UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA

SCUOLA DI MEDICINA

Faculty of Medicine

Emergency Medicine specialty training program
Director: Prof. Walter Ageno

Accuracy levels of Emergency Department discharge diagnosis coding in the Royal Infirmary of Edinburgh

Supervisor:
Prof. Walter Ageno

Co-Supervisor:
Prof. Matthew Reed

Specialty thesis of:

Dr. Noga Dotan
Student n. 738628

Academic year 2021/2022

INDEX

1. INTRODUCTION

Recording data from Emergency Department (ED) attendance, particularly emergency care diagnoses, plays a crucial role in many different areas of health care and guides its improvement and advancement. The analysis of this data is widely applied at national and local levels for healthcare resources management, funding strategies, monitoring of disease outbreaks and trending, audit and research. Hence, it is of great importance that the data collected is of high quality, reliable and accurate.

1.1 Clinical coding

Clinical coding is the process by which crude clinical information is translated into numeric or alphanumeric codes which subsequently allows better handling, analysis and use of the clinical data. Clinical codes overall represent a standardised manner of capturing clinical information. It encompasses both diagnosis codes and procedure codes. The diagnosis codes are used as a tool to group and identify clinical conditions, diseases, disorders, injuries, symptoms, intoxications, adverse drug effects and other incidences or pathological conditions. They are translated from descriptive notions into codes with different diagnosis classification systems used worldwide. Procedure codes on the other hand are used to identify and record specific surgical, medical, or diagnostic interventions.

This thesis project will focus on diagnosis codes, in particular in the context of the ED in Scotland.

Diagnosis Coding Sets

Across the world, different code sets are used, with the World Health Organization (WHO) International Classification of Disease (ICD) being the most common one [1–3]. In the UK, a poly-hierarchical new generation clinical terminology tool called Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [4,5]. It allows a better clinical content coverage, clinical-orientation, flexible data entry and retrieval capabilities [6]. Additionally, through dedicated mapping tools, the SNOMED CT codes, together with consideration of co-morbidity, patient demographics and other variables can, in most cases, be inferred to ICD codes.

For ED diagnosis codes a designated code set is used in the UK since 2015, called Emergency Care Data Set (ECDS), which is a truncated version of SNOMED [7,8].

Coding process

The actual process of assigning discharge codes is complex. Coding strategies and practice differs greatly between countries and even between hospitals in the same region, moreover, it may be setting specific [9,10]. Most commonly, coding is performed by dedicated trained coders, who are non-medical staff with strong terminology skills, using different resources of information recorded by physicians such as discharge summaries and abstracts, case notes, clinical database or local registries. This makes quality of coded data highly dependent on the quality and precision of the original patient notes or abstracts written by the physician, as well as the experience and expertise of the individual coder assigning the codes [11–13].

1.2 The use of clinical coded data

Funding and reimbursement

Originally, nosology and the first disease 'codes' were established with the aim of tracking causes of sickness and death in the population. However, later on, in the mid-late twentieth century, medical insurance programs and healthcare billing companies made nosology and coding a matter of great interest to public and private payers of health care [10]. Indeed, nowadays, one of the central uses of data derived from clinical coding is for the purpose of financing strategies, funding and reimbursement. Different countries and different health systems have distinct processes and methods to use and analyse this data for funding purposes.

Funding and reimbursement in the UK and Scotland

Historically, in the UK, payment was often linked to hospital activity. However, such activity-based payments were extensively criticised, as they did not cover all aspects of care. Concerns that funding based purely on activity may actually incentivise unnecessary admissions to hospital have been raised. These concerns have led to the development of ‘tariffs’ that reflect payment for a whole patient pathway, including post admission care or care to prevent admissions. This concept of reflecting payment on improved outcomes, and on the patient's pathway, is called 'Payment by results' (PbR) and is based on coded clinical information recordings. The process starts by the collection of information from hospital activity through coding of clinical notes to create the Commissioning Datasets (CDS). The CDS is then sent from the hospital to an external national data warehouse (Secondary Uses Service or SUS). Extracts from SUS are in turn used for different requirements such as PbR and Hospital Episode Statistics (HES). While PbR is a system of paying healthcare providers a standard national price or tariff for each patient seen or treated, the HES in England, and similarly the Scottish Morbidity Record (SMR), is designed to enable secondary use for non-clinical purposes of this administrative data. Purposes such as epidemiology, health trends and service development, informing governments and monitoring quality of health care [14,15]. HES or SMR include data from all hospital admissions, ED attendances and outpatient attendances in secondary and tertiary care. Unlike PbR, HES is subjected to additional processing to clean and de-duplicate the data.

The health care funding distribution in Scotland differs from the rest of the UK. It is unique in that it is not tariff based. The Scottish National Health System (NHS Scotland) consists of 14 regional NHS Boards which are responsible for the protection and improvement of their population's health and their healthcare provision. Funding to each NHS Board is allocated based on weighted capitation formulas that calculate target shares (percentages) for each NHS Board. The formula starts with the number of people resident in each NHS Board area, followed by formula adjustments for the age/sex profile, their additional needs based on morbidity and life circumstances (such as deprivation) and the excess costs of providing services in different geographical areas. The target shares calculated by the formula are further modified by way of 'differential growth', whereby all Boards would continue to enjoy real-terms growth in their allocations year-on-year based on their parity (i.e. above or below their formula target share the previous year) [15].
This means that Scotland NHS does not use coded data for direct funding purposes as in the rest of the UK. It can therefore be implied that the interest of codes accuracy in Scotland is not of financial interest, but mainly for secondary non-clinical purposes such as research and healthcare planning.

Healthcare planning and outcome management

Beyond funding and reimbursement, data from clinical coding provides crucial information for resource distribution and healthcare organization [16]. Healthcare costs are rising, partially due to an aging and growing population, but also driven by developments in medical technologies associated with increase in the cost of care-providing. In order to guarantee quality and safety of care, appropriate planning and distribution strategies for use of limited resources are required. As an example, thrombectomy, a highly effective stroke treatment, is not available in all hospitals in Scotland, or at all times of the day or week. It is a time-sensitive procedure, being most effective and relevant in the first few hours of stroke presentation. However, this service must have adequate specialised staffing and diagnostic resources to ensure its availability. Epidemiological data of local and global incidence patterns of stroke onset, extracted from discharge codes, guides the planning of this service availability and distribution. It is therefore imperative that diagnostic coding be accurate and of high quality to guarantee patient safety and best possible care through adequate planning and resource management.

Research

Equally important, coded data is harvested for research use, for it is readily available and relatively inexpensive to acquire. It is used to identify large groups of patients or samples with defined inclusion and exclusion criteria, report the incidence of complications, group populations according to their diagnoses and more. Moreover, using administrative data from clinical coding in Electronic Health Records (EHRs) for longitudinal research, provides many potential advantages for risk prediction and follow up [17–20].

The validity of research findings based on such data depends however on the accuracy and reliability of this data. For example, Frolova et al. demonstrated that acute heart failure (AHF) diagnoses codes from ED have high positive predictive value and thus could be used in outcomes research to establish cohorts of AHF [21]. Similarly, a Californian study, suggested that when a specific strategy to extrapolate administrative data is used, ICD coded data are sufficient for identification of ischaemic stroke hospitalisations [22]. Therefore, such data is possibly also adequate for stroke population sampling for research. On the other hand, fewer promising findings were observed when Burles et al. aimed to evaluate pulmonary embolism (PE) coding accuracy and validity for the use of research. The study revealed that a subset of a group of patients identified using PE codes on ED discharge were not actually diagnosed with PE. Hence, PE codes were incorrectly assigned in some cases, demonstrating a total of 17.7 % of false positive when sampling by diagnostic codes. This means that there is a potential threat to validity of PE studies or initiatives that rely on this administrative data

[23]. In the context of this study, accuracy of PE diagnosis was not measured against imaging but rather reviewed cases assessing whether ICD-10 code assigned by professional coders was congruent with the physician's discharge diagnosis. Most false positives in the study were cases in which the physician's written discharge diagnosis was unclear, such as "query PE" or " rule out PE". In another study, accuracy of ICD-10 coding for bleeding events in anticoagulated patients was assessed. The methodology involved reviewing medical charts to determine the presence or absence of bleeding. The results indicated high sensitivity and a negative predictive value (91.4% and 98.9%). However, the positive predictive value was unacceptably low, standing at a mere 52.5%.[24]. Erroneously, a large number of bleeding events were recorded (false positives), making this data unreliable for the identification of bleeding complications. Furthermore, Fleet at al. recoded a low sensitivity of hyperkalaemia diagnosis coding when compared to the laboratory results of every single patient. Thus, implying the incidence of true hyperkalaemia was underestimated by coded data, making such data inadequate for research use [25]. Overall, based on what is currently known when research is based on administrative data, a high degree of caution should be applied.

In the UK, growing numbers of researchers from a variety of specialties indeed rely on HES and SMR data for use in their research, indicating its perceived value [26]. Furthermore, coded clinical events from HES are included in the UK BIOBANK (UKB), a large-scale biomedical database and research resource [27]. However, there are some discrepancies regarding the accuracy of HES and SMR data for research use [28]. The reason for that is probably multifactorial and still under study. It is thought that some variability is probably related to a specific disease or the outcome assessed. For example, Holt et al. found that while HES data contains inaccuracies through incorrect or incomplete coding, it is still sufficiently accurate regarding hospital related death recording, suggesting it as an appropriate source for studying mortality [29]. An additional example is a study by Rannikm̈ae et al. suggesting that stroke cases can be ascertained in UKB through linked coded data with sufficient accuracy for use in many genetic and epidemiologic research studies without further expert validation [30].

Communication

"The difference between the almost right word and the right word is the difference between the lightning bug and the lightning" [Mark Twain]. Using a common language and terminology throughout the healthcare system reduces ambiguity and increases precision.

As super specialised medicine and hospital medicine expands, the transfer of responsibility for patient care between hospital-based physicians and primary care physicians, as well as between specialists, is a critical point in the patient's care pathway. Communicating timely, accurately and efficiently are of paramount importance in guaranteeing correct and best quality continuity of care [31]. As so, the use of common standardised terminology and accurate coding to record clinical data, promotes better handover of patient care within the healthcare system transferring information in an effective and precise way [32].

1.3 Accuracy

The vast use and application of data from clinical coding makes its accuracy highly important and relevant. Code accuracy is influenced by the many different steps along a dynamic interplay between the patient as he or she progresses through the health care and the creation of the medical record, followed by its interpretation [10]. These many steps and participants mean, that the coding process is subjected to numerous opportunities for errors and inaccuracies [9–11,13]. For this reason, coding accuracy has been a subject of many studies addressing its different aspects and strategies. Interestingly, these studies record a very wide variability of accuracy levels between studies, depending on coding method, classification set used, specialty, specific disease or diagnosis, setting, coder-training, method of measuring accuracy and more [9,28,33–35].

The impact of coding strategy on accuracy levels

The translation process itself, in which professional coders interpret clinical notes into diagnostic codes, carries many of its potential errors [33]. The quality of original clinical documentation is key in determining level of accuracy of codes assigned [12]. Incomplete and disorganized clinical documentation, as well as lack of good communication with clinicians, has been demonstrated to have a significant impact on the quality of clinical coding [13,36–38]. Additionally, as coding practice differs greatly between countries and even hospitals, the type of medical documentation used to extract the codes from are different, impacting on levels of accuracy. Tsopra et al. found that the accuracy of diagnosis codes improved when coders used either case notes or medical support, in addition to the discharge summary alone [9]. Reviewing case notes and discharge summaries may be however time consuming and inefficient. As an alternative, Walaraven et al. conducted a study where a clinical database was created from short forms completed by physicians. These forms were then used by clinical coders instead of reviewing clinical notes to extract codes, demonstrating improved accuracy [39]. Furthermore, accuracy levels are greatly influenced by whether and how medical doctors were involved in the coding process [40]. Mahbubani et al. found that the involvement of a doctor led to coding changes in 55.3% of cases reviewed, improving the overall accuracy of the coding process [41]. Lastly, experience and expertise of the coder are essential for this process and may be responsible for the huge variance in accuracy of discharge coding.

Accuracy levels in different specialties or disease

Wide variability in clinical coding accuracy was also recorded across specialties with different diseases showing different coding accuracy ranging from below 20% to higher than 95% [16,23,42–53]. This is related to the diagnosis process and requirements for each single disease, as well as its prevalence, its different possible outcomes, specificity of its clinical presentation and level of clinical interpretation required for diagnosis. For example, Wab et al., noted that the diagnosis coding for influenza demonstrated high sensitivity, specificity, PPV and NPV for influenza diagnosis codes against laboratory results, when test results were available before discharge. However, the accuracy of coding for influenza was substantially lower with a sensitivity of only 32.7% for patients whose test results were not available at the time of discharge [54]. Similarly, pathologies that do not have specific dichotomic diagnostic test, or have complex diagnostic patterns such as in mental health pathologies, have very low diagnostic code accuracies [45].

How accuracy is measured

Lastly, it is crucial to acknowledge that the measurement of accuracy for routinely collected data can be approached in various ways and evaluated against different standards. One common method typically involves engaging external coders, with or without the assistance of a physician, to re-code clinical records. Subsequently, the newly assigned codes are compared to the original codes given by local coders for the same records. It is worth noting that in this context, the assessment of code accuracy revolves around comparing the diagnosis code with the diagnosis documented in the medical records, rather than relying on actual test results that confirm the diagnosis. Consequently, the evaluation primarily focuses on the accuracy of the coding process itself, which entails translating medical notes into codes, rather than directly assessing the true diagnosis of the patient compared to the diagnosis made by the physician.

Accuracy levels are frequently expressed by means of sensitivity and positive predictive value (especially for a single disease code) or as degree of agreement between external and original coders. Otherwise, in some studies, the significance of inaccuracy is measured by evaluating its impact on funding and reimbursement [35,41,55]. For example, determining the financial difference between the reimbursement based on the original codes and the reimbursement based on the new codes (by external coders) for the same group of patients. However, data inaccuracies that may have an important financial impact might still be adequate for some non-clinical purposes. Similarly, coding practice driven by financial impact in the first place may yield inaccurate data for epidemiology and research purposes [56]. It is therefore difficult to quantify the disparate impact of accuracy level measured on reimbursement versus research use, of the same data [28]. This observation is particularly relevant in the context of this study, since the Scottish health care funding is not based on administrative coded data. Therefore the interest in code accuracy is driven by research, epidemiological and healthcare planning purposes.

Accuracy of coding in the Emergency Department

Emergency care, which represents one of the largest volumes of patient activities, is distinct from inpatient care and ambulatory care in many aspects. Moreover, ED coding process is often different from other settings. The accuracy of emergency medicine diagnostic coding however is poorly studied with only few studies conducted in this setting [21,25,42,45,57]. The majority of studies assessing discharge coding accuracy examine hospitalised and ambulatory patients' diagnosis codes when assigned by professional coders [28]. As ED data handling and coding methods are often different from inpatient care and ambulatory care with subsequent different sources of errors [58], ED discharge codes accuracy is unlikely to be represented by such studies.

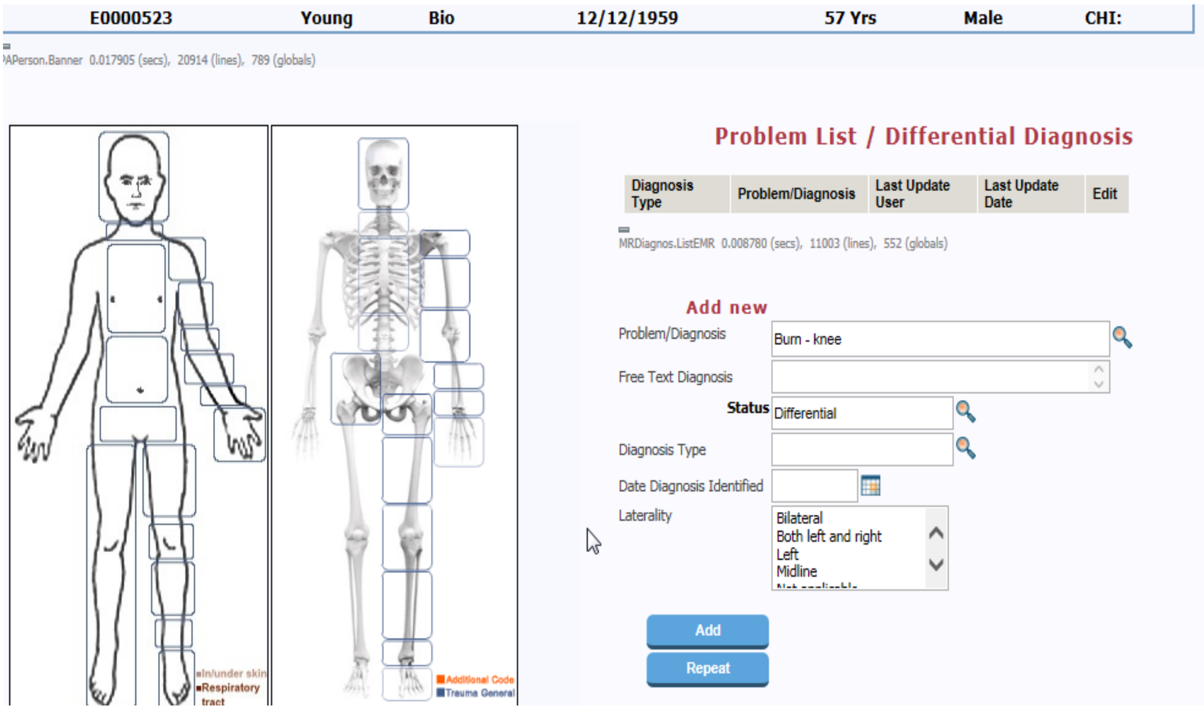
1.4 Study setting and background

Before December 2017, there was no '*Presenting Complaint*' or '*Discharge Diagnosis*' coding facility at the ED of the Royal Infirmary of Edinburgh (RIE). This changed as part of the NHS Lothian Annotated Bioresource Consortium (EMERGE Blood BioResource). This initiative allows disease specific excess blood samples, taken from patients at the earliest stages of their acute illness, to be stored under controlled conditions with robust quality assurance processes that protect and defend the samples' identification and integrity. Each sample can be linked to the patient's routinely collected hospital data, including ED presenting complaint and ED discharge diagnosis to generate linked phenotypicp data. Once surplus to clinical requirements, samples can then be fully anonymised and released to potential researchers.

A key part of the EMERGE Blood BioResource was the introduction of Presenting Complaint (PC) and Discharge Diagnosis code (DDC) to the software used in RIE ED (*TrakCare®*). This was designed, based on the Emergency Care Dataset v2.4 (ECDS)[7] and was launched in RIE on 7/12/2017 and then launched at St John's Hospital and the Royal Hospital for Sick Children on 7/3/2018.

In the RIE ED, diagnostic codes are assigned at time of discharge from the ED or admission to hospital, by different health care workers such as physicians, nurses, assistant physicians (AP) and advanced nurse practitioners (ANP). Upon discharge or admission, the individual responsible for assigning the diagnostic code, selects the most appropriate code from a predefined list presented in a Discharge Diagnosis questionnaire (Figure 1). The DDC list contains 945 different DDCs divided into different categories to facilitate proper code assignment. Overall, 74 different categories exist and they are based on either the body system affected or speciality, body district or mechanism of injury/pathology.
In this coding process no professional coders are involved. It is a point-of-care coding strategy, done immediately, directly by the team providing care for the patient who are not trained in coding or nosology. This process is similar across EDs in the UK and in some other countries, however, it is significantly different from the common coding practice in the inpatients and ambulatory care settings.

Figure 1: Discharge Diagnosis questionnaire screen presented to users upon ED discharge/admission



1.5 Aim of the present study

Discharge coding accuracy has a meaningful impact on different areas of health care. The current literature regarding ED discharge code accuracy is limited and inconclusive. Furthermore, accuracy of ED coded data assigned directly by healthcare providers who are non-coders, is poorly investigated [58]. As multiple variables impact accuracy levels, it is difficult to conclude the local ED coded data accuracy levels.

Therefore, the objective of our study is to assess the accuracy of discharge diagnosis codes (DDCs) in the Emergency Department of the RIE hospital. Specifically, we aim to assess the appropriateness of primary DDCs, selected and assigned by ED healthcare professionals, compared to the inferred discharge diagnosis derived from the clinical notes. Our aspiration is to establish the accuracy and validity of our administrative data, thereby enabling us to improve and upgrade research in areas where it is based on clinical coding data.

2. METHODS:

2.1 Study Design and Setting

This is a single-centre retrospective cohort study, conducted at the Royal Infirmary of Edinburgh (RIE) Emergency Department (ED). The facility is a large adult ED, which sees 125,000 patients per annum, within a university hospital in Scotland, UK.

2.2 Population

All patients presenting to the ED on Wednesday, May 3rd, 2023.

<u>Inclusion criteria:</u>

- Adult patients (≥16 years old)
- Attending to the ED or to the Acute Medicine team based in ED
- Patient having an ED assessment prior to discharge or admission

<u>Exclusion criteria:</u>

- Patients **not** assessed in the ED, include cases where:
    - Patient did not wait to be seen.
    - Self-Discharge.
    - Visit Not Appropriate.
    - Dead on arrival.
    - Direct admission to a specialty.
    - Episode created in error.
    - Left before clinical assessment.
    - Redirected to alternative community service.
    - Redirected to own GP practice.
    - Attendance for specimen collection.
    - Attention to, or removal of surgical sutures or dressing.
    - Request of medical certificate.

2.3 Methodology

The ED software, *TrakCare®* (InterSystems) was interrogated to obtain a list of all consecutive patients who presented to the ED in a whole 24-hour period, from 00:00 to 23:59, on May 3rd 2023. For each patient, data was extrapolated electronically from *TrakCare®* including: presenting time, age, sex, presenting complaint, primary discharge diagnosis code (DDC), secondary discharge diagnosis code if present, health professional assigning DDC, ED clinical notes, care provider and final destination (Admission/ Discharge/ Observation unit).
All patients were then screened to ensure they met inclusion/exclusion criteria.

The clinical notes were then reviewed and thoroughly examined by a single clinician ('the investigator'). The clinician meticulously identified and determined the primary diagnosis as well as any additional diagnoses mentioned within the clinical notes. Subsequently, these diagnoses were carefully assessed against assigned primary DDC. The reviewing clinician evaluated the appropriateness and accuracy of the chosen primary DDC compared to the primary diagnosis inferred from the care provider's clinical notes. Accuracy level of the primary DDC was measured using a scoring system specifically designed for this study; A categorical scale for accuracy levels based on the appropriateness level of the allocated DDC, as presented in Table 1.

Table 1: *categorical scale for primary discharge diagnosis code accuracy level.*

Accuracy level	Category description
1	**Most appropriate** discharge diagnosis code available (or one of the most appropriate diagnoses codes)
2	**Appropriate** discharge diagnosis code **but** more appropriate discharge diagnosis code available
3	**Appropriate** diagnostic code **category** but more appropriate discharge diagnosis code available - or - **Only partly appropriate** discharge diagnosis code with more specific/better discharge diagnosis code available
4	**Inappropriate** discharge diagnosis code, a much better appropriate discharge diagnosis code available
5	**Completely inappropriate** discharge diagnosis code

A second clinician reviewed a sample of 20 patients to assess agreement and validity of the scoring system used.

Additionally, presence of secondary discharge diagnosis implied from clinical notes were noted and corresponding secondary DDC recording rates were collected.

2.4 Data and statistical analysis

After manual chart reviews were completed, all patient-data were entered to a specially designed Microsoft excel database.

Descriptive statistics were used for patient demographics and reported as median and interquartile range. Categorical data was reported as n values and percentage.

VassarStats, an online calculator, was used for Chi-square test to assess differences between groups for categorical data and Yates value correction was applied when appropriate for continuity. Interrater-rater variability was assessed using Cohen's kappa coefficient.

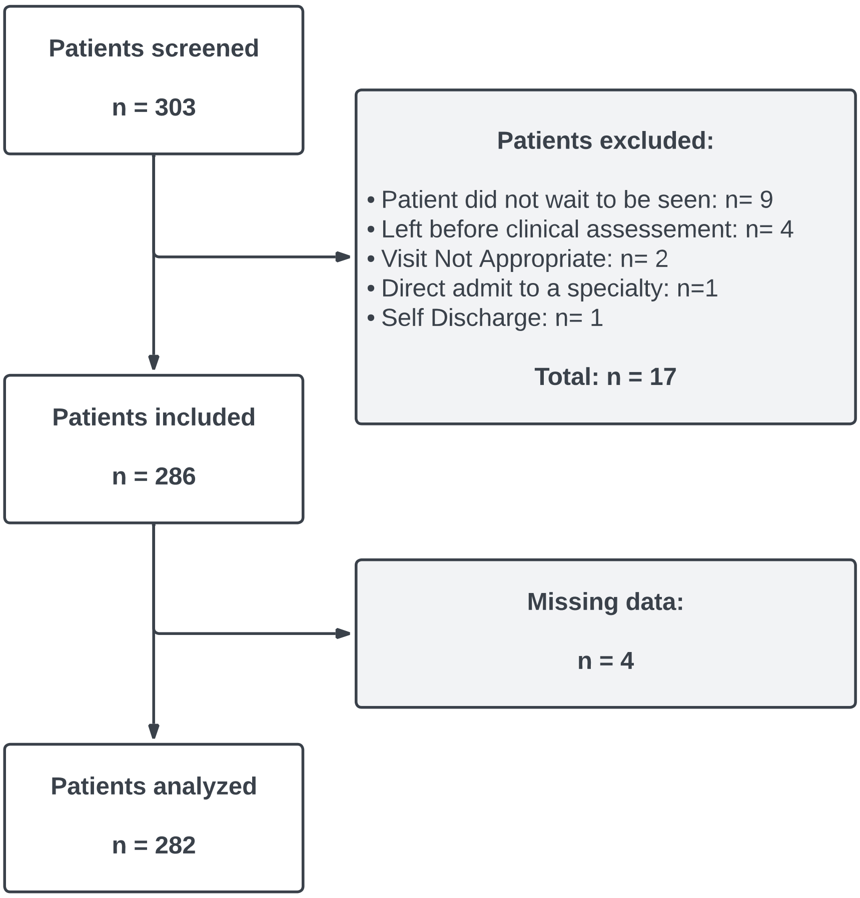
A P-value ≤ 0.05 was considered statistically significant.

3. RESULTS

3.1 Study population

In the course of the 24-hour period on May 3rd 2023, all records of 303 consecutive patients attending the ED were screened. Of these, 17 patients were excluded since they were not assessed in the ED (see study flow represented in Figure 2 for specific exclusion criteria). Additional 4 patients were excluded for missing data, since either DDC was not assigned or clinical notes were not recorded. In all, a total of 282 patients were analysed.
Among the 282 patients analysed, the median age was 52 years [IQR 33-72] and 140 (49.6%) were males. Most patients were assessed in the ED whereas only a small portion (8.5%; n=24) were assessed by the Acute Medicine team in the ED. At the end of the assessment, the majority of patients (n = 194; 68.8%) were discharged from the ED, while 21.1% (n = 60) of patients were admitted to the hospital and 9.9% (n = 28) of patients were transferred either to the ED observation unit or to the Surgical observation unit (n = 23 and n = 5 respectively).

Figure 2: *study flow*



Overall, the 20 most common presenting complaints (PCs) are represented in Table 2, accounting for the PCs in 76.8% (n = 217) of patients included in the study. The 20 most common primary discharge diagnosis codes (DDCs) in the study are represented in Table 3. Out of which, the three most common primary DDCs given were *'No abnormality detected'*, *'Non-specific chest pain'* and *'Non-specific abdominal pain'*. These three DDCs were assigned in slightly over a quarter (n = 75; 26.6 %) of the population sample. Collectively, a total of 119 different primary DDCs were recorded in the study population. These different primary DDCs account for 12.5% of the total 945 DDCs available in the ECDS list used in the ED.

Table 2: *The most common presenting complaints during the study period*

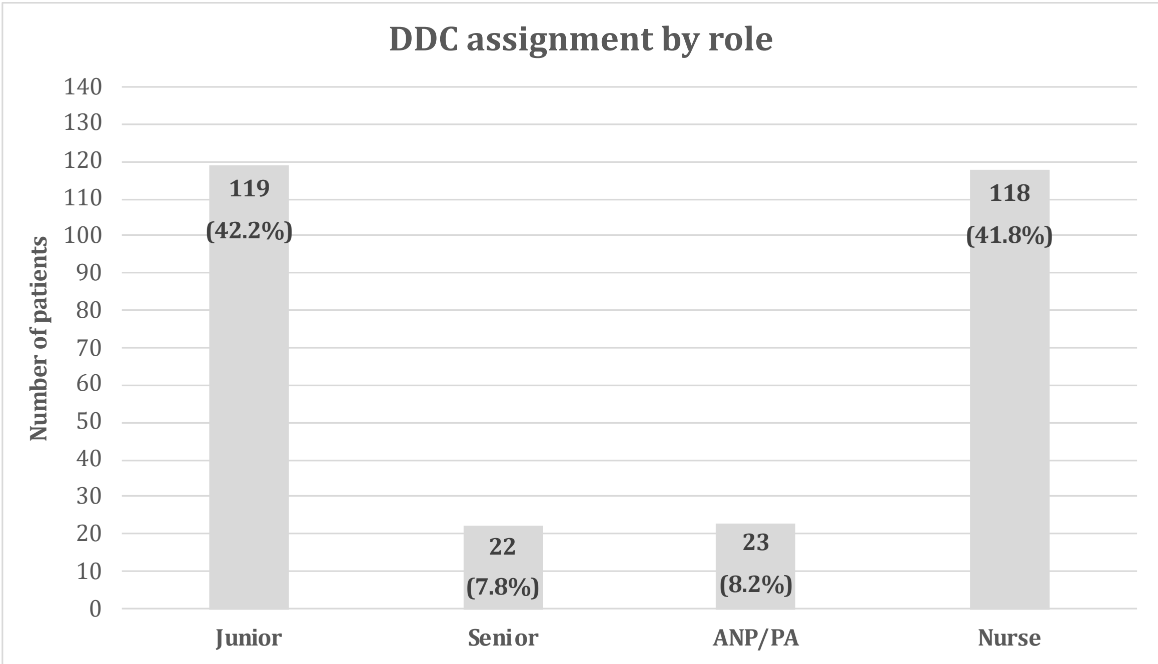
Most common presenting complaints	*n=*	*% (out of 282 patients)*
Chest pain	54	19.1
Abdominal pain	39	13.8
Short of breath / Difficulty breathing	18	6.4
Head injury	15	5.3
Pain in hip / leg / knee / ankle / foot	10	3.5
Injury of hip / leg / knee / ankle / foot	8	2.8
Localised swelling / redness / lumps / bumps	7	2.5
Palpitations	7	2.5
Backache (no recent injury)	6	2.1
Drug / alcohol intoxication / withdrawal	6	2.1
Vomiting +/- nausea	6	2.1
Frequent urination	5	1.8
Injury of back	5	1.8
Injury of shoulder / arm / elbow / wrist/ hand	5	1.8
Overdose	5	1.8
Seizure (fit)	5	1.8
Blood in stools	4	1.4
Fainting episode	4	1.4
Headache	4	1.4
Injury of neck	4	1.4
TOTAL	217	76.8

Table 3: *The most common primary discharge diagnosis codes during the study period*

Most common discharge diagnosis codes	*n=*	*% (out of 282 patients)*
No abnormality detected	27	9.6
Non-specific chest pain	26	9.2
Non-specific abdominal pain	22	7.8
Musculoskeletal chest pain	8	2.8
Urinary tract infection	7	2.5
Alcohol (ethanol) intoxication	6	2.1
Lower respiratory tract infection	6	2.1
Vasovagal syncope	6	2.1
Stomach: Muscle injury : lower back	5	1.8
Atrial fibrillation / flutter	4	1.4
Cauda equina syndrome	4	1.4
Cellulitis	4	1.4
Gastritis	4	1.4
Sepsis	4	1.4
Acute renal failure	3	1.1
Asthma	3	1.1
Bruise / contusion / abrasion : abdomen	3	1.1
Bruise / contusion / abrasion : head	3	1.1
Infectious gastroenteritis	3	1.1
Non specific headache	3	1.1
TOTAL	151	53.6

During the 24-hour period of the study, DDCs were assigned by 69 different healthcare professionals. The majority were junior doctors (n = 29; 42.1%), followed by nurses (n = 24; 34.8%), senior doctors (n = 9; 13%) and Advanced Nurse Practitioners or Physician Assistances (ANP/PA) (n = 7; 10.1%). In most cases, primary DDC was indeed assigned by junior doctors or nurses with the complete distribution of DDC assignment by professional role represented in figure 3.

Figure 3: *role distribution of the different health professionals assigning the primary DDC, presented as count and %.*



DDC- Discharge Diagnosis code, ANP- advanced nurse practitioners PA- assistant physicians, junior- Junior doctors, senior- Senior doctors

3.2 Primary Discharge Diagnosis Code Accuracy

Of the 282 patients' records reviewed and analysed, in 80.5% (n = 227) of cases, the DDC was considered to be either '*the most appropriate DDC'* (level 1) or '*appropriate DDC'* (level 2). A comprehensive distribution of the different DDC accuracy levels, as per the scale used in the study, is represented in Figure 4.

Interrater-rater variability was substantial (observed Cohen's kappa coefficient (κ) = 0.78 and observed as proportion of maximum attainable kappa (κ/κ_{max}) = 0.92).

For sample size reliability, cumulative percentage of level 1 accuracy for increasing sample size was calculated, as illustrated in Figure 5. Reaching a plateau with increasing number of patients beyond 29 patients with level 1 accuracy percentage ranging between 70%-75% steadily.

Figure 4: pie chart illustrating DDC accuracy levels, presented as count and %.

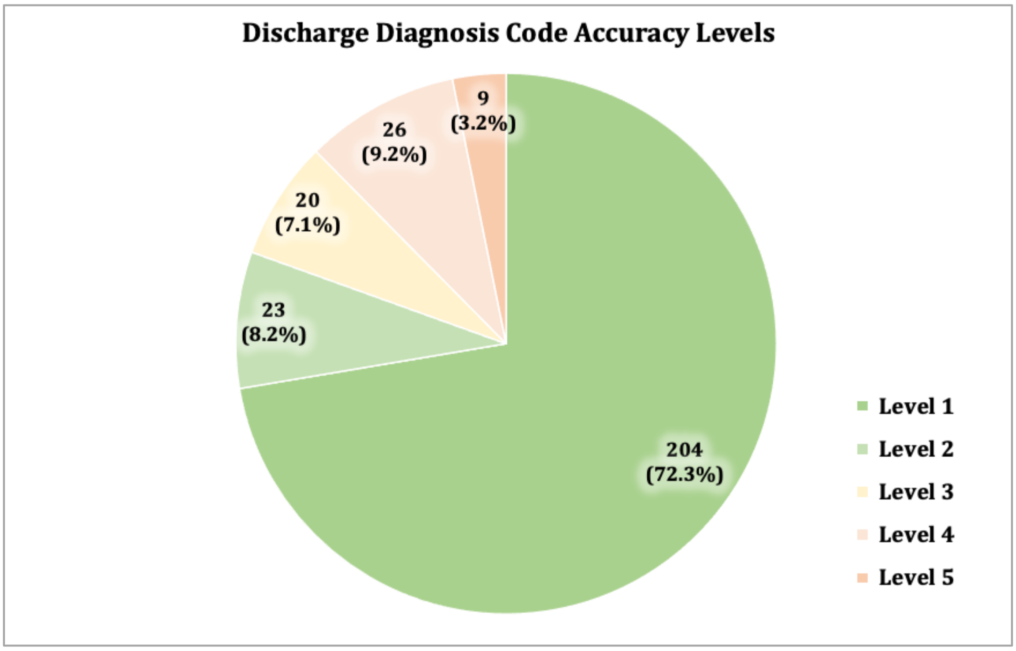
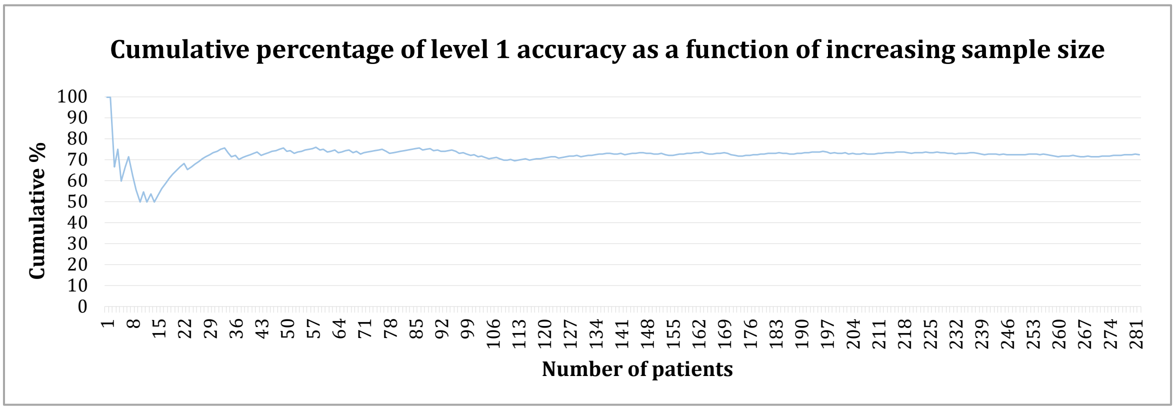


Figure 5: *a line plot of level 1 accuracy cumulative percentage expressing frequency distribution for increasing number of patients.*



Overall, in 131 (46.5%) of the cases, the health professional who was directly providing care to the patient and documenting their clinical notes, was also the individual assigning the primary DDC (referred to as 'same CP'). Within this particular group, a higher level of accuracy in DDC assignment was observed compared to cases in which the individual assigning the DDC was not the same person providing care and writing the clinical notes (referred to as 'not same CP'), p= 0.001, as illustrated in Figure 6. Furthermore, a statistically significant ($p<0.001$) higher level of accuracy was recorded when doctors or ANP/PA assigned DDC rather than nurses. However, when accuracy levels were specifically analysed only within the cases where the DDC was assigned by a different person than the one providing care ('not same CP'), no statistically significant accuracy level difference (p=0.377) was observed between nurses and the other healthcare workers roles assigning the DDC (see table 4).

No difference was found in accuracy level between admitted and discharged patients (p= 0.162).

Figure 6: *comparison of DDC accuracy levels between cases where the health professional responsible for providing care and documenting clinical notes was the same individual assigning the DDC (referred to as 'same CP') and cases where the DDC was assigned by a different individual (referred to as 'not same CP') presented as count (%). (CP- care provider)*

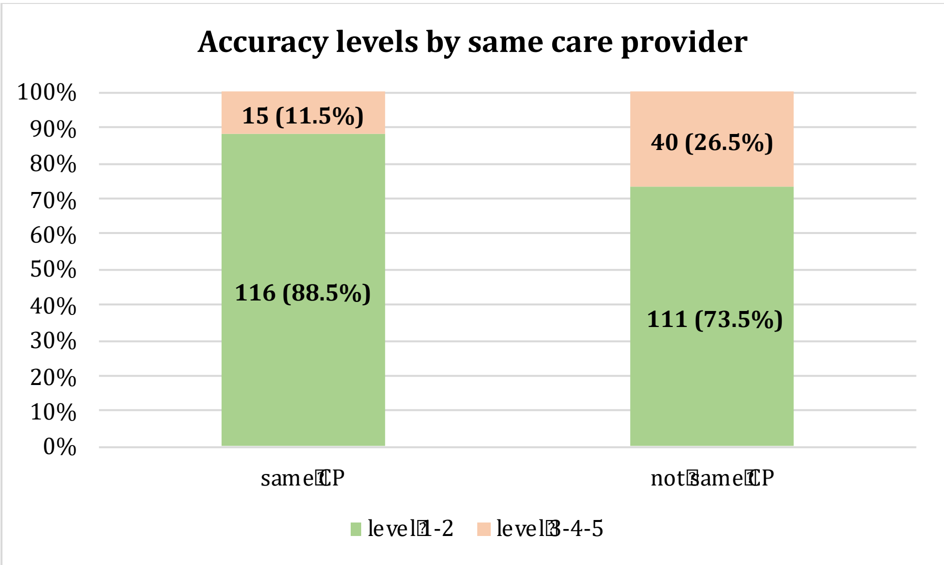


Table 4: *represents a comparison of DDC accuracy levels as a factor of the role of the different health professionals assigning the DDC. Both a comparison for all cases included in the study, as well as a comparison for only the cases where DDC was assigned by another individual ('not same CP') is presented with their respective statistical significance.*

	All cases				**Not same CP cases only**		
	Nurse n= (%) Total 118	**Doctors/ ANP/PA** n= (%) Total 164	**Chi square sig.** p =		**Nurse** n= (%) Total 115	**Doctors/ ANP/PA** n= (%) Total 36	**Chi square sig.** p =
level 1+2	83 (70%)	144 (88%)	p <0.001	**level 1+2**	82 (71.3%)	29 (80.5%)	p = 0.271
level 3+4+5	35 (30%)	20 (12%)		**level 3+4+5**	33 (28.7%)	7 (19.5%)	

Analysis of data revealed that when non-specific DDCs, such as *non-specific chest pain*, were grouped together, they exhibited a statistically significant higher frequency of low accuracy levels (scoring level 3, 4 or 5) when compared to the sum of all other DDCs in the study (p= 0.008), as per Table 5. There was no significant difference in the occurrence of non-specific DDCs assignment between the groups categorized as 'same CP' and 'not same CP' (p= 0.257).

***Table 5**: A two-by-two table comparing accuracy levels of non-specific DDCs together vs the sum of all other DDCs. Non-specific DDCs include: non-specific abdominal pain, non-specific chest pain, non- specific headache, no abnormality found. P value = 0.008.*

Accuracy level	**Non-specific DDCs** n= (%)	**Other DDCs** n= (%)	*Total*
level 1+2	55 (70.5%)	172 (84%)	227
level 3+4+5	23 (29.5%)	32 (16%)	55
Total	78	204	282

3.3 Secondary Discharge Diagnosis Code

Out of the 282 cases analysed, no secondary DDC were recorded. The clinician reviewing the clinical notes (the investigator), identified at least 79 patients (28%) with ≥1 possible secondary diagnosis inferred from the clinical notes.

4. DISCUSSION

4.1 Key findings

Our study assessed the accuracy level of our local ED administrative data, specifically examining ED discharge diagnosis codes accuracy.

We report 2 key findings:

1. Overall, the ED primary discharge diagnosis codes in the RIE hospital were found to have a high accuracy level of 81%.
2. The accuracy of DDCs were higher when the DDC was assigned by the same healthcare professional directly providing care to the patient.

4.2 Literature comparison and interpretation of key study findings

Currently, there is no consensus regarding the acceptable data accuracy level or what is the most appropriate way to assess it. Findings of our studies are consistent with Burns et al. systemic review, which demonstrated a wide variety of accuracy levels across the UK, reporting a median DDC accuracy of 80.3% (IQR: 63.3–94.1%) [28]. In this systemic review, however, no or very little ED data were included and coding strategies were largely different from the strategy used in our ED. In the context of ED, Peng et al. assessed accuracy levels in different Canadian emergency departments and found an ICD diagnosis coding agreement of 86.5% and 82.2% at 3 and 4 code digits levels respectively [57]. However, once again, our local coding strategy and set (ECDS) differs from the one performed in the EDs in the latter study, where coding was performed by professional coders, as in other ED clinical coding accuracy studies [23].

The comparison of our findings with previous literature is limited due to the unique coding method employed in our ED (and across most EDs in the UK), wherein clinicians perform coding at the time of ED discharge/admission. Despite the overall similarity of the findings with the presented literature, this coding approach restricts the extent of comparison.

The higher accuracy level seen when the same individual who provided the patient's care also coded the DDC could be attributed to the lack of a second person's interpretation of clinical notes. Indeed, within the cases of 'no same CP' where DDCs assigned by a different person than the direct care provider, lower accuracy levels were observed when the clinical notes were of poor quality or incomplete (diagnostic impression was not clearly stated or not updated after investigation results). Similarly, many studies demonstrated how poor, vague or incomplete clinical documentation has led to lower DDC accuracy when coded by professional coders [10,13,33,36–38].

We believe that the lower accuracy levels recorded for 'non-specific DDCs' could be explained by their inherently low accuracy and specificity, being non-specific entities harbouring several potential possibilities. These 'non-specific DDCs' are clinically driven DDCs which were added to the adapted ECDS code set used in our ED, to match ED practice needs. Such DDCs are crucial and inevitable in the ED, where the aim is not to diagnose all types of diseases but rather exclude life threatening and severe diseases.

4.3 Strengths and limitations

Our study has several strengths. It is a retrospective cohort study, with an unbiased sampling technique, screening all consecutive cases of a random predefined 24h period of a mid-weekday. Users assigning codes were not aware of assessment of coding accuracy hence excluding possible performance bias. Although this is a single centre study, limiting its generalizability, the RIE ED follows the same coding guidelines and practices as other EDs across the UK [7]. Therefore, our overall emergency department data quality could possibly be representative for other UK EDs. Furthermore, no financial bias on determining accuracy level is present in this study and accuracy level assessed in this study is not measured in terms of reimbursement impact.
This study has the following limitations. Not all possible DDCs were coded during the study period, and most DDCs that were represented had a low frequency, suggesting possibly small sample size. However, the observation that the frequency of level 1 accuracy for increasing sample size, calculated as cumulative percentage, reached a plateau after 30 patients, suggests that our sample size of 282 cases was more than adequate to generate an estimation of general ED DDC accuracy. Another limitation of our study is related to the fact that 'the

investigator', the clinician reviewing the clinical notes and assessing DDC accuracy, has no formal training in coding. Hence, possible introduction of errors when assessing coding accuracy exists. An additional bias may have originated from the fact that the reviewing clinician, was not blinded to the DDC given before determining the inferred diagnosis from the clinical notes and its corresponding most appropriate DDC available. Lastly, observer reliability was tested solely on 20 cases rather than on the whole sample and by only one additional rater.

4.4 Future perspectives and implications

This study aimed to assess our administrative data accuracy, i.e., how well the assigned discharge diagnosis code reflects the discharge diagnosis in the clinical documents. The validity of the actual diagnosis was not assessed in this study, i.e., how well the diagnosis established by the care provider reflects the true diagnosis of the patient. Therefore, the question in hand is whether our health professionals are adequately coding rather than adequately diagnosing.

Based on the observations of this research, we propose several possible strategies which might improve coding accuracy in our settings;

Firstly, positive reinforcement strategies should be used to encourage DDC assignment directly by the treating care provider and not by another individual.

Secondly, basic coding practice training and DDC set revision might be effective in improving accuracy levels [11].

Furthermore, a possible reason for inaccuracies has been attributed to professional's lack of interest in clinical coding [38], therefore, brief educational sessions to promote awareness of coding importance and impact should be provided to staff assigning codes.

Lastly, no secondary diagnoses codes were recorded during our study. We believe that raising general awareness for DDC significance and training, will also improve secondary DDC recording rate.

Our study provides distinctive and important novel data regarding the accuracy of ED DDCs with the specific coding strategy used across EDs in the UK and other countries. Nevertheless, to provide for better data consistency and meta-analysis, further research is necessary from multiple centres alongside larger sample sizes. This should ideally be followed by trials of best accuracy improvement strategies for ED coding, especially when the DDC is assigned by non-coders.

5. CONCLUSION

RIE hospital's ED administrative data from discharge diagnosis codes has a substantial degree of accuracy of 80.5%. Accuracy levels were improved by 20% when the same health professional, who was directly providing care to the patient, was the individual assigning the discharge diagnosis code.

We hope this study will contribute to the comprehension and improvement of administrative data quality from the ED, eventually leading naturally, to better and safer patient care.

6. BIBLIOGRAPHY

1. International Classification of Diseases (ICD). https://www.who.int/standards/classifications/classification-of-diseases.

2. Harrison, J. E., Weber, S., Jakob, R. & Chute, C. G. ICD-11: an international classification of diseases for the twenty-first century. *BMC Med Inform Decis Mak* **21**, (2021).

3. Masys, D. R. Of codes and keywords: standards for biomedical nomenclature. *Acad Med* **65**, 627–9 (1990).

4. SNOMED CT - NHS Digital. https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct.

5. SNOMED CT - Home. https://browser.ihtsdotools.org/?

6. Overview of SNOMED CT. https://www-nlm-nih-gov.insubria.idm.oclc.org/healthit/snomedct/snomed_overview.html.

7. Emergency Care Data Set (ECDS) - NHS Digital. https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/emergency-care-data-set-ecds.

8. 5.18 Data Group: Emergency Care Diagnosis (SNOMED CT) - NHS Digital. https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-sets/emergency-care-data-set-ecds/user-guide/data-group-emergency-care-diagnosis-snomed-ct.

9. Tsopra, R. *et al.* Level of accuracy of diagnoses recorded in discharge summaries: A cohort study in three respiratory wards. *J Eval Clin Pract* **25**, 36–43 (2019).

10. O'Malley, K. J. *et al.* Measuring Diagnoses: ICD Code Accuracy. *Health Serv Res* **40**, 1620–1639 (2005).

11. Lorenzoni, L., Da Cas, R. & Aparo, U. L. The quality of abstracting medical information from the medical record: the impact of training programmes. *Int J Qual Health Care* **11**, 209–213 (1999).

12. Alonso, V. *et al.* Health records as the basis of clinical coding: Is the quality adequate? A qualitative study of medical coders' perceptions. *Health Information Management Journal* **49**, 28–37 (2020).

13. Doktorchik, C., Lu, M., Quan, H., Ringham, C. & Eastwood, C. A qualitative evaluation of clinically coded data quality from health information manager perspectives. *Health Information Management Journal* **49**, 19–27 (2020).

14. Hospital Episode Statistics (HES) - NHS Digital. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics.

15. National Data Catalogue | National Datasets | ISD Scotland | Information Services Division. https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=5.

16. Nouraei, S. A. R. *et al.* Accuracy of clinician-clinical coder information handover following acute medical admissions: implication for using administrative datasets in clinical outcomes management. *J Public Health (Oxf)* **38**, 352–362 (2016).

17. Evans, E., Grella, C. E., Murphy, D. A. & Hser, Y. I. Using Administrative Data for Longitudinal Substance Abuse Research. *J Behav Health Serv Res* **37**, 252 (2010).

18. Roos, L. L., Nicol, J. P. & Cageorge, S. M. Using administrative data for longitudinal research: comparisons with primary data collection. *J Chronic Dis* **40**, 41–49 (1987).

19. Goldstein, B. A., Navar, A. M., Pencina, M. J. & Ioannidis, J. P. A. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* **24**, 198–208 (2017).

20. Darke, P. *et al.* Curating a longitudinal research resource using linked primary care EHR data-a UK Biobank case study. *J Am Med Inform Assoc* **29**, 546–552 (2022).

21. Frolova, N. *et al.* Assessing the use of international classification of diseases-10th revision codes from the emergency department for the identification of acute heart failure. *JACC Heart Fail* **3**, 386–391 (2015).

22. Zachrison, K. S. *et al.* Strategy for reliable identification of ischaemic stroke, thrombolytics and thrombectomy in large administrative databases. *Stroke Vasc Neurol* **6**, 194–200 (2021).

23. Burles, K., Innes, G., Senior, K., Lang, E. & McRae, A. Limitations of pulmonary embolism ICD-10 codes in emergency department administrative data: Let the buyer beware. *BMC Med Res Methodol* **17**, 1–8 (2017).

24. Joos, C., Lawrence, K., Jones, A. E., Johnson, S. A. & Witt, D. M. Accuracy of ICD-10 codes for identifying hospitalizations for acute anticoagulation therapy-related bleeding events. *Thromb Res* **181**, 71–76 (2019).

25. Fleet, J. L. *et al.* Validity of the International Classification of Diseases 10th revision code for hyperkalaemia in elderly patients at presentation to an emergency department and at hospital admission. *BMJ Open* **2**, e002011 (2012).

26. Chaudhry, Z. *et al.* Research outputs of England's Hospital Episode Statistics (HES) database: a bibliometric analysis. *BMJ Health Care Inform* **24**, 329–333 (2017).

27. UK Biobank - UK Biobank. https://www.ukbiobank.ac.uk/.

28. Burns, E. M. *et al.* Systematic review of discharge coding accuracy. *J Public Health (Oxf)* **34**, 138–148 (2012).

29. Holt, P. J. E., Poloniecki, J. D. & Thompson, M. M. Multicentre study of the quality of a large administrative data set and implications for comparing death rates. *British Journal of Surgery* **99**, 58–65 (2012).

30. Rannikmäe, K. *et al.* Accuracy of identifying incident stroke cases from linked health care data in UK Biobank. *Neurology* **95**, E697–E707 (2020).

31. Kripalani, S. *et al.* Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. *JAMA* **297**, 831–841 (2007).

32. Vuokko, R., Vakkuri, A. & Palojoki, S. Systematized Nomenclature of Medicine–Clinical Terminology (SNOMED CT) Clinical Use Cases in the Context of Electronic Health Record Systems: Systematic Literature Review. *JMIR Med Inform* **11**, (2023).

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

34. Quan, H. *et al.* Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* **43**, 1424–1441 (2008).

35. *The quality of clinical coding in the NHS*. (2014).

36. Lloyd, S. S. & Rissing, J. P. Physician and Coding Errors in Patient Records. *JAMA* **254**, 1330–1336 (1985).

37. Tang, K. L., Lucyk, K. & Quan, H. Coder perspectives on physician-related barriers to producing high-quality administrative data: a qualitative study. *CMAJ Open* **5**, E617 (2017).

38. Williams, J. G. & Mann, R. Y. Hospital episode statistics: time for clinicians to get involved? *Clin Med (Lond)* **2**, 34–37 (2002).

39. Van Walraven, C. & Demers, S. V. Coding diagnoses and procedures using a high-quality clinical database instead of a medical record review. *J Eval Clin Pract* **7**, 289–297 (2001).

40. Yeoh, C. & Davies, H. Clinical coding: completeness and accuracy when doctors take it on. *BMJ : British Medical Journal* **306**, 972 (1993).

41. Mahbubani, K. *et al.* Clinician-directed improvement in the accuracy of hospital clinical coding. *Future Healthc J* **5**, 47 (2018).

42. Yu, A. Y. X. *et al.* Moderate sensitivity and high specificity of emergency department administrative data for transient ischemic attacks. *BMC Health Serv Res* **17**, 1–7 (2017).

43. Kirkman, M. A., Mahattanakul, W., Gregson, B. A. & Mendelow, A. D. The accuracy of hospital discharge coding for hemorrhagic stroke. *Acta Neurol Belg* **109**, 114–119 (2009).

44. Petersen, L. A., Wright, S., Normand, S. L. T. & Daley, J. Positive predictive value of the diagnosis of acute myocardial infarction in an administrative database. *J Gen Intern Med* **14**, 555–558 (1999).

45. Vandyk, A. D., Graham, I. D., Vandenkerkhof, E. G. & Harrison, M. B. Accuracy of ICD Codes for Persons with Considerable Emergency Department Use for Mental Health Complaints. *International Journal of Emergency Mental Health and Human Resilience 2015 17:3* **17**, 1–2 (1970).

46. Reddy-Kolanu, G. R. & Hogg, R. P. Accuracy of clinical coding in ENT day surgery. *Clin Otolaryngol* **34**, 405–407 (2009).

47. Ratelle, S. *et al.* Predictive value of clinical diagnostic codes for the CDC case definition of pelvic inflammatory disease (PID): implications for surveillance. *Sex Transm Dis* **30**, 866–870 (2003).

48. Nouraei, S. A. R. *et al.* A Study of Clinical Coding Accuracy in Surgery: Implications for the Use of Administrative Big Data for Outcomes Management. *Ann Surg* **261**, 1096–1107 (2015).

49. Anaya, D. A., Becker, N. S., Richardson, P. & Abraham, N. S. Use of administrative data to identify colorectal liver metastasis. *J Surg Res* **176**, 141–146 (2012).

50. Chong, W. F., Ding, Y. Y. & Heng, B. H. A comparison of comorbidities obtained from hospital administrative data and medical charts in older patients with pneumonia. *BMC Health Serv Res* **11**, (2011).

51. Hussey, K., Siddiqui, T., Burton, P., Welch, G. H. & Stuart, W. P. Understanding administrative abdominal aortic aneurysm mortality data. *Eur J Vasc Endovasc Surg* **49**, 277–282 (2015).

52. Kokotailo, R. A. & Hill, M. D. Coding of Stroke and Stroke Risk Factors Using International Classification of Diseases, Revisions 9 and 10. *Stroke* **36**, 1776–1781 (2005).

53. Hall, R., Mondor, L., Porter, J., Fang, J. & Kapral, M. K. Accuracy of administrative data for the coding of acute stroke and TIAs. *Canadian Journal of Neurological Sciences* **43**, 765–773 (2016).

54. Wabe, N. *et al.* Evaluation of the accuracy of diagnostic coding for influenza compared to laboratory results: the availability of test results before hospital discharge facilitates improved coding accuracy. *BMC Med Inform Decis Mak* **21**, 1–10 (2021).

55. Cheng, P., Gilchrist, A., Robinson, K. M. & Paul, L. The risk and consequences of clinical miscoding due to inadequate medical documentation: a case study of the impact on health services funding. *Health Inf Manag* **38**, 35–46 (2009).

56. Derby, C. A., Lapane, K. L., Feldman, H. A. & Carleton, R. A. Possible Effect of DRGs on the Classification of Stroke : Implications for Epidemiological Surveillance. *Stroke* **32**, 1487–1491 (2001).

57. Peng, M. *et al.* Coding reliability and agreement of International Classification of Disease, 10th revision (ICD-10) codes in emergency department data. *Int J Popul Data Sci* **3**, 14 (2018).

58. Dickson, J. M., Mason, S. M. & Bailey, A. Emergency department diagnostic codes: useful data? *Emergency Medicine Journal* **34**, 627–627 (2017).