



FEBRUARY 2004

• The NHS Improvers

A study of the Commission for Health Improvement

Patricia Day and Rudolf Klein

The NHS Improvers

A study of the Commission for Health Improvement

Patricia Day and Rudolf Klein

This study of the Commission for Health Improvement's inspection process was undertaken by independent researchers with a grant from the King's Fund. The authors alone are responsible for the views expressed.

About the authors

Patricia Day is a Senior Research Fellow at Bath University. Her publications include, as co-author, *How Organisations Measure Success* (Routledge 1992), *Why Regulate?* (The Policy Press 1996) and *Auditing the Auditors* (The Nuffield Trust 2001).

Rudolf Klein is Emeritus Professor of Social Policy at Bath University, and Visiting Professor at the London School of Economics and the London School of Hygiene and Tropical Medicine. He is author of *The Politics of the NHS* as well as other books, and has published numerous articles in a variety of academic and other journals.

Published by:

King's Fund

11–13 Cavendish Square

London W1G 0AN

Tel: 020 7307 2591

Fax: 020 7307 2801

www.kingsfund.org.uk

Charity registration number: 207401

First published 2004

Contents

Tables	iv
Acknowledgements	v
Foreword	vii
Executive summary	1
Inventing a new institution	1
Inspection methods	1
Impact and reactions	2
Conclusions and implications	3
Introduction	5
The context and launch of CHI	6
Dilemmas of regulation	8
CHI invents itself	11
CHI in action	17
How solid are the pillars?	17
Searching for the patient	19
The search for consistency	23
Scoring the reviews	25
To whom is CHI talking?	29
What difference has CHI made?	31
The NHS on CHI	37
General views about CHI	38
The calibre of review teams	39
Why do reviewers see mainly bad things?	40
Pre-visit data	40
Tools and instruments	42
CHI's administrative and organisational standards	44
Conclusions	46
Appendix: Translating CHI reviews into evidence about London and the NHS	50
What do CHI reports explain, other than clinical governance?	54
Trust profile issues in CHI reports	58
Appendix conclusions	61
Bibliography	64

Tables

Table 1:	CHI scores v. scores in patient surveys: outpatients survey	22
Table 2:	CHI scores v. scores in patient surveys: emergency department survey	23
Table 3:	The distribution of CHI scores	27
Table 4:	Total clinical governance scores for London, and for all trusts	52
Table 5:	London clinical governance scores	53
Table 6:	Clinical governance scores of all trusts	53
Table 7:	Clinical governance scores, by geographical area	54
Table 8:	Most frequently mentioned reviewer impressions in London acute and specialty trusts	55
Table 9:	Most frequently mentioned issues in Cheshire and Merseyside acute and specialty trusts	56
Table 10:	Background and profile information most frequently mentioned in London	59
Table 11:	Background and profile information most frequently mentioned in Cheshire and Merseyside	59

Acknowledgements

Patricia Day's work was supported by a grant from the King's Fund, for which we are grateful. We also thank CHI for its co-operation in giving us access to documents and sparing us the time of senior staff. Similarly, we would like to express our appreciation of the help given by a large cast of Department of Health and NHS personnel. In all cases, our policy was to interview on a 'lobby basis' – in other words, to use the material generated by the interviews without identifying the specific source, so we do not acknowledge by name any of the many individuals who helped us, but our thanks go out to all those who contributed. In all, we conducted some 60 individual and a further 10 group interviews.

Foreword

This King's Fund grant-funded study brings an independent and critical eye to the Commission for Health Improvement's (CHI's) short lifespan. The work brings scrutiny to bear on health policy in practice. Generating informed debate over the future shape of the UK health care system transcends vested political or institutional concerns and requires independent analysis conducted in the public interest.

As soon as new approaches to health care are adopted they evolve and adapt. As CHI is about to be superseded by the Commission for Healthcare Audit and Inspection (CHAI) there is merit in asking how experience could inform the next round of change. This report points to five areas where the inspection process for hospitals could be improved. In short, the process needs to:

- establish clearer standards with a stronger conceptual foundation
- develop methods which make it easier to compare quality of health care across organisations, services or localities
- develop the skills required to address issues of effectiveness and value for money
- establish more effective ways to target proportional inspections on different organisations
- develop ways of assessing the quality of care for those patients who require integrated health care from across a range of local organisations.

While complimenting CHI on the achievement of an enormous inspection task, the report provides a critical commentary on the inspection process. It challenges the conceptual framework for inspection as being too focused on a 'taxonomy' of quality rather than an appreciation of the underlying 'dynamics' of high quality organisations. It highlights a lack of analytical purchase on important but nebulous aspects of 'good organisations', such as culture and leadership. And, in the absence of a more sophisticated framework for assessment, the authors question the process by which 'a rag bag of reflections' (including variable attempts to capture the patient's experience) are converted to global and quantified assessments.

The report also provides a specific analysis of CHI's work in the capital, comparing inspection reports from different regions of the country and demonstrating the complex staffing problems with much higher vacancy and turnover rates for London – as well as the consistent concern of the inspectors over the poor cleanliness of London's hospitals. However, there appears to be no direct relationship between such problems and resource levels.

Underlying the work is one question that provides a consistent touchstone for the analysis: for whom are these inspections conducted? The authors rightly question the espoused aspiration to create reports that are directly useful for the public, pointing out that people either want to know about specific services or have an interest in the totality of care in their locality (rather than judgements over single institutions). The authors argue that more accessible and nationally comparable information will be needed at the level of individual services.

This report provides the first tangible and substantial analysis of the CHI inspection process in practice. It informs timely debates over the purpose, impact, mechanics and politics of inspection. It places the question of proportional and effective regulation as

a mechanism for improving local health (rather than a narrow, institutionally defined concern with systems to ensure health care quality) back at the top of the health policy agenda.

The King's Fund supports the authors in hoping that this paper will spark the constructive reflection befitting this further time of change to the overarching regulatory structure for the UK health system.

Steve Dewar
Director of Health Policy , King's Fund

Executive summary

Inventing a new institution

The Commission for Health Improvement (CHI) was launched in November 1999 as the monitor of quality in the National Health Service. It will be wound up in April 2004, when its responsibilities will be passed to the newly created Commission for Healthcare Audit and Inspection (CHAI). This three-year study analyses CHI's methodology, activities and impact. Our aim was to use CHI's experience to illuminate the challenges and dilemmas of inspection in the NHS, rather than to provide a comprehensive evaluation of the commission's work. Our focus was on CHI's reviews of the acute trusts, excluding its many other responsibilities.

Inherent in CHI's creation were some tensions. First, the legislation that set it up defined its role as being to report on the 'arrangements' of NHS trusts for monitoring and improving the quality of their services – in other words, whether or not they had set up an effective system of clinical governance. The expectations of ministers, on the other hand, was that CHI would report on the quality of the services, although the assumption that good clinical governance could be equated with good performance rested on an act of faith. Second, ministers saw CHI as an inspectorate, informing the government and the public alike, based on the model of OFSTED. In contrast, CHI, anxious to make itself acceptable in the NHS, repudiated the inspectorate label, insisting that its role was to help trusts to improve, not name and shame.

The tensions between ministerial expectations and CHI's self-perception persisted throughout its existence. The relationship between the Department of Health and the commission was one of mutual exasperation. CHI had to accept – albeit resentfully – the department's role in setting its performance targets (in other words, the number of annual reviews to be carried out), but it guarded its independence jealously. Conversely, the department felt frustrated by what it saw as CHI's resistance to its legitimate demands, and its exclusion from detailed discussion of CHI's methodology.

CHI invented itself from scratch. By 2003, it had built up a headquarters staff of more than 400 administrators, review managers and analysts, with a budget nudging £33 million, and had recruited and trained more than 1,000 part-time reviewers. It also successfully achieved the overall target that the Department of Health had set for it – to complete the inspection of all trusts within a four-year cycle – as well as carrying out the extra responsibilities thrust on it during its existence.

Inspection methods

The Department of Health set CHI the task of judging the performance of trusts on seven dimensions – the so-called seven pillars of clinical governance:

- patient involvement
- risk management
- clinical audit

- staffing and management
- education and training
- clinical effectiveness (formerly research and effectiveness)
- use of information.

These pillars provided the framework of – and the focus for – its inspections. However, starting without any explicit standards or criteria, CHI had to develop its own methods for assessing the performance of trusts.

The methodology that developed has been refined but has not changed in its essentials over the years. It has three stages:

- Trusts are asked to supply data about their own performance for processing by CHI's analysts.
- A CHI team, made up of a mix of NHS professionals and lay people, conducts an inspection that involves interviews with staff and patients.
- A review report is published and an action plan is agreed.

The aim of the inspection is to examine not only the structure of clinical governance, but also the extent to which it is implemented throughout the organisations.

CHI's reports score each trust on each of the seven dimensions, and the total scores are subsequently incorporated into the 'star ratings' exercise. However, not all the pillars are useful proxies for quality of service performance. Each pillar carries the same weighting, though there is no evidence about their relative contribution to the quality of the delivery of services to patients. A further problem is that there sometimes appears to be a discrepancy between the text of the report and the score awarded.

From the start, CHI has stressed that 'the patient experience' lay at the heart of its approach. However, in practice, CHI has found it difficult to devise robust methods for achieving this aim, given the heterogeneity of trusts and of patients. For example, the patient involvement dimension is not synonymous with patient-centred service delivery. Furthermore, patient involvement incorporates two components without any logical connection: the extent to which patients have a say in their own treatment, and the extent to which a trust engages in consulting users and community representatives.

Like other inspectorates, CHI has had difficulty ensuring consistency between review teams. CHI has developed an elaborate coding system for recording the evidence collected by the teams, and for cross-checking the impressions gathered. Even so, reports are only consistent in being inconsistent. They vary in the issues identified, in the way the evidence is presented and in the analysis – or lack of analysis – of data.

Impact and reactions

One of CHI's goals was that by 2004 most of the general public would be aware of its work. This has not been achieved. In contrast to OFSTED, most of the public remain ignorant about CHI. Given the general absence of choice, and the fact that potential or actual users are likely to be interested in the quality of specific services rather than institutional governance, there may be little incentive for the public to pay heed to CHI's activities.

CHI's main measure of success was that there would be 'demonstrable improvements in the quality of care' provided by reviewed trusts. In practice, it is difficult to measure quality of care. Even if this were an option, it is impossible to disentangle the impact of CHI inspections from all the other factors, such as extra resources, that affect changes in service provision or improvements in quality. The way CHI's prescriptions for improvements (action plans) are implemented is monitored by the strategic health authorities, rather than CHI. But the trusts hold the view that on the whole, these plans do no more than ratify proposals for improvement that were already on the agenda – although they may change the degree of urgency and order of priorities.

Interviews with trust staff suggest that CHI's main impact lies in the way that inspections change institutional dynamics. The prospect of an inspection tends to prompt institutional self-examination, concentrating management's minds on clinical governance. In short, CHI's role is largely catalytic. In addition, CHI inspections give extra leverage to those institutional actors who are seeking to change organisational practice and attitudes.

Trusts vary in their views about CHI inspections, with reactions ranging from the enthusiastically positive to the critically ambivalent. Overall, CHI has succeeded in making the notion of inspection acceptable in the NHS, overcoming initial suspicions and some hostility. However, there is also criticism on specific issues – in particular, the use of anecdotal evidence in reports, the lack of consistency, and what trusts see as excessive emphasis on the negative.

The attitudes of trusts are also influenced by how the review teams have performed. Here, the main complaint has been that members of review teams often lack the required seniority and experience. CHI has sought to address this problem with an energetic recruitment drive, but its solution probably depends on the NHS giving greater priority and recognition to participation in inspectorial activity.

The trusts also complain about CHI's collection and interpretation of performance data. Trusts see themselves swamped by demands for data, while also often being critical of the way in which the information is subsequently interpreted by CHI's analysts. Their comments about the amount of data required are difficult to interpret. While CHI was undoubtedly initially fishing for data without a clear idea of the key elements of performance, the difficulties that some trusts have had in meeting data requirements may reflect their own lack of capacity to organise and handle information. Nevertheless, in response, CHI has slimmed down its demands for data, as well as streamlining the process of inspection.

Conclusions and implications

Overall, CHI is an example of successful institution-building. Its replacement by CHAI should be seen not as an execution, but as policy adaptation. The problems of CHI's methodology largely (though not wholly) reflect the fact that it has to work within the framework of clinical governance. CHAI's remit will be to look at the quality of care, rather than clinical governance as such, and to inspect against standards. This represents a rebalancing of the regulatory task, which will give the new body greater freedom to develop its strategy and methods.

Unlike CHI, CHAI will have the flexibility to adopt the principles of targeting and proportionality in devising its inspection programme. The success of such a strategy will depend crucially on developing a parsimonious set of indicators that signal actual or impending problems and identify chronically under-performing organisations or services. The act of faith represented by the seven pillars of clinical governance needs to be translated into an empirical analysis, in which key aspects of structure and process are related to good outcomes.

CHI was not set up to provide comparative information about the performance of trusts. This has further limited the usefulness of its reports to policy-makers, purchasers and the public. The aim of the new inspectorate should be to provide information that is both consistent and comparable. This will be all the more important if the Government succeeds in implementing its goal of patient choice. Equally important, CHAI should also be able to supply consistent information about the performance of the NHS as a whole.

CHI's reports provide no analysis of how trusts manage their resources. It is, therefore, impossible to know whether any shortcomings reflect inadequate resources, or poor use of adequate resources. For this reason, the transfer to CHAI of the Audit Commission's responsibility (and staff) for studies of the efficiency, effectiveness and economy with resources should be a great source of strength for the new body.

CHI's inspections have been institution-specific, but the emphasis in the NHS is increasingly on promoting the flow of patients between different sectors and institutions. Here, the population-focused methodology developed jointly by CHI and the Audit Commission for reviewing the implementation of national service frameworks may provide a model, but there is a risk that this may create a vacuum of accountability. If there are failures of co-ordination, who is to be held responsible?

CHI has made much of being a learning organisation, and has indeed done a great deal of internal evaluation and adaptation of its methods and processes. Its record, however, suggests that it may have excessively equated learning with introspection, and with commissioning evaluations in the academic mould while neglecting more informal (and quicker) ways of generating feedback. CHAI's aim should be to establish a dialogue with the reviewers and the reviewed alike from the start.

From the start, CHI was under pressure to deliver. Its record suggests that this sort of pressure may be counter-productive, producing an attitude of protective self-defence, and inhibiting experiment and debate. As CHI has demonstrated, developing inspectorial strategies and methodologies is an evolutionary process, so it is important that CHAI should be allotted time for doing so.

Introduction

In April 2000, the Commission for Health Improvement started its work programme, having been formally launched six months earlier by the prime minister. In April 2004 its existence will come to an end, and its functions will be absorbed by the newly created Commission for Healthcare Audit and Inspection, with its larger and wider responsibilities. This, then, seems to an appropriate time to report not only on how CHI set about its task, but also on what its experience tells us about inspection in the National Health Service. So this paper has two aims: to identify and analyse the challenges faced by CHI and the way it responded, and to draw out some implications for the future.

The paper is based on a three-year study of CHI using a variety of strategies:

- interviewing policy-makers and CHI executives and staff
- observing meetings of the commission and its boards
- analysing the reports produced by the commission
- interviewing the inspected – representatives of the trusts reviewed by CHI – as well as the inspectors.

The result is more than an obituary because our interest is not just in the life of an NHS institution, but also in what that life has to tell us about the style and methodology of inspection in the NHS. It is less than an evaluation because it makes little sense to try to draw up a balance sheet of CHI's achievements and failures. CHI's impact cannot be disentangled from all the other factors bearing on the performance of the NHS in a period of rapid change. The paper also has the less ambitious, but perhaps more useful, goal of helping the reader to draw up his or her own balance sheet. Given different perspectives, there will inevitably be different verdicts.

One limitation of our study, and another good reason for not offering this as an evaluation of CHI in the round, must be noted. Our focus in this report is on CHI's clinical governance reviews of acute hospital trusts. This was the first task allocated to CHI when it was set up, and when we started our project. Subsequently, CHI's work programme and responsibilities expanded greatly, to cover:

- primary care
- ambulance and mental health trusts
- conducting investigations
- monitoring the implementation of national service frameworks
- taking over responsibility for the 'star-rating' exercise.

However, while CHI's activities increased in scale and scope, research resources remained static. CHI's budget was elastic, but ours was not. So, while we note the various dimensions of CHI's work, we make no attempt to be comprehensive in our analysis of its activities. To this extent, we understate CHI's achievements.

The context and launch of CHI

The Commission for Health Improvement had a mixed parentage. Its creation reflected a number of different concerns and policy goals that were explicit and implicit in the 1997 white paper published by the incoming Labour government within months of its taking office (Secretary of State for Health 1997). There was a concern to ensure 'greater consistency in the availability and quality of services' by defining and monitoring national standards. There was also an emphasis on achieving greater accountability by using a strengthened battery of performance measures to give greater public visibility to the activities of the NHS, as well as informing central government.

If the white paper committed itself to apparently contradictory aims – strengthening central control on the one hand, while also promising to devolve responsibility to health care providers on the other – then perhaps new regulatory institutions could square the circle by informing the managerial hierarchy of the NHS, and providing tools for it, without being part of that hierarchy: a hands-off form of control. The creation of CHI neatly fitted into this policy pattern. The new institution promised to monitor local systems of quality assurance, to provide greater accountability, and to illuminate the opaque depths of the NHS for ministers.

There were other influences. Over the previous decade or so, the welfare state had increasingly been transformed into the regulatory state (Day and Klein 1987a), with an explosion in audit and inspectorial activities (Power 1997). For example, in 1990 the Audit Commission's remit had already been extended to the NHS. So the creation of what was in effect, if not in name, an inspectorate for the NHS – in other words, CHI – was riding the existing wave of regulatory enthusiasm. At the same time, there was increasing concern about quality: in opposition, Labour had flirted with the idea of extending the Audit Commission's remit beyond the 'three Es' (economy, efficiency and effectiveness) by reconstituting it as a quality commission.

Soon after the Labour government came into office, the General Medical Council started its hearings of the case against three Bristol doctors, prompted by the deaths of a number of babies and small children following cardiac surgery. Quality of care was high on the political agenda, and was kept there by a series of other cases demonstrating lack of competence by doctors. Further, the notion of continuous quality improvement, imported from industry and expounded by its missionaries cross-nationally, was becoming increasingly influential, and provided intellectual backing for political concerns. In turn, the notion of continuous quality improvement was translated into the concept of clinical governance as the means for bringing about the desired changes in the culture of the NHS (Scully and Donaldson 1998) – a concept that largely shaped CHI, as we shall see.

However, CHI was only one element in the Government's overall design (Secretary of State for Health 1998). Central to the design was the production of 'clear standards of service'. Accordingly, a National Institute for Clinical Excellence (NICE) was set up to produce guidelines based on evidence of clinical- and cost-effectiveness. Further, a set of national service frameworks, providing blueprints for the organisation and delivery of specific services, were to be produced.

At the provider level, a statutory duty of accountability for quality was placed on the chief executives of trusts. To carry out this duty, they were required to develop a clinical

governance system designed to monitor clinical care and promote quality improvement activity. To ensure that this system actually worked as intended, CHI was created. Its role would be to 'independently scrutinise local clinical governance arrangements to support, promote and deliver high quality services, through a rolling programme of local reviews of services' (Secretary of State for Health 1998, p 52). It would also monitor the implementation of the national service frameworks and the guidance from NICE.

The Health Act 1999 translated these aims into legislation, and gave CHI three main functions:

- to conduct reviews of, and report on, 'the arrangements' by primary care trusts or NHS trusts for monitoring and improving the quality of health care for which they had responsibility
- to carry out investigations into 'the management, provision or quality of health care' in cases of serious service failure
- to review the implementation of national service frameworks in collaboration with the Audit Commission. (However, nothing explicit was said about the implementation of NICE guidelines in the legislation, nor in the accompanying press handouts and explanatory notes).

CHI was also to provide advice and guidance to NHS organisations and staff. Subsequently, the NHS Reform and Health Care Professions Act 2002 made CHI responsible for publishing an annual report on the state of the NHS, and created the Office for Information on Healthcare Performance within the commission. The latter was to be responsible for:

- producing the annual NHS 'star ratings' for trusts
- commissioning national clinical audits
- conducting annual staff and patient surveys.

CHI was also subsequently given some responsibilities for monitoring child protection procedures.

CHI was formally established as an executive non-departmental public body on November 1, 1999. Much was made of its independence. However, there were limitations on that independence. The Department of Health determined its budget and, in the words of the 1999 Act, 'The Secretary of State may give directions with the exercise of any functions of the commission' (Health Act 1999, pp 25–30). So there was a certain ambiguity about just how independent it would be in practice, and this became a source of tension between the department and CHI in future years. Nor was this the only tension implicit in the circumstances of its launch.

The christening ceremony for CHI was a drum-banging occasion, with a speech by the prime minister. In his speech, the prime minister compared the commission to OFSTED – the inspectorial body for schools. But CHI's newly appointed chairman and director Dame Deirdre Hine (previously chief medical officer for Wales) and Dr Peter Homa (previously chief executive of an NHS trust) were at the same time distancing themselves from any suggestion that the new organisation would be anything like OFSTED.

This was because in public mythology, if not in reality, OFSTED symbolised heavy-handed intervention in the running of schools, naming and shaming those who did not meet inspectorial standards. Dame Deirdre and Peter Homa, anxious above all to make CHI acceptable to those working in the NHS, repudiated the very notion that CHI would be

an inspectorate, let alone another OFSTED. CHI, they stressed, was about promoting improvement, not about rooting out delinquency.

This rhetoric had long antecedents, since it echoed the philosophy of the NHS's first quasi-inspectorate – the Health Advisory Service, set up in 1969 to report on conditions in long-stay institutions (Klein and Hall 1974). Indeed, for a long time CHI remained an inspectorate that dared not speak its name, and only acknowledged that title some three years after its launch. There was a further ambiguity. In his speech, the prime minister defined CHI's role as being 'to examine the quality of care from the point of view of patients' (Prime Minister 1999, p 3). But was that, in fact, CHI's role?

The legislation, as we have seen, defined its role rather differently. It was to examine the 'arrangements' made by trusts – in other words, their systems of clinical governance, rather than the quality of care itself. The assumption appeared to be that good clinical governance could be equated with good performance. But could it?

Later in this paper, we examine how CHI dealt with these tensions and ambiguities in practice in developing its style and methodology. But not all the challenges to CHI were unique to it. Far from it: other inspectors and regulators, both in the UK and elsewhere, had long faced many of the same or similar challenges. Accordingly, we first briefly review the generic dilemmas of regulation and inspection to provide context for CHI's experiment in self-invention.

Dilemmas of regulation

There is no generally accepted definition of regulation and of what – if anything – distinguishes it from inspection (Walshe 2003). Rigorously defined, regulation should in our view be taken to mean the function of monitoring the performance of the regulated bodies, and imposing sanctions if they fail to comply with required conditions. Inspection, however, has a narrower meaning: to cite the Oxford English Dictionary, it is 'to view closely and critically', usually involving a hands-on process of inquiry. Inspection may be part of regulation – though not necessarily so – but it is not identical with it.

Strictly speaking, CHI is an inspectorate, not a regulator. It cannot impose sanctions, although it may recommend 'special measures' to the Secretary of State. Unlike the Housing Corporation or the Care Standards Commission (responsible for the private sector of health care, until it is replaced by CHAI), CHI cannot strike a provider off the list of sanctioned institutions. It relies on the managerial hierarchy of the NHS to take any remedial action that might be required following its inspections. However, in practice, the distinction between regulation and inspection has become blurred in public discussion. This is why we have used the two terms interchangeably in what has gone before, and continue to do so now.

In the wider sense, regulation can best be seen as a balancing or juggling act. Regulators have to pursue different (not necessarily compatible) aims and strategies. They have to satisfy different interest groups or constituencies. They have to choose between different methods. They have to decide on their style of intervention. In each case, there may be difficult choices or dilemmas (Day et al 1996).

Different aims

The aim of regulation can be:

- to avoid the eruption of scandals
- to ensure the achievement of minimum standards
- to improve quality
- to achieve value for money
- some combination of all four.

Different constituencies

Regulators – in the case of the NHS, at any rate – may have to try to satisfy:

- ministers
- the public
- professionals
- managers
- their own organisation (regulatory bureaucracies – like all bureaucracies – have their own organisational interests to pursue).

Different methods

Regulators have to choose between using:

- professional inspectors or professional peers of those being inspected
- routine rounds of inspection or selective visiting based on desktop monitoring
- standard-based inspections or relying on the judgement of individual inspectorial teams.

Equally, they have to decide whether their focus should be on the inspected organisation's own system of control, or whether and to what extent they should carry out a reality check and investigate how that system of control actually works in practice.

Different styles Regulators have to choose between policing or a developmental/consultancy role. If they choose the former, their function is simply to ensure that the rules or standards are followed. If they choose the latter, their role is to help the inspected to improve their performance. Further, their choice of styles will be influenced by whether they put more weight on achieving consistency or allowing inspectors discretion to use their judgement in the light of each individual case.

Two points about this schematic analysis of the regulatory task need to be stressed. First, there is a close relationship between the different dimensions of the regulatory task. To the extent that regulators want to reassure the professionals working in the service being inspected, they are likely to use teams of peers rather than professional inspectors. To the extent that they see themselves as consultants rather than policemen, they are likely to emphasise judgement and discretion.

Second, the choices are rarely absolute, and frequently not even explicit. Consider the policing versus developmental antithesis. In practice, no one (with the possible exception of traffic wardens, but certainly not the police) enforces laws or rules automatically: discretion is usually required in their interpretation, and decisions are made about their applicability to individual situations. And, as always, there is a gap between policy and implementation. Even in regulatory agencies where the organisational policy is one of policing, the street-level bureaucrats tend to adopt a consultancy mode when inspecting (Day and Klein 1987b).

Again, when it comes to choosing between desktop monitoring or hands-on inspection, or between concentrating on the inspected organisation's own control system and carrying out reality checks by digging deep into how that organisation is operating, regulatory agencies do not pursue an 'either/or' strategy, but try to achieve a balance between different methods.

Achieving such a balance is not a once-and-for-all matter. Regulatory agencies tend to change their strategies and methods over time. Such changes may be prompted by a variety of factors. Experience may show weaknesses in the methods being used; the inspected may learn to game the system; ministers may make new demands for consistency or rigour; and the inspected may complain about overload. Regulation is a process of experiment, adjustment and change, and one test of a regulatory agency may in fact be its ability to demonstrate flexibility and adaptability. In short, there was no ideal, or generally accepted, model available for CHI when it started work. It had to make its own choices about strategies and methods.

CHI invents itself

CHI started from scratch, but grew rapidly. Its first-year budget was a mere £1.5 million. In 2001 the figure rose to £11.3 million, which doubled to £22.7 million the following year. In the 2003/04 financial year, it is set to reach £32.8 million. Much of the rise was attributable to the introduction of the Office for Information and Healthcare Performance (Commission for Health Improvement 2003a).

The number of staff at CHI's City of London headquarters shows the same pattern of rapid expansion: in mid-2003 it stood at 428 whole-time equivalents (WTEs). Much effort, therefore, had to be invested in recruiting staff and building up the organisation. Staff came from a variety of backgrounds: predominantly managerial, medical and nursing. Some leading members – including the director of policy and development, and subsequent acting chief executive, Dr Jocelyn Cornwell – were recruited from the Audit Commission, and brought their experience of inspection in the NHS (Day and Klein 2001) with them. Many had worked in the NHS and the Department of Health. A few had academic backgrounds.

The members of the commission itself – who included a handful of academics, a general practitioner and several with a background in charity or community work – had to be organised in a complex system of boards, each responsible for overlooking particular aspects of CHI's work. Only one commissioner other than the chairman had experience of frontline NHS management.

From the start, the new commission was under pressure from the Department of Health to deliver quickly. Indeed, throughout CHI's existence, relations between the commission and the department can best be described as one of mutual exasperation, with occasional sparks of outright hostility. CHI resented what it saw as the department's attempts to performance-manage it: for example, by prescribing the number of reviews that should be carried out in any one year. The department felt frustrated by what it saw as CHI's resistance to its legitimate demands, and its exclusion from detailed discussions of CHI's methodology.

CHI, on the other hand, was highly critical of the department's first 'star-ratings' exercise because it cut across, and ignored, CHI's own emphasis on quality in assessing performance. The department, conveying ministerial views, was critical of CHI's failure to recruit top-level managers to its band of reviewers – a criticism strongly resented by CHI, which argued that this was not for want of trying. The ambiguities of CHI's status – independent, yet dependent on government funding – haunted the relationship: the department was paying the piper, but the piper appeared insistent on calling the tune. Neither party was happy in the outcome.

Whether or not because of prompting from the department, CHI decided on the broad outlines of its inspection methodology with remarkable speed, and with very fast and only limited outside consultation within six months of being set up. Indeed, one characteristic of CHI from the start – perhaps because of the challenge of having to build up a large organisation quickly – was a certain degree of introversion: a preoccupation with ensuring the integrity of its own procedures and, armed with a strong sense of its own rigour and rectitude, a reluctance to draw a larger constituency of NHS professionals or others into any discussion of its strategy or methods. The commission chairman and leading members of staff were very active in addressing meetings in line with CHI's much

emphasised commitment to the principle of openness, but this was an exercise in providing explanation and reassurance, rather than consultation.

In this respect, as in others noted later, CHI differed from its Scottish counterpart – the Clinical Standards Board, since absorbed into NHS Improvement Scotland. Helped, no doubt, by Scotland's smaller size, NHS Quality Improvement's generic clinical governance standards were produced with the help of a large and representative project group, backed by an even larger review team, to examine the implementation of the inspection process. Maybe, however, more than a Scottish factor was involved. CHI's successor, CHAI, issued a paper outlining its strategic choices, and inviting comments even before it had formally come into existence (Commission for Healthcare Audit and Inspection 2003). This even included a questionnaire, although it was not clear who was expected to fill it in.

In any case, much of the agenda was set for CHI, as was clear from the paper presented to the commission on 9 March 2000 (Commission for Health Inspection 2000a). This outlined the 'principles and methodology' for reviews. The paper is worth examining here in some detail, since it determined (with surprisingly little modification) the way CHI was to set about its task in the years to come. Much of the task was predetermined by the Department of Health's remit to CHI. Reviews of NHS trusts were to take place every four years, and performance would be judged against the pillars of clinical governance, as proposed by the NHS Executive. These were:

- clinical audit
- risk management
- staffing and staff management
- uses of clinical and other information in management
- education and continuing personal development
- research and effectiveness (now replaced by clinical effectiveness)
- patients' experience.

The first six pillars were to provide the framework for all CHI reviews, and the reports subsequently published. But CHI added a seventh – patient and public involvement – when it moved on to elaborating its methodology. (See 'What is covered by a CHI review?', p 17). It was on these seven criteria that the performance of trusts was – and is – scored, on a I-to-IV scale. CHI incorporated two unscored elements into its reviews: each report carries an introductory section on the patient experience and a final section on the trust's capacity for improvement. We explore these characteristics of the reviews, and the subsequent reports, in greater detail in what follows.

The March 2000 paper was also crucial for two decisions – one implicit, and the other explicit. The implicit decision was that, in contrast to the National Care Standards Commission and CHI's Scottish counterpart, CHI would not inspect against a set of explicit standards. No such standards for clinical governance existed, so they would have had to have been devised by the Department of Health – a task that would have set back CHI's schedule of carrying out a complete round of trust reviews within four years.

In any case, CHI rejected such an approach, on the grounds that it would mean a box-ticking, mechanistic policing approach, incompatible with its proposed style of working. In contrast, CHI's approach would be to treat the review process as a 'collaborative activity with developmental aims' (Commission for Health Improvement 2000a, p 3). While the National Care Standards Commission had been set up to ensure minimum standards in the private sector of health care as a form of consumer protection, CHI was

about promoting continuous quality improvement. In the event, over time, CHI put flesh on the bones of clinical governance by unpacking the individual pillars into checklists for reviewers.

The explicit decision was to use review teams drawn from professionals working in the NHS, with a leavening of lay members, opting for what was in effect largely a peer-judgement model. The teams were to consist of one-to-two doctors, two-to-three nurses or professions ancillary to medicine (PAMs), one manager and one-to-two lay people. They would also include a review manager who would be on CHI's staff, while the others would serve on a part-time, occasional basis.

The model for review teams outlined in the March 2000 paper has survived largely intact. The paper estimated that a pool of 400–600 reviewers would eventually be needed. An energetic recruitment campaign was launched, offering prospective reviewers £250 per day. In the outcome, the upper figure was reached by 2002 – and has almost doubled since, now standing at over 1,000. The mix and size varies slightly from review to review. From time to time, an extra member of CHI's staff is drafted in, whether to gain first-hand experience or to beef up the team, and occasionally numbers are expanded when a particularly large or complex trust is reviewed.

But the original model holds. To take at random a 2003 review report on one of the largest acute trusts in the NHS, the team consisted of a director of pharmacy in an NHS hospital, an independent consultant surgeon, an NHS consultant oncologist, a retired headmaster, the chief executive of a primary care trust, the director of nursing services in a large NHS trust, a retired naval officer and an independent occupational therapist. In another case, the team's membership was a clinical director for professional services and head of physiotherapy in an NHS hospital, a retired archdeacon, a consultant paediatrician, a university lecturer, a director of corporate development in an NHS trust and a maternity services manager.

The quality and calibre of reviewers was to become a matter of concern in the NHS. But there was another issue, inherent in the nature of the model of inspection chosen by CHI, that was to emerge as even more important: that of consistency. CHI's March 2000 paper recognised the centrality of this challenge, but was somewhat vague as to how it was to be met: 'CHI will adopt a common and consistent approach in all its reviews and will train all members of review teams in the approach.' (Commission for Health Improvement 2000d, p 4)

In the event, would-be reviewers were assessed and, if accepted, sent on a three-day residential training course, which most of them found helpful, particularly in terms of introducing them to the human dynamics of the review process. Role-playing actors proved a great success with the trainees. But to achieve consistency required more than that – it required clarity and precision about what review teams should look at, and the criteria to be used in making judgements. This was an issue that continued to haunt CHI throughout its existence, as we shall see.

The original model also proved enduring in another respect. The framework for conducting reviews outlined in the March 2000 paper survived, albeit with modifications in the light of the first pilot reviews and subsequent experience. The first step in any review would be to collect available data about the trust's performance. This would draw on national data sources, reports of reviews carried out by other bodies, and consultations with local stakeholders, including GPs, social services, community health services, patient groups, and voluntary organisations. CHI's analysis of this information

would then shape the team's focus during its visit to the trust. The review would examine:

- what actually happens to patients and the organisation of care
- clinical outcomes, where possible
- views and comments of patients and carers
- aspects of the environment in which care is provided, including physical access and the protection of privacy and dignity.

In carrying out its task, the team would 'explore the actions and manner in which a sample of clinical teams control, influence and improve the quality of care on the ground' (Commission for Health Improvement 2000a, p 4). There would be interviews with clinical teams and senior staff. The visit would 'produce the information on which the visiting team base their judgements about the strengths and weaknesses of the leadership and corporate arrangements for clinical governance'. (Ibid, p 5)

After the visit and the production of the draft report, there would be a meeting with the trust to provide feedback to discuss the draft and agree on an action plan. The implementation of the action plan would be followed up by the regional office or health authority as appropriate. Both the report and the action plan would be published. CHI also proposed to commission an external evaluation of the review process, to be carried out by an external organisation, from the autumn of 2000.

The external evaluation did not take place. Only in mid-2003 did CHI commission an external evaluation of its activities. Otherwise, the proposals formed the basis, in an elaborated form, of the review process (Commission for Health Improvement 2001a). The data-collection exercise was extended to include information about the trust's profile, strategies and business plans, and about the individual components of clinical governance, such as clinical audit and patient surveys, as well as the minutes of board meetings.

To assess the patient experience (which, as CHI invariably stressed in all its publications, lay at the heart of its work), CHI introduced the use of patient diaries. As part of each review, patient diaries were distributed to a random selection of 200 people who had been in, or had attended, the hospital in the previous two months. The process for selecting the 'tracer' clinical teams was elaborated: the norm was to select three teams (two chosen by the trust and one picked by CHI). The aim in following through these teams was to test how clinical governance was working at the grassroots level, rather than picking a sample representative of the whole trust. The brief for reviews and reports included identifying examples of good practice – worthy of imitation in the NHS – as well as shortcomings calling for remedial action.

While CHI was hesitant about commissioning external scrutinies, it made much of its dedication 'to a process of constant self-examination and improvement as part of its commitment to being a high-performing learning organisation' (Commission for Health Improvement 2002a, p 8). Indeed, internally, CHI appeared to be engaged in a constant process of self-questioning, although the many papers produced did not always appear to make much difference to practice.

However, one product did make a significant impact on the review process: 'Project Redesign', launched in mid-2001, and carried into effect the following year. This was a resource-driven project whose primary aim was to increase productivity by cutting the

investment of staff time in each review. The result was to reduce the demands on trusts for data, to compress the time cycle of each review, and to produce shorter reports.

Previously reviews had taken around 24 weeks to complete, from start to preparing a report. Now the aim was to complete the process in 16 weeks. The site visit was to be carried out in five days in week eight of the review, during which time all interviews were to be carried out and a presentation made to the trust on initial findings. And while earlier reports tended to be between 30 and 50 pages long, post-redesign reports have hovered around the 20 page mark.

The result has been a gain in readability but offset by a loss of detail, making it more difficult for the reader to assess the evidence on which conclusions are based. For example, there are no longer any details about the number of people interviewed or meetings held. Finally, as part of the redesign project, patient diaries were dropped. Given the low completion rates, there had long been doubts at CHI about their usefulness.

There were other changes, too. The assessment scores became more sophisticated. CHI's internal system of quality control was tightened: it appointed managers to assure the quality of the review process and reports, and devised a quality-standards manual. It produced self-assessment forms for trusts too. These were to have played an important part in the review process, as envisaged in CHI's original March 2000 paper, but never became part of it (in contrast to the developments in Scotland). The 'research and effectiveness' pillar was relabelled 'clinical effectiveness', reflecting the realisation that research was not necessarily related to the quality of services delivered. Most recently, a computer-assisted system for weighting and validating scores has been developed.

However, neither Project Redesign nor any of the other changes involved a fundamental rethink of the review methodology, nor the appropriateness of the clinical governance framework for inspection. This was despite the fact that an internal evaluation of Project Redesign (not made available to the public or to researchers, despite CHI's emphasis on openness) suggested that quite a few problems remained. If CHI's life had not been cut short by the Government, there may have been more radical changes and a reassessment of its strategy, although this might well have required a change in the legislative mandate for the commission. As it was, these tasks were left to its successor body.

Overall, then, in its first three years, CHI had been triumphantly successful in delivering products, as the chairman could announce in her annual report for 2002/03 (Commission for Health Improvement 2003b). The commission had completed more than 260 clinical governance reviews, carried out ten investigations into alleged serious service failures, and made a study of the implementation of the National Service Framework for Cancer in co-operation with the Audit Commission, as well as laying the groundwork for future NSF studies. In addition, it was about to publish a report on the state of the NHS and the first of the 'star-rating' exercises for which it was responsible. But could productivity be equated with effectiveness? How rigorous and fair (to use CHI's own criteria for its work) were its methods? How were CHI's activities seen in the NHS? In the following sections, we address these questions by analysing the reviews and the experience of acute trusts. We explore these issues against the background of the organisational goals that CHI set itself, as set out in the box overleaf.

CHI's high-level organisational goals

Goal 1 By 2004, CHI will have made a significant contribution towards achieving improvement in the quality of NHS patient health care and social care, including:

- demonstrable improvements in the quality of care provided by the organisations CHI has assessed
- evidence that NHS organisations are learning from and acting on the outcomes across all CHI's work
- bringing together partner organisations to reduce the regulatory burden on the NHS.

Goal 2 By 2004, CHI will have fully integrated NHS patients and the public, especially vulnerable and marginalised people, in all aspects of its work.

Goal 3 By 2004, most of the general public will be aware of CHI's work and how they can use it to understand their local NHS.

Goal 4 By 2004, CHI will have made a significant contribution towards better and more appropriate public information on the quality of the NHS.

Goal 5 By 2004, CHI will be fully established as a high-performing enterprise achieving continuous improvement of its own work.

Source: Commission for Health Improvement (2001b). Corporate Plan, 2001–2004. London: CHI

CHI in action

In this section, we examine how CHI's strategy and methodology worked out in practice. First, we discuss the extent to which the seven pillars of clinical governance provide an adequate conceptual foundation for the reviews, and whether they are the appropriate 'tin-openers'. Second, we set out the problems encountered in achieving CHI's aim of putting the patient at the centre of its activities, and ask whether or not CHI achieved consistency, and whether the different review teams use the same lenses when looking at the performance of different trusts. Third, we analyse the outcomes of the reviews, the scores awarded, and the difficulties in interpreting them. Finally, we ask to whom the reports were addressed: what is the audience for CHI's output?

How solid are the pillars?

Unlike the ten commandments, the seven pillars do not claim divine inspiration. But their origin is almost as mysterious. If anyone can claim the role of Moses, it is probably Sir Liam Donaldson. One of the first prophets of clinical governance (Scally and Donaldson 1998), he subsequently became chief medical officer of the Department of Health, and continued to pursue his vision (Halligan and Donaldson 2001). The pillars are perhaps best seen as a pragmatic distillation and formalisation by officials of current thinking about the promotion of quality in health care, drawing as much on common sense as on theory or evidence.

The box below sets out the standard statement of the seven pillars, and the other areas covered by reviews, as it appears in all CHI reports. The rationale for most of them is self-evident. If a trust does not have adequate information systems, it cannot monitor the quality of patient care. If it does not apply evidence of effectiveness, it is unlikely to promote best practice. If clinical audit is not carried out, there is no check on standards or evidence about how to improve them. If a trust does not systematically identify risks and if there is not a readiness to learn from mistakes, patients and staff may suffer unnecessary exposure to danger. If there is inadequate training and education, staff may not be up to date with developments in their field. If the trust does not manage the recruitment and deployment of its staff well, or if there are poor working conditions, patients may suffer.

What is covered by a CHI review?	
<p>Seven areas of clinical governance:</p> <ul style="list-style-type: none"> ■ patient involvement ■ risk management ■ clinical audit ■ staffing and management ■ education and training ■ clinical effectiveness (formerly research and effectiveness) ■ use of information 	<p>Two additional areas:</p> <ul style="list-style-type: none"> ■ the patient experience ■ the trust's strategic capacity for developing and implementing clinical governance

Source: CHI reports

The odd one out in the list of pillars is patient involvement, or the trust's systems for patient, service user, carer and public involvement, as the CHI reports put it. This, as already noted, was a late addition to the six existing pillars. We shall discuss this further (see 'Searching for the patient', p 19) when we address the specific question of CHI's strivings to put the patient at the centre of its activities.

Here, it is worth noting that this pillar does not fit comfortably into the logic of the approach as a whole. It conflates two very different issues that have no necessary relationship to each other. The first is the extent to which 'patients are enabled to be partners in making decisions about their own care', as the Scottish standards put it (NHS Improvement Scotland 2003, p 3). The second is the extent of wider involvement in the design and planning of services. Both may be considered to be desirable, but they have very different justifications, and the ability of patients to have a say in their own treatment is surely a product of an effective system of clinical governance— an indicator, as it were, of whether quality care is being delivered, as distinct from being one of the components of that system.

In this respect, this pillar is different from the others. Yet in CHI's scoring system it receives the same weighting as the other pillars. More important still, neither of these concepts of patient involvement bear any relationship to patient-centred services, which is what CHI originally set out to achieve (see 'Searching for the patient', p 19).

The fact that all seven areas receive the same weighting in CHI's assessments of trusts points to a further issue – that the seven-pillar approach is essentially a taxonomy of what good practice requires. It does not attempt to analyse the dynamics of the process – in other words, to ask about the relative contribution of the different factors and the interaction between them. So, for example, it might be argued that a trust's ability to inform itself about its activities is more fundamental than the other dimensions of clinical governance. Without a good information system, can a trust be relied upon to deliver good quality care?

Again, while the seven pillars (or at least six of them, including the revised clinical effectiveness) may describe the prerequisites necessary for delivering good quality care and striving for improvement, it does not follow that the desired results will inevitably flow from putting the structures in place. Here, we return to the tension in CHI's original mandate. This is that its task was defined as being to review 'the arrangements' for clinical governance, not the quality of the resulting performance. In this, CHI's dilemma mirrors that of all regulatory bodies: the balance between looking at the formal machinery of self-regulation (often an end in itself, as well as a discrete and flourishing industry) in the inspected organisations, and scrutinising how that machinery works in practice.

CHI's approach has been to test the effectiveness (or otherwise) of the formal systems of clinical governance by digging into the operation: by supplementing evidence about the systems with evidence about their impact on the ground, by interviewing staff and patients. This is one of its strengths, but also the source of its main problems: how to secure evidence about the patient experience, and how to achieve consistency.

There is a further difficulty. To return to the article that launched the concept of clinical governance: 'The feature that distinguishes the best health organisations is their culture... However, evidence on how to define a 'good culture' and on the methods required to promote one is largely lacking in the healthcare field.' (Scully and Donaldson 1998, p 63). And, again: 'Most observers would identify leadership as an equally

important ingredient in successful organisations. However, leadership too is a rather vague concept'. (Ibid)

The seven pillars clearly do not capture these crucial – if elusive – characteristics of culture and leadership. So, in effect, CHI has been left to fill this vacuum, which it has done by including a section on the trust's strategic capacity in its review reports. This section is not scored (although CHI has been working on methods for so doing), and in the outcome, this part of the reports tends to be a something of a rag bag of reflections on the trust's leadership style, committee structure and relations with other bodies.

Searching for the patient

If constant invocation of the principle of patient-centredness could make it so, CHI would have no problems. But, like other inspectorates, CHI has found that basing its reviews on the patient (or client or consumer, as the case might be) experience is easier said than done. This has not been for want of trying: CHI has invested much effort in developing its methods and continues to do so. There has been the experiment with patient diaries, since abandoned in favour of national patient surveys. There has been, from the start, an emphasis on seeking the views of the local population by inviting comments from users and carers, as well as consulting a variety of stakeholders such as voluntary organisations, in advance of a review. In addition, of course, review teams have drawn on their own observations and interviews with patients as well as staff.

All the methods have weaknesses or drawbacks. The most recent, streamlined, review reports give no details of the response to CHI's invitations to patients and others to give their views. But a sample of earlier reports that do give these details suggests that the response tends to be stuttering and variable. There is no queue of users eager to give their views to CHI. So, for example, the number of meetings with individuals and groups ranged from two (Taunton and Somerset) to 72 (Whittington), and the number of letters and emails received ranged from six (Winchester and Eastleigh) to 169 (Bart's and The London), while the completion rates of patient diaries varied from 18 in Homerton to 175 in Royal West Sussex, though very few topped the 100 mark.

While surveys tend to achieve higher response rates, they have other drawbacks in exploring the experience of patients using specific services – coronary care, outpatients, A&E – but not necessarily giving an all-round picture. Moreover, little is known about the stability or otherwise of services: is patient experience at a particular point in time still relevant six months or a year later? Lastly, observation and interviewing is bound to be selective – all the more so given the complexity and heterogeneity of the NHS.

In making its assessments of the patient experience, CHI has therefore been largely driven to use indirect methods and selective insights into how trusts treat patients. CHI's own conceptual framework for assessing the patient experience (as set out on its website) has five dimensions:

- clinical effectiveness
- access to services
- organisation of care
- humanity of care
- environment.

In turn, the framework lists six information sources for each of the dimensions:

- standardised indicators
- information provided by the trust
- statutory community and voluntary organisations
- the views of patients and carers
- site observation
- staff input.

The relative importance of the information sources varies between the different dimensions. In the case of clinical effectiveness, national and trust data make the most significant contribution, supplemented by the views of users and staff. In the case of access, standard national and trust data about waiting times, missed appointments and cancelled operations is central, supplemented by the experience of patients and the observations of reviewers (for example, disabled access or signage). In the case of humanity of care, trust information about its policies for care provision is supplemented by patient and staff accounts, as well as the views of reviewers about whether patients are treated with dignity.

The published reports give an impressionistic picture of care delivery: a skeleton of hard data is fleshed out by the team's interpretation of what they see and hear. So a typical report will comment on:

- dignity and privacy often drawing attention to mixed-sex wards
- accessibility largely based on the trust's record in achieving performance targets, but also taking into account patient views about parking and transport
- standards of cleanliness, food and facilities relying largely on patient views and observation
- outcomes using mortality and readmission data
- the organisation of care raising issues such as cancelled operations, trolley waits and the number of 'outliers' – in other words, patients not in their specialty ward.

However, within this standardised framework, reports differ considerably. Some give specific details about outcomes by comparing the trust's performance to the national average. Others simply remark that the trust's rates are 'higher than expected'. Generally, reports do not attempt to resolve disputes about the meaning of the figures (for example, if mortality is above average, does this reflect case mix, mistakes in coding, or poor performance?) but call upon the trust to resolve the puzzle. This may be understandable, given the complexities of sorting out such perplexities, but is not helpful to the lay reader wondering about how good his or her local trust is.

Further, like the NHS itself, patients are heterogeneous, and their views differ. The result is that reports often draw no conclusions: 'Patients' views on the quality and quantity of food varied from 'excellent' to 'the food was cold and unappetising.' In this case (Worcestershire), CHI's reviewers gallantly sampled the food and found it to be 'acceptable'. The frequently invoked formula of 'Some patients complain...' or 'CHI received several reports...' is less than helpful in trying to gain an overall impression of what a trust is like in the absence of any attempt to establish how far these complaints or reports represent the usual patient experience.

Turning from the patient experience to patient involvement (a scored element in all CHI reviews), somewhat different issues present themselves. The first has already been

touched on – that the category incorporates two dimensions that do not have any necessary link between them. To quote the definition which appears in CHI reports, ‘Patient involvement describes how individual patients can have a say in their own treatment and how collectively they and patient organisations can have a say in the way that services are provided.’ Addressing the question of how far individual patients have a say in their own treatment presents all the problems already rehearsed. Like the patient experience, it is a butterfly of a concept – difficult to pin down and measure in practice. And, in practice, CHI reports concentrate on assessing the extent of patient and public involvement.

At a descriptive level, looking at patient and public involvement presents few problems. A range of indicators are used in reports, including:

- whether a strategy document exists
- patient and public representation on trust committees
- the extent of consultation with user organisations
- the availability of information in different languages
- the way complaints are handled.

It is more difficult still to move from description to dynamics: the extent to which patient and public involvement influences policy or practice. Here, reports rely chiefly on evidence from trusts. The onus is on the inspected organisation to provide evidence of changes consequent on consulting patient and public views.

The central question remains: does good performance on involving users and public equate with good performance in terms of patients having a say in their own treatment?

Tables 1 and 2 (see pp 22–23) provide a tentative answer to this question. In these tables, we have taken a random selection of eight trusts, half of which received the minimum score of I for patient involvement in CHI reviews, while the other half received a score of III. We next extracted their scores in the user surveys of outpatient (Table 1) and emergency (Table 2) departments from the CHI website. In each case, we show the results for two questions: the first about involvement in decisions about care and treatment, the second about overall satisfaction with the care received.

If CHI’s methodology for assessing trusts on this dimension is sufficiently appropriate and sensitive, we would expect the two sets of scores to match. Trusts with high CHI scores should also have high scores in the surveys, and vice versa. In the outcome, as the tables show, this is not the case. Only one of the highly rated trusts (Doncaster) has an unambiguously better survey performance, allowing for confidence intervals, than the low-rated one. Conversely, only one of the low-rated trusts (Whipp’s Cross) seems to have an unambiguously worse performance than the high-rated trusts. Otherwise, the picture is mixed, with some of the low-rated trusts doing better than the high-rated ones.

We do not want to make too much of these results because our sample is very small, the surveys cover only two NHS services, and other factors are involved. However, the results do reinforce scepticism, and suggest the need for CHI (or its successor) to carry out more sophisticated analyses of the relationship between its scores and other indicators of performance.

Table 1: CHI scores v. scores in patient surveys: outpatients survey

Trusts with a score of I	Trusts with a score of III
United Lincolnshire a: * 82 (79–84)** b: 83 (82–85)	Northampton a: 82 (80–85) b: 79 (78–81)
Whipp's Cross a: 78 (75–81) b: 71 (69–73)	Doncaster a: 84 (81–86) b: 83 (81–84)
Essex Rivers a: 79 (77–82) b: 81 (79–82)	Portsmouth a: 80 (77–83) b: 83 (81–84)
Mid-Essex a: 84 (81–86) b: 77 (76–79)	Queen's Medical Centre, Nottingham a: 82 (79–85) b: 81 (79–83)
<p>* The CHI scores are for 'the trust's systems for patient, service user, carer and public involvement'. The patient survey scores are for the responses to two questions: a: Were you involved in decisions about your care and treatment? b: Overall, how would you rate the care received? ** The figures in brackets give the confidence intervals. In all cases, the threshold for the top 20 per cent of trusts is 84.</p>	

Source: CHI reports; CHI website

Table 2: CHI scores v. scores in patient surveys: emergency department survey

Trusts with a score of I	Trusts with a score of III
United Lincolnshire a: * 79 (76–82)** b: 78 (76–80)	Northampton a: 78 (74–81) b: 73 (70–75)
Whipp's Cross a: 68 (64–72) b: 68 (65–70)	Doncaster a: 83 (79–87) b: 81 (78–84)
Essex Rivers a: 78 (75–81) b: 74 (72–76)	Portsmouth a: 71 (68–75) b: 72 (69–74)
Mid-Essex a: 78 (74–81) b: 73 (71–76)	Queen's Medical Centre, Nottingham a: 71 (67–75) b: 71 (68–73)
*The CHI ratings are for 'the trust's systems for patient, service user and public involvement'. The patient survey scores are for the responses to two questions: a: Were you involved in decisions about your care and treatment? b: Overall, how would you rate the care received? **The figures in brackets give the confidence intervals. In all cases, the threshold for the top 20 per cent of trusts is 80.	

Source: CHI reports; CHI website

The search for consistency

The achievement of consistency is the (usually elusive) quest for the Holy Grail of all inspectorates. In the case of CHI, it is a quest made more difficult by the way it decided to carry out its remit, as noted earlier (see 'How solid are the pillars?', p 17). CHI has opted against inspecting against explicit standards or criteria. It uses part-time reviewers, who usually carry out no more than two or three inspections a year and inevitably bring different perspectives to bear on their task. So the burden of ensuring consistency of approach falls on:

- the analysts who prepare the briefs for the visits
- the review managers who orchestrate the visit and prepare the report
- CHI's directorate, responsible for reviewing reports before they are published.

Our interim report on CHI (Day and Klein 2002) concluded that CHI's reports were consistent only in being inconsistent. Does this verdict still hold? In answering this question, it is important to unpack the notion of consistency. To achieve consistency (in other words, to ensure that all trusts are treated alike), there would seem to be three requirements:

- that all inspection teams look at the same aspects of performance and use the same criteria in assessing them
- that they weigh and interpret the evidence in the same way

- that they present their findings and conclusions in a way which allows comparison between trusts.

Consistency must not be confused with uniformity. Trusts come in all shapes and sizes. They may well have characteristics or problems special to themselves. To the extent that reports identify these particular characteristics or problems, consistency is not compromised. However, consistency does require that differences between reports reflect variations between trusts, not variations in the interests or focus of the teams, or in the flow of information to them.

If the searchlight of inspection is directed arbitrarily or erratically and if review teams are idiosyncratic in their focus, then consistency will be compromised, and what is revealed will reflect the team's characteristics rather than the trust's performance. Equally, consistency requires that once an issue is identified, it should be analysed and presented in the same way, otherwise it is impossible to put a trust into a comparative context.

To an extent, CHI's methodology does ensure that all review teams look at the same aspects of performance. CHI sends out a 39-page questionnaire to trusts before reviews, setting out the information it requires, and in effect providing a checklist of the issues to be covered (Commission for Health Improvement 2002c).

For each of the seven pillars, there is a list of questions. These range from the general (What are the main strategic priorities of the trust?) to the particular (Does the trust have exclusively single-sex wards and bathroom facilities?). They cover systems and processes (Does the trust have an infection control strategy?) and their implementation (Does the trust audit the implementation of evidence-based practice?). There is also a coding system for entering the findings of the review team on a computer during the visit to provide a systematic record of the evidence collected.

Finally, there is guidance on how to weigh the evidence. For a finding to be included in the review report, it has to be classified as either 'very confident' (drawn from a number of sources, including data, documents, interviews and observations) or as 'confident' (drawing from several items of information of the same source type – such as interviews – from different areas or organisations, or from one interview or observation confirmed by an independent source). If it is classified as 'some confidence' (drawn from several items of information from the same source type from the same area or organisation) it may or may not qualify for inclusion in the report.

Despite this, the reports show differences that cannot be explained by the characteristics of the trusts being reviewed. Consider, for example, a significant component of risk management with obvious relevance to the patient experience: the control of infections and bed sores. Some reports comment on both, others report on one or the other but not on both, and a few report on neither. Most reports concentrate exclusively on the processes of control and the number of staff in the relevant teams, while a few give details of the outcomes.

The way in which outcomes are presented varies yet again. Sometimes, the report simply records whether the rate of infections and pressure sores has increased or decreased over time (which is meaningless without knowing what the baseline was). Others compare the rates to those of similar trusts. Some reports comment on whether the review team observed staff washing their hands, while others do not. One (University College) comments on the extent of non-compliance with the dress code in theatres,

while others do not. Does this mean we should assume that, with the exception of this one trust, compliance is no problem in the rest of the NHS?

Let us take another of the seven pillars – staffing – with relevance to the patient experience. Many reviews report staff shortages – particularly of nurses – and stress, and these comments are usually assessed by staff themselves. Few reviews take the issue any further. One exception is an old-style, full report (Homerton) that compares the trust's staffing level with the national index. This is an example of what can be done with the data, but it does not appear to have been followed elsewhere.

Similarly, some reviews give vacancy, sickness and turnover rates but do not compare them to other, similar trusts or national data, but most do not even provide the basic information. Many reports comment on the high level of agency staff. But some do not. The only consistent element here is the absence of any attempt to analyse the staffing situation – to ask, in particular, the extent to which stress or the use of agency nurses reflect the way in which management deploys its staff or the trust's financial situation.

Further, there are some idiosyncracies in the issues picked up by the teams: one report draws attention to locked fire doors and another makes a point about vending machines. It may be that these seemingly trivial matters may be symptomatic of larger organisational weaknesses. These observations may indicate that managers are reluctant to patrol the corridors or to listen to staff and patients.

However, there are also indications that it is the composition of the team that influences what issues are picked up. For example, one report (Hinchinbrooke) comments, unusually, that 'The pharmacy is not used effectively. It is only used as a supply service, and staff do not achieve their full potential for providing a clinical advisory role.' In this case, a chief pharmacist was on the review team. In another report, much is made (again unusually) about staffing levels among therapy staff and the absence of physiotherapy in obstetrics (Burnley). In this case, a physiotherapy manager was on the review team. Does silence on these points in other reports indicate that there were no relevant experts on the teams or that there were no problems?

The examples of inconsistency could be multiplied. Some of them may appear minor, and many could be eliminated by tighter editing when producing reports. But they raise a larger, more fundamental issue: can absence of evidence be taken as evidence of absence? If we cannot be sure that different review teams are consistent in the way they approach their task – a challenge not only to CHI, but to all inspectorates – we cannot give a confident answer to this question.

Scoring the reviews

Trusts have strong incentives to do well in CHI reviews. Getting good scores matters greatly to them. This is not only a matter of pride in their reputation: poor review scores may on occasion prompt the departure of the trust's chief executive, followed sometimes by the chairman. Above all, the CHI scores play an important part in the annual performance ratings exercise, initially carried out by the Department of Health but now the responsibility of CHI (Commission for Health Improvement 2003d). High-performing trusts get three stars, while poor-performing ones get no stars.

The awards are the product of a complex 'snakes and ladders' exercise, involving the trust's performance in a number of dimensions. One of these dimensions is the result of the CHI review. So a trust that would have been awarded three stars on the basis of its performance in the other dimensions slips down to two if the CHI review indicates some weaknesses. Conversely, a trust that would otherwise have been awarded two stars climbs the ladder to three given a favourable CHI review. In turn, star ratings have significant implications for the trust: three stars mean more autonomy and the option of applying for foundation status.

CHI scores each of the seven pillars on a one-to-four-point scale (see 'CHI's scoring system', below). In deciding on their scores, review teams are guided by an assessment matrix that sets out the criteria to be used for each of the pillars (Commission for Health Improvement 2000b). These criteria set out requirements for each organisational level, from the board down to the clinical team. So, to take the example of clinical audit, a trust will score only I if:

- there is no lead responsibility
- clinical audit is not included in the business plan
- audits do not involve patients
- no specific resources are allocated to audits
- audit recommendations are not disseminated through the organisation.

The converse also follows. If the trust complies with all these requirements, it will score more highly – though the actual mark will depend on the degree of compliance and the extent to which the activity permeates the whole organisation.

CHI's scoring system

Each of the seven pillars is scored on a four-point scale:

- I Little or no progress at strategic and planning levels or at operational level
- II
 - a Worthwhile progress and development at strategic and planning level but not at operational level, or
 - b Worthwhile progress and development at operational level but not at strategic and planning level, or
 - c Worthwhile progress and development at strategic and planning levels and at operational level, but not across the whole organisation
- III Good strategic grasp and substantial implementation. Alignment of activity and development across the strategic and planning levels and operational level of the trust.
- IV Excellence – co-ordinated activity and development across the organisation and with partner organisations in the local health economy that is demonstrably leading to improvement.

In the outcome, the reports tend to be cautious in their scoring. The most frequent mark is II, followed by I. There is a fair sprinkling of IIIs, but IVs are extremely rare – so rare, indeed, that in Table 3 (opposite), which shows the distribution of scores, there is no column for the top mark: once the percentages have been rounded off to one decimal point, the IV score disappears from sight and is not included.

As can be seen, there is a considerable difference between the seven pillars. Use of information has the highest percentage of its scores in the bottom category, while education and training has the lowest. No trust has achieved anything like the maximum possible score of 28, though one scored the lowest possible – seven (Epsom and St Helier). The scores of the best performing trusts tend to hover around the 18–20 mark, but many three-star trusts did not even reach this total.

The CHI scores for the first six three star trusts in the 2003 list (alphabetically) were respectively: Addenbrooke's (17), Airedale (13), Barnsley (14), Basildon and Thurrock (16), Birmingham Heartlands (17) and Bradford (17).

No trust comes anywhere near to recording all-round excellence. These results are perhaps all the more disappointing given that the Modernisation Agency – the NHS's missionary management consultancy – has invested considerable effort into promoting clinical governance and providing coaching for trusts.

Table 3: The distribution of CHI scores

Component	I	II	III
Clinical audit	14.6%	72.8%	12.6%
Risk management	19.2%	65.6%	15.2%
Patient and public involvement	22.5%	71.5%	6.0%
Education and training	3.3%	64.2%	32.5%
Research and effectiveness	21.2%	61.6%	17.2%
Staff and staff management	17.9%	68.9%	13.2%
Use of information	33.1%	57.6%	9.3%

Source: CHI acute trusts reviews, up to August 2003

Interpreting these results is difficult. A number of different conclusions might be drawn. One is that clinical governance is still in its infancy and that much development is needed – changing the culture of the NHS, and the attitudes of those working in it, cannot be done overnight. Another is that trusts may perform well and deliver high-quality care even though CHI does not rate their clinical governance systems highly. The assumption of a direct relationship between such systems and the quality of care represents an act of faith. A third conclusion might be that CHI's scores reflect as much the weaknesses of its own methodology as those of the organisations being inspected and that, as a consequence, they short-change at least some of the trusts.

The three interpretations are not necessarily incompatible, although the emphasis given to each is likely to depend on one's point of view. The champions of clinical governance are likely to favour the first conclusion, while many trusts will undoubtedly subscribe to the third. On the basis of our evidence, we cannot resolve this question. Our evidence does not allow us to determine whether or not CHI's reviews are systematically biased towards parsimony in their scoring.

However, it does allow us to explore further the kind of problems inherent in making inspectorial assessments. In this respect, CHI is no different from other inspectorates

required to translate qualitative judgements into quantitative scores. No one has yet discovered an infallible method for doing so, and criticism and recrimination is the norm. So the fact that some CHI scores are challenged by the trusts concerned – and a few are indeed changed following (occasionally angry) representations – does not tell us very much.

There is, however, a puzzle. This is that the scores given often cannot be surmised from the text of the reports. Consider, for example, the section on patient, service user, carer and public involvement in the report on the Royal Brompton and Harefield Trust: 'There is commitment at both strategic and operational levels to develop patient and public involvement', the report says, and goes on to describe a host of activities and initiatives. But this earns the trust only a II (c) rating.

Conversely, the section on risk management in the report on University Hospitals of Leicester Trust comments on the complexity of the corporate structures to support risk management, and points out that 'most staff do not receive feedback' from the analysis of incidents and that 'risk assessments are sporadic and focus on existing rather than potential problems'. However, despite these criticisms, the trust is given a score of III.

In short, the problem of inconsistency in gathering evidence is compounded when it comes to interpreting and weighing that evidence in order to decide on a score. Splitting the II score into three categories (a, b and c) recognises that there is a difference between devising strategies and implementing them operationally, and that there can be bottom-up activity in the absence of top-down strategic planning, but it does not solve the problem of how to give weight to the different categories.

In addition, review teams appear to differ on the relative importance attached to formal mechanisms and the operational reality: some seem to mark trusts down if they pursue a devolved approach. There are also differences, not only between review teams but also within them, when members disagree with each other, or with the review manager. There are also cases of review team agreements being subsequently overruled by CHI headquarters.

There is a further factor: that of the 'inspector's nose'. Our previous work in other regulatory agencies suggests that all inspectors make an instinctive and sensory judgement about the organisation being reviewed almost as soon as they come through the door. They rely on a variety of signals – clues and hints – to give them a sense of what the organisation is like. CHI inspectors are no different, as their responses showed when we asked them about how they shaped their first impressions of a trust.

One reviewer explained:

What leaps out when you first walk into a trust are things like its cleanness and its brightness and the way the staff are welcoming.

Another said:

Once you begin interviewing, you can tell whether the trust is an open environment or whether staff are worried or frightened to speak. Staff who don't want to speak to CHI are signs of a trust with problems.

Again, the way the CHI team is looked after provides a signal about the state of the organisation:

If rooms allocated are small and dirty or drab, it's a bad sign. If general courtesies are absent, this is a bad sign – and often a sign of hostility. If the treatment of reviewers is too opulent, this is also a bad sign. If we are allocated newly decorated rooms smelling of new paint – this is probably put on just for CHI. A bad sign.

Signals like these inevitably shape the perceptions and interpretations of reviewers when they come to look for, and at, the evidence. So, too, do the personalities involved: if the chief executive and chairman carry conviction and authority, this is likely to influence the verdict reached. If the evidence is ambiguous or conflicting (as it often is), or if there is a question of giving the benefit of the doubt to an organisation, then the way the reviewers/analysts have 'framed' a trust is likely to be decisive – and explains what at times seem puzzling scores.

Another factor may influence the scores. The final section of each report discusses the trust's 'strategic capacity for improvement'. This is unscored, although CHI has been working on devising a scoring methodology following the example of the Audit Commission, which puts much emphasis on the capacity for improvement in its local authority reviews. As always, reports vary in what they include in this section, but most comment on the quality of leadership, the structure of governance, and relations with outside organisations. The tone tends to be rather 'headmasterly': a verdict is being given on whether the pupil is promising or needs to work harder. It may well be that these general conclusions influence the scores given.

To whom is CHI talking?

In CHI's 2001–04 Corporate Plan (2001b), the commission set itself five goals to be achieved by 2004 (see 'CHI's high-level organisational goals', p 16). In this section, we are concerned with CHI's strategy and style in communicating its work. Are most of the general public aware of CHI's work? The answer, as shown by a MORI poll (Office of Public Services Reform 2003), is an emphatic 'No'. When a representative sample of the public were asked what inspectorates or watchdogs they had heard of, only 1 per cent named CHI. When they were subsequently shown a prompt card with various inspectorates listed, the figure rose to 20 per cent. The equivalent figures for OFSTED were 17 per cent and 65 per cent respectively.

CHI's emphasis on this point underlines the importance it attaches to communication. But to whom is CHI talking, and who is listening? Again, there is no conclusive evidence, but there are some clues. Like many other institutions, CHI faces the problem of multiple potential audiences:

- ministers and civil servants
- those working in the inspected trusts
- those working in other trusts and in the wider NHS
- local authority members responsible for scrutinising their local NHS
- the public, or, more accurately, many different publics with different perspectives:
 - users, actual or potential, of a specific trust
 - members of pressure groups concerned with specific patient groups
 - activists with a general interest in improving NHS services.

It seems unlikely that any report – however clearly presented – can satisfy all these audiences. Our interviews with policy-makers suggest that ministers and civil servants are less than happy, for reasons that have little to do with presentation and everything to do with the kind of problems already discussed: lack of consistency and the consequent difficulties with problems of interpretation and comparison. The reactions of trusts vary, naturally enough, with the precise contents and scoring of individual reports. Our interviews suggest that there is often irritation when, in the view of the trust, CHI's text misrepresents the situation. Some trusts are also critical of CHI's press releases accompanying the publication of reports which, as they see it, tend to accentuate the negative.

This last point is important because it is largely through press releases – and the way they are picked up by the media – that the public's impression of CHI's reports is formed. The local press, in particular, give extensive coverage to CHI reports. In effect, the public is informed through the media. Around 300 copies of each report are printed. About two-thirds of these go to the trust concerned, the Department of Health, strategic health authorities, community health councils and organisations that have been involved in the consultation process preceding the review. The others are available for purchase, at a price of £6 each. However, demand for these is limited.

More important than the circulation of the reports is free access to CHI's website. Unfortunately, CHI's own information about the extent to which the public use the website is limited. All evidence of interest is based on the 'number of clicks' recorded by the server. The number varies greatly: ranging from several thousand for some reports to zero for others. The figures are difficult to interpret. It may be that some trusts use servers that do not make 'click' returns. All that can be concluded is that electronic circulation is much greater than the printed distribution.

CHI continues to make efforts to create a wider public audience for its reports. But there remains the question of what incentive the general public has to read CHI reports. Why bother to read a CHI report if there is no opportunity to exert any influence on the trust – or, in the case of a critical report, if there is no choice of an alternative hospital? In any case, the actual or potential user is more likely to be concerned about the quality of the particular service relevant to his or her needs, rather than about the overall performance of a hospital. There may be rather greater incentives when the Government implements its pledge to give patients choice. The opportunities and the challenge to CHAI will be greater.

What difference has CHI made?

From the start, CHI has emphasised that its guiding aim is not to criticise or chastise, but to help trusts to improve. As the commission's 2001 corporate plan (Commission for Health Improvement 2001b, p 11) put it: 'CHI will measure its impact by demonstrable improvements in the quality of care provided by the organisations CHI has assessed.' (See 'CHI's high-level organisational goals', p 16). However, in assessing CHI's impact there is a fundamental problem: it is difficult – if not impossible – to disentangle CHI's influence from all the other factors bearing on the performance of reviewed trusts.

If a trust has improved the quality of care subsequent to a CHI review, it could be because it already was on an upward trajectory. Alternatively, it could be due to the infusion of extra funding into the NHS, the arrival of a new chief executive or the activities of the Modernisation Agency, among other factors. In short, CHI's inspections are part of a complex process. We, therefore, do not attempt to address the question of CHI's impact directly. Instead, we examine the dynamics of the process, and explore the roles of different actors both within the reviewed trusts and within the wider NHS.

Every CHI report carries a list of prescriptions for improvement, under the rubric of 'What are key areas for action that the trust needs to address to improve its clinical governance systems?'. There are usually between four and nine of these, supplemented by additional suggestions about what should be done, which are addressed in the rest of the text. The box overleaf gives some examples of these 'key areas for action', illustrating their diversity.

Some are indeed about strengthening the system of clinical governance, for example, the recommendation for strengthening the mechanisms of risk management. Others have nothing to do with improving the clinical governance system but focus on specific service weaknesses – for example, the recommendations about the patient environment and the scheduling of trauma operations. Some are extremely general; others are very specific. Some urge the trust to take new initiatives, while others simply call for continued or sustained progress.

Examples of CHI prescriptions for action

The trust needs to ensure that progress on clinical governance is sustained and that strategies and plans are managed to produce the desired outcomes.

(University College)

The trust must take urgent action to ensure that collection, storage and disposal of clinical waste is safe and effective.

(St George's)

The trust board should urgently review mechanisms for identifying and reporting significant clinical and non-clinical risks.

(Whipps Cross)

The trust should ensure there is a clear framework for clinical governance with robust monitoring and evaluation systems in place.

(East Cheshire)

The trust needs to take action to ensure the appropriateness of all inpatient and outpatient care environments.

(Guy's and St Thomas')

The needs to widen its understanding of cultural issues to ensure it is meeting the needs of its local ethnic minority groups.

(Kingston)

The trust should ensure effective leadership and integration of the elements of clinical governance at corporate level.

(Northern Lincolnshire and Goole)

The scheduling of trauma operations at Chancellor Wing, St James' University Hospital, must be urgently reviewed.

(Leeds)

The trust needs to urgently develop mechanisms to give effective support to staff working hard in busy clinical and support areas.

(South Manchester)

The trust should ensure that its patient and public involvement strategy is implemented and sufficiently resourced.

(Norfolk and Norwich)

The trust needs to see the help of external facilitators to help resolve some of the dysfunctional relationships amongst senior staff in general surgery.

(Heatherwood and Wrexham)

These recommendations feed into the final stage of the review process: the preparation of an agreed action plan for remedying weaknesses identified during the inspection. The action plans typically specify in some detail what needs to be done, who is to be responsible, and when the changes are to be delivered. That formally ends CHI's involvement. Its role is diagnostic, not therapeutic. Responsibility for monitoring the implementation of the action plan falls on the relevant strategic health authority, while responsibility for giving advice and support on how to carry out improvements falls on the Modernisation Agency.

CHI cannot sanction trusts that fail to comply, which is why it is more accurately described as an inspectorate rather than a regulator. Only since taking over responsibility for producing the star ratings has CHI carried some stick. As part of the star-rating process, CHI calls for progress reports from strategic health authorities in order to update its review findings. Perhaps surprisingly, CHI put only a handful of trusts in the 'significant areas of weakness' category. Given the very significant weaknesses identified in some of the reports, this suggests either that there had been some dramatic improvements or that CHI was reluctant to push trusts into the zero-star category.

Assuming that all the review prescriptions for improvement were implemented – and their total number now runs well in excess of 1,000 – it might be that CHI could indeed claim a large share of credit for raising the quality of NHS services. But if we change the perspective of our analysis and look at the review process from the bottom up, the picture changes somewhat.

Unlike its Scottish counterpart, CHI does not use a system of formal self-assessment by the trusts concerned as the starting point for its reviews, although it has considered doing so. But, in practice, it depends on the information supplied by the trusts, supplemented though it is from other sources. In effect, trusts carry out an implicit and informal self-assessment, both in deciding what data to supply in advance of the review and how to present themselves during it.

As anyone who has ever been involved in being inspected knows (the experience of universities, for example, is relevant), this is a highly tactical process. It requires balancing self-criticism and the acknowledgement of weaknesses on the one hand, with emphasis that plans for remedying these weaknesses are already in train on the other. The premise is that it is better to come clean about one's weaknesses than to have them uncovered by the inspection. The hair-shirt has to be balanced by the halo.

From the trusts' point of view, CHI reviews are largely an exercise in playing back to them what they already know, and urging them to do what they wanted to do or intended doing anyway. There are often disputes about CHI's interpretation of the evidence that trusts submit, and about what counts as a 'fact'. Similarly, some trusts also resent what they see as CHI exploitation of their self-revelatory honesty, by claiming credit for drawing attention to, and emphasising, their shortcomings.

But there are few surprises. One participant, voicing what appeared to be the majority view among those interviewed, said:

CHI did not tell us anything we did not already know.
(Trust review co-ordinator)

And, in the words of a trust's lead director for one of the first CHI reviews to use the streamlined methodology:

The CHI review did not, on the whole, provide us with information that we did not already have. It is more likely that CHI as a new organisation gained more from the experience than the trust.

(Trust lead director)

This view was echoed at strategic health authority level too.

But this is not say that CHI reviews have no impact – they do. But it is not measured in the currency of prescriptions for improvement. It is an impact that has little to do with CHI's specific methods or style but everything to do with the impact of most (perhaps all) inspections on inspected bodies. This 'general inspection effect' takes two forms. First, the process of preparing for the inspection forces the organisation to engage in a systematic exercise in self-examination, which can turn into a pre-emptive exercise in self-improvement. Second, the publication of the inspection report changes the dynamics of the organisation concerned, giving extra leverage to those committed to bringing about change. In what follows, we illustrate both points.

Trusts found preparing for the visit a painful experience, and not just because of what they saw as CHI's excessive and insufficiently focused demands for data. Responding to CHI meant holding up a mirror to their own shortcomings. As one review co-ordinator put it:

The original questionnaire from CHI about our organisation has been good for the trust. It allowed us to get our act together by bringing information which gave us a coherent picture of ourselves – including gaps in our information about what we do and also gaps in what we should be doing.

(Review co-ordinator)

Or, in the words of a chief executive: 'Most hospitals, and certainly this one, spend quite a lot of time every year looking at how their benchmarking is doing against all sorts of measures, and all that happens when something like CHI comes in is that it rather crystallises your thinking about some of those, and leads you to be perhaps more proactive than you might otherwise be.'

The prospect of a CHI review also forced trusts to pay attention to their clinical governance systems. As a clinical director said:

It makes us look at our services to see whether we are doing things in exactly the way we should in accordance with clinical governance. It allows you to define where you are doing well and give people credit for that, and it also allows you to define where you could improve.

(Clinical director, cited in Easton 2002, p 28)

A trust review co-ordinator explained:

CHI checks to see whether the systems are being translated into clinical practice. Our trust had many systems in place which were not actually functioning at clinical level. CHI gave us the push to take systems down into practice.

(Trust review co-ordinator)

As the previous quote indicates, either preparing for a CHI visit or dealing with its aftermath provides ammunition for those wanting change. The same review co-ordinator said:

The benefit of CHI is its existence. We used the CHI review as a lever to make changes in the trust... Doctors have no corporate image of clinical services. It was good to tell doctors that they need to do more than write clinical papers as their contribution to audit and effectiveness. They need to close the loop by seeing that effectiveness gets taken down to patient level.

(Ibid)

Conversely, one chief executive interviewed complained that CHI had been excessively complimentary about a part of his trust which he considered to be weak, so depriving him of the opportunity to use the review to engineer change. Trusts, however, tended to emphasise that the CHI prescriptions, and the accompanying action plan, are in effect endorsements of decisions already taken by the trust. They may give a higher profile to some issues, change the order of priorities, reinforce management's determination or give it new tools of persuasion, but they do not, by and large, represent new initiatives that would not have been taken if CHI did not exist.

There may, of course, be a self-serving element in this insistence. But even allowing for this, the conclusion would seem to be that CHI's role is largely catalytic. CHI certainly has an impact, but it is a diffuse and indirect one: it changes the way in which trusts see themselves.

The difficulties of isolating the impact of reviews on individual trusts are compounded when it comes to the commission's more general goal (see 'CHI's high-level organisational goals', p 16) – that by 2004 there should be 'evidence that NHS organisations are learning from and acting on the outcomes across all CHI's work'.

CHI does indeed put much emphasis on drawing out general lessons from its reviews. It identifies general themes, pinpointing generic weaknesses in the NHS. Unsurprisingly, poor information systems emerges as one such theme. And each report carries a section on what the 'rest of the NHS can learn' from the experience of, or initiatives taken by, the reviewed trust. The box overleaf gives some examples. It is, however, far from clear how far the good practices, held out for imitation by the rest of the NHS, are in fact unique to the trust concerned – let alone followed by other organisations.

In our interviews with NHS staff designed to explore their perceptions of CHI's role (see 'The NHS on CHI', p 37) we found evidence that trusts read CHI reports attentively as part of their preparations for being reviewed, but no indication that they scan them systematically to learn about good practices. They want to learn about CHI's approach to inspection rather than about what other trusts are doing. This in itself may help them to concentrate their minds, though learning how to manage a review successfully should not be confused with learning how to improve the quality of care.

So our conclusion on this point must be agnostic. As researchers, we can only sympathise with CHI's problem in obtaining 'evidence' about its effectiveness, because the difficulties in so doing are remarkably similar to attempts to show that research can influence policy or practice. Research tends to be used selectively when it meets the needs of policy-makers and practitioners, and may also have a general 'sedimentary effect' in changing perceptions and attitudes. It may have its greatest impact when the source is forgotten, and the same may be true of CHI.

Lessons for the NHS identified by reviewers

- The trust has developed a horizon-scanning service with the Evidence Based Child Health Unit and the university that provides frontline staff with the most recent evidence in relation to research.
- The trust has appointed a bereavement care co-ordinator, and has developed a bereavement care pathway to support the care of dying children and their parents or carers.
- Follow-up clinics for stroke patients at three months, six months and one year were well run. Patients and carers valued this service highly. The trust is actively using the experiences of current stroke patients to develop and improve services for the future.
- All patient information is to go through a readership panel for quality checking. Patients are to be members of this panel.
- The trust has introduced 'greening teams' to modernise services across the trust. The term 'greening' is linked to traffic light colours, and green indicates that performance is of high quality and meeting targets. These teams are multi-disciplinary, and use staff expertise to identify areas for improvement in clinical service delivery. This work is supported by in-house facilitators, and feedback from staff is very positive.
- Volunteers have job descriptions, paid travel expenses and access to staff benefits. The service is linked to the patient advice and liaison service (PALS), and volunteers are given training to extend their roles – for example, as 'hand holders' in theatre.

Source: CHI reports

The NHS on CHI

This section looks at the experiences of trusts in being reviewed and inspected. The trusts that gave evidence to the study provided acute and specialty services in all parts of England. They included high and low scorers on CHI's clinical governance reckoning, and – in theory – the sample allows for bias for or against CHI and the reviews. As it turned out, most of the trusts' observations and comments were both positive and negative, irrespective of whether they had had 'successful' or 'unsuccessful' results.

Our analysis is primarily based on interviews with staff at 25 acute and specialty trusts that had been reviewed by CHI. The trust staff interviewed were recruited in a variety of ways: by letter and telephone as well as at seminars and meetings. In some cases, staff volunteered themselves, having heard of the study. Discussions were held mainly face-to-face at meetings, while some were carried out by phone. Information from trusts and other interested bodies was collected throughout the duration of the study, and began soon after the first reviews were completed.

We asked chief executives and other senior managers, clinical governance leads and some senior medical and nursing staff about pre-review, review and post-review experiences. We also interviewed field reviewers and CHI review managers (most of whom were health service staff, ex-health service staff or had allied health connections) about their review experiences. Our discussion is further informed by some of CHI's own 'feedback from the field' that we recorded at CHI meetings, and from interviews with CHI staff.

The methodology was designed to illuminate the range of review experiences in the NHS, rather than to identify a representative sample of trusts. In what follows, we do not make any attempt to quantify the different views expressed but use them to illustrate the issues that came up.

We also draw on the findings of the two NHS Confederation surveys. The first was carried out in November 2001 with a response rate of 70 per cent (NHS Confederation 2002). The second was carried out roughly a year later, with a response rate of 37 per cent (NHS Confederation 2003). The first survey was followed by a seminar hosted by the Nuffield Trust, and attended by trust chairs, chief executives and other senior staff, at which the written responses were discussed with the participation of the researchers.

The confederation also allowed us to follow up their second survey by contacting respondents – with the permission of the trusts concerned – and exploring further their experiences of being reviewed. The second confederation survey was aimed at trusts reviewed in 2002. Almost one-third of the respondents had experienced CHI's revised and streamlined review processes. These changes made by CHI – in part, at least, in response to earlier criticism – included slimming down the data requirements, compressing the review process and shortening the final report on the assumption that the new system would be less burdensome for both the reviewed and the reviewers.

General views

Most of the chief executives and senior staff interviewed were convinced of the logic and sense and – to a greater extent, the inevitability – of an inspected and reviewed NHS. Moreover, many trusts found the act of being reviewed a mainly useful – if strenuous – experience. A few managers said they were provided with insights into their own organisations before and during the review, while by far the largest number said that they did not learn anything they did not already know but that it gave them a lever with which to crank up staff and systems. Trusts also acknowledged that CHI was a new and inexperienced organisation, and that they themselves had some responsibility for making the review system work.

In line with our own discussions with trust staff, the first confederation survey showed that 60 per cent of responding trusts were overall positive about being reviewed, 20 per cent were negative and 20 per cent were neutral. The second confederation survey, with its much lower response rate, showed rather less satisfaction and rather more criticism: 44 per cent were positive, 25 per cent were negative and 27 per cent were neutral. But when asked about the impact of the CHI review on clinical governance, more trusts were positive (50 per cent) than those who said it made no difference (25 per cent), and only 13 per cent were negative, saying that the review had held back progress with the development of clinical governance.

The confederation's interpretation of this was that many of the trusts criticising CHI's methods nevertheless accepted the principles behind the CHI review as necessary, at this stage, to the development of the NHS. Inspection and review were seen to be a fundamental means of changing and improving health services.

Our interviews confirm and amplify the survey findings. As the following quotations show, attitudes towards CHI tend to be a mixture of support in principle and criticism on points of execution:

Everybody in the trust was ambivalent about being 'CHled'. No one was anti – just thought it may be unnecessary or a waste of time.

(Clinical governance lead director)

Inspection and inspectors are very good things in the NHS. They make us do things we ought to have done anyway. And the preparation for inspection really concentrates our minds on getting our organisations into top condition.

(Trust chairman)

The second, more positive, trust quoted above had had a mixed experience of the review processes. Aside from believing NHS inspection to be a good thing, and in spite of doing well with generally high scores, its perception of CHI and the review processes was more negative:

There was a meanness and pettiness about the CHI process. We have not done badly and we are not a paranoid organisation, but we do feel we have been marked down.

(Chief executive)

There were some criticisms also of CHI's ambiguous inspection role: professedly developmental while practicing in a punitive and policing manner:

... CHI pretends to be a nurse – lots of caring and seemingly non-judgemental talk but then acts like a customs inspector – lots of cold, hard requests for information and evidence of innocence. CHI pretends to do one thing and then does another. CHI is there to punish but pretends to help and develop us. Almost as if they don't have the competence to do the latter.

(Trust chairman)

We discuss these points further in the following sections.

The calibre of review teams

From the start, it was clear that the credibility of the review team was crucially important to the smooth running of an inspection visit. Reviewers were aware, through the NHS grapevine, that whether an encounter between review team and trust was 'successful' or 'unsuccessful' depended largely on the status and experience of team members. The delicate balance in the relationship was seen as being influenced in particular by team managers.

Reflecting on his experience, one reviewer commented that the inspection had been successful essentially because trust staff had been impressed by the high calibre of the team members. Their ability to communicate with senior managers, talk the same language and have a shared understanding of the difference between systems in place and the practice of clinical governance, for example, had recommended them to trust managers:

What we learned most critically was that the review team must have credibility with the trust's top managers, and credibility for trust staff is associated with high-ranking team members.

(Medical reviewer)

A chief executive who was reviewed early on was aware, as a CHI reviewer himself, of the continuing variability of reports, even after the new, shorter ones were introduced after the redesign of reviews. His conclusion was that the quality of the review process and the report alike hinged on the quality of the review manager – as did the impact of the draft report presentation to the trust. CHI itself was aware by autumn 2001 that some trusts did not find review teams credible and as a consequence, pressed for more high-calibre reviewers to be recruited. It was even suggested at CHI headquarters that recruitment policies should aim to establish a CHI reviewing attachment as part of the career ladder for senior NHS staff – particularly chief executives.

The first survey of the NHS confederation elicited similar comments about the variable quality of review teams. The most common concern raised was that the review team was not acceptable to trusts. Some were seen as inadequately qualified, as well as too negative and critical, making the review experience less developmental and supportive than expected. The concern about the quality of reviewers was raised not only in the written survey but also in the seminar, where participants – mainly chief executives –

expressed concern about the lack of experience and expertise of review teams sent to inspect them.

It was clear that senior staff in trusts were not prepared to be judged by fellow NHS workers whom they considered to be ill-equipped for the job of inspecting NHS organisations. The calibre of review managers was thought to be crucial to the success or failure of a review process.

Why do reviewers see mainly bad things?

In its training and its presentation of its philosophy, CHI puts much emphasis on accentuating the positive as much as the negative. And indeed, all review reports carry a section on 'areas of notable practice'. The quotes listed in the box below suggest that some report writers may have to dredge rather desperately to find encouraging examples. Nevertheless, one theme to emerge from our interviews was a tendency to skew inspectoral scrutiny to issues of concern rather than incidences of good practice.

The review manager keeps reminding reviewers that the search is for positive and negative service practice, but it is as if there is a natural tendency in inspection to go for the negative.

(Chief executive and CHI reviewer)

Good practice cited in CHI reports

- The volunteers project is a successful partnership funded via a government bid. Volunteers are given an induction and formal training.
- The trust has set up a programme for internal staff secondment to the patient advice and liaison service (PALS).
- The orthopaedic department is active in developing patient information, some of which is also available on the intranet, and staff orientation materials.
- The trust has established a service-user council, which is a sub-group of the clinical governance board. The council is the main decision-making forum for issues relating to service-user involvement, and is responsible for strengthening and developing service-user involvement across the trust.
- The trust recognises the multicultural diversity of patients, carers and staff, and actively promotes the ethos that staff should be courteous and respectful at all times.
- The patients' forum produces reports relating to patient and public involvement that are made public after being reported to the trust board.

Pre-visit data

CHI's demands for, and use of, data prompt two kinds of complaint. The first is about the quantity. The second is about the quality of the analysis. To start with the issue of quantity, the 2001 confederation survey (NHS Confederation 2002) reported that a 'large majority of trusts' believed that inappropriate or unnecessary documents were requested, imposing a burdensome workload on trusts. The volume of documents and data was listed as the second highest priority for CHI's attention. Subsequently, CHI did

slim down its demands for data quite dramatically. But the 2002 confederation survey (NHS Confederation 2003) showed continued concern among NHS trusts, identifying precisely the same issues as in the previous year. Not only were the requirements excessive but, some trusts complained, they were also unclear. All these themes emerged in our interviews:

The data collection itself was massive and a huge amount of work for the trust. There were boxes and boxes of data.

(Director of quality improvement)

The data collection seemed to be mammoth and endless. CHI kept on asking for additional information.

(Clinical governance manager)

The pre-visit data collection consisted of collecting volumes and volumes of papers and figures. CHI did ask for key areas of information but it was not specific. The trust struggled with this work.

(CHI co-ordinator)

The interpretation of such comments is, however, complex. If trusts had complaints about CHI's demands, CHI could cite examples of trusts that had not got their information production in good order, not even for their own management use:

Complaints from trusts about the vague and unfocused requests for information from CHI say as much about their poor management of information as about CHI's lack of clear focus.

(CHI director)

So one conclusion might be that, certainly at the beginning, CHI was fishing for data without having a sense of what it was looking for: there appeared to be no clear conception of the key elements of performance. Another, however, might be that complaining trusts lacked the data systems required to manage themselves and assess their performance and, like other inspected bodies, translated their own sins of omission into sins of commission by the inspectors. If trusts send boxes and boxes of data, it may be as much a gesture of protest as a sign of regulatory unreasonableness.

There was also much criticism of the quality of data analysis. The issue emerged strongly in both the confederation surveys and again in our interviews. Some trusts focused on this problem with particular irritation. In their view, the assembly and analysis of pre-visit data by CHI was incorrect and misleading.

It might be expected that the trusts reviewed after CHI's redesign exercise would have been subjected to a more sophisticated, and therefore more acceptable, data collection and analysis process. But while there is some evidence from our interviews that the change from a hand-cranked to a more streamlined system did indeed improve matters, trusts maintained that some flaws remained. The first quote comes from a trust reviewed in 2001, pre-reformation. The second comes from a trust reviewed in late 2002.

Data collection was very difficult for the trust. We were given an awful questionnaire with lots of repetition – and badly worded. We complained, and the questionnaire was given some improvements. After the questionnaire was completed, the trust was

visited by the review manager and an analyst, who were not very helpful, and their data analysis was poorly done.

(Chief executive)

After we got our data analysed and we had evidence tables returned to us, we could see that a lot of the analysis was inaccurate. Some tables were worse than others in accuracy. CHI's idea of factual accuracy was not ours. The tables were very variable in quality. We had then to drip-feed the changes to the tables back to CHI as we made them more accurate. We were then left with those tables that we could not reach agreement on. CHI did not agree with our changes, and we did not agree with CHI's analysis...

We argued about what is 'factual accuracy', where I explained that the data analysts were taking four or five sentences and condensing them into one sentence, regardless of the actual meaning of the information. This was, for us, changing facts into factual inaccuracy. I was of the impression that CHI was using word searches to arrive at data analysis and were losing the meaning of the data.

(CHI co-ordinator)

Once again, the interpretation of such comments is not simple. The examples of complaints by trusts that were reviewed early on may say as much about trusts themselves as they do about CHI's methods, as the CHI director quoted above observed. Equally, later complaints about CHI's poor analysis of data may also say as much about trusts as about CHI, but with an important difference. Perhaps the trusts asking for higher quality analysis from CHI are those that have their own high-quality data production. Both the quotes above come from trusts with average-to-good clinical governance scores. Similarly, the trusts that did not complain about poor data analysis may have had lower expectations of data all round, including those from CHI, as well as those in their own organisations.

Tools and instruments

There was a range of comments from trusts about the tools and instruments that CHI used for assessing health services. There were criticisms – for example, of the clinical governance components as indicators of performance, and also of their inconsistent application between trusts. Some trusts had doubts about the relationship between the unscored parts of a review and the scored components. Others were quite positive about the use of clinical governance by CHI, approving of their ability to embrace all aspects of service delivery. The following quotes illustrate the range of views expressed.

On the positive side, a clinical governance lead said:

Many trusts and hospitals don't have systems of clinical governance in place, and CHI has established that they must, and has seen that systems have to be set up. And where hospitals have systems in place, CHI checks to see whether the systems are being translated into clinical practice. Our trust has systems in place which are not actually functioning at clinical level, and CHI gave us the push to take systems down into practice.

(Clinical governance lead)

On the critical side, a chief executive who was also a CHI reviewer thought that clinical governance components were insufficient to assess performance, and felt that this was shown by the way CHI and review teams slipped in and out of using unscored information to inform the review and the scores achieved by trusts:

If asked whether score are influenced by factors external to clinical governance pillars' scores, the answer is sometimes yes and sometimes no. The effect of influences of the environment and expectations of trusts varies between different reviews. The general state of the trust – mergers, de-mergers, crises, the impressions of reviewers of the cleanliness of the trust, the feelings of staff interviewed, the attitudes and feelings of the board and management etc – these things will possibly alter clinical governance grades up or down slightly, but not in extreme cases.

(Chief executive)

Other interviewees were not at all impressed by clinical governance as a measure of health service performance. A senior trust manager thought that CHI's reviews did not look further than systems in place:

Inspection of our clinical governance did not say anything about actual clinical and managerial work being done by the trust. The review observed only the machinery of clinical governance, and did not look at the impact on clinical practice. What CHI sees as inspection does not capture the whole story at a trust.

(Clinical governance lead)

A trust chairman who was interviewed challenged CHI's components and definition of clinical governance. He argued that his trust's own version of clinical governance – a system set up to monitor clinical improvement – was better than CHI's. He was also convinced that the trust had been downgraded because of this, since the CHI report had declared it to have 'no overall strategy for clinical governance':

Clinical governance is our hospital strategy, but we call it 'clinical improvement'. When we put this to CHI reviewers, they said 'we are only concerned with clinical governance and nothing else can be substituted'. It seems that the idea of clinical governance has taken over from the idea of 'quality' which is also an industry in the NHS.

(Trust chairman)

One chief executive complained about the selection of the clinical governance components, asking: 'Why these? Why not others?'. He argued that there had been no discussion of the components, and that he and his fellow chief executives could offer suitable additions or alternatives, based on their knowledge and experience of the NHS. He offered CHI some leading questions about performance measures, and pleaded for more consultation within the NHS. The staff turnover component on the performance checklist was, he argued, a nonsense, and in need of re-modelling, because in the NHS staff come and go for positive as well as negative reasons:

Why are some areas of performance turned into clinical governance components and scored for performance, and not others? What is the relationship between each component? Why does CHI not explain and discuss the ideas behind its choice of performance measures? The seven 'pillars' in total do not give an indication of clinical quality. Just how do we arrive at clinical quality, and what other qualities should we

be including in NHS performance assessment? CHI makes statements about methods of assessment without justifying them – and expects trusts to accept its word.

(Chief executive)

A further theme to emerge from our interviews – as it did from the confederation surveys – was scepticism about the methods used by CHI reviewers to come to their conclusions and recommendations. There were many complaints about reviewers relying on anecdotes – complaints not stilled by CHI's invocation of 'triangulation' – a research method that requires three pieces of independently observed evidence. Even a CHI reviewer admitted:

Some of CHI's recommendations for action following on from reviews do not stand up, since they are a mixture of the formal clinical governance findings and the informal impressions of reviewers. In other words, the CHI approach creates a contradiction in methods. Some of CHI's recommendations for action are more obsessional than practical and reality based.

(CHI reviewer)

Two other criticisms about CHI's methods emerged from the confederation surveys as well as our interviews. First, trusts were concerned that the clinical governance scores did not make sufficient allowance for the different circumstances of trusts – especially for those involved in reorganisations (mergers, de-mergers, and so on) and for major changes in senior management. Moreover, many trusts saw CHI as inconsistent and uneven in reviewing and reporting, while at the same time insufficiently flexible to judge particular circumstances of individual trusts and allow for these variations. This inconsistency lying alongside rigid inspections makes for the worst of all possible regulatory worlds.

Second, in the words of a chief executive who is also an experienced reviewer, one cause of confusion is that CHI does not know what it is looking for, or what it ought to be looking for, in NHS performance. This strongly felt view was shared by many other managers in the NHS.

CHI's administrative and organisational standards

Interviews with trusts pointed to some administrative friction with CHI. Given that CHI was busy setting up and expanding its tasks and operations for most of its organisational life, this is not surprising. There were comments in interviews about CHI changing review dates and timetables at short notice and leaving trusts to re-jig meetings and interviews with busy staff:

The review visit was eventually moved twice by CHI, and as review co-ordinator I had the job of explaining to health professionals that old appointments were gone and new ones had to be fixed, when they had already carefully allocated their diary time. I also had the difficult job of co-ordinating the whole thing. The visit-arranging exercise was a shambles. We wondered about the continuity back at CHI headquarters.

(Clinical governance manager)

One trust was rebuked for being late with its pre-visit data collection, only to be met with an absence of staff when hastily collected data was delivered by taxi to CHI headquarters:

We were two days late with the data collection, which we thought was okay, but CHI was appalled and disapproving. We eventually took the collection in a taxi to CHI's headquarters, only to find that our analysts were away and no one was actually waiting to use the data. We were told, however, that punctuality is a question of principle, and lateness reflected badly on the trust rather than being an inconvenience to the analysts.

(CHI co-ordinator)

One interviewee made a criticism of CHI's poor internal communications:

I am finished with reviewing, I think, unless I get some tender loving care from CHI. CHI's administration is not hot, and CHI is not good at communications and management. Its communications are actually bad. If you get in touch with them for whatever reason, they are slow in responding. Reviewers are kept waiting for replies to queries.

(Chief executive and reviewer)

Criticisms of CHI's administrative standards were also raised in the second confederation survey with examples given of poor communications and delayed and failed responses to trusts, telephone calls and enquiries sent to CHI headquarters. One chief executive remarked:

If I were doing a report on CHI, it would be along the lines 'could do a lot better'.

(Chief executive)

In the concluding section, we discuss in what respects CHI – or rather, its successor body – could do better, in the light both of our own analysis and of the comments from the inspected.

Conclusions

The balance sheet

The Commission for Health Audit and Inspection represents a new birth rather than a reincarnation. CHAI is not a revised model of CHI. It has a new remit, an expanded role and a somewhat different philosophy (Commission for Healthcare Audit and Inspection 2003). The break with the past is not complete, of course – there are elements of continuity. CHAI's emphasis on promoting improvement echoes CHI's. Many of CHI's staff will move across to the new body, though the fact that Dr Peter Homa, from CHI, survived only a matter of weeks as CHAI's chief inspector may have symbolic significance. In the short term, the new commission may have to lean heavily on the experience and methods of its predecessor while it develops its own strategy and tools.

It would be wrong, however, to interpret this as a total repudiation of CHI, or as evidence of its failure. The balance sheet of its achievements and failures is complex, as our findings suggest. On the debit side, CHI failed to satisfy all the expectations of ministers and the Department of Health. Its methodology for inspecting trusts has significant weaknesses. It continues to attract criticism from within the NHS. The NHS still sees itself as a collection of over-inspected bodies, with different regulators making overlapping and competing demands for data and time. The agenda for rationalising the system remains long (Public Sector Team/Regulatory Impact Unit 2003). It has not resolved the tension between looking at the 'arrangements' for ensuring quality (structures) as distinct from looking at quality directly (processes and outcomes).

On the credit side, starting with a blank drawing board, CHI built up a new organisation: a massive managerial undertaking. It has carried out its remit of inspecting all trusts within the timescale set for it. It has made the notion of inspection acceptable to NHS managers and professionals, despite their reservations, overcoming considerable initial suspicion. It has taken on a whole series of new tasks. Our study, focusing as it does exclusively on CHI's reviews of acute trusts, understates its achievements in this respect.

Some of the items on the debit side reflect circumstances outside CHI's control. In developing its methodology, it had to work within the straitjacket of the seven pillars. It invested considerable effort in negotiating memoranda of understanding with other inspectorates, but lacked the authority to compel, for instance, the development of a common set of data requirements.

Perhaps the best way of interpreting the demise of CHI and the creation of CHAI is as an example of policy learning. In part, the learning has been based on CHI's own experience, on seeing the strengths and weaknesses of its approach. It partly reflects changes in the way the regulatory task is perceived and defined, influenced by the Bristol Inquiry (Kennedy 2001) and other factors. So, most importantly, CHAI's remit is to report on the quality of care, rather than on clinical governance as such, and to inspect against standards – quite apart from taking over the Care Standards Commission's responsibilities for the private acute sector of care and some of the Audit Commission's functions.

In short, the substitution of CHI by CHAI can be seen as part of the constant process of re-balancing or re-calibrating the regulatory process. The following points draw out some of the implications of our study – informed by the more general literature on inspection – as a contribution to the policy learning cycle.

Inspecting against standards

The decision that CHAI is to inspect against standards (yet to be devised) represents welcome recognition that the lack of specificity and precision in CHI's methodology has been a source of weakness. However, inspecting against standards – fleshed out by criteria for assessing whether those standards have been reached – has its own pathology. It can (as CHI rightly recognised) lead to a box-ticking approach.

Compare the reports of the Standards Commission (see, for example, National Care Standards Commission 2003) on private hospitals with CHI's reviews. The former certainly score on precision and detail. But they give no overall picture to the reader – in particular, to the lay reader – of whether good quality care is being delivered. The standards, and the complementary criteria, may correctly identify the structures and process required to ensure that the care does not fall below an acceptable minimum, but they give no sense of whether good outcomes are being achieved. But neither do CHI's clinical governance components assess outcomes.

Assessing performance

Assessing performance by outcomes is, of course, the elusive aspiration of all inspectorates. It is elusive because, as yet, outcome measures tend to be ambiguous, contestable or difficult to design – witness CHI's difficulties in capturing the patient experience. Given the inadequacy of what we have, it is probably inevitable that proxy measures have to be used. But if they are to be used, it is important to devise a parsimonious model based on identifying key characteristics that define an organisation that is well-run or poorly performing. Such key characteristics might, for example, include staff absences, patient outliers and the use of agency staff (how many, which services and which shifts).

Similarly, we need to know more about the relationship between systems and performance. The act of faith represented by the seven pillars of clinical governance needs to be translated into an empirical inquiry into which aspects of structure and process are related to good outcomes. For a study of the implementation of clinical governance which, however, does not question the concept itself, see National Audit Office (2003).

Consistency and comparability

A recurring theme in our analysis has been the difficulty of using CHI reports for purposes of comparison, whether between individual trusts or geographical regions. This reflects the fact that CHI was not set up to provide this kind of 'benchmarking' information, and its review methods were not designed to do so. However, this inevitably limits the usefulness of reports to policy makers, purchasers (or 'commissioners', as they are cosmetically now known) and the public. So the aim of the new inspectorate should be to provide information that is both consistent and comparable, if only on a limited

number of critical dimensions. This will be all the more important if the Government succeeds in implementing its goal of patient choice.

Working with the Audit Commission

One of the difficulties in interpreting CHI reports is that they provide no analysis of how trusts manage their resources. There are frequent references to shortages of staff or capacity, but such shortages could mean either that the trusts concerned are under-funded, or that they are not using their resources to best effect. As the Audit Commission's studies have demonstrated (Day and Klein 2001), in many cases the latter explanation seems plausible.

The transfer to CHAI of the Audit Commission's responsibility – and staff – for studies of the efficiency, effectiveness and economy with which resources are used should be a source of strength for the new body. It will give CHAI the opportunity to add another dimension to its inspections. However, the Audit Commission's strength derives from its network of district auditors. It is their annual audits that provide the necessary context for understanding how trusts are performing in terms of managing their resources. Much will depend on the two commissions maintaining a close relationship, so that CHAI can draw on the expertise and local knowledge of district audit.

Challenges for CHAI

CHI was set the task of reviewing all trusts on a four-year cycle. In contrast, CHAI will have more freedom. It will have the flexibility and scope for moving towards the two principles laid out by the Better Regulation Task Force: targeting and proportionality (Better Regulation Taskforce 1999). This means it will be able to tailor the frequency and 'depth' of inspections to the circumstances of individual institutions. A new balance will have to be struck between desktop and hands-on inspection. Once again, however, this will depend crucially – as CHAI has already recognised – on developing a parsimonious set of indicators (see 'Inspecting against standards', above) which signal actual or impending problems and identify chronically under-performing organisations.

Equally, it will require a strong team of in-house analysts. Starting from scratch, CHI had no such team, and initially it contracted out much of the work. The result, as we have seen, was considerable friction, as trusts challenged CHI's interpretation of the evidence – particularly in the early days of inspection, but also latterly, in spite of the 'beefing up' of the analysis function. To avoid such unnecessary friction, and to ensure CHAI's legitimacy, it is essential that any data analysis is expert and authoritative from the start.

Examining relationships between institutions

Until now, inspection has been institution-specific. However, the emphasis in the NHS is increasingly on creating networks and promoting the smooth flow of patients between different sectors and institutions. So a new balance will have to be struck between inspecting institutions and inspecting the relationship between them as it affects the population being served. Here, the methodology developed jointly by CHI and the Audit Commission for reviewing the implementation of national service frameworks – which is largely population based – may provide a model.

However, in pursuing such a strategy, it will be important to recognise that it may create a vacuum of accountability. If there are failures of co-ordination, who is to be held responsible?

From evaluation to dialogue

CHI has made much of being a learning organisation, and it seems likely that CHAI will develop the same theme. Indeed, CHI has done a great deal of introspective evaluation of its methods and processes. However, there is an important difference between holding up a mirror to one's own performance and looking at how others see that performance. CHI's record suggests that learning may be excessively equated with commissioning evaluations and rigorous research in the academic mould, to the neglect of more informal ways of generating feedback. So, for example, CHI only organised a conference of reviewers in November 2002 – which, as one reviewer remarked (cited in Commission for Health Improvement 2003b, p 30), was 'long overdue'.

Similarly, it was left to the NHS Confederation to organise a survey to gather evidence about the reaction of trusts themselves. Learning from the experience of the reviewed and the reviewers alike, eliciting their views in seminars and meetings, should be a matter of routine. Only by establishing such a dialogue can an inspectorate inform itself about how its ways of operating work out in practice.

The dangers of rhetoric

It is important for any inspectorate to exercise verbal restraint. Incantatory prose – whether about being patient-centred or being developmental – can backfire. It may set up expectations that cannot be satisfied. Modesty in claims made and restraint in the use of words may carry more conviction than rhetoric presented on glossy pages and padded out with photographs. In this, CHI was conforming to the latter-day NHS (and beyond) style of relying heavily on feel-good words and the magic of presentation (see, for example, Commission for Health Improvement 2003c). In this respect, certainly, its successor should represent a new departure.

Applying realistic timescales

CHI was under pressure to deliver from the moment it was set up. Its record suggests that this sort of pressure may be counter-productive. It fosters an attitude of protective self-defence, and inhibits experiment and debate. CHI has been most innovative in precisely those of its activities – not covered by our study – where it has not been on the treadmill of producing inspection reports on a tight timetable. Developing inspectorial strategies and methodologies is an evolutionary process, as CHI has demonstrated. It is therefore important that CHAI should be allowed time for so doing.

Appendix: Translating CHI reviews into evidence about London and the NHS

This appendix considers what light CHI reviews and reports have thrown on London's health services. Is it possible, for example, to draw a picture of London trusts and make comparisons with those in other parts of the country? This question formed part of the original design of our study, at a time when it was not clear whether CHI's methodology would allow it to be answered. In the outcome, CHI's methodology was not devised with the primary aim of geographical analysis. And, as shown in the main text, there are formidable problems even in making comparisons between individual trusts because of problems of inconsistency and subjectivity.

What follows is an attempt to use CHI reports for a purpose for which they were not intended, and is inevitably impressionistic in character – though the analysis does yield some significant leads for anyone interested in London's health services. Our perspective is that of an imaginary lay reader who – unlike policy-makers or professional healthcare watchers – has to rely exclusively on CHI reports for information about London.

CHI itself has commissioned regular reports that track issues arising out of the analysis of its reviews. These reports were based on clinical governance components and the unscored areas of review reports, including patient's experiences, strategic capacity and examples of best practice. These 'tracking reports' also provided an analysis of geographical variations, based on clinical governance scores.

For example, the October 2002 tracking report showed the frequency of scores by health region. It found that London trusts were more likely than organisations in England and Wales as a whole to have been assessed with the minimum score of one for the clinical governance components of information: patient involvement and clinical audit. Taking all components together, they were as likely as organisations in England and Wales as a whole to be assessed with scores of I and slightly more likely to receive scores of III or IV.

This report also showed that trusts in the north were less likely than those in England and Wales as a whole to have been assessed with low scores for every component of clinical governance. Furthermore, taking all components together, trusts in the north were less likely than organisations in England and Wales as a whole to be assessed with I scores, and slightly less likely to receive scores of III or IV (Walters 2002).

Based on this information, with additional evidence from analysis of trusts' action plans (which were required to be drawn up after reviews), and using the information about good practice observed in reviews, CHI issued a press release entitled 'South North Divide in the NHS' (Commission for Health Improvement 2002b). Using the October tracking report on 175 inspected hospital trusts and other NHS organisations, CHI concluded that typically trusts in the north and Midlands had fewer areas of concern and stronger overall clinical governance assessment scores. The chief executive, Dr Peter Homa, was quoted as saying: 'From our inspections a clear picture is emerging. NHS bodies in the North and Midlands have better working systems in place to deliver high quality care than those in the South' (Ibid).

Although CHI protested that this was only a broad picture, there were doubts about the implication that somehow London fared worse in health services than other parts of the country. It was argued that CHI's review methods were not actually designed to make these kinds of comparisons, neither between organisations, nor between geographical areas. This prompts a number of questions:

- How reliable a picture do we get from CHI's analysis of its own reports?
- Is CHI's press release in November 2002 right in suggesting that the population of London has, in some ways, a worse deal in health services than trusts elsewhere?
- How much do we know aside from CHI reports about health and health services in London?

This part of our study sets out to interrogate the CHI reports further, and to ask first what they say about London, and second whether London is noticeably different from other parts of the country. In order to answer these questions, we began by looking at the reports of acute and specialty trusts in London, and then examining some reports from trusts outside London. What follows is an analysis and discussion based on this material. During the period of this study, many trusts (including those within the London area) were engaged in mergers, part-mergers or other large-scale organisational rearrangements. So we decided to scrutinise whatever CHI reports on acute and specialty trusts were identifiable, available and appropriate at the point of investigation. In the event, we analysed 31 London reports. This was the number of reviews published by April 2003.

As discussed earlier, the theory of the CHI review processes is that trusts are assessed on the basis of their performance, against the 'seven pillars' of clinical governance. Reviewers do look at other aspects of performance besides clinical governance, and, of course, the informal 'inspector's nose' operates (see p 28). But since clinical governance scores are supposed to seal the fate of trusts in a CHI review, it was sensible to start by examining these factors, much as CHI itself has done regularly in its tracking reports.

There are, however, one or two cautionary notes to be issued beforehand. First, CHI has revised its review methods during the three years it has been operating and has reorganised the components, changing the titles and the emphasis of some 'pillars'. So there is no way that it can be assumed that clinical governance scores in earlier reports are entirely consistent with later ones. Second, we have noticed that scores in one London report had been inaccurately transcribed into CHI's published lists of performance scores. Some scores were added to the wrong columns on the score sheets which, although not serious in the overall scheme of things, did cast a slight doubt on the published scores of individual pillars of clinical governance. However, the net result of overall scores for each trust remained the same.

These reservations on clinical governance do mean allowing a degree of elasticity in their interpretation. So what follows is an analysis of clinical governance ratings as they stand. We are, nevertheless, confident that tables 4–7 give an accurate picture of trusts' performance as scored by CHI's clinical governance components – which is, of course, different from saying that the clinical governance scores necessarily give an accurate picture of the performance of trusts. Note that the national figures for all trusts differ slightly from those that appear in Table 3 (p 27). This is because the figures used here refer to the period ending April 2003, rather than August 2003, to make them comparable to the London data.

In Table 4, and in line with CHI's own interpretation, our analysis of clinical governance scores for London shows that the profile of the greater metropolis is largely similar to the rest of the country. That is, London appears to be an average player in health services performance, with the middle II score (some progress in achieving clinical governance) predominating for each of the seven technical components, with the III score (good progress) and the I score (no progress) occurring less often. The IV score (very good) had only been awarded four times in the whole of the CHI acute and specialty trust reviews as this paper was being compiled, and two of these were in London trusts.

As also recorded in the CHI analysis, the difference between the clinical governance performance profile of London trusts and that of the country as a whole is that London has a smaller proportion of trusts that are making little to no progress overall in clinical governance (14 per cent), compared to the national average of 19 per cent. Of the IV scores gained by London trusts and the Midlands and Eastern Group trusts, three were awarded for research and effectiveness, and one for clinical audit.

Table 4: Total clinical governance scores for London, and for all trusts

	I	II	III	IV
London clinical governance scores	14%	65%	20%	0.8%
All trusts' clinical governance scores	19%	63%	17%	0.6%

Source: CHI reports for London and CHI's own analysis of scores

For individual components, as Tables 5 and 6 (see overleaf) show, there are a few factors of note that imply differences between London and the country as a whole. For example, London more frequently scores well (III) for clinical effectiveness (32 per cent) and education and training (45 per cent), compared to the countywide scores of 22 per cent and 30 per cent respectively. This is partly explained, however, by the clinical effectiveness component also having been referred to as 'research and effectiveness' earlier, and probably reflects the concentration of research and teaching in London. Staffing and staff management issues in London are given a low score less often (3 per cent), compared with 19 per cent countrywide. As discussed earlier, the problem with the staffing component in clinical governance is too widely defined to lend itself to illuminating analysis.

Table 5: London clinical governance scores

Components	I	II	III	IV
Clinical audit	16%	74%	1%	–
Risk management	13%	81%	1%	–
Patient and public information and involvement	16%	68%	16%	–
Education and training	–	55%	45%	–
Clinical effectiveness (formerly 'research and effectiveness')	10%	52%	32%	6%
Staff and staff management	3%	81%	16%	–
Use of information	39%	48%	13%	–

Source: CHI reports for London

Table 6: Clinical governance scores of all trusts

Components	I	II	III	IV
Clinical audit	15%	68%	13%	1%
Risk management	20%	60%	20%	–
Patient and public information and involvement	23%	69%	8%	–
Education and training	4%	66%	30%	–
Clinical effectiveness (formerly 'research and effectiveness')	22%	54%	22%	2%
Staff and staff management	19%	65%	16%	–
Use of information	31%	58%	11%	–

Source: CHI's own analysis of scores

So far, our analysis of clinical governance scores is similar to CHI's own, while our interpretation is that few insights into London health service performance can be gleaned from this. The next questions are: how does London look if its trusts are compared with others, and which group of trusts would provide an appropriate comparison?

As London is often described as being deprived in terms of both health status and health services, comparison with, for example, the Northern Group of trusts, with its various socio-economic and health deprivations, might well be appropriate. However, the newly convened Northern group has too great a number of heterogeneous areas within its bounds. This makes a comparison between the whole of the North Directorate and London problematic.

One single health authority within the Northern group was then considered as a possible alternative, and Cheshire and Merseyside strategic health authority was selected. The disadvantage of this choice was that Cheshire and Merseyside had fewer acute and specialty trusts than London (Commission for Health Improvement 2001a). However, the advantage was that its trusts operated within a variety of environments and populations,

ranging from affluent to deprived, rather similarly to London. Given that no comparators are perfectly matched to London, and given that we do not propose any sophisticated statistical analysis, but adopt a broad-brush approach, this seemed a reasonable second-best strategy.

Table 7: Clinical governance scores, by geographical area

Geographical area	I	II	III	IV
London	14%	65%	20%	0.8%.
Cheshire and Merseyside	9%	76%	15%	–
The north	19%	69%	12%	–
All trusts	19%	63%	17%	0.7%

Source: CHI's own internal analysis of scores

Table 4 compares the London and Cheshire and Merseyside scores, within the wider context of the north as a whole and the country as a whole. The distributions are broadly similar for all areas with II scores predominating and ranged around with fewer I scores and III scores. Of some interest, Cheshire has fewer I scores than London – 9 per cent compared to 14 per cent, and slightly less than half of the Northern Group as a whole (19 per cent). Cheshire has 76 per cent II scores compared with London's 65 per cent and 69 per cent in the Northern Group. Cheshire also has 15 per cent III scores compared to London's 20 per cent and Northern Authority's 12 per cent. Neither Cheshire nor its Northern Authority have any IV scores compared.

In summary, given the dominance of II scores, the differences in clinical governance scores between all areas are not remarkable. They tell us very little about London trusts or about those in the rest of the country. As we have already discussed in the main text, it is not clear what the large numbers of II scores in CHI reviews actually mean: whether the heavy incidence of II scores for all reviews says something about NHS trusts, something about reviewers and the nature of inspecting or something about clinical governance components as indicators of health service performance. These questions remain unaddressed so far.

What do CHI reports explain, other than clinical governance?

If an analysis of clinical governance scores alone cannot tell a satisfactory tale of how things are in, and for, NHS trusts, what can we find out from looking at other sections of review reports – unscored and less obvious – as indicators of performance? There are several parts to the review report that receive, on the whole, less attention than clinical governance assessment:

- the background information gathered for reports, which contains introductory details of population and geography, and profiles of the area in which trusts operate
- reviewers' impressions of the environment and general observations made during a review, including assessments of patient experiences
- trusts' strategic capacity for improvement, recorded by reviewers but not yet part of the scored assessment
- a list of actions that each trust must follow to improve, as well as a list of good points that commend them, some of which are recommended to the rest of the NHS.

This study chose to focus on the first two points in this list: the reviewers' general impressions and observations, and their background information. There are, of course, some reservations in using this non-scored material from reports. Just as we raised cautionary tones with clinical governance scores as indicators of service performance, we do the same for these issues raised in reports.

First, just as clinical governance scores can be influenced by extraneous factors, so, in turn, observations and issues raised can be influenced by individual reviewer perceptions. Second, the non-scored detail within the reports varies over time, with earlier reports being more fulsome on detail than the later revised and shortened reports. So the question is: can unscored details in CHI reports either add to, or give us more – and better – information about London's health services than clinical governance scores do?

The answer to the question is that no information is ideal, and no data analysis is completely without flaws. As we have discussed earlier, it is difficult to separate out the different factors that influence the judgements made by inspectors or reviewers on the performances of organisations. Nor have we yet devised the definitive system for scoring these performances.

This is not unique, however, to CHI reviews – it applies equally to inspectors of other services. CHI itself has tended to seek impeccable theories of performance measurement in order to justify its findings. In spite of this search, noticeable inconsistencies of reporting remain, as do levels of incompatibility between what the scores represent at individual trusts and what the rest of the report says. Variations in and between reports, for example, occur not only on issues mentioned (but not scored) but also in the way these issues are handled.

Given these variations, all that we would claim here is that our experimental strategy of looking at the unscored elements in reports may shed some light on the situation in London. Accordingly, Tables 8 and 9 list the main issues mentioned by reviewers – either positively or negatively – in review reports in London's acute and specialty trusts, and in a 'matching' list from Cheshire and Merseyside. Many, but not all, the issues mentioned in the London and the North West trusts were similar. What is interesting, however, are the differences between positive and negative comments.

Table 8: Most frequently mentioned reviewer impressions in London acute and specialty trusts

Issue	Total number of reports where mentioned	Mentioned positively	Mentioned negatively
Environment/cleanliness	20	3	17
Staffing	16	2	14
Bank and agency staff use	14	1 (low)	13 (high)
	–	–	–
Accident and emergency	8	–	8
Food	8	2	6

Source: London acute and specialty trust reports

Table 9: Most frequently mentioned issues in Cheshire and Merseyside acute and specialty trusts

Issue	Total number of reports where mentioned	Mentioned positively	Mentioned negatively
Environment/cleanliness	13	12	1
Public/patient information and involvement	13	4	9 (but patients happy with services)
Staffing	13	5	8 (4 mixed)
Patient and user satisfaction	11	11	–
Information use	11	8	3
Staff attitudes	8	8	–

Source: Cheshire and Merseyside acute and specialty trusts

The first point to emerge from the tables is that the environment and cleanliness dominate in London and in Cheshire and Merseyside alike. The issue is mentioned in more reports than any other issues in London, and is equal with the three other top issues in trusts in Cheshire and Merseyside. But the overwhelming difference is that in London, levels of cleanliness have a mainly negative impact on reviewers. Reviewers noticed problems of dirty and uncared for environments in 17 London acute and specialty trusts. Three trusts were mentioned as coping well with cleanliness issues. In contrast, the Cheshire and Merseyside environments had a positive impact on reviewers in 12 out of 13 reports.

The second most frequently mentioned issue in London is staffing, which is not only a clinical governance component but was also recorded as a separate issue in many reports. In London, it is the second most frequently mentioned issue (16 out of 31 reports), albeit mainly negatively (14 out of 16 mentions), while in Cheshire and Merseyside it is mentioned as frequently within the issue of the environment and cleanliness – mainly negatively, but also with a more mixed tone, suggesting trusts having problems but attempting to solve them.

What we conclude is that staffing as an issue looms large in London, as elsewhere in the NHS. The broad heading needs to be unpacked, however, in order to reveal the wider, and more interesting, subset of staffing issues operating in trusts. The straightforward labels of staff shortages and turnover, for example, are inadequate to explain observations from reviewers in London and Cheshire and Merseyside alike, which suggest that more complex staffing issues are at work in the NHS.

For example, reviewers noticed and remarked on concerns about the standards of nursing care in some trusts (North Middlesex) and the inappropriate nursing skill mixes in specific specialties (Queen Elizabeth in London and North Cheshire). The use of bank and agency staff was described as 'high' in 13 reports in London, and 'low' in only one report. Bank and agency staff were mentioned in three Cheshire and Merseyside reports – two as high use and one as low.

Shortages of staff appeared to be centred on evenings and weekends in some trusts in London, with high uses of agency and bank staff being concentrated at these times (for example, in Redbridge). It was observed that bank and agency staff work permanently on night shifts in some trusts. At Guy's and St Thomas', high levels of reliance on bank and agency staff was also noted by reviewers.

The new emphasis on using NHS based bank staff rather than external agency staff was not without problems. In London at the South London and Maudsley Trust, sickness absence of staff and the excessive hours of those members of permanent staff who also work as bank staff is currently under review. Outside London, the Mid-Cheshire Trust was noted by reviewers as having low numbers of nurses on its permanent staff as they prefer instead to work as bank and agency staff. The report mentioned an earlier Audit Commission report on ward staffing that commented on the low rates of skilled nursing staff employed actually employed by the trust.

The staff issue recorded by CHI reviewers at the Liverpool Women's Hospital raised an equally interesting issue. Within an under-developed staff management system at the trust, a culture had been established of recruiting nursing staff to fill vacancies at low grades while internal staff are promoted to higher grades. There are obviously more issues at stake in NHS staffing than shortages, turnover and the high cost of bank and agency staff. So it is not surprising that CHI reviewers commented on the lack of adequate information recorded, or made available, on sickness absence and the use of bank and agency staff in some trusts that were reviewed.

As far as we could see, the issue of patient satisfaction was not raised in any of the 31 London trusts, and after examining the Cheshire and Merseyside trusts (where patient satisfaction was recorded positively in 11 out of 14 reports), only became notable by its absence. No negative impressions were recorded by reviewers.

Again we have to consider the meaning of the silence factor in reviews. Does no mention of an issue mean a positive impression, without need for remark, or does it mean that reviewers of London trusts did not think it important to record? It surely cannot be a variation between reviewers, since CHI reviewers are not area-specific. The nearest that London reports came to raising patient satisfaction issues were negative comments in two reports about staff attitudes and poor quality of services causing patient distress (North Middlesex and North West London Hospitals). Staff attitudes and the management of staff was mentioned in eight of the 14 Cheshire and Merseyside reports, and all references were positive. (Mid-Cheshire and the Wirral and Walton Centre for Neurology found staff caring and responsive).

The public and patient involvement issue, which is also a clinical governance component, appeared in only two London reports – on both instances as being not insufficiently developed structurally. One of the trusts was said to be without a patient involvement system in place but having good things going on in practice (Chelsea and Westminster). This lack of comment by reviewers might be because it was not seen as an important issue in London health services, for whatever reason, or it might have been omitted simply because it was 'going along nicely', according to clinical governance requirements. (For a discussion of the meaning of the silence and omissions factor in inspection, see pp 24–25).

Turning to Cheshire and Merseyside, patient and public involvement issues were mentioned in 13 out of the 14 reports – four positive, with lots of involvement going on at the trusts (for example, East Cheshire), while nine were negative, with four reporting that

patient involvement was patchy and partial (for example, Stockport). In all nine negative comments, however, there was the proviso that patients were nevertheless reported as being happy with the trusts' services (for example, the Countess of Chester and the Liverpool Women's Hospital). In summary, the existence of patient involvement mechanisms in trusts is not synonymous with patient satisfaction or patient-centred services, as noted earlier.

Accident and emergency conditions in London were mentioned in eight reports – all negatively. These problems included long waiting times as well as dirty and unhygienic conditions. CHI's survey of accident and emergency (A&E) services found that patients consider some of London's A&E units to be the dirtiest in the country (National Patients Survey 2003a). Nothing was said, however, in the other 23 trust reports.

As discussed earlier, this is another dilemma of the meaning of silence. Does it imply here that 23 of 31 acute and specialty trusts in London have no problems, so that reviewers felt no pressure to comment? Or is it the case that reviewers did not pick up A&E as an issue, whether positive or negative? CHI's national survey of patient's views of A&E services also found that some of London's A&E units received the lowest ratings for cleanliness and were deemed to be the dirtiest in the country (National Patients Survey 2003b). The six reports in Cheshire and Merseyside mentioning the state of A&E services were also all negative, but listed problems mainly of staffing, poor services and long waiting times, with only one report of a dirty environment.

For reviewers in London, the production and use of information looms less large outside of clinical governance. The issue was mentioned in only six out of 31 reports, all rated negatively by reviewers. (The Royal Free has poor medical records in some departments, as does Bart's and The London, while University College has poorly integrated information systems, and Chelsea and Westminster is reported as having such poor information that it is holding back the organisation's capacity for improvement.) In contrast, the issue was raised in 11 of 14 reports – eight positively and three negatively – in Cheshire and Merseyside. The Wirral Trust's integrated and electronic information system, both within and without the trust, is held up for high praise, and Stockport Trust is mentioned as particularly good at information production and dissemination.

Trust profile issues in CHI reports

Although they are not officially part of reviewers' impressions, the backgrounds and profiles of trusts can provide insights into the environments in which they operate. So information supplied by reviews on the financial positions of trusts and the socio-economic culture in which they operate may help to create a picture of the settings in which London provides health services. Our imaginary lay reader would want information about the financial situations of trusts, as well as the composition of the populations that reflect pressures on services, as without this background information, it is difficult to interpret how well trusts are doing.

The first observation is that in London, as in Cheshire and Merseyside, profile information is not raised uniformly in all reports, and these variations are not entirely due to the evolution of the report format (from the original full versions to the new, slim publications). So although these profiles are provided as background for all reviews, they tend to mirror the inconsistencies found in other parts of reports.

Table 10: Background and profile information most frequently mentioned in London

Issue	Total number of reports where mentioned	Mentioned positively	Mentioned negatively
Finances	14	11 (resources good)	2 (trust in debt or other financial problems)
Mortality rates in area	8	6 (low)	2 (high)
Population profile in area	6 (high proportion of ethnic minorities)	–	–
Deprivation in area	6	–	6 (high levels of deprivation)
Population age in area	6 (5 young, typical London, 1 ageing)	–	–

Source: London acute and specialty trust reports

Table 11: Background and profile information most frequently mentioned in Cheshire and Merseyside

Issue	Total number of reports where mentioned	Mentioned positively	Mentioned negatively
Finances	10	9	1
Population in area	11 (predominately white, or low numbers of black and minority ethnic groups)		
Deprivation in area	8		8 (high levels)
Mortality rates in area	7	2 (low levels)	5 (high levels) (North group 10% higher than English average)
Morbidity rates in area	5	2 (lower than English average)	3 (higher than national average)

Source: Cheshire and Merseyside acute and specialty trust reports

Tables 10 and 11 show the results of our analysis. The finances of trusts were mentioned in 14 of 31 reports in London. Eleven of these said positively that finances were sound and trusts were well funded, and only two reports said that the trusts were in bad debt or had other financial problems. Trust finances were also mentioned in 10 of 14 reports in Cheshire and Merseyside, eight positive (Mid-Cheshire being the lowest cost trust in the North West), and only one said that the trust's financial situation was difficult.

This reported impression of London trusts is backed up by interview evidence from the London District Audit. These interviews underlined the vast range of standard and non-standard resources available to London trusts. In particular, many of the medical schools and specialty hospitals are prosperous organisations. Some trusts (such as Great Ormond Street, for example) have access not only to the standard government funding weighted for population numbers and status, but also to teaching and research funds and charitable resources.

Other teaching hospitals, such as Guy's and St Thomas', Bart's and King's, have millions of pounds of assets in trust funds, which were originally set up to provide benefits for staff through subsidising research, important work and conferences, but which indirectly benefit patients through service provision. (University College Hospitals Trust has bought a heart hospital with assets that have provided patients with 30–50 extra beds in a state-of-the-art facility.)

In London, high levels of deprivation are mentioned as issues in only six reports (including Homerton, Newham and King's College), although there are no specific details of what kind of deprivations. Five trusts in London are in areas mentioned by CHI reviewers as specifically having low population-morbidity rates, and five also operate in areas of low population-mortality rates. There is a wide variation in mortality and morbidity rates across the London trusts, however:

- Chelsea and Westminster has relatively low levels of morbidity and mortality alike.
- North West London Hospitals Trust is in an area with high levels of deprivation but a healthier-than-average surrounding population.
- The West Middlesex Trust is in an area with a healthier-than-average population but higher-than-average death rates.
- Homerton Trust operates partly in one of the most deprived areas in England, and has high mortality rates boosted artificially by high rates of deaths of newborn babies and still-born babies.

Black and minority ethnic groups and asylum seekers are mentioned as issues in four trusts, but it is not clear from the reviews what this signifies for their health status or what these groups need from the trusts' health services. The Chelsea and Westminster review, for example, cites large numbers of black and minority ethnic populations served by the trust, but also says that the area in which the trust operates has lower levels of morbidity and mortality than the average. St Mary's Trust is also said to serve a younger, fitter population than the national average, while having large populations of asylum seekers and refugees. None of the London reports identify the combination of high rates of deprivation, sickness and death in any one trust area.

In Cheshire and Merseyside, eight of 14 reports mention high levels of deprivation. Mortality rates are mentioned in seven trust reports – five with high rates (the North Group of trusts has a mortality rate 10 per cent higher than the English average), and two with low rates. Morbidity rates are mentioned in five trust reports – three as higher than average, including the Wirral which is a three-star trust, and two with lower-than-English-average death rates.

As with London reports, the picture that emerges is unclear. Clatterbridge Centre for Oncology, for example, which serves the North West with specialist services, is reported to be in an area of significant poverty, but with a smaller-than-average ethnic minority population. Aintree Trust – a regional centre for specialist surgery plus a university

teaching hospital – is also said to be serving a less healthy and less wealthy population than the national average, with significantly lower black and minority ethnic populations. The Mid-Cheshire Trust report describes an ageing population being served in the area, but with lower rates of illness than the national average, and predominately white.

Whether described as smaller-than-average black and minority ethnic populations or predominately white populations, these factors are mentioned in the area profiles of 11 of the 14 Cheshire and Merseyside reports but, as with London trusts, without explanations of the implications for health services.

CHI's inconsistent references to deprivations, white populations and black and minority ethnic populations, alongside equally uneven information on age, mortality and morbidity in different areas means that the lay reader cannot relate specific deprivations to specific service demands. As all these figures are available, and could easily be provided in a standard form in all reports, these inconsistencies are unfortunate.

Appendix conclusions

What can we tell from all this about London trusts' health service performance? Is London's health service performance better or worse than that of trusts outside London? The CHI reports may not pass all tests for data reliability, but the issues they raise do provide clues as to what is going on in London. The sketches we have drawn from them give some idea of London's underlying health culture, and its customs and practices in acute health services delivery. The other question is: can we also know whether and what makes London different from the rest of the country? In this section, we examine the evidence from the issues raised in the CHI reports, backed up with occasional references to district audit and Audit Commission work in the London area.

The first, and most obvious, point to emerge is that the environment and cleanliness are big issues in trusts. Judging by the numbers of times these are brought up, CHI reviewers pay a great deal of attention to the environment in which services are delivered. Standards of cleanliness are mentioned in 20 of 31 London reports, and in all 14 of the Cheshire and Merseyside reports. The second point is that London has a high proportion of dirty health service delivery environments and that Cheshire and Merseyside have more cleaner and well-cared-for trusts.

The image of London's NHS finances emerging from CHI reports is mixed but largely positive, particularly in inner London, with teaching hospitals having few-to-no financial problems and others – particularly borough-based general hospitals – having more of a struggle. But even among less prestigious trusts, only two were mentioned as being in debt – one to a minor degree and one major. Resource levels alone, however, do not appear to account for performance differences in London. Given that most trusts appear to have adequate resources, some do less well in maintaining service delivery standards.

In view of the variations between trusts within London in access to services, quality of services, and recruitment and retention of staff, it appears that some organisations are more successful than others at embracing the challenges. The problem of comparing and assessing causal factors in health service performance is notoriously tricky. One frequently quoted example of the resources and services dilemma is that of Newham and Homerton – both London trusts, and both operating with similar supply-and-demand

factors, but with entirely different performances. Homerton is a high-performing trust, while Newham has huge performance problems.

Next to the environment and cleanliness, the most frequently mentioned issue in London is staffing. Sixteen of 31 reports mention this issue – 14 as problematic, and two positively. Cheshire and Merseyside reports also mention staff frequently, and mainly negatively, but give more clues as to underlying problems. From our analysis, it appears that staffing problems in Cheshire and Merseyside are more widespread than they are in London, and are more complex than simply recruitment and retention.

It is clear also that in staffing issues, nurses feature most highly. Reviewers raised issues of standards of nursing care, inappropriate skill mixes, and shortages of nursing staff centred on evenings and weekends, with a subsequent heavy reliance on bank and agency staff. The bank and agency reliance issue was also unpicked by reviewers, who noted that there are concerns in some trusts in London about bank and agency staff working permanently on night shifts and some bank staff working excessive hours while also being permanently employed by the trust. Oddly enough, the reviews do not make anything of the fact that London – with its large number of teaching hospitals training both doctors and nurses – has traditionally had a high degree of mobility among staff, which may well be desirable.

The confusion between the issues of patient involvement and patient satisfaction emerged from many reports. In the London reports, patient satisfaction was not mentioned at all, while it was recorded in 11 of 14 reports in Cheshire and Merseyside – all positively. Patient and public involvement (also a clinical governance component) was mentioned in only two reports in London – both negatively – while it was mentioned in 13 reports in Cheshire and Merseyside, negatively in eight (but with patients happy with services) and positively in five.

It is difficult to know what to make of this information – particularly the silence factor in London – but the comments in the Cheshire and Merseyside reports make it clear that even where patient involvement was low, they were happy and satisfied with services. This might tell us that patient involvement is a poor indicator of patient satisfaction and, as such, is of more interest to the NHS inspectors than to the trust managers.

Overall, our imaginary lay reader is likely to be left unenlightened and confused as to what to make of London health services on the basis of CHI reports. If we assume that high levels of deprivation, high mortality and morbidity rates are all part of the external pressures on trusts and their performance, the picture that emerges of London is far from clear, and it does not seem possible to label London as a whole as having health services burdened by needy populations.

If we move on to the needs of black and minority ethnic populations (sometimes thought of as being synonymous with deprivation and therefore high levels of health needs), the picture of London is still unclear. As the analysis of issues raised in reports shows, population deprivations are being conflated with information about mortality and morbidity in a most uninformative way.

So the picture of London's health needs, and how trusts are meeting them, is ill-defined and unclear, although the same can be said of the Cheshire and Merseyside trust profiles. What is more, London is extremely heterogeneous. The variations – of poverty and demand for services, as well as supply – within London itself are at least as great as the differences between London and the rest of the country.

The picture we have drawn of London health service provision in the acute and specialty trusts may be impressionistic, but it is not incompatible with CHI's own conclusions from its analysis of clinical governance scores, and tends to back up the brief but controversial press release of November 2002 (Commission for Health Improvement 2002b) that London trusts may have some problems with their health services delivery not experienced in the rest of the country.

The picture of Cheshire and Merseyside trusts emerging from reports suggests that London is not unique in its problems, but that they are more pronounced in the capital. For example, London trusts have considerably more difficulties in ensuring a clean environment for patients, and there are staffing problems which cannot be measured purely in the currency of vacancies and turnover. However, the decisive factor in London appears to be not a shortage of resources, but the way in which they are used.

Bibliography

- Better Regulation Taskforce (1999). *Principles of Good Regulation*. London: Cabinet Office.
- Commission for Health Improvement (2003a). *Corporate Plan 2003/2004*. London: CHI.
- Commission for Health Improvement (2003b). *Delivering Improvement: Annual report 2002/2003*. London: CHI.
- Commission for Health Improvement (2003c). *Getting Better? A report on the NHS*. London: CHI.
- Commission for Health Improvement (2003d). *NHS Performance Ratings, 2002/2003*. London: CHI.
- Commission for Health Improvement (2002a). *Growing a new organisation: CHI's combined annual report and accounts 2001–2002*. London: Stationery Office.
- Commission for Health Improvement (2002b). *South North Divide in the NHS*, Press Release, 27 November 2002. London: CHI.
- Commission for Health Improvement (2002c). *Trust Questionnaire*. London: CHI.
- Commission for Health Improvement (2001a). *A Guide to Clinical Governance Reviews in NHS Acute Trusts*. London: CHI.
- Commission for Health Improvement (2001b). *Corporate Plan, 2001–2004*. London: CHI.
- Commission for Health Improvement (2000a). *Clinical Governance Reviews: Principles and methodology*. London: CHI.
- Commission for Health Improvement (2000b). *The Assessment Framework for Clinical Governance*. London: CHI.
- Commission for Healthcare Audit and Inspection (2003). *CHAI: A new organisation*. London: CHAI.
- Day P, Klein R (2002). 'Who nose best?'. *Health Service Journal*, 4 April, pp 26–9.
- Day P, Klein R (2001). *Auditing the Auditors*. London: Stationery Office/Nuffield Trust.
- Day P, Klein R (1987a). 'The business of welfare'. *New Society*, 19 June, pp 11–13.
- Day P, Klein R (1987b). 'The regulation of nursing homes: a comparative perspective'. *Milbank Quarterly* vol 65, no 3, pp 304–47.
- Day P, Klein R, Redmayne S (1996). *Why Regulate?* Bristol: Policy Press.
- Easton M (2002). *The Review and Inspection of NHS Trusts: What value does a CHI clinical governance review add?*, unpublished MSc thesis. London: The Management School, Imperial College, University of London.
- Halligan A, Donaldson LJ (2001). 'Implementing clinical governance: turning vision into reality'. *British Medical Journal* vol 322, pp 1413–17.
- Health Act (1999). London: The Stationery Office.
- Hood C, Scott C, James O, Jones G, Travers A. *Regulation Inside Government*. Oxford: Oxford University Press.
- Kennedy I (2001). *Learning from Bristol: The report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984–1995*. London: The Stationery Office.
- Klein R, Hall P (1974). *Caring for Quality in the Caring Services*. London: Centre for Studies in Social Policy.
- National Audit Office (2003). *Achieving Improvements through Clinical Governance: A progress report on implementation by NHS trusts*. London: Stationery Office.

- National Care Standards Commission (2003). *The Portland Hospital for Women and Children*. London: NCSC.
- National Patient Survey (2003a). *Emergency Department Survey*. London: Commission for Health Improvement.
- National Patient Survey (2003b). *Outpatients Department Survey*. London: Commission for Health Improvement.
- NHS Confederation (2003). *Re-viewing the Reviewers: The NHS Confederation's second survey of NHS trust experience of CHI clinical governance reviews*. London: NHS Confederation.
- NHS Confederation (2002). *Reviewing the Reviewers: NHS experience of CHI clinical governance reviews*. London: NHS Confederation.
- NHS Improvement Scotland (2003). *Safe and Effective Patient Care: Generic clinical governance standards*. Edinburgh: NHS Improvement Scotland.
- Office of Public Services Reform (2003). *Inspecting for Improvement: Developing a customer focused approach. Results of the MORI poll 'Attitudes to public services inspections', carried out in December 2002*. London: OPSR.
- Power M (1997). *The Audit Society*. Oxford: Oxford University Press. (See also Hood et al (1999).
- Prime Minister (1999). *Speech at launch of Commission for Health Improvement*. Press notice, 10 Downing Street, 28 October.
- Public Sector Team/Regulatory Impact Unit (2003). *Making a Difference: Reducing burdens in healthcare inspection and monitoring*. London: Department of Health/Cabinet Office.
- Scally G, Donaldson LJ (1998). 'Clinical governance and the drive for quality improvement in the new NHS in England'. *British Medical Journal* vol 317, pp 61–5.
- Secretary of State for Health (1998). *A First Class Service: Quality in the new NHS*. London: Department of Health.
- Secretary of State for Health (1997). *The New NHS: Modern – dependable*. London: Stationery Office.
- Walshe K (2003). *Regulating Healthcare*. Maidenhead: Open University Press.
- Walters R (2002). *Clinical Governance Reports to September 2002. A Report to the Commission for Health Improvement*. London: Commission for Health Improvement.