

HOW MUCH SHOULD WE SPEND ON THE NHS?

ISSUES AND CHALLENGES ARISING FROM THE WANLESS REVIEW OF FUTURE HEALTH CARE SPENDING

Edited by John Appleby, Nancy Devlin and Diane Dawson



King's Fund
11-13 Cavendish Square
London W1G 0AW
www.kingsfund.org.uk



Office of Health
Economics
12 Whitehall
London SW1A 2DY
www.ohe.org



Centre for Health
Economics
The University of York
Heslington
York YO10 5DD
www.york.ac.uk

© King's Fund, Office of Health Economics and University of York Centre for Health Economics, July 2004

Any part of this publication may be reproduced without permission for educational and non-profit purposes only if the source is acknowledged. All other rights reserved, including the right of reproduction in whole or in part in any form.

Price £12.50

ISBN 1 899040 87 0

Printed by BSC Print, London.

Office of Health Economics

The Office of Health Economics (OHE) was founded in 1962. Its terms of reference are to:

- commission and undertake research on the economics of health and health care;
- collect and analyse health and health care data from the UK and other countries;
- disseminate the results of this work and stimulate discussion of them and their policy implications.

The OHE is supported by an annual grant from the Association of the British Pharmaceutical Industry and income from commissioned research, consultancy and sales of its publications.

The King's Fund

The King's Fund is an independent charitable foundation working for better health, especially in London. We carry out research, policy analysis and development activities, working on our own, in partnerships and through grants. We are a major resource to people working in health, offering leadership and development courses; seminars and workshops; publications; information and library services; and conference and meeting facilities.

Centre for Health Economics

The Centre for Health Economics (CHE) is a specialist health economics research unit within the University of York. Now in its 21st year, CHE has continually expanded since its inception in 1983. Providing research of worldwide repute, its principal areas of activity include: outcome measurement, economic evaluation of health technologies, primary care, addiction and health promotion, health economics in low and middle income countries, resource allocation and Health policy.

Acknowledgements

<p>The chapters in this book are based on the proceedings of a workshop on November 13th 2002, jointly organised by the OHE, King's Fund and Centre for Health Economics.</p>

We are grateful to Tony Culyer, Jon Sussex, John Appleby and Alan Williams for chairing the sessions in that workshop. A list of participants in the workshop is provided in Appendix 1.

CONTENTS

Contributors	5
Preface	6
Introduction Adrian Towse	8

Section 1: Modelling health care productivity

1 Measuring productivity in the UK and trends in the service and non-market sectors Mary O'Mahony	15
2 Is health care productivity cost-reducing or quality-enhancing? Diane Dawson and Adrian Towse	26
3 Discussion: modelling health care productivity	37

Section 2: Modelling the demand for health care

4 UK evidence on 'healthy lifestyles' and their effect on the demand for health care Klim McPherson	44
5 Age, proximity to death and future demands on the NHS Meena Seshamani and Alastair Gray	56
6 Discussion: modelling the demand for health care	70

Section 3: Modelling reductions in waiting

7 Resource and service implications of short waiting times: the French experience Stephane Jacobzone	76
8 UK evidence on waiting John Yates	94
9 Discussion: modelling reduction in waiting	107

CONTENTS

Section 4: Modelling NHS spending: what can we learn from elsewhere?

10	Modelling health care spending in the US Mark Freeland	116
11	Lessons from macroeconomic modelling in the UK Martin Weale	124
12	Discussion: what can we learn from elsewhere?	133

Section 5: Conclusions

	Appendix	144
	List of attendees	

CONTRIBUTORS

Diane Dawson

Centre for Health Economics, University of York

Mark Freeland

Office of the Actuary, Centers for Medicare & Medicaid Services, US

Alastair Gray

Institute of Health Sciences, University of Oxford

Stephane Jacobzone

OECD

Klim McPherson

University of Bristol

Mary O'Mahony

NIESR, London

Meena Seshamani

Institute of Health Sciences, University of Oxford

Adrian Towse

Office of Health Economics, London

Martin Weale

NIESR, London

John Yates

University of Birmingham

Editors

John Appleby

King's Fund, London

Diane Dawson

Centre for Health Economics, University of York

Nancy Devlin

City Health Economics Centre, City University, London

PREFACE

Prior to the 2002 Spending Review Derek Wanless was asked by the Chancellor of The Exchequer to assess the 'financial and other resources required to ensure that the NHS can provide a publicly funded, comprehensive, high quality service on the basis of clinical need and not ability to pay.' The resultant report on the future course of NHS spending was immensely important. Not only did it directly inform the Chancellor's spending plans for the NHS for the next five years – which gave the NHS an unprecedented real increase of over 40% by 2007/8 – but it also laid out a framework for revisiting a fundamental public policy question: how much do we want to spend on health care?

There is, of course, no objectively correct answer to this question; it is a matter of choice – albeit one that, in an economist's rational world, is informed by the value we place on the benefits derived from spending on the NHS compared with other uses of society's scarce resources.

Derek Wanless and his team tackled this high-level resource allocation problem by firstly defining and quantifying what a 'high quality' NHS would look like, and then estimating what such a service would cost. By varying assumptions about factors which would affect the costings – population size, demographic structure, healthiness, NHS productivity – three NHS spending paths (or 'scenarios') to 2022 were produced¹.

Derek Wanless stated that his review of future NHS spending should not be a one-off exercise, but should be revisited to update estimates, assumptions and to refine spending models as new data emerges and knowledge expands.

As a contribution to future reviews and research in this area, a one day seminar organised by the Office of Health Economics, the King's Fund

¹ In February 2004, a further Wanless report was published exploring in more detail the challenges in implementing the lower-cost, 'fully engaged' scenario, with a particular focus on cross-departmental work on preventative health measures and health inequalities ('Securing good health for the whole population': http://www.hm-treasury.gov.uk/consultations_and_legislation/wanless/consult_wanless04_final.cfm.)

PREFACE

and the Centre for Health Economics was held on 13 November 2002 to critically examine the assumptions, estimates and models underpinning the recommendations of Derek Wanless' review. This book brings together the outcomes of the seminar.

John Appleby
Nancy Devlin
Diane Dawson

Introduction

ADRIAN TOWSE

This chapter looks at:

- Wanless' scenario projections;
- Wanless' approach to modelling – a four stage model;
- issues arising from the Wanless report;
- lessons for future modelling.

Introduction

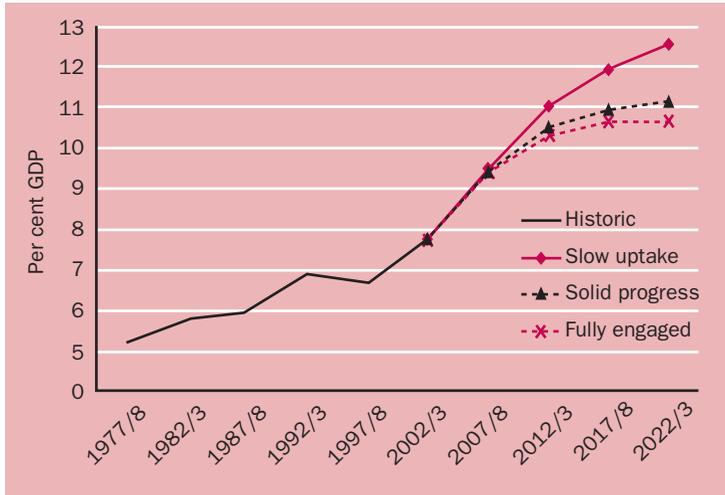
Derek Wanless was commissioned by the Chancellor of the Exchequer to investigate the resources that the NHS would require over the next 20 years. These estimates were to be made on the basis that the NHS would remain a publicly funded service and that health care would continue to be available on the basis of clinical need and not ability to pay. The Wanless report (2002) was published at the time of the 2002 Budget. The importance of the report was immediately evident: the unprecedented increases in spending on the NHS announced by the Chancellor in the 2002 Budget (HM Treasury 2002) were directly in line with those Wanless recommended.

Wanless scenario projections

Wanless and his team produced three projected spending paths for the NHS, one for each of the scenarios shown in Figure 1. These scenarios differed in terms of the degree of 'engagement' of both producers and consumers of health care, that is the extent to which people are aware of the impact of their own behaviours on their health; and the extent to which producers are aware of and responsive to opportunities to improve productivity and efficiency.

The 'fully engaged' scenario, which is based on the highest life expectancy, has the lowest health care expenditure as a share of GDP (10.6% by 2022/23). 'Slow uptake', which has the poorest health

Figure 1 Anticipated Annual Health Care Costs £ Billion



Source: Wanless, 2002.

outcomes, has the highest projected spending of 12.5% of GDP by 2022/23.

There are two reasons why the ‘fully engaged’ scenario turns out to be the lowest cost. First, in this scenario people are healthier – in part because they are assumed to take better care of their own health and therefore have fewer episodes of illness. Second, under the ‘fully engaged’ scenario the NHS embraces productivity improvements, in particular the use of new information technologies, in such a way that it is able to improve productivity and reduce costs.

On the other hand, there are some offsetting costs associated with this scenario: for example it is assumed that, as part of being more engaged with their own health, individuals make greater use of some health services, such as check-ups. Overall, however, the scenario in which people are healthiest and in which the NHS is investing more in technology, generates the lowest projections of required increases in NHS spending over the 20 year period.

INTRODUCTION

The Chancellor in his 2002 Budget announced that he was increasing NHS funding in line with the 'fully engaged' scenario. NHS expenditure is planned to grow by 7.5% annually in real terms over the five year period 2002/03 to 2007/08. This commitment has been sustained in the 2003 and 2004 Budgets.

A four stage model

Wanless essentially employs a four-stage approach to modelling future NHS costs. The starting point is a projection of the impact of demographic change at constant rates of use and quality of care. The second step is to change one variable: improving the quality of care in line with the National Service Frameworks (NSFs). Five NSFs have already been published and assumptions can be made about the NSFs still to be introduced throughout other disease areas. The third step is to explore the effect of changes in age-specific use of health care, driven primarily by the degree of engagement of consumers as noted above, i.e. to look at the impact of changes in demand for health care for a given level of need. Finally, the impact of improved access, technological change and productivity gains (taking account of the degree of engagement of NHS professionals) are taken into account in the spending projections.

Key issues in modelling NHS spending

It is worth noting how crucial the cost-reducing productivity assumptions are to the Wanless projections. Under the 'solid progress' scenario, a one percentage point per annum difference in the assumptions about cost reducing productivity means that at the end of 20 years there is potentially a two to four point difference in the health care share of GDP. Chapters 1 to 3 examine issues surrounding productivity measurement, the evidence on productivity gains and its implications for health care spending.

The impact of lifestyle and ageing are also crucial. A key Wanless assumption, already noted, is that healthy lifestyles lead to an overall reduction in the demand for NHS services. The Wanless report also proceeds on the basis of population ageing having its principal effect

on health and health care via a ‘compression of morbidity’ effect. This impacts on modelling in two ways. First, there is a concentration in health care spending at the end of life (whenever that may be and regardless of age per se). Second, that up until that point, people are healthier than they have been before. Chapters 4 to 6 explore these issues in detail, including the evidence of what happens in the final years of life in terms of the concentration of health care spending.

Another key assumption is that inpatient waiting times are to be reduced to two weeks. Evidently, the existing Department of Health waiting times model simply could not cope with projecting a reduction in waiting time of such magnitude. Therefore, the Wanless team proceeded to model this by assuming that increases in activity would reduce waiting times. The Wanless team therefore looked at examples of countries that had waiting times they regarded as close to where the NHS ought to be – an obvious example being France – and the activity rates in these countries. Thus the model essentially assumed that if the NHS was able to achieve the activity rates of these countries it would also be capable of achieving comparable waiting times. Chapters 7 to 9 explore waiting time models, discuss the implications of France’s experience with waiting times and, in the light of evidence on waiting and waiting times in the NHS, question both the low waiting time ‘vision’ for the UK and the links between activity and waiting that are apparent in the Wanless model.

Lessons for future modelling

The Wanless report recommends that the modelling exercise should be repeated in five years. There is an issue as to whether more could be made of the existing exercise if the full assumptions and models were put in the public domain. This would enable academics and other interested parties to evaluate the impact of alternative assumptions. There has been some debate between the Treasury and the Department of Health as to whether that is (a) practical, and (b) a good thing. To date there has been no further public disclosure and the Treasury and Department of Health team that supported Wanless was quickly disbanded.

Chapters 10 to 12 touch on the debate about whether such modelling exercises ought to be ‘open’ (for example, with the encouragement of rival or competing models) in order to improve the modelling process and what may be learned from it. As discussed in chapter 10, in the US there are competing models of both Medicare and overall US health care spending. In the UK, there are, in principle, competing macroeconomic models of the economy.

An important prior question, however, is what is the purpose of modelling health care spending? Are we simply trying to understand better the links between, for example, waiting times and activity levels, or funding levels and health outcomes? Or, a related issue, are we trying to project expenditure so that we can ascertain how much the NHS ought to be spending in order to achieve certain goals? Alternative models and modelling approaches can achieve different things; the complexity of the model might need to be different, depending on the objectives of the exercise.

Building on Wanless

The Wanless modelling exercise was undertaken under extremely tight political and time constraints. As a result, it had to make a number of important assumptions about the future course of health status and the relationships between activity and waiting times. There are aspects of the model that could undoubtedly be improved. However, the set of models employed in the Wanless report has already played a valuable role by making links between inputs, environment and outputs that have really not been explored properly before in the context of the NHS, helping us to better understand the implications of key assumptions on overall health care spending.

The aim of this book is to consider in detail the issues and evidence surrounding the determinants and modelling of health care spending – and what remains to be done if the Wanless report is to provide the foundations for a future research programme. It should be noted that the views expressed by the authors of succeeding chapters are given without full knowledge of the details of the Wanless set of models as these have not, at time of writing, been released by the Treasury.

INTRODUCTION

REFERENCES

Wanless D (2002) *Securing our Future Health: Taking a Long-Term View. Final Report.* HM Treasury: London.

HM Treasury (2002) *Spending Review. Opportunity and Security for All: Investing in an Enterprising, Fairer Britain. New Public Spending Plans 2003-2006.* HM Treasury: London.

HM Treasury (2003) *Budget Report 2003.* HM Treasury: London.

Section 1

MODELLING HEALTH CARE PRODUCTIVITY

Chapter 1

Measuring productivity in the UK and trends in the service and non-market sectors

MARY O'MAHONY

This chapter looks at:

- alternative measures of productivity;
- comparisons of health sector productivity with other sectors of the economy;
- the problem of adjusting for quality of output.

Introduction

This chapter highlights five key issues pertinent to the Wanless model. First, what do we mean by productivity and how do we measure it in the public sector? Second, what have been the trends in productivity growth in the UK and US since 1979 for the service sector as a whole and for health services in particular? Third, what has been the impact of changes in the mix of inputs in productivity? Fourth, how should outputs be adjusted for changes in quality over time and, finally should we measure outputs or outcomes?

Wanless on productivity

The Wanless report did not make it clear how the review team had defined productivity in the NHS; there appear to have been a number of different measures used. My assumption from reading the report is that Wanless used a definition of productivity usually known as *total factor productivity (TFP)*, and that different assumptions concerning changes in TFP in part defined the three scenarios set out by the review. For example, the slow uptake scenario assumed that health service productivity would rise at about 1.5% per annum, matching performance in the wider UK service sector and roughly what has been achieved over the past 20 years, (productivity in the private

sector has probably been slightly higher than this). The *solid progress* and *fully engaged* scenarios have higher productivity assumptions, starting at 2% per annum over the next five years, rising to 3% per annum over time.

The Wanless review also split this productivity growth into cost reductions and quality improvements – an important issue further examined in the next chapter.

Productivity: definitions and measurement

There are a number of measures of productivity which can be used when examining national accounts data. The simplest, and one that has been the mainstay of research on productivity, is labour productivity, for example, output per hour worked. This is easy to measure and, at the national level, is associated with increases in the standard of living. But labour productivity is only a partial measure. For example, much of the advance in productivity in health care is not due to improvements in labour productivity but to new technology – which is part of capital input – or drugs – which are part of materials input.

In order to capture the full range of factors affecting productivity a broader measure is required – such as total factor productivity².

The conventional measure of TFP utilises a Törnqvist Index (see Box 1.1). In essence, this index measures productivity as the difference between the growth in outputs minus the growth in inputs between two time periods. Inputs are usually divided into three categories: labour, capital and material, and the input growths are weighted by their importance in the value of output, i.e. their share of the value of output.

The Törnqvist equation is based on a number of very restrictive assumptions about perfect markets, and has been in use for nearly 50 years (see Solow, 1957). There has been a great deal of work in the

² TFP has also been called: multi factor productivity, residual productivity, the Solow residual and underlying productivity.

Box 1.1 Productivity: definitions and measurement

The conventional measure of total factor productivity (TFP) as expressed by the Törnqvist equation:

$$\Delta \ln (A) = \Delta \ln (Q) - \alpha^L \Delta \ln (L) - \alpha^K \Delta \ln (K) - \alpha^M \Delta \ln (M)$$

TFP = Output – Labour – Capital – Materials

Where, α^i = input i 's share in the value of output (averaged over periods t and $(t-1)$)

TFP is the difference between real output growth and (share weighted) input growth.

academic literature on this approach to measuring productivity and it has remained the mainstay of productivity analysis and measurement over the last five decades.

Unfortunately, this measure of TFP cannot be used with national accounts data because we generally do not have information on material input. However, there is a slightly different measure, the value added measure, which uses just two inputs, labour and capital, and it is this which forms the basis of the productivity measurement which follows. Possible differences between this and the broader measure of productivity are discussed later.

A more general measure still would allow for changes in the quality of the inputs. For example, it is possible to look at a mix of various types of labour, capital and materials and the effect they have on output growth and productivity. Some estimates of productivity based on variations in the quality of inputs are presented below.

Estimates of productivity

Labour productivity

Table 1.1, below, shows data for two time periods – one that underlies the Wanless projections in the 1979 to 2000 period and the final ten years of this period – and the growth in 1990-2000 minus the

Table 1.1 Labour productivity growth, UK and US: 1979-2000

%	UK			US		
	1979-2000	1990-2000	1990-00 minus 1979-90	1979-2000	1990-2000	1990-00 minus 1979-90
Total economy	2.07	2.14	0.13	1.45	1.58	0.26
Market services*	2.06	2.77	1.35	1.98	2.82	1.59
Distribution	2.00	2.68	1.29	2.84	4.01	2.23
Finance and Business	0.79	1.42	1.21	0.86	1.65	1.52
Non-market**	0.37	0.97	1.14	-0.29	-0.48	-0.36
Health	1.00	2.73	3.30	-1.59	-1.33	0.48

*Sum over transport, communications, distribution, financial services, business services and personal services;

** Sum over public administration, education and health.

Sources: Based on National Accounts data, UK: Blue Book, ONS; US: National Income and Product Accounts, BEA. For source details see O'Mahony (1999).

growth in 1979-1990. Labour productivity growth for the 1979 to 2000 period for the total economy was about 2% in the UK – a figure that has not varied for many years (see O'Mahony, 1999). In comparison with the US, productivity growth tends to be higher in the UK. Although the US has the highest productivity levels in the world, European countries tend to have higher growth rates as they catch up with the US. The acceleration in growth in the 1990s relative to the 1980s was marginally higher in the US, mostly due to rapid US growth after 1995 coinciding with a slowdown in UK labour productivity growth.

Now consider total market services. Here we see labour productivity growth rates fairly much on a par with the total economy for the entire period, with somewhat greater growth in the final decade. But

there is some variation within this group. Table 1.1 also shows estimates for two of the largest sub-sectors of market services. Productivity growth rates in distribution in the UK tend to be similar to those achieved in general market services. But in financial and business services, labour productivity growth rates are relatively low in both countries. This is related to the way we measure output, since it is difficult to define what is being produced by these sectors.

Similar problems apply to non-market services where measurement problems are compounded by the lack of market prices. In the past, national statistical offices tended to measure outputs largely by inputs so that labour productivity growth rates tended also to be low in these sectors. Given the measurement difficulties, in comparing both across time and countries, the acceleration measures shown in Table 1.1 probably give the more accurate picture of cross country performance.

In the last row of Table 1.1 we look at the health sector component of non-market services. The numbers refer to health and social services since UK national accounts data do not permit further disaggregation. Here there is a significant difference between the two countries in growth rates, but both show improved performance in the 1990s relative to the 1980s. The negative growth rates in the US largely reflect problems in the measurement of price deflators, which do not adequately account for quality changes. (See the discussion in Triplett, 2001 where it is suggested that more reasonable price deflators would change the US picture from negative to large positive labour productivity growth rates).

Total factor productivity

TFP growth rates (see Table 1.2) allow additionally for changes in capital input, so they are lower than the labour productivity growth rates in Table 1.1. This shows a similar picture across time, and comparing the two countries, as for labour productivity. In both countries we see an acceleration of total factor productivity growth rates in the health sector. In the UK this is greater than that achieved in the total economy and the market service sectors. In the US, the TFP acceleration is also greater than that for the total economy but lower

Table 1.2 Total Factor Productivity (TFP) growth, 1979-2000

%	UK			US		
	1979-2000	1990-2000	1990-00 minus 1979-90	1979-2000	1990-2000	1990-00 minus 1979-90
Total economy	1.41	1.49	0.16	1.18	1.23	0.09
Market services*	1.22	2.03	1.55	1.56	2.28	1.37
Distribution	0.79	1.37	1.10	2.23	3.30	2.05
Finance and Business	-0.05	1.22	2.44	0.40	1.28	1.69
Non-market**	0.14	0.69	1.05	-0.25	-0.53	-0.53
Health	0.40	1.96	2.98	-1.90	-1.62	0.55

*Sum over transport, communications, distribution, financial services, business services and personal services;

** Sum over public administration, education and health.

Real investment was cumulated to a capital stock measure using the method outlined in O'Mahony (1999). Capital input was weighted by one minus labour's share of value added with the latter equal to labour compensation plus imputed compensation for self employed divided by value added.

Underlying sources as for Table 1.1.

than in market services including both distribution and financial and business services.

These numbers might seem a bit surprising as there has been much discussion of decreasing productivity and efficiency in the UK health service. It is interesting to break down the last period into two halves. Here we see slightly higher TFP growth in the UK health and social services sector since 1995, about 2.3% against 1.7% in the period 1990-95, an acceleration of 0.6 percentage points. At the same time TFP growth rates were decelerating in the UK economy in general and showed only a small acceleration in market services. So measured against general performance UK health sector TFP was relatively

strong. In the US the acceleration in the health and social services sector was even stronger across these two time periods, equal to about 1.4 percentage points, and about twice that achieved in the total economy. Hence if we focus on improvements in productivity growth rates across time, i.e. the growth acceleration, then the UK health sector has performed well relative to achievements in the remainder of the economy but not as well as the comparable sector in the US.

Office of National Statistics health services productivity measure

An alternative measure of health sector productivity that suggested a decrease in efficiency has been published recently by Pritchard (2002). This estimate is for health services alone and does not include social services. Using a cost-weighted activity index (which is the basis for the output increase both in this measure and in the national accounts), Pritchard reported productivity growth between 1995 and 1999 of around 2.2% per annum. As real expenditures on labour, capital and materials have been growing at about 4.1% per annum, this implies a reduction in TFP of around 1.8% per annum – somewhat different from the growth in TFP of 2.3% which underlies the figures in Table 1.2.

So why the difference between these two measures of health sector productivity? First, it is perhaps debatable whether the measure used by Pritchard is in fact a true measure of productivity. However, the two measures are based on more or less the same data, so it should be possible to reconcile them. There are some obvious differences between the two approaches. First, the ONS measure is for health alone whereas the TFP measure, using national accounts data, includes social services. Second, the cost-weighted activity index measures only part of the health sector whereas TFP includes the whole sector. Although ONS were not specific about the size of this difference, it probably amounts to around 1 percentage point. This leaves a total difference of around 3 percentage points between the TFP measure and the measure used by Pritchard.

Another reason for differences in our measures of productivity is that the national accounts use a value added measure which excludes

changes in material inputs, which have been growing faster than output. Important examples of this are the use of new and more expensive drugs. This may well account for much of the difference in the two productivity measures. Finally, any remaining difference will be due to changes in the mix of labour and capital inputs. Generally, there have been shifts to using more highly skilled labour and high technology equipment, but the input expenditure deflator that is used by ONS does not take this into account.

But how important are these changes in the mix of labour and capital input? At the moment I cannot answer that for the health services, as I do not have sufficient data. Recent estimates by the National Institute of Economic and Social Research, however, suggest about 85% of TFP growth in total non-market services can be accounted for by a combination of increases in the quality of labour due to employing more highly skilled workers and to greater use of information and communications technology capital. Applying this ratio to health would account for much of the remaining difference between the estimates in Pritchard (2002) and the national accounts based estimates presented here.

It is worth adding that the input quality adjustments we have made probably do not adequately account for this element in the health service, because our estimates only distinguish graduates from people with intermediate qualifications, and miss any quality rise within these broad categories. In addition, ICT capital is probably not the biggest driver in the health sector, which utilises a wide range of high technology equipment.

Starting with labour productivity and moving to adding the contributions to output growth from additional inputs, and changes in the quality of those inputs, highlights the sources of output growth. The focus is then on the (positive) output change and the channels through which this has been achieved. It suggests that raising output growth requires expenditure on higher quality inputs with little by way of costless improvements. I favour this method of measuring TFP as it is the most common approach employed in the private sector.

Quality of output

So far I have not discussed a major measurement issue: the need to adjust output for quality change. This is probably the most important problem in measuring productivity in health services. Since I cannot present any empirical evidence on this, I will merely outline the issues involved and the ideal concepts to measure.

Output should be measured as the number of units produced of a good or service, adjusted for quality change. In the private sector this is most commonly derived as the value of output deflated by a price index. This is how we measure almost all of the private sector output, so the quality adjustment can be incorporated through the price deflator. But in some parts of the private sector quantity indicators of output change have also been used in the national accounts. This was most commonly the case for transport services in the past, where it was usual to employ indicators such as passenger kilometres travelled. Over time national accounts have been moving as much as possible away from using quantity indicators and more towards employing deflated measures of output. Nevertheless, in some parts of the financial business services sector, quantity indicators are still employed. When using quantity indicators of output, it is necessary to incorporate the quality adjustment directly in the quantity indicators.

In the UK health services a quantity indicator is employed: the cost weighted activity index. This is necessary as there are no market prices, so a deflated value series cannot be used. But the general conclusion amongst analysts is that this can be a sensible measure to use in international comparisons. It is valid to compare quantity indicators in a country such as the UK (where the service is publicly provided) with a deflated value series in a country such as the US (where the service is privately provided). The main problem in both series is the necessity to take account of quality change. As mentioned earlier, research by US economists suggests that the US deflators did not adequately account for quality change (e.g. papers in Cutler and Berndt, 2001). Deflators are generally based on inputs such as the cost per day of a stay in hospital but should also take account of, say, increasing use of outpatient care.

Even this approach is not really measuring quality because the deflators are still based on input costs. How should we account for output quality change? One solution, associated with, amongst others, Jack Triplett of the Brookings Institute, is to move to using outcome measures such as quality adjusted life years (QALYs). The argument here is that the consumer cannot directly evaluate the merits of different types of treatment, so it is necessary to look directly at the effectiveness of these treatments. Triplett, in an interesting paper titled 'What is the difference in health? Human repair versus car repair' (2000), compared the two service industries of human repair and car repair. His main argument is that you can scrap the car but you cannot scrap the human. Therefore, it is necessary to look directly at outcomes in health services in order to capture the effectiveness of treatments. In the case of car repairs it is not necessary to examine outcomes as the market already does so. The ability of the consumer to scrap or sell off the car means they are in a position to forego the service (the repair) if they deem it inappropriate.

This is probably one step too far for the national accounts. The national accounts measure the value of transactions in a given period and health care outcomes typically stretch for longer than this. When using outcome measures it is necessary to take account of exogenous lifestyle factors. In some cases it may be possible to measure outcomes on a disease by disease basis so that calculations are based on a homogenous population of patients, or by the use of regression analysis to adjust for extraneous influences. But in general this goes way beyond the boundary of the national accounts. Hence the use of outcome measures would require a major research input.

Conclusions

National accounts estimates suggest reasonable productivity growth in the UK health services. Much of this change, given the output measure in the national accounts, is probably due to the quality mix of inputs. If account is also taken of changes in purchased inputs, this could result in negative underlying TFP growth. But this result is not unusual. Negative TFP growth rates arise over short time periods in

many sectors, due to lagged responses. The national accounts' estimates do not adequately adjust for changes in the quality of output. An alternative estimate based on outcomes could lead to estimates of positive residual productivity growth in the health services. However, an examination of performance in the health care sector based on outcomes probably needs to be carried out by independent researchers, since national accounts statisticians will not cross the traditional boundaries entailed in outcome measurement.

REFERENCES

Cutler DM and Berndt ER, eds. (2001) *Medical Care Output and Productivity*. Chicago University Press: Chicago.

O'Mahony M (1999) *Britain's Productivity Performance 1950-1996*. NIESR: London.

Pritchard A (2002) 'Measuring productivity change in the provision of public services', *Economic Trends*, The Stationery Office: London.

Triplett JE (2001) 'What's different about health? Human repair and car repair in national accounts and national health accounts', in Cutler DM and Berndt ER eds. (2001) *Medical Care Output and Productivity*. Chicago University Press: Chicago.

Solow RM (1957) 'Technical change and the aggregate production function', *Review of Economics and Statistics*, vol. 39, pp.312-320.

Chapter 2

Is health care productivity cost-reducing or quality-enhancing?

DIANE DAWSON and ADRIAN TOWSE

This chapter looks at:

- the Wanless assumptions on productivity change;
- the evidence on unit cost reduction in the NHS;
- recent US methodologies for estimating quality change;
- approaches to measuring quality change in the UK.

Introduction

The 2002 Wanless report pointed to productivity change as one of the areas where key assumptions were needed in order to develop long term forecasts of expenditure trends within the NHS. The review seemed to see changes in productivity as being driven by two different engines. One was technical change, which, for the UK, may be exogenous in terms of the development of new technologies. Wanless assumed medical technology added 2-3% per annum to NHS cost. However, this technology would increase productivity if it added more than 2-3% to quality adjusted output. The other engine of change was organisation for service delivery. There is, however, an interaction with technical change as organisational behaviour and incentives can have an impact on the rate at which technological innovations are adopted within a health care system. The importance of understanding productivity change, if we are to forecast health care expenditure, is beyond dispute. What is at issue is the evidence base for making these forecasts and the methodologies needed to measure the impact of technical change on expenditure and outcomes.

The Wanless review

The 2002 Wanless report incorporated a range of assumptions on the rate of productivity change over the next 20 years in the NHS. It also made the specific assumption that productivity growth would

Table 2.1 Wanless productivity assumptions, per cent per annum

	Unit cost reduction		Quality improvement		Quality adjusted productivity	
	Lower	Higher	Lower	Higher	Lower	Higher
2003/04 - 2007/08	0.75	1.00	0.75	1.00	1.50	2.0
2008/09 - 2012/13	0.75	1.25	0.75	1.25	1.50	2.5
2013/14 - 2017/18	1.00	1.50	0.75	1.50	1.75	3.0
2018/19 - 2022/23	1.00	1.50	0.75	1.50	1.75	3.0

Source: Wanless, 2002.

manifest itself in unit cost reduction and quality improvements in roughly equal parts. Table 2.1 is taken from the report and shows that there were both high and low estimates of productivity change, but in all cases a judgement was made as to how the gain will appear: productivity change that is reducing unit cost and productivity change that is enhancing quality.

One of the problems of assessing the usefulness of these productivity assumptions is that there is no evidence from the UK – or for that matter from other health care systems – on the impact of trend changes in productivity in terms of unit cost reduction or quality improvement. One of the reasons we do not have evidence to support the Wanless assumptions is that in the past we have not tried to measure annual improvements in health outcomes and the quality of care. There is no basis for examining whether the Wanless estimates are reasonable to incorporate into an expenditure forecasting exercise.

NHS unit cost reduction

Almost all of the research in the UK has been on unit cost reduction.

Table 2.2 Unit cost reduction targets for the NHS

Efficiency Gain (%)		
	Target	Actual
1992/93	2.00	Achieved
1993/94	2.00	Exceeded
1994/95	2.25	On course
1995/96	3.00	Achieved
1996/97	3.00	Not available
1997/98	2.70	1.70
1998/99	3.00	Not available
1999/2000	3.00	Not available
2000/01	3.00	Not available
2001/02	2.00	Not available

The NHS has planned for unit costs to fall between 2% and 3% per annum for at least the last decade.

These targets were regularly reported in Department of Health reports or in evidence to the House of Commons Health Committee. There is some suggestion that in earlier years these targets were met. There is virtually no reported evidence of that for later periods (see Table 2.2).

The methodology for measuring unit costs and changes in unit cost is similar to that adopted by the Office for National Statistics (ONS) and discussed in the previous chapter. It is an activity based measure: cost-weighted activity divided by total expenditure. Some of the work carried out at the Centre for Health Economics at the University of York has suggested no evidence of cost reducing efficiency gains over the period from 1994 to 2000 (Dawson and Jacobs, 2001). This is consistent with the results reported by ONS. The question is, therefore, what is the evidence base for the Wanless assumption of annual unit cost reductions of 0.75% to 1.5% per annum? Given our

lack of evidence of unit cost reductions and the NIESR estimates of positive productivity gains, there is a clear need to examine the methodologies appropriate for measuring and forecasting future changes in unit costs.

One issue to be addressed is the time scale over which we expect technical change to lead to changes in unit costs. During the 1970s and 1980s many technical changes were accumulating and impinging on average length of hospital stay. The consequent reduction in average length of stay led to reductions in unit cost. Since that particular source of productivity gain has receded, researchers have not identified another driver that would deliver a similar widespread change in medical practice leading to falling treatment costs. We may have exhausted the cost saving benefits of one generation of technical innovation – but we have yet to identify the next. There is no reason to assume that the rate of technical innovation will be smooth with predictable annual change.

The Treasury tends to set very short term targets for productivity change measured as unit cost reduction. We know from other sectors of the economy that sustainable cost reduction is a long term issue, however. The current value for money public service agreement (PSA) for the Department of Health requires the NHS in England to deliver a 1% unit cost reduction every year for the next three years. If it fails, according to the PSA, the Secretary of State for Health will be held responsible. Not only does existing research fail to indicate where or how these cost reductions will emerge, but the present Departmental investment plans will in fact cause unit costs to rise. A great deal of expenditure is on activities designed to improve access (that is, reduced waiting times) and on the quality of the NHS infrastructure. So, currently at least, there is a Departmental expenditure programme that is inconsistent with both the PSA target and the Wanless assumptions.

Quality improvement

So far our discussion has focused on one half of the Wanless productivity assumption – the reduction in unit costs. However, in the

UK we have not even begun to address the issues raised by the other half of the Wanless productivity assumption – increases in quality of services at a rate of 0.75% to 1.5% per annum. To get some idea of what might be done we consider two examples from the US literature on heart attacks and depression. We examine what the evidence suggests about technological change as a driver of cost, the extent to which we can say it is quality enhancing. We then ask if we can measure quality in an NHS context, and the implications for modelling future productivity changes.

Heart attacks

The first example is based on work by Cutler and McClellan (2001), who attempted to estimate the monetary return on investment in technology for treating heart attacks over the last 15 years – for example, thrombolytic drugs, catheterisation, various procedures for revascularisation and long term drug therapies. They looked at evidence for Medicare beneficiaries (primarily the elderly) over a 15 year period, and then monitored treatment costs 12 months after an initial heart attack. Table 2.3 sets out Cutler and McClellan’s basic calculations.

What they found was a significant increase in expenditure, of which around half was due to greater use of the newer technologies. So, newer, more expensive technology was being used over the period

Table 2.3 **Heart attacks**

	1984	1998	Change
Total spend	\$3,000 million	\$4,000 million	3.4% pa
Number of cases	246,000	221,000	-0.8% pa
Cost per case	\$12,000	\$22,000	4.2% pa (\$10,000 over 15 years)
Survival	5 years	6 years	1 year over 15 years

Source: Derived from Cutler and McClellan, 2001.

and it was the change in take-up of newer technologies that was driving the increase in costs, rather than the increase in prices of the technologies.

Looking at the last column of Table 2.3, it can be seen that total spending on treatment rose by 3.4% per annum and the number of cases was decreasing by 0.8% per annum. So, overall, cost per case increased by \$10,000 (from \$12,000 to \$22,000). Post-treatment survival rates increased by one year (from five years to six years).

It would appear that, on average, Medicare patients have gained an extra life year at a cost of \$10,000. If we assume that this life year is lived in perfect health and is equivalent to one quality adjusted life year (QALY), and that, further, the value of a QALY is in the region of \$50,000 (that is, equivalent to an approximate value of £30,000³), then the net value of the benefit arising from extra investment in new cardiac technologies is equivalent to \$40,000 (i.e. \$50,000 minus \$10,000).

This looks to be a reasonable return. However, we need to think carefully about what is being measured here. For example, the cost per case has risen by 80% (from \$12,000 to \$22,000) but outcomes have only increased by 20% (that is, survival has increased from 5 to 6 years). But there are two problems with this view. Firstly, the investment per case of \$12,000 was not in fact 'buying' the full five years survival, it was only buying a very small part of that survival outcome. And the extra year of life achieved by the end of the 15 year period is not all due to the interventions that took place in that 12 month period after the episode. Many other things will have contributed to increased life expectancy of heart attack patients. These include reductions in risk factors caused by smoking cessation, better management of high blood pressure, improved diet and more exercise. Assumptions have to be made about how much of that extra year is actually attributable to the technologies. Cutler and McClellan estimate that 70% of the improvement comes from use of these better technologies.

3 This figure is derived from decisions made by the National Institute for Clinical Excellence (NICE) for the NHS in England and Wales (see, for example, Towse and Pritchard, 2002).

Secondly, what we are really interested in is the extra cost and the value we place on the extra benefit. So what is really needed here is an incremental approach to assessing the change in health outcomes arising from changed patterns of investment. At the margin the extra expenditure represented good value for money, even if we assume that only 70% of the life year gained is attributable to the extra expenditure of \$10,000. It will increase the quality adjusted productivity of the US health care system if the marginal cost per life year gained is less than the average. Cutler *et al.* (2001) also present the results in the form of a fall in the quality adjusted price of treating heart attacks – i.e. at the disease level quality adjusted productivity is improving.

Depression

Berndt and colleagues in a series of articles (2000, 2001, 2002) used a claims database in the US covering 13,000 episodes of depression over the period 1991-6 to investigate the impact of three technologies: SSRI medicines, tricyclic medicines and therapy. The researchers tracked an overall decline in the price of treating depression but also noted that the number of diagnosed and treated patients doubled over the time period of their study.

Berndt *et al.* estimated treatment costs to be around \$1,000 per patient. They further estimated that the health benefit of treatment was, on average, eight weeks of reduced depression, which they valued at 0.4 of a QALY if extended over a whole year. As with the estimate of the value of the benefits from treating heart attack patients, assuming one QALY is worth \$50,000, the value of 0.4 of a QALY is around \$3,000⁴. In essence, for an investment of \$1,000 \$3,000 of benefit were obtained. So, while total costs rose because more patients were treated for depression, according to these calculations extra value was obtained as a consequence.

Valuing quality improvements in the NHS

Given these examples, what are the implications for the NHS? In terms of the new PSA value for money target, it is possible to calculate

⁴ That is, (8 weeks/52 weeks) x (0.4 QALYx \$50,000) = \$3,077.

an approximate increase in benefits that the NHS needs to achieve over the next three years. If, for example, we take current total annual expenditure on the NHS in the UK as £60,000 million, then, in order to meet the PSA target of a 1% increase in quality, the NHS needs to produce extra benefits (that is, improved health outcomes) each year equivalent to a value of £600 million. In terms of the size of the health outcome this represents, measured, perhaps, in terms of QALYs, then, assuming the monetary value of a QALY is around £30,000, the NHS needs to produce an extra 20,000 QALYs per year in order to meet its PSA target.

There are, however, a number of issues about how we measure whether or not the quality improvement is being achieved. Some of these arose in the US examples above. At its simplest, the question is whether we should take an *ex ante* approach (the way NICE does for new technologies early in their life cycles), or an *ex post* approach as the US examples adopted? The *ex ante* approach essentially asks what is expected from a new technology in terms of costs and benefits to patients. If we follow this approach, we can multiply the measure of expected outcome by the expected quantity of use to obtain some measure of total benefits. The *ex post* approach relies on available data on the actual impact on costs and outcomes associated with a technology already in use.

A further issue is that even if outcomes can be identified and quantified, how confident can we be about how much can be attributed to the interventions that have taken place within the health service – that is, what exactly is the relationship between the financial inputs to the service and eventual health outcomes?

Finally, there is a question of how we are going to measure treatment cost. For example, should we use health care resource groups (HRGs – the UK equivalent of diagnostic resource groups)? Do such process/output measures actually pick up the patient pathway in an accurate enough way? This brings us back to something that is implicit in the Wanless approach, which is the need to move towards a disease-based approach, where costs and benefits, and the changes in both, are identified on a disease basis.

Issues for modelling NHS expenditures and productivity

We return now to our original question: is productivity cost-reducing or quality-enhancing? The answer is that it is both, but the problem is that we do not know what the split is. The US evidence suggests variation across different interventions: in one case it was cost-increasing and quality-increasing; in another it was cost neutral and quality-increasing. The evidence (such as it is) suggests there is no reason to suppose, as the Wanless review did, that there is an even split between the cost-reducing and quality-enhancing aspects of productivity increases.

Three modelling issues are raised by our analysis of the Wanless report. The first research issue for any future expenditure on a modelling exercise has to be the reasonableness of the 2 to 3% expenditure growth per annum that the Wanless review assumed is attributable to new technology. If the NHS is essentially a technology taker, i.e. few technologies will be developed just for the NHS but rather will arise on a global basis primarily driven by the opportunities in the US marketplace, then US evidence suggests that most of the extra cost associated with new technologies will not be linked to their price, but rather the new opportunities to treat people – which will lead to an expansion in the volume of care delivered by the NHS. Quantifying this relationship between the development of new technologies and additional NHS costs is difficult, however.

Part of the answer is in understanding how the NHS takes decisions as to whether or not to use new technologies and how quickly they diffuse through the system. Unfortunately, there is a lack of understanding in both areas. In the case of NICE, for example, when a technology passes the appropriate cost-benefit threshold, does this effectively turn on the technology tap? We know that so far, in terms of NICE guidance, this has not been the case. The question remains, therefore, about the way and the speed with which new technologies are taken up by the NHS.

The second and third issues relate to how to measure quality adjusted productivity.

A particular problem in making this adjustment is that we lack detail on the health outcomes associated with NHS spending on a disease by disease basis. There is a need for the NHS to invest in research to identify this information. It may be that routine data collection on health outcomes will need to be carried out to identify the changes in health status that NHS treatment is delivering (alongside data on the costs of treatment that is, in principle, already being collected).

The third research issue is that improved quality cannot simply be measured in terms of improved health outcomes. As we noted above, much investment in 'access' and in the environment is designed to improve patient experiences, with very limited immediate impact on their health. Other methods will have to be used to measure the value of these expenditures to patients.

REFERENCES

Berndt ER, Bir A, Busch SH, Frank RG and Normand S-LT (2002) 'The medical treatment of depression, 1991-1996: productive inefficiency, expected outcome variations, and price indexes. *Journal of Health Economics*, Vol.21, No.3, pp373-396.

Berndt E, Busch S and Frank R (2001) 'Treatment price indexes for acute phase major depression', in Cutler D and Berndt E, eds., *Medical Output and Productivity*. University of Chicago Press for the National Bureau of Economic Research: Chicago, pp463-505.

Berndt E, Cutler D, Frank R, Griliches Z and Triplett J (2000) 'Medical care prices and output', in Culyer A and Newhouse J, eds. *Handbook of Health Economics Volume 1*. Elsevier Science.

Cutler D and McClellan M (2001) 'Is technological change in medicine worth it? *Health Affairs*, 20; 5: pp11-29.

Cutler D, McClellan M, Newhouse J and Remler D. (2001) 'Pricing heart attack treatments', in Cutler D and Berndt E, eds., *Medical Output and Productivity*. University of Chicago Press for the National Bureau of Economic Research: Chicago, pp305-362.

Dawson D and Jacobs R (2001) *Hospital Efficiency Targets: a Report for the Department of Health*. University of York.

Towse A and Pritchard C (2002) 'Does NICE have a threshold? An external view', in Towse A, Pritchard C and Devlin N (2003) *Cost-effectiveness Thresholds: Economic and Ethical Issues*, Office of Health Economics and Kings Fund: London.

Wanless D (2002) *Securing our Future Health: Taking a Long-term View*. HM Treasury: London.

Chapter 3

Discussion: modelling health care productivity

Topics covered in the discussion on modelling health care productivity included:

- the value of macro data in productivity studies;
- the need for health outcome data, not activity rates, when discussing productivity;
- the type of target most relevant to planning and forecasting expenditure;
- the problems of existing disease level studies of productivity.

Alan Williams: It seems to me that the great weakness that is emerging in the productivity results is that they are not about improvements in health. They are about volume of health care activity. They are based on the implicit assumption that the more activity there is and the cheaper it is the better. I cannot accept either of those propositions, because all the micro studies that have been done cast grave doubt upon those two assumptions.

If we are going to move forward we have to find some way of bringing all that we know from detailed microeconomic studies to bear upon likely future trends in NHS costs. That requires us to know what the cost per QALY threshold is going to be for the NHS, because that will determine the diffusion of technology. Starting from national accounts and adopting a broad macro approach seems a fruitless way of proceeding.

Mary O'Mahony: We do need to have measures that will monitor improvements in the health of people. The national accounts data at present do not incorporate this, but I would not dismiss the national accounts entirely. They do tell us something about how much efficiency has been increasing in terms of cost reduction and in terms of the inputs that are used. But they are telling just one aspect of the story, so, in that respect are deficient.

Martin Weale: The whole national accounting structure and valuation at market prices is, as Mary O'Mahony mentioned, based on the premise that markets are efficient; that people want what they are buying rather than something else. However, in the health service, where the good is not traded as in a market, bodies like NICE are needed to ensure that resources are allocated efficiently. In those circumstances, productivity studies can do a very valuable job in telling us about whether the production of the services that are provided is becoming more efficient or less efficient.

Diane Dawson: We have two different issues here. We can debate whether and how the health care sector should be included in the national accounts, and if it is, whether we should attempt to put values on output or not. I have the impression that what is happening in the UK is that we are trying to improve the reporting of the health sector and other non-market sectors in the national accounts so that we have a better *ex post* view of what is happening to the economy as a whole. I do not think that Alan William's points are really criticisms of the way in which we should be trying to improve the national accounts. What he is after is a different approach to planning public expenditure, the resources that should go into the National Health Service as opposed to other uses of taxpayers' money.

Martin Weale: I think what has been highlighted with regard to the NHS value for money target is a problem that arises with all quantitative targets without distinction. Under the former Soviet Union's chief planning mechanism, Gosplan, for example, there was a target concerning the number of miles that the railways had to run – and running empty trains was easier than running full trains! I would not go quite from there to say one should not have targets because some may be of help in a situation where we do not have market disciplines. My guess is that under this Government we have got far too many quantitative targets, however, and not enough arrangements for local flexibility allowing achievement on one target to be traded off against underachievement on another.

Mary O'Mahony: There are many different productivity measures available. I think all of them are useful, because all measures answer

different questions. The national accounts based measure is essentially designed to look at the question of how we improve our understanding of what is going on in the aggregate economy. But if you want to look at health performance, it is necessary to go beyond the national accounts. A commonly employed measure is TFP adjusted for changes in the quality of inputs, but it is also interesting to look at labour productivity or productivity per quantity of some other input. Each time you get more information about what is happening in the system. The ONS has done a very good job in terms of putting output measures into the national accounts, but if we want some more on performance in health care then the work has to be done by independent researchers rather than by national accounts statisticians.

Nancy Devlin: It is worth thinking about what the role of health is in the Wanless report. Health is seemingly considered to be exogenous; its role in the Wanless model is simply to determine health care requirements, or the demand for health care. Everything else in the model is just about estimating what the population's health care requirements are and how much service we need in order to meet those requirements. So there is nothing in the approach that really allows health care to determine health. There may be some reasons for being sceptical about the amount of health that one gets from spending more on health care, but the point is the model does not even allow us to get at those sorts of trade-offs.

Adrian Towse: I would like to make three points. The first concerns the Department of Health's public service agreement (PSA). It seems to me incredibly important that there should be a quality-enhancing aspect to the PSA value for money target. There are enormous implications requiring the NHS to identify £600 million worth of quality-enhancing service provision.

Secondly, Diane Dawson's and my analysis is saying that the cost-reducing assumptions of productivity improvement in the Wanless report look, by historic trends, to be too high. The implication of that is quite significant. Remember, variations in the productivity assumptions gives rise to a variation in projected NHS spend of two percentage points of GDP (equivalent to around £20 billion at today's

prices on total GDP of around £1,000 billion). If the wrong productivity assumptions are in the model in terms of cost reduction then NHS expenditure under the Wanless projections is going to be 2% of GDP lower than it needs to be to achieve what Wanless wants the NHS to achieve, which is a huge amount of money.

Finally, in terms of the service projections, how we cost National Service Frameworks, is also very important. I think the key message is that we need disease-based costing and disease-based outcome measurements. That has to be the fundamental building block for any future modelling in expenditure exercise.

Mark Freeland: This work on productivity is very important. We are extremely interested in it in the US too. For example, as Mary O'Mahony pointed out, Jack Triplett produced what seemed to be a good analogy between automobile repair and repair of humans with health problems. He hypothesised that if health outcomes were properly measured, and real outputs were measured properly, positive productivity would be evident in the health industry just as he assumed it must be for automobile repair. Zvi Griliches, a productivity expert from Harvard University, who was a discussant of Triplett's paper at a National Bureau for Economic Research meeting, got unpublished data on productivity from the US GDP accounts for both automobile repair and health care. To Triplett's (and the audience's) surprise, productivity changes for auto repair and health care were almost identical, but significantly negative for both industries. That, of course, does not mean that either of these industries had correct measures for inputs and outputs. But it also gives pause for consideration on what the true rate of increase is for health productivity.

David Cutler, Ernst Berndt, Jack Triplett and Joe Newhouse, very prominent economists in the US, are making very significant contributions on the conceptual and empirical basis for utility-based price indexes for the health care sector⁵. These utility-based price indexes are very important for cost-benefit analysis of the value of

5 See for example, David M Cutler and Ernst R. Berndt, *Medical Care Output and Productivity*, University of Chicago Press, 2001.

health spending relative to other priorities in government budgets. Utility-based price indexes should not be confused with production-based price indexes used in GDP accounts all over the world to account for price and productivity changes. Some caveats of the utility-based health care price indexes are rarely made in the presentations and articles, however. The distinction between utility-based price indexes (reflecting improved quality of life and length of life for example) and production-based price indexes (reflecting transaction market prices adjusted for output characteristics) is typically not mentioned.

This group of prominent economists has tended to focus on types of medical care treatments (heart attacks, depression, cataract surgery etc), which have achieved the biggest economies of scale and cost-reducing technological advances over the period examined. A couple of years ago at the American Economic Association, Cutler gave a paper on breast cancer and the benefit-cost ratios were less than the earlier group of selectively chosen highly productive treatments. There is great value in what they are doing, but there are many complex issues which have not been given proper caveats and attention thus far.

Mary O'Mahony: I agree that there are many problems with the US studies. Everybody who reads them can straight away point out that they do not adequately account for exogenous lifestyle factors. But, in their defence, it is necessary to start somewhere. I think the contribution of these studies is that they begin the research process. Subsequently, others in the field refine the measures, deal with the problems and so ultimately produce better measures. I think that has been the main contribution of these kind of studies.

Stephane Jacobzone: I am close to the US approach since we have carried out collaborative work with Mark McClellan and David Cutler and colleagues over the past couple of years. I think, as Mary O'Mahony pointed out, the contribution of the US approach can only remain limited in the field of health in the national accounts because of the international standardisation of methodology. To my

knowledge, it is only in some cases where hedonic prices⁶ can be used, for example for pharmaceuticals, in national accounts. Productivity improvements can also be assessed through proxies such as rates of intervention, which would give an idea of the level and intensity of medical activity in various countries, even if at that level, no judgement can be made in terms of appropriateness.

What is interesting when you look at a disease like stroke is that pure activity rates, using technology such as MRI investigations, may not appear effective in generating productivity in terms of QALYS as these are purely diagnostic. By contrast some of the Scandinavian approaches treating stroke with fully co-ordinated teams of professionals appear very effective. With early discharge and early rehabilitation you can significantly reduce the disability rates of patients in the long term. It seems to me that further emphasis on this type of evidence in determining disease-based indicators can be very fruitful. They also need to be linked with some outcome data, and some of our work, using some of the Oxford linkage study data for the UK provided to us by Michael Goldacre, shows that we usually find a curve with decreasing marginal returns in technology used across countries. For heart attack a reduction in mortality occurs, but the more the technology is being used, the less the reduction in mortality that is being observed at the margin. There is a social choice to be made, which is how much of a marginal reduction in heart attacks a society is ready to buy, given the opportunity cost of public funds.

6 The basic premise of the hedonic pricing method is that the price of a marketed good is related to its characteristics, or the services it provides. For example, the price of a car reflects the characteristics of that car – transportation, comfort, style, luxury, fuel economy, etc. Therefore, we can value the individual characteristics of a car or other good by looking at how the price people are willing to pay for it changes when the characteristics change.

Section 2

MODELLING THE DEMAND FOR HEALTH CARE

Chapter 4

UK evidence on ‘healthy lifestyles’ and their effect on the demand for health care

KLIM McPHERSON

This chapter looks at:

- healthy lifestyles and their potential to reduce demand for health care;
- impediments to effective public health;
- future directions for public health in the UK.

Introduction

Healthy lifestyles can have a big effect on the demand for health care and consequently health care expenditure (Naidoo *et al*, 2000). Indeed, improved lifestyles are already having a big effect – although I would contend that this is in spite of the established public health infrastructure rather than because of it. Notwithstanding the influence of evidence from McKeown (1979) and others, public health in the UK is currently too close to health care, too far from health and often too late.

Enabling further behaviour change involves full engagement, political commitment, an effective workforce and, most importantly, enabling *real* choices in a complex world. The spending projections contained in the Wanless Report (2002) are based on the fundamental assumption that if we make a real commitment to public health we could reduce the demand for health care substantially. The question is: by how much and what do we need to do differently in public health to achieve these outcomes?

An interim draft of the Wanless report received a great deal of feedback from the public health community. In essence they expressed concern that, in focusing on health care, the review might be missing the point. Health is more than health care: it is about prevention and public health too. Further, if public health were properly organized we would not actually need to spend that much

more on health care. The influence of these views is reflected in the final report's approach to modelling health care spending under a range of scenarios, concerning, among other things, health behaviours with a fully engaged population.

Impediments to effective public health

The sort of commitment to public health envisaged by Wanless faces a number of impediments. The Wanless review's 'fully engaged' scenario for health sets a clear aim of a dramatic improvement in the public's engagement, driven by widespread access to information and to real choices. But public health strategies are prone to policy pendulum swings and often appear condescending. They are constrained by the resources available to the public health agenda and by the fact that the UK public health workforce is divided and weak. Further, there are strong anti-health vested interests that impede progress towards public health goals.

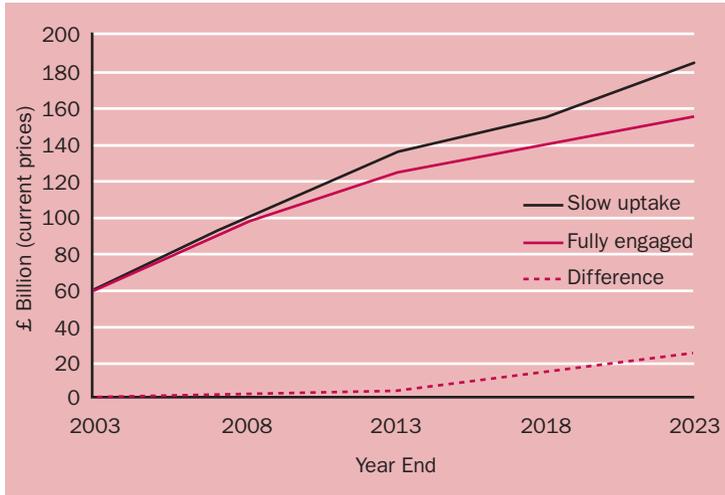
Demand-led health care leads to a disproportionate share of the health budget going on health care as a consequence of the immediate demands of ill people who need care. Short-term political expediency leads inexorably to the treatment of ill health continuing to dominate health resource allocation.

Finally, the evidence base for public health is somewhat problematic. In part, this is a consequence of the long term focus of public health and its outcomes being influenced by a wide range of interactions with society, politics, fashion and money (McPherson, 2001).

Fully engaged scenario

The fully engaged scenario has significance other than as a set of modelling assumptions underpinning the spending estimates. It reflects a commitment by the Treasury to a world-class health service (HM Treasury, 2002). The Chancellor of the Exchequer has subsequently referred to the fully engaged scenario as being a key policy instrument. And it has since received further examination in a subsequent report from Derek Wanless (Wanless, 2004).

Figure 4.1 Anticipated Annual Health Care Costs £ Billion



Source: Wanless, 2002.

The essential components of the fully engaged scenario are: people’s engagement in relation to their health should be high; life expectancy should increase; health status should improve; people should be confident in the health system and demand quality care. On the supply side, the health service should be responsive to high rates of technology uptake and in particular to disease prevention.

Projected NHS expenditure in the Wanless review suggests (on his modelling) a £30bn saving (at current prices) by 2020 as a consequence of the implementation of the fully engaged scenario, relative to the ‘slow uptake’ scenario (see Figure 4.1).

Where do we go from here?

The vital, and unconfirmed, postulate underpinning the fully engaged scenario is that we can have more health and thus less health care. This is not what people tend to believe, particularly Secretaries of State for Health, who tend only to think in terms of more, quicker and more

efficient health care as the dominant means to achieve more health. The Wanless Report thus usefully highlighted the potential role that public health can play in reducing demand for care. We need sophisticated epidemiological models to predict the long-term impact of improving public health, which we do not have, and further evidence is required to fully understand the relationships between health and inequalities.

The Wanless Report sets out a powerful case for investing in public health, and the prevention of disease, as a means to improve outcomes and health services and to yield savings to the NHS and social care services. Such a step change in the investment and delivery of public health needs to be accompanied by efficient implementation and reliable analysis of the consequences. We need to examine complex issues of public health supply in relation to changes in NHS demand; no longer assuming that demand is constant, not amenable to change, or is changing according to unknown or uncontrollable influences.

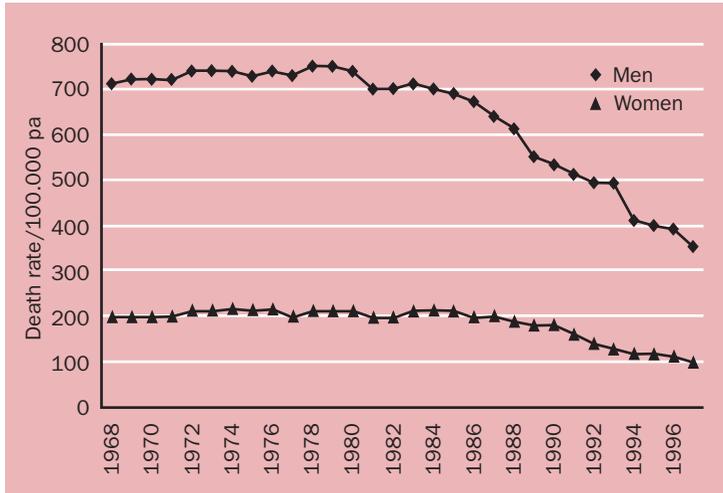
The potential of 'proper public health'

The Wanless Report contains a number of references to the potential of 'proper public health'. The second Wanless report, concentrating on public health (Wanless, 2004), examined these issues in detail in a UK context.

Coronary Heart Disease (CHD) provides an excellent example of this potential. CHD is all but avoidable for those under the age of 70. If we were able to eliminate premature CHD – as we were able to eliminate smallpox – this would be as a consequence of life course choices and changes among populations starting very young indeed. It might also involve drugs or functional foods. But these changes largely involve diet, obesity, tobacco and exercise. Making these changes means providing people with real and popular choices and opportunities in their own context. To do so requires a very wide, but possibly highly cost effective, policy agenda.

Research provides evidence on the factors that can explain the incidence of, and reductions in, CHD. Figure 5.2 shows the reduction in CHD deaths since 1968 for the 55 to 64 age group in England and Wales.

Figure 4.2 CHD Deaths Aged 55-64, England and Wales



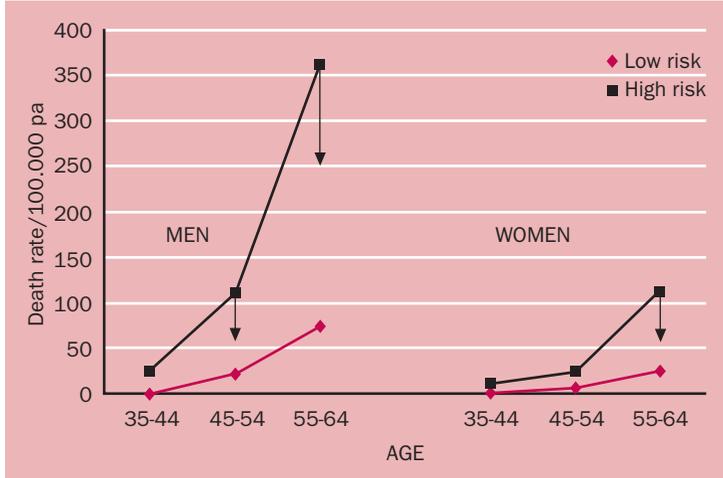
Source: British Heart Foundation, 2003.

The reduction in CHD deaths is quite dramatic – and is not at all due to improved treatment. It is not even mostly due to treatment (Capewell et al, 1999). It is largely attributable to changes in lifestyles that populations in this country have somewhat belatedly come to grips with and been able to enact so that at the ages of 55 to 64 they succumb less to CHD (British Heart Foundation, 2003). Note that Figure 4.2 shows only the effect of this on *deaths*. Not everybody who has a coronary event dies, but about half do. Therefore the incidence of people getting a coronary event is about twice the number in that age group, and is also reducing dramatically.

Figure 4.3 shows current CHD mortality rates by age and gender, and points to potential decreases if achievable changes in diet, exercise and tobacco did occur (Stamler et al, 1999). Things are already happening which are changing the number of people dying from CHD.

Figure 4.4 shows standardised death rates from CHD up to age 64 by social class. It demonstrates that there are important social class

Figure 4.3 CHD Mortality Rates and the Potential for Change



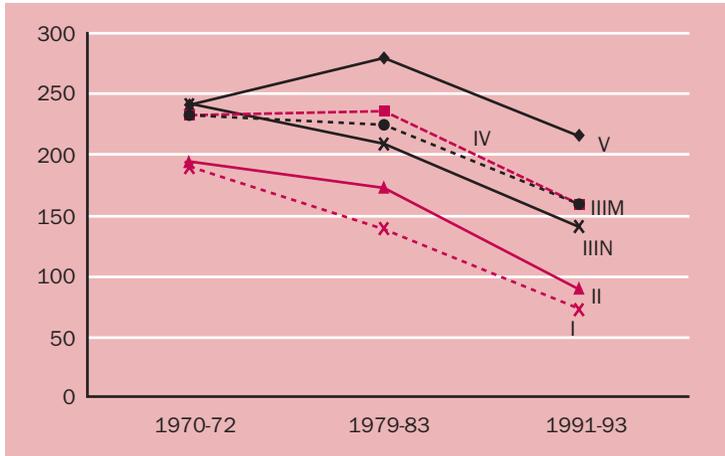
Source: Stamler et al, 1999.

differences: deaths from CHD under age 65 in social class 1 men are currently quite rare. They have information on how to prevent CHD and have been able to act on it. Arguably, everyone knows how to avoid CHD. It is not that difficult, but takes a long time to implement for an individual. It is a matter of changing aspects of our exposure that enable us to live without experiencing cardiac illness. That requires cheaper healthy food options, appropriate exercising facilities and smoke free environments.

Estimating the changes in risk factors

Estimates of the effect on CHD risk of plausible changes in risk factors have recently been published (McPherson et al, 2002). Essentially we know what causes CHD quite well (see Figure 4.4). Premature death from CHD is largely determined by cholesterol, physical inactivity, smoking and blood pressure. All these things are amenable to change without a great deal of pain in the longer term, given proper options in individual contexts.

Figure 4.4 **Standardised (to European population) Death Rates from Coronary Heart Disease, Men Aged 20-64, by Social Class, at Three Time Periods, 1970-1993, England and Wales**



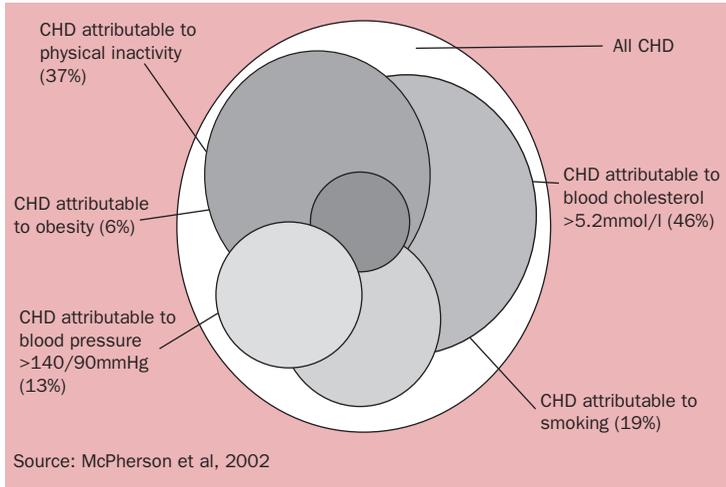
Source: Drever and Whitehead 1997.

How much change is possible in practice depends on there being real choices for people with regard to their health. School children are not given real choices: they are barraged with advertisements on the television or at school telling them they should be eating high fat and high salt products (Hastings *et al*, 2003). Local governments sell off school playing fields, so people's exercise opportunities are limited. The marketing of cigarettes makes it quite difficult for teenagers and young children to quit by 25, which they would need to do if they did not want to get a coronary (Doll *et al*, 1994). These are important issues concerning whether or not there are real choices for those people. If there were a commitment to enable real choices for those people over a life course, it could have dramatic effects on the number of CHD episodes. Quick fixes are usually not appropriate.

The demand for health care

A sustained focus on public health, which provides real choices for

Figure 4.5 Risk Factors for CHD

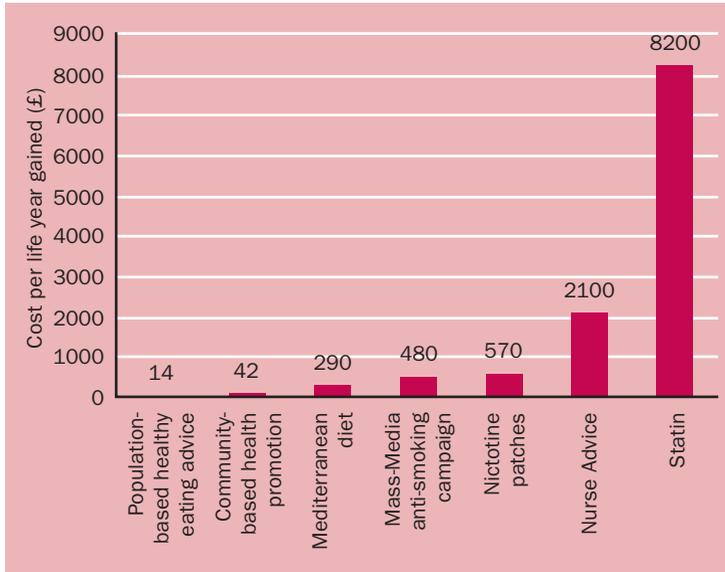


Source: McPherson et al, 2002.

people, could be expected to have a substantial impact on the demand for health care. For example, by giving people choices concerning aspects of their lifestyle relevant to CHD, the number of coronary events in those under 75 could be markedly reduced. That would translate unequivocally into a massive reduction in demand for NHS treatment. People would not be lined up to be treated in coronary care units if they were able to have lifestyles throughout their entire lives which essentially eliminated the possibility of having a coronary before aged 75. This has already almost occurred for men in social class 1. The challenge is to put policies in place that effectively makes these changes possible across all social classes. This requires a policy commitment and coordination that has hitherto been lacking.

We have already seen that changing lifestyle factors can have an enormous effect on the incidence of disease and premature death. Policies and interventions that change these behavioral factors are also likely to be cost-effective.

Figure 4.6 **Estimated Cost Effectiveness of Interventions to Prevent CHD**



Source: Brunner *et al*, 2001.

One way to stop people having coronaries is to give everyone some statins at age 40, because this will certainly lower their cholesterol level and reduce their CHD risk. However, while this might be cost-effective for those people who have particularly high cholesterol, or have particularly high risk, in the general population it is not a cost-effective intervention. It may even be a silly intervention since the consequence of effective LDL – C lowering drugs on people at moderate risk may be to encourage a less healthy lifestyle – the net effect of which may be to increase risk at considerable cost. There are many other much more cost-effective interventions which could give rise to those kind of changes (Brunner *et al*, 2001) (see Figure 4.6).

Accreditation and the public health infrastructure

In my opinion we have had a sub-optimal public health infrastructure

in this country. People who practise public health have worked from different silos and had little contact with one another. Each saw themselves as having different responsibilities and different effects.

There was an unjustified view that the people in charge of public health have to be doctors. To make real progress, competent public health enthusiasts from any core discipline must take the lead. But accreditation remains a real problem, because until very recently it was impossible for anybody who was not a doctor to become accredited in the specialist practice of public health. That is only now changing but is essential for a meritocratic profession with real responsibility for the health of the public.

There is no centre of excellence for practitioners in public health. The interaction between academia and practice too often has to go through a single conduit: public health medicine. Unfortunately, because public health doctors may not understand much about important parts of public health, other conduits between academic public health and practice are required, which we are just beginning to establish.

The future of public health inside and outside the NHS

If the Wanless review's vision of a 'fully engaged' population is to be achieved, there remains much to do. There is a good case to further review public health demands and investment, to guide the Department of Health. The current emphasis in public health is to partition itself into three domains; health protection, health improvement and health services. The new Health Protection Agency provides new opportunities to get at least one domain of public health right with respect to the crucial balance between responsibility, scientific endeavour and public engagement with health protection. How much longer might it be before the much greater predictable disease burden amenable to diligent and sensible health improvement is provided with the same opportunities? They will be different and more complex since they will require the more active participation of communities and more diverse policy instruments.

This could all assist the delivery of the NHS Plan and other current policy documents to which the Government is fully committed. In

particular, there is a lot of work to be done in modelling the impact of public health interventions outside the NHS. Arguably, the benefits from a greater focus on public health would be even greater were there broader investment across both national and local government.

REFERENCES

British Heart Foundation. Editors: Petersen S, Peto V and Rayner M (2003) *Coronary Heart Disease Statistics*. British Heart Foundation: London.

Brunner EJ, Cohen D and Toon L (2001) 'Cost-effectiveness of CVD prevention strategies: a perspective on EU food based dietary guidelines'. *Public Health Nutrition*, Vol. 4, Special Issue 2B, pp711-715.

Capewell S, Morrison CE and McMurray JJ (1999) 'Contribution of modern cardiovascular treatment and risk factor changes to the decline in coronary heart disease mortality in Scotland between 1975 and 1994'. *Heart*; 81:pp360-386.

Wanless D (2002) *Securing our Future Health: Taking a Long-Term View. Final Report*. HM Treasury: London.

Wanless D (2004) *Securing Good Health for the Whole Population. Final Report*. HM Treasury: London.

Doll R, Peto R and Wheatley K et al. (1994) 'Mortality in relation to smoking: 40 years' observation on male British doctors'. *BMJ*; 3309:pp 901-911.

Drever F and Whitehead M (1997) *Health Inequalities: Decennial Supplement*. The Stationery Office: London.

Hastings G et al. (2003) *Review of Research on the Effects of Food Promotion to Children*. Final report prepared for the Food Standards Agency, University of Strathclyde: Glasgow.

HM Treasury (2002) *Opportunity and Security for All: Investing in an Enterprising, Fairer Britain. New Spending Plans 2003-2006*. HM Treasury: London

McKeown T (1979) *The Role of Medicine: Dream, Mirage, or Nemesis?* Second edition. Princeton University Press.

McPherson K (2001) 'Are disease prevention initiatives working? *The Lancet*; 357; 1790-1792.

McPherson K, Britton A and Causer L (2002) *Coronary Heart Disease; Estimating the Impact of Changes in Risk Factors*. National Heart Forum. The Stationery Office: London.

Naidoo B, Stevens W and McPherson K (2000) 'Modelling the short term consequences of smoking cessation in England on the hospitalisation rates for acute myocardial infarction and stroke'. *Tobacco Control*; 9:pp387-400.

Stamler J et al. (1999) 'Low risk factor profile and long-term cardiovascular and noncardiovascular mortality and life expectancy: findings for 5 large cohorts of young adult and middle-aged men and women' *JAMA*; 282; pp 2012-2018.

Chapter 5

Age, proximity to death and future demands on the NHS

MEENA SESHAMANI and ALASTAIR GRAY⁷

This chapter looks at:

- the relationship between age and health care costs;
- problems with the ‘traditional’ view of ageing and health care costs;
- evidence to support the ‘compression of morbidity’ effect assumed in the Wanless report’s projections of NHS spending over the next 20 years.

Introduction

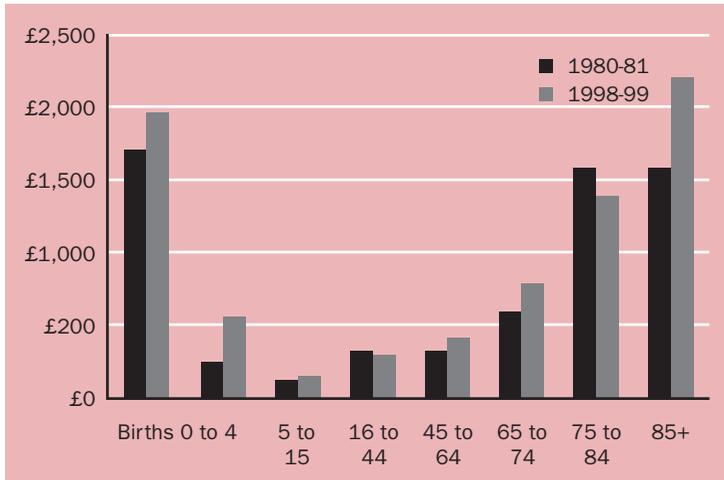
It has long been evident that there is a relationship between age and health care expenditure. Health spending by age follows a u-shaped curve, whereby the young and old have higher per capita costs than the population aged 5-64 years. There is evidence in several countries that this pattern of higher costs in the older ages has become more pronounced over time. Consequently, there are concerns about the impact that an ageing population might have on future health care spending, in terms of increased numbers but also increased intensity of health care use (Culyer, 1988).

Figure 5.1 shows NHS hospital and community health service (HCHS) expenditure (i.e., not all NHS expenditure) per person by age group in 1980-81 and 1998-99.

The data in Figure 5.1 demonstrate the u-shaped relationship with respect to age and provide some evidence of it deepening over time in the UK – although this is not as pronounced as many people think (Seshamani and Gray, 2002).

⁷ Paper was presented by Alastair Gray

Figure 5.1 **HCHS Expenditure per Person by Age Group in 1980-81 and 1998-99**



Source: Seshamani and Gray, 2002.

How have these data traditionally been used to analyse the impact of demographic change on health expenditure? Examples are provided by the Department of Health's annual expenditure plan publications from the mid-1990s. Data on age-related health expenditure, taken together with projections of demographic change, were run forward by multiplying projected population figures by constant age-specific per capita costs, and converted into an index. This gave an estimate of the impact of future demographic change on health care expenditure, purely as a result of changes in the age composition of the population.

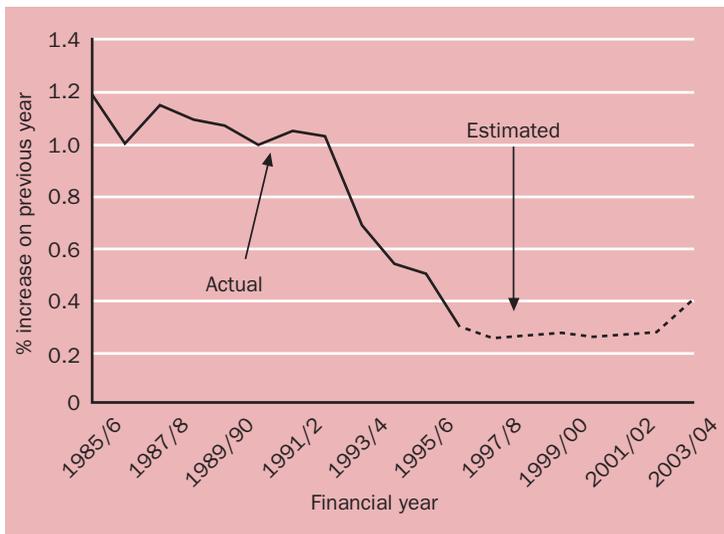
In the late 1980s and early 1990s, the annual impact was estimated to be in the order of 1% per annum. Estimates varied considerably, due to different sizes of age cohorts moving through the population. Such results were generally interpreted to mean that the age structure of the population is changing and, therefore, that even if we just continue to do exactly what we are doing at present in terms of health care, we will need to increase health expenditure by 1% per annum in real terms to stand still.

Problems with the ‘traditional’ approach to the effect of age on health care cost

Questions began to arise about this traditional approach for three reasons. First, research by the OECD and others suggested that, looking retrospectively at the experience of a number of countries over 10-20 years, demographic change seemed to be a minor factor in explaining past health expenditure increases (OECD, 1988; Barer *et al.*, 1989; Fuchs, 1998; Gerdtham, 1993). The implication was that this was probably also likely to be the case in the future (see Figure 5.2). Intensity of health service use and technical change seemed to be much more important influences on health care spending trends than ageing.

Second, a debate opened in the epidemiological literature from the 1980s onwards on the centrality of morbidity within the process of

Figure 5.2 **Traditional Analysis: Age-Related Expenditure + Demographic Change = Increased Expenditure**



Source: Department of Health, 1998.

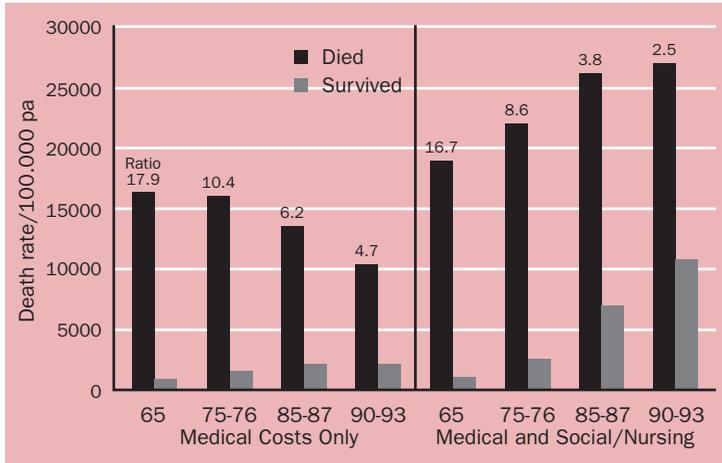
ageing. The compression of morbidity hypothesis (Fries, 1980) proposed that as people live longer over time, the onset of disability is postponed, such that morbidity gets pushed to the end of life. Empirical studies from the US, France, and other countries corroborate this hypothesis, indicating that increases in life expectancy from the mid-1980s to the mid-1990s have been accompanied by increases in healthy life expectancy, and that the prevalence of disability may be decreasing over time in older age groups (Fries, 2000). These developments emphasised that one could not simply look at age when examining the impact of ageing on health care use and costs; one had to consider morbidity as well.

Associated with the epidemiological literature on the postponement of morbidity towards the end of life is a series of recent studies from Canada, the US and elsewhere that have each found health care costs to be concentrated at the end of life (McGrail *et al.*, 2000; Roos *et al.*, 1987; Lubitz and Prihoda, 1984; Busse *et al.*, 2002; Stooker *et al.*, 2001). Studies of US Medicare have consistently estimated that the 6% of Medicare recipients who die in a given year are responsible for about 30% of all the Medicare expenditures in that year (Lubitz and Prihoda, 1984). Depending on age group and whether nursing costs are considered in analyses, patients in their last year of life can cost up to 18 times more than patients who are not in the last year of their life (see Figure 5.3) (McGrail *et al.*, 2000).

This suggests that the traditional approach of applying age-related expenditures to changing population figures may be a misspecification of the projection method. Health expenditures do increase with age, but the probability of dying also increases with age. Since health expenditures are heavily concentrated near the end of life, increased health care costs in the older ages may be due to increasing proximity to death. If this is the case, then proximity to death needs to be incorporated into improved expenditure projection models.

In a seminal piece of research, Zweifel *et al* (1999) put forth a robust argument that age might indeed be a 'red herring' for the underlying association between health care costs and time to death (Zweifel *et al*,

Figure 5.3 **Costs in Last Year of Life among the Aged: Comparison of Costs between Survivors and Decedents**



Source: Adapted from McGrail et al, 2000.

1999). Using Swiss health insurance data at the individual patient level over the two years prior to death, quarterly health care expenditure was modelled as a function of age and of the number of quarters remaining until death. Results suggested that the quarter from death was strongly significant as an explanatory variable, but that patients' age was not significant, in explaining health care costs.

This research did, however, have some limitations. First, the data only covered two, or in some cases up to five, years from death. It is quite possible that there are age-related chronic diseases whose effects could extend much further back. The study also did not examine any interactions between the effects of age and time to death, whereby the trajectory of costs with approaching death could be different for someone in the youngest old versus the oldest old. There were also several econometric limitations to the analyses, many of which are beyond the scope of this paper (Salas and Raftery, 2001; Dow and Norton, 2002; Seshamani and Gray, 2004b), but in particular the quarterly hospital costs were treated as independent observations,

rather than as a set of associated observations from an individual patient over time. Finally, there is always a question of whether the results from a study conducted in a particular country with a particular data set can be generalised and replicated in other health care settings.

Results from the Oxford research programme

The research programme developed in Oxford began by attempting to replicate Zweifel's approach. Data were obtained from a longitudinal hospital dataset called the Oxford Record Linkage Study (ORLS). Unique in England, this dataset links all general and psychiatric hospital inpatient and day case episodes and death records for every individual within a geographical area from 1963 to the present. Cost data for the study were 1997-99 specialty-specific costs per day obtained from the Department of Health.

We identified a cohort of patients who were aged 65 or over in 1970, and followed them until death or until 1999. Doing so enabled the identification of a cohort which as far as possible would have died by the end of the observation period, mitigating the problems of 'right censoring' that have plagued other studies in this area.

The sample comprised about 96,000 individuals (see Table 5.1). Typically, each person was in hospital between two and three times from 1970 to death, although quite a significant minority of patients lived out their life span without being in hospital at all. Males typically spent about 55 days and females 84 days in hospital over the time period.

Using a two part model, the probability of being hospitalised at least once in any given quarter was first modelled as a function of the patients' age, sex, quarter from death, social class and cause of death. Then, conditional on being hospitalised, the health care costs incurred were modelled as a function of these same variables, plus diagnosis, source of admission, place of discharge and marital status.

Results were consistent with Zweifel et al (1999) (Seshamani and Gray, 2004). The probability of being hospitalised steadily increased with proximity to death, and increased dramatically in the quarter

Table 5.1 Oxford Study Descriptive Statistics, Mean (and Standard Deviation)

	Males	Females	Both
Number of patients	38,501	57,399	95,900
Age at death	81.1 (7.1)	84.7 (7.2)	83.3 (7.4)
Average number of hospitalisations from 1970 to death	2.5 (3.1)	2.6 (4.9)	2.6 (3.6)
Average days in hospital from 1970 to death	54.7 (162.0)	84.2 (250.9)	72.3 (220.1)
Average hospital costs from 1970 to death	£9,738 (£24,086)	£14,008 (£36,540)	£12,294 (£32,193)

immediately before death. Once admitted to hospital, costs increased gradually with proximity to death, but fell in the last quarter. This decrease is likely due to the curtailment of the final hospital stay by death. The effects of proximity to death on the probability of hospitalisation and on costs once admitted are summarised in Figure 5.4.

Combining the effects of proximity to death on the probability of hospitalisation and on costs once in hospital allows us to show the effect of proximity to death on expected costs (see Figure 5.5). From five years prior to the time of death, there are steadily increasing costs, with a very steep increase immediately before death resulting from the jump in likelihood of being hospitalised.

Turning now to the expected costs in the last quarter of life by patient age and sex (see Figure 5.6), it is evident that the extent to which cost changes with patient age, is much smaller than the quadrupling of costs that occur when approaching death in the last year of life. Additionally, once people get beyond a certain age, health care costs actually decrease, perhaps due to age rationing or a shift of care to long term care facilities (Seshamani and Gray, 2004a,b).

Figure 5.4 **Effect of Proximity to Death on Probability and Conditional Cost of Hospitalisation**

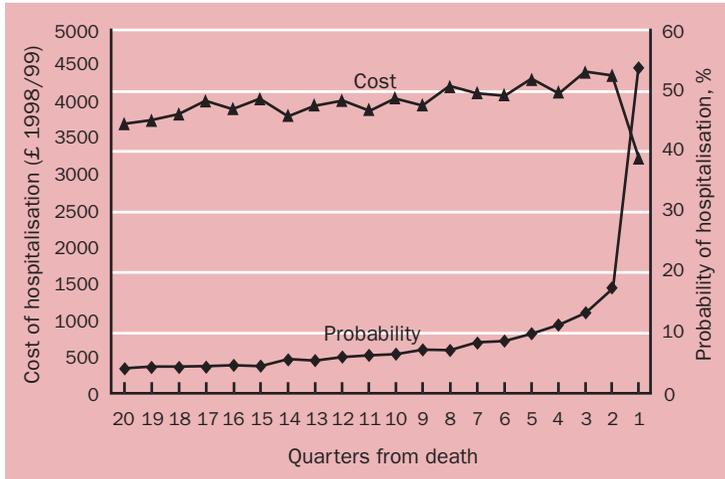
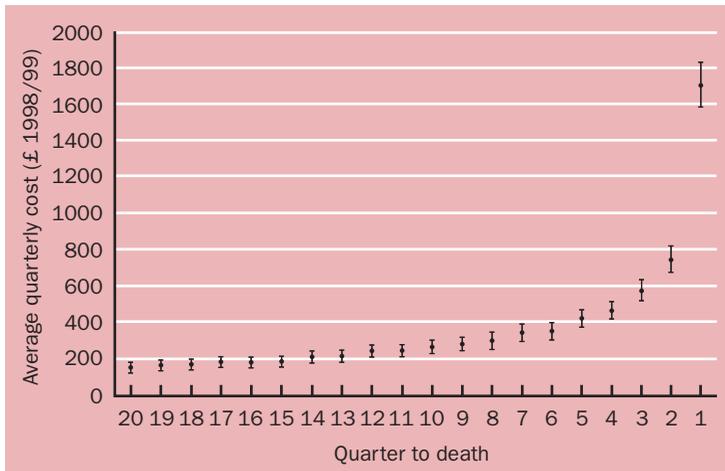
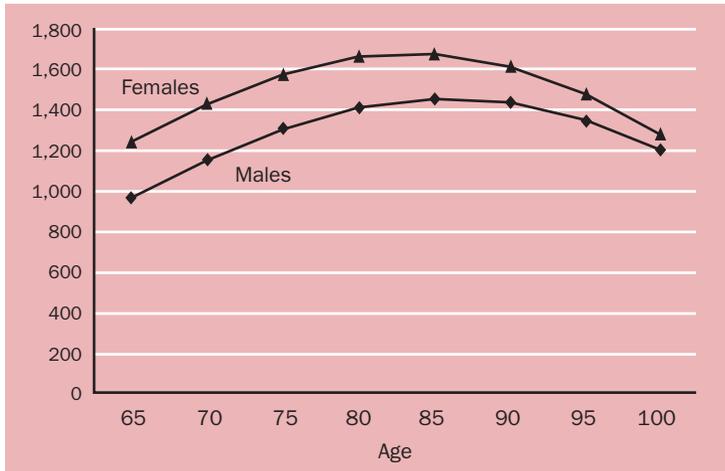


Figure 5.5 **Effect of Proximity to Death on Expected Costs**



Source: Seshamani and Gray, 2004b.

Figure 5.6 **Expected Cost by Age and Sex in the Last Quarter of Life (£ 1998/99)**

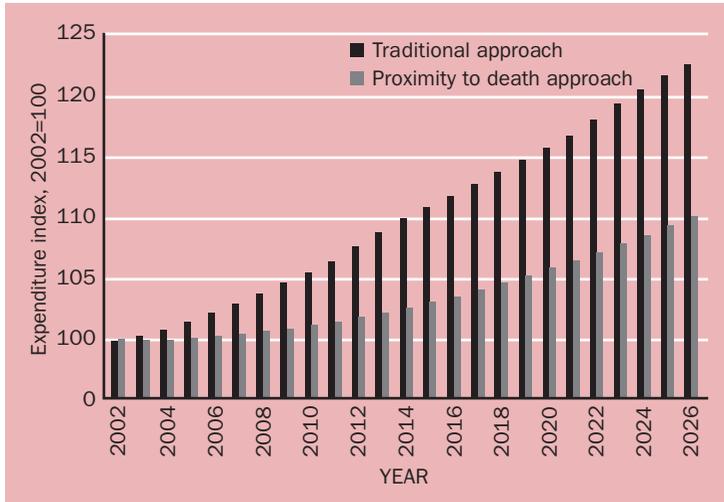


Source: Seshamani and Gray, 2004b.

Other demographic characteristics, such as not being married and being from a manual social class, were associated with higher costs, presumably because these patients are less likely to have the resources to facilitate discharge to home. Being transferred to the hospital from a long term care facility, and being transferred out to another hospital at the end of the hospital episode, were also associated with higher costs. Finally, heart disease as a cause of death was associated with a decreased likelihood of hospitalisation and decreased costs once in hospital compared to cancer or respiratory disease. This is intuitively plausible, as more people die immediately from heart disease than from other types of disease (Seshamani and Gray, 2004a).

Our subsequent research extended these results to explore the effects of proximity to death and age up to 15 years prior to death, also considering any interactions between these effects. The results were then applied to population projections for England up to 2026. We compared this updated expenditure model with results obtained from using a more traditional expenditure projection model (see Figure 5.7).

Figure 5.7 Projecting Impact of Demographic Change on Aggregate Hospital Costs: ‘Traditional’ versus Proximity to Death Models, 2002–2026



A striking finding is that if one were to simply take constant age-specific per capita health expenditures and extrapolate forward using changing population figures, health expenditures would have to go up by about 22% over this period for the health service to ‘stand still’, i.e. to accommodate demographic change alone. In contrast, the projection based on proximity to death more than halves this expenditure growth estimate. Taking proximity to death into account, the impact of ageing populations on health expenditure is much reduced.

This different result stems from the decrease in death rates across all age groups, and the subsequent decrease over time in the proportion of individuals in each age group who are in their last year of life (see Table 5.2). Since there is a tremendous concentration of health expenditure on individuals who are in their last year of life, a postponement of death pushes out associated costs, lessening the anticipated effect of population ageing.

Table 5.2 Decedents and Expenditure, 2002 and 2026

Age group	2002		2026	
	% of population in last year of life	% of expenditure on those in last year of life	% of population in last year of life	% of expenditure on those in last year of life
0-4	0.03	1.54	0.02	1.02
5-15	0.01	0.65	0.01	0.44
16-44	0.09	3.83	0.07	3.10
45-64	0.56	18.97	0.47	16.48
65-74	2.35	43.06	1.68	36.81
75-84	6.24	55.94	4.63	51.44
85+	15.90	64.63	13.47	63.04
All ages		28.88		27.98

An obvious question to ask is why so much money is spent on people in their last year of life? Two points should be noted. First, hindsight is 20/20; prospectively, it is difficult to predict when an individual will die. For instance, one recent study found physicians able to estimate accurately the remaining survival time of terminally ill patients upon hospice admission only 20% of the time (Christakis and Lamont, 2001). Second, we must recognise that the care being delivered in the last year of life may not be intended to improve life expectancy, but instead to offer significant palliation and pain relief.

What are the implications for Wanless?

The Wanless report was right to conclude that age is, in part, a red herring in relation to health expenditure. Proximity to death is the more important demand driver, and this is in line with epidemiological evidence on the compression of morbidity. Hence, inclusion of remaining life expectancy can significantly improve

expenditure projections. Extending this approach to other areas, it is also worth considering the allocation of resources to regional populations based on remaining life expectancy rather than age composition.

There are, however, several qualifications. First, all the work discussed in this chapter, and all of the work that underpinned the Wanless report, was based on hospital costs. This limitation may not be a major cause for concern as more than two thirds of NHS expenditures are hospital related (Seshamani and Gray, 2002). However, there may be different associations between age and proximity to death in primary care, and long term care.

Second, proximity to death is itself a proxy for underlying factors; in particular, disease progression. It is quite possible that if we had good measures of disease progression, proximity to death might become a less important benchmark for health service demand. Some evidence of that is available from a much smaller cohort study of patients with Alzheimer's Disease, where repeated measures of their cognitive ability and behavioural characteristics in activities of daily living were available (Wolstenholme et al, 2002). Once disease progression was taken into account, proximity to death became nonsignificant.

Finally, it is important to remember that neither age nor proximity to death are ultimately the dominant drivers of health expenditure. The key issues are intensity of use of health care inputs and the role of technological change. However, it is worth noting that technological change interacts with age. For example, improved surgical and anaesthetic techniques have allowed surgery to be undertaken safely at older ages, leading to increased technology diffusion in the older age groups, which may compound demographic effects (Dozet et al., 2002; Fuchs, 1999). Understanding the determinants and effects of all of these variables will become increasingly important to projecting future health care spending more accurately.

REFERENCES

Barer ML, Pulcins IR, Evans RG, Hertzman C, Lomas J and Anderson GM (1989) 'Trends in use of medical services by the elderly in British Columbia'. *Canadian Medical Association Journal*, 141, pp39-45.

Busse R, Krauth C and Schwartz FW (2002) 'Use of acute hospital beds does not increase as the population ages: results from a seven year cohort study in Germany'. *Journal of Epidemiology and Community Health* 56, pp289-293.

Christakis NA and Lamont EB (2001) 'Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study', *BMJ*, 320; pp469-473.

Culyer A J (1988) *Health Care Expenditures in Canada: Myth and Reality; Past and Future*. Canadian Tax Paper Series, No 82, 30-31, Canadian Tax Foundation: Toronto.

Department of Health (1998) *The Government's Expenditure Plans – 1998-1999*, Departmental Report. Cm 3912. The Stationery Office: London.

Dow WH and Norton EC (2002). The red herring that eats cake: Heckit versus two-part model redux. Triangle Health Economics Working Paper Series.

Dozet A, Lyttkens CH and Nystedt P (2002) 'Health care for the elderly: two cases of technology diffusion', *Social Science and Medicine* 54, pp49-64.

Fries JF (1980) 'Ageing, natural death, and the compression of morbidity'. *The New England Journal of Medicine*, 303, pp130-135.

Fries JF (2000) 'Compression of morbidity in the elderly'. *Vaccine* 18, pp1584-1589.

Fuchs V (1998) *Provide, Provide: The Economics of Ageing*. NBER Working Paper Series, No. 6642; National Bureau of Economic Research: Cambridge MA.

FuchsV (1999) 'Health care for the elderly: how much? Who will pay for it?' *Health Affairs*, 18, pp11-21.

Gerdtham U-G (1993) 'The impact of ageing on health care expenditure in Sweden'. *Health Policy* 24, pp1-8.

Lubitz J and Prihoda R (1984) 'The use and costs of medicare services in the last 2 years of life'. *Health Care Financing Review*, 5, pp117-131.

McGrail K, Green B, Barer ML, Evans RG, Hertzman C and Normand C (2000) 'Age, costs of acute and long-term care and proximity to death: evidence for 1987-88 and 1994-95 in British Columbia'. *Age and Ageing* 29, pp249-253.

OECD (1988) *Ageing Populations: the Social Policy Implications*. OECD: Paris.

Roos NP, Montgomery P and Roos LL (1987) 'Health care utilisation in the years prior to death'. *The Milbank Quarterly* 65, pp231-254.

Salas C and Raftery JP (2001) 'Econometric issues in testing the age neutrality of health care expenditure', *Health Economics Letters* 5, pp12-15.

Seshamani M and Gray A (2002) 'The impact of ageing on expenditures in the National Health Service'. *Age and Ageing*, 31, pp287-294.

Seshamani M and Gray A (2004a) 'A longitudinal study of the effects of age and time to death on hospital costs'. *Journal of Health Economics*, Vol 23; (2); pp217-235.

Seshamani M and Gray A (2004b) 'Ageing and health-care expenditure: the red herring argument revisited? *Health Economics*, Vol 13; pp303-314.

Stooker T, van Acht JW, van Barneveld EM, van Vliet RCJA, van Hout BA., Hessing DJ and Busschbach JJV (2001) 'Costs in the last year of life in the Netherlands'. *Inquiry*, 38, pp73-80.

Wolstenholme J, Fenn P, Gray A, Keene J, Jacoby R and Hope T (2002) 'Estimating the relationship between disease progression and cost of care in dementia', *British Journal of Psychiatry*, 181; pp36-42.

Zweifel P, Felder S, and Meiers M (1999) 'Ageing of population and health care expenditure: a red herring?' *Health Economics* 8, pp485-496.

Chapter 6

Discussion: modelling the demand for health care

Topics covered in the discussion of modelling demand for health care included:

- does preventing one disease and reducing associated health care costs lead to more deaths, and more health care costs, from other diseases?;
- who demands health care: patients or doctors?;
- how do incentives facing doctors affect overall demand and spending?;
- what is the relationship between need and demand?

Barry McCormick: Klim McPherson has argued that better preventive care will lead to less CHD and consequently less spending on the NHS. The problem with this kind of argument is that it does not include a counterfactual. If people do not die of CHD then they die of something else and the total demand for health services may in fact rise.

Alan Williams: I think when discussing demand, we need to be clearer about who is demanding what. A large tranche of medical sociology literature concerns what leads people to go to a GP in the first place. There is an extensive literature about lay referral systems: who people consult before they actually ever get into the health care system. Women and men employ very different lay referral systems. Arguably men ought to be going to the doctor more often and women perhaps less often. You need to understand something about the social setting within which that initial consultation takes place in order to manage demand for health care.

Actually, I do not think it is patients who are demanding health care: it is doctors who are in charge, and it is very difficult for patients to exercise autonomy. I wonder whether all the expenditures in the last year of life are demanded by patients or by doctors?

Hugh Gravelle: We know that health care professionals respond to incentives, including financial incentives. The experiences with prescribing incentive schemes and fund-holding schemes bear that out. Therefore, in attempting to predict future demand for, and spending on, health care, we need to take account of potential changes to the incentive system. There is a push to introduce budgets attached to real financial incentives for general practices in terms of their HCHS. We would expect that to do something to the use of health care services.

Stephane Jacobzone: I have some concerns about the optimism arising from evidence on compression of morbidity. There is, perhaps, a view that maybe we will not have to increase health expenditure by much despite an ageing population. For example, the Wanless report suggests that if you get both a healthy lifestyle and compression of morbidity, you will have to spend less. But when you look at detailed longitudinal data and the improvement and changes in health care, as Mark McClellan has done using the national long term care survey in the United States, you see that reductions in severe disability occur with increased rates of interventions and hence expenditure. The longer people live the greater will be the demand for, and spending on, these interventions. The UK u-curve has been fairly flat compared with other countries which have quicker uptake of new technologies.

Jon Sussex: The relationship between demand and need requires some thought. In the Wanless report there is, in the initial forecast years, a presumption that the gap between demand and need will close: we will do some catching up if the resources are made available. Some of the demand hitherto stifled by the lack of doctors is assured to be overcome in the next ten years or so. Thereafter the Wanless report projects a new kind of dynamic equilibrium where the rate of growth of health spending will slow down a bit, because we have got rid of that backlog. I wonder whether it has struck anyone that this same idea underpinned the foundation of the NHS: that is, the idea that you can cut through a backlog with an initial hump of expenditure and then get down to a more peaceful steady state. Do we believe that story?

Alan Williams: Put another way, the question is: will the health professionals get less ambitious?

Adrian Towse: The real problem is that we do not have a consensual model about what determines waiting lists. There is no model that explains the big stylized facts about waiting list differences between countries, across specialties and between different types of people. I think it is waiting lists that are primarily driving the increase in Government spend on health. That is the thing that most concerns patients, along with the quality of care that they feel they are given when they are treated. The lack of a consensual model means that until we know more about that process we really will not be able to answer whether it is a 'hump' in health care spending, or whether it is long term growth, that is required.

Jon Sussex: I would like to ask Alastair Gray and Klim McPherson if they have any thoughts, given the wide range of comments we have just heard. Alastair, are we all agreed that the relationship between age and spend is as you described it? And Klim, you argue that by modifying health behaviours, we can reduce treatment needs and therefore the demand for health care, at least in some disease areas. Is this too optimistic a picture?

Alastair Gray: The question about what are we actually doing when we are spending money towards the end of people's lives is a very interesting one. It probably is something to do with improving people's quality of life, but not in a way that would be easily addressed by what we currently think of as technology assessment. It would be hard to imagine people doing randomised studies of palliation for people with end stage cancers, for example. We should be looking more seriously at what benefits we are getting from the expenditure in that part of the life span.

Secondly, in moving away from age as the key variable influencing health care spending, we should not forget about some other demographic factors that potentially are going to be significant in the future. In particular, the role of marital status and the strong evidence of relationships between people's domestic circumstances and their health care use, is something we have not paid enough attention to.

If one were to project what household composition might look like in another 15 or 20 years, it is likely to be much more atomised than it is now. It is possible that changes in household composition will turn out to be much more important influences on health care expenditure than age.

Klim McPherson: My agenda has to do with changes in perceptions about what health is. There are many examples of ways in which this is currently wrong. One is the CHD national service framework, which talked endlessly about defibrillators and drugs and treatment and surgery, but hardly at all about what it is that causes people to get CHD and how much of that is amenable to change. As I have emphasized, change is possible: not by telling people what to do or being authoritarian, but by enabling people to have real choices.

I think it is a question of using policy levers in a way that will change people's view about what health is all about. One view is that we should enhance preventive care. But I do not really want to enhance preventive care! I want to enhance the opportunities for people to make up their own minds about what they might or might not die of. That means all sorts of very fundamental changes in the way we run our society.

Finally, I agree that if you do save a lot of people from having coronaries, they clearly will have to die or become ill from something else. That kind of analysis has yet to be done. I accept that point completely.

Section 3

MODELLING REDUCTIONS IN WAITING

Chapter 7

Resource and service implications of short waiting times: the French experience

STEPHANE JACOBZONE⁸

This chapter looks at:

- the main reasons for differences in waiting times between the French and the UK health care systems;
- the underlying economic incentives facing providers of inpatient care in France and their impact on the responsiveness of the system to patients' demands;
- the implications for the UK health care system.

Introduction

It may seem odd to discuss the issue of waiting times in relation to the French health care system. Although waiting lists do exist, there is little political and public concern about waiting as an issue. But, of course, this lack of apparent policy concern in France contrasts strongly with the situation in the UK where waiting lists and waiting times are much longer. Why this difference exists is an important question.

While it is well known that France devotes more financial resources to health care than the UK, with which it buys significantly more inputs (see Table 7.1), there are other differences in the way the two systems are organised and structured which account for differences in waiting. Increasing resources in the public health care sector without changing the underlying incentives may prove insufficient to fully tackle the issue of waiting times.

8 The views expressed here are the author's and do not necessarily reflect those of the OECD. I would like to thank my colleagues at the OECD, Jeremy Hurst and Luigi Siciliani, for ideas and thoughts on the subject of waiting times. Their international study of waiting times has been released as a working paper by the OECD in 2003: 'Tackling Excessive Waiting Times for Elective Surgery: A Comparative Analysis of Policies in 12 OECD countries. (See www.oecd.org/health). I would also like to express my warm gratitude to John Appleby from King's Fund for his assistance in writing up my presentation.

Table 7.1 Health system resource differences between France and the UK

	France		UK	
	1990	2000	1990	2000
Total beds (per 1,000 population)	9.7	8.2	5.9	4.1
Acute Care Beds (per 1,000 population)	5.2	4.2	2.8	3.3
Physicians (per 1,000 population)	3.1	3.3	1.4	1.8
Nurses (per 1,000 population)	5.4	6.5	5.3	5.3
Total public spending (per cent of GDP)	6.6	7.2	5.0	5.9
Total private spending (per cent of GDP)	2.0	2.3	1.0	1.4
Total spending (per cent of GDP)	8.6	9.5	6.0	7.3

Source: OECD health data file (2002).

Clearing the market

One way of modelling waiting times is to think about a market which is out of balance: crudely, demand exceeds supply given the resources devoted to health care (that is, the budget set for the public system). The market must therefore find non-price ways to equilibrium. The obvious non-price mechanism used by the UK NHS to clear the market is waiting times.

The demand for health care will be influenced by the level of ill health in the population, technologies available to deal with diseases, government policies on rationing and a whole host of other factors including the level of supply of care. Demand may be dampened by

the level of co-payment required from patients, but this effect is not very strong in the long run, particularly for inpatient care where price is less relevant. Supply, on the other hand, will depend on capacity in terms of beds or high-tech equipment or the workforce; payment incentives for providers; productive efficiency; and so on.

Given this, the focus here is on the microeconomics and the underlying economic mechanisms at work in the French system, which have permitted France to avoid significant waiting times. This simple economist's model of demand and supply is of course too simple a way of representing all the intricacies of the health care market. However, with a bit of elaboration it can help illustrate some important issues concerning waiting, and help highlight important differences between the UK and France.

The UK situation

The essential difficulty faced in the UK is that the level of resources publicly allocated to the health care system is set below that which would clear the 'market'; there is 'excess demand'. The NHS has traditionally dealt with this imbalance by delaying treatment, in effect reducing demand to match the predetermined level of supply. But waiting times are not the only solution to the problem of excess demand. Other choices include loosening the resource constraint – that is, increasing the publicly funded resources to bring supply and demand into line. These publicly-funded resources could either go to publicly-provided or privately-provided care. Another option is to increase the resources brought to the system through additional private input, either by encouraging individuals to buy their own care, or to increase their subscription to private/supplemental insurance. The impact of the additional resources depends on the elasticity of supply with regards to additional resources. With a low elasticity of supply, spending would have to increase substantially even to gain a relatively small increase in health care.

The French situation

The situation in France is not, as some may believe, that there are no

waits and that the system fully meets all demands. Waiting does exist, but the excess demand in France is significantly smaller than in the UK and hence waiting as a rationing device is simply not seen as particularly problematic.

When the public sectors are compared, France faces problems that are in most respects similar to those experienced in the UK for inpatient care. All public hospitals are effectively managed centrally following fairly rigid public sector rules, with national pay and staffing guidelines and, until recently, very weak links between activity and funding. Everyone can access public hospitals nearly freely, but waiting exists. For example, it is not unusual to have to wait more than six months for elective treatment at a Parisian hospital. Nevertheless, how has France managed to keep waiting times within politically acceptable limits? Why is France different?

There are a number of reasons why system differences in France may account for differences in waiting times. First, France spends more on its public health care system: around 1.3 or 1.4% of GDP more than the UK. The second difference in terms of resources is physicians, where France has a significant edge over the UK, with 80% more doctors. However, a third set of differences is more important and concerns the nature and size of the private specialised and inpatient care sector, competition, the dual specialisation of clinics and hospitals and the types of payment incentives in the system.

There are two major differences here between France and the UK:

- The first is the existence of private ambulatory care specialists. There is generally no need to wait for a consultant in a hospital, when the same person, or a person with a similar level of qualification, can be seen directly and privately. This is possible because of the large number of physicians in France, but also because of the fee for service payment scheme for private physicians, which makes them very receptive to patients' demands⁹.

9 It should be noted also that these ambulatory care specialists are paid under an official fee schedule, where patients are nearly fully reimbursed, given the near universal supplemental insurance, for those specialists respecting the official fee

- The second is the private inpatient care sector, which has about one third of all secondary care beds in France. It is one of the major players in the hospital system. The origins of this sector are historically rooted in the weaknesses of the public hospitals after World War II, and before the major reshuffling introduced in 1958 by the Debre Reform. Then many entrepreneurial private physicians chose to set up their own clinics. The clinics also involve private not for profit religious units and units from the Red Cross and other organisations. Several private groups have now acquired chains of clinics.

This system involves a purely fee for service reimbursement, with intense competition among private clinics and between private clinics and public hospitals. There are bankruptcies, mergers and acquisitions. Even public hospitals have to close sometimes, or merge some of their underperforming surgical services with nearby clinics, as requested by regional hospital authorities. Most clinics contract both with public insurance funds and with private supplemental insurance, where most patients would only be required to make very minor co-payments, except for a private room or other amenities.

That the prices paid by the public insurance funds, both for the majority of private physicians and for nearly all clinics, are essentially fixed. Competition is thus on the basis of volumes. This means that additional quantities of care can be purchased comparatively cheaply. Prices have often not been re-evaluated for many years. This means that additional productivity gains have to be realised every year, and that the system functions like a 'price cap' for public utilities. Fixed prices present the private sector with a particular incentive: to specialise in high volume elective work where they can achieve

schedule (Around 23 Euros per visit in 2003, and nearly 75% of all specialists). However, the patient may be only partly reimbursed, given the relative generosity of his or her supplemental insurance, if the specialist is allowed to set their own prices (around 25% of all specialists). There have been increasing pressures in the French medical profession in 2002 and 2003 to ask for greater freedom in setting the level of fees, disregarding the official fee schedule.

productivity gains, that allow for a profit, even taking into account the moderate price-cap. By responding to this incentive, the private sector both provides the additional outputs the system needs in order to balance, and helps to clear the waits in the system.

How health care is financed in France

It is important to understand how the French pay for their care, as this has a bearing not only on the relative unimportance of waiting as a problem in France, but also on other potential problems of coverage and geographical inequalities in access to care.

France has a two tier social insurance system. Interestingly, while some of the arguments in the UK public debate support such a system, France has been moving away from funding based on social insurance (although in practice formal arrangements still largely resemble those of a social insurance system). In France there is increasing reliance on general taxation, mainly because of the detrimental impact on the labour market of the high social insurance contribution. This simply means that the existing employee social insurance contributions have been replaced by contributions levied on all incomes. This tax, called the CSG and introduced in 1987, also applies to capital earnings, unemployment benefits, pensions, and anything that is above the minimum income. The tax is only slightly progressive.

Overall, quasi fiscal and direct fiscal resources represent over 40% of the public insurance funds' resources today. However, a significant proportion of the revenue continues to derive from the employer social insurance contribution: over 12% of salaries. As a whole, the fiscal burden resulting from the public health insurance system imposed on French salaried workers is very high: over 20% of gross labour costs. This does not include private health insurance, which may increase contributions to around 22%. There is an ongoing debate in France as to the appropriate level of taxation and the productive and redistributive implications of changing taxes further to eliminate fully the notion of a contribution purely levied on salaries.

In addition to the 'basic' funding system, supplemental insurance coverage exists. This has a very different history to private medical

insurance in the UK. In 1970, 65% of the population was covered through the *mutuelles* or supplemental insurance. Supplemental coverage has always been important because the public system generally covers only 80% of hospital costs. Additionally, the public system only reimburses about half of ambulatory (primary) care costs in France. In order to cover the remaining gap, *mutuelles*' take up had risen to more than 80% of the population by the early 1990s. For the very poor, a means tested supplemental coverage, the *Couverture Maladie Universelle* (CMU) has been introduced in recent years. Consequently, the only people who do not have supplemental coverage now are the near poor – people earning around £350-600 a month, many of them working poor, salaried agriculture workers or employees of small firms in the service sector.

In a sense, this system is filling the gaps, neutralising the co-payments introduced by the public system, acting in a complementary way to the insurance provided by the public system. This is different from the UK where private insurance generally finances services that are also provided or funded by the public system.

In France, fee-for-service payments for ambulatory care are made directly by the patients themselves, who are subsequently reimbursed. The French are critical about this because of its inflationary effects, but it is difficult to change given the reluctance of the medical profession to receive direct payment from the public insurance funds, and the wish to preserve this system of '*médecine libérale*'. However, low fee levels have led to GP strikes. The level of fees for GPs had not been re-evaluated in the past eight years: it has stayed the same in nominal terms between 1993 and 2001 at between €15 or €16 per visit. As a result of the strikes, it has been increased to €20 for a consultation overall, while payments for home visits were reduced. However, at the time of writing, it had not been increased for specialists, which remains at around €23 in spite of recurrent protests.

Secondary care is financed through a dual system, which at first sight can be seen as not being very rational. Public hospitals are funded on a global budget basis, with budgets being decided on a historical basis. Recently, with the introduction of diagnosis related groups

(DRGs), there have been modest attempts to readjust these budgets according to activity as measured through notional points by the DRG system.

Private clinics receive monies from the public insurance funds on a fee for service basis. Reimbursement for a stay at a private clinic is similar to that in the US: hospital 'hotel' costs are paid separately by the insurance funds and on a different basis to payments for medical services from physicians.

In terms of spending levels, the aggregate level of expenditure on health care in France is about 2.2% of GDP higher than in the UK (equivalent to £20 billion). It was 2.6% of GDP higher ten years ago. The deficit of the main social security scheme financing health care was about 0.4% of GDP in 2002, and increased significantly in 2003, which means that health care is now part of a major policy debate in meeting the Stability and Growth Pact requirements of the Euro area and associated problems of managing fiscal stability as part of the European Union. Health care is to a much greater extent than hitherto, caught up in a macroeconomic debate. The Juppe plan 1995 tried to implement a binding financial constraint on all ambulatory and inpatient care expenditure but it has never really worked. It has been challenged in courts, including Conseil D'Etat and the constitutional court, where the Government has lost several times, which meant that the new scheme was not implemented.

Macro incentives in France and the UK

Although there are concerns in France over the level and rate of increase in health care funding arising from the nature of the budget for health care, there are also positive benefits arising from having non-binding financial constraints or open-ended budgets. For example, and apart from the recent and moderate fee increase for GPs, additional expenditure (over and above that predicted or planned) incurred in the French ambulatory care and private inpatient care systems is always associated with quantity increases and therefore results from meeting additional demand. This is not necessarily true, however, for additional expenditure injected in the public inpatient

care system, and particularly in the case of the additional costs of the 35-hour week imposed on public hospitals.

In the UK, on the other hand, there are very different types of incentive associated with global, constrained budgets. For example, increasing spending in the UK system means setting a budget at the beginning of the year. In other words, this means showing the money up front and then hoping people will do something useful with it. But in practice this can induce ‘capturing’ behaviours. The extra resources could end up being wholly or partially absorbed by higher wage demands by public sector unions (who, because the global budget is known, also know what employers – the NHS – can afford to pay). There is therefore a risk of ‘rent dissipating’ when allocating resources to a system in such a way, without any clear and binding link between funding and activity.

With the open-ended French system, additional expenditures result in a deficit known popularly as the ‘hole of the sécu’¹⁰. This is a recurrent topic in public debate. On the one hand, this may be perceived negatively – as a public system which can never be accountable – but as fees cannot be increased (given the precarious status of the health insurance finances), if physicians want to earn more they have to do more.

This system has also been implicitly used to rebalance overall funding, which is excessively geared towards hospitals compared with other countries. At the end of the year, the public hospital budgets are generally in line with planned spending, by definition, whereas ambulatory care expenditure has generally overshot all plans in recent years. Therefore, at the margin, this helps to rebalance funding away from hospitals.

Resources

As is clear from Table 7.1, as a percentage of its GDP, France spends a quarter more on its publicly-funded health care system than the UK and nearly double the amount on private health care. Such differences

10 The ‘Secu’ is the popular nick-name for the *securité sociale*.

in spending levels (which stretch back many decades) explain the fact that France also has around twice the overall number of hospital beds than the UK, a third more acute care beds, three quarters more physicians and a fifth more nurses. However, the gap was slightly reduced over the 1990s, particularly in terms of acute care beds, and also in terms of physicians. However, the nurse gap was not filled at all, even though France is currently experiencing shortages of nurses at current levels.

As far as levels of technological resources are concerned, France also tends to be better equipped than the UK, with respect to mammography machines, revascularisation units for coronary care, scanners and so on, although it remains underprovided in MRI machines (leading to waiting)¹¹. Greater levels of resource inputs also lead to greater levels of utilisation. For example, hospital inpatient admissions per 1,000 population in France in 1998 were 231.0 compared with around 150.9 in the UK.

One observation, given these resource differences, and given the underlying financial incentives for providers, is that the recent trends in the UK to increase health care spending, are at risk of creating inflationary pressures, unless the number of physicians can be drastically increased.

Competition in delivery

Perhaps the most important difference between France and the UK is competition in delivery. In France, the fact that the state pays for both public and private care is not a contentious ideological issue. Private clinics work under contract to the public insurance funds. Given the dual incentives generated by the system, clinics have had to specialise in high volume elective surgery in order to survive and prosper.

11 The 1999 levels were 9.6 per million in France against 6.1 in the UK for CT scanners, 2.8 against 4.5 for MRI, 7.2 against 4.8 for radiation therapy equipment (OECD Health Data 2002). For mammography machines according to results from the recent OECD Ageing Related Disease project, the gap is very high, with more than six times more machines in France than in the UK: 175 against 25 per million women aged 40 and over. The French rate is even above the US. (These figures do not account for the relative obsolescence of some of these machines).

For surgical activity, particularly non-emergency interventions such as heart valve operations, private clinics account for more than half of all procedures. In urban areas the private sector's market share in elective surgery can be anywhere between 50% and 80%. There is a split, however, in terms of the types of patients who use private and public health care services, with the well-off tending to attend private clinics and psychiatric patients, the poor and older patients using public hospitals. Public hospitals also have a monopoly on emergency care, and on research and teaching.

In the UK, use of and access to the private sector is much more limited. Only about 10% of the population has private insurance. Care in a private hospital can be considered as a luxury good. In addition, private hospitals set their own prices and tend to operate in a very expensive way compared to many of private hospitals under contract in France.

In France, the cost of going outside the public system and accessing a private clinic is very low. Referral by a specialist to a private clinic only depends on whether a patient has supplemental insurance. But with 80 to 90% of the population holding such insurance the barrier in terms of access to the private sector is lower than in the UK.

The nature of the market relations – and particularly the essentially fixed prices paid – means that private clinics have to work in an entrepreneurial way. In essence this means specialising in high-volume, low risk elective work in order to maximise revenues. As a result, private clinics are relatively productive in terms of the work they undertake, and certainly more efficient in carrying out this work than the public hospitals (the opportunity costs of increasing activity in the public hospitals in order to reduce excess demand would, therefore, be higher). No official direct cost comparisons have been published to date between private clinics and public hospitals, but unofficial estimates reveal the latter to be around 15% more costly than the former. However, this estimate does not take account of case mix differences, and the significant social disadvantage or higher age of patients in the public sector, which results in higher costs. The comparison is also made difficult by the fact that physicians' income is recorded separately and is often not included in private clinics'

accounts, whereas it is included, as with all other salary costs, in public hospitals budgets.

However, public and private hospitals coexist. So, for example, in most areas there will be a large public hospital and a number of small private clinics carrying out elective surgery (cataracts, heart valve operations and so on). The public hospital remains the only place where a full range of activities can be found and is still the place of last resort, something no one wants to undermine. There is therefore competition, but at the margin.

One particular issue of note concerning private clinics is that their capital investment has to be financed privately in private markets, at market conditions. The Government does not subsidise the way clinics are financed, nor does it refinance their capital. This is quite a problem for many clinics, and one of the reasons why there are so many mergers and acquisitions – many simply do not have the resources (or access to resources) for the huge capital investments that are required in health care.

As for public hospitals, the central government has little control, centrally, on the level of investments that they make. Public hospitals borrow from markets at market rates, often through the *credit local de France* (DEXIA) and their loans have to be guaranteed by local authorities¹², themselves under the scrutiny of rating agencies and financial markets. Banks look at the financial health of local government and cities before giving loans. In this way, capital spending is to a large extent constrained through traditional market lending rules. In addition, until recently, administrative requirements have been imposed on purchases of major hospital equipment, through the ‘*Carte Sanitaire*’. This planning mechanism (one of the first to be introduced) was abolished in 2003, as it led to much political bargaining, and political decisions about equipment in a minor provincial city ending on a ministerial desk in Paris. Planning is now to be conducted at regional level, and under the auspices of the newly created hospital regional authorities (*Agences Régionales de l’Hospitalisation*).

¹² It should be noted that the head of the city council is also the chairman of the hospital board.

Micro incentives

Within the public and private sectors, the work incentives of physicians differ. Physicians in public hospitals retain a monopoly and enjoy a certain prestige and status in terms of, for example, their teaching roles. The status of hospital physicians was greatly improved by the Debre Reform in 1958, with the possibility of a double affiliation to the medical school and to the hospital, with a double system of pay. As a result, if physicians obtain a teaching post in a university, based on their research achievements, they also enjoy a reasonably high income, particularly outside Paris. Following the introduction of the 35 hour week, both teaching and non-teaching physicians have also seen their potential leisure time greatly increased. Alternatively, physicians can enter private practice to earn more, but they must also be more productive, increasing their income mainly by increasing their activity.

Although the incentives faced by physicians working in private clinics encourages productive efficiency, budget caps have been introduced since 1993, which can mean that fees are adjusted downwards if increases in activity are too high (similar to the German-style floating point system). From the point of view of the public insurance funds and the contributors this is a reasonable situation. Private clinics act as a buffer and absorb excess demand at lower cost, which shifts the burden off the shoulders of the public system, which would otherwise be overwhelmed. Each sector in a sense specialises in their field of excellence. In addition, they increasingly co-operate at a local level and enter into economic partnerships, exchanging activities, or sharing some major equipment.

Coverage, equity and other problems in France

However, problems remain with the funding and reimbursement system. Although there appears to be a low level of unmet need, in reality there are regional variations in health and access. For example, in Toulouse about 90% of acute myocardial infarctions received a revascularisation; whereas in Lille only 30% were revascularised.

There is also a growing awareness of social and socio-economic issues associated with the French health care system. With the universal

double coverage of almost all the population, progress has been made in improving access across social groups, but much less attention paid to geographical equity in France. The distribution of physicians across the country tends to follow the distribution of sunny days during the year and the level of per capita income in various cities and regions for specialists.

There are also regional variations in the state of the capital stock – a result of historical factors which the current funding system does not properly address. Overall, the state of technical facilities is falling behind, prompting the government to announce a plan to spend £500m to rebuild hospital technological facilities in the next couple of years. But generally, health care resources are not distributed on the basis of need. When deprived areas of France are being examined, then waiting lists and times start to become an issue. This is particularly true in the north of the country. However, the scope of the problem might appear moderate, as even the north seems to some extent to be in a privileged situation in comparison with the UK.

Other problems in the French system include huge gaps in terms of, for example, rehabilitation and also nursing homes. In the case of Alzheimer's disease or dementia, one would have to travel 100 miles away from Paris to find a residential or nursing home at a reasonable price. And although you can obtain a new hip, or receive a CT scan, or be treated for your stroke in Paris, rehabilitation (physiotherapy and so on) is much harder to access. So there are still a number of gaps in the French system which tend to be forgotten about – partly because those who need these services are very old, or very poor, and often very sick people who lack the clout to make politicians act when the market fails them. Some of the gaps in cancer treatment have, however, become a policy concern where the government has made an official pledge in 2003 and announced significant increases in resources and co-ordination.

There are additional pressures on the French health care system arising from, for example, moves to reduce the working week to 35 hours. It has resulted in a reduction of the net capacity of public hospitals and some disruption in the organisation of inpatient care, as is commonly

reported in the press, Although, an official quantitative estimate of the impact still remains to be done. There are also problems with nurse shortages in some areas – particularly Paris, where private clinic pay rates are low relative to the high cost of living. With the 35 hour week, and additional posts, these nurses have been able to find better employment opportunities in other regions and in the public sector.

Reforming health care in France

Overall, it is clear that while waiting has not been a particular problem in France, it is an emerging problem in certain areas. Moreover, while the particular financial and economic environment in which the health care system operates has served to go some way in meeting some excess demand, problems in coverage, access and geographical equity remain. Some steps towards financing health care from general taxation rather than social insurance contributions have been made but remain very partial at this stage. There is also currently a debate about changing the reimbursement system for hospitals, but it remains to be seen whether in a very constrained financial environment it can result in more than cosmetic reform.

Private clinics and sickness funds have been pushing for a change to DRG-related payment, as they would anticipate that this would both favour them and reduce expenditure. However, in the absence of a proper system to fully account for factors determining cost differences between hospitals, and with fears over the non-surgical sectors and psychiatric activity, this move has been resisted until now. Studies have also shown that patients with a lower socio-economic status within the same DRG rating had an average hospitalisation cost that were 30% higher than those of other patients¹³. As long as the payment mechanism does not take all these characteristics into account in a proper way, it is very difficult to envisage anything but a cosmetic change.

13 Mathy C and Bensadon M (2000), 'Prendre en compte la précarité dans la comparaison des performances des établissements de santé: les leçons d'une étude', Mission PMSI, Direction des Hôpitaux, Centre de traitement de l'information du PMSI, Université de Jussieu, Colloque international des économistes français de la santé: L'Etat de la Réforme, February.

Lessons for the UK?

What are the implications for the UK of the French situation? While NHS patients treated abroad tend to grab the newspaper headlines, this method of boosting activity is not really a long term option for the system. Elderly patients in need of surgical interventions would not necessarily enjoy being treated abroad, away from their family and in a different environment. The other option, as it operates in France, is to encourage greater private provision, paid largely from public funds, and under publicly-controlled prices to minimise providers' rents. But as far as the author of this article understands the UK situation, the private sector there is small, current fee levels are high, access is limited and the relationship with the NHS remains a contentious issue.

Nevertheless, one option would be to reduce the marginal opportunity cost of going to the private sector, transforming it into something that would be more mainstream, with lower costs and higher volumes, and would be something that every patient could enjoy in the UK. This would enhance competition on the supply side, whilst using more privately provided care, resulting in increased access and lowering prices. This could be perceived as controversial by some analysts, and not everybody would agree with this, but from the French experience it could yield welfare benefits for the UK, far greater than other alternatives such as increases in resources for the public sector only, or increases in private financing, for example through additional fiscal incentives for private health insurance.

In terms of resources and finance, the UK could make good use of greater levels of funding, but rather than distribute the entire budget at the beginning of the year (risking capture by health care workers and inflationary pressures), maybe it would be wise to have an earmarked fund for privately provided care. The fund would pay any of the care that would be required, at pre-announced rates. At the end of the year, the fund would then be balanced by financing from a general tax, levied on all incomes in a non-progressive way. The level of the financing contribution could be fixed by an independent regulatory authority whose mandate would be to oversee the system,

ensure its financial integrity, and licence the private providers working under contract for the public system. This would only require some initial arrangements for the first few years, and would also offer private investors the necessary visibility for increased investment in capacity in health care that would be necessary.

This system would need to be accompanied by a very significant increase in medical and nurse outputs from medical schools and universities. Without it, the UK would have to cherry pick resources in poorer countries, or to compete with the very high fee levels of the US. While it seems that the UK currently has a willingness to pay for more health care, a question arises as to how long this will last. Without significantly increasing staff, there is a risk that the additional funding will be absorbed by higher wage settlements.

Conclusions

The institutional features of the French health care system make it highly responsive to patients' needs, although relatively costly. The dual public/private structure is mirrored in other systems, but France is unique in allowing open-ended public financing of both the public and the private supply. But can it last? The current financing gap, the largest since the acute crisis of 1993-1994, puts the entire system at risk – although nobody is politically willing to change it, for it remains very popular with the entire population. Pressures arising from an ageing population, renewed pressures from the providers, constraints on public financing levels and the macroeconomic necessities associated with membership of the Euro are putting the whole balance in question. This also means that some of the gaps in the French system, such as rehabilitation services, care of the elderly and so on, have a long way to go before they are filled.

What is the future for the French health care system? What could result could be either implicit emerging waiting times rampant in the public system, or a smaller basket of public goods offered for reimbursement and a greater share for private health insurance. Another option, following the Spanish and Italian example, is greater scope for decentralisation of responsibilities to regional levels. This

will involve intense discussions on the scope of the social contract and is part of the continuing public policy debate.

Chapter 8

UK evidence on waiting

JOHN YATES

This chapter looks at:

- the UK's historical record on waiting lists;
- strategies to reduce waiting lists and waiting times;
- trends in clinical productivity.

Wanless on waiting

Before taking a look at the history of waiting in the NHS, there are three observations to be made about the Wanless review. First, it can hardly be in dispute that there is public dissatisfaction with waiting times. There has been for rather a long time and it should come as no surprise that patients want faster access. However, the suggestions Wanless put forward for future maximum waiting times are radical. Never in the history of the NHS has anybody suggested a two week waiting time. For those not familiar with the waiting times targets, Table 8.1 shows the main access targets set by the NHS Plan and suggested by the Wanless review.

Table 8.1 Reducing Waiting Times: NHS Plan and Wanless Review Targets

	Maximum outpatient waiting time*	Maximum inpatient waiting time
2002/3	6 months	15 months
2005/6	3 months	6 months**
2008/9	3 months	3 months
2022/3	2 weeks	2 weeks

* Excludes cancer.

**With all admissions booked.

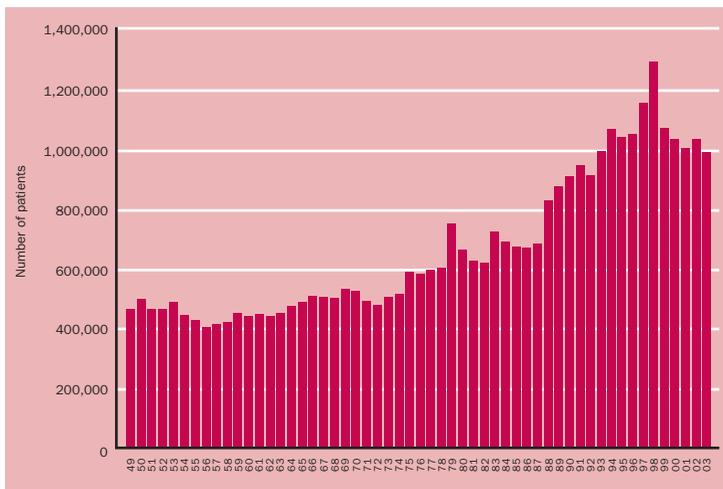
Secondly, the big reductions in waiting times proposed by Wanless will require extra activity which will need to be achieved by a substantial increase in productivity and additional capacity. Whether substantial increases in productivity are feasible is open to question.

Thirdly, the new NHS consultants' contract may help stimulate greater activity and productivity. However the contract plays out in practice, there needs to be a significant increase in contact time for health care professionals with each patient and this will have a negative effect on 'crude' productivity.

A history of waiting

In 1948 the NHS inherited about half a million people on the waiting list, and for 25 years it remained fairly constant. Then came a quarter of a century of growth in waiting list numbers (Figure 8.1). The peaks and sudden blips in the trend are due usually to industrial action or changes in statistical definitions. For example, industrial action in 1979 and 1983 are reflected in larger than average increases in

Figure 8.1 **Total Waiting Lists: England 1949-2003**



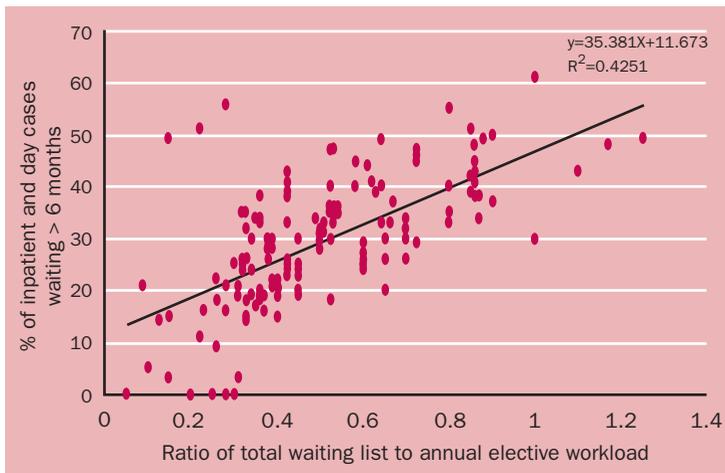
numbers of patients waiting, and the inclusion of day cases in the waiting list statistics in 1988 account for a step change in numbers.

The rising trend over the last 25 years was reversed after 1998 following the new Labour government’s pledge to reduce the list to a level 100,000 less than they inherited in May 1997 – equivalent to a reduction of around 9% in numbers waiting.

Although it has been said that statistics concerning the numbers waiting on a list are irrelevant to patients and that it is waiting time that is important, it should not be forgotten that there is an association between waiting numbers and waiting time. For example, Figure 8.2 implies some sort of association in orthopaedics between waiting time and the size of the total waiting list.

Trusts with the lowest proportion of orthopaedic patients waiting over six months tended also to have small total orthopaedic waiting lists relative to their elective workload. The group of trusts in the bottom left hand corner of Figure 8.2 are those that need to be

Figure 8.2 **Relationship between Waiting Time and Length of Waiting Lists, Orthopaedics, 2000/01, England**

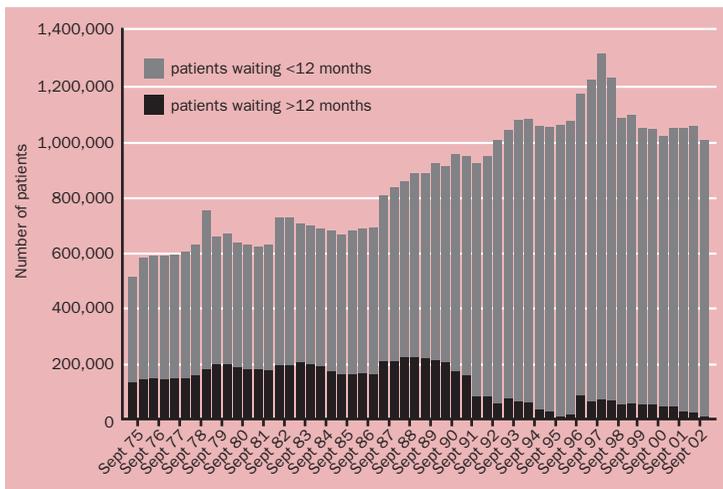


studied more carefully to understand how they got there and why others cannot move towards that position.

An associated technical matter concerns how we measure waiting times. Very different pictures emerge of waiting times depending on the measurement method adopted. Numbers waiting can be calculated on the basis of the waiting time experience of those admitted from the list (i.e. those whose wait has ended), or derived from a snapshot or census count of all those on the waiting list at a point in time (i.e. those whose wait has not yet ended). Mean and median waiting times in the case of the former tend to be rather shorter than in the latter case.

The actual recording of the length of time people were waiting on waiting lists only began in 1975 – and then only in a very crude way with the list split into those waiting over one year and those waiting less than that. As Figure 8.3 shows, until 1990 there were about 200,000 people in England waiting over a year for treatment.

Figure 8.3 Inpatient and Day Case Waiting Lists in England: September 1975 - March 2003



However, this figure included people on the waiting list who were dead, who had had the operation, who had gone somewhere else, who had got better, and so forth. Nevertheless, it remained a substantial number.

The impact of a determined effort to reduce the number of people waiting over a year can be seen in the period from 1991 to 1996. Numbers were reduced substantially from just under a quarter of a million to around 4,500 in September 1996. Since then numbers have risen and then fallen again. There is some evidence that we can tackle the longer end of the waiting list in Britain, and things are not quite what they were 20 years ago, or even ten years ago.

On the alternative measures of mean and median waiting times, rather than numbers on waiting lists, the picture looks somewhat different, however. As Figure 8.4 shows, mean waiting times nearly halved between 1988 and 1992, but since then have remained fairly constant at around 4 to 4.5 months. There is a similar, if slightly less dramatic, picture in terms of the median waiting time.

Figure 8.4 Mean and Median Waiting Times, England, 1988-2001

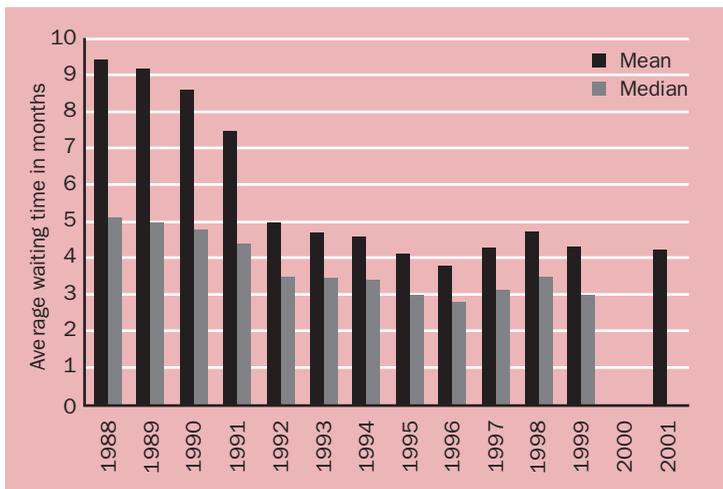
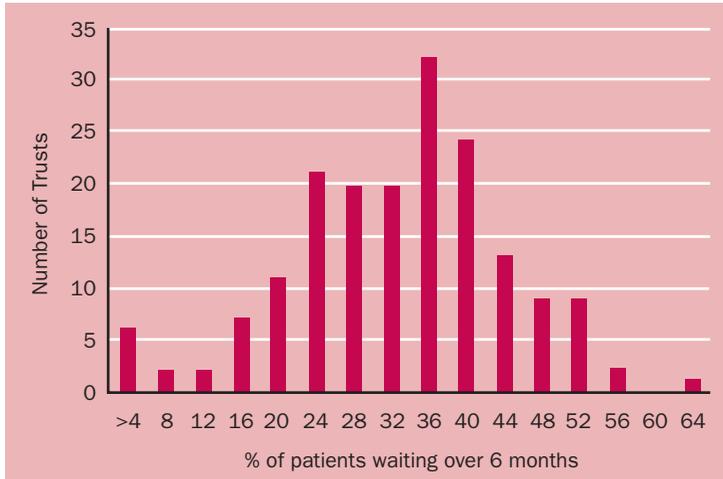


Figure 8.5 **Distribution of Proportion of People Waiting Over Six Months, Orthopaedics, 180 English NHS Trusts, March 2001**



All these national trends over time – no matter which measure of waiting is used – conceal considerable geographical variations. For example, why is it that for England as a whole 33% of those waiting wait over six months, while none wait so long in Dorset? As Figure 8.5 reveals, in one NHS trust in England in March 2001, 64% of people waiting in orthopaedics were waiting over six months for treatment while for four trusts down in Dorset nobody was waiting over six months. There is clearly a lot to learn about why these variations exist and how trusts at the right hand end of the distribution in Figure 8.5 can be more like those at the left hand end. We have a National Health Service in name only. In reality it comprises a whole set of local health services that run things very differently. This variation has been a persistent feature of waiting lists for many years, but defies simple explanation.

Strategies to reduce waiting lists and times

The history of strategies to reduce waiting lists and times includes the

Table 8.2 Progress Towards Waiting List/Time Targets: England

Quarter ending	Total				
	waiting	> 6 months	> 9 months	> 12 months	> 15 months
June 2001	1,037,875	267,160	125,355	46,689	12,319
Sept 2001	1,035,302	277,456	125,491	44,670	10,323
Dec 2001	1,050,221	259,123	117,318	31,769	4,222
Mar 2002	1,034,719	241,262	98,093	22,195	104 (0)*
June 2002	1,054,739	251,467	98,669	20,969	105 (0)*
Sept 2002	1,048,089	253,298	94,039	17,111	61 (0)*
Dec 2002	1,056,648	236,928	86,810	11,002	49 (0)*
Mar 2003	992,045	192,422	54,475	229 (0)*	50 (0)*
June 2003	992,604	188,343	51,487	220 (0)*	56 (0)*
Sept 2003	973,600	180,403	43,603	309 (0)*	103 (0)*
Dec 2003	821,500	151,568	35,168	368 (0)*	68 (0)*
Mar 2004			Target = 0		
June 2004					
Sept 2004					
Dec 2004					
Mar 2005			Target = 0		

*Some people still waiting over 12+ months despite target deadlines

use of list validation (removing people who have died, moved away or for whatever reason no longer require an operation) and ‘tailgunning’ – tackling the long waiters at the right hand end of the waiting list distribution. Setting targets in terms of reducing maximum waiting times is still the preferred method for reducing waiting times. The target for the end of March 2002, for example, was to make sure nobody waited over 15 months for treatment. As Table 8.2 shows, there were around 12,000 people waiting over 15 months in June 2001 and by March 2002 this target had all but been achieved.

Box 8.1 Uncertainty and Variation in Clinical Decision Making

The issue of uncertainty is rarely acknowledged by the medical profession, and by the 'population at large', who are subsequently encouraged to have unrealistic expectations about the anticipated outcomes of clinical decision-making.

(Hatt, 1998)

Both the medical profession and health care policy-makers downplay the appearance of doubt in clinical decision-making.

(McKee and Clarke, 1995)

Traditional medical education has instilled in its students the belief that uncertainty is a manifestation of ignorance, weakness or failure, and this view is often held by patients who have been led to believe in the objectivity and precision of clinical decision-making.

(Hatt, 1998)

A patient of any age who is admitted to a mental hospital in the UK is ten times more likely to be diagnosed as manic depressive than an individual exhibiting the same symptoms in the US.

(Hatt, 1998, referring to Turner, 1995)

There are wide variations in both the definition of 'normal', 'high' and 'low' blood pressures and the treatment of 'abnormal' blood pressure levels in different cultural contexts within what we might call 'Western medicine'.

(Hatt, 1998)

Hospitals, for example, vary enormously in the percentage of normal appendices that are removed. Some surgeons obviously cut too readily, as reflected by the fact that more than half of the appendices removed are normal; in other hospitals, the number of normal appendices removed is so low as to suggest that decisions concerning surgical intervention may be too conservative, and that willingness to assume the risk of ruptured appendices is perhaps too high.

(Robinson, 1978).

Much of clinical medicine remains empirical, and everyday practice is characterised by wide variations that have no basis in clinical science.

(Wennberg, 2002)

QED – quick and early diagnosis – was the title of a paper in *The Lancet* describing patient assessment in a one-stop visit to a hospital unit within two weeks of referral. It described the concept and results and concluded, 'QED is an adaptable concept which could form the basis for a rapid diagnostic service in any hospital The next step is to determine for which disorders earlier diagnosis means better prognosis.'

(Kendall et al, 1996)

The subsequent target milestone for March 2003 required the NHS to have no one waiting over 12 months – again, largely achieved. The target for March 2004 requires a further reduction of around 75,000 patients, and the following year, 2005, a further reduction of around 140,000 to meet the six month maximum waiting time target will be needed. By 2005, nearly a quarter of a million people waiting over six months will have to come off the waiting list (see Table 8.2). It is a major task.

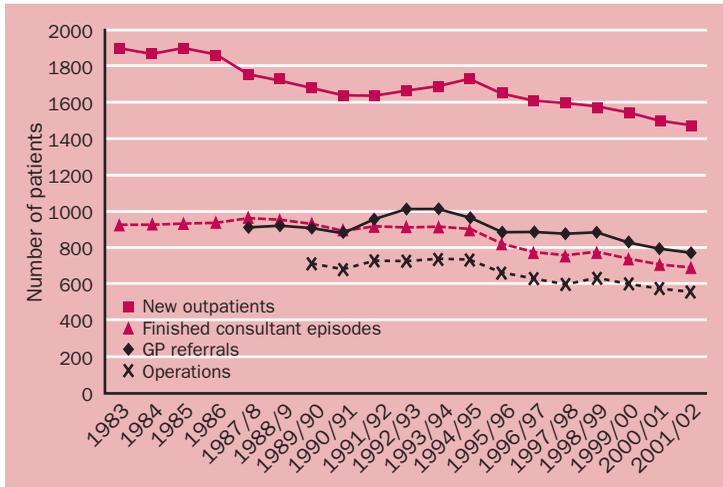
Shorter waiting times... but some notes of caution

Having made various observations about the history of waiting times, the tactics used to reduce waiting lists and times and, as I see it, the difficulty of meeting the NHS Plan waiting times targets, I would make some comment on the waiting times issues and goals raised in the Wanless review.

First, on Wanless' vision of a maximum waiting time of just two weeks by 2022, an obvious question to ask is whether patients and doctors could cope with such a minimal wait for an operation. Two of the key functions of medicine, as Thomas McKeown has pointed out, are to diagnose and to make a prognosis (McKeown, 1971). These activities are invariably somewhat unscientific and there is much evidence to suggest a lot of clinical uncertainty around both diagnosis and prognosis (see Box 8.1). For example, one of the most common interventions in the twentieth century was tonsillectomy. But evidence suggests that most tonsillectomies were unnecessary, and studies as early as the 1930s indicated considerable variation in clinical opinion as to who would benefit from such an operation (Glover, 1938). Compressing clinical decision making to accommodate a maximum two week wait may not be conducive to good decisions in some cases.

Secondly, the Wanless review makes assumptions about increases in clinical productivity over the next 20 years which require reversing the historic trend for falling productivity in some key specialties. For example, in trauma and orthopaedics, Figure 8.6 shows that trauma and orthopaedic consultant productivity has actually fallen over the

Figure 8.6 Activity per Consultant Team in Trauma and Orthopaedics, England, 1983-2001/2

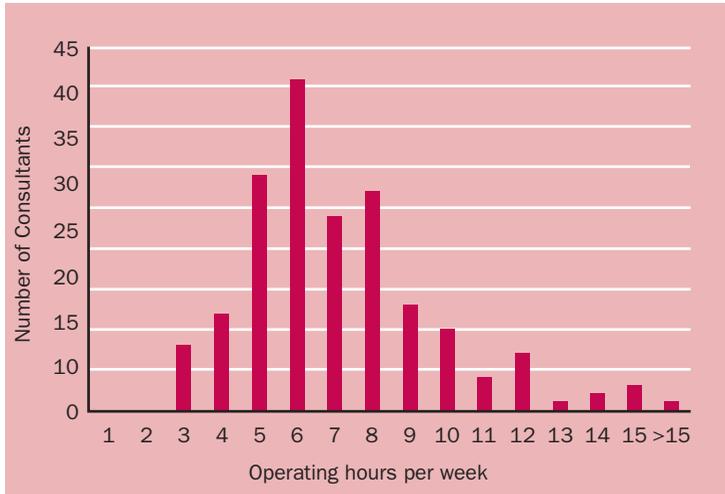


last 20 years. This is due to a faster rise in the number of consultants than in activities such as outpatient attendances and operations.

Similar trends are reflected in some other surgical disciplines. There are many reasons for the observed falls in productivity which I will not go into here. But the point is that if we are expecting (or projecting) an increase in productivity we should recognise that this would be going against a trend of falling productivity stretching back over 15 years.

However, perhaps of more interest than the downward national aggregate trends in productivity is the variation in productivity across consultants. Taking the average operating time in theatre (knife to skin time, not including anaesthetic time) we find that the average time in theatre for surgeons in orthopaedics is about seven hours (see Figure 8.7). But the range of time spent in theatre is considerable, some consultants spending 14 or more hours in theatre each week, and others spending three or four. Just as we need to properly explain and

Figure 8.7 Average Operating Hours per Week per Consultant Trauma and Orthopaedic Surgeon, England (sample: 182 whole time or maximum part time consultants)



tackle the reasons for variations in waiting times (see above), if we want to increase productivity we have really got to examine – and tackle – the reasons why productivity varies so much across the health service.

Thirdly, we need to be sure that very short waiting times are actually what patients always want, given other goals and objectives for the NHS. People do seem to want short waiting lists and short waiting times, but many are just as concerned about equity as they are about waiting. A maximum wait of two weeks may, in many cases, not be necessary if some people are prepared to wait three, four or five months for certain conditions (for example, cataracts, hips, etc). This is unlikely to be the case for other conditions such as cancer, but the general point is the validity of the Wanless ‘vision’ for very short waits, and the distinct possibility that many people would be willing to trade off some reduction in waiting time for a higher degree of fairness with regard to access to care. What concerns and aggravates

many people has been the inequity in the British health care system arising from the ability of those with the financial means to jump the queue.

Finally, it seems a remarkable thing to talk about reducing waiting times to two weeks when we are currently finding it difficult to meet current NHS Plan targets up to 18 times as long. Twenty-five years ago we focused on people waiting over a year for treatment in this country and it has taken us that long to get close to eradicating such long waits. The NHS is now aiming for a maximum wait of nine months and, subsequently, a maximum of six months. Currently, there is no NHS hospital in England that can offer a maximum three months waiting time – the NHS Plan target for 2008. Regardless of the merits and costs of a two week wait, practically, this looks a highly optimistic goal.

REFERENCES

- Glover JA (1938) 'The incidence of tonsillectomy in school children', *Proceedings of the Royal Society of Medicine*; XXXI: pp. 1219-1235.
- Hatt G (1998) 'Uncertainty in medical decision-making', in Petersen A and Waddell C, *Health Matters: a Sociology of Illness, Prevention and Care*. Open University Press: Buckingham; pp.223-239.
- Kendall MJ, Toescu V and Wallace DMA (1996) 'QED: quick and early diagnosis', *The Lancet*; 348: pp.528-529.
- McKee M and Clarke A (1995) 'Guidelines, enthusiasms, uncertainty, and the limits to purchasing', *British Medical Journal*; 310: pp.101-104.
- McKeown T (1971), 'A sociological approach to the history of medicine', in *Medical History and Medical Care – a Symposium of Perspectives*. Oxford University Press for the Nuffield Provincial Hospitals Trust: Oxford, pp.1-16.
- Robinson D (1978) *Patients, Practitioners, and Medical Care: Aspects of Medical Sociology*, 2nd edition, William Heineman Medical Books Ltd: London.

Turner BS (1995) *Medical Power and Social Knowledge*, 2nd edition, Sage: London.

Wennberg JE (2002) 'Unwarranted variations in health care delivery: implications for academic medical centres', *BMJ*; 325: pp.961-964.

Chapter 9

Discussion: modelling reductions in waiting

Topics covered in the discussion of modelling reduction in waiting times included:

- rationing system alternatives to waiting;
- private sector role;
- payment systems and incentives to promote activity;
- measurement and modelling data issues;
- variations in waiting times;
- opportunity costs of reducing waiting times;
- meeting NHS Plan targets.

Hugh Gravelle: Stephane Jacobzone has focused on one method of reducing waiting times: effectively increasing supply, or activity rates – either putting more resources in, or making resources more productive. But an alternative is to change the method of rationing.

To determine how much more we need to put into the health service in order to reduce waiting times, we need an estimate of demand elasticities with respect to waiting times. All the ones I have seen are pretty low; the magnitude is somewhere between about 0.1 and about 0.25 to 0.3. So maybe we do not need such an enormous increase in the volume of resources to bring about significant reductions in waiting times through greater activity.

But the question is why do we want a target of two weeks anyway? There is an alternative, long advocated by my distinguished colleague Professor Alan Williams, which is to change the method of rationing.

Essentially what we have at the moment is an implicit points system. Patients are treated (taken off the list) when they have a threshold

number of points. You get implicit points for time spent on a waiting list. But if you look at the distributions of waiting times across different providers you observe very different distributions, which suggests that there are also implicit prices, or implicit rationing points, for different need factors being applied in different ways by different providers.

The greater the weight you place on the implicit needs points that you accumulate when you are initially placed on the waiting list, and the lower the relative weights on the waiting time, then the shorter is the average length of time that people will wait. Patients with high initial need points have shorter waiting times than those with low initial need points who have to wait for longer in order to accumulate the threshold number required for treatment.

In the limit, you would not accumulate any points for waiting on the waiting list, you just add all your needs points from having particular characteristics, and if everybody was very clever the number of people entering on the waiting list would just equal the number of people treated in each period, and there would be no waiting at all. Hence you would not need any more resources in this system.

Luigi Siciliani: Following up on the issue about alternative ways of rationing, there is one country, New Zealand, which is trying to implement such a policy. Given that resources are limited, rationing is carried out on the basis of need. This has proved an effective way to reduce the time patients wait.

Nancy Devlin: I was involved in the evaluation of the New Zealand points schemes whereby patients are ranked in terms of scores on measures of clinical need. Patients with points below the clinical threshold are told that their clinical condition is not severe enough to justify surgery; this in itself reduces waiting lists. Given the budget for elective surgery, you rank the remaining patients in terms of their points. You treat all of them who have a sufficient number of points to be treated given the available budget – which defines the ‘financial threshold’. If, as a society, we decide somebody with a certain number of points is not getting treatment when they ought to, you then have a sensible basis upon which to decide what the shift in resources

should be between competing uses. Explicit prioritisation of patients therefore facilitates higher level decisions about resource allocation between different types of health care.

Those whose ‘points’ are above the financial threshold will get treatment rapidly and waiting lists can be virtually eliminated. Such a system provides a superior means of managing waiting compared to very crude waiting times targets.

Klim McPherson: The first thing that strikes me is that we have had two masterly presentations, one telling us that we need to take more advantage of the private sector – as the French do – and one about problems of reducing waiting times in the NHS which did not even mention the private sector. I wonder if we should perhaps worry a bit about the private sector and its effects on waiting lists, or the role of waiting lists with respect to the private sector.

Tony Culyer: Whereas the waiting times issue may partly be to do with rationing criteria, there are also major behavioural issues concerning the incentive structure. An important part of that in the UK system is the way the private and public sectors interact. That needs addressing and building into any modelling of waiting times reductions.

Diane Dawson: Stephane Jacobzone asked why the UK does not look at introducing a greater competitive edge between the private and public sector, where the public sector funds, but the private sector provides – at cheaper prices – for certain elective care. That is the system the UK Government is currently putting into place in England. With the move also to use fixed price tariffs for both public and private sector care, then in many respects the English NHS system is going to be quite similar to what was described in the French system.

But does France share its public sector doctors with the private sector, as happens in the UK?

Stephane Jacobzone: The private and public labour markets are not separate in France either, but their interaction is complex. Physicians all do their internship and their specialist training in public hospitals. Once they have finished their internship, a small proportion will stay in the public hospitals, with a tiny elite pursuing a double

university/teaching career, and many of the others will go and find a job in the private sector, as ambulatory care physicians. They will work full-time in private practice or part-time in a clinic. Many of them will maintain their links with a public hospital to maintain their skills and keep contacts with former colleagues, through keeping some consultations in public hospitals.

There is a lot of cross-over between the public and private sectors. For example, a doctor may work, say, half a day once or twice a week in a private clinic, then they may have private consultations at another practice where they work two or three days a week, and then they may work another two days a week at a public hospital where they do their procedures. When a doctor works for the public system they are paid by the hospital and when they work for the private system they are reimbursed by the sickness funds.

What is more complex is that some of the physicians can maintain a private practice within public hospitals, where patients have to pay out of pocket. That is tolerated by the political authorities as a way to retain some of the best physicians in the public sector. However, this is not statistically significant.

There has been a lot of interest in Britain in overseas experiences of public/private mix in delivery. But it seems to me that it is difficult to gear up the private sector in Britain as it is effectively, a cottage industry. Moreover, while you may want the private sector to start running cataract factories with high volumes of care, this will take time. They will need a stable regulatory environment over several years; some certainty to make investments.

Adrian Towse: The UK Government is indeed encouraging elective procedure factories which will be sub-contracted to the private sector. That could well be a way of producing incremental capacity at a relatively lower cost, as in France. Fixed prices based on health care resource groups may well provide the incentive to increase volumes.

At the moment the way we are using the private sector is mostly not on a long term basis, but is short term, at the margins, and it is expensive. We are using agency staff, we are discharging to private

nursing homes – but it is ad hoc. Long term relationships at lower incremental cost are just not there at the moment.

Barry McCormick: Could Stephane Jacobzone tell us the nature of the doctors' contractual relationship in the French public sector? Are public sector physicians paid on a session basis, or are they paid an annual salary?

Stephane Jacobzone: Physicians are paid on a salary basis in public hospitals. They are on the career scale like all other civil servants, which means that the scale of pay is determined on a national basis. The rules are the general rules for the public sector and are fairly rigid. However, those physicians with a teaching affiliation also receive in addition a salary for their teaching. As a result, the financial conditions for teaching physicians have been made attractive in public hospitals, to keep them competitive with the level of income generated by the fee-for service payment in the private sector. The other thing that can drive careers in public hospitals is research work. That is how physicians gain promotion and higher pay.

Barry McCormick: Does it follow that access to public elective care is rationed?

Stephane Jacobzone: No. There is some implicit rationing in public hospitals: when physicians are spending less time in their practice, this tends also to generate a waiting list. If you want an appointment with a particular named consultant in a Parisian public hospital, for example, you may have to wait up to six months, as happened to me, because these people do not spend a lot of time receiving their patients as they are largely focused on research. The difference with the UK though, is that I can see a physician in the private sector or in ambulatory practice the next day if I wish to.

The French public hospitals' systems also suffers a number of rigidities. The French Government is currently trying to implement, at the margin, a DRG-type payment system. Each hospital produces so many DRG adjusted points every year. The budget is then divided by the number of points, to see how costly it is to produce a point in a given hospital. This is used to see which hospitals are relatively costly

and which are relatively efficient. This can then be used to shift some resources at the margin from one hospital to another from year to year, in a very constrained budget environment. This is currently performed through regional planning mechanisms where the head of the regional hospital authority can reallocate resources across various hospitals, public or private, in a given geographical area.

Sean Boyle: There is an issue about outpatient waiting times. There is a wait for outpatient appointments as well as a wait for inpatient appointments. We need to bring the two kinds of waits together. It is amazing that you cannot actually get data that links them. If you look at performance on outpatient appointments you find that things are going the wrong way: people are waiting longer on average. That outweighs the shorter average wait for inpatient appointments.

James Raftery: In terms of taking forward the productivity issues that John Yates has identified, a critical part of that is having data at consultant level. John can do his work only with great difficulty. He can get the data at consultant level for particular specialties only if they are released by hospital chief executives. This is a crazy way to run a system.

Secondly, an associated data issue is where are the models used by the Wanless team in their review? Can I have them? Did somebody model the two week wait? If so, what assumptions did they make? What will it cost? If they did not, that is okay, but can we still see the modelling that was done, not just the assumptions, but all the data?

Alan Maynard: To reply to James' first point about consultant level data: there is a page on the Department of Health web site which shows their clear intentions to change the fields of the Hospital Episode Statistics (HES) so that you can clearly see who is doing the operation; to link HES to a private sector equivalent of HES, which will enable us to model public and private activity; and to link HES to Office of National Statistics 'registration of death' certificates, so that you can follow people out into the community. The intention is to make this data available in the next three years at the consultant level.

Sean Boyle: I am involved in a project looking at getting sustained improvements in waiting times, and have been visiting Trusts to explore this. I think that there is a lot to be said for looking at why waiting times vary across the country. My feeling is that there is substantial room for improvement in terms of better use of operating sessions and better use of outpatient clinics so as to reduce productivity variations across trusts. I suspect that there is plenty of capacity already in the NHS and that the first step is to realise that capacity (through improved productivity) rather than turning immediately to the private sector for extra capacity.

Nancy Devlin: There is a danger in focusing on the feasibility of achieving a given waiting times target. I think that there is something to be said for more fundamentally questioning whether or not waiting times targets are the way to go. For example, with regard to the two week waiting times target, if we can figure out roughly what amount of activity and spending we need to achieve that, the relevant question is: what is the opportunity cost of that in terms of the health gain foregone from using those resources elsewhere? That is fundamentally what we should be concerned about. Elective surgery tends to be politically and popularly very prominent, but waiting times targets are not necessarily going to guide a sensible use of resources.

Adrian Towse: The benefits of a two week waiting time cannot be quantified in terms of improved health outcomes. It is about whether people are actually willing to pay for the extra resources in terms of the benefits they see of a two-week wait, and maybe they do not want a uniform two weeks' waiting. It may be they want that for some forms of treatments, but not others. But I think the general point is that we need to have some understanding of what length of waiting people regard as an appropriate trade off with other uses of the scarce resources required.

Section 4

MODELLING NHS SPENDING: WHAT CAN WE LEARN FROM ELSEWHERE?

Chapter 10

Modelling health care spending in the US

MARK FREELAND¹⁴

This chapter looks at:

- the objectives of modelling health care spending in the US compared with the UK;
- approaches to long term modelling of health care expenditure used in the US;
- issues encountered in modelling US health care spending;
- what might be learnt from the US for future UK modelling exercises?

The purpose of modelling

The UK and the US obviously have very different health care systems and substantially different levels of health spending relative to the GDP. This in turn, is reflected in somewhat different objectives for modelling health care spending in each system. In the UK the focus of modelling exercises, such as the recent Wanless Report, is ‘What resources are required to close ‘performance gaps’?’ In the US, the greater role of the market economy in the health sector means that modelling instead focuses on the question ‘How do interactions between market forces and government institutions affect policy objectives, policy outcomes, spending, and budgets’?

Models of health care spending play an important role in US health policy. They facilitate informed dialogue and decision making. They allow policy makers to examine economic, budget, and access-to-care impacts of alternative health policies, and inform assessments of whether existing programmes should be modified or dropped, or new programmes added.

¹⁴ The author gratefully acknowledges assistance from Gregory Won, economist in the Office of the Actuary. The views expressed in this paper are those of the author and do not necessarily reflect the views of the Office of the Actuary or the Centers for Medicare & Medicaid Services in the Department of Health and Human Services.

Given the importance of the health sector in the US economy, projected health care spending has significant implications for living standards and economic growth. There is little doubt that past growth in US health care spending is unsustainable – if it stayed on its current track the US economy would essentially collapse. Clearly, a policy response is required: carefully developed projections allow policy makers to anticipate future fiscal imbalances and therefore, in principle, to craft rational cost containment strategies that are phased in to balance future budgets.

In the US there are two groups doing baseline and ‘alternative scenarios’ projections of health care spending. Seventy-five-year actuarial projections are developed by the Medicare actuaries (executive branch of government) and they also do ten-year projections for the Federal budget. The Congressional Budget Office (legislative branch of government) also produces ten-year projections of health spending on the basis of current laws and then for every new benefit proposal concerning, for example, Medicare or Medicaid.

The existence of competing modelling exercises by the Office of the Actuary in the Department of Health and Human Services (executive branch) and the Congressional Budget Office (legislative branch) has served to sharpen the accountability and transparency of modelling. Projections by both organisations are closely scrutinised by lobby groups representing the stakeholder interests of the elderly, hospitals, insurance industry, pharmaceutical industry, and so on – each of which hire top academics and consultants to pick apart the projections. The existence of simultaneous, competing models and strong external critical review imposes an important discipline on the modellers and the modelling.

Different models address different policy problems. By its very nature, there are diverse and complex issues in health economics, so different models are required depending upon the objectives addressed. None of the models is as accurate or as good as we would like. However, the models provide discipline in the form of the consistency of assumptions and structures of interrelated components. Finally, the models provide insights for policy makers that would not be available without the framework of the models.

Knowledge and data are powerful, but there is no single, best way to model health spending and budgets. Model building for projecting the future is intrinsically an exercise in humility. As the historical data come in to replace the projections, the complexity and inherent unknowability of the future is forced upon the model builders.

Approaches to modelling

Ten-year projections of national health spending produced by the Office of the Actuary proceed as follows. Econometric projections are produced by a team that includes economists, statisticians and actuaries. Medicare actuarial projections (mentioned above) are included as an exogenous input into the national health expenditure projections. The private sector is modelled conditional on the exogenous Medicare projections, using econometric methods, employing a ‘bottom up’ approach by type of service (physician, hospital, etc.) and a ‘top down’ approach, looking at total private spending per capita. The top down and bottom up approaches are then reconciled.

There is, of course, nothing magic about econometric forecasting. It cannot provide definitive predictions for an infinitely complex world. We do what essentially all econometric forecasters do: we ‘add-factor’ the equations for relevant omitted variables by adjusting the constant terms of the equations using our best judgement. There are always things going on in the economy that cannot be measured.

Our work is transparent in one sense, i.e. people can access and examine our add-factored equations. However, even if they had our forecasting model and equations, that does not mean that they can simply take our model next year and run it successfully. They could mechanically run the model, but it does not mean that it would provide reasonably precise forecasts. This is because the process involves both the econometric ‘science’ of equation estimation and the art of expert judgement (add-factoring constant terms) on what is dynamically going on industry by industry over time, but which is not measured by the equations explicitly. This is the same process that Lawrence Klein used in his Nobel prize-winning econometric model

building: before he would release econometric forecasts, he would have invited subject matter experts from each industry, for example the coal industry, to assess the reasonableness of the forecasts. The industry experts may say ‘Your model did not adequately take into account factors X and Y.’ The constant terms of the forecasting equations are then expertly adjusted to incorporate those omitted factors. That is also the approach used in the US to model and forecast health care spending.

Projections of US health care spending

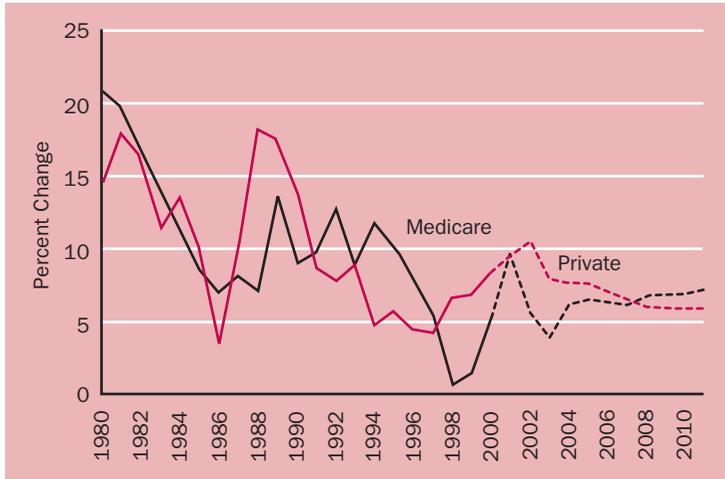
Figure 10.1 shows the ten-year projections (2001-2011) for national health care spending as a proportion of GDP. There are two important influences to note regarding the trends apparent in this figure. First, the increased penetration of private managed care had what we predict is primarily a one-time, distributed, downward effect on the private sector spending growth. It substantially reduced physician fee levels and it got patients out of hospitals into outpatient and

Figure 10.1 **Ten-year National Health Expenditure Projections: Total Public and Private Spending as % of GDP**



Source: Heffler et al (2002).

Figure 10.2 **Ten-year National Health Expenditure Projections: Growth in Private Health Insurance and Medicare Spending**



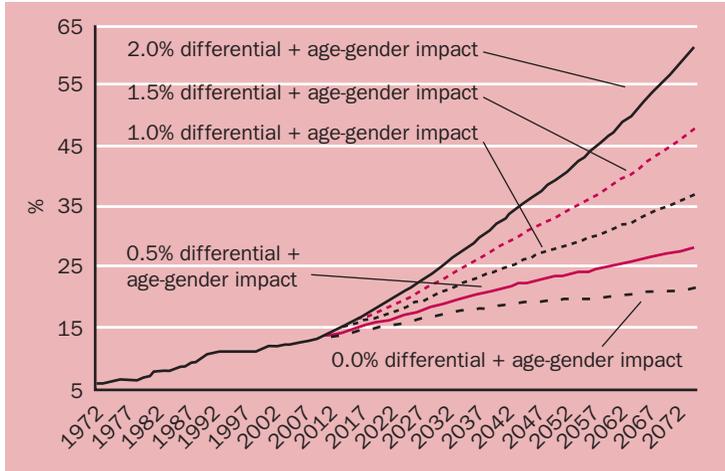
Source: Heffler et al (2002).

ambulatory settings. These impacts cannot be repeated beyond a point. Second, substantial cuts in the Medicare programme enacted by Congress reduced Government spending growth.

Figure 10.2 demonstrates that, historically, health spending in the public sector and the private sector are closely associated. This probably has something to do with equity conditions and highlights the importance of models incorporating interactions between the government and private sectors. One of the first things Congress always wants to know is ‘What’s happening to Medicare spending relative to private sector spending?’

One of the most problematic and difficult issues is that by law the Office of the Actuary is required to make 75-year projections of Medicare spending and trust fund balances. The reason for this is that Medicare is an intergenerational transfer programme. Today’s working population primarily pays for today’s Medicare beneficiaries. In doing such long term projections, we encounter many of the same issues that

Figure 10.3 Long-term Inter-industry Forecasting Tool (LIFT) Projections of Nominal Health Sector GDP Shares Under Various Scenarios



are evident in the Wanless review: very small changes in assumptions regarding productivity and technology have substantial effects on projected health spending and its implications for the economy.

Figure 10.3 presents the scenarios produced using an input/output model. This shows the differences in projections we got as a fraction of GDP as a result of various assumptions regarding the growth in health spending relative to GDP; and the effect of changes in demography, such as age and gender effects. Historically, the differential in annual growth of per capita health care spending relative to per capita GDP has been higher than 2%.

Underpinning these aggregate models of health care spending discussed above are related modelling exercises performed in the Federal Government by another agency that projects workforce requirements for physicians, nurses, dentists and so on. These workforce projections are integrated into the ten-year national health expenditure projections.

The Office of the Actuary also has additional models related to staffing mix, staffing ratios and quality of care. These detailed models of the health care workforce enable us to address some quite complex policy issues. For example, we had to model compensation effects that arise from expanding the staffing ratios in nursing homes in tight labour markets and how that affects wage rates due to the interaction between nursing home and hospital labour markets.

Other macro-level models of physician behaviour used in the Office of the Actuary incorporate micro-level modelling of individual physician behaviour. To predict physician-induced service effects, associated with reductions in physicians' fees, detailed specialty/procedure-specific data are used. Exploration of such models in the UK context may be useful for projections of NHS spending.

Another important focus of US models of health care spending is variations in regional spending among states, metropolitan areas, and counties.

I note that the Wanless Report does not, in general, present the historical data for the variables modelled. It uses a historical baseline and, from this, projects three scenarios out to 2020. It is important, for such projections, that users can see the historical time series for each of the explanatory and predicted variables. The more transparent a carefully crafted model (including input data) is to users, the more confidence they will have in its projections. In particular, if policy makers are to act on the projections, they need to be confident that there is nothing 'underneath it' that they do not understand.

Conclusions

The US cannot offer any magical solutions to modelling and projecting health care spending. We confront many of the same issues that are evident in the Wanless Report projections of health care expenditures over long time periods. Indeed, in one important respect, analysis in the UK is superior: health outcomes are explicitly recognised as a factor to project. Outcomes are extremely important for cost-benefit analysis and for formulating health policy, yet these have typically not played a role in US projection models.

There are, however, some important differences between the approach to modelling health spending in the US and that evident in the Wanless review, and from which future UK modelling exercises might arguably learn.

First, in the US competing models are used by different branches of the Government (executive and legislative), all of which are transparent (to various degrees) and often subject to vigorous, detailed and independent critical review by academics and consultants. This imposes a discipline on the quality and consistency of the projections provided from modelling.

Second, the models of health care spending are themselves built upon and informed by a range of related and sometimes detailed auxiliary models. For example, productivity, workforce requirements and demographic models may be integrated into aggregate models. Models include both macro-level and micro-simulation models. It is not clear from the Wanless review how underlying factors have been modelled there or what the evidence base is for some of the crucial assumptions supporting its projections of NHS spending.

Third, there are aspects of models of health care spending that are important in the US but are not apparent in the UK model; for example the focus on regional variation in health spending that is a function of private sector income and health resources in the US. The NHS simulations assume a constant and small role for the private sector. The difference probably reflects the contrasts between the US and UK health care systems and the different objectives in projecting future health care spending.

REFERENCE

Heffler S, Smith S, Won G, Kent Clemens M, Keehan S and Zezza M (2002) 'Health spending projections for 2001-2011: the latest outlook'. *Health Affairs*, March/April; pp207-218.

Chapter 11

Lessons from macroeconomic modelling in the UK

MARTIN WEALE

This chapter looks at:

- the history and experience of macroeconomic modelling in the UK;
- confusion between ‘modelling’ and ‘forecasting’, and between ‘good’ and ‘bad’ forecasts;
- openness and transparency in public sector forecasting;
- ‘Gresham’s Law’: bad modelling drives out good;
- third party use of models;
- lessons for modelling health sector spending.

Introduction

This chapter is in some sense an outlier compared to the others. Earlier chapters have all been to do with modelling aspects of the health industry. In this chapter, I describe the experience of macroeconomic modelling in the UK, and consider what lessons might or might not be drawn from that for modelling the health sector.

History of macroeconomic modelling in the UK

Until 1982 there was a variety of macroeconomic models in the public sector, funded on an ad hoc basis by the Economic and Social Research Council (ESRC, formerly the Social Science Research Council). The main players were the National Institute of Economic and Social Research (NIESR), the Cambridge Growth Project and the Cambridge Economic Policy Group, the London Business School, the Southampton Group (closed in the late 1970s), and a group in Liverpool. Some of these organisations also drew support from Government departments: the Treasury and the Department of Trade and Industry.

In 1982 the ESRC decided to establish a more structured basis for this work. They set up a macroeconomic modelling consortium, funded largely by them and the Treasury with the Bank of England providing some additional support. The consortium supported large teams at most of the centres and a small team at the Cambridge Economic Policy Group.

A very important point is that this funding also supported a bureau in Warwick which was designed to carry out and facilitate *comparisons* between macroeconomic models. It was hoped at that time that the existence of the Warwick bureau might facilitate some standardisation and would result in the macroeconomic models being used more widely in the academic community. With a few exceptions, that did not happen. Over a period of 16 years, there was a sort of stand off between the macro modellers and other academics, who plainly felt that whatever questions it was they wanted to answer, macroeconomic models did not help.

In the early stages, the modellers had to produce forecasts as well as models, and the UK Treasury took the view this was a complement to its own modelling work. The Treasury had to produce projections and at that stage, 20 years or more ago, civil servants and more particularly ministers did not have the confidence that their projections were better (i.e. more accurate) than other people's. They wanted to see that outsiders' views were either coherent with their own, or to understand why they were different. Certainly Treasury civil servants regarded the work done by the forecasters, and the forecasts produced by them, as a valuable way of checking what they in the Treasury were doing, and therefore the way in which they were setting both fiscal and monetary policy.

Over the next 16 years following the establishment of the consortium the position was gradually eroded. The Research Council took the view, not unreasonably, that *forecasting* as opposed to modelling was not research and therefore should not be funded from research grants.

At the same time there has been rapid technical progress, and the maintenance of models is much cheaper and efficient than was the case in the 1980s when data needed to be inputted by hand.

Computers were then much slower and modelling was terribly time consuming. Technical progress has reduced the staff needed to run and develop models of this sort.

But it is also the case that the modelling groups were not often involved in developing new models for forecasting. They had models which they saw the need to maintain and sometimes develop, but their existing models represented a sort of intellectual baggage they carried with them. Thus it was not logically possible for one of the modelling group to move outside existing approaches and try new, different and perhaps better ideas about how to look at the economy. The modelling framework thus inhibited evolution and made it difficult for the macro modellers to develop new methods. Successful academics are innovators and this was perhaps one reason why the rest of the academic community tended to think that macro modellers were lagging behind best practice.

Another explanation for lack of innovation in macroeconomic modelling in the UK relates to the Treasury wanting standards for comparison. That meant that for the work to be of interest to the Treasury the structure had to be rather similar to that in use by the Treasury. For example, the Cambridge Growth Project was an input/output model: it modelled 40 industries and added up the forecasts for those to give an estimate to what was going to happen to the whole economy. The Treasury was not interested in the 40 industries and so the Cambridge Growth Project was one of the earlier casualties of the four-yearly reviews which took place. This generated an unhealthy degree of standardisation and was perhaps inimical to the sort of modelling innovation that might otherwise have taken place. For example, the consortium supported only a limited amount of general equilibrium modelling.

Contributing to these difficulties was a real confusion between *models* and *forecasts*. *Models* tended then to be thought of (and possibly still are) as being 'right' or 'wrong' because *forecasts* were either 'right' or 'wrong'. To those who appreciate continuous distributions, obviously the forecasts were always wrong: the interesting question was how wrong they were. It seems perfectly obvious to me that you cannot say

whether a forecast is right or wrong; you can only say whether it is good or bad – and you can only establish whether it is good or bad by comparing one forecasting procedure with an alternative forecasting procedure. Further, you obviously cannot come to any judgement on the basis of an outturn in a single year.

However, misunderstandings of this nature are common. For example, from time to time the *Financial Times* has a leader article explaining why forecasts are no use. One of the reasons it once identified for forecasts being no use was that they were particularly inaccurate in uncertain economic times. Now, how you tell that times are uncertain except by seeing that forecasting is particularly inaccurate, I have absolutely no idea.

Models are frequently criticised for not forecasting recessions. If these are rare events, as they are, then it is perfectly possible that they should be the consequence of extreme forecast errors at times when modellers forecast growth rates of only just above zero. The interesting question is not whether one fails to forecast recessions, but whether the forecast errors are so large that, compared to some alternative forecasting model, the process is bad.

So, added to the confusion between forecasts and models, there is fundamental confusion about what is a ‘good’ forecast and what is a ‘bad’ forecast. The position of the macro models was complicated further – although this issue would probably not arise with respect to health sector modelling – by the increasing number of forecasts produced by the private sector. This led to people questioning the need for public support for modelling. Again, this reflects the confusion between forecasts and modelling. Short term economic forecasting is as much a question of judgement as a question of having a model. The essence of having a good model or a bad model is whether it lets you address interesting policy questions in a coherent way.

The modellers did little to help themselves until the mid 1990s. For example, there was no serious attempt to put error margins around forecasts. Modellers were criticised for giving point projections which were described as having spurious accuracy, and the invention of the

word processor meant that we worked in terms of a tenth of a percentage point, instead of halves and quarters, so technical progress made spurious accuracy worse. At the NIESR it had been argued that if we produced confidence intervals then people would think our forecasts were worse than other forecasts and it would damage our reputation.

The first thing I did when joining the NIESR in 1995 was to display probability distributions for growth and inflation. The Bank of England started publishing similar information shortly afterwards. There are obviously questions about how you calculate and display these. When we started doing this, we hoped it would generate a debate among modellers about how best to do this and that some sort of consensus would emerge. Unfortunately that did not happen because we started doing this at a time when the publicly funded modelling industry was declining fairly rapidly. There was little interest in the private sector in these developments, as it had no incentive to do this while users were getting on perfectly happily without it. But to make a contribution to the public environment, forecasts need some sort of probability distribution represented in a coherent way.

As an example of the importance of that, the Chancellor of the Exchequer Gordon Brown, in his May 2002 Budget, forecast a surplus on the current account in five years' time of £9bn. The NIESR forecast a deficit in five years' time of around £9bn. To put this into perspective, this is about one standard deviation from Gordon Brown's forecast. He is saying he thinks there is a 60% chance of him meeting his fiscal target. We are saying we think there is a 40% chance. Put like that, is the difference so great? Maybe we are both saying we do not know very much about the Government's finances in five year's time. This is a fair conclusion and one that should be communicated rather more clearly. When we are pressed on our numbers, our answer is 'People should ask Gordon Brown what he thinks the chance of a number the wrong side of zero is, given what he has projected'. That seems to me much the most interesting question.

I have said earlier that the models became a piece of intellectual baggage. From the point of view of a research body supporting research, what tended to go wrong was that the model became the

goal, instead of the questions being the goal. Having observed this, when I read research proposals that say they want to construct a model of this or that, rather than to answer this or that interesting question, then I am very reluctant to support them. Researchers should be funded to answer interesting questions, not to produce models. Models are tools and not goals: a model may serve to answer an interesting question, it is a means and not an end in itself.

Openness and transparency in public sector forecasts

In 1997 the Bank of England became independent. The Treasury no longer set interest rates, but it was still required, under the 1977 Industry Act, to produce two macroeconomic forecasts a year and it needs to make fiscal projections. The Bank of England did not feel the same need as the Treasury had done previously for an independent view against which to compare its model.

The modelling consortium also lost Treasury support because the Treasury no longer felt it needed it. As a consequence of these developments, in 1999 the modelling consortium shut down. The NIESR is now the only group supported by the ESRC to do research connected to a macroeconomic model. The issues we are looking at include an analysis of the effects of uncertainty and volatility, using our models as a tool to help us.

The Bank of England, for the first two to three years of its independent existence, claimed to have a suite of models and to come to decisions using judgement across them. One would have no objection to that, but as I and others questioned the Bank of England about what their models looked like, the near blank sheets of paper were rather surprising, and the argument seemed to be, 'We can't release any one of our models, because people have to see them overall'.

The Bank of England Act went through Parliament, giving statutory basis to the change that the Chancellor had announced just after the 1997 election. At the Committee stage of the Act, in the House of Lords, Labour Peers threatened to write into the legislation that the Bank of England should publish its models. Instead, the Bank of England came to a voluntary agreement to publish its models but the

work was not completed until 1999 and revised in 2000 (Bank of England, 2000). It was not clear why it took two years to prepare the work if the model was properly documented in 1997. Since then they have commissioned a report (Pagan, 2003) on their modelling work, which made a number of suggestions for improvement.

There is a strong case for arguing that models developed and used in the formation of public policy should be open to scrutiny. Similar concerns exist regarding the Wanless Report on Britain's future health care expenditure needs (Wanless, 2002). As I understand it, the models which produced the forecast spending for each scenario have not been released. I have also been involved in various aspects of the statistical service where the same issue arises. Statisticians produce very short term forecasts, with good reason. Are they willing to explain to people how they produce their forecasts and therefore generate a public debate how those forecasts might be improved? The answer, at least so far, is 'no'.

Despite the appeal of openness, third party use of models needs to be handled with care. Third parties can ask questions from models that the models were not designed to answer. In consequence they produce silly results. This suggests a need to be careful about dissemination and the way that models are used.

'Gresham's Law': bad modelling drives out good

In macroeconomic modelling there has been a form of Gresham's Law in operation: the bad tends to force out the good. There are a number of reasons for this.

First, as I have noted above, good *forecasting tools* are not necessarily good *models*.

Second, private producers of models are more interested in forecasts than in policy analysis. You do see a little policy analysis from the private sector, but only a little.

Third, the private sector has no reason to support technically sound models, and in macro modelling an important issue is model consistent expectation. But nevertheless we have seen from the private sector the suggestion, for example, that if the inflation targets were

reduced interest rates would be higher. They might be so in the very short term, but there is no reason to think this is possible in the long run. High interest rates compensate savers for a declining value of money and are low at times of low inflation. This sort of result would not emerge from a coherent model.

Private sector forecasts may from time to time be *parti pris*. City forecasts of a recession are likely to be somewhat influenced by the fact that the bank would make profits from a reduction to interest rates. In some sense, of course, that is a sensible position for people to be in. If a bank is in the job of taking a position on interest rates, then if it expects interest rates to come down further than the market does, it will make money. That is its job. But how far should you trust an economic forecast from someone in that position?

The question of being *parti pris* is, I think, an important one. There is a further issue in the private sector: people are producing forecasts essentially as an advertising measure. They are under pressure to say things that will attract attention, rather than necessarily because they believe them to be correct.

Gresham's Law is less likely to operate with health sector modelling, but the arguments for an independent assessment are as important in the health sector as they are with macroeconomic modelling. This requires that the data and modelling processes employed in their analysis are made available.

Conclusions: an ideal framework

In macroeconomics, modelling work has undoubtedly contributed to the policy debate; the discussion on the five tests of the Euro is a good example of that (HM Treasury, 2003). It is less clear how far policy in other areas has been influenced by outsiders. Given the profusion of private sector forecasts, there is such a diversity of views that it is easy for the Government to find an outsider somewhere supporting its policy views.

It is difficult to see that the experience with public funding of macroeconomic modelling provides a satisfactory example to be

replicated in the support of modelling health expenditure. There are obvious benefits from having independent comment on Government policy analysis, whether in the macro-economic or health sphere.

Perhaps the best mechanism in both cases would be for parliamentary select committees to support the necessary work. They are not in a position to conduct substantial analysis for themselves. However, if they were to make money available for this purpose, they could then seek academic advice, from a body such as the ESRC on how to allocate that money to meet the need for independent advice. Even though the ESRC has been reluctant to see this sort of modelling as a research activity, they are almost certainly the best people to ensure that funds intended for that purpose are allocated sensibly.

Whether this structure has any prospect of being set up, I am rather doubtful. The US has bodies (such as the Congressional Budget Office) whose function is to provide Congress with advice and information independently of the Government. The suggestion above is for something on a much more modest scale. But the need for Parliament to have access to independent analysis is as great in the UK as in the US.

REFERENCES

Bank of England (2000). *Economic Models at the Bank of England*. September 2000 update.

HM Treasury (2003) *UK Membership of the Single Currency: an Assessment of the Five Tests*.

www.hm-treasury.gov.uk/documents/the_euro/assessment/report/euro_assess03_repindex.cfm

Pagan A (2003) *Report on Modelling and Forecasting at the Bank of England*. www.bankofengland.co.uk/pressreleases/2003/paganreport.pdf

Wanless D (2002) *Securing our Future Health: Taking a Long-term View*. HM Treasury.

http://www.hm-treasury.gov.uk/Consultations_and_Legislation/wanless/consult_wanless_final.cfm

Chapter 12

Discussion: what can we learn from elsewhere?

Topics covered in the discussion of international modelling experience included:

- how are confidence intervals around forecast spending to be estimated and presented? How is uncertainty to be dealt with?;
- would competing models of health care spending, as in the US, give more and better information?;
- should Wanless' model be made public and open to inspection?

Mark Freeland: The issue of confidence levels, raised by Martin Weale is interesting. We have done a lot of thinking about confidence limits for forecasting health spending, including alternative ways to portray uncertainty. However, we do not think there is one 'technically' correct way to develop confidence limits surrounding health care expenditure projections.

Sir John Hicks' book, *Causality in Economics* (Basic Books, 1979), refers to two kinds of probabilities: probability based on risk and subjective probabilities that incorporate uncertainty. He is building upon Frank Knight's concepts of risk and uncertainty. Classical mathematical statistics and econometrics text are based on frequency theory or risk probabilities. The nature of much economic forecasting on the other hand requires a broader concept that includes uncertainty. John Maynard Keynes' book, *Treatise on Probability* (1921) and Harold Jeffrey's book, *Theory of Probability* (1939), discuss this broader concept of probability.

In the US some major private sector forecasting firms take the uncertainty dimension into account by forecasting alternative scenarios and then assigning each scenario or forecast a subjective probability. For example, DRI-WEFA in its October 2002 *US Economic Outlook* publication makes three alternative scenario forecasts: 1) 'Baseline' with a subjective probability of 55 percent, 2) 'A More

Bullish Recovery' with a subjective probability of 15 percent, and 3)' The Recovery Holds, But Barely' with a subjective probability of 30 percent.

Forecasts and projections need to take uncertainty into account in some way. Both the actuaries/economists making projections and the readers of the final report need to have an understanding of how uncertainty impacts on the projections.

The major forecasting project within the Office the Actuary, the annual Medicare Trustees Report, takes uncertainty into account in two different ways. We publish alternative 75 year (as required by statute) scenarios of the projections: High Cost, Intermediate Cost, and Low Cost. We do not provide subjective probabilities for each scenario, however. In addition, in an Appendix to the Medicare Trustees Report, we publish '95-Percent Projection Intervals' around the Intermediate Cost projections for ten years out. These '95-percent projection intervals', developed using Stochastic projection intervals, are carefully explained (see <http://www.cms.hhs.gov/publications/trusteesreport/>).

Martin Weale: Problems can arise if you think there are bimodal distributions: will a particular event occur or not? I do not think we have a good methodology for addressing that sort of problem.

Secondly, our forecast errors and confidence intervals are based on our past performance. But if you have just started making 75 year projections you cannot say very much about your past performance at that.

Sean Boyle: I was interested in Martin Weale's comments regarding parliamentary select committees. I have been advising the Select Committee on Health for over ten years now, and at one time we were continually asking to see the Department of Health's expenditure model. The response to that has always been that 'We can't show you the model because it is the model we use in our discussions with the Treasury'. I guess a more modern response is 'If we reveal our hand in terms of negotiations with staff groups that could be a problem'. I was just wondering what kind of answer to that response Martin Weale might make.

Martin Weale: As someone very much on the outside rather than the inside, I think open government, like western civilisation, would be a good idea. There is obviously something very badly wrong with the structure if the Department of Health needs a secret model in order to negotiate with the Treasury. The Treasury quite possibly or very probably has its own secret model. If the model is needed from one side, a model of that sort must also be needed from the other side. But the Department of Health's case might be more convincing to the Treasury if it explained fully how it arrived at it. Similarly the Treasury's rebuttals might be more convincing if the Treasury explained fully how it arrived at them. I should have thought you were therefore more likely to get a consensus and therefore more likely to have public money sensibly used than where departments are hiding things from each other.

Can Mark Freeland tell us what kind of dialogue takes place between the two main US modelling groups? Do you learn at all from one another's experience and one another's critiques?

Mark Freeland: The Office of the Actuary and Congressional Budget Office (CBO) are not always totally independent. There are sometimes discussions between the Office of the Actuary and CBO when each group is developing its own baselines and alternative scenarios. Once forecasts for each group are publicly released, Congress, the executive branch, and special interest groups may well request a detailed explanation or reconciliation for what factors are causing the differences. This is a learning experience for both the Office of the Actuary and CBO and may lead to future forecasts being more similar if technically it seems advisable.

One reason that CBO may have higher projections is that it often assumes higher economy-wide price inflation than is assumed by the Executive Office of the President and used by the Office of the Actuary. So it is an important factor to examine and adjust for. After this adjustment, how close are the two forecasts to each other? Often the public dialogue is associated with different behavioural assumptions for the elasticities of demand and supply. For example, in pharmaceuticals a really big question is the assumed elasticity of

demand. When additional insurance coverage lowers the out-of-pocket price to consumers, what is the incremental utilisation of drugs, and the effect on pharmaceutical pricing? In the longer run how does that affect research and development and the new drugs coming on line?

So if the Office of the Actuary and CBO have substantial differences in the technical behavioral relationships and trends in their models, it can actually slow down or stop public policy. The policy makers may have difficulty deciding which forecast is likely to be more correct and may fear that the budget may not be able to handle the more expensive scenario. In some cases this may be prudent.

Adrian Towse: When this steering group of OHE, King's Fund and Centre for Health Economics were planning this meeting we had two purposes. One was to look at the components in the Wanless model: the key assumptions, to try to understand some of the interactions and how important they were for policy purposes. That is what we tried to do in the first three sessions.

It does seem to me a couple of things have come out clearly. First, it is vital that the Wanless model is made available. It is, of course, a Treasury model, so it is not fair to blame the Department of Health for the fact the model has not been released. Department of Health economists were involved, but technically they were seconded to the Treasury for that purpose. So it is the Treasury's call.

Going back to Martin Weale's point about the Bank of England coming up with a model after the event. I think there is an issue about whether we want ex ante, 'It is going to be clear that the model will be in the public domain', or whether we are talking about ex post, 'That is very interesting. Why don't you give us the details'. I know when I was in consultancy if the client wanted the model handed over to them at the end of the exercise then you approached it very differently than if they just wanted to see the results and ask you a few questions about it.

So I think there is an issue about what it is reasonable to get out of the Wanless exercise retrospectively, but nonetheless it would be very

helpful to have more information. I think the key point where it is interesting is that Martin Weale and Mark Freeland keep a degree of consensus, Mark is saying that in the US experience the key thing is that the political and the executive both have their models and both of those are in the public domain. That is what happens in the US, and Martin is saying it should happen in the UK in the context of macroeconomic policy, and it is not. This is a potential model which would give academic groups and others the opportunity to look at those models and to question some of the assumptions in order to have an informed debate.

Conclusions

JOHN APPLEBY, DIANE DAWSON and NANCY DEVLIN

'All models are wrong, but some are useful.'

George Box

In sharp contrast to the amount of detailed attention and analytical effort devoted in the UK to decisions about the allocation of resources between regions (e.g. RAWP) and between therapies (e.g. NICE), there have been relatively few attempts systematically to grapple with the most high-level, fundamental allocative decision: how much should be spent on the NHS? Decisions have, in the past, been based either on ad hoc incrementalism (this year's spending = last year's spending, with tweaks) or the dubious logic of international comparisons of health care spending.

Derek Wanless's review of future NHS spending therefore represents a very important step forward. The headline answer the review produced was that more – much more – needs to be spent in order to enable the NHS to provide a 'world class' health service. This may come as no surprise to those within the NHS who have been calling for increasing funding for many years and it is consistent with the preferences of the general public as expressed in surveys, which have shown consistent support for increasing spending. But, for the first time, specific amounts of increased spending have been justified by a comprehensive review that links spending requirements to specified objectives for the NHS (the Wanless 'vision'). The Wanless recommendations have had a direct impact on Government spending decisions: the spending review carried out in 2002 set out a future expenditure path for the NHS directly in line with Wanless' 'fully engaged' scenario. The first report has prompted a follow up review, published in February 2004 requested by the Chancellor to inform the 2004 spending review and setting out more detailed work on the public health investment and other changes needed to bring about the fully engaged scenario.

The Wanless attempt to model and plan future health care spending is a remarkably ambitious undertaking, involving complex interactions

CONCLUSIONS

combined with a certain amount of crystal ball gazing. The Wanless report itself provides a tantalizing glimpse of the way the models work and the variables that drive spending under each scenario. But, as the models have never been released, independent analysts have been unable to determine their veracity, or to find ways of building on and strengthening that work. The Wanless report is also remarkable in another way: in addition to projecting the spending requirements of future policy, it also appears to be the first occasion on which the resource implications of some extant policies have been costed. For example, quality improvement initiatives such as national service frameworks and waiting times targets clearly have substantial impacts on resource use – it seems extraordinary that such initiatives can be introduced without careful scrutiny of their costs and value for money.

While there is no doubt as to the importance and impact the Wanless review has had, as this collection of chapters reveals, there is also no doubt that the review's assumptions, methods, models and data could be improved. In this final chapter we¹⁵ draw three broad conclusions about future attempts to provide the public and policymakers with evidence-based planning of future NHS spending. First, there is a prior argument for repeating a Wanless-type modelling exercise at regular intervals; second, there are ways in which the modelling approach could be improved; and finally we note the importance of raising awareness amongst the public and taxpayers of the issues involved in such exercises and the possibility of greater involvement of the public in this work.

A Wanless exercise should be conducted on a regular basis

The Wanless review could be considered a one-off attempt to set out the broad future path for NHS spending. However, as the original review noted, there are a number of reasons why such an exercise should become a regular activity. Future resource requirements will

¹⁵ This section reflects the views of the editors and not necessarily those of other contributors.

CONCLUSIONS

change in the light of: changing health care needs; changes in medical technology; changes in the way health care is delivered; and the accumulation of greater knowledge and information concerning key aspects of the models and assumptions used to plot future health care spending paths.

More particularly, as Mark Freeland notes in the context of the work of the US Office of the Actuary and the Congressional Budget Office, regular updates of long term spending forecasts are used as a baseline to analyse the impact of proposed changes to the health care ‘benefit package’ and the impact of cost trends. In the UK context, a well-developed health care spending model could provide similar opportunities to analyse not only policy changes but also changes in basic assumptions underlying the model. For example, technical change in health care and the diffusion of new technologies is continuous (via NICE, medical practice, etc) so the baseline is constantly changing. This implies a need not only to get the model right in the first place (cf Dawson and Towse, above) but also to understand how relationships in this particular area change over time.

Moreover, and on the immediate NHS agenda, new policies such as patient choice and the separation of elective surgical treatment from A&E and emergency medical services are likely to change costs and medical workforce demands – to name the most obvious.

More broadly, a regular review of future spending would inject some science into debates about future funding and take some of the political heat out of arguments over spending.

Overall, of course, there cannot be a technical solution to the question of NHS spending, but considered modelling can be useful not only in separating fact from political judgement, but also in highlighting important issues such as productivity measurement (cf O’Mahony) and the difficult problem of quantifying quality.

Improving Wanless

If a Wanless-style review of future spending is to be undertaken at regular intervals there is a need to improve its methodological basis.

CONCLUSIONS

A first step should be to publish the methods used to model forecasts in the first Wanless exercise. Open dissemination of the original methodology should encourage independent researchers to seek ways of improving on all or parts of the approach.

Neither the Department of Health nor the Treasury has the resources (or, perhaps, the independence or public credibility) needed to examine alternative data and methods over the large range of topics covered by Derek Wanless. Macroeconomic modelling in the Treasury, for example, has benefited from the availability of independent modelling and forecasting (cf Weale). Similarly, modelling future health care spending should be an open, joint exercise, involving not only the Department of Health, its counterparts in Northern Ireland, Scotland and Wales, and the Treasury, but also academics and others with skills and expertise in areas pertinent to the exercise.

Wanless's particular approach was to specify a 'vision' for the NHS in 20 years' time (for example, inpatient waiting times of two weeks, comprehensive national service frameworks) and then cost it. While such an approach has some merits – not least as a pragmatic response to the need to carry out the review to a tight timetable – a preferable method would be to establish a proper baseline against which policy change could be compared.

A future review should better reflect the main business outcome of the NHS i.e. health (rather than process and activity)¹⁶. It should therefore include analysis of the health gain associated with current policies and proposed changes. There is then the related issue of the value society would want to attach to such gains in health, not just in the aggregate, but also in terms of its distribution. There will also be questions concerning the values attached to gains in other areas of performance (such as waiting times, improved facilities etc.) and the trade-offs that might be necessary in achieving competing desired objectives (cf Yates). Data on willingness to pay for health gains and other related benefits could provide a guide to the size of future

¹⁶ This is not to say that intermediate performance measures are of no importance but to refocus on the health service's main objective.

CONCLUSIONS

investments in the NHS, while also providing indications of priorities for investment within the NHS.

An important area in need of much greater research and understanding is productivity (cf O'Mahony). In the absence of any significant evidence base, the Wanless review necessarily made assumptions about the scope for productivity improvement. Small deviations from these assumptions can have large absolute effects on forecast expenditure. It is therefore vital to understand more about the potential for improvements in productivity.

Finally, there was little or no evidence base for the probability of any of the three scenarios described by the Wanless review actually materialising. The follow-up review (published in February 2004) was asked by the Chancellor to address this question for the 'fully engaged' scenario. However, while there exists some evidence for effective public health interventions (cf McPherson), it is not overwhelming. As with productivity, assumptions about the population's future health (and hence health care needs) and the extent of their 'engagement' with their own health and health services lead to large differences in future spending requirements. It is important, therefore, that knowledge in this area is improved.

Involving the public and public awareness

Finally, there is a need to engage the public not only in their own health, but in the very exercise of modelling future expenditure on health care. A straightforward reason for this is the fact that the NHS is tax funded and as such requires a democratic input to future spending decisions over and above that afforded by voting. This is, of course, a potential argument for any government spending of tax payers' money. However, the scale of NHS funding, the fundamental importance of health and health care and the political heat that funding debates generate perhaps makes the NHS a special case.

There are, however, additional reasons for engaging the public in this matter. As already alluded to, it is not possible to reduce a review of future spending to a technical exercise (albeit a very complicated one). There are questions of value (particularly in a non-marketed

CONCLUSIONS

service such as the NHS) which are best addressed not by economists, politicians or Treasury policy wonks, but by society more generally. There are also prior questions about the sort of NHS we want – the ‘vision’ – and what we are prepared to sacrifice in order to achieve it. No amount of modelling, no matter how sophisticated, will help answer such vital questions – rather, they require consideration by those to whom the NHS is ultimately accountable: the public and patients.

Appendix

Participants in the workshop held on 13th November 2002

Chairs:

John Appleby	King's Fund
Tony Culyer	University of York
Jon Sussex	Office of Health Economics
Alan Williams	University of York

Speakers:

Diane Dawson	University of York
Mark Freeland	Office of the Actuary, Center for Medicare & Medicaid Services
Alastair Gray	University of Oxford
Stephane Jacobzone	OECD
Klim McPherson	University of Bristol
Mary O'Mahony	NIESR
Adrian Towse	Office of Health Economics
Martin Weale	NIESR
John Yates	University of Birmingham

Attendees:

Martin Anderson	ABPI
Rifat Atun	Imperial College Management School
Mark Bassett	BUPA
Sean Boyle	LSE Health & Social Care
David Buck	Department of Health
Martin Buxton	HERG – Brunel University
Neil Craig	NACF Glasgow
Amanda Croft	The NHS Confederation
Nancy Devlin	City University
Stephen Dunn	Department of Health – Strategy Unit
Julien Forder	Department of Health – Strategy Unit
Hugh Gravelle	University of York
Chris Ham	Department of Health – Strategy Unit
John Henderson	Department of Health
Stuart Hurst	Pfizer

APPENDIX

Duncan Innes	Boots Health care Development
Benedict Irvine	Civitas
Trevor Jones	ABPI
Panos Kanavos	LSE
Julian Le Grand	LSE Health & Social Care
Lindsey Lewis	Department of Health
Nora Lukova	BUPA
Alan Maynard	University of York
Barry McCormick	Department of Health
Jorge Mestre-Ferrandiz	Office of Health Economics
Richard Murray	Department of Health
Charles Normand	London School of Hygiene and Tropical Medicine
David Parkin	City University
Clive Pritchard	Office of Health Economics
James Raftery	University of Birmingham
Peter Robinson	IPPR
Ray Robinson	London School of Economics
Luigi Siciliani	OECD
Matthew Swindells	Cabinet Office
Andrew Walker	Robertson Centre for Biostatistics, University of Glasgow
Nick Wells	Pfizer Ltd
Tessa Westbrook	Office of Health Economics
Tim Williams	Merck Sharp & Dohme
Frank Windmeijer	Institute for Fiscal Studies
Melanie York	Office of Health Economics
Nick York	Department of Health