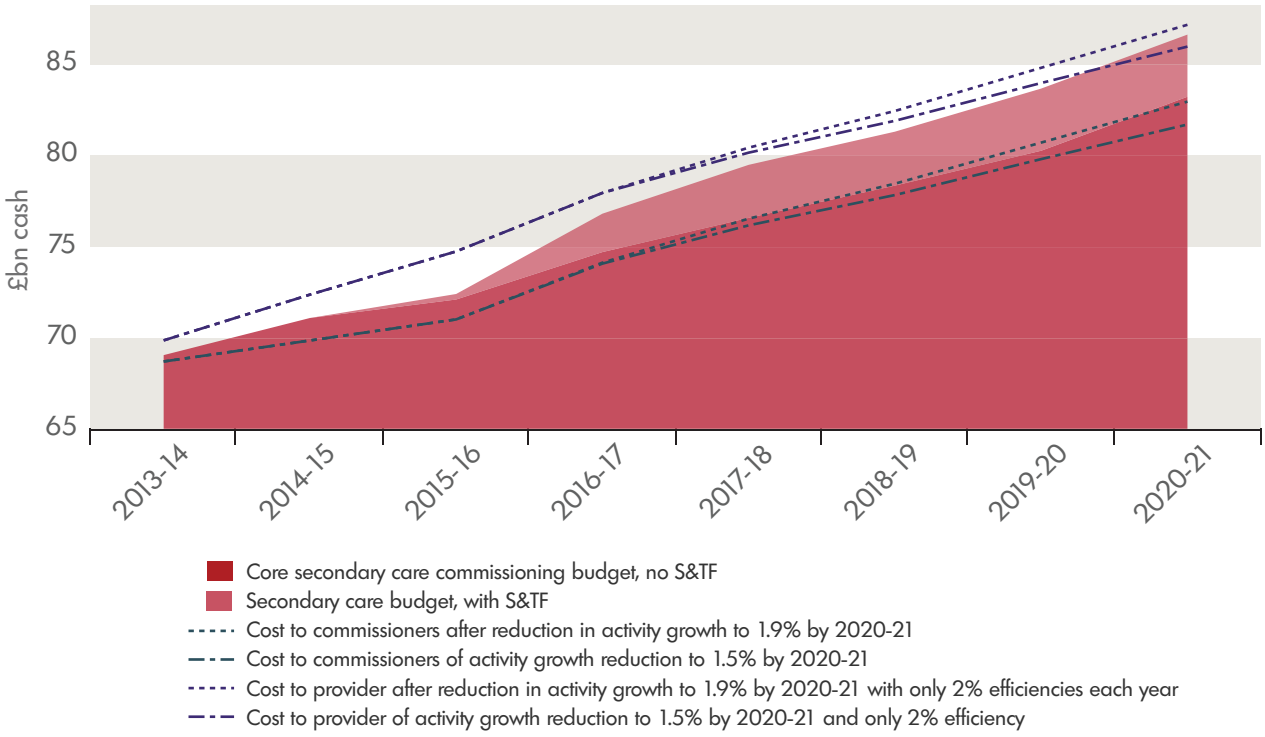
Aside from the ethical, political and long-term health implications of such a move, how viable would crude rationing be as a cost-saving measure in the short term?

According to NHS England, the current secondary care activity growth rate of just under 3% a year is comprised of 1.5% growth due to population growth and ageing and the remaining 1.5% growth due to so-called 'non-demographic growth', which is the pressure to increase activity levels due to factors such as rising public expectations (for example, that X condition will be treated, rather than endured), improvements in quality and clinical advancements.[10]

Reducing the rate of activity growth to below the 1.5% driven by population growth and change would involve drastic measures: it would effectively mean denying today's 74-year-olds the same hip replacement their 75-year-old neighbours received last year. But commissioners might, in the absence of sufficient investment funds, be tempted to make inroads into the 1.5 percentage points of growth above that level, which might be seen as more discretionary; which we could loosely term as being driven by improvements in quality and access.

Figure 8 shows the impact of the same staged reduction in the activity growth rate to 1.9% explored above, but with provider efficiencies at 'just' 2% a year. But as the graph shows, that won't work: attempting to balance the NHS budget through a one percentage point reduction in activity alone is not a viable option: provider costs by 2020–21 would still be £4 billion above the core secondary care commissioning budget.

Figure 8: Funding gap in secondary care: 'demographic growth-only' scenario



10 NHS England's breakdown of the components of activity growth are set out in chapter 2 of its May 2016 technical note on the Five Year Forward View. See also John Appleby's 2013 report on long-term trends in health spending for The King's Fund, which identified several problems in isolating individual drivers of activity growth.

Even after spending all of the S&TF on plugging provider deficits, the system as a whole would be spending £0.5 billion a year more than its total available resource. To bring the NHS into financial balance without provider efficiencies above 2% a year would require crude rationing to go even further and to reduce the rate of growth right down to the 1.5% needed to keep up with only that element of demand driven by demographic change.

Doing that, as the dot-dashed lines in Figure 8 on the previous page show, would bring provider costs down to below the level of the total budget. Those costs would still be significantly higher than their tariff income, and the core commissioning budget, but in such a scenario the NHS would have dispensed with the ambition to invest in genuinely reducing the growth in demand through improved services and population health, and so could instead simply spend the entire S&TF on plugging the gap in provider finances (or alternatively use it to increase tariff rates to meet provider costs). But this precarious balance would be achieved at the cost of preserving the NHS in aspic, standing still and deliberately halting any further advancement in health care quality, such as the adoption of new treatments available elsewhere in the developed world.

The political acceptability of that – following a Brexit campaign which highlighted a potential £350 million extra for the NHS a week – is highly questionable.

About the author

Sally Gainsbury joined the Nuffield Trust in October 2015 as Senior Policy Analyst. Her focus is on health and social care funding and the NHS financing system. She also contributes to the Trust's rapid response and analysis of emerging policy issues.

Prior to joining the Trust Sally was an investigative journalist at the *Financial Times*, working on UK and international investigations spanning public spending, tax avoidance and money laundering. Before joining the *FT* Sally was chief reporter and news editor at *Health Service Journal*.

Sally has a PhD in history and a Master's degree in politics.

evidence for
better health care

nuffieldtrust

For more information about the Nuffield Trust, including details of our latest research and analysis, please visit www.nuffieldtrust.org.uk

Download this report from
www.nuffieldtrust.org.uk/publications

Subscribe to our newsletter:
www.nuffieldtrust.org.uk/newsletter

Follow us on Twitter: Twitter.com/NuffieldTrust

Nuffield Trust is an independent health charity. We aim to improve the quality of health care in the UK by providing evidence-based research and policy analysis and informing and generating debate.

59 New Cavendish Street
London W1G 7LP
Telephone: 020 7631 8450
Email: info@nuffieldtrust.org.uk

www.nuffieldtrust.org.uk

Published by the Nuffield Trust.