Health

National report

November 2006

audit commission

Payment by results assurance framework

Pilot results and recommendations

Final report to the Department of Health

The Audit Commission is an independent body responsible for ensuring that public money is spent economically, efficiently and effectively, to achieve high-quality local services for the public. Our remit covers around 11,000 bodies in England, which between them spend more than £180 billion of public money each year. Our work covers local government, health, housing, community safety and fire and rescue services.

As an independent watchdog, we provide important information on the quality of public services. As a driving force for improvement in those services, we provide practical recommendations and spread best practice. As an independent auditor, we ensure that public services are good value for money and that public money is properly spent.

For further information about the Audit Commission, visit our website at www.audit-commission.gov.uk

Design and production by the Audit Commission Publishing Team

Executive summary 2

1 Introduction 14

2 Approach 17

3 Findings 22

4 Other issues 47

5 Recommendations 51

6 Implementation 55

Appendix 1: Benchmarking methodology 57

Appendix 2: Clinical coding audit process 64

Appendix 3: Coding audit results by trust 65

Appendix 4: Breakdown of coder versus non-coder error 91

Appendix 5: Data definitions 92

First published in November 2006 by the Audit Commission for local authorities and the National Health Service in England, 1st Floor, Millbank Tower, Millbank, London SW1P 4HQ

Executive summary

Introduction

1 In January 2006, the Department of Health asked the Audit Commission to develop and test aspects of an assurance framework for payment by results (PbR), based on the proposal that was submitted to ministers in December 2005. The main objective of the proposed assurance framework was to support the accuracy of payments and the underpinning data under the new financial regime.

2 The proposed PbR assurance framework requires primary care trusts (PCTs) and trusts to take action locally to assure themselves of the validity of data and therefore the accuracy of the payments being made. This should be supplemented by an external programme of data quality audits, focused on clinical coding, using national benchmarking of providers to identify anomalies and target the audits to address specific issues.

3 Between April and August, the Audit Commission piloted the external clinical coding audit and benchmarking components of the PbR assurance framework in two health economies – South Yorkshire and Avon, Gloucestershire and Wiltshire (AGW). The aim of the pilots was to assess the level and nature of clinical coding error and the value of the proposed approach, and to inform a decision about the nature and implementation of a PbR assurance framework for the NHS. In addition, we explored with the pilot sites the issues which NHS bodies were facing locally in monitoring and ensuring accuracy of payments and the areas that they felt should be addressed by the PbR assurance framework.

4 The pilots cost £153,000, which is in line with the estimates provided in the December proposal (£75,000 per pilot).

Key findings

5 The pilots show a relatively high level of clinical coding error, which is leading to inaccurate payments under PbR. The HRG error rate, averaged across the 12 participating acute trusts, is 11.9 per cent[I]. There is also considerable variation across trusts, ranging from an overall Healthcare Resource Group (HRG) error rate of 3.5 per cent to 28 per cent, and across specialties. Diagnosis coding has a higher error rate than procedure coding and the error is more likely to impact on payments. There is greater and more material error among non-elective episodes.

6 The coding error has resulted in both overpayment and underpayment to trusts for individual episodes – underpayment for seven trusts and overpayment for five trusts. The absolute impact of the error on individual trust payments ranges from £7,091 to £67,698 for the sampled 200 episodes, representing between 5 per cent and 14 per cent of the total sample value. Although it is difficult to extrapolate the impact on accuracy of payments to a trust overall with any confidence, we can say that clinical coding error is clearly impacting negatively on payment accuracy and that mechanisms should be in place to address this.

7 Both coder and non-coder errors are contributing to the high error rate. Quality of clinical documentation and clinician involvement are some of the non-coder issues that are contributing to the error rate; while low numbers of accredited coders, insufficient training, and insufficient guidance on coding practice locally (for example, absence of a complete and up-to-date policy and procedure document) are the main causes of coder error. Error rates are noticeably higher where there is a low proportion of accredited coders or a poor coding process (for example, coders are not coding from case notes).

8 There is evidence of overcoding (recording more diagnosis/procedure codes than is necessary) in some trusts, but in some cases this has been a historic practice, and without access to prior audit reports it is difficult to attribute to the financial incentives under PbR. There was also evidence of trusts actively working to optimise their coding to maximise income, but within existing coding rules (that is, the practice is not fraudulent).

I This rate is the percentage of episodes where one or more errors in the coding of diagnoses and procedures resulted in an incorrect HRG being assigned.

There is some concerning practice emerging, for example, where the clinical coding team reports to the director of finance, and is involved in discussions about the financial impact of coding changes on a regular basis. This increases the likelihood of manipulation and it will be important to have a protocol in place to refer cases of suspected fraud to the NHS Counter Fraud and Security Management Service (CFSMS).

9 Coding auditors have also raised concerns about the appropriateness of admissions in some trusts, where it appeared from the case notes that patients should have been treated as outpatients and there was no obvious reason for admission. This is a big issue in both pilots. However, current data definitions are ambiguous and their application is subjective, leading to different local practices in the treatment of day cases/outpatients – until this is resolved it will be difficult to address this area.

10 While there is no agreed standard or target for clinical coding error in the NHS, the reported level of HRG error seems intuitively too high, particularly in a system where clinical coding is the primary determinant of payment[l]. The level and nature of clinical coding error is not so high as to destabilise the PbR regime, but it does undermine it, raising concerns about the accuracy and fairness of funding flows. In addition, it has major data quality, epidemiological and clinical implications.

11 The pilots demonstrate that there is a need for a centrally coordinated external clinical coding audit programme, which ensures objectivity and independence of results. Even where there are strong local monitoring arrangements in place at commissioner and provider level, there is a need for an external and independent audit to supplement and provide additional impetus to improve data quality. Both pilot sites have commented on the value of the audits in highlighting data quality and payment issues from an independent and objective perspective, and in opening dialogue about this between commissioners and providers and across the health economy.

---

l The percentage HRG change across Victorian hospitals in Australia for the 2000-01 clinical coding audit (the most recent report available) was 9.8 per cent (Source: 2000-01 Audit Report), which is a comparable level of error with the NHS but which the Victorian government has been actively working to reduce through external audit and other mechanisms.

For US Medicare, in November 2005, 5.2 per cent of the dollars paid nationally did not comply with one or more of Medicare coverage, coding, billing, and payment rules. Projected overpayments were $11.2 billion and the underpayments were $0.9 billion (Source: CMS website).

12 In addition to the local benefits, a comprehensive audit programme has an obvious national benefit in improving the quality of data in the NHS and refining the policy framework (for example, the audits have highlighted inconsistencies in coding rules and/or the HRG grouping algorithm). There is also a benefit for patients. It is the first time that this kind of information has been made available, and various stakeholders have commented on its value and the opportunity to share lessons learnt and good practice.

Lessons learnt

13 A number of lessons have been learnt during the pilots, including specific issues relating to the process and methodology, which are documented in the main report.

14 Two key issues have resurfaced, both of which were also raised during the design phase. The first is the scope of the PbR assurance framework, what it does and should cover, and the extent of external involvement. The second and related issue is the purpose of the clinical coding audits, and how they are used to drive improvements in data quality and payment accuracy.

Scope of the PbR assurance framework

15 Although the purpose of the PbR assurance framework is to support accuracy of payments and data quality locally, the term has raised the expectations of the pilot sites that the framework will provide assurance around the PbR policy itself, particularly the tariff. A decision is required on whether the scope of the PbR assurance framework is widened to include assurance of the tariff itself, incorporating the actions of the Department of Health as a result of the Lawlor review, or whether it focuses on monitoring payments and data quality. Regardless, feedback from the pilots reinforces the Audit Commission's earlier recommendation in the recent report to the Secretary of State for Health that 'some external assurance on the figuring and rationale underlying the tariff' is needed[l]. Depending on the decision, the Department of Health may wish to reconsider use of the term 'assurance framework'. The work that the Audit Commission has undertaken is more in line with a monitoring programme.

l *Audit Commission Review of the NHS Financial Management and Accounting Regime: A Report to the Secretary of State for Health*, Audit Commission, 2006.

16 Second, the pilots have indicated that clinical coding is too narrow a focus, and external audits should cover broader aspects of whether activity has been correctly recorded (including adherence to data standards and other data quality errors that affect payments). The focus of the pilots on clinical coding was based on the assumption that local PCT arrangements should already cover, to some extent, the areas which are important for comprehensive monitoring of payments, including:

- inappropriate admissions/transfers;
- medically unnecessary diagnoses and procedures;
- re-admissions due to incomplete care/premature discharge; and
- other errors that affect payment (for example, wrong commissioner code, duplicates).

17 The external clinical coding audit programme was intended to provide additional support to identify and address coding error and upcoding, rather than to comprehensively support payment accuracy. However, findings from the pilots indicate that external audits should also cover wider aspects of whether activity has been correctly recorded. Although coding auditors are able to identify some of these issues, the audit methodology would need to be refined to properly address these areas.

18 Before any audit of these issues can be effective, ambiguity around data definitions (for example, the definition of an admission) and inconsistency in how these are applied on the ground needs to be addressed as a priority. The current ambiguity inherent in the definitions is creating difficulties and it is not sufficient to leave these to local interpretation.

19 Finally, the audit of reference costs and a review of behaviours against the PbR Code of Conduct were raised as additional areas that should be included in the PbR assurance framework. Both of these areas were included in the original proposal to the Department of Health. There is merit in introducing an audit of reference costs. This was recommended in the Audit Commission's recent report to the Secretary of State for Health on the review of the NHS financial management and accounting regime, and would support the Department of Health's actions stemming from the Lawlor review to improve the calculation of the national tariff. However, behaviours are more difficult to address through external checks and we have not made any recommendations for external action as part of the PbR assurance framework.

Role of the external clinical coding audits

20 The audits that were undertaken for the pilots aimed to test the level and nature of clinical coding at acute trusts and assess the impact this was having on payments. In practice, they enabled us to identify data quality issues which are impacting on payment accuracy and which warrant further investigation, to identify potential manipulation or abuse of the system and to identify perverse incentives in the policy framework which need refinement (although there are some improvements that can be made to the methodology).

21 Initially, some of the organisations participating in pilots raised questions about the purpose of the external clinical coding audit programme and whether it should aim to provide assurance on clinical coding and payments, rather than simply identifying issues. However, there were mixed views about the value and cost-effectiveness of this. The size and cost of the audits required to provide assurance would be prohibitive and this approach is not used internationally. By the end of the pilots, and in light of the findings, the overwhelming view was that the primary aim of the external audits should be to improve data quality – if we achieve this, other objectives such as payment accuracy will follow.

22 In response, our recommendation is that the clinical coding audit approach should be data quality-led, but should also support policy refinement and payment accuracy by :

- identifying the key coding issues that are impacting on payments locally and which warrant further investigation; and
- identifying upcoding and potential manipulation.

23 The audits should contribute to the overall objective of the PbR assurance framework to support payment accuracy, in conjunction with other components of the framework. It should be the responsibility of local NHS bodies to investigate current issues in order to reconcile payments at the end of each quarter.

24 Where possible, it is important that the audits support collaborative working between commissioners and providers and across the health economy, rather than polarising organisations and leading to disputes. Financial implications of error will need to be managed carefully, and consideration will need to be given to how reporting is handled.

Actions from the clinical coding audits

25 As the coding audits were retrospective (generally covering the last quarter of 2005), the results are not able to inform actual payment reconciliation under the current policy (in other words flex and freeze rules which mean that payment adjustment is only possible for the current quarter) unless the audits are much more responsive and targeted. However, they will inform future adjustments if a systematic issue is uncovered and the PCT undertakes future investigation.

26 If payment reconciliation is unlikely to be a direct action stemming from the audits (although they should influence future reconciliation), there will need to be clear consequences associated with the coding audits in order to achieve their objectives (see recommendations).

27 Importantly, although penalty clauses are already included in some foundation trust contracts, the model contract has no existing provision for penalties for incorrect data that could be applied in response to these audits. A standard penalty clause should be developed and incorporated into the revised model contract which the Department of Health is developing, and applied to all NHS providers. The contract clause should be phrased in such a way that poor results from an external audit of coding or data quality will attract a penalty, most likely paid to the host commissioner. To avoid disputes and ensure that trusts are not penalised unfairly, the nature and level of the penalty will need to be considered carefully – for example, should it be applied for results above a certain error rate, or for failure to improve? This should be the subject of consultation with the NHS.

Strengthening local arrangements

28 PCTs' arrangements for monitoring data quality and payments have progressed, and are expected to further develop with reorganisation. However, they are still variable and strengthening them is a continuing priority.

29 The Department of Health's fitness for purpose assessments (the monitoring diagnostic tool in particular) and the Healthcare Commission's Annual Health Check will go some way towards this, but further support which gives greater direction and specifically addresses PbR is required. Guidance that the Audit Commission intends to issue to auditors during 2006/07, which will include steps that auditors should take to check arrangements for ensuring accurate payments before signing off the accounts, will be one

mechanism for achieving this. In addition, PCTs would benefit from specific support and guidance to develop local monitoring arrangements, for example, a tool or checklist which enables them to check their arrangements against best practice.

Recommendations

30 Based on the findings from the pilots, we have the following recommendations for the Department of Health:

a. An external clinical coding audit programme should be implemented as a core part of the PBR assurance framework, but with some changes to the approach used in the pilots, including:

- a larger sample size, aiming for statistical significance;
- spells-based rather than episode-based; and
- extension of scope to include outpatient and critical care data.

This should replace the annual information governance toolkit audit currently commissioned by Trusts, but should be underpinned by a requirement for Trusts to conduct regular internal audits.

b. The benchmarking methodology should be further developed and used to identify areas of concern and target the audits. The targeting of audits should vary from quite broad (trust-wide), where there are wider concerns about data quality, to very precise (focusing on individual HRGs), where specific issues have been identified as impacting on data quality and payments. Themed audits, where the same area is covered across all NHS providers, should be an integral part of the programme.

c. A proper analytical tool to support a more sophisticated approach to national benchmarking should be developed in partnership with the Information Centre. Ideally, this tool should support routine national analysis of hospital activity data for a range of purposes, not just targeting data quality audits, with the data returned systematically to NHS bodies, and could be incorporated into the Secondary Uses Service (SUS).

d. The external audit programme should be extended to encompass the broader recording of activity. A distinct audit methodology should be developed to address areas such as the correct use of data standards and data definitions, drawing on approaches that are already being used locally. This audit would be less intensive and would not require the use of clinical coding auditors, although they could be trained to undertake this work as an extension of the coding audit.

e. The following actions should result from the coding audits:
   - penalties for poor data quality;
   - sharing of good practice and lessons learnt;
   - influencing auditors' local evaluation (ALE) assessments and the Annual Health Check;
   - publication of a league table showing error rates and possibly a national error rate statistic;[l]
   - further investigation by the PCT, which may influence future payments;
   - further investigation at the expense of the trust, where data quality is poor or specific concerns are raised; and
   - referral of cases of suspected fraud to CFSMS for further investigation in line with an agreed protocol.

f. There should be a standardised penalty clause in all contracts with NHS providers, including foundation trusts and the independent sector, which can be invoked by the PCT in response to a poor result from external audit. This may be phased in gradually, and should not apply to the first audit. The Department of Health should consult with the NHS on this issue.

g. The Department of Health should consider whether assurance of PbR policy, including the tariff, should be included in the assurance framework. Regardless of whether it is included in the framework, there should be some external assurance on the calculation of the final tariff structure. Consideration should be given to whether assurance framework is the correct term for what we are trying to achieve, and whether an actual assurance framework should be published as a result of this work.

h. Inconsistencies and ambiguities in current data definitions and standards as identified through the pilots should be addressed as an immediate priority, and a clear process put in place so that similar issues which are raised by NHS bodies in future can be properly addressed. While it may take some time for these to be reflected in the information infrastructure, the Department should issue interim guidance before the end of the year on how these cases should be treated.

l This would need to be handled carefully, as NHS trusts and the independent sector may interpret this as the benchmark.

i. There should be an independent audit of NHS trusts' reference costs, which would assess the application of the national costing methodology at local level and the quality of the reference cost data, and the information should be published.

j. A dedicated tool or checklist should be developed for PCTs, enabling them to check their arrangements for monitoring activity and payment under PbR against best practice.

k. The Audit Commission continues to orchestrate the external components of the PbR assurance framework and to support strengthening of local arrangements through routine audit work and the development of the PCT tool.

l. The external programme should be reviewed on an annual basis, as data quality improves and local arrangements strengthen.

31 In addition, we propose that the findings from the pilots are published in a report aimed at the NHS, with the emphasis on good practice and lessons learnt.

Implementation of the external programme

32 Should these recommendations be accepted by the Department of Health, we propose to adopt a phased implementation, commencing in October.

Phase 1 (to March 2007)

- Developing the benchmarking tool in conjunction with the Information Centre.
- Preparation of the external clinical coding audit programme to roll out across all NHS providers for acute inpatient episodes, including refinements to the methodology and further audits to be undertaken towards the end of 2006/07.
- Developing the audit methodology for broader recording of activity (data standards and data definitions).
- Developing the audit methodology for outpatient and critical care data.

Phase 2 (2007/08)

- NHS-wide implementation of the external clinical coding audit programme (covering approximately 74,000 episodes).

- Testing and implementation of audits on broader recording of activity (integration with the clinical coding audits or a separate audit programme).
- Ongoing refinements to the benchmarking tool.
- Testing and implementation of the audit programme for outpatient and critical care data.
- Developing the audit methodology for mental health data.

33 In addition, the audits will need to be extended to cover independent sector providers of NHS services. This should ideally commence in 2007/08, but further work is required to determine how the approach needs to be adapted to support this, particularly given the different information infrastructure, and a feasible timescale for implementation.

34 The following actions should be a priority for the Department of Health to action in the remainder of 2006/07:

- refinement of data definitions;
- establishing a robust external audit of reference costs; and
- development of a standardised penalty clause for poor data quality.

35 The estimated annual cost of the targeted clinical coding audit programme is £4.45 million, or an average cost of approximately £30,000 per PCT[l]. There would be implementation costs of £2.02 million in the remainder of 2006/07, which includes a further 30 audits. Extending the audit programme to address wider recording of activity would result in an additional annual cost in the order of £2.25 million (based on five days per audit). Inclusion of the independent sector, outpatient and critical care data will need to be costed separately once the approach to implementation has been agreed.

36 There are several options for how the audit programme could be organised initially. At the minimum, there should be a three-year rolling programme, covering every provider of NHS services over this period, and the option to do further follow-up or targeted work at poorly performing trusts. This would increase the coverage and sample size at individual trusts. Alternatively, all providers may be audited in the first year to establish a baseline, covering 500 episodes per trust, with a risk-based approach applied in following years.

l This should not be read as the fee charged to PCTs for the work. The approach to setting fees for this work is yet to be agreed and would be part of the Audit Commission's annual fee consultation for 2007/08.

37 The detailed programme will need to be defined more clearly in the early stages of implementation, and will depend on the results from future benchmarking work, the balance between the data quality and payment accuracy focus and what is acceptable in the NHS.

1

Introduction

Background

38 In 2004, the Audit Commission raised concerns about the absence of a robust strategy for monitoring accuracy of payments in the NHS under PbR. The potential for inaccurate payments resulting from poor quality activity and cost data was seen as a key risk to PbR and to the NHS as a whole. There was also concern about the potential for deliberate manipulation of activity and data to influence payments under PbR which has been a central feature of similar systems in other countries. Overall, there appeared to be an absence of robust mechanisms within the NHS to support improvements in data quality in the new environment, to support accuracy of payments and to identify and address manipulation, if it occurred.

39 These concerns were highlighted in our 2004 report *Introducing Payment by Results: Getting the Balance Right for the NHS and Taxpayers*. They were also raised directly with the Department of Health, along with recommendations for the development of a comprehensive 'policing' function to support systems integrity.

40 This led the Department of Health to establish a programme of work to design a robust PbR assurance framework, led by the Audit Commission under the direction of the Assurance Steering Group. The work was undertaken in 2005, and involved a thorough review of current arrangements for monitoring payments and data quality and stakeholder perceptions on the need for an assurance framework. This culminated in a final proposal to the Department of Health in December 2005, which was then submitted to ministers for consideration.

41 The proposal scope covered: strengthening PCT monitoring arrangements and arrangements for data quality, as well as recommendations for an external programme of data quality audits, focused on clinical coding (including an initial audit to test the level and nature of coding error and its impact on funding flows) and development of a benchmarking tool to analyse variations in provider data and enable targeting of audits. Reference costs and behaviours (in line with the PbR Code of Conduct) were also included.

42 In January 2006, the Department of Health asked the Audit Commission to begin work on the foundations of a PbR assurance framework for the NHS in 2006/07.

Objectives

43 The following five objectives were defined for the PbR assurance framework in the initial proposal:

1) to provide assurance on the appropriateness of payments made to providers under PbR;

2) to support improvements in data quality and costing across the NHS;

3) to inform refinements of the policy and technical framework;

4) to limit opportunities for manipulation or abuse of the system, and interface with the existing fraud framework; and

5) to promote appropriate behaviour under PbR, in line with good practice.

44 A number of actions are necessary to achieve these objectives. The approach proposed by the Audit Commission rests on PCTs and trusts taking action locally to assure themselves of the validity of data and therefore the accuracy of the payments being made, supplemented by external checks on key aspects of data quality.

45 Due to tight timescales for implementation in 2006/07, it was agreed that the focus would be on putting the basic building blocks in place, while providing the foundation for a longer-term approach.

46 It was agreed that the Audit Commission's work would focus initially on a pilot of external clinical coding audits in two health economies, and the development of an analytical tool to benchmark providers, identify potential data quality issues and target the audits.

47 The objectives of the pilots were to test the benchmarking and audit methodology in the pilots, collate evidence on the level and nature of error in data quality and payments, and inform a decision about full implementation of a PbR assurance framework during the remainder of 2006/07 – in particular, whether a centrally managed external clinical coding audit programme would be of value, and any other areas that need to be covered by external checks. The need to test whether the focus of the audits on accuracy of clinical coding was appropriate was highlighted at the outset.

Purpose of the report

48 This report summarises findings from the PbR assurance pilot work undertaken by the Audit Commission between April and August 2006. It makes recommendations on implementation of the approach, and on what the wider PbR assurance framework should cover. It has been prepared for the Department of Health, and is intended to inform a decision about the future of the PbR assurance framework.

49 Section 2 sets out the approach and methodology used for the PbR assurance pilots in more detail. The results of the benchmarking analysis and the external clinical coding audits are covered in Section 3. Other issues that have arisen in the course of the pilots, including developments in local arrangements, are discussed in Section 4. Finally, Section 5 sets out the Audit Commission's recommendations and Section 6 discusses implementation.

2

Approach

50 The approach and methodology for the PbR assurance pilots were discussed and agreed with the External Advisory Group (EAG) in April. The EAG was set up to oversee the Audit Commission's PbR assurance work programme in 2006/07, comprising external stakeholders from the NHS and national organisations, including the Department of Health.

51 As noted earlier, the pilots focused on two key activities:

- national benchmarking, to target audits at areas where there is the greatest risk of poor data quality, and ensure that audits are proportionate and impactful; and
- external data quality (clinical coding) audits.

52 Pilots were undertaken from April to August 2006 in two health economies – South Yorkshire and AGW (based on the old strategic health authority areas). South Yorkshire was selected by virtue of its status as the PbR laboratory for the Department of Health. The second pilot site was chosen to reflect different characteristics.

53 The pilots covered 12 acute trusts (including foundation trusts) and 21 PCTs. The Royal National Hospital for Rheumatic Diseases was excluded from the pilots due to difficulties scheduling the audit.

54 Initial analysis of national hospital data and development of the preliminary benchmarking tool for the pilots was undertaken between April and June, along with the development of the audit approach. A report on the benchmarking results was provided to the relevant commissioner and provider.

55 Workshops were held in AGW and South Yorkshire on 7 and 9 June respectively to discuss the benchmarking and audit methodology and obtain feedback. Where possible, the issues were reflected in the approach to the pilots. However, much of the feedback was on the scope of the PbR assurance framework and should be used to inform implementation.

56 Benchmarking data was used to target the clinical coding audits. Where time permitted, the proposed target areas were discussed with the host PCT and coding experts prior to finalising the audit sample, to validate the approach and ensure any local intelligence on the causes of the variation had been taken into account.

57 The clinical coding audits were undertaken from 24 July to 25 August. Although the audits themselves were performed at the acute trusts, and recommendations for improving data quality were made to the acute trusts, the audits were undertaken from the perspective of the commissioner, with the main report prepared for the PCT.

58 Feedback workshops were held for the two pilot sites on 8 and 12 September, to discuss the overall findings, proposed recommendations and enable feedback on the value of the pilots and how the approach could be improved.

59 The pilots cost approximately £153,000, which is in line with the estimates provided in the December proposal (£75,000 per pilot).

60 The remainder of this section provides more detail on the scope and methodology that were applied.

Scope

61 The pilots focused on the accuracy of recorded diagnosis and procedure codes[l] for acute inpatient episodes. Outpatient data, critical care, mental health and community care were excluded from the audits.

62 They did not consider broader data quality issues (general management arrangements, adherence to data standards), the correctness of other data elements that impact on payments, accuracy of reference costs, or the behaviours of commissioners and providers under PbR (except their behaviour with respect to ensuring and monitoring data quality).

Benchmarking

63 The main purpose of the benchmarking work is to identify potential data quality issues by comparing acute trusts against their peers for a core set of indicators. By identifying a dynamic set of indicators and benchmarking between providers and over time, we can

l The latest version of the procedure codes, OPCS 4.3, is not within the scope of the audit, as this classification was only used from 2006/07.

analyse patterns and identify anomalies that may indicate potential data quality issues and warrant further investigation. This will enable us primarily to target external data quality audits. By providing the results back to commissioners and providers, the data can also support local monitoring.

64 There are limitations to this approach which were recognised at the outset. Until the SUS is fully implemented, there are still discrepancies between national datasets and the data used locally for payment purposes. Secondly, until PbR beds in and the HRG classification, data definitions and coding rules are refined and stable, there will be a lot of variation in the system, making it difficult to identify data quality issues.

65 The pilots were used to develop indicators and the associated analysis and test the approach. Details of the data and methodology used are set out in Appendix 1. The acute trusts participating in the pilots were benchmarked against their peer group for a set of indicators which cover admission and referral patterns, data quality and HRG complexity. Significant variation against their peers or change over time was identified.

66 The intention was to analyse both the NHS-wide Clearing Service (NWCS) and Hospital Episode Statistics (HES) data on hospital activity for the pilots, and to determine the most appropriate data source to use for the analysis in the absence of SUS data. However, there were substantial delays in receiving NWCS data as resources were being used to validate SUS, and only the HES dataset was obtained. The delays impacted on the methodology that could be developed, with the following changes:

- only 2.5 years of hospital activity data was used (rather than the full 3 years);
- age-sex standardisation was not applied;
- indicators were prioritised and the number was reduced;
- cost data was not used; and
- a less sophisticated approach to identifying variation in the data was used (for example, Statistical Process Control was not used to identify changes over time).

67 The likely impact of these changes is described in Section 3.

Clinical coding audits

68 External data quality audits, focused on the accuracy of clinical coding and the impact of this on payments, were at the heart of proposals for a PbR assurance framework, and the main focus of the pilots.

69 Although NHS trusts already commission annual audits of their clinical coding, as part of the information governance toolkit, these are internal audits (albeit usually undertaken by auditors external to the trust), are not systematically undertaken by all trusts and are undertaken for the benefit of the trust alone, with the results not routinely available to commissioners or national bodies with an interest in data quality. The PbR environment, where clinical coding is a key determinant of payment, warrants a fully independent and systematic programme of external clinical coding audits.

70 The purpose of the clinical coding audits undertaken for the pilots was to measure the level and nature of clinical coding error in acute trusts, and the impact of error on payments, and test the value of a centrally managed external audit programme.

71 Audits were undertaken at the 12 acute trusts participating in the pilots. The audits followed the latest version of the national standard clinical coding audit methodology – NHS Connecting for Health Clinical Coding Audit Methodology (published April 2006) – which involves a review of the trust's coding against the case notes. Auditors were asked to fully analyse the case-mix implications of coding errors and include any issues related to the wider recording of activity (for example, where day cases may have been more appropriately recorded as outpatients).

72 Further information on the audit process is set out in Appendix 2.

73 Each audit looked at a random sample of 200 Finished Consultant Episodes (FCEs) over a given quarter, drawn from the top three areas of interest as identified in the benchmarking exercise. Time constraints and availability of accredited clinical coding auditors meant it was not possible to cover 2 per cent of each trust's episodes as originally intended.

74 Separate reports were produced for the trust and the PCT. A draft report was agreed with the trust initially, to confirm factual accuracy and agree the conclusions. This report focused entirely on data quality, with no financial information included. The findings were

then translated into a report for the host PCT, in which the financial impact of coding error was considered and specific recommendations to commissioners were formed.

75 Both reports are being shared with the relevant PCTs and trusts for the purposes of the pilots. The reporting protocol would need to be reviewed prior to full implementation of the approach. Feedback from the pilots indicates that a single report to both providers and commissioners would be appropriate.

76 Fully accredited coding auditors were contracted by the Audit Commission to undertake the audits. Where possible, auditors were assigned to trusts where they had no recent audit experience, to ensure full independence.

77 To address concerns raised previously by Phil Walker of Connecting for Health about external contractors having access to confidential patient information, we held a workshop on auditors' responsibilities in working for the Commission, focusing on confidentiality and data protection issues. The coding auditors were also required to sign a notice confirming their familiarity with the requirements of Section 49 of the 1998 Audit Commission Act and the duty of confidentiality. We issued guidance on adhering to the safeguards set out in Section 60 of the Health and Social Care Act 2001: to ensure only sufficient and not excessive information is obtained and processed, to remove any personal identifiers not required, to take steps to prevent unauthorised processing of the information and to ensure patient information is held securely and not retained for longer than necessary.

3

Findings

Benchmarking

Results

78 The analysis identified a number of anomalies for each acute trust. Two or three specialties were then selected for audit, on the basis of materiality, whether the variation was explainable (for example, due to local characteristics) and the importance of the indicator from the PCT perspective. Where time permitted, the benchmarking results were calibrated with the PCT and the coding auditors.

79 **Table 1** summarises the specialties that were targeted in each audit. One of the benefits of orchestrating the coding audits centrally is the ability to compare practice across different organisations. Therefore, General Medicine was selected as a theme and covered in all acute trusts, with the exception of Sheffield Children's Hospital.

80 The analysis typically identified a subset of activity within a specialty area as being of interest (for example, looking at non-elective patients within HRG chapter L). Therefore, in some cases, the audit was more narrowly targeted, while in others the focus was the specialty. This has given us an indication of whether it is more effective to target audits narrowly – for example, down to a specific HRG chapter, individual HRGs or even diagnosis/procedure codes – or more broadly.

81 As the size of the audit sample was limited to 200 episodes for the pilots, only one-quarter was selected for audit. If the sample size had been larger, we would have selected two periods to allow comparability of coding between periods and identify any step changes. For most acute trusts the period selected for audit was October-December 2005, the most recent period for which national data was available.

82 Examples of the benchmarking reports which were provided to PCTs and trusts in the pilots are available from the Audit Commission.

Table 1
Summary of areas selected for the audits

	General Medicine	Paediatrics	Obstetrics	Cardiology	Nephrology	Geriatric Medicine	Urology	Gynaecology	Paediatric Surgery	Trauma and Orthopaedics	Ear, Nose and Throat	Medical Oncology
Trust 1	■	■					■					
Trust 2	■		■			■						
Trust 3		■							■	■		
Trust 4	■				■			■				
Trust 5	■	■									■	
Trust 6	■		■				■					
Trust 7	■			■	■							
Trust 8	■					■						
Trust 9	■			■				■				
Trust 10	■		■		■							
Trust 11	■			■								■
Trust 12	■			■			■	■				
Total number of case notes audited	1000	200	200	250	150	150	125	150	50	50	50	50

Evaluation

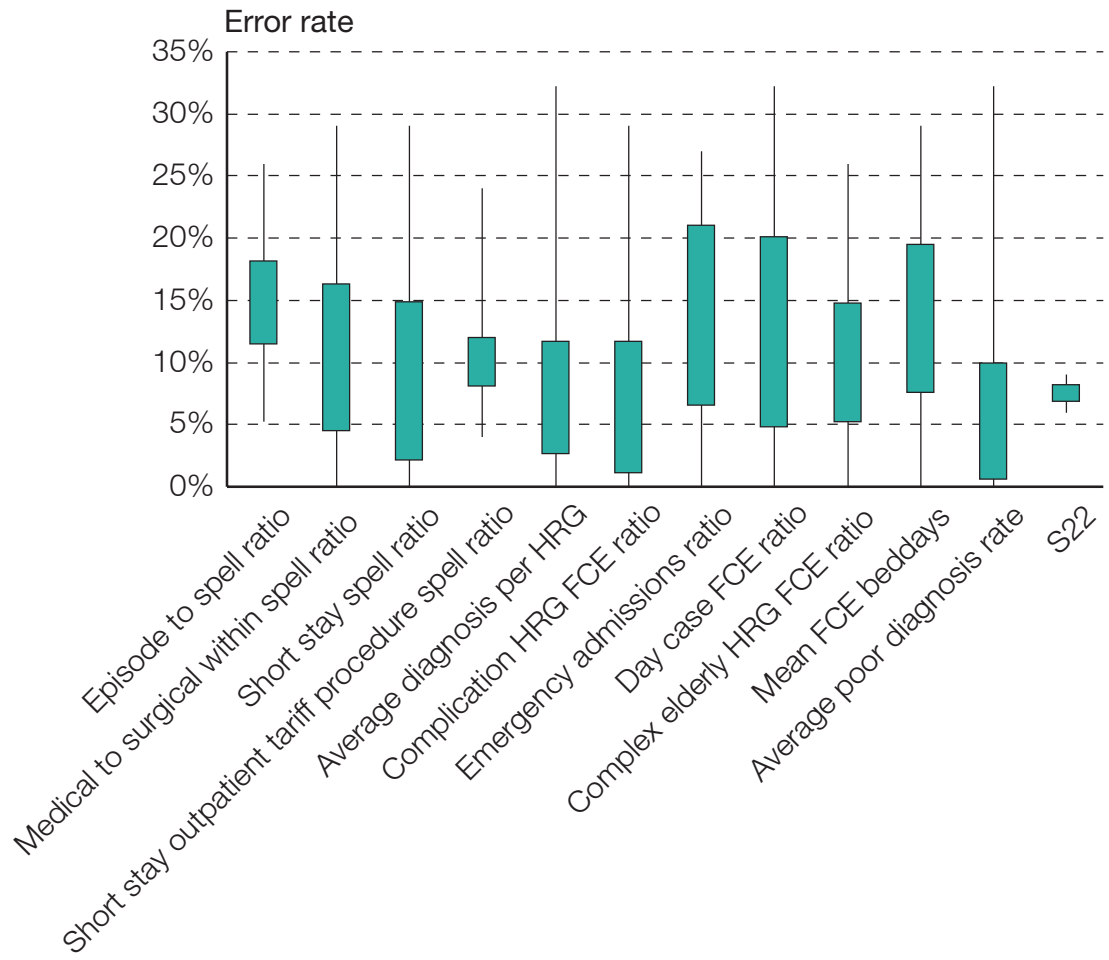
83 It is difficult to fully evaluate the usefulness of the benchmarking as the basis for targeting the audits, as we did not have a control group where the audits were not targeted. However, we have reviewed the results from the coding audits to identify which indicators were the most effective at highlighting data quality problems and to assess whether the errors found were related to the indicators used. **Table 2** illustrates the frequency of indicators used to target the audits (multiple indicators were used per trust).

Table 2
Frequency of indicators used to target the audits

Indicator	N
Short stay spell ratio	23
Average poor diagnosis rate	14
Medical to surgical within spell ratio	12
Mean FCE beddays	11
Day case FCE ratio	11
Complication HRG FCE ratio	11
Emergency admissions ratio	10
Complex elderly HRG FCE ratio	7
Average diagnosis per HRG	7
Short stay outpatient tariff procedure spell ratio	5
Episode to spell ratio	5
Planned procedures not carried out (HRG S22)	2

Figure 1 summarises the range in the resultant HRG error rate for the 12 indicators, when they were used as the basis for targeting the audits.

Figure 1
Analysis of error rate against benchmarking indicators used



84 The indicators that tend to result in a lower HRG error rate (less than 10 per cent) are short stay outpatient tariff procedure spell ratio, average diagnosis per HRG, complication HRG ratio, average poor diagnosis rate and recording of S22. The episode to spell ratio HRG consistently results in a higher error rate (greater than 10 per cent), suggesting that this may be the most effective predictor of error[l]. The other indicators have a wider range in error rate but do on the whole have an error rate greater than 10 per cent. Finally, there was no relationship between the number of indicators identified for a speciality and the error rate.

l This may be due to the fact that where there are multiple episodes per spell, any errors that have been made in the first episode tend to be repeated in other episodes within the spell.

85 There are a number of limitations to this analysis. Notably, some of the indicators, such as outpatient tariff procedure spell ratio, are likely to reflect wider data quality issues (for example, appropriateness of admissions) which are not always picked up in a clinical coding audit. It is not possible to identify how good the indicators were as a predictor of HRG error rates, as there is no control group. It is also difficult to conclude whether the error is linked to indicators. Finally, not all trusts followed the requirements for the sample selection set out in the pre-audit letter, which meant that the audits did not always look at the areas intended.

86 Overall, despite the limitations of the approach and the difficulties in evaluating the precise indicators, the findings from the pilots support the general principle of using benchmarking to target the audits to high-risk areas. The rates of coding error identified in the audits were generally higher than the results from audits previously undertaken by the trust, suggesting that high-risk areas had been identified. And in most cases, the findings from the coding audit confirmed the benchmarking results. However, this was not always due to error or unusual recording practices but rather due to unusual local configuration or population (for example, a high number of elderly patients). This confirms the importance of collating comprehensive local intelligence from both the trust and PCT before finalising the audit sample and the need to build age-sex standardisation into the methodology going forward (as originally intended).

87 In the absence of SUS data, the HES dataset is an acceptable alternative, but the time lag before these datasets are available (three to six months from data collection) does limit the exercise.

88 As noted earlier, the approach that was used for the pilot had some recognised limitations difficulties obtaining national data (NWCS) and the resultant limited timescales. We also received detailed feedback from the pilot sites on the benchmarking methodology which will enable us to refine the approach. The following refinements will need to be made:

- the indicators will need to be extended and refined;
- the approach to identifying significant variation needs to be refined, including the use of a technique such as Statistical Process Control (SPC) to identify anomalies over time;
- age-sex standardisation will need to be applied;

- the focus on specialties with higher volume activity will need to be reviewed, as coding errors in lower volume areas may be equally or more interesting (and more beneficial) to review;
- consultant specialty rather than treatment specialty was used as the first level of analysis – both have inherent local variation, and the most appropriate starting point needs further consideration; and
- in conducting spell-based analysis, the first episode within the spell as the dominant FCE is currently being used as a proxy, pending full spell linkage being carried out on the whole dataset.

89 It is clear from the audit results that age-sex standardisation would have had some bearing on the targeting of audits. However, it would not have changed the targeting substantially. The use of a SPC would have enabled us to target the audits more sharply to a specific period of time, but again, would not have substantially affected the areas selected.

90 Finally, there is an obvious trade-off between a narrowly targeted audit and a wider audit. Where we focused on an entire specialty, we have picked up a broader range of errors and data quality issues than would have been identified through a closely targeted audit. Conversely, a narrowly targeted audit is more likely to identify systematic problems and potential gaming (assuming the analysis behind it is accurate).

91 In conclusion, we recommend that the benchmarking is used as the basis for targeting external data quality audits in the future. The specification should be refined, as originally intended, and a benchmarking tool developed. This should ideally be led by the Information Centre, as this type of analysis of national activity data has a number of other uses. The intention would still be to return the data locally to NHS bodies to use in their own monitoring arrangements.

92 In developing the approach, SUS data should be used as the primary data source. The data will be used locally as the basis for monitoring payments, providing us with a consistency which the existing HES and NWCS datasets do not offer. It also contains additional data fields, for example, robust derived spell identifiers and costed activity, will support more effective analysis. SUS currently holds spell-based data from 2004/05 and will contain three years of data by the end of 2006/07.

93 There is also potential to build the indicators which we are interested in analysing into SUS (potentially for the June 2007 release). These indicators would then be available for other users, notably NHS bodies themselves, to use. This would be the preferred approach. However, a source of funding would need to be identified for this and we would need to move quickly to have the indicators incorporated.

Clinical coding audits

Results

94 The findings from the 12 clinical coding audits which were undertaken in the pilots show a significant level of coding error **(Table 3)**. The high accuracy of secondary procedure coding is in part due to the relatively small number of secondary procedures that were audited.

Table 3
Summary of coding audit results by trust

	% HRG error	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	No. unsafe to audit
South Yorkshire						
Trust 1	10	86	73.5	96.5	93	2
Trust 2	13.5	80.5	78	91	94	2
Trust 3	11.17	68	50	66	n/a	2
Trust 4	11	86.1	86.2	83.8	80.5	13
Trust 5	7	94.4	94.1	91.0	100	1
South Yorkshire average	*10.5*	*83*	*76.4*	*85.7*	*73.5*	*4*
Avon, Gloucestershire and Wiltshire (AGW)						
Trust 6	3.5	86.9	91.2	97.2	81.9	2
Trust 7	23.6	74.9	81.5	94.6	100	9
Trust 8	28	71.5	74.3	76.2	87.5	18
Trust 9	10.5	75.5	64.2	75.9	79.5	1
Trust 10	10.2	84.4	77	93.2	84.3	0
Trust 11	13	73	53	88.5	75.6	4
Trust 12	4.6	94.9	94.1	92.5	86.7	3
AGW average	*13.3*	*80.2*	*76.5*	*88.3*	*85.1*	*5.3*
Average of the two pilots	11.9	81.6	76.5	87	79.3	4.7

Note: in some cases the unsafe to audit episodes were counted in the 200 sample audited, in other cases they were excluded.

Appendix 3 contains a breakdown of the findings by trust.

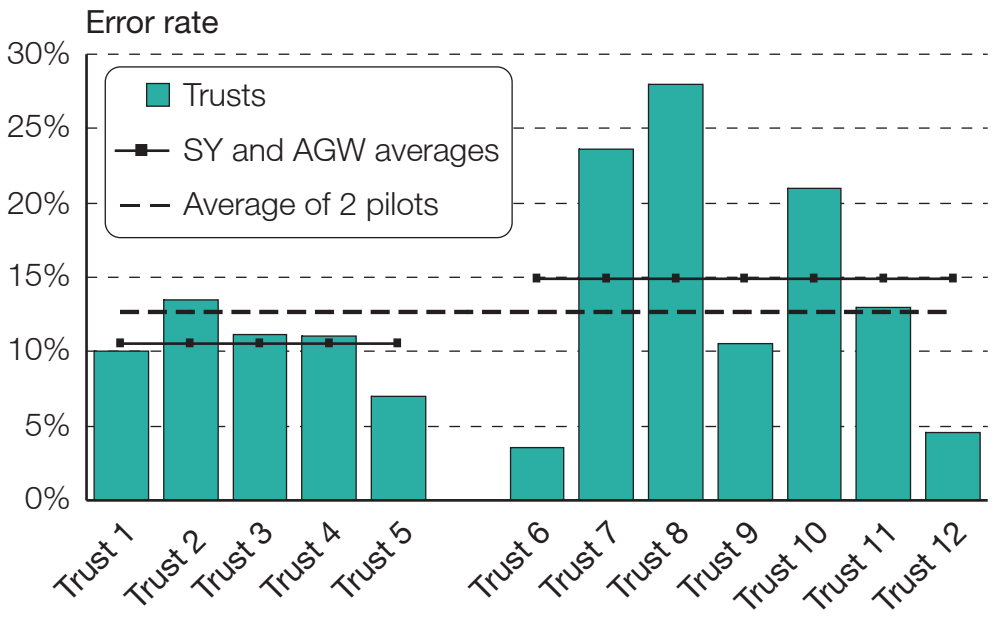
HRG error rate

95 The average HRG error rate, the proportion of episodes audited where coding error has resulted in an incorrect HRG, was 11.9 per cent. Due to the nature of the coding audit methodology, which counts errors in diagnosis and procedure codes rather than in overall episodes, we do not have data on the number of episodes that had at least one error for all trusts. The data we have indicates that this error rate was higher (around 30-40 per cent error).

96 There was considerable variation across trusts, ranging from an overall HRG error rate of 3.5 per cent to 28 per cent, as illustrated in **Figure 2**. There does not seem to be a clear relationship between error rate and foundation trust status, although there is much less variation among foundation trusts and the two worst-performing trusts are not foundation trusts. The variation is not explained by case-mix complexity (using teaching hospital status as a proxy), or number of sites. The two best-performing trusts had a high proportion of accredited coders and comparably large coding teams.

97 In considering the variation in error rates between trusts, it is important to remember that different specialties were targeted at different trusts. Although this does not explain the variation (General Medicine was covered at all trusts), it may have been a contributing factor. For example, the worst-performing trust had two of the less straightforward specialties audited (General Medicine and Geriatric Medicine).

Figure 2
HRG error rate by trust



98 The results also disguise considerable variation across and within specialties. **Figure 3**, overleaf, illustrates the overall error rate by specialty, which averages the results for all trusts where the specialty was audited, while **Figure 4** illustrates the variation in error rates for General Medicine across individual trusts.

Figure 3
Average HRG error rate by specialty

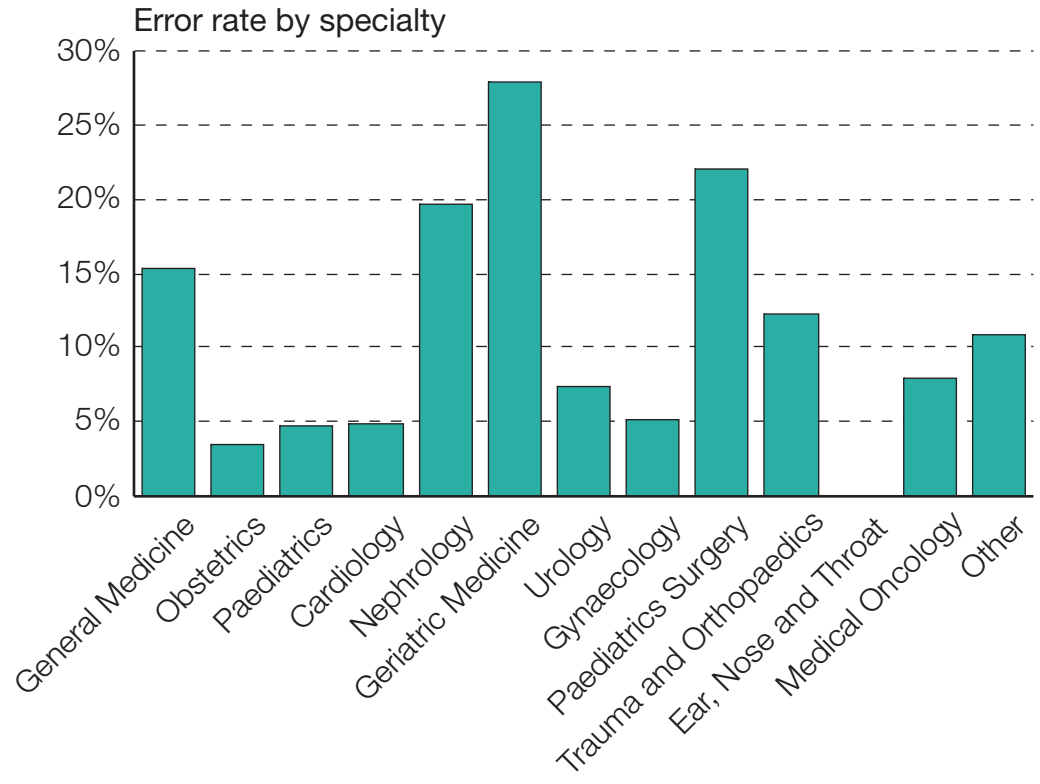
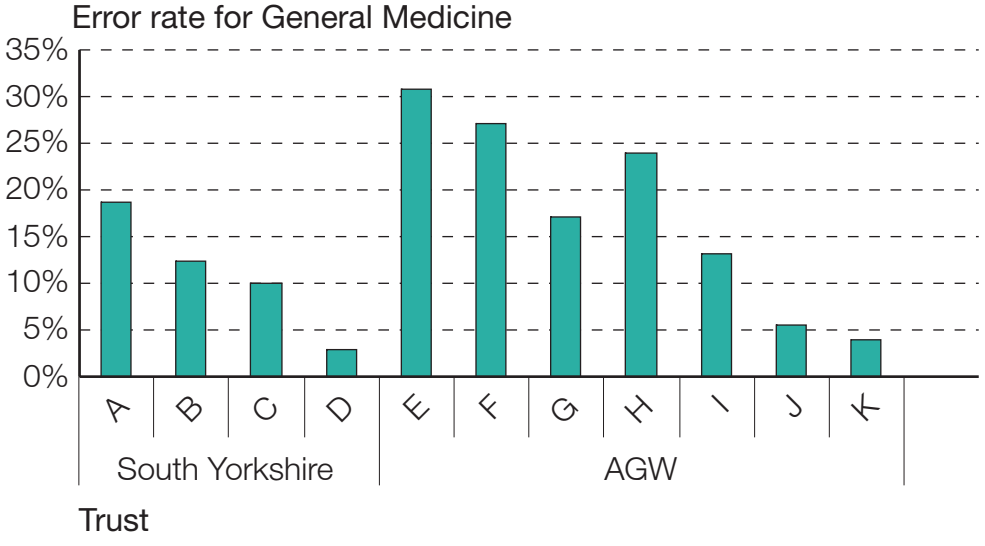


Figure 4
HRG error rate for General Medicine



99 Overall, the two specialties with the greatest level of HRG error across all trusts are General Medicine and Geriatric Medicine. This is not surprising, as they cover a broad range of patients, and coding is, as a rule, more difficult (there tend to be a greater number of comorbidities, for example). Overall, the more complex medical specialties (from a coding perspective) tend to have a higher error rate than the surgical specialties, but there are some exceptions to this (cardiology has a comparatively low level of HRG error, while Paediatric Surgery has a high error rate).

Causes of error

100 Overall, diagnosis coding has a higher error rate than procedure coding, and the error is more likely to impact on payments. There is greater and more material error among non-elective episodes.

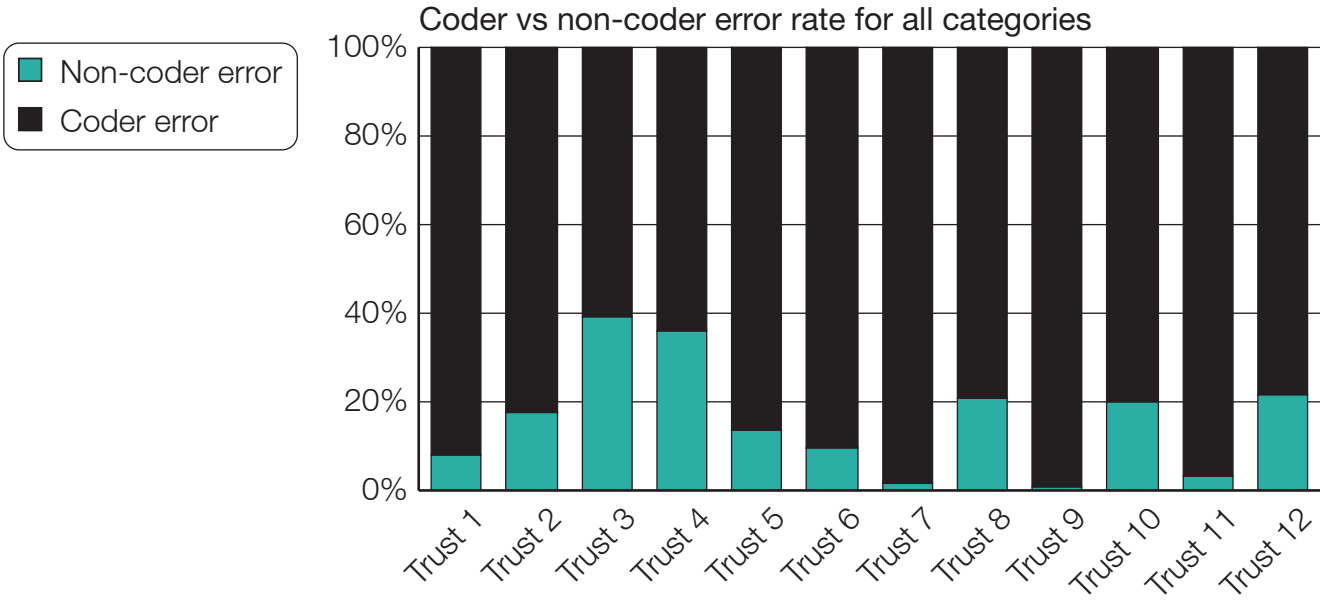
101 Both coder and non-coder errors are contributing to the high error rate. Non-coder error covers wider system issues, such as the information not being available at the time of coding, poor clinical documentation or coding to the specification of the clinician or management. Coder error generally relates to education and training issues, or simply human error (for example, transposing characters).

102 **Figure 5** (overleaf) shows the split between coder and non-coder error overall, while Appendix 4 contains a breakdown by the type of error (primary diagnosis, secondary diagnosis, primary procedure and secondary procedure).

103 While coder error accounts for 87 per cent of all errors identified overall, the level of non-coder error is as high as 40 per cent in some trusts. This highlights the need to address wider information issues, including the quality and availability of clinical documentation, clinical engagement in coding, and other issues relating to the coding process, if coding accuracy is to reach an acceptable standard. This requires ownership of the issues at trust management level.

104 The results indicate there is a substantial coder training and education agenda. A number of trusts have very few accredited coders, and this clearly has an impact on coding (and therefore payment) accuracy. In some cases, training and annual coding audit services are provided by the same company. This results in a lack of independence, in some cases a reinforcement of errors and a potential conflict of interest given the close relationship between audit and training.

Figure 5
Coder vs non-coder error by trust



105 Insufficient guidance or documentation on local coding practices (for example, absence of a complete and up-to-date policy and procedure document) was also a factor and coding managers should address this as a priority. The individual audit reports also point to a lack of or limited clinician involvement in coding and pressure to code to trust timetables as being factors.

Table 4 summarises the key characteristics of the coding environment.

Table 4
Key characteristics of the coding environment

Trust	HRG error	FCEs per coder p.day (based on 250 days)	Dedicated clinical coding manager	Proportion accredited (ACC qualified)	Coding deadline (no. of days after month end)	Trainer onsite	Policy and procedure document	Internal audit programme
Trust 1	10%	32.6	Yes	0%	5	Yes	Yes (under review)	Yes (monthly)
Trust 2	13.5%	27.8	Partial	0%	Month end	No	Yes (limited use)	Ad hoc
Trust 3	11.2%	16.7	Yes	0%	5	No	No (short coding procedure manual)	Ad hoc
Trust 4	8.6%	56.7	Yes	0%	7	No	Yes (not complete)	No
Trust 5	6%	39.2	Partial	36%	5th week	No	Yes	Ad hoc
Trust 6	5.1%	19.9	Yes	70%	5 days after discharge	No	Yes (under review)	Yes (monthly)
Trust 7	23.6%	45.3	Yes	20%	15	Yes	Yes	No (ad hoc)
Trust 8	28.0%	33.1	Yes	33%	6	N/a	Yes	Yes
Trust 9	10.5%	33.0	No	33%	8	No	Yes (under review)	Yes (quarterly)
Trust 10	10.2%	32.4	Yes	N/a	8	No	Yes	
Trust 11	13%	42.2	Yes	40%	14	Yes	Yes (needs updating)	Yes (quarterly)
Trust 12	4.6%	16.4	N/a	N/a	10 days after discharge	Yes	Yes (needs updating)	Ad hoc

Impact on payment

106 In calculating the impact of error on payment, we have had to assume that the episode HRG equates to the HRG for the spell. Further, in applying the tariff, we did not have information on length of stay so were unable to apply the short stay tariff. As a result, these figures need to be treated with caution.

107 The coding error has resulted in both overpayment and underpayment to trusts on individual episodes, overall underpayment for seven trusts and overpayment for five trusts. The absolute impact of the error on individual trust payments ranges from £7,091 to £67,698 for the sampled 200 episodes, representing between 5 per cent and 14 per cent of the total value of the sample.

108 Although it is not possible to extrapolate the findings to look at the impact on accuracy of payments to a trust overall with any confidence, due to the targeting and the size of the sample, it is clear that clinical coding error is impacting negatively on payment accuracy and that mechanisms should be in place to address this.

Adherence to data standards

109 Coding auditors have also raised concerns about the appropriateness of admissions in some trusts, where it appeared from the case notes that patients should have been treated as outpatients and there was no obvious reason for admission. This is a big issue in both pilots. However, current data definitions are ambiguous and their application is subjective, leading to different local practices in the treatment of day cases/outpatients – until this is resolved it will be difficult to address this area.

Overcoding/manipulation

110 There is evidence of overcoding (recording more diagnosis/procedure codes than is necessary) in some trusts, for example, inappropriate use of the living alone code, which has impacted on the HRG. However, in some cases this has been a historic practice, which is difficult to attribute directly to the financial incentives under PbR. There was no evidence that this is a deliberate strategy to increase payment.

Table 5
Impact of coding error on payment

	Overall impact on payment	Absolute impact of error on payment	Payment adjustment	% net difference pre/post audit	% absolute payment error of total sample value
Trust 1	(£628)	£10,744	-	2	5.39
Trust 2	£7,668	£28,385	+	12	7.01
Trust 3	(£5,211)	£15,199	-	26	n/a
Trust 4	£4,323	£8,227	+	21	n/a
Trust 5	£358	£13,906	+	2	4.08
Trust 6	£1,179	£6,383	+	12.3	2.87
Trust 7	£1,794	£43576	+	2	n/a
Trust 8	(£44,450)	£67,698	-	36	14.60
Trust 9	(£9,533)	£18,861	-	33	7.09
Trust 10	£2,983	£20,721	+	8	6.57
Trust 11	£3,148	£30,322	+	1.3	12.76
Trust 12	(£100)	n/a	-	1	n/a

111 There was also evidence of trusts actively working to optimise their coding to maximise income, but where this is happening they are doing so within existing coding rules (that is, the practice is not fraudulent). There is some poor practice emerging in this area – for example, in one trust the clinical coding team reports to the director of finance, and is involved in discussions about the financial impact of coding changes on a regular basis. Consideration needs to be given to whether this is appropriate and how it should be addressed.

112 In the process of conducting the pilots, we received anecdotal feedback from the auditors about coding issues in other parts of the NHS that may be of concern. For example, in one trust coders are being encouraged to put Z codes from ICD-10 on obstetrics episodes to move the episode to a more complex HRG. In another trust, the coding manager is reported to be under pressure to ensure the levels of complex HRGs in obstetrics are the same as peer trusts. A commercial company has advised them which Z codes could be added to achieve this.

113 It was agreed with the NHS Counter Fraud and Security Management Service (CFSMS) at the beginning of the pilots that any cases of suspected fraud would be referred to them for further investigation. The audits identified no such cases. There may be benefit in providing specific guidance to coding auditors on how to identify potential fraud and incorporating this into the methodology in future.

Perverse incentives

114 There was concern among coding managers and coding auditors about idiosyncrasies in the HRG grouping algorithm and/or coding rules and the resulting disincentives to code correctly. The coding audits have highlighted examples where correct coding disadvantages a trust and leads to unfair payment. In addition to undermining the fairness of the PbR regime, it acts as a disincentive for accurate coding. Coding auditors have raised concerns that HRGs and the grouping algorithm may not be constructed with any coding input. While there is clinical input, clinicians do not usually have a detailed understanding of coding rules, and will be less able to identify perverse incentives from a coding perspective.

115 Examples have been reported where more accurate coding of a complex patient results in a lower payment – which means it is more lucrative to apply general rather than specific coding (when coder education is working in the opposite direction).

116 A paper setting out the specific issues identified in the pilots will be submitted to the Department of Health and Connecting for Health separately.

Evaluation

117 While there is not an agreed standard or target for HRG error in the NHS, the reported level of 11.9 per cent seems intuitively too high in a system where clinical coding is the primary determinant of payment[l]. From a data quality perspective, the standards of accuracy set out in the information governance toolkit for achieving Levels 1 and 2 are 90 per cent accuracy for primary diagnosis/procedures and 80 per cent for secondary

l The percentage HRG change across Victorian hospitals in Australia for the 2000-01 clinical coding audit, which is the most recent data available to us, was 9.8 per cent (Source: 2000-01 Audit Report). While this is a comparable level of error with the NHS, the Victorian government has been actively working to reduce this through external audit and other mechanisms.

For US Medicare, in November 2005, 5.2 per cent of the dollars paid nationally did not comply with one or more of Medicare coverage, coding, billing and payment rules. Projected overpayments were $11.2 billion and the underpayments were $0.9 billion (Source: CMS website).

diagnosis and procedures. The average accuracy rates reported for the pilots are below these levels, with performance varying widely across individual trusts.

118 The level and nature of clinical coding error is not so high as to destabilise the PbR regime, but it does undermine it, raising concerns about the accuracy and fairness of funding flows. In addition, it has major data quality, epidemiological and clinical implications.

119 The pilots demonstrate that there is a need for a centrally coordinated external clinical coding audit programme, which ensures objectivity and independence of results. This should aim to improve accuracy levels and, in particular, reducing the level of variation across different providers. Even where there are strong local monitoring arrangements in place at commissioner and provider level, there is a need for an external and independent audit to supplement and provide additional impetus to improve data quality. Both pilot sites have commented on the value of the audits in highlighting data quality and payment issues from an independent and objective perspective, and in opening dialogue about this between commissioners and providers and across the health economy.

120 In addition to the local benefits, a comprehensive audit programme has a clear national benefit in improving the quality of data in the NHS and refining the policy framework (for example, the audits have highlighted inconsistencies in coding rules and/or the HRG grouping algorithm). It is also the first time that information on clinical coding accuracy has been presented in this way, and national stakeholders, including Connecting for Health and the Information Centre, have commented on its value.

121 Finally, central coordination and oversight has highlighted differences in the application of the national standard methodology across clinical coding auditors, including differences in reporting and treatment of cases which are unsafe to audit. It has also highlighted an absence of robust quality assurance around the process. Ensuring consistency and accuracy of the audits that are undertaken is an important benefit of a centrally coordinated approach.

122 The proposed clinical coding audit programme should replace the requirement for trusts to commission annual audits as part of the information governance toolkit. These audits are often referred to as external audits, but they are actually internal audits, undertaken by the trust themselves using auditors who are external to the trust. The proposed external audit programme should be supplemented by regular internal audit (ideally monthly), and trusts may choose to continue commissioning work from an external auditor for this purpose.

Role of clinical coding audits

123 The audits that were undertaken for the pilots aimed to test the level and nature of clinical coding at acute trusts and assess the impact this was having on payments. In practice, they enabled us to identify data quality issues which are impacting on payment accuracy and which warrant further investigation, to identify potential manipulation or abuse of the system and to identify perverse incentives in the policy framework which need refinement (although there are some improvements that can be made to the methodology).

124 Initially, some of the organisations participating in pilots raised questions about the purpose of the external clinical coding audit programme and whether it should aim to provide assurance on clinical coding and payments, rather than simply identifying issues. However, there were mixed views about the value and cost-effectiveness of this. The size and cost of the audits required to provide assurance would be prohibitive and this approach is not used internationally. By the end of the pilots, there was recognition that external coding audit alone should not be a vehicle for achieving assurance. The overwhelming view was that the primary aim of the external audits should be to improve data quality – if we achieve this, other objectives such as payment accuracy will follow.

125 In response, our recommendation is that the clinical coding audit approach should be data quality-led, but should also support policy refinement and payment accuracy by:

- identifying the key coding issues that are impacting on payments locally and which warrant further investigation; and
- identifying upcoding and potential manipulation.

126 The audits should contribute to the overall objective of the PbR assurance framework to support payment accuracy, in conjunction with other components of the framework. It should be the responsibility of local NHS bodies to investigate current issues in order to reconcile payments at the end of each quarter. The audits should also aim to identify cases of potential fraud, and refer these to the CFSMS in line with an agreed protocol.

127 In order to support both payment accuracy and improvements in data quality, the audit programme will need to include both narrow and more widely targeted audits. A narrowly targeted audit will be able to address specific issues which are impacting on payments and pick up any manipulation or gaming (assuming the audit is correctly targeted). Conversely, a wider audit is more likely to pick up a range of data quality issues. Themed audits across different organisations will enable us to address differences in recording and enable comparison.

Actions from the clinical coding audits

128 The audits provide a picture of the state of coding within a trust for the specialties targeted (the level and nature of coding error, and the reasons for the error), some insight into broader data quality issues (for example, adherence to data definitions); and the impact that the error has on overpayment and underpayment in the areas audited.

129 As the coding audits were retrospective (generally covering the last quarter of 2005), the results are not able to inform actual payment reconciliation under the current policy (that is, flex and freeze rules which mean that payment adjustment is only possible for the current quarter). The audits would need to be much more responsive, covering the most recent quarterly, and more narrowly targeted to achieve this. However, they highlight areas that PCTs may wish to scrutinise in the future to ensure correct payment for episodes and inform future adjustments.

130 If payment reconciliation is unlikely to be a direct action stemming from the audits (although they should influence future reconciliation), there will need to be clear consequences associated with the coding audits in order to achieve their objectives. These should include:

- penalties for poor data quality;
- sharing of good practice and lessons learnt;
- influencing ALE assessments and the Annual Health Check;
- publication of a league table showing error rates and possibly a national error rate statistic;[l]
- further investigation by the PCT, which may influence future payments;
- further investigation at the expense of the trust, where data quality is poor or specific concerns are raised; and
- referral of cases of suspected fraud to CFSMS for further investigation in line with an agreed protocol.

l This would need to be handled carefully, as NHS trusts and the independent sector may interpret this as the benchmark.

131 While penalty clauses are already included in some foundation trust contracts, the model contract has no existing provision for penalties for incorrect data that could be applied in response to these audits and no penalties are in place for other NHS trusts.

132 The idea of applying penalties for poor data quality was a subject of much debate among the pilot sites and other external commentators. While some organisations (particularly trusts) were hesitant about the application of penalties, there was a general recognition that they do work and that they should be introduced as long as this was done fairly, gave an opportunity for improvement, and did not penalise trusts for factors beyond their control (such as highly complex case mix, which is more difficult to code accurately). The fact that error impacts not only on payments but on the ability to plan and manage effectively gives further weight to the use of penalties.

133 A standard penalty clause should be developed and incorporated into the revised model contract which the Department of Health is developing, and applied to all NHS providers. The contract clause should be phrased in such a way that poor results from an external audit of coding or data quality will attract a penalty, most likely paid to the host commissioner. To avoid disputes and ensure that Trusts are not penalised unfairly, the nature and level of penalties will need to be considered carefully – for example, should they be applied for results above a certain error rate, or for failure to improve. There should be time for improvement before penalties are imposed, so penalties should not apply to the first audit. However, the view from commissioners was that sanctions should come into force fairly quickly, given that inaccurate data undermines their ability to plan. Penalties should be the subject of consultation with the NHS.

Extending the scope

134 Prior to commencement of the pilots, we highlighted the need to test the focus of the external audit programme on the accuracy of clinical coding – whether it was appropriate or needed to be broader.

135 Both pilot sites have indicated that clinical coding is too narrow a focus, and external audits should cover broader aspects of whether activity has been correctly recorded (including adherence to data standards and other data quality errors that affect payments). The focus of the pilots on clinical coding was based on the assumption that local PCT arrangements should already cover, to some extent, the areas which are important for comprehensive monitoring of payments, including:

- inappropriate admissions/transfers;
- medically unnecessary diagnoses and procedures;
- re-admissions due to incomplete care/premature discharge; and
- other errors that affect payment (for example, wrong commissioner code, duplicates).

136 The external clinical coding audit programme was intended to provide additional support to identify and address coding error and upcoding, rather than to comprehensively support payment accuracy. However, findings from the pilots indicate that external audits should also cover wider aspects of whether activity has been correctly recorded. Although coding auditors are able to identify some of these issues, the audit methodology would need to be refined to properly address these areas.

Changes to the methodology

137 There are a number of issues that have arisen with the detailed audit methodology that will need to be addressed in moving to implementation.

138 As noted above, the methodology should be amended so that it can genuinely address the following:

1) Was the patient actually treated in the hospital?
2) Was the activity (and patient details) accurately recorded?

139 There are methodologies which are used by commissioners – for example, in Birmingham and the Black Country – as well as bespoke work that the Audit Commission has undertaken previously, which could be adapted to meet this purpose.

140 Other issues that have been identified include:

141 **Sample size** While 200 episodes is twice the size of audits currently undertaken by individual trusts under the information governance toolkit, to have confidence in the results when extrapolated across the trust from a statistical perspective, and reach firm conclusions on the error rate, the sample size will need to be larger in most cases. Concerns about the size of the sample were raised at both pilot workshops.

142 The sample sizes used in the pilots, for the most part, mean that we can only be confident that the actual error rate lies within quite a broad range. For example, for Trust 2, the sample size for General Medicine was 75 episodes. If the actual error rate was 15 per cent, we are 95 per cent confident that the error rate for all General Medicine episodes is between 7.6 per cent and 22.4 per cent. To reduce the range, we would need to take a larger sample.

143 The size of the sample that we should be looking at depends on how narrowly the audits are targeted (the size of the population), how narrow we want the confidence intervals to be and the actual error rate. If the actual error rate is low, we can look at a smaller sample and be confident in extrapolating the results.

144 In the example used above, to have a confidence interval of +/- 5 per cent from the identified error rate, we would need to sample 140 records; for +/- 1 per cent, we would need to sample 140 records. Alternatively, if we did not target the audit at General Medicine, and wanted to look at the error rate in Trust 2 overall for the entire year, to be 95 per cent confident that the error rate was within +/- 5 per cent, we would need to sample 195 records. This increases to 533 for +/- 3 per cent and 4,088 for +/- 1 per cent.

145 Feedback from stakeholders has been that having a statistically significant sample is important for credibility, particularly among clinicians, although the main message is that the sample size should be fit for purpose. Although the actual sample size will vary, to be confident in the results, it would be closer to 500 episodes per audit. This is consistent with the size of the audits undertaken in both the United States and Australia. As a result, this sample size has been used in the cost estimates given for implementation in Section 6. However, there are different statistical approaches that may be used, for example, using smaller samples and control charts, and these should be explored further in implementation.

146 **Sample selection** The episodes contained in the audit sample are based on a randomly generated picking list from the hospital's Patient Administration System (PAS). Therefore, although the population for the sample is independently selected, the actual cases are selected by trust. The selection is supposed to be random, but the trust has the opportunity to self-select. We also found that not all trusts followed the specification that was given for extracting the random sample. The alternative is to independently select the exact episodes to be audited. Although this creates additional work at the outset and may create problems if the required case notes are in use or otherwise unavailable, it is a more rigorous process.

147 **Episodes versus spells** Coding audits are by their nature episode-based – they reflect episodes recorded in the case notes. This makes conclusions about the impact of error on payments more difficult, as payment is spells-based. The methodology should be adapted to enable reporting on spells-based going forward.

148 **Electronic coding audit** Use of software to support the coding audit and increase automation would make the audit process much more efficient. There are at least two existing software packages that could be used. Connecting for Health is also exploring this avenue at present and we propose to work closely with them on this.

149 **Reviewing PCT arrangements** It is difficult to translate the reports on the audits undertaken at the acute trusts into a meaningful report for PCTs without an understanding of their existing monitoring arrangements. A basic review is being undertaken as part of the pilots, gathering intelligence from the appointed auditor/audit manager of the lead commissioning PCT. In future it would be beneficial to build a short review of these arrangements into the audit methodology to ensure that recommendations to PCTs are contextualised. This would be undertaken by the Audit Commission itself rather than the coding auditors.

150 **Other care types** Most of the organisations participating in the pilots felt there would be a lot of value in extending the PbR assurance framework fairly quickly to incorporate care types other than acute inpatient which are either already covered by PbR, or will be in the near future – outpatients and critical care, in particular.

151 **Identifying overcoding** While overcoding is identified in the audits, it is not included in the overall error rate and coding auditors find it difficult to identify overcoding, particularly in relation to comorbidities. Creating rules around the recording of comorbidities has been historically difficult. In general, comorbidities should only be recorded where they affect the care, influence the health status or are the reason for the hospital stay. The minimum number of codes which accurately reflect the patients' care should be recorded. Background or chronic problems which are no longer active or do not influence the care provided should not be recorded. However, coders find it difficult to identify which comorbidities are relevant, particularly where the case notes are not clear. They are also required to code any conditions that the clinician has identified as relevant and are not able to question this. As a result, both coders and coding auditors are not confident in challenging the recording of comorbidities and overcoding. In an environment where clinicians are increasing aware of HRGs and the financial impact of coding, a link to clinical audit may be of benefit in identifying overcoding.

152 Finally, the culture of clinical coding auditors and the audit methodology are by their nature data quality-focused. These are as they should be. However, it does create difficulties in applying the methodology with full independence from the trust and from the perspective of the commissioner. Auditors tend to report positively on most trusts, stating that the coding is of a high or very high standard, while the actual error levels vary considerably.

153 There may be some merit in developing a consistent scale for how trusts are rated.

154 There are also aspects of the coding rules and methodology that need to be resolved. For example, we found inconsistent treatment of episodes which are unsafe to audit[l], with some auditors counting them as part of the sample (but not including them in the error rate), and others excluding them. Connecting for Health has advised that it would be inappropriate to count these cases as errors, as there is no information available to prove or disprove coding accuracy. However, they should be clearly highlighted for the medico-legal implications they present.

l 'Unsafe to audit' means there was insufficient information available to audit the episode, usually because the auditor could not find any evidence of the episode in the case notes.

4

Other issues

Data definitions

155 The area identified by the pilot sites as having the biggest impact on payment accuracy was data definitions. Several key definitions are open to interpretation (notably, the definition of an admission) and so are not consistently applied.

156 Before any audit of these issues can be effective, ambiguity around data definitions (for example, the definition of an admission) and inconsistency in how these are applied on the ground need to be addressed as priorities. The current ambiguity inherent in the definitions is creating difficulties and it is not sufficient to leave these to local interpretation. Work that Connecting for Health has undertaken with users in developing SUS has also identified critical data definition issues and has reiterated the need to address these. In particular, it is important to ensure that definitions are operationally viable, fully incorporated in the data dictionary, and that the data dictionary and reference cost definitions are aligned.

157 This is not a new issue. It was also raised in the South Yorkshire PbR laboratory report in March 2006, which stated that 'A national initiative on data definitions is urgently needed to provide clearer data standards'. The definition of a day case has been problematic for many years.

158 Appendix 4 contains a summary of the issues that are proving problematic in the two pilot sites, many of which were also raised in a separate paper to the Department of Health from the South Yorkshire PbR laboratory. The Department of Health is considering how to resolve them. A clear process, involving Connecting for Health and the Information Centre is required, along with urgent resolution of current issues. The pilot sites commented on the reduction in responsive and support from the centre in resolving questions and ambiguities with respect to data definitions and how certain cases should be treated where there is local disagreement. Suggestions for resolving this included a more responsive PbR team at the Department of Health, and reinstatement of the frequently asked questions (FAQs) for NHS bodies dedicated to information issues. Where clarifications are provided by the national PbR team in response to queries, these should be published on a national website.

159 A similar issue exists with clinical coding rules, with a suggestion that coding queries and the response from Connecting for Health should be made available to all trusts.

Reference costs

160 The audit of reference costs and a review of behaviours against the PbR Code of Conduct were raised as additional areas that should be included in the PbR assurance framework. Both of these areas were included in the original proposal to the Department of Health. Feedback from the pilots has highlighted concerns about the quality and consistency of reference costs and the need for more robust checks. External checks on reference costs was recommended in the Audit Commission's recent report to the Secretary of State for Health on the review of the NHS financial management and accounting regime, and would support the Department of Health's actions stemming from the Lawlor review to improve the calculation of the national tariff.

Behaviours

161 Behaviours are more difficult. It is hard to see how external audit or other checks could reinforce appropriate behaviours effectively, for example, through reviewing behaviours against the Code of Conduct. Therefore, we have not made any recommendations for action in this area as part of the PbR assurance framework.

Strengthening PCT monitoring

162 In addition to the pilot work, the scope of the PbR assurance framework also involved strengthening local arrangements for ensuring data quality and accurate payments – notably PCT monitoring arrangements.

163 PCTs' arrangements for monitoring data quality and payments have progressed since the early work undertaken in 2005 to develop the proposal for the PbR assurance framework. Some PCTs are now undertaking regularly reviews of data at HRG level and undertaking their own coding reviews. However they are still variable and strengthening them is a continuing priority if PbR is to operate effectively at the local level.

164 The proposal was to develop and feed proposed minimum standards for PCT monitoring of payments and data quality into the fitness for purpose assessments and competency framework that has been developed by the Department of Health.

165 The fitness for purpose assessment framework (the monitoring diagnostic tool, in particular) goes some way towards this. It broadly covers the areas that are important for monitoring payments under PbR. However, the complexity of the overall process and the lack of specificity will make it difficult to use as a tool for strengthening arrangements with respect to PbR. The minimum standards set out are also too low for PbR assurance purposes. The message from the pilots was that further support which gives greater direction and specifically addresses PbR would be of benefit to commissioners, although this would not necessarily need to be part of the PbR assurance framework.

166 Guidance that the Audit Commission intends to issue to auditors during 2006/07, which will include steps that auditors should take to check arrangements for ensuring accurate payments before signing off the accounts, will be one mechanism for achieving this (see Development of an audit guide). In addition, PCTs would benefit from more specific support and guidance to develop local monitoring arrangements, for example, a tool or checklist which enables them to check their arrangements against best practice.

Development of an audit guide

167 Auditors need to satisfy themselves that income is not materially overstated or understated in the accounts of NHS trusts and that the corresponding expenditure is not materially overstated or understated in the accounts of PCTs, and that such expenditure has been applied for the purposes intended by Parliament. With the introduction of PbR, auditors need to satisfy themselves that robust corporate systems and internal controls to ensure the accurate capture and recording of all activity (NHS trusts) and for challenging and accepting the reasonableness of charges (PCTs) are in place, from which they can take assurance in giving their opinions on the financial statements.

168 By spring 2007, the Audit Commission intends to develop guidance, in the form of an audit tool, on the steps that auditors may wish to take to satisfy themselves that income/expenditure is not materially over or understated in the accounts of NHS bodies and that bodies have put in place adequate and effective systems to ensure data quality.

169 This work needs to be undertaken regardless of whether the PbR assurance framework is implemented and what form it takes, and will be included in the Audit Commission's annual audit fee, but is very closely related and has a number of interdependencies. This would be an important way of reviewing whether PCTs and trusts are taking the right action locally to monitor and assure data quality and payments. It may also rely on other

parts of the PbR assurance framework, notably any external clinical coding or data quality audits undertaken. The absence of a PbR assurance framework is likely to increase the amount of work that auditors need to undertake.

Strengthening trust arrangements

170 The Audit Commission has been working with Connecting for Health to refine the information governance toolkit so it is fit for purpose in the PbR environment and provides a greater level of assurance on data quality in the NHS.

171 We are yet to review the process followed in the Annual Health Check for validating trusts' declarations of compliance against Standards for Better Health (SfBH) Standard C9. Findings from the first year will be used to determine whether additional, broader work on data quality is required in the future and what this should look like.

Assurance of the tariff

172 There continue to be a number of concerns about the accuracy of the tariff and the assurance around this and the fairness of the overall policy framework.

173 The term 'PbR assurance framework' raises expectations that the framework will provide assurance around the policy itself. As a result, organisations participating in the pilots have raised issues such as the validity of the specialist top-ups and assurance on the accuracy of the tariff itself, which they expect the assurance framework to be covering.

174 A decision is required on whether the scope of the PbR assurance framework is widened to include assurance of the tariff itself, incorporating the actions of the Department of Health as a result of the Lawlor review, or whether it focuses on monitoring payments and data quality. In this case, the Department of Health may wish to reconsider use of the term 'assurance framework'.

175 Regardless of whether actions are considered as part of the assurance framework, feedback from the pilots reinforces the Audit Commission's recommendation in the recent report to the Secretary of State for Health that 'some external assurance on the figuring and rationale underlying the tariff' in its recent report to the Secretary of State for Health.[I]

I *Audit Commission Review of the NHS Financial Management and Accounting Regime: A Report to the Secretary of State for Health*, Audit Commission, 2006.

5

Recommendations

Summary

176 The PbR assurance framework pilots have demonstrated that there is a higher level of clinical coding error than is desirable and this is having an impact on funding flows and the accuracy of payments.

177 The pilots have confirmed that, at least over the next few years, there need to be stronger arrangements in place to drive improvements in clinical coding, wider data quality and to support refinements in the policy framework. This will benefit NHS bodies locally, through more accurate payments, better data for clinical and management purposes, and a clearer framework which results in fewer disputes. It will also have benefits nationally, for the Department of Health, patients and taxpayers.

178 It is clear that the focus of any external audits should not be limited to clinical coding. Broader data quality issues – the correct counting and recording of activity, not just the coding of diagnoses and procedures – need to be part of the approach. The audit methodology that was applied in the pilots will need to be modified to address this.

179 Despite its limitations, the benchmarking data has proven useful in identifying areas to target for the clinical coding audits. However, there is a question about how narrowly any future audits should be targeted. To be useful in assessing payment accuracy, they may need to be targeted more narrowly (at specific HRGs); alternatively, a wider audit will make a greater contribution to data quality.

180 The pilots have also demonstrated the immediate need to resolve some of the issues with data definitions and work towards addressing perverse incentives that currently exist in the grouping methodology and tariff structure. The coding audits have picked up some specific issues relating to the grouper and coding rules.

181 Pilot organisations have also raised concerns about the accuracy of reference cost data and adherence to the Code of Conduct – both areas were included within the original scope of the PbR assurance framework and there is clearly merit in addressing reference costs in particular.

182 The participating organisations recognised the importance of strengthening local arrangements. In particular, some PCTs felt it would be valuable to have some guidance or a checklist on what they should be doing to monitoring PbR locally and what issues they should focus on, to ensure they were doing everything they should be.

Individual recommendations

183 Based on the findings from the pilots, we have the following recommendations for the Department of Health:

1. An external clinical coding audit programme should be implemented as a core part of the PBR assurance framework, but with some changes to the approach used in the pilots, including:
   - a larger sample size;
   - spell-based rather than episode-based;
   - use of coding audit software; and
   - extension of scope to include outpatient and critical care data.

   This should replace the annual information governance toolkit audits currently commissioned by trusts, but should be underpinned by a requirement for trusts to conduct regular internal audits.
2. The benchmarking methodology should be further developed and used to identify areas of concern and target the audits. The targeting of audits should vary from quite broad (trust-wide) where there are wider concerns about data quality to very precise (focusing on individual HRGs) where specific issues have been identified as impacting on data quality and payments. Themed audits, where the same area is covered across all NHS providers, should be an integral part of the programme.
3. A proper analytical tool to support a more sophisticated approach to national benchmarking should be developed in partnership with the Information Centre. Ideally this tool should support routine national analysis of hospital activity data for a range of purposes, not just targeting data quality audits, with the data returned systematically to NHS bodies.
4. The external audit programme should be extended to encompass the broader recording of activity. A distinct audit methodology should be developed to address areas such as the correct use of data standards and data definitions, drawing on

approaches that are already being used locally. This audit would be less intensive and would not require the use of clinical coding auditors, although they could be trained to undertake this work as an extension of the coding audit. It may also encompass testing of basic data fields, such as duplicates and use of correct PCT codes, although this should continue to be a key responsibility locally and will be partly addressed by the introduction of SUS.

5. The following actions should result from the coding audits:
   - penalties for poor data quality;
   - inform ALE assessments and the Annual Health Check;
   - publication of a league table showing error rates and possibly a national error rate statistic – it may be worth considering a cost-weighted HRG error rate;[l]
   - further investigation by the PCT, which may influence future payments; and
   - further investigation at the expense of the trust, where data quality is poor or specific concerns are raised.
6. There should be a standardised penalty clause in all contracts with NHS providers, including foundation trusts and the independent sector, which can be invoked by the PCT in response to a poor result from external audit. This may be phased in gradually, and should not apply to the first audit. The Department of Health should consult with the NHS on this issue.
7. The Department of Health should consider whether assurance of PbR policy, including the tariff, should be included in the PbR assurance framework. Regardless of whether it is included, there should be some external assurance on the calculation of the final tariff structure which is published. Consideration should be given to whether 'PbR assurance framework' is the current term for what we are trying to achieve, and whether an actual document called the assurance framework should be published as a result of this work.
8. Inconsistencies and ambiguities in current data definitions and standards as identified through the pilots should be addressed as an immediate priority, and a clear process put in place so that similar issues which are raised by NHS bodies in future can be properly addressed. While it may take some time for these to be reflected in the information infrastructure, the Department should issue interim guidance before the end of the year on how these cases should be treated.

l This would need to be handled carefully, as NHS trusts and the independent sector may interpret this as the benchmark.

9. There should be an independent audit of NHS trusts' reference costs, which would assess the application of the national costing methodology at local level and the quality of the reference cost data, and the information should be published.
10. A dedicated tool or checklist should be developed for PCTs, enabling them to check their arrangements for monitoring activity and payment under PbR against best practice.
11. The Audit Commission continues to orchestrate the external components of the PbR assurance framework and to support strengthening of local arrangements through routine audit work and the development of the PCT tool.
12. The scope and targeting of the external programme should be reviewed on an annual basis, as data quality improves and local arrangements strengthen.

184 In addition to the recommendations above, we propose that the findings from the pilots are published in a report aimed at the NHS, with the emphasis on good practice and lessons learnt.

6

Implementation

185 Should our recommendations be accepted by the Department of Health, we propose to adopt a phased implementation approach, commencing in October.

Phase 1 (to March 2007)

- Developing the benchmarking tool in conjunction with the Information Centre.
- Roll-out of the external clinical coding audit programme across all NHS providers for acute inpatient episodes, including refinements to the methodology.
- Developing the audit methodology for broader recording of activity (data standards and data definitions).
- Developing the audit methodology for outpatient and critical care data.

Phase 2 (2007/08)

- Ongoing refinements to the benchmarking tool.
- Comprehensive external audit of clinical coding (covering approximately 74,000 episodes).
- Testing and implementing the audit programme for outpatient and critical care data.
- Developing the audit methodology for mental health data.

186 In addition, the audits will need to be extended to cover independent sector providers of NHS services. This should ideally commence in 2007/08, but further work is required to determine how the approach needs to be adapted to support this, particularly given the different information infrastructure, and a feasible timescale for implementation.

187 In addition, the following should be a priority for the Department of Health to action in the remainder of 2006/07:

- refinement of data definitions;
- external audit of reference costs; and
- development of a standardised penalty clause for poor data quality.

188 The estimated annual cost of the targeted clinical coding audit programme is £4.45 million, or approximately £30,000 per PCT. There would be implementation costs of

£2.02 million in the remainder of 2006/07, which includes a further 30 audits. The costs are based on an assumption that 75,000 episodes will be audited per year (around 0.5 per cent of all episodes). Based on the average episode to spell ratio across NHS providers, this translates to 67,000 spells.

189 There will be an additional cost of extending the audit programme to address wider recording of activity. This will depend on how the audit is organised – in particular, whether it is undertaken as a separate audit or as an extension of the clinical coding audit – but is expected to be in the order of £2.25 million (based on five days per audit).

190 Further work needs to be undertaken to estimate the cost of extending the audits to cover independent sector provider and outpatient and critical care data.

191 There are several options for how the audit programme could be organised initially. At the minimum, there should be a three-year rolling programme, covering every provider of NHS services over this period, and the option to do further follow-up or targeted work at poorly performing trusts. This would increase the coverage and sample size at individual trusts. Alternatively, all providers may be audited in the first year to establish a baseline, covering 500 episodes per trust, with a risk-based approach applied in the following years.

192 The approach will need to be defined more clearly in the early stages of implementation, and will depend on the results from future benchmarking work, the balance between the data quality and payment accuracy focus and political acceptability.

193 We would implement the external audit programme as part of the annual PCT audit under existing Audit Commission powers. A number of issues will need resolution prior to implementation, including the sharing of reports and whether a single report is produced for the host PCT which is then shared with other interested commissioners, or whether several PCT reports are produced for a given trust.

194 One possible constraint on full implementation of the annual audit programme is the availability of clinical coding auditors. We estimate that approximately 14 Full Time Equivalent (FTE) clinical coding auditors will be required for implementation. Although there are currently 36 accredited coding auditors in the NHS and the numbers are increasing, two-thirds of these are employed by the NHS. It may be necessary to adopt a phased approach to implementation in the first year, increasing the scale of the programme as more resources become available.

Appendix 1: Benchmarking methodology

Data

The data used for the benchmarking exercise is a download of the Department of Health's HES database, covering ten quarters in total (2003/04 Quarter 1 to 2005/06 Quarter 2). The final two quarters for 2005/06 were not available at the time of the analysis. The data was provided at FCE level and included:

- clinical information about diagnoses and operations;
- information about the patient, such as age group, gender and ethnic category;
- administrative information, such as time waited and date of admission; and
- geographical information on where the patient was treated and the area in which they lived.

An On Line Analytical Processing (OLAP) database was created from the HES records, containing only the fields required for analysis, to enable more efficient analysis.

Scope

The analysis focused on the acute trusts within the two pilot sites, covering:

- South Yorkshire Strategic Health Authority
  - Barnsley District General NHS Foundation Trust
  - Doncaster and Bassetlaw Hospitals NHS Foundation Trust
  - Rotherham NHS Foundation Trust
  - Sheffield Teaching Hospitals NHS Foundation Trust
  - Sheffield Children's NHS Trust
- Avon, Gloucestershire and Wiltshire Strategic Health Authority
  - North Bristol NHS Trust
  - Royal National Hospital for Rheumatic Diseases NHS Trust
  - Weston Area Health NHS Trust

- Salisbury Healthcare NHS Trust
- United Bristol Healthcare NHS Trust
- Royal United Hospital Bath NHS Trust
- Swindon and Marlborough NHS Trust
- Gloucestershire Hospitals NHS Trust

Indicators

The following indicators were used for the pilots:

- admittance/referral patterns;
  - mean spell beddays
  - day case ratio
  - elective/emergency ratio
  - medical to surgical within spell episode ratio
  - short stay rates
  - FCE to spell ratio
  - outpatient procedure rates
- coding quality;
  - mean diagnoses recorded
  - not elsewhere classified diagnosis rates
  - not specified diagnosis rates
  - not elsewhere classified procedure rates
  - not specified procedure rates
  - non-specific chest pain diagnosis rates
- HRG complexity;
  - complication rates
  - complicating diagnosis rates
  - complex elderly rate
  - HRG S22 rates (in other words, planned procedures not carried out).

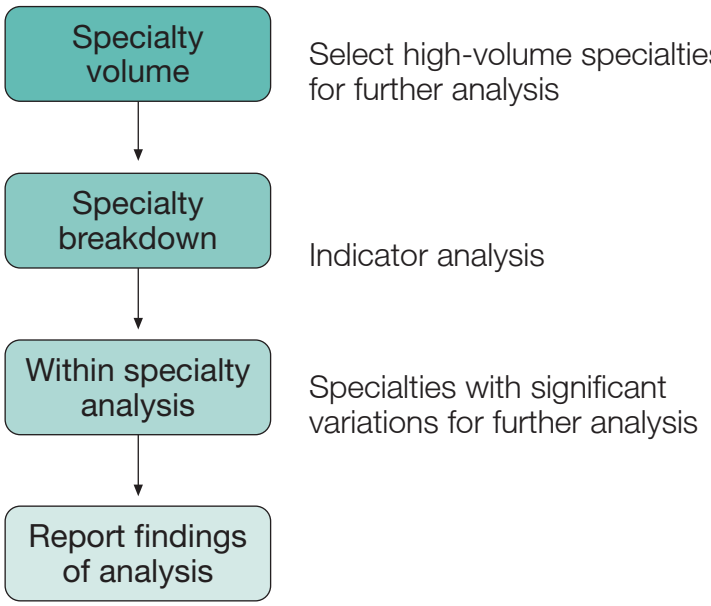
Methodology

The analysis has been undertaken at FCE level on a set of indicators for each provider, comparing them against their peer group (using provider categories as published by the Healthcare Commission) to identify significant variation.

A 25 per cent variation rule was applied to determine which areas were significant. For example, where the peer group shows a day case rate of 50 per cent for a particular specialty and the target provider shows a day case rate of below 25 per cent or above 75 per cent for the same specialty. This difference would deem to be significant and further analysis would be conducted on the specialty to explain the difference – this, typically, would be to drill-down to HRG level.

The diagram below depicts the analytical process. A hierarchical approach was adopted.

The initial step was to identify the top six to ten high-volume specialties, covering around 90 per cent of patient activity for subsequent more-detailed analyses. While in practice this may not be where the greatest error lies and we know there are likely to be particular



issues in the lower-volume specialties, any errors identified in the higher-volume specialties are likely to have greater materiality. This analysis is also useful for identifying trends in volumes of activity over time for each specialty at a provider.

High-volume specialties were analysed against the identified, benchmarking the acute provider against its peer group.

Within each specialty, for indicators showing variation, further analysis is performed by HRG chapter, admission method and in some cases individual HRG.

The following measures were typically examined for each acute trust:

FCE count

Simple count of episodes within the data cohort.

FCE beddays

The total length of stay (EPIDUR) for the episodes within the observation.

Mean FCE beddays

The total number of observed episode beddays divided by the number of episodes.

CLASSPAT day case FCE count

A count of episodes that are marked as being day cases by the CLASSPAT field.

INTMANIG day case FCE count

A count of episodes that are marked as being day cases by the INTMANIG field.

Day case FCE Ratio – CLASSPAT

The ratio of CLASSPAT day cases to non-day cases within the cohort.

Day case FCE ratio – INTMANIG

The ratio of INTMANIG day cases to non-day cases within the cohort.

Diagnosis count

The total number of diagnoses (including primary) within the cohort.

Poor diagnosis count

The total number of diagnoses (including primary) within the cohort that end in an 8 (other specified) or 9 (unspecified).

Poor diagnosis ratio

The ratio of poor diagnoses to total diagnoses within the observed records.

Average diagnoses per FCE

The number of diagnoses, divided by the number of episodes. This figure can serve as a rough guide as to the completeness of coding being performed when compared against peers with a similar casemix.

Average poor diagnosis per FCE

The total number of poor diagnoses (including primary) within the cohort that end in an 8 (other specified) or 9 (unspecified).

Complication HRG FCE count

A count of every episode within the cohort that has been assigned to a 'with complication' (w cc) HRG.

Complication HRG FCE ratio

The proportion of complication HRGs to the overall number of episodes within the cohort. Upon casemix adjusted data, a particularly high or low value may indicate issues with gaming or coding completeness respectively.

Complicating diagnosis rate

Defined as the complication HRG count divided by the product of the average number of diagnoses and the number of episodes within the observed records. Coupled with the average diagnoses per episode indicator, it can give a feel for whether the number of complication codes being hit is high or low for the relative level of coding completeness. A low number of diagnoses per episode, but a high rate of complicating diagnoses may indicate HRG optimisation, or coding from complication oriented pick-lists. Conversely, a low rate of complicating diagnoses may indicate poor non-specific coding.

Complex elderly HRG FCE count

A count of every episode within the cohort that has been assigned to a *99 HRG.

Complex elderly HRG FCE ratio

The proportion of complex elderly HRGs to the overall number of episodes within the cohort. Upon casemix adjusted data, a particularly high or low value may indicate issues with gaming or coding completeness.

Medical FCE count

The number of episodes that have been assigned on the basis of diagnoses (as opposed to those assigned by a procedure).

Spell count

A count of the number of spells within the cohort.

Spell beddays

The total length of stay for the spells within the observation.

Mean spell beddays

The total number of observed spell beddays divided by the number of spells. Can be used as a crude indicator of whether patients are being discharged comparatively early or late.

Episode to spell ratio

A ratio of episodes to spells within the cohort. A high value may indicate simple episode inflation.

Short stay spell count

A count of spells that have zero spell beddays.

Short stay spell ratio

The ratio of spells with zero beddays to other spells within the observation. A high short stay spell ratio suggests that patients are being admitted that may well have been treated as outpatients. This measure can be of particular interest when the observation consists of emergency patients.

Short stay outpatient tariff procedure spell count

A count of episodes that have been grouped on the basis of procedures that have outpatient tariffs within a spell that has a zero length of stay.

Short stay outpatient tariff procedure spell ratio

The ratio of short stay spells with outpatient procedures of the spells within the data cohort. A high value here may indicate over-admission of potential outpatients in order to attract the higher inpatient tariff.

Medical to surgical within spell ratio

The ratio of the number of diagnostically driven episodes to procedurally driven episodes within each spell in the data cohort. This may be useful as an additional indicator as to whether deliberate episode inflation is taking place.

Appendix 2: Clinical coding audit process

The clinical coding audits followed the latest version of the national standard clinical coding audit methodology, NHS Connecting for Health Clinical Coding Audit Methodology (published April 2006), which involves a review of the trust's coding against the case notes. The full methodology is available from Connecting for Health.

The audit objectives were defined as follows:

1. To assess the accuracy of clinical coding and its impact on casemix.
2. To identify the causes of coding errors.
3. To assess adherence to national standards pertaining to clinical coding and data definitions.

A pre-audit letter and questionnaire were sent to the acute trust a minimum of two weeks in advance of the audit. This included details of the population from which the sample case notes were to be drawn.

Trusts were requested to select episodes for audit from a randomly generated picking list by the hospital's PAS system. The coding auditors audited the case notes against the codes recorded in this print-out.

The coding auditors produced an audit report for the trust, focusing entirely on data quality issues. A draft report was agreed with the trust before being finalised, to ensure there was agreement over the errors documented, the conclusions and recommendations. The findings were then translated into a report for the host PCT, which involved applying the tariff, drawing conclusions about the impact of any error on payments, and a basic check on the monitoring arrangements in place locally to inform recommendations for PCTs.

Appendix 3: Coding audit results by trust

Trust 1

From a sample of 199 episodes from the period 1 October 2005 and 1 December 2005, 2 were deemed unsafe to audit. Therefore, the results presented are based on 197 episodes.

The accuracy rate overall was 86 per cent of primary diagnoses but only 73.5 per cent of secondary diagnoses correct. Within procedures 96.5 per cent of primary procedures and 93 per cent of secondary procedures were correct. The number of episodes with errors which would have changed the HRG was 20, representing 10 per cent of the total. **Table 6** provides a breakdown of the results by specialty.

Table 6

Percentage breakdown of error levels by specialty

	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	86	73.5	96.5	93	10
General Medicine	79	77	95	87.5	12
Urology	92	59	98	100	14
Paediatrics	98	88	100 (only 1 episode)	N/a	2

General Medicine

General Medicine was targeted due to:

- higher than average diagnoses per episode;
- a significantly higher proportion of HRGs with complications;

- comparative volatility between quarters for complex elderly cases;
- a substantially higher rate of short stay emergency patients; and
- a significantly higher FCE to spell ratio (possibility of certain episode inflation taking place).

Ninety-eight episodes were audited in total (two episodes were deemed unsafe to audit).

Coding error impacted on HRG assignment in 12 cases, all due to errors in primary and secondary diagnosis. Eighty-two per cent of the errors in primary diagnosis were due to coder-related errors, the remaining 18 per cent due to problems with the documentation. The majority of secondary diagnosis errors related to codes being admitted (33 out of 38 instances), but only 1 of these errors impacted on the HRG.

There were ten instances of irrelevant coding identified in this specialty, five of which involved inappropriate use of the living alone code. This code should only be used when it is clear that this is the reason for an extended length of stay, and tends to change the HRG assignment. In this case, inappropriate use of the code moved two patients to a more complex HRG, resulting in the trust being overpaid £2,147.

Procedure coding had a high level of accuracy. Out of 37 primary procedures only 2 were coded incorrectly.

Urology

Urology was targeted due to:

- a significantly higher proportion of HRGs with complications.

Fifty-one episodes were audited in total.

Coding error impacted on HRG assignment in seven cases, all due to errors in primary and secondary diagnosis. Most of these errors were coder error.

The majority of errors in secondary diagnosis arose from omitted codes (21 errors in total). There is little guidance available nationally on the treatment of comorbidities, which causes difficulty in this area. However, there was also a lack of consistency within the trust.

There were 48 episodes with procedures and only 1 error was found, relating to the primary procedure.

Paediatrics

Paediatrics was targeted due to:

- a significantly higher proportion of HRGs with complications; and
- a high rate of emergency admissions.

Fifty episodes were audited in total.

Coding accuracy was very good in paediatrics, with only one primary diagnosis error and five secondary diagnosis errors (all related to missed diagnoses). There was only one episode involving a procedure in this specialty and that was correctly coded. There was only one episode where the coding error led to incorrect HRG assignment.

Trust 2

From a sample of 202 episodes from the period 1 October 2005 and 1 December 2005, 2 were deemed unsafe to audit. Therefore, the results presented are based on 200 episodes.

The accuracy rate overall was 80.5 per cent of primary diagnoses and 78 per cent of secondary diagnoses correct. Within procedures 91 per cent of primary and 94 per cent of secondary procedures were correct. The number of episodes with errors which would have changed the HRG was 27, representing 13.5 per cent of the total. **Table 7** provides a breakdown of the results by specialty.

Table 7
Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	80.5	78	91	94	13.5
General Medicine	85	79	83	*	19
Obstetrics	78	75	97	94	0
Geriatric medicine	78	78	*	*	26

* Numbers are too small to have relevance in percentages

General Medicine

General Medicine was targeted due to:

- a significantly lower average length of stay when compared to its peers;
- a sharp increase in the proportion of day cases since January 2004;
- a sharp increase in short stay admissions (defined by a length of stay of zero days) since September 2006; and
- a low FCE to spell ratio (which could be due to good practice rather than coding issues).

Seventy-five episodes were audited in total.

The majority of errors in General Medicine which impacted on the HRG (14 episodes in total) were to due to errors relating to primary diagnosis. The majority resulted in underpayment to the trust. Thirty-six per cent of these errors were due to non-coder errors, relating to inconsistent information provided or information not being available at the time. The audit report also identifies some training issues for the coders.

With respect to secondary diagnosis errors, the majority were coder errors where a relevant secondary diagnosis had been omitted, in some cases affecting the HRG.

There was also evidence of overcoding – assigning a code that is considered not relevant to an episode – which impacted on HRG assignment. Under the current audit methodology, this is not counted in the total of errors as only information compared to the auditor codes is counted.

There were very few procedures in the audit sample, due to the nature of the specialty. The errors were minor and did not impact on HRG assignment.

Obstetrics

Obstetrics was targeted due to:

- a sharp increase in the proportion of day cases from the start of 2004/05 onwards; and
- relatedly, a sharp increase in short stay admissions (defined by a length of stay of zero days) in 2004/05.

Seventy-five episodes were audited in total.

Although the error rate in obstetrics appears high for diagnoses, none of the errors impacted on HRG assignment. Half of the errors were non-coder errors relating to how and when the information has to be coded; the other half were coding training issues. Coding for obstetrics is devolved to ward level, and normal deliveries are coded from a ward book, which often provides incomplete clinical information.

Geriatric Medicine

Geriatric Medicine was targeted due to:

- a significantly lower average length of stay for Geriatric Medicine;
- a higher rate of complex elderly cases; and
- a low FCE to spell ratio (which could be due to good practice rather than coding issues).

Fifty episodes were audited in total.

In 13 cases, coding error impacted on the HRG assignment, the vast majority of which were due to errors in the primary diagnosis. The cases notes were found to be poorly ordered and unclear.

The auditor reports a significant number of irrelevant codes entered as secondary diagnoses (21). Notably the use of the living alone code Z60.2 was evidenced seven times with no clear indication in the case notes that this had extended the length of stay of the patient.

Again, being a medical specialty, there were very few procedures to code, but there was a high level of accuracy and the few errors identified had no impact on HRG assignment.

Trust 3

From a sample of 200 episodes from the period 1 October 2005 and 1 December 2005, 3 were deemed unsafe to audit. Therefore, the results presented are based on 197 episodes.

The accuracy rate overall was 68 per cent of primary diagnoses and 50 per cent of secondary diagnoses correct. 66 per cent of primary procedures were correct. The number of episodes with errors which would have changed the HRG was 22, representing 11.2 per cent of the total. **Table 8** provides a breakdown of the results by specialty.

Table 8
Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	68	50	66	56	11.2
Paediatrics	89	75	20	33	5.1
Orthopaedics	27	34	80	92	12.25
Paediatric Surgery	68	31	47	64	22.0

There were a number of reasons for the relatively low accuracy rates at the time of the sample; insufficient coder training had an impact (there are no accredited coders), but it appeared that the flow and quality of information to support coding is a major contributory factor. For example, diagnoses were generally coded from flimsy discharge summaries for diagnoses, often incomplete and/or illegible, and procedures were being coded from theatre sheets. Coders also had difficulty accessing case notes.

During 2006, the trust has implemented a number of improvements in the structure of, and processes in, clinical coding. These include a review of the size of and skills within the coding team, requiring that full case notes are available for coding, and implementing a clinical coding action plan.

Paediatrics

Paediatrics was targeted due to:

- a high short stay rate as compared to peers; and
- a low complicating diagnosis rate.

One hundred episodes were audited in total.

The number of errors in Paediatrics which impacted on the HRG were relatively low (five in total), despite poor accuracy in procedure coding. The errors which changed the HRG assignment were a combination of incorrect or incorrectly sequenced diagnoses and omission of primary procedures. All five errors resulted in overpayment to the trust, amounting to £1,922 in total.

Orthopaedics

Orthopaedics was targeted due to:

- a low day case rate;
- a low medical/surgical ratio.

Fifty episodes were audited in total.

The error rate for primary and secondary diagnosis coding is notably high for Orthopaedics. However, there were only six errors which led to incorrect HRG assignment, all of which were due to error in the primary diagnosis. All of the errors except one resulted in a higher tariff than should have been paid, amounting to £3,437 overall.

Of the total errors identified, 30 per cent were non-coder errors associated with insufficient and poor-quality clinical information.

Paediatric Surgery

Paediatric Surgery was targeted due to:

- a higher than average short stay rate.

Fifty episodes were audited in total.

Both diagnosis and procedure coding in this specialty were poor, resulting in incorrect HRG assignment in 11 cases (22 per cent). Again, the majority of errors that changed the HRG were due to errors in primary diagnosis, although secondary diagnosis and omission of procedures were also a factor. Paediatric Surgery had very high levels of non-coder error (67 per cent), reflecting incomplete or poor-quality clinical information available at the time of coding. This is likely to reflect the practice of coding procedures from theatre sheets.

Trust 4

From a sample of 187 episodes from 1 October 2005 and 1 December 2005, 13 were deemed unsafe to audit, all of which were Nephrology cases. This is a high level of incomplete documentation and this may highlight other data quality issues. Therefore, the results presented are based on 187 episodes.

The accuracy rate overall was 86 per cent of both primary and secondary diagnosis codes, with 84 per cent of primary procedure codes correct but only 80 per cent of secondary procedure codes. The number of episodes with errors that would have changed the HRG was 16, representing 9 per cent of the total; the majority of which were either primary or secondary diagnosis. Eleven errors were identified where coding had been applied without full information being available. **Table 9** provides a breakdown of the results by specialty.

Table 9

Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	86	86	84	80	9
General Medicine	94	94	100	N/a	3
Gynaecology	86	68	90	75	4
Nephrology	72	66	57	0	30

General Medicine

General Medicine was targeted due to:

- significantly higher than average lengths of stay when compared to similar trusts.

One hundred episodes were audited in total.

The coding in General Medicine was of a good standard. The accuracy rate overall was 94 per cent in both primary and secondary diagnosis, with 100 per cent of primary procedures accurately coded. Coding error impacted on the HRG assignment of three episodes. Two of these were coder-related error and were as a result of information being unavailable to coders.

Gynaecology

Gynaecology was targeted due to:

- a significant increase in the number of short stay elective patients with procedures accruing specific outpatients tariffs.

Fifty episodes were audited in total.

Coding error impacted on HRG assignment in two cases, one in primary diagnosis and one in secondary procedure. The accuracy rate overall was better in the primary diagnosis and procedures (86 per cent and 90 per cent respectively) than in secondary diagnosis and procedures (68 per cent and 75 per cent respectively).

Nephrology

Nephrology was targeted due to:

- a significantly lower average length of stay when compared to similar trusts;
- a high day case to non-day case ratio and high proportion of short stay elective admission; and
- a sharp increase in short stay emergency admissions; higher than average rate of emergency admission with low length of stay.

Thirty-seven episodes were audited in total (13 were deemed unsafe to audit).

The accuracy rate overall was poor, with 76 per cent of primary and 66 per cent of secondary diagnosis coded accurately; of the primary procedures 57 per cent were coded accurately, with none of the secondary procedures (seven episodes) coded accurately. A total of 61 errors were found overall. The HRG assignment was changed in 305 episodes (11 cases).

Nine of the episodes had been coded without the case notes being available which suggests there is a problem in getting information to the coders when they require it.

Trust 5

From a sample of 200 episodes from the period 1 July to 30 September 2005, 1 was deemed unsafe to audit. Therefore, the results presented are based on 197 episodes.

The accuracy rate overall was 94.4 per cent in primary and 94.1 per cent in secondary diagnosis codes; of the primary procedures 91 per cent were coded accurately with all secondary procedures being accurately coded. The number of episodes with errors that would have changed the HRG was 13, representing 6.5 per cent of the total.

Of the 13 errors identified, 23 per cent were the result of poor documentation. Of the remaining 77 per cent attributable to coder-related error, 31 per cent were due to inappropriate use of 'operation cancelled' codes. **Table 10** below provides a breakdown by specialty.

Table 10

Percentage breakdown of error levels by specialty

Speciality	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
General Medicine	94	Unknown	96	100	10
Paediatrics	96	Unknown	79	100	6
ENT	98	Unknown	93	100	0
Total (Actual)	199		55	37	6.5

General Medicine

General Medicine was targeted due to:

- a high rate of planned procedures not carried out.

The accuracy rate overall was 94 per cent in primary diagnosis, and there was a good standard of coding in procedures. There were ten episodes (10 per cent) where coding error impacted on the HRG assignment. Errors were fairly balanced between coder-related error and information error.

Most of the 47 instances of irrelevant coding identified were in General Medicine. The majority of these resulted from the incorrect use of a cancelled operation code which should only be used when a procedure has been cancelled due to lack of resources to undertake it.

Paediatrics

Paediatrics was targeted due to:

- a low number of elective admissions and high number of non-elective admissions.

The accuracy rate was 96 per cent in primary diagnosis, 79 per cent in primary procedures and 100 per cent in secondary procedures. The errors identified in Paediatrics that affected the HRG assignment were all in primary diagnosis; there were three (2 per cent) in total and they were all coder-related errors. There were no HRG errors in secondary diagnosis or in primary or secondary procedures.

Ear, Nose and Throat (ENT)

ENT was targeted due to:

- a high rate of day cases and significantly variation in diagnosis per episode.

The coding of ENT cases in the audit was very good, with no errors affecting the HRG assignment. Only one primary diagnosis was incorrect. The information provided for ENT admissions allowed for accurate information to be recoded and the audit found very few unnecessary or additional codes being used in this specialty.

In relation to the high day case rate of admissions, the audit suggested that this could be due to some outpatient attendances being reclassified as a daycase when a procedure is performed.

Trust 6

From a sample of 200 episodes from the period 1 October 2005 and 1 December 2005, 2 were deemed unsafe to audit. Therefore, the results presented are based on 197 episodes.

The accuracy rate overall was 86.9 per cent of primary diagnoses and 91.2 per cent of secondary diagnoses correct. Within procedures 97.2 per cent of primary procedures and 81.9 per cent of secondary procedures were correct. The number of episodes with errors that would have changed the HRG was seven, representing 3.5 per cent of the total. **Table 11** provides a breakdown of the results by specialty.

Table 11

Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	86.9	91.2	97.2	81.9	3.5
General Medicine	93	94	100	N/a	4.1
Urology	91.8	92.1	97.7	100	2
Obstetrics	77.3	85.9	95.7	75	4

General Medicine

General Medicine was targeted due to:

- a high level of poor diagnosis.

Seventy-four episodes were audited in total (one episode was deemed unsafe to audit).

Coding error impacted on HRG assignment in three cases, all due to errors in primary and secondary diagnosis. Of the total errors in primary diagnosis, 60 per cent were due to coder error, and 40 per cent to documentation issues. All errors identified in secondary diagnosis were due to coder error.

Procedure coding had a high level of accuracy. All 21 procedures were coded correctly.

Urology

- A high medical/surgical ratio of HRGs;
- a high level of HRGs with complications; and
- a high level of emergency admissions with a short stay.

Forty-nine episodes were audited in total (one episode was deemed unfit to audit).

The error identified in Urology which amended HRG assignment was due to an error in recording primary diagnosis. With one exception, all errors were in primary and secondary diagnosis, and all were as a result of coder error.

Obstetrics

- A high level of patients with a short stay.

Seventy-five episodes were audited in total.

Coding error impacted on HRG assignment in three cases, each due to coder-related errors in primary diagnosis. Seventy-six per cent of the errors were due to coder-related errors, the remainder due to problems with availability of information at the time of coding; this issue has now been addressed by the trust. Nearly all errors in secondary diagnosis were coder-related errors (88 per cent), relating to local coder practices which did not comply with trust practice; this has also been addressed.

Trust 7

From a sample of 199 episodes from the period 1 October 2005 and 1 December 2005, 9 were deemed unsafe to audit, and 1 was outside the scope of the audit. Therefore, the results presented are based on 189 episodes.

The vast majority of errors were in primary and secondary diagnosis, with the accuracy rate for primary diagnosis at 74.9 per cent, and secondary diagnosis at 81.5 per cent. Within procedures, 94.6 per cent of primary procedure coding was accurate and there were no errors in secondary procedures.

A major issue identified by the audit was the number of irrelevant codes for many of the episodes with a total of 282 irrelevant codes, to a large extent caused by a practice of creating a new FCE when patients are moved from one consultant to another within the same specialty (as opposed to normal practice of only doing this when patients move between specialties). Where two consultants share a ward, patients are transferred on a regular basis for no clinical reason. This has created complications in undertaking the coding audit, especially in relation to assessing comorbidities. This practice will certainly inflate the number of episodes and potentially the number of spells and is likely to impact on payments. **Table 12** provides a breakdown of the results by specialty.

Table 12
Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	74.9	81.5	94.6	100	23
General Medicine	68.5	79.5	100	100	31
Nephrology	78.8	86	88.2	100	10
Cardiology	87.2	85	92	100	12

General Medicine

General Medicine was targeted due to:

- a high day case rate in elective;
- high diagnosis, and high poor coding rate.

One hundred and seventeen episodes were audited in total (five were deemed as unsafe to audit).

In 36 cases (31 per cent) coding error impacted on the HRG assignment – all were primary and secondary diagnosis errors. While all but two were coder-related errors, many of them were created by the issue highlighted above relating to inappropriate

recording of FCEs. This is compounded where patients are recoded as having new FCEs when transferred from admissions wards, or transferred to discharge areas.

Coders were also not recognising 'geriatric falls' or palliative care as a legitimate diagnosis.

Nephrology

Nephrology was targeted due to:

- a low short stay rate;
- high diagnoses; and
- a high poor coding rate.

Thirty-nine episodes were audited in total (two were deemed to be unsafe to audit).

Coding error impacted on HRG assignment in four cases (10 per cent), all due to coder error in primary and secondary diagnosis. There were issues with the patient's main diagnosis being described in a number of ways, causing confusion as to how to code. Also, there were errors in the coding of hypertension. There were only two errors in procedure coding.

Cardiology

Cardiology was targeted due to:

- a high complicating diagnosis rate, and complex elderly rate in elective cases;
- a high emergency length of stay, high emergency rate, high FCE/spell ratio;
- a high emergency diagnosis, high poor coding rate, high complication rate.

Thirty-three episodes were audited in total (two were deemed to be unsafe to audit).

Coding error impacted on HRG assignment in four cases (12 per cent), with the majority of the errors being coder-related error in secondary diagnosis. There were no errors in procedure coding.

Again, in Cardiology there were issues with the way information was being interpreted in relation to hypertension, with inappropriate coding to this diagnosis. Patients were also being kept on the PAS system when in fact being transferred to another trust for angiograms.

Trust 8

From a sample of 204 episodes from the period 1 October 2005 and 1 December 2005, 18 were deemed unsafe to audit. Therefore, the results are based on 186 episodes.

The accuracy rate overall was 71.5 per cent in primary diagnosis and 74.3 per cent in secondary diagnosis. The accuracy rate for primary and secondary procedures was 76.2 per cent and 87.5 per cent respectively. Coding error impacted on the HRG assignment in 52 episodes, which is 28 per cent of the sample.

The trust has a PAS system which restricts coding to a maximum of seven diagnosis and seven procedures, which had an adverse affect on the results of the audit, although a new information system is planned which will remove this issue. A large number of case notes were found to be unsafe to audit (9 per cent) which may suggest further data quality issues; many case notes were in a poor state, both in terms of misfiling and physical condition. There are some issues relating to availability of adequate information to the coders, including problems accessing case notes from out-lying hospitals, and timeliness of discharge summaries, especially in General Medicine. **Table 13** provides a breakdown of the results by specialty.

Table 13

Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	71.5	74.3	76.2	87.5	28
General Medicine	73	64.7	80	100	27
Geriatric Medicine	69.8	80.9	72.7	75	29

General Medicine

General Medicine was targeted due to:

- a high rate of emergency admissions.

There were 104 episodes examined of which 4 were deemed unsafe to audit. Of the 27 HRG errors, the majority were in primary diagnosis; 4 were non-coder errors related to documentation issues, 8 errors were at the 3 or 4 character level, 9 errors occurred where the diagnosis was incorrectly sequenced and 6 primary diagnoses were omitted.

Geriatric Medicine

Geriatric Medicine was targeted due to:

- a higher than average length of stay;
- a higher FCE/spell ratio; and
- a lower rate of short stay non-day case patients.

Of the 100 episodes examined, 14 (14 per cent) were deemed unsafe to audit (raising data quality concerns), leaving 86 to be audited. There were 25 episodes (29 per cent per cent error rate) where the HRG was amended as a result of incorrect coding. The majority of HRG errors were in the primary diagnosis.

Trust 9

From a sample of 201 episodes from the period 1 October and 1 December 2005, 1 was deemed to be unsafe to audit. Therefore, the results presented are based on 200 episodes. The accuracy rate overall was 75.5 per cent primary diagnosis and 64.2 per cent secondary diagnoses correct. 75.9 per cent of primary and 79.5 per cent of secondary procedures were correct. There were 21 episodes (10.5 per cent) with HRG errors.

The largest number of errors were as a result of omission of secondary codes, inappropriate sequencing of primary diagnosis, and irrelevant codes, which would suggest a need to enhance formal training for clinical coders. **Table 14** provides a breakdown of the results by specialty.

Table 14
Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	75.5	64.2	75.9	79.5	10.5
General Medicine	72	63.2	94.6	100	17
Cardiology	78	67.4	55.9	100	0
Gynaecology	80	58.2	75.6	77.5	8

General Medicine
General was targeted due to

- a lower than average length of stay in particular for emergency admissions; and
- higher than average short stay non-day case patients mainly for emergency admissions.

In General Medicine there were 17 episodes where the HRG was amended as a result of incorrect coding; the total number of case notes reviewed was 100. Most of the errors were in the primary diagnosis, and related mainly to coder-related error.

Cardiology
Cardiology was targeted due to:

- a significantly higher length of stay; and
- a lower than average medical to surgical episodes ratio.

Although there were no HRG errors, there were 44 per cent of primary procedures which had been incorrectly coded. Coders encountered ambiguities in terminology used on coronary angiograms/catheter reports that made it difficult to determine which investigation the patient was undergoing.

Gynaecology

Gynaecology was targeted due to:

- a significantly above average rate of emergency admissions; and
- a significantly higher rate of short stay patients undergoing procedures for which there are specific outpatient tariffs.

In Gynaecology there were four episodes (8 per cent) where the HRG was amended as a result of incorrect coding; the total number of case notes reviewed was 50. The errors were spread between primary and secondary procedure and secondary diagnosis.

Trust 10

A sample of 204 episodes was taken from the period 1 October 2005 and 1 December 2005, all episodes were considered fit to audit.

The accuracy rate overall was 84.4 per cent of primary diagnoses and 77 per cent of secondary diagnoses correct. Of the procedure codes, 93.2 per cent of primary and 84.3 per cent were correct. There were 21, representing 10.2 per cent of the sample, where the errors would have changed the HRG assignment. **Table 15** provides a breakdown of the results by specialty. The three specialties originally identified for audit were General Medicine, Obstetrics and Nephrology, for the reasons outlined below. However, there had been insufficient FCEs in the period, and a sample of 'other' case notes was selected covering Nephrology, Cardiology and Gynaecology to make up the shortfall. Therefore, the audit assessed 204 episodes, 83 in General Medicine, 49 in Obstetrics, 17 in Gynaecology and 55 in 'other'.

Table 15
Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	84	77	93	84	10
General Medicine	86	77	93	84	13
Obstetrics	82	81	96	87	8
Gynaecology	88	82	87	88	0
Other	86	79	92	90	11

General Medicine
General Medicine was targeted due to:

- a significantly higher average length of stay for General Medicine when compared to other similar trusts;
- a substantially lower rate of short stay patients when compared to similar trusts; and
- a significantly higher proportion of HRGs with complications to the overall number of episodes within the cohort, when compared to similar trusts.

All the errors occurred in primary and secondary diagnosis in non-elective cases and were spread across a range of HRGs, in other words not concentrated in particular areas. The majority of the errors in primary diagnosis were coder errors, with the main cause being the coding of several FCEs to one code when in fact each should be coded separately. The majority of the secondary diagnosis errors related to codes being omitted. There were 24 instances of irrelevant secondary diagnosis but none of these errors impacted on the HRG.

Obstetrics

Obstetrics was targeted due to:

- high rates of outpatient tariff procedures can be seen in Obstetrics at the trust. In particular, there is a high rate of short stay elective admissions within HRG chapter M; and
- a low ratio of medical to surgical episodes within a spell.

There were four HRG errors; in two of the four cases the HRG assigned pre-audit was N12 (non-delivery maternity event) and there appear to be specific issues relating to the use of this code in a number of trusts.

Nephrology

Nephrology was targeted due to:

- a significantly higher average length of stay for Nephrology;
- Nephrology has a higher rate of emergency admissions; and
- the trust shows a substantially higher than expected rate of short stay patients when compared to similar trusts.

Other

As outlined above, it became necessary to select a varied sample to achieve numbers required. Within these, the errors occurred in primary diagnosis, involved non-elective cases and were spread across a wide range of HRGs. All the errors were coding errors. There were seven instances of an irrelevant secondary code being included. As in General Medicine, a cause of the high error rate was incorrect coding of several FCEs to one code when in fact each should be coded separately.

Trust 11

From a sample of 204 episodes from the period 1 October 2005 to 31 December 2005, 4 were deemed unsafe to audit. Therefore, the results presented are based on 200 episodes.

The accuracy rate overall was 73 per cent in primary diagnosis but only 53 per cent of secondary diagnosis correct. The accuracy rate was higher for procedures, with 88.5 per cent and 75.6 per cent respectively for primary and secondary procedures.

Coding error impacted on the HRG assignment in 26 episodes. The majority (69 per cent) were as a result of coder-related primary diagnosis errors, with 15 per cent as a result of coder-related secondary diagnosis errors, and the remainder are procedure coding errors.

The audit highlighted a particular issue with a lack of clinical involvement, both in terms of support to the coding team and a lack of clarity in clinical case notes as to diagnosis or synopsis to assist coders. **Table 16** provides a breakdown of the results by specialty.

Table 16
Percentage breakdown of error levels by specialty

Total	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	73	53	88.5	75.6	13
General Medicine	58.7	54.3	82.1	71.4	24
Cardiology	74.7	46.7	93.1	98.4	5.3
Medical Oncology	86	56.3	90.2	57.1	8

General Medicine

General Medicine was targeted due to:

- a high rate of short stay patients where outpatient tariffs exist;
- a low rate of emergency admissions; and
- a high rate of emergency short non-day case patients.

In General Medicine there were 18 episodes where there was an HRG error, from a total of 75 episodes. There were no errors in procedures although this is not surprising given that this is a medical (as opposed to a surgical) specialty. All the errors occurred in primary and secondary diagnosis and, with one exception, occurred in non-elective cases.

Of the 18 errors in General Medicine only 1 occurred as a result of poor documentation, the others were the consequence of coder error. The error rate in General Medicine for primary and secondary diagnosis is high even where the HRG was not affected.

Cardiology

Cardiology was targeted due to:

- a high rate of emergency short stay non-day case patients.

Overall coding in Cardiology was of a reasonable standard. There were four HRG errors identified in cardiology; these were all in primary and secondary diagnosis. Again there were no errors in either primary or secondary procedures. The errors were all coder errors and there was no identifiable pattern to them.

Medical Oncology

Medical Oncology was targeted due to:

- a high number of elective day cases;
- a high day case to non-day case rate; and
- a low rate of short non-day case patients.

There were four HRG errors, which were a mixture of diagnosis and procedure errors.

Trust 12

From a sample of 201 episodes from the period 1 October 2005 and 1 December 2005, 1 was deemed unsafe to audit. Therefore, the results are based on 200 episodes.

The accuracy rate overall was 94.9 per cent in primary diagnosis and 94.1 per cent in secondary diagnosis. The accuracy rate for primary and secondary procedures was 92.5 per cent and 86.7 per cent respectively. There was an HRG error in 21 episodes, which is 10.5 per cent of the sample.

The majority of HRG errors were due to errors in the primary diagnosis. There was evidence of some overcoding, which may be due to clinical coders also being trained in clinical audit giving them a wider awareness of comorbidities. The FCEs to spell ratio may be high due to patients being admitted to admission wards before being transferred to a medical ward at a later date. This has the effect of increasing the volume of coding required as the patient is coded at each transfer.

Table 17
Percentage breakdown of error levels by specialty

Speciality	% Primary diagnoses correct	% Secondary diagnoses correct	% Primary procedures correct	% Secondary procedures correct	% of episodes changing HRG
Total	94.9	94.1	92.5	86.7	4.6
General Medicine	93.4	93.4	88.2	81.8	5.2
Gynaecology	95.9	96.1	91.1	88.5	3.5
Urology	96.6	93.3	100	84.6	0
Cardiology	92.6	90.0	90.5	100	3.7

General Medicine
General Medicine was targeted due to:

- a higher length of stay due to admissions whilst awaiting a specialist transfer and then because of discharge back to Weston; and
- a higher average number of diagnoses per FCE.

In General Medicine there were 5 HRG errors from a total of 91 episodes. Of these, 4 were coder-related errors, and 1 was due to the information not being available at the time of coding.

Gynaecology
Gynaecology was targeted due to:

- a substantially higher rate of short stay non-day case patients when compared to similar trusts.

In Gynaecology there were 3 HRG errors from a total of 49 episodes. The errors were spread between procedures and diagnoses.

Urology

Urology was targeted due to:

- a significantly higher rate of short stay patients undergoing procedures for which there are specific outpatient tariffs; and
- a higher volume of unspecified diagnoses in comparison to similar trusts.

In Urology there were no episodes where the HRG was amended as a result of incorrect coding; the total number of case notes reviewed was 29.

Most of the errors in Urology were in the secondary diagnosis. Of the five HRG errors, two were errors relating to a lack of information and the remaining three were coder-related errors.

Cardiology

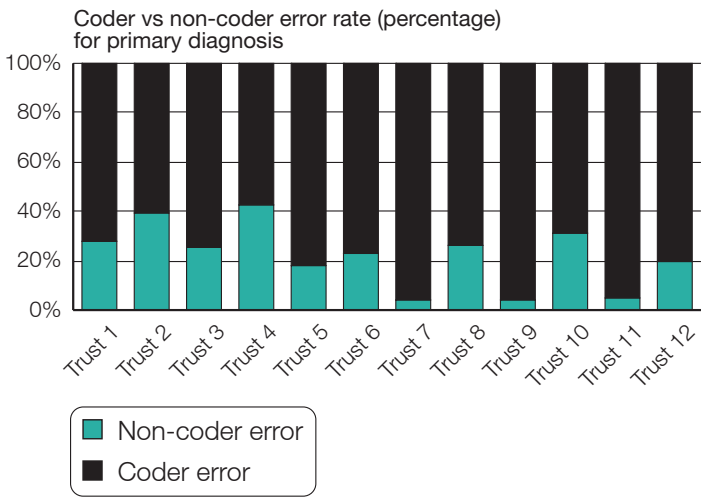
Cardiology was targeted due to:

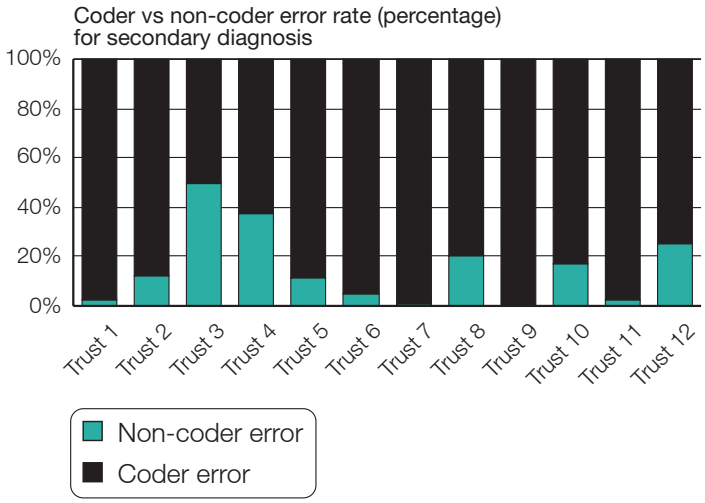
- a higher rate of emergency admissions; and
- a higher volume of unspecified diagnoses when compared to similar trusts.

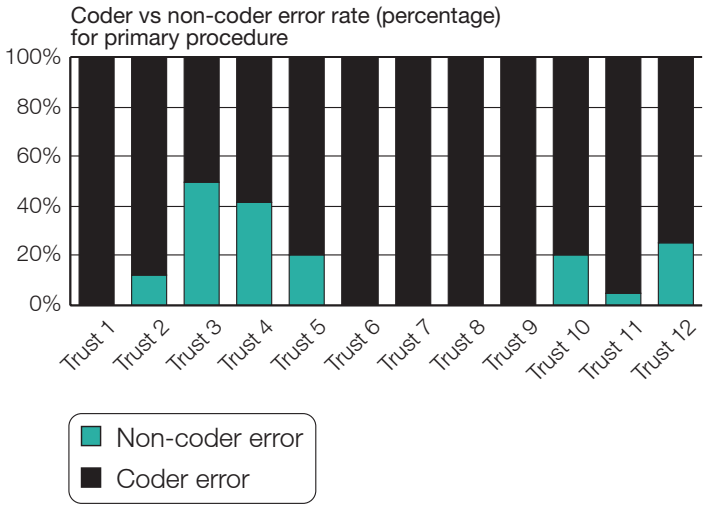
In Cardiology there was 1 HRG error from a total of 27 episodes; this was due to information not being available at the time of coding.

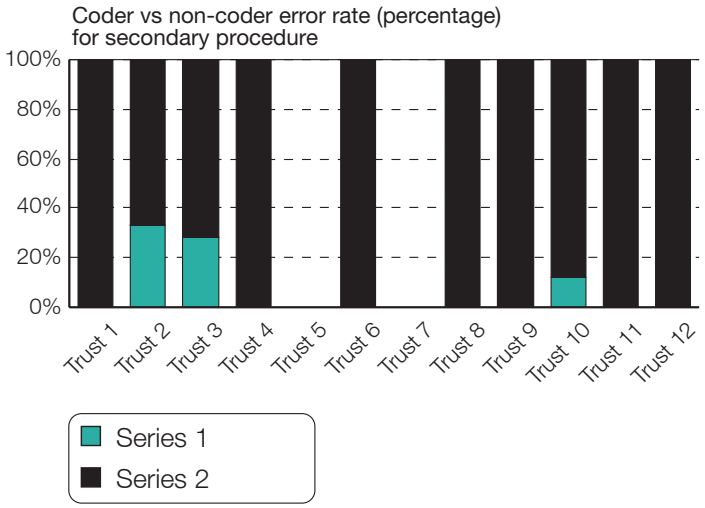
Although not affecting the HRG, there were 11 instances of irrelevant secondary diagnosis coding in Cardiology which implies a requirement for further training.

Appendix 4: Breakdown of coder versus non-coder error









Appendix 5: Data definitions

This Appendix sets out the specific problems with data definitions that were identified during the PbR assurance pilots and which the Audit Commission is recommending are addressed as a priority. In particular, it is important that the definitions are operationally viable, fully incorporated in the Data Dictionary, and that the Data Dictionary and reference cost definitions are aligned. The priority is to achieve consistency in how activity is recorded. We recommend that national guidance, clarifying the position on the issues set out below, is published by the PbR team, in conjunction with the Information Centre and Connecting for Health, where appropriate.

1. Admitted patient care

Admissions

The biggest single issue with respect to admitted patient care is the definition of an 'admission'. This underpins many of the specific problems highlighted in the pilots, particularly with respect to the outpatient and day case interface (what is a day case?).

For example:

- How should activity be recorded for Medical and Surgical Assessment Units? These are treated as admitted patients in some trusts, outpatient contacts in others.
- How should activity be recorded for observation beds in Accident and Emergency departments? Some trusts classify this as 'admitted patient care', others only count it after a 'decision to admit' is taken, some do not charge at all.[l]
- In particular, how should HRG N12 'antenatal admissions not related to delivery event' activity be billed at Day Attendance Units? There is evidence from the pilots that some of this activity would be more appropriately recorded as an admission rather than a day case.

l Paragraph 40 of the Reference Costs Changes document (April 2006) appears to suggest that all activity on admission/pre-admission, assessment units and observation ward admissions that go on to be admitted to the main body of the hospital 'should be counted as FCEs allocated to an appropriate (V prefix) version 3.5 HRGs and costed accordingly'. This will need to be incorporated in the Data Dictionary. Does the phrase 'that go on to be admitted' just relate to observation ward admissions or cover admission/pre-admission and assessment units?

- How should activity in maternity admission units be counted?

Rehabilitation

How should activity for rehabilitation and intermediate care be recorded and charged? Should rehabilitation activity be charged separately or is it in the tariff? Organisations interpret this differently, especially when they have different arrangements for the provision of rehabilitation services (whether it is provided in-house of externally. Therefore, the same care can result in a new spell or a longer length of stay, depending on the trust. PCTs in the pilot have commented that trim points seemed so large for rehab-related HRGs that rehabilitation days must have been included in reference costs.

Neonatal

How should neonatal care levels, particularly with regard to intensive care, be counted and charged for PbR and reference cost purposes? The costs of neonatal care activity carried out on wards tend to be excluded from reference costs, which results in a lower tariff, suggesting this care should be charged separately.

2. Outpatient/day case interface

Closely related to the definition of an admission:

- When should a procedure be classified as an outpatient procedure and when should it be classified as a day case? Illustrative examples highlighted include endoscopy, colposcopy, various gynaecology procedures, minor operations in oral surgery/dentistry and orthodontic procedures and minor dermatology procedures.
- Can outpatient procedures which are covered by the PbR tariff be performed and charged for as day cases?
- Are there plans to alter the definition of an elective day case from discharged before midnight to discharged within 24 hours of admission? (in other words to cater for patients who are admitted the evening before for procedures very early in the morning the following day).
- When can patients attending for diagnostics be counted as day cases?
- When should chemotherapy be treated as a day case or outpatient? This has been a consistent theme highlighted in the pilots with examples often quoted of patients being treated for half an hour being counted as day cases.

Commissioners, in particular, have commented that it would be useful to publish (and keep under review) lists of procedures, which clinical practice would predominantly expect to be done on a day case basis and a list of procedures, which clinical practice would predominantly expect to be done on an outpatient basis.

3. Outpatients

The main ambiguity with respect to outpatient data relates to the definition of 'consultant led' and 'consultant responsible' clinics, and what activity can be counted under PbR.

- What are the definitions of 'consultant led' and 'consultant responsible' clinics and which are excluded from PbR?
- How should pre-assessment units be treated? Are these a follow-up to the initial outpatient contact or should they be included in the first contact?
- When is Allied Health Professional (AHP) and Nursing Services clinic activity chargeable? When is an AHP 'consultation' an outpatient attendance and when is it a contact?[l]
- How should national screening programme work by AHPs be recorded? This is currently charged outside of tariff. Are there any circumstances in which it can be legitimately charged for at a higher outpatient specialty rate?
- When is a nursing 'consultation' an outpatient attendance or a contact? What is a 'nurse-run' clinic? What criteria need to be met for a clinic to be classified as 'nurse-run'? Can clinics still be nurse-run if a consultant retains overall responsibility for the patient's care? Are all midwife clinics automatically nurse-run or does this depend on whether the care is midwife led?
- When do nursing 'consultations' form part of a composite consultant outpatient charge and when are they separately chargeable when care is provided as part of an outpatient episode?
- If a clinical nurse specialist sees patients at a consultant clinic (at which the consultant is present and the consultant remains in charge of the case) in lieu of a junior doctor (say, for example, in a gastroenterology clinic) can this be classified and billed as an outpatient follow-up attendance in that particular specialty?

l Several queries have been received about the interpretation of the reference costs guidance regarding the presence of a doctor possibly being a deciding factor. For example, if a patient sees an AHP at the request of a consultant at a 'clinic session' but the patient does not see the consultant or any other medical member of the consultant's team, can this be classified as an independent outpatient attendance and charged for separately?

- What criteria need to be satisfied for a nurse clinic or AHP clinic provided by hospital trusts in the community to be classified as a hospital 'outreach clinic'? Are these criteria different from consultant clinics?
- When can trusts charge for an outpatient attendance taking place during inpatient episodes? The Data Dictionary states that patients, who are seen by a consultant from a different main specialty on their ward during their inpatient episode (hospital provider), should be classified as outpatient attendances where no transfer of responsibility for the care has taken place. Examples include:-
  - pre-operative visits by an anaesthetist (often on the eve of the operation;
  - emergency medical patients may be seen by members of various medical sub-specialty firms as part of the assessment and diagnosis process;
  - visits by a consultant or member of his firm to assess a patient, which leads to a decision to transfer (in other words the visit is prior to the actual transfer date); and
  - surgical teams from more than one specialty are present in the theatre for an operation, does this count as an outpatient attendance?
- How should one stop clinics be recorded and paid for. For example, if the consultant sees the patient a second time during the visit to discuss test results, thus avoiding the need for a second visit to the hospital, does this count as a follow-up attendance?
- What is the correct recording practice for joint clinics – when two consultants see a patient at the same clinic (for example, a surgeon and an oncologist)? Do these count as two separate outpatient attendances for charging purposes?
- When do telephone and other telemedicine contacts count as outpatient activity and when do they not?
- Cystic fibrosis – there is a need to establish clearer definitions and recording methodologies for outpatient attendances involving such patients.
- What is a treatment attendance for leukaemia/Oncology services – an individual appointment or a treatment cycle?
- How should outpatient activity be recorded for patients with dual diagnoses, which include learning disabilities, for PbR charging purposes? If a patient has learning disabilities does this mean that treatment is outside of tariff?

In addition, ward attenders are considered to be outpatients under PbR but ward attenders still exist as a separate category in the Data Dictionary. When is a 'ward attendance' covered by an inpatient spell charge and when is it separately chargeable? For example, if a patient pays a pre-operative visit to a ward and sees a nurse and this is recorded as a ward attendance, does this visit now qualify for an extra payment?

5. Reference costs

The following issues with reference cost guidance have been specifically identified:

- Reference costs have introduced a new category of 'direct access' 'outpatient attendances' for certain types of AHP activity (occupational therapy, physiotherapy, speech and language therapy and dietetics). This categorisation is to be used in instances where referrals for treatment carried out in a hospital setting have been made to a clinical/ other professional, including a GP or self-referral and where the patient attends a discreet therapy clinic solely for the purpose of receiving therapy treatment. The Data Dictionary still, however, continues to state these are contacts and does not recognise this category. How should this activity be recorded?
- Queries have also been raised about the significance placed on the term 'pre-bookable' in the reference costs guidance. For example, if a patient was given an appointment time to see an AHP or nurse (as distinct from the time being left open) at a consultant clinic session, does this constitute a separate AHP or nurse-run clinic running concurrently to the consultant clinic, which can be charged for separately?
- Paragraph 147 of reference costs appears to introduce a more flexible definition of outpatient services – 'Outpatient clinics have previously related to those clinics with pre-booked appointments for which a consultant is clinically responsible whether they are present at the clinic or not.' 'The above definition may not capture all types of appropriate activity, especially where changes in service provision and delivery mean that services are provided in a variety of ways and by a variety of clinical and medical staff. Thus, for 2006, where a patient sees a health care professional in a clinic setting, for which the patient has an appointment, and that patient receives healthcare treatment, including a clinical consultation, such activity can be counted as valid outpatient activity.' It would be helpful if the centre could clarify what this actually means in practice for reporting and charging purposes.
- The current HRG 3.5 structure means that patients are often assigned to the higher-priced HRG of a paired HRG on the basis of comorbidities/complications. PCTs had

believed that where, for example, known comorbidities exist, the higher HRG inpatient tariff was intended to cover the cost of bringing in expertise from other specialties, if required. (For example, if surgery was carried out on a diabetic patient, it covered the cost of bringing in the physician with a special interest in diabetics, or if a patient had a heart problem, the cardiologist). Do visits from an associate consultant now have to be paid for on top of the HRG inpatient tariff?

- Although not currently covered by PbR, reference costing requires the continued collection of data items which may eventually come under the tariff. Some of these have been removed from the Data Dictionary (for example, total contacts, post-natal visits and so on) following the discontinuation of central Korner returns and for which there are consequently no definitions. If these items are required, can the definitions and collection be reinstated? Comparative analysis by the strategic health authority's finance department suggests that there are considerable and unexpected variations in community nursing costs. It is unlikely this is due to service models and we suspect that different organisations are interpreting the guidance in different ways.
- The new reference costs collection also appears to have mandated the collection of non face-to-face nursing contacts. This item was not centrally collected under the former Korner returns. Examples of some of the problems this causes are highlighted below:
    - It is not clear how to record activity at group sessions and organisations are interpreting the requirement in different ways.
    - Some aspects of health visiting and other specialist nurse activity have been traditionally reported under Korner returns relating to maternity care or other structured programmes such as vaccination or immunisation programmes and excluded from the main activity return counts (in other words the Korner returns were specifically set up to avoid overlap between the different returns.) It is no longer clear if this reporting structure still applies to reference costs. Assuming it does, one cannot be certain that organisations are aware of and still acting on this guidance (in other words data may be being supplied on an inconsistent basis.)
    - It is no longer clear if different interventions under different programme budgets in the same visit count as multiple or single visits. It is also not clear if nursing tasks requiring the attendance of two or more staff count as single or multiple visits.

  - Likewise, previous guidance stated that if a joint health visit: midwife visit was carried out, it should only be recorded on one Korner return. It is not clear how this should be treated under reference costs.

- There is no clear distinction in the reference costs guidance between chiropody and podiatry services.
- The reference costs guidance also needs to be more specific about what constitutes an acceptable sampling timeframe to give a representative 12 months' activity or an informed 'clinical estimate' if organisations are not able to supply actual 12 months' figures. Some organisations are sampling simultaneously for both programme budget and reference cost return purposes and the advice needs to be the same for both collections.
- Implementation of PbR has shown that trusts are interpreting the reference costs guidance in different ways as to what is excluded from standard HRG inpatient or outpatient costs. In particular, the equipment and device definitions appear to have been a source of misunderstanding – for example, Ilizarov Frames, Statement Z and Continuous Positive Airway Pressure (CPAP)/Bilevel Positive Airway Pressure (BiPAP) replacements, spare parts and consumables. This guidance would benefit from a review.
- It is not always clear when Data Set Control Notices (DSCNs) are published whether definitional changes are being introduced for tariff purposes or whether they only apply to reference costs. It would be helpful if this could be clarified in the DSCN.

6. Other

The technical guidance is currently vague for non-contracted activity and organisations would welcome more specific guidance. For example, where critical care is provided as non-contracted activity, should the trust charge for it?

There are a number of grey areas in Patient Transport Services' (PTS) definitions. It is suggested that these are reviewed before PTS is split off from the main tariff as they are likely to become the source of local disputes.

In the light of recent changes to specialty coding, should the phrase 'main specialty' still be interpreted to mean main specialty or does it mean 'treatment function' speciality? Because of current levels of sub-specialisation, a changeover to treatment function could make a significant difference to recording and thus charging arrangements.

Specialist centres have flagged concerns that genetics is developing very fast as both a diagnostic and therapeutic application and this needs to be addressed in both definitional and PbR development work.

A significant amount of medical and other health professional time is spent providing advice to other NHS personnel on the management of individual patients without physically seeing or speaking to the patient. These personnel can be within their own trust, in neighbouring trusts or members of primary health care teams or community health services. Trusts do not receive any payment for these services at present.

There is a need to develop a better definition of Tier 4 Child Protection services before this becomes part of tariff.

Skill mix redesign is beginning to move responsibility for carrying out certain clinical tasks between different professionals and between qualified and non-qualified staff. This is likely to have implications for data capture and needs to be addressed in both definitional and PbR development work. The pace at which this is occurring is likely to be accelerated by delivery of the 18-week patient pathway.

Audit Commission
1st Floor, Millbank Tower,
Millbank, London SW1P 4HQ
Tel: 020 7828 1212 Fax: 020 7976 6187
Textphone (minicom): 020 7630 0421
www.audit-commission.gov.uk