
Article:
Tsopra, R, Peckham, D orcid.org/0000-0001-7723-1868, Beirne, P et al. (8 more authors) (2018) The impact of three discharge coding methods on the accuracy of diagnostic coding and hospital reimbursement for inpatient medical care. International Journal of Medical Informatics, 115. pp. 35-42. ISSN 1386-5056

https://doi.org/10.1016/j.ijmedinf.2018.03.015

TITLE PAGE

TITLE

The impact of three discharge coding methods on the accuracy of diagnostic coding and hospital reimbursement for inpatient medical care.

AUTHORS' NAMES

Rosy TSOPRA, Daniel PECKHAM, Paul BEIRNE, Kirsty RODGER, Matthew CALLISTER, Helen WHITE, Jean-Phillipe JAIS, Dipansu GHOSH, Paul WHITAKER, Ian J CLIFTON, Jeremy C WYATT

Rosy TSOPRA
Postdoctoral Fellow, MD, PhD
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
Leeds Institute of Health Sciences, 101 Clarendon Rd, Leeds LS2 9LJ, United Kingdom
LIMICS, INSERM, U1142, Université Paris 13, Sorbonne Paris Cité, F75006 Paris, France
AP-HP, Assistance Publique des Hôpitaux de Paris, Paris, France
rosy.tsopra@aphp.fr

Daniel PECKHAM
Professor of Respiratory Medicine
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
daniel.peckham@nhs.net

Paul BEIRNE
Consultant Respiratory Physician
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
p.beirne@nhs.net

Kirsty RODGER
Consultant Respiratory Physician
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
kirsty.rodger1@nhs.net

Matthew CALLISTER
Consultant Respiratory Physician
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
matthew.callister@nhs.net

Helen WHITE
Ph D Doctor
School of Clinical and Applied Science, Leeds Beckett University, Leeds LS1 3HE, United Kingdom
H.White@leedsbeckett.ac.uk

Jean-Philippe JAIS
Consultant Public Health
INSERM UMRS 1138 Team 22 ; Paris Descartes University, Sorbonne Paris Cité ;
Assistance Publique Hôpitaux de Paris, Necker Enfants Malades Hospital, Biostatistics Unit, Paris, France

Dipansu GHOSH
Consultant Respiratory Physician
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
dipansughosh@nhs.net

Paul WHITAKER
Consultant Respiratory Physician
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
p.whitaker@nhs.net

Ian J CLIFTON
Consultant Respiratory Physician
Leeds Centre for Respiratory Medicine, St James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom
i.clifton@nhs.net

Jeremy C WYATT
Professor of Digital Healthcare
Wessex Institute of Health & Research, Faculty of Medicine, University of Southampton, SO16 7NS, United Kingdom
j.c.wyatt@soton.ac.uk

CORRESPONDING AUTHOR

Dr Rosy TSOPRA

Postal address:

LIMICS

INSERM, U1142, F-75006 Paris, France
Université Paris 13, Sorbonne Paris Cité, UMR_S 1142, F93000 Bobigny, France
Sorbonne Universités, UPMC Université Paris 06, UMR_S 1142, F75006 Paris

e-mail: rosy.tsopra@aphp.fr

Telephone number: +33 1 48 38 73 34

NUMBER OF FIGURES: 1

NUMBER OF TABLES: 5

NUMBER OF SUPPLEMENTARY FILES FOR ONLINE PUBLICATION: 0

NUMBER OF REFERENCES: 36

ABSTRACT

Background

Coding of diagnoses is important for patient care, hospital management and research. However coding accuracy is often poor and may reflect methods of coding. This study investigates the impact of three alternative coding methods on the inaccuracy of diagnosis codes and hospital reimbursement.

Methods:

Comparisons of coding inaccuracy were made between a list of coded diagnoses obtained by a coder using (i)the discharge summary alone, (ii)case notes and discharge summary, and (iii)discharge summary with the addition of medical input. For each method, inaccuracy was determined for the primary, secondary diagnoses, Healthcare Resource Group (HRG) and estimated hospital reimbursement. These data were then compared with a gold standard derived by a consultant and coder.

Results:

107 consecutive patient discharges were analysed. Inaccuracy of diagnosis codes was highest when a coder used the discharge summary alone, and decreased significantly when the coder used the case notes (70% vs 58% respectively, $p<0.0001$) or coded from the discharge summary with medical support (70% vs 60% respectively, $p<0.0001$). When compared with the gold standard, the percentage of incorrect HRGs was 42% for discharge summary alone, 31% for coding with case notes, and 35% for coding with medical support. The three coding methods resulted in an annual estimated loss of hospital remuneration of between £1.8M and £16.5M.

Conclusion:

The accuracy of diagnosis codes and percentage of correct HRGs improved when coders used either case notes or medical support in addition to the discharge summary. Further emphasis needs to be placed on improving the standard of information recorded in discharge summaries.

KEYWORDS

Diagnosis, Data Accuracy, Clinical coding and Quality of health care

ABBREVIATIONS

HRG: Healthcare Resource Group

MAIN TEXT

1. INTRODUCTION

In most health systems across the world, patient diagnoses are translated into suitable codes at hospital discharge using a coding scheme such as ICD 10 [1,2], the most widely used terminology. The resulting diagnosis codes are used by several organizations for differing purposes: (i) hospitals or health insurers, to justify and receive financial remuneration (tariffs attributed according to the Healthcare Resource Group (HRG) within UK and to the Diagnosis Related Group (DRG) throughout other countries. HRG (or DRG) are clinically meaningful groups of diagnoses and interventions considered as consuming similar levels of financial resources) [3], (ii) health systems, to monitor disease outbreaks, report mortality and plan national strategies for improving the quality and safety of healthcare (e.g. the Centre for Disease Control in USA) [4,5], (iii) companies, to measure doctor and hospital performance (e.g. Dr Foster in the UK) [6–8], (iv) researchers, to carry out epidemiology and health services research [9]. Despite the importance of recording accurate data, there remains significant variation in the reported accuracy of diagnosis codes which can range from 51% to 98% [10,11]. This may reflect differences in coding practices between hospitals.

There is significant variation in coding practice between countries and even hospitals in the same health care system [12,13], with diagnoses being collected either through remote or point-of-care coding. Remote coding is done entirely by dedicated coders, who are non-medical staff with strong terminology skills, using discharge summaries with or without case notes [12]. The discharge summary is often used as the sole source of information. While these are concise documents, they can be written retrospectively due to time constraints and the information in summaries can be inaccurate [14,15]. In contrast, the case notes contain in-depth prospectively recorded information. However, this is often voluminous, disorganized and contains multiple abbreviations, making it difficult for coders to extract the information they need [16]. Point-of-care coding is undertaken by medical doctors and coders, usually from discharge summaries alone [12]. Since medical doctors

often lack knowledge of coding terminology, coders usually check and complete the list of codes generated by the doctor.

Despite the importance of accurately recording and coding discharge diagnoses, few studies [17–19] have assessed the impact of varying methods of data capture on the accuracy of diagnosis codes. It remains unclear if coders should refer to case notes and/or discharge summaries and whether and how medical doctors should be involved. To address these questions, we conducted a prospective study comparing the impact of three coding methods on the inaccuracy of diagnosis coding against a gold standard (or criterion standard), and the consequent impact on calculated hospital remuneration.

2.METHODS

2.1 Study design

This was a comparative study using data from a prospective cohort of consecutive patients discharged from three adult respiratory wards at St James University Hospital Leeds during March 2015. Exclusion criteria included the absence of a primary respiratory diagnosis, a missing discharge summary or an ambulatory patient attending for a day case procedure such as a bronchoscopy.

For each patient, we generated four lists of diagnosis codes (Table 1):

- The gold standard list (or criterion standard). This was derived soon after discharge by the doctor responsible for the care episode, working with a coder using the case notes
- The remote coded list with case notes. This was derived by a coder using the paper case notes in addition to the electronic discharge summaries
- The remote coded list. This was derived by the coder using the electronic discharge summary, which had been generated by junior doctors following discharge using a basic template.

- The point of care coded list with doctor. This was derived by a doctor naïve to the case and the coder using the electronic discharge summary alone

One author (RT) who did not participate in the coding process compared the four lists of codes for each patient.

Table 1: Definition of source of coding and personal involvement in code generation

		Features	**Gold standard**	**Remote coding**	**Remote coding with case notes**	**Point of care coding with doctor**
Materials	Case notes	Paper format	x		x	
	Discharge summaries	Electronic basic template fill by junior doctors for all patients in discharge. It includes the following sections: - Administrative patient data - Drug allergies and sensitivities - Primary diagnosis/procedure and advice to GPs - Information on medication - Follow up arrangements		x	x	x
Persons involved in coding	Coder	Individual (without medical knowledge) having the UK National Clinical Coding Qualification, and 8 years of experience in coding	x	x		x
	Coders	Two individuals (without medical knowledge) having			x	

		the UK National Clinical Coding Qualification, and 8 years of experience in coding				
	Consultant	Responsible for the care episode	x			
	Consultant	Naïve to the care episode				x

2.2 Generation of the four code lists for each patient

2.2.1 Generation of the gold standard list

First, the doctor responsible for that patient's inpatient stay identified the primary and secondary diagnoses for their patient using the case notes, test results and their knowledge about the patient. These diagnoses were then converted into ICD 10 codes by the coder in the presence of the doctor. Then, after reading the case notes alone, the coder suggested new or modified codes. The doctor could then decide whether to accept the changes or not. The resulting list was taken as the gold standard for each patient. During this process, the team was blinded to the contents of the discharge summary.

2.2.2 Generation of remote coding with case notes

The coders used the case notes and electronic discharge summaries to generate this list during the routine hospital coding process. Coders were blind to the three other code lists.

2.2.3 Remote coding with discharge summary and Point of care coding by coder and doctor with electronic discharge summaries

A doctor naïve to the clinical case and the coder independently and simultaneously generated a list of diagnoses from the anonymised discharge summary. The coder generated a list of codes from the electronic discharge summary and converted them into ICD10 codes. This corresponds to the remote coding with discharge summaries. Then, the doctor and coder compared their lists and generated a complete list of codes that was taken as the point of care coding list. Anonymised discharge summaries and a wash-out period of at least three weeks after the derivation of the gold standard were used to reduce memory effects in the coder.

2.3 Comparison of coding methods

For each patient, the remote coded list using discharge summary, remote coded list using case notes and discharge summary and the point of care coding were compared to the gold standard.

The comparison of code accuracy for each method was undertaken by matching each ICD10 code at the level of 3 or 4 characters, the ICD block title or clinical meaning alone (An example of ICD 10 structure and explanation of these terms is shown in Figure 1). Six categories of accuracy compared to the gold standard coding were defined:

- Accurate, if the same four ICD10 characters appeared on both lists
- Partially accurate, if the same three ICD10 characters appeared
- Inaccurate, if the codes appeared under same ICD10 block titles only
- Seriously inaccurate, if codes had the same clinical meaning only
- Missing code, if the gold standard codes were not present in the coded list
- Wrong code, if a code was present in the list but not present in the gold standard list

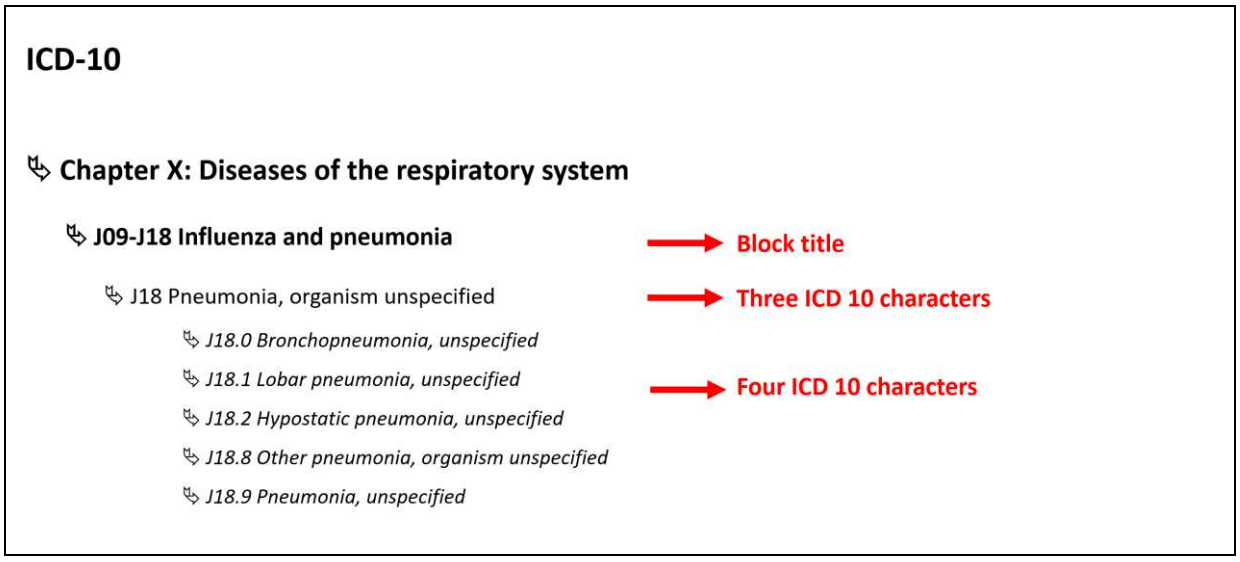

Figure 1: ICD 10 structure. Example

2.4 Analysis

2.4.1 Inaccuracy rate per patient

For each of the three coding methods, the inaccuracy rate per patient was measured for each of the six categories of accuracy defined above. A global inaccuracy rate was also defined as follows:

$$p\% = \frac{\text{Partially inaccurate codes} + \text{Inaccurate codes} + \text{Seriously inaccurate codes} + \text{Missing codes} + \text{Wrong codes}}{\text{Accurate codes} + \text{Partially inaccurate codes} + \text{Inaccurate codes} + \text{Seriously inaccurate codes} + \text{Missing codes} + \text{Wrong codes}}$$

Two kinds of inaccuracy were defined: the inaccuracy for all diagnoses, and the inaccuracy for the primary diagnosis alone. For the primary diagnosis, we matched the primary diagnosis code from each of the three lists with the gold standard primary diagnosis code. When no match was retrieved, we investigated if the primary diagnosis code for each list matched the gold standard secondary diagnosis code.

2.4.2 Calculation of hospital remuneration

For each of the four code lists and each patient stay we identified the HRG that best matched the list of diagnosis codes using the NHS HRG4-2014-2015 Grouper [20]. This grouper is NHS software used by coders for training them to calculate the HRG from a list of codes and demographics [20]. As the grouper is limited to patients with less than 14 discharge diagnoses, we were only able to obtain the HRG for 73 (68%) of the 107 patients. NHS tariff costs were then calculated using NHS tariff information linking HRG and spell tariff [21]. For each coding method, we estimated the impact on total hospital remuneration per year for a typical 125 000 inpatients, as:

- no impact, i.e. when the incorrect HRG led to the same remuneration as the HRG based on the gold standard

- loss of remuneration, i.e. when the incorrect HRG led to a lower remuneration than the HRG based on the gold standard
- incorrectly high remuneration, i.e. when the incorrect HRG led to a higher remuneration than the HRG obtained using the gold standard

Results are presented with 95% confidence intervals. A p value of <0.05 was taken as significant (Software used: R version 3.2.2).

The inaccuracy rate for all diagnosis codes were compared using the Wilcoxon test on paired data for all diagnoses. Inaccuracy rates for primary diagnosis codes and differences in HRG were compared using a McNemar test on paired data for primary diagnosis. The Bonferroni correction was used to take account of multiple comparisons.

2.4.3 Qualitative analysis of coding errors

Where possible, we assessed the reasons for coding errors in each of the three lists, using comments made by coders and doctor recorded during the coding process.

3. RESULTS

The initial patient cohort included 142 patients. Thirty-two patients were excluded because of a missing discharge summary or no primary respiratory diagnosis. Data were analyzed for 107/110 (97%) cases. The delay in receiving case notes was too long for two patients and a breach in study protocol occurred for one patient.

3.1 Coding inaccuracy

Inaccuracy rates are shown in table 2 for all patient diagnosis codes, and in table 3 for primary diagnosis codes.

3.1.2 All patient diagnosis codes (Table 2)

Remote coding using the discharge summary alone was more inaccurate than remote coding with notes (70% vs 58% respectively, $p < 0.0001$) and point of care coding with a doctor (70% vs 60% respectively, $p < 0.0001$). Most inaccuracies related to missing codes, which were approximately 10% greater in remote coding using the discharge summary alone. There were no significant differences in any category of inaccuracy between remote coding with case notes and point of care coding.

3.1.3 Primary diagnosis code (Table 3)

Remote coding using the discharge summary alone was more inaccurate than remote coding with notes (65% vs 50% respectively, $p < 0.002$) and point of care coding with a doctor (65% vs 57% respectively, $p < 0.02$).

The primary diagnosis failed to match the primary gold standard diagnosis code for 29% [20% to 38%], 22% [14% to 29%], or 22% [14% to 29%] of patients for remote coding with discharge summary, remote coding with notes and point of care coding, respectively.

There were no significant differences in the inaccuracy of primary diagnosis codes between remote coding with case notes and point of care coding.

Table 2: Inaccuracy rate for combined primary and secondary diagnosis codes for the three coding methods compared with the gold standard

	Remote coding (%)	Remote coding with case notes (%)	Point of care coding with doctor (%)	Remote coding versus Remote coding with case notes (p value)	Remote coding versus Point of care coding with doctor (p value)	Remote coding with case notes versus Point of care coding with doctor (p value)
Accurate (same 4 digits)	30% [27 to 33]	42% [39 to 45]	40% [37 to 43]	**< 0.0001**	**< 0.0001**	1
Partially Inaccurate (same 3 digits)	6% [5 to 8]	5% [3 to 6]	5% [4 to 6]	**0.01**	0.06	0.8
Inaccurate (same block title)	4% [2 to 5]	3% [2 to 5]	4% [2 to 5]	1	1	1
Seriously inaccurate (clinical connection only)	3% [2 to 5]	3% [2 to 4]	3% [2 to 5]	1	1	1

Missing	52% [48 to 55]	41% [38 to 44]	41% [38 to 44]	**< 0.0001**	**< 0.0001**	1
Wrong	5% [3 to 7]	6% [4 to 8]	7% [5 to 9]	0.9	**0.02**	0.8

Table 3 : Inaccuracy rate for primary diagnosis code for the three coding methods compared with the gold standard

		Remote coding (%)	Remote coding with case notes (%)	Point of care coding with doctor (%)	Remote coding versus Remote coding with case notes (p value)	Remote coding versus Point of care coding with doctor (p value)	Remote coding with case notes versus Point of care coding with doctor (p value)
Match with the gold standard primary diagnosis code	Accurate (same 4 digits)	35% [26 to 44]	50% [41 to 60]	43% [34 to 52]	**0.002**	**0.02**	0.3
	Partially Inaccurate (same 3 digits)	10% [5 to 16]	7% [2 to 11]	11% [5 to 17]	1	1	0.4
	Inaccurate (same block title)	21% [14 to 29]	19% [12 to 27]	20% [13 to 28]	1	1	1
	Seriously inaccurate (clinical connection only)	5% [1 to 9]	2% [0 to 4]	4% [0 to 7]	1	1	1

Match with one of the gold standard secondary diagnosis codes	Accurate (same 4 digits)	14% [7 to 21]	9% [4 to 15]	12% [6 to 18]	0.5	1	1
	Partially Inaccurate (same 3 digits)	2% [0 to 4]	2% [0 to 4]	1% [0 to 3]	1	1	1
	Inaccurate (same block title)	3% [0 to 6]	2% [0 to 4]	2% [0 to 4]	1	1	1
	Seriously inaccurate (clinical connection only)	3% [0 to 6]	4% [0 to 7]	2% [0 to 4]	1	1	1
No match with the gold standard diagnosis codes	Wrong	7% [2 to 12]	5% [1 to 9]	5% [1 to 9]	1	0.7	1

3.2 Impact on estimated hospital remuneration

When compared with the gold standard, the percentage of incorrect HRGs was 42% for remote coding, 31% for remote coding with case notes, and 35% for point of care coding, respectively. No significant difference was found between the three methods of coding (Table 4).

Scaling these figures up to a typical large teaching hospital with 125 thousand inpatient discharges per year, the three methods of coding led to an estimated loss of remuneration per year of £16.5M for remote coding from discharge summaries alone, £1.8M for remote coding with case notes, and £15.4M for point of care coding. The loss of remuneration was less for remote coding with case notes because the loss was compensated by an incorrectly high remuneration for some patients having an incorrect HRG (Table 4).

Table 4: Impact of coding on Health Resource Groups (HRG) and hospital remuneration predicted by year (125 000 inpatients by year [36])

	Remote coding		Remote coding with case notes		Point of care coding with doctor	
	Percentage of incorrect HRGs (%)	Impact on hospital remuneration predicted by year	Percentage of incorrect HRGs (%)	Impact on hospital remuneration predicted by year	Percentage of incorrect HRGs (%)	Impact on hospital remuneration predicted by year
Incorrect HRGs (%) leading to correct remuneration (£)	1% [0 to 4]	0£	0%	0£	1% [0 to 4]	0£
Incorrect HRGs (%) leading to incorrect high remuneration (£)	16% [8 to 25]	+£13.1M	16% [8 to 25]	+£21.5M	12% [5 to 20]	+£10.1M
Incorrect HRGs (%) leading to incorrect low remuneration (£)	25% [15 to 35]	-£29.6M	15% [7 to 23]	-£23.3M	22% [12 to 31]	-£25.5M
Incorrect HRGs in total (%) impacting on remuneration (£)	42% [31 to 54]	-£16.5M£	31% [21 to 42]	-£1.8M	35% [25 to 47]	-£15.4M

3.3 Subjective feedback from the doctors and coders

The inaccuracy of remote coding from discharge summaries appears to be related to the lack of medical knowledge in the coders, leading to difficulty in deducing the diagnoses from, for example, the drug chart or clinical test results. For instance, “acidosis” was not coded, even when blood gas results clearly indicating this were written in the discharge summary (Table 5).

The inaccuracy of remote coding with case notes seems to be related to the inability of the coder to judge if a given diagnosis within several diagnoses is relevant enough to be coded, again because of their lack of medical knowledge. Further explanations are their lack of training in medical reasoning, leading to variable interpretation of medical terms, and difficulty managing the paper notes, which are voluminous and difficult to read.

Inaccuracies in the point of care coding by the doctor seem to be related to the doctor’s lack of involvement in the care process, which makes it harder for them to know if the diagnoses were relevant enough to be coded, and by ambiguity in the discharge summaries. For example, we noted a lack of accurate or sufficiently detailed diagnoses, a mix between past and current diagnoses, and the presence of queries in the list of diagnoses.

Table 5: Subjective feedback from the doctors and coders. Reasons for coding errors according to the coding method (n= number of reasons that were documented in total)

Problems related to		Remote coding (n=186)	Remote coding with notes (n=99)	Point of carer coding with doctor (n=41)	Example of errors
No involvement in the care process	Relevance	8.5%	29.5%	41.5%	Coder coded personal history of infectious and parasitic diseases (Z86.1), because he thought it was relevant for the patient
Medical reasoning	Symptoms	2.5%	15%	7.5%	Coder coded hallucination (R44.3) and disorientation (R41.0) instead of the diagnosis delirium (F05.9)
	Interpretation	35%	20%	-	Coder coded lobar pneumoniae (J18.1) instead of pneumonia due to streptococcus pneumoniae (J13.X), because he couldn't interpret the information "blood culture positive to S pneumoniae"
	Drug chart	32.5%	-	17%	Coder didn't code hypercholesterolemia (E78.0) from discharge

					summary, because it could only be deduced from drug chart (treatment by statins)
Content of documents	Inaccurate description in notes or discharge summary	2%	10%	10%	Coder coded left ventricular failure (I50.1) instead of left ventricular hypertrophy (I51.7), because of a misunderstanding of the abbreviation "LVH" in the notes Coder and doctor both coded bladder disorder (N32.9) instead of bladder neoplasm (C67.9), because "bladder disorder" was written in the discharge summary
	Past and current diagnosis	6%	16.5%	22%	Coder and doctor both coded stroke (I64) instead of history of stroke (Z86.7), because "stroke" was written in the discharge summary
	Hypotheses	4%	-	2%	Coder code lobar pneumoniae (J18.1) instead of pneumonia due to mycoplasma pneumoniae (J15.7), because "likely due to mycoplasma pneumoniae" was written in the discharge summary
ICD 10	Coding	-	9%	-	Coder coded obesity (E66.8) and sleep apnoea (G47.3) instead of

knowledge	terminology				the unique code “extreme obesity with alveolar hypoventilation (E66.2)”
Other	Forgetting	9.5%	-	-	Coder forget to code recurrent depressive disorder (F33.3) whereas it was written in the document

4. DISCUSSION

To our knowledge, this is the first study investigating the impact of various professional roles (coder or doctor) and sources of data (notes or discharge summaries) on the inaccuracy of coding discharge diagnoses and resulting impact on hospital remuneration. Despite the heterogeneity of coding practices between countries, and between hospitals within the same country, there have been no studies establishing best practice for coding discharge diagnoses. In our study the highest diagnosis code inaccuracy occurred when coding was done by coders alone using the hospital discharge summary as their only source of information. Inaccuracy was decreased by the addition of a doctor's input and when coders used a combination of the case notes and discharge summary. Either of these changes resulted in a 10% absolute decrease in coding inaccuracy for all diagnoses, and an 8% to 15% absolute decrease for the primary diagnosis codes. The use of case notes also minimized the percentage of incorrect HRGs at 31%. The three methods of coding led to a loss of remuneration to the hospital estimated at between £1.8M and £16.5M.

The main strengths of our study are that we used a rigorous sequential sampling methodology for recruiting participants, a well-defined gold standard resulting from an appropriately multidisciplinary process, and an objective comparison process. Our study was carried out using predefined rigorous methodology [22] which included consecutive cases (only 2.7% of patients were omitted), clear determination of the gold standard list of diagnosis codes, generation of three independent diagnostic lists corresponding to the three coding practices, an objective measurement of diagnostic inaccuracy using ICD-10 codes compared by a person not involved in the generation of the lists using a 6-item taxonomy, and an objective calculation of the impact on remuneration using the NHS approved HRG4 Grouper [20].

Our study has acknowledged limitations. Firstly, bias may have occurred through the use of different coders to generate the remote coding from case notes than for the three other lists [33]. ICD 10

complexity may lead to inter-rater variability between different coders [34,35], and intra-rater variability for the same coder's decisions at different time points [33]. Whilst this methodology was used to enhance time efficiency and to avoid memory effects in the coder allocating the gold standard list of diagnosis codes, we also limited this bias by involving coders in the same coding department with similar coding experience. The greater accuracy shown by coders from case notes compared to coders from discharge summaries, would support this view. A second bias may have been introduced by the doctor's very limited knowledge of ICD 10 terms. Doctors habitually describe patient diagnoses using medical, not ICD 10, terms. This gap between medical thinking and ICD 10 terminology [33] was bridged by the use of a trained coder, who translated medical terms into appropriate ICD 10 terms. Third, our study was conducted solely on three respiratory wards in one large UK teaching hospital, and over a one month period. Although our respiratory cohort had significant comorbidities (a mean of 12 diagnoses per patient), it is likely to reflect general medicine patients in their complexity. However, further studies in other hospitals and across differing specialities and time periods should be conducted to confirm our results.

A few studies have assessed the inaccuracy of clinical coding practices [17–19]. Some of these reported higher errors in HRG with the use of discharge summaries [17]. Others reported lower inaccuracy rates with point of care coding with a doctor compared to remote coding [18]. However, all these studies suffered from one or more serious methodological issues, specifically: (i) the robustness of the gold standard was low [17], (ii) the samples were too small [17], (iii) too many patients were omitted, leading to suspicion of selection bias [18], (iv) the conclusions were based on HRG change (not on the inaccuracy of discharge diagnoses) [17], and (v) some studies focused on only one diagnosis [17]. In addition, these studies focused on either the documents that should be used (discharge summaries or notes) [17,19] or on who should be involved in coding (coder or doctor) [18]. No study directly compared remote coding with or without notes and point of care coding with a doctor against a robust gold standard in the same cohort of sequential patients, as we have.

Our results objectively demonstrate that diagnostic coding of inpatients requires both clinical and ICD 10 terminology knowledge. Remote coding by the coder alone from discharge summaries leads to higher inaccuracy, probably because interpretation of the patient data in discharge summaries was often influenced by the coder's lack of medical knowledge. We have shown that this can be offset by the use the case records, which contain more detailed and explicit patient data, or the addition of medical input to interpret and make explicit unclear data in discharge summaries. Point-of-care coding with a doctor using discharge summaries was significantly quicker than remote coding with notes, taking approximately 5 minutes and 30 minutes per patient, respectively, according to our estimate. Because of the study context, we were not able to measure the exact time for each method. Indeed, the study was undertaken during the working time of consultants and coders, and subject to "breaks" during the process of coding. While the addition of a medical input to analyse the discharge summary generates additional costs, improving the quality, accuracy and content of the discharge summary at source is likely to be cost effective. More accurate discharge summaries could be used by coders to translate the medical terminology into ICD 10 codes. Writing more accurate, explicit discharge summaries is not time wasting for doctors, since it would also impact on improving patient management, safety of care, hospital remuneration [23] and decrease the risk of rehospitalization [24]. Alternatively, medical time could be saved using automatic completion of discharge summaries from templates or electronic care records [25–28], allowing more accurate and time-efficient summaries to be generated at the time of discharge [29–32].

5. CONCLUSION

Our results report disappointingly high inaccuracy in diagnostic coding, the main source of data used for hospital funding, research (Hospital Episode Statistics), and quality improvement (e.g. hospital standardised mortality rates). The diagnostic inaccuracy was highest (70%) when we restricted the coders to using the discharge summary only, a common scenario in many hospitals across the globe.

However, when doctors unfamiliar with the patients used the same summary, they were able to restore the inaccuracy to the baseline of around 60%. This suggests that the necessary patient data were actually present in the summary but not in a format that non-medical coders could interpret to assemble a complete list of coded diagnoses. This could be solved either by hospital doctors spending more time writing the discharge summary, or by electronic patient records harvesting diagnoses prospectively from progress notes, lab reports etc. during the inpatient stay. The resulting increase in hospital reimbursement would more than cover the costs of implementing such digital record and summary writing systems.

ACKNOWLEDGMENTS

We would like to thank members of the coding department of St James's University Hospital for their help and support. Michael Routledge, the coder, who coded all the notes and discharge summaries. Natasha Noble and Rebecca Marshall, the runners, who collected the notes. Ben Phillipskirk and Catherine Hutchinson, the coders, who were involved in the official coding. Tracey Conroy and Victoria Anne Macwhirter, the coder leaders for their support and expert advice.

FUNDING

Funding for RT salary from the Yorkshire & Humber NIHR CLAHRC

CONFLICTS OF INTEREST

None declared

KEYPOINTS

- Accuracy of coding is important for hospital reimbursement, audit and research
- Inaccuracy of coding is related to the method of coding within hospitals
- Inaccuracy of coding decreased with the use of case notes or with medical support
- Clinicians should write discharge summaries accurately to improve coding quality

ETHICAL APPROVAL

None

CONTRIBUTORSHIP STATEMENT

Design of the study protocol: RT, DP, JW
Setting up the study in hospital: RT, DP
Derivation of the gold standard: IC, PB, PW, KR, MC, DG
Extraction and coding of diagnoses from discharge summaries: IC, PW, MC, DG
Analysis of Data and statistical analysis: RT
Writing the manuscript: RT
Revising the manuscript critically: DP, JW
Adding relevant suggestions to improve the manuscript: IC, PB, PW, KR, MC, DG, HW
Agreement for all aspects of the work and approval of the final version to be published: RT, DP, JW, IC, PB, PW, KR, MC, DG, JPJ, HW

REFERENCES

[1] WHO | International Classification of Diseases, WHO. (n.d.). http://www.who.int/classifications/icd/en/ (accessed October 16, 2016).

[2] J.J. Cimino, Desiderata for Controlled Medical Vocabularies in the Twenty-First Century, Methods Inf. Med. 37 (1998) 394–403.

[3] R. Rogers, S. Williams, B. Jarman, P. Aylin, "HRG drift" and payment by results, BMJ. 330 (2005) 563. doi:10.1136/bmj.330.7491.563.

[4] P. Cheng, A. Gilchrist, K.M. Robinson, L. Paul, The risk and consequences of clinical miscoding due to inadequate medical documentation: a case study of the impact on health services funding, HIM J. 38 (2009) 35–46.

[5] P.L. Yudkin, C.W. Redman, Obstetric audit using routinely collected computerised data, BMJ. 301 (1990) 1371–1373.

[6] S. Jameson, M.R. Reed, Payment by results and coding practice in the National Health Service. The importance for orthopaedic surgeons, J. Bone Joint Surg. Br. 89 (2007) 1427–1430. doi:10.1302/0301-620X.89B11.19609.

[7] M.A. Mohammed, J.J. Deeks, A. Girling, G. Rudge, M. Carmalt, A.J. Stevens, R.J. Lilford, Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals, BMJ. 338 (2009) b780.

[8] Dr Foster — Intelligence in healthcare, (n.d.). http://www.drfoster.com/ (accessed December 28, 2015).

[9] J.D. Lewsey, A.H. Leyland, G.D. Murray, F.A. Boddy, Using routine data to complement and enhance the results of randomised controlled trials, Health Technol. Assess. Winch. Engl. 4 (2000) 1–55.

[10] S.E. Campbell, M.K. Campbell, J.M. Grimshaw, A.E. Walker, A systematic review of discharge coding accuracy, J. Public Health Med. 23 (2001) 205–211.

[11] E.M. Burns, E. Rigby, R. Mamidanna, A. Bottle, P. Aylin, P. Ziprin, O.D. Faiz, Systematic review of discharge coding accuracy, J. Public Health Oxf. Engl. 34 (2012) 138–148. doi:10.1093/pubmed/fdr054.

[12] K. Pongpirul, D.G. Walker, P.J. Winch, C. Robinson, A qualitative study of DRG coding practice in hospitals under the Thai Universal Coverage scheme, BMC Health Serv. Res. 11 (2011) 71. doi:10.1186/1472-6963-11-71.

[13] S. a. R. Nouraei, S. O'Hanlon, C.R. Butler, A. Hadovsky, E. Donald, E. Benjamin, G.S. Sandhu, A multidisciplinary audit of clinical coding accuracy in otolaryngology: financial, managerial and clinical governance considerations under payment-by-results, Clin. Otolaryngol. Off. J. ENT-UK Off. J. Neth. Soc. Oto-Rhino-Laryngol. Cervico-Facial Surg. 34 (2009) 43–51. doi:10.1111/j.1749-4486.2008.01863.x.

[14] K. Legault, J. Ostro, Z. Khalid, P. Wasi, J.J. You, Quality of discharge summaries prepared by first year internal medicine residents, BMC Med. Educ. 12 (2012) 77. doi:10.1186/1472-6920-12-77.

[15] J.L. Callen, M. Alderton, J. McIntosh, Evaluation of electronic discharge summaries: a comparison of documentation in electronic and handwritten discharge summaries, Int. J. Med. Inf. 77 (2008) 613–620. doi:10.1016/j.ijmedinf.2007.12.002.

[16] K.J. O'Malley, K.F. Cook, M.D. Price, K.R. Wildes, J.F. Hurdle, C.M. Ashton, Measuring diagnoses: ICD code accuracy, Health Serv. Res. 40 (2005) 1620–1639. doi:10.1111/j.1475-6773.2005.00444.x.

[17] S. Jameson, A. Nargol, M. Reed, Payment by results for fractured neck of femur in two NHS Secondary Care Trusts, Bull. R. Coll. Surg. Engl. 89 (2007) 318–320. doi:10.1308/147363507X238830.

[18] C. Yeoh, H. Davies, Clinical coding: completeness and accuracy when doctors take it on., BMJ. 306 (1993) 972–972. doi:10.1136/bmj.306.6883.972.

[19] N. Chin, P. Perera, A. Roberts, R. Nagappan, Review of medical discharge summaries and medical documentation in a metropolitan hospital: impact on diagnostic-related groups and

Weighted Inlier Equivalent Separation, Intern. Med. J. 43 (2013) 767–771. doi:10.1111/imj.12084.

[20] 1 Trevelyan Square NHS Digital, HRG4 2014/15 Payment Grouper, (2014). http://content.digital.nhs.uk/article/3938/HRG4-201415-Payment-Grouper?tabid=1 (accessed October 16, 2016).

[21] National tariff payment system 2014/15 - Publications - GOV.UK, (n.d.). https://www.gov.uk/government/publications/national-tariff-payment-system-2014-to-2015 (accessed October 16, 2016).

[22] W.R. Hogan, M.M. Wagner, Accuracy of Data in Computer-based Patient Records, J. Am. Med. Inform. Assoc. 4 (1997) 342–355.

[23] S. Kripalani, F. LeFevre, C.O. Phillips, M.V. Williams, P. Basaviah, D.W. Baker, Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care, JAMA. 297 (2007) 831–841. doi:10.1001/jama.297.8.831.

[24] C. van Walraven, R. Seth, P.C. Austin, A. Laupacis, Effect of Discharge Summary Availability During Post-discharge Visits on Hospital Readmission, J. Gen. Intern. Med. 17 (2002) 186–192. doi:10.1046/j.1525-1497.2002.10741.x.

[25] R. Crosswhite, S.H. Beckham, P. Gray, P.R. Hawkins, J. Hughes, Using a multidisciplinary automated discharge summary process to improve information management across the system, Am. J. Manag. Care. 3 (1997) 473–479.

[26] K. Bischoff, A. Goel, H. Hollander, S.R. Ranji, M. Mourad, The Housestaff Incentive Program: improving the timeliness and quality of discharge summaries by engaging residents in quality improvement, BMJ Qual. Saf. 22 (2013) 768–774. doi:10.1136/bmjqs-2012-001671.

[27] S. Chan, A.P. Maurice, C.W. Pollard, S.J. Ayre, D.L. Walters, H.E. Ward, Improving the efficiency of discharge summary completion by linking to preexisiting patient information databases, BMJ Qual. Improv. Rep. 3 (2014) u200548.w2006. doi:10.1136/bmjquality.u200548.w2006.

[28] K. Kortüm, C. Hirneiß, M. Müller, A. Babenko, A. Kampik, T.C. Kreutzer, The influence of a specific ophthalmological electronic health record on ICD-10 coding, BMC Med. Inform. Decis. Mak. 16 (2016) 100. doi:10.1186/s12911-016-0340-1.

[29] K.J. O'Leary, D.M. Liebovitz, J. Feinglass, D.T. Liss, D.B. Evans, N. Kulkarni, M.P. Landler, D.W. Baker, Creating a better discharge summary: improvement in quality and timeliness using an electronic discharge summary, J. Hosp. Med. 4 (2009) 219–225. doi:10.1002/jhm.425.

[30] H.B. Burke, L.L. Sessums, A. Hoang, D.A. Becher, P. Fontelo, F. Liu, M. Stephens, L.N. Pangaro, P.G. O'Malley, N.S. Baxi, C.W. Bunt, V.F. Capaldi, J.M. Chen, B.A. Cooper, D.A. Djuric, J.A. Hodge, S. Kane, C. Magee, Z.R. Makary, R.M. Mallory, T. Miller, A. Saperstein, J. Servey, R.W. Gimbel, Electronic health records improve clinical note quality, J. Am. Med. Inform. Assoc. (2014) amiajnl-2014-002726. doi:10.1136/amiajnl-2014-002726.

[31] P.C. Tang, M.P. LaRosa, S.M. Gorden, Use of Computer-based Records, Completeness of Documentation, and Appropriateness of Documented Clinical Decisions, J. Am. Med. Inform. Assoc. 6 (1999) 245–251. doi:10.1136/jamia.1999.0060245.

[32] S.T. Rosenbloom, J.C. Denny, H. Xu, N. Lorenzi, W.W. Stead, K.B. Johnson, Data from clinical notes: a perspective on the tension between structure and flexible documentation, J. Am. Med. Inform. Assoc. 18 (2011) 181–186. doi:10.1136/jamia.2010.007237.

[33] J. Stausberg, N. Lehmann, D. Kaczmarek, M. Stein, Reliability of diagnoses coding with ICD-10, Int. J. Med. Inf. 77 (2008) 50–57. doi:10.1016/j.ijmedinf.2006.11.005.

[34] B. Misset, D. Nakache, A. Vesin, M. Darmon, M. Garrouste-Orgeas, B. Mourvillier, C. Adrie, S. Pease, M.-A.C. de Beauregard, D. Goldgran-Toledano, E. Métais, J.-F. Timsit, Reliability of diagnostic coding in intensive care patients, Crit. Care. 12 (2008) R95. doi:10.1186/cc6969.

[35] P. Hjerpe, J. Merlo, H. Ohlsson, K. Bengtsson Boström, U. Lindblad, Validity of registration of ICD codes and prescriptions in a research database in Swedish primary care: a cross-sectional study in Skaraborg primary care database, BMC Med. Inform. Decis. Mak. 10 (2010) 23. doi:10.1186/1472-6947-10-23.

[36] The Leeds Teaching Hospitals NHS Trust, Annual Report and Accounts 2014/15, (2014).