Introduction

The NHS has moved from an overall net deficit to a net surplus within a year, according to the figures released by the government in June 2007 (Department of Health 2007a), reversing a three-year trend towards increasingly large gross deficits.

According to the unaudited figures released for the financial year 2006/7, the gross surplus was £1.42 billion and the gross deficit £911 million, which leaves an overall net surplus of £510 million. This represents the first net surplus since 2003/4, as the chart below shows (figure 1).
The government argues that these latest figures show that the NHS as a whole has turned a corner in its ability to manage its finances well, even though gross deficit (which is the sum of all the deficits posted by organisations who failed to break even or post a surplus) is still quite large (£911 million) and a small number of NHS organisations remain in serious financial difficulties. This briefing analyses the latest figures and data extracted from NHS trust boards, to assess the scale of the challenges to the financial security of the NHS in the future.

**Scale and distribution of surpluses and deficits**

Compared to the audited accounts from the previous financial year (2005/6), fewer organisations have forecast deficits for 2006/7 and more are in surplus (see figures 2 and 3).

The deficit has become slightly more concentrated in fewer organisations: 80 per cent of the gross deficit is found in 10 per cent of NHS organisations for 2006/7 compared to 70 per cent of the gross deficit in 11 per cent of organisations the previous financial year.

Another change from the previous year is the shift in location of the gross deficit towards primary care trusts (PCTs) and away from NHS trusts (hospitals, mental health trusts and care trusts): nearly 70 per cent of the gross deficit now rests with PCTs compared with just under half (47 per cent) in 2005/6.
size of the gross deficit among PCTs as a whole was slightly higher in 2006/7 (£633 million) than in 2005/6 (£616 million).

NHS trusts delivered a gross deficit of £278 million in 2006/7, an improvement from the £696 million gross deficit reported in 2005/6.

The disparity in performance between PCTs and NHS trusts is partly due to a change to the financial regime known as Resource Accounting and Budgeting or RAB (some of the negative effects of which have been reversed for trusts but not PCTs) and the effect of ‘top slicing’ funds to help fight the deficits, which applied to only PCTs, not trusts (both are explored in greater detail below).

As in 2005/6, the accounts of the strategic health authorities (SHAs) show a large net surplus, amounting to £961 million, a key factor in the reversal of the previous year’s poor financial performance of the NHS as a whole (figure 4).

![Figure 4: Financial performance by type of NHS organisation, unaudited figures 2006/7](chart)

These figures exclude the financial performance of foundation trusts. Figures for foundation trusts for 2006/7 have been published by the regulating body, Monitor. They show that, taken as a group, foundation trusts have delivered a net surplus of £130 million, with only three organisations in deficit. Foundation trusts also have £995 million in cash reserves and are showing ‘significant caution’ in reinvesting this money because of ‘uncertainties’ about funding and financial risks in the future (Monitor 2007).

Accounts for the NHS in Scotland, Wales and Northern Ireland for 2006/7 have yet to be published. NHS Scotland successfully moved from a net deficit in 2003/4 to a small net surplus in 2005/6 (£70.6 million of a £9 billion budget) with only 2 of the 24 NHS bodies failing to meet their financial targets (Audit Scotland 2006). By contrast, the NHS in Wales was forecasting ‘a significant overall deficit position’ for the final accounts of 2005/6 (Wales Audit Office 2006). The NHS in Northern Ireland expected to break even in 2005/6 (Department of Health, Social Services and Public Safety 2006).
How has the NHS managed to move from overall deficit to surplus?

One important element of this year's net surplus in England was the creation of a 'contingency' of £450 million, generated from budgets held by the strategic health authorities (SHAs). This contingency was created from savings to a set of budgets worth £5.5 billion that are designed, according to the Department of Health, to be spent on items that include 'specific public health programmes, medical education, non-medical, clinical training, GP performance reimbursement, clinical excellence awards, services such as walk-in centres, out of hours, NHS direct' (Department of Health 2006c). The Department of Health gives no details about how the £450 million was saved so it is not known whether (and on what scale) specific training or public health programmes were scaled back during the financial year.

It is not entirely clear from the Department of Health's documents where the bulk of this £450 million is now located in the NHS accounts at an organisational level. The bulletin on NHS finances for the first quarter of 2006/7 stated that 'in line with the policy of not bailing out over-spending organisations, the contingency fund will be held centrally by the NHS Bank on behalf of SHAs and not passed out to individual trusts and PCTs' (Department of Health 2006c). However, Patricia Hewitt, then Secretary of State for Health, told parliament at the end of March 2007 that this £450 million was to be returned to SHAs (Hansard 2007) and the most recent (and final) bulletin for the year states that the final figures for the year are 'after distribution of the £450 million' (Department of Health 2007a).

Some of the contingency fund (£178 million) has been used to offset the effects of the Resource and Accounting Budget (RAB) regime. According to the Audit Commission, RAB had been inconsistently applied by SHAs to some NHS trusts. In some cases, trusts that had run up a deficit in one financial year were faced with a reduction in income (by the same amount) the next and were also expected to generate a surplus to cover the deficit, which effectively penalised them twice over for their deficits (the so-called double-deficit effect (Audit Commission 2006)). In March 2007, the Department of Health announced that £178 million would be ‘returned' to 28 hospital trusts, reversing their income deductions for 2006/7 and so reducing or eliminating their deficits (Department of Health 2007b).

Another tactic used for 2006/7 was a ‘top slice' of PCT resources, whereby amounts totalling £1.14 billion were removed from PCT budgets and lodged with SHAs. This was designed to make PCTs operate within more limited budgets from the beginning of the year. The Department of Health states that some of this total (£319 million) has already been returned to PCTs, but the remainder is still with SHAs and will be returned to PCTs in the future (Department of Health 2007a). According to the Department of Health, the ‘top slice' was a ‘temporary measure to stabilise the NHS'. It is not clear what this money would have been used for if it had not been set aside as a reserve.

In addition to the top slice for PCTs, all trusts were expected to make savings and to improve their efficiency and productivity in order to generate a surplus (Department of Health 2006d). Some trusts and PCTs were given extra help with this task through 'turnaround teams', which in some cases were composed of outside consultants. There is no national level data available on what NHS organisations have done to reduce or eliminate their deficits or what the effects of these actions might have been on patient care. However, the Department of Health identified lack of control over recruitment, pay, agency staff and high-cost procedures as a problem for trusts and recommended 'improved budgetary control, capacity reductions, recruitment bans for back office functions, tighter authorisation controls for agency spend, improved control over high cost supplies and deferral of discretionary spend'. For PCTs, better control of hospital demand was also a recommended strategy (Department of Health 2006b).

The government has argued that patient care has not been adversely affected and that there has been no slippage of high-level access targets, such as inpatient or outpatient waiting times (Department of Health 2007a).
There is no doubt that some costs have been reduced through job losses. According to the government, 2,330 compulsory redundancies have been reported during 2006/7, of which 1,912 (82 per cent) relate to non-clinical staff (Department of Health 2007a). This is a tiny proportion of the 1.3 million staff working in the service. The latest workforce survey for the NHS shows a slight (1.3 per cent) reduction in overall staff numbers in 2006 compared to 2005, the first fall since 1997 (Information Centre for Health and Social Care 2007). The reductions occurred principally among non-clinical administrative, managerial and infrastructure support staff, although ambulance staff numbers also fell by 10 per cent between 2005 and 2006 (from 17,417 to 15,723 measured as full-time equivalents) (Information Centre for Health and Social Care 2007).

Some professional bodies have claimed that the government’s figures do not reflect the scale of the impact that financial strains have had on jobs. The Royal College of Nursing (RCN) has been conducting the most comprehensive survey of the situation in individual trusts on an ongoing basis since June 2005. The RCN collates reports from their members of proposed compulsory and voluntary redundancies and of the deletion of vacant posts; their London office then verifies these through trusts’ board papers and press releases. They estimate that by April 2007 22,363 posts had been lost as a result of deficits (Royal College of Nursing 2007). However, the fact that there is no official, national data on voluntary redundancies or the deletion of vacant posts means that it is not possible to verify this figure.

There are also concerns that some services to patients might have been scaled back in clinical areas that are not the subject of national targets and are therefore less visible in the national data. In 2006/7 mental health trusts delivered an overall surplus (unlike acute hospital trusts, which account for the bulk of the non-PCT deficit) (Department of Health 2007a). It is not clear what this money might otherwise have been spent on.

**Is the NHS in better underlying financial health?**

Although the NHS has achieved a surplus this year, there remains an important question: do these latest figures represent another short-term ‘fix’ (through the use of temporary measures such as the top slice or contingency reserve) or do they suggest that the NHS has finally reversed a tendency to overspend in some areas? In reality, the future financial health of the NHS depends on many factors, not all of them within the control of individual NHS organisations, for instance, growth in wage costs or the cost of pharmaceutical products, or the overall impact of system reforms such as Payment by Results (PbR) (a change to the way hospitals and other providers of health care are paid for the services they deliver).

Nevertheless, improving the quality of financial management within individual organisations has been the subject of intense reform effort since the deficits began to emerge. In December 2005, as the scale of the 2005/6 deficit was becoming apparent, the government made financial management a priority and made it clear to NHS organisations that they must not only learn to balance their books as a matter of routine, but must also aim to generate surpluses in the future (Department of Health 2006d). The rationale was two-fold: first, the NHS needed to prepare for the slowdown in the growth of funding expected after 2008 and, second, better financial management was needed in order to allow the NHS system reforms to function properly. These reforms include PbR and increased patient choice, more diversity of provision and institutional autonomy in the shape of foundation trusts, and better commissioning at PCT and GP level. They all hinge on the ability of NHS organisations to function more like businesses, avoiding losses and generating surpluses for re-investment.

Better financial management was to be achieved by a number of mechanisms, including generating much better (and more frequently collected) data about expenditure and improving the financial expertise at trust board level. One specific challenge for the NHS was to rein in spending on acute hospital services and this task had become all the more urgent, as the new hospital payment system,
PbR, contains strong financial incentives for hospital trusts to perform more operations and deliver more treatments.

Since December 2005 the Department of Health has issued a series of policy documents aimed to equip PCTs (who handle 80 per cent of the NHS budget) with the means to reduce the demand for hospital services, without compromising waiting times or other quality targets. The tactics include rolling out practice-based commissioning (to enable GP practices to design and deliver non-hospital treatment alternatives), reducing the rate of GP referrals into the hospital system, reducing the number of ‘consultant to consultant’ referrals (that cannot be controlled by the PCT or GPs), reviewing what range of operations should be available locally and giving PCTs new powers to penalise trusts who over- or underperform (Department of Health 2006a, d).

The strength of this year’s financial results has led the government to argue that their strategy has worked and that the overall financial health of the NHS is better, with more organisations balancing their books within the year and delivering surpluses or breakeven at the end of the year. To shed some light on the nature of this effort to contain hospital demand during 2006/7, we analysed the most recent board papers from England’s 152 PCTs during March 2007 (the end of the financial year). We attempted to locate the most recent financial reports and other relevant documents for each PCT. Recent board papers were found on the websites of 124 PCTs. A number of themes emerged from the board papers.

- Many PCTs reported that acute hospital trusts were ‘overperforming’ (treating more patients than the PCT had planned for) or referred to a risk of overperformance, which might negatively affect the PCTs’ efforts to achieve financial balance (97 PCTs or 78 per cent of PCTs with board papers).
- There were disputes between PCTs and hospital trusts over some of this activity or references to ‘challenges’ over hospital trust invoices or data: this was mentioned by 47 PCTs (that is, nearly half of the PCTs referring to ‘overperformance’).
- Demand management initiatives: 72 PCTs referred to some sort of demand management initiatives in their board papers, including explicit references to:
  - encouraging GPs to reduce their referrals to outpatients (33 PCTs)
  - reducing the rate of consultant-to-consultant referrals (20 PCTs)
  - reducing the rate of outpatient follow-up appointments (21 PCTs)
  - reviewing the effectiveness of certain hospital procedures/operations (17 PCTs).

A further 14 PCTs referred to strategies to temporarily slow down the rate of operations or procedures performed by their local hospitals while staying within the overall national waiting time targets. There was evidence in some PCT board papers of attempts to forecast the impact of meeting the 18-week referral-to-treatment target [www.kingsfund.org.uk/go_rm?id=2284], which in some cases involved increases in the amount of activity performed by hospital trusts.

**Hospital ‘overperformance’ and disputes between PCTs and hospital trusts**

Sometimes overperformance was described in the board papers as having a legitimate cause (such as pressures on beds in winter or an exceptional number of intensive care cases). However, 47 PCTs referred to initiating some challenge to the figures given to them by their hospital trusts, sometimes on the grounds that the activity undertaken by trusts exceeded the agreed range set in the contracts with PCT and sometimes because the PCTs appeared to question either the accuracy or the legitimacy of the data submitted by trusts. For example, PCT papers referred to challenges about unexplained or sudden rises in the rate of day cases, short-stay patients or accident and emergency (A&E) admissions. There were also references to challenges to hospital trusts that had apparently ‘recategorised’ certain
procedures to their own financial advantage (for instance, classifying endoscopic investigations as day cases where they had previously been outpatient appointments) or ‘upcoding’ procedures (as being ‘with complications’ which carries a higher tariff) or charging PCTs at consultant rate for procedures that were being carried out by nurses.

There were also frequent references to hospital trusts sending in higher than expected bills for activities or treatments not covered by the PbR national tariff, such as high-cost drugs – one PCT referred to an ‘unprecedented increase’ in such invoices.

It is difficult to disentangle from the board papers what might be driving this, in particular whether it is the result of the intentional incentives of the PbR policy (hospital trusts now have a much better grip on their activity and cost data) or whether it is the product of the perverse incentives of PbR (gaming for financial gain). One PCT summed up its own analysis of the situation in these terms:

*A common feature of the overperformance is the better capture of activity by trusts and in particular the inclusion of nurse led clinics. In the main this falls within the scope of PbR and as such the PCT is liable for reimbursement despite intensive challenge. [But] The recorded increase in activity at most providers is not readily explicable in service or waiting list terms.... GP referral patterns do not appear to have changed significantly during the period.... As a consequence the level and profile of data analysis has significantly increased in the PCT during the second half of the year with the number of challenges to reported activity changes commensurately increasing.*

Another PCT was more blunt about the motives of hospitals, accusing local trusts of ‘artificially inflating’ the figures while another complained of a hospital trust making changes to its activity ‘months after month end, to the detriment of the PCT’.

DEMAND MANAGEMENT

Over half of all the PCTs studied referred to some sort of activity to manage demand. Some PCTs drew a distinction between short-term demand management strategies (such as reducing the rate of GP referrals) and longer-term ones, like using community matrons to keep people out of hospital or investing in new alternatives to hospital-based treatment under practice-based commissioning (PBC) schemes. PBC was referred to by several PCTs as an important tool to reduce reliance on hospital-based services in the future but, in the main, it was too early to be generating actual savings.

GP REFERRALS

Board papers from 33 PCTs referred to some efforts to reduce the number of patients being referred to hospital outpatient departments by GPs. Sometimes, this was in the context of PBC initiatives to manage conditions in a primary care setting, for example

*We are working with practices to take forward a broad range of initiatives with the general aim of finding better ways to manage a range of conditions within primary care such as orthopaedics, ophthalmology, dermatology and gynaecology. GP referrals for a list of conditions are being used to ensure that as many referrals as appropriate are managed in primary care....*

More often the board papers referred to a simpler approach involving setting notional targets for GP practices, for example:

*If we can save one admission per practice per week for the final three months we will reduce costs by over £1.25m*

*...contracts have been agreed at levels that assume GPs save one referral per week for each practice. Failure to contain demand to these baselines will see over performance on contracts*
Some PCTs reported that it was too soon to see any results from these initiatives, while others had already quantified the reduction: one PCT had reduced GP referrals by 9 per cent in 2006/7 and projects a further fall of 5 per cent next year, equivalent to 2,000 outpatient attendances. Another (inner city) PCT calculated that there is the ‘potential to reduce 35 per cent to 50 per cent of total outpatient attendances in selected specialties’ worth £4.3 million over the next two years.

It is not possible to tell from most board papers whether the ‘saved’ referrals are targeted at particular specialties (although some PCTs are reviewing the efficacy of some specific treatments, see below) or whether they were calculated to take into account differing levels of need for health care services locally. Only two PCTs explicitly raised the question of whether high referral rates might be a reflection of genuine need. However, one PCT concluded that:

_We are currently 15 per cent above national average for secondary care referrals, and whilst population profile might suggest[that PCT] would be above average need for secondary care this would suggest significant ‘headroom’ for reducing referrals._

**REDDUCING CONSULTANT-TO-CONSULTANT REFERRALS AND FOLLOW-UP RATES**

As well as encouraging GPs to reduce their initial referrals into hospitals, some PCTs were also attempting to reduce the rate of referrals within the hospital system, either from one hospital consultant to another (21 PCTs) or the number of follow-up appointments that patients sometimes receive after an operation or treatment in hospital (20 PCTs). Both of these have been underpinned by recent policy initiatives, in particular the new powers that PCTs have been given to scrutinise these referrals in the model contracts (Department of Health 2006a). It is not clear nationally what proportion of all referrals are from consultant to consultant: one PCT, however, reported that 30 per cent of all its referrals fell into this category.

**REVIEWING THE USE OF CERTAIN PROCEDURES OR TREATMENTS**

A small number of PCTs referred to plans to review the use of operations or procedures where there is limited clinical evidence of benefit. Although the drive to reduce the use of clinically ineffective procedures predates the deficits in the NHS, the need to make savings appears to have focused efforts on rethinking what sort of procedures are carried out, when and to what effect. Some PCTs referred to reducing the use of certain procedures: tonsillectomies, hysterectomies, paediatric circumcision, breast reduction, breast augmentation, IVF, orthodontics, carpal tunnel release and vasectomies were all mentioned in board papers. PCTs also referred to reviewing whether some procedures should be publicly available on the NHS at all, such as cosmetic plastic surgery. Most of the references in board papers were to future plans rather than to actual restrictions in place now, although one PCT reported that it had achieved a 50 per cent reduction in the ‘low priority’ treatments of hernias and varicose veins (against a target of 70 per cent) for 2006/7.

**SLOWING DOWN TREATMENT RATES**

There was some evidence from PCT board papers that PCTs and hospital trusts had slowed down the rate at which patients are treated in the acute sector, a tactic that attracted media attention in early 2007 (for example, Hull 2007, Timmins 2007). Fourteen PCTs referred to some sort of slowing down using slightly different language, for example, ‘freezing non-essential elective work’ or ‘slowing down elective work’. There was no evidence of trusts failing to treat patients within the national inpatient waiting time target (26 weeks) but in a few cases the slowdown took the form of trusts abandoning more stringent local waiting times targets (20 weeks, for example) as a way of saving money:

_To avoid worsening the financial position, it is not now planned to meet the 20 week milestone but to maintain achievement at 26 weeks as per the current national target._
In one case a PCT reported in its board papers that plans ‘freezing all non-essential elective activity, so that only urgent patients would be treated in the last 2-3 months of the financial year’ were overturned by the strategic health authority, despite planned savings worth £6.7 million.

**FUTURE PRESSURES ON DEMAND MANAGEMENT: 18-WEEK REFERRAL-TO-TREATMENT TARGET**

Forty PCTs mentioned work to model or forecast the impact of meeting the 18-week referral-to-treatment target (which needs to be met by the end of 2008). Not all the PCTs gave figures; however, a common theme is that meeting the target is expected to be costly, because extra activity will need to be purchased from hospitals.

> The cost of delivering the 18 weeks milestone target for elective care is one of the largest calls on resources in 2007/8.

> One of the most significant areas of cost pressure for 2007/08 is commissioning additional activity to meet the 18 week referral to treatment target.

Several PCTs commented that their forecasts also assumed successful ‘demand management’, in other words, it would be easier to meet the target if fewer people were being referred to hospital in the first place.

> To make this target both affordable and deliverable are all the demand management schemes for emergency activity, elective and outpatient activity. Achieving these demand management initiatives remains the key to delivering on the 18 week target.

Other PCT papers also referred to the need to redesign of services to make them more streamlined and to reduce waiting times for patients, particularly at the diagnostic test point of the patient pathway, where some of the worst delays currently occur (see 18-week waiting times target briefing: [www.kingsfund.org.uk/go.rm?id=2284](http://www.kingsfund.org.uk/go.rm?id=2284)).

**Conclusion**

It seems clear that the record deficits of 2005/6 and the government’s response to them have had a powerful effect on NHS organisations. A combination of assertive financial management at SHA level (top slicing and generating a contingency reserve) coupled with intense pressure on trusts to manage their finances more accurately and effectively, does seem to have successfully reversed a very large net deficit. While some of the tactics are undeniably temporary in nature, there has also been considerable effort expended, as the analysis of PCT board papers has shown, to tackle some of the underlying drivers of overspending, in particular the tendency for hospitals to perform more operations than local budgets have allowed. It is also clear that some of the more important methods of curbing ‘demand’ for hospital services, by designing better non-hospital alternatives for instance, are embryonic at this stage.

The PCT board papers reveal an often uneasy relationship between the PCTs and acute trusts, the latter now experiencing strong incentives, under the current reform strategy, to increase their income through extra activity. This relationship will be tested further in 2007/8 and beyond. In March 2008 hospital trusts are expected to meet and then maintain the testing new waiting time target (18 weeks from GP referral to treatment) which will almost certainly require an expansion of activity. PCTs, on the other hand, will come under increasing pressure to keep hospital activity under control, particularly as the big increases in NHS funding are due to end after 2008.
