

MESTRADO INTEGRADO EM MEDICINA

2020/2021

Roberto da Silva Carvalho

Análise de causas de problemas que afetam a qualidade de dados hospitalares baseados em Grupos de Diagnósticos Homogêneos: revisão sistemática e diagrama Ishikawa / Analysis of root causes of problems affecting the quality of the Diagnosis Related Group-based hospital data: a systematic review and Ishikawa diagram

Março, 2021

Roberto da Silva Carvalho

Análise de causas de problemas que afetam a qualidade de dados hospitalares
baseados em Grupos de Diagnósticos Homogêneos: revisão sistemática e diagrama
Ishikawa / Analysis of root causes of problems affecting the quality of the Diagnosis
Related Group-based hospital data: a systematic review and Ishikawa diagram

Mestrado Integrado em Medicina

Área: Ciências da Saúde

Tipologia: Dissertação

Trabalho efetuado sob a Orientação de:

Dr. Alberto Freitas

E sob a Coorientação de:

Dr. João Vasco Santos

Trabalho organizado de acordo com as normas da revista:

International Journal of Medical Informatics

Março, 2021

Eu, Roberto da Silva Carvalho, abaixo-assinado, nº mecanográfico 201403572, estudante do 6º ano do Ciclo de Estudos Integrado em Medicina, na Faculdade de Medicina da Universidade do Porto, declaro ter atuado com absoluta integridade na elaboração deste projeto de opção.

Neste sentido, confirmo que **NÃO** incorri em plágio (ato pelo qual um indivíduo, mesmo por omissão, assume a autoria de um determinado trabalho intelectual, ou partes dele). Mais declaro que todas as frases que retirei de trabalhos anteriores pertencentes a outros autores, foram referenciadas, ou redigidas com novas palavras, tendo colocado, neste caso, a citação da fonte bibliográfica.

Faculdade de Medicina da Universidade do Porto, 25/03/2021

Assinatura conforme cartão de identificação:

Roberto da Silva Carvalho

NOME

Roberto da Silva Carvalho

NÚMERO DE ESTUDANTE

201403572

E-MAIL

robertocarvalho.06@gmail.com

DESIGNAÇÃO DA ÁREA DO PROJECTO

Ciências da saúde

TÍTULO DISSERTAÇÃO/~~MONOGRAFIA~~ (riscar o que não interessa)

Análise de causas de problemas que afetam a qualidade de dados hospitalares baseados em Grupos de Diagnósticos Homogêneos: revisão sistemática e diagrama Ishikawa

ORIENTADOR

José Alberto da Silva Freitas

COORIENTADOR (se aplicável)

João Vasco Nunes dos Santos

ASSINALE APENAS UMA DAS OPÇÕES:

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTES TRABALHOS APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.	☒
É AUTORIZADA A REPRODUÇÃO PARCIAL DESTES TRABALHOS (INDICAR, CASO TAL SEJA NECESSÁRIO, Nº MÁXIMO DE PÁGINAS, ILUSTRAÇÕES, GRÁFICOS, ETC.) APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.	☐
DE ACORDO COM A LEGISLAÇÃO EM VIGOR, (INDICAR, CASO TAL SEJA NECESSÁRIO, Nº MÁXIMO DE PÁGINAS, ILUSTRAÇÕES, GRÁFICOS, ETC.) NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTES TRABALHOS.	☐

Faculdade de Medicina da Universidade do Porto, 25 / 03 / 2021

Assinatura conforme cartão de identificação:

Roberto da Silva Carvalho

Analysis of Root Causes of Problems Affecting the Quality of the Diagnosis-Related Group-Based Hospital Data: A Systematic Review and Ishikawa Diagram

Roberto Carvalho¹, Mariana Lobo^{1,2}, Mariana Oliveira¹, Ana Raquel Oliveira¹, Fernando Lopes^{1,2}, Júlio Souza^{1,2}, André Ramalho^{1,2}, João Viana^{1,2}, Vera Alonso², Ismael Caballero⁴, João Vasco Santos^{1,2,3}, Alberto Freitas^{1,2}

1 – MEDCIDS – Department of Community Medicine, Information and Health Decision Sciences, Faculty of Medicine, University of Porto, Portugal

2 – CINTESIS - Centre for Health Technology and Services Research, Portugal

3 – Public Health Unit, ACES Grande Porto VIII – Espinho/Gaia, ARS Norte, Portugal

4 – Institute of Information Systems and Technologies (ITSI), University of Castilla-La Mancha, Ciudad Real, Castilla-La Mancha, Ciudad Real, Spain

Abstract:

Introduction: Administrative hospital databases built around Diagnosis-Related Groups (DRGs) represent an important tool for hospital financing in many national health systems and are also an important data source for clinical, epidemiological and health services research. Therefore, the data quality of such databases is of utmost importance. This paper aims to present a systematic review of root causes of data quality problems affecting DRG-based hospital data, creating a catalogue of potential issues for data quality analysts to explore.

Methods: The MEDLINE and Scopus databases were searched using inclusion criteria based on two following concept blocks: (1) administrative hospital databases and (2) data quality. Studies' titles and abstracts were screened by two reviewers independently. Three researchers independently selected the screened studies based on their full texts and then extracted the root causes. These were subsequently classified according to the Ishikawa model based on 6 categories: "Man", "Material", "Method", "Machine", "Mission" and "Management".

Results: The result of our investigation and the contribution of this paper is a classification of the potential (105) root causes found through a systematic review of the 77 relevant studies we have identified and analyzed. The result was represented by an Ishikawa diagram. Most (25.7%) of the root causes were associated with the category “Man” – people’s knowledge, preferences, education and culture, mostly related to clinical coders and health care providers activities. The quality of hospital documentation, within category “Material”, and aspects related to financial incentives or disincentives, within category “Mission”, were also frequently cited in the literature as relevant root causes for data quality issues.

Conclusions: The resultant catalogue of root causes, systematized using the Ishikawa framework, provides a causal ground for data quality issues to be considered when reusing these data but also points out for possible data quality assessments, which can be further used as a tool to delineate actions to address and mitigate existing issues.

Keywords: Data quality; Data quality issue root cause; Diagnosis-related group; Administrative hospital data; Clinical coding;

1. Introduction

First developed in 1970s, the concept of Diagnosis-Related Groups (DRGs) has been progressively adopted since the 1990s in national health systems around the world for hospital financing, turning DRG-based administrative hospital databases into an important tool for hospital planning and budgeting, and for clinical, epidemiological and health services research. [1, 2].

These DRG-based databases usually have a nationwide coverage and contain comprehensive information, including demographic, clinical, administrative and reimbursement hospitalization data. The information results from the abstraction of structured and unstructured information documented in clinical records (including discharge summaries, surgical reports, pathology and image studies) and administrative records covering each patient’s episode of care [3].

The clinical information gathered and registered during the episode of care is classified using clinical classification systems, such as International Classification of Disease 10th Revision (ICD-10), which, along with demographic data, is used to group episodes of care into DRGs - coherent groups regarding clinical terms and resources utilization.

For a number of reasons (errors, lack of standards, financial interests) fallible data can be included in these types of databases, eliciting data quality problems. This can potentially compromise hospital financing/reimbursement and their reutilization for epidemiological or health services research [4-6].

Within the research context, this source of data entails enormous potential. In fact, it is systematically collected, encompassing a large population and a broad geographic coverage. Sometimes, it is the only feasible source available to study specific subgroups or to allow international comparisons. It may also provide important insights that can be used to improve the quality of care provided to patients, reducing access disparities, improving patient outcomes, and better allocating resources [7]. To maximize the application of these data in view of the much-appreciated potential, the highest level of quality must be pursued, considering the perspective of the different stakeholders. Thus, we follow the definition of “*fitness for use*” of data quality in this work. This definition emphasizes the users’ viewpoint and intended application of the data in a given context of use, while requiring the judgment of various dimensions of quality for a given dataset [8].

There are numerous studies identifying data quality issues in hospital administrative databases, mainly related with accuracy problems [9-11]. However, as data quality goes beyond the accuracy or completeness dimensions, other dimensions not usually considered, such as credibility, currentness and consistency, should also be addressed when it comes to identify root causes of data quality-related problems in the healthcare services domain, an area where a systematic compilation and analysis is lacking. We felt motivated to conduct an investigation to bridge this gap, aiming to systematically identify causes of problems affecting the quality of DRG-based administrative hospital data, whether these relate with hospital financing, epidemiological or health services research, and to propose a framework to derive and analyze a list of root causes.

2. Methods

2.1 Protocol

This study implemented a systematic review, following a previously published PRISMA-based protocol [12, 13].

2.2 Information sources and search strategy

The MEDLINE and Scopus scientific research databases have been searched to identify potentially relevant studies reporting causes of data quality problems of DRG-based administrative hospital databases. The search was performed on studies published from 1990 to 2019, given that this attests to some degree of quality of the studies and reflects the period when this type of data had a wide implementation around the world, and it was concluded on 31st September 2019. The drafted search strategy was based on two concept blocks: (1) administrative hospital databases and (2) data quality. The final search strategy for Medline can be found as supplementary material (appendix A, Table A1). The search results have been exported into EndNote X9.2, where most of the duplicates were handled.

2.3 Eligibility criteria

The studies were subsequently screen according to the following inclusion criteria:

- Peer-reviewed studies (including conference papers), written in English.
- Studies referring to data quality of DRG-based administrative hospital databases. This might include inpatient or outpatient hospital episodes and systematic or random failures resulting in data quality issues of these databases, including those originated in processes regarding data generation/acquisition, processing (e.g. clinical coding), storage, and utilization of these databases and their data sources (e.g. medical records), or encompassing the different perspective of producers, custodians and users (e.g. clinical coders, physicians, hospital managers, health ministry agents, researchers).
- Claims or billing data are to be included because data quality problems of these and the DRG-based administrative databases share the same root causes [14-16], namely those related to the process of deriving clinical codes from hospital encounters.

- Studies in which the authors provide objective evidence of root causes of data quality problems of the DRG-based administrative hospital database.

The studies covering data regarding emergency, primary care or long-term care settings, or studies focusing databases of patient/disease registries, health surveys, or clinical trials data are to be excluded from our review, as the main objective of this systematic review is to identify root causes of data quality issues impacting DRG-based systems, which are independent of these types of healthcare data sources.

Evaluation of the quality of the studies was not considered as an exclusion criterion as we privileged comprehensiveness over study quality. We believe that the quality of the studies will have a small impact in our conclusions since our goal was to systematically identify root causes of problems affecting the quality of DRG-based administrative hospital data, regardless of the study quality.

2.4 Selection of sources of evidence and data charting.

Three reviewers (ARO, ML, MO) independently screened the titles and abstracts, two for each study. Disagreements were solved by consensus with a third reviewer. The Rayyan online free application assisted the screening phase [17].

Three reviewers (ML, MO, RC) then read all the full texts of the studies retained in the screening phase, applying the same eligibility criteria. Any disagreements (if at least one reviewer had a different opinion) on study selection were resolved by consensus.

Data-charting consisted in extracting information relevant for the study of root causes of data quality problems. This was accomplished using a predefined form including several items: 1) study title, 2) type of study (i.e. observational, experimental, qualitative), 3) country of origin of the data, 4) data reporting years, 5) study description, 6) the name of the root cause, 7) the description of root causes identified, including, when possible, an example of the root cause, 8) a sequential identification number, and 9) the study where the root cause had been extracted from. For this task, a root cause has been defined as any identifiable process prone to result in data quality problems of DRG-based administrative hospital databases and it was abstracted from the studies citing the manuscripts' texts as much as possible.

Three reviewers independently charted the data (ML, MO, RC). One of the reviewers (RC) collated the root causes extracted by each of the reviewers into a single list of root causes, removing repeated root causes extracted from the same study by different reviewers, and matching root causes across studies. Any doubts about this process were solved by consensus among the three reviewers.

2.5 Strategy for data synthesis

After matching duplicates across studies, we analyzed the list of root causes of data quality problems by mapping them into major and sub-categories. For the major categories, we borrowed the 6Ms of Ishikawa diagram framework:

- Man - everything that has to do with people's knowledge, preferences, education, culture;
- Material - includes raw material, consumables, and information;
- Method – processes performed in the making or treatment of the data;
- Machine- tools/technology available to generate/collect and use data;
- Mission - purpose, environment;
- Management - everything related with leadership.

This diagram is a cause analysis tool that identifies many possible causes for an effect or problem, graphically sorting ideas into useful categories [18].

Each root cause was independently assigned to one and only one Ishikawa diagram branch by three reviewers. Any conflicts were discussed and solved by consensus. If 1 out of 3 reviewers had a different opinion from the others this would be considered a disagreement.

The subcategories were not part of the Ishikawa framework. These were subjectively conceived by the reviewers, based on the studies themselves and the work of Nouraei et al [19], in efforts to specify similar groups of root causes within each branch (see column 3, Tables 1-6 for a list of all of the considered subcategories). This provides further insights about the causes of data quality issues while affording some detail of the individual root causes without being as exhaustive as the full list.

3. Results

From the total of 2009 retrieved studies, 238 studies were assessed for eligibility based on their full-texts and 77 studies were included for extraction of root causes (Figure 1). Most abstracts and full texts were excluded either due to wrong type of data, wrong study type or research field, wrong outcomes (no root causes), or a combination of these reasons. The characteristics of the studies are presented in Table B1, appendix B. According to Table B1, the included studies cover several study types and reflect root causes perceived in data from countries like the USA (31.2%), UK (14.3%), Portugal (10.4%), Australia (9.1%) and France (5.2%).

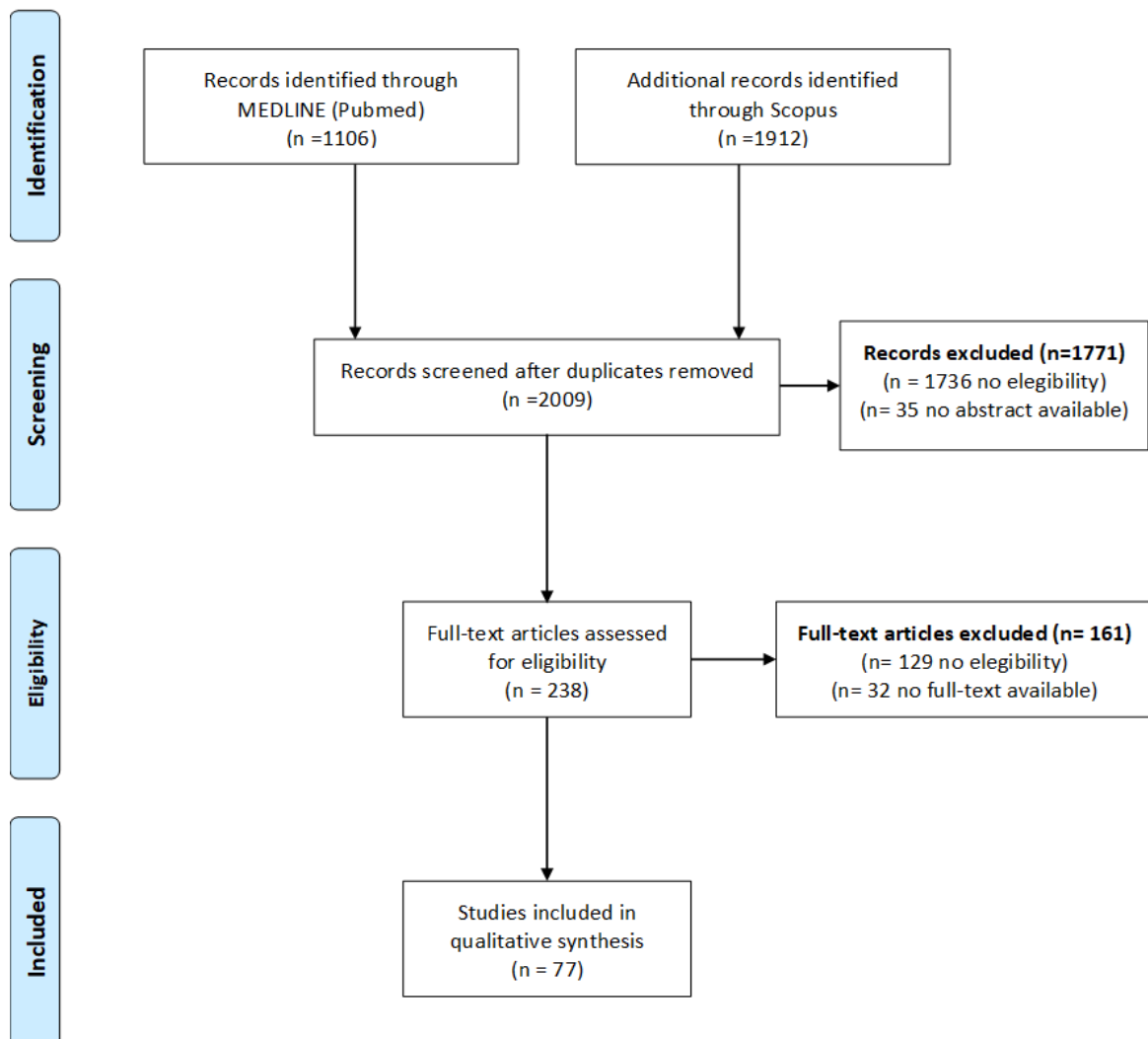


Figure 1 – PRISMA flow diagram for the work conducted in this investigation

The included studies initially yielded 258 root causes of data quality problems. Specifically, root causes related to financial incentives or disincentives (n=24), incomplete or missing documentation (n=15), poor-quality documentation (n=13) and miscoding/misclassification (n=12) had the highest number of duplicates. After the sorting and combination of the duplicates we

arrived at a final set of 105 unique root causes (Tables 1-6). The full description and examples for each root cause can be found in appendix B, Table B2.

Table 1- Root causes of data quality of DRG-based databases assigned to the Ishikawa's

Machine branch

Root Cause #	Number of root causes	Ishikawa subcategory	Root cause name	References
1	4	Clinical coding classification	Complexity of coding classification / rules	[20-22]
2	5		ICD terminology limitations / lack of discriminatory detail	[23-27]
3	2		Inefficient management of reviewed / discontinued ICD codes	[28, 29]
4	2		Transition between coding classification systems	[20, 30]
5	1		Use of two classifications	[31]
6	1	Guidelines and consensus	Deficient guidelines for coding comorbidities	[32]
7	1		Lack of guidelines and consensus	[20]
8	1		Multiple sources of advice	[33]
9	3	Tools related with source information	Unstandardized / ambiguous coding guidelines	[33, 34]
10	1		Impossibility of making explanatory drawings in EHR	[35]
11	1		Inadequate discharge summaries standards	[36]
12	1		Increased information technology adoption	[37]
13	3		Limitations of the computer system	[35, 38, 39]

Table 2- Root causes of data quality of DRG-based databases assigned to the Ishikawa's

Man branch

Root Cause #	Number of root causes	Ishikawa subcategory	Root cause name	References
14	1	Coders and health care providers	Limited awareness of the impact of inaccurate complication coding	[33]
15	4		Mismatch of coding terminology and clinical language	[33, 40, 41]
16	1		Polarized perspectives of coders and physicians	[21]
17	5		Coders' experience	[27, 33, 42-44]
18	2		Coders' insufficient clinical knowledge	[36, 45]
19	1		Coders' lack of education	[46]
20	4		Coders' lack of knowledge on a specific condition	[23, 27, 47, 48]
21	1		Coding bias	[32]
22	2		Coding infrequent events	[49, 50]
23	1		Coding learning curve	[51]
24	6	Clinical coders	Coding subjectivity and variability	[19, 24, 49, 52-54]
25	1		Definition of secondary diagnosis is subject to interpretation	[55]
26	1		Inadequate procedural coding education	[56]
27	4		Inter-coder variation	[38, 51, 57, 58]
28	2		Lack of knowledge of coding guidelines	[39, 59]
29	1		Lack of understanding of procedure coding	[53]
30	1		Medical record administrators' lack of education	[43]
31	3		Poor training and support interventions to coders	[41, 60, 61]
32	1		Reluctance to record more than one code	[42]
33	1		Clinicians' experience	[62]
34	6	Health care providers	Clinicians' lack of awareness / education on the coding process and purpose	[23, 28, 36, 45, 63, 64]
35	2		Clinicians' lack of time and motivation to document records	[28, 65]
36	1		Clinicians' unawareness of documentation needs	[66]
37	5		Difficulties / uncertainty in establishing a diagnosis	[23, 26, 31, 67, 68]
38	2		Lack of assertiveness in diagnostic documentation	[35, 54]
39	1		Student documentation	[69]
40	3		Undercoding of diagnoses and morbidities	[45, 70, 71]

Table 3- Root causes of data quality of DRG-based databases assigned to the Ishikawa's Management branch

Root Cause #	Number of root causes	Ishikawa subcategory	Root cause name	References
41	1	Leadership involvement	Lack of feedback	[61]
42	2		Lack of knowledge of health information management systems	[59, 72]
43	2		Lack of senior leadership involvement	[59, 72]
44	2		Lack of support and respect for the coding function	[59, 72]
45	1		Pressure from administration and health care actors	[73]
46	1		Unsatisfactory education and engagement of hospital administrator, information technologists and researchers	[74]
47	1	Limited resources	Coders' quotas and expectations	[52]
48	5		Coders' time constraints and workload	[27, 39, 40, 42, 75]
49	1		Lack of budget and time	[34]
50	2		Poor investment in documentation / education of physicians	[59, 72]
51	1		Priority of coding	[65]
52	1	Other	Condition not formally acknowledged as to complicate the clinical care of the patient	[55]
53	1		Lack of incentives for comprehensive coding	[40]
54	4		Lack of routine and systematic internal coding audits	[20, 28, 76, 77]

Table 4- Root causes of data quality of DRG-based databases assigned to the Ishikawa's Material branch

Root Cause #	Number of root causes	Ishikawa subcategory	Root cause name	References
55	15	Incomplete / missing information	Incomplete / missing documentation of diagnostic information	[28, 32, 35, 36, 38, 40, 44, 46, 52, 66, 78-80]
56	1		Lack of contextual information	[79]
57	2		Lack of meaningful codes	[57, 58]
58	2		Limited clinical and socio-economic detail	[67, 76]
59	6		Missing / incorrect codes or values	[50, 65, 67, 77, 81, 82]
60	3	Inconsistent information	Inconsistency	[78, 81, 82]
61	2		Inconsistent / conflicting information in the record	[24, 75]
62	3		Inconsistent or incomplete racial / ethnic classification	[83, 84]
63	1	Information fragmentation	Information fragmentation / lack of a standard organization of EHR	[79]
64	1		Information fragmentation / lack of unique patient identifier	[67]
65	1	Medical record factors	Availability of notes	[62]
66	3		Medical record organization	[52, 62, 75]
67	13		Poor-quality documentation	[14, 24, 33, 34, 39, 47, 48, 50, 62, 71, 75, 77, 85]
68	1	Other information issues	Ambiguous documentation	[54]
69	1		Duplicated episodes	[82]
70	1		Imprecise discharge summary values for DRG grouping	[78]
71	1		Incorrect or absent documentation of a procedure	[53]
72	1		Lack of standardized recognition and documentation of transfers	[74]
73	1		Non-verifiable ex post information	[86]
74	1		Updates to definitions of clinical diagnoses	[55]
75	2	Readability	Illegibility	[35, 52]
76	2		Unreadability	[27, 42]
77	7		Use of abbreviations / acronyms / synonyms / nonstandard terminology	[23, 28, 31, 35, 44, 63]

Table 5- Root causes of data quality of DRG-based databases assigned to the Ishikawa's Method branch

Root Cause #	Number of root causes	Ishikawa subcategory	Root cause name	References
78	1	Basis for coding	Ambiguous classification criteria	[87]
79	1		Code sequencing rule	[58]
80	1		Coding based on case notes only	[88]
81	2		Coding based on codebook	[27, 42]
82	1		Coding based on discharge summary	[88]
83	1		Coding based on face sheet and memory -based coding	[42]
84	1		Coding team	[64]
85	1		No checks with physicians for clinical relevance	[33]
86	1	Coding errors	Poor usage of coding materials	[43]
87	1		Errors at the point of attestation	[44]
88	5		Induction errors / Misspecification	[44, 46, 89, 90]
89	12		Miscoding / Misclassification	[29, 33, 38, 44, 48, 66, 70, 74, 77, 82, 84, 89]
90	3		Resequencing	[44, 89, 90]
91	2		Unbundling	[44, 56]
92	1	Quality of filling	Errors assigning information	[81]
93	1		Transcription errors	[44]
94	2		Use of copy and paste	[35, 79]
95	1	Retrospective information	Point in time of the initial classification	[87]
96	1		Retrospective coding / Information collected at discharge	[57]
97	1		Retrospective queries for coding	[28]
98	1		Retrospective writing of discharge summary	[88]

Table 6- Root causes of data quality of DRG-based databases assigned to the Ishikawa's Mission branch

Root Cause #	Number of root causes	Ishikawa subcategory	Root cause name	References
99	1	Coding purpose	Coding purpose	[64]
100	1		Misclassification of severity / risk	[76]
101	24	Financial incentives or disincentives	Financial incentives or disincentives	[25, 29, 33, 44, 47, 49, 51, 56, 57, 60, 64, 67, 68, 86, 87, 91-94]
102	1		Reimbursement system	[26]
103	1	Variability of practices	Fast turnover and day-case nature	[63]
104	3		Variability in coding practices	[26, 27, 54]
105	1		Variability of the quality of health records	[35]

The analysis of root causes provided a structured classification of root causes into categories (Ishikawa's categories) and subcategories (Table 1-6). Most (25.7%) root causes were associated with people's knowledge, preferences, education and culture – "Man", and the fewest (6.7%) were associated with purpose or environment – "Mission" (Figure 2).

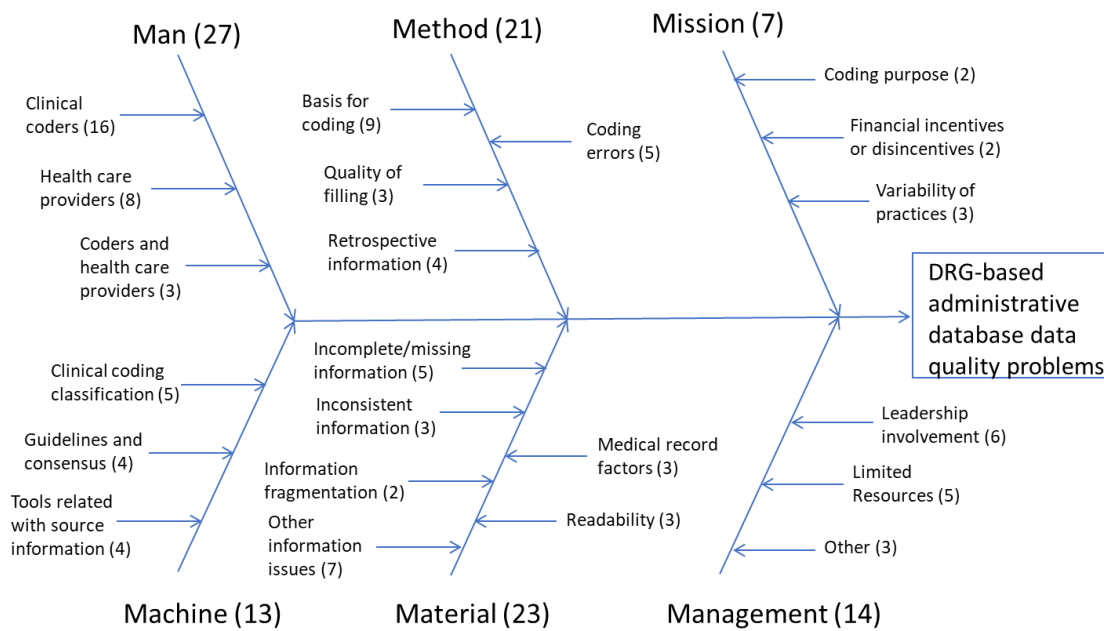


Figure 2 – Ishikawa diagram of root causes of data quality problems affecting DRG-based hospital data: (n) number of root causes identified in each category and subcategory.

Within the Ishikawa branch “Man”, three subcategories related with the professionals involved in generating, collecting and processing the health information (i.e. the clinical coders, and the health care providers) were abstracted. Within the “Machine” branch, three subcategories have emerged. These covered root causes related with tools assisting on the information collection such as clinical coding classification (e.g. ICD terminology limitations), guidelines and consensus (e.g. unstandardized coding guidelines) and tools related with information management (e.g. limitations of the computer system). The “Management” branch was subdivided into three subcategories: leadership involvement which included root causes such as lack of feedback and lack of senior leadership involvement; limited resources, which included any root causes related with budget or human resources constraints; and a miscellaneous subcategory (e.g. lack of incentives for comprehensive coding). The “Material” branch included six subcategories describing different issues that have been reported about the data in health records – the data sources of the DRG-based administrative hospital databases. These subcategories were: incomplete/missing information, inconsistent information, information fragmentation, medical record factors, readability and other information issues (e.g. duplicated episodes). The “Method” branch comprised four subcategories of processes related to health records’ data collection and abstraction into the database, the latter mostly covering issues that arise in the coding process of

clinical information. Finally, within the “Mission” branch, the three considered subdivisions encompass root causes linked to the envisioned utilization of data (i.e. “coding purpose”, “financial incentives or disincentives”) which promote particular documentation practices (e.g. documenting information used to compare hospital performance), as well as root causes linked to the environment in which data is collected, such as variability of practices induced by different hospital contexts.

4. Discussion

This study provides a comprehensive list of root causes of data quality problems affecting DRG-based administrative hospital databases and proposes a framework (the Ishikawa classification) for deriving and analyzing such list.

Root causes related to documentation issues (“Material”) and to financial incentives or disincentives (“Mission”) play an important role in contributing to data quality problem, being highly present in the literature. In fact, we obtained a high number of duplicates for these causes. Regarding the documentation issues, we found that missing or incomplete documentation of diagnostic information and poor-quality of documentation were the most frequently encountered causes of data quality problems in DRG-based administrative hospital databases. Other sources of data quality problems included in the “Material” category are also linked to documentation issues such as inconsistent or fragmented information, and lack of readability of health records. The transition from paper to electronic health records has been determinant in improving the legibility of health records as well as their information organization and presentation. However, electronic records remain susceptible to the use of abbreviations, acronyms and non-standard terminology, and are prone to create other problems such as the use of copy and paste, also identified in our literature review, that may lead to the repetition of large uninformative chunks of text, sometimes including erroneous information, making coding more difficult and more time-consuming [35, 79]. Moreover, the definition of a standard content for medical records that is fit for coding purposes has also been suggested to address problems of missing/incomplete information [35, 79].

DRG-based hospital data are primarily used for billing and reimbursement purposes, hence clinical coding is frequently financially driven. Among the studies reporting root causes related to the financial incentives or disincentives, upcoding (the misreporting of episodes of care into higher payment DRGs) was the most frequently assessed root cause. Upcoding may result from different coding practices with varying legal implications, including coding comorbidities comprehensively to raise treatment costs; substituting the primary diagnosis by a secondary diagnosis; and adding comorbidities that are not documented. The case of upcoding through birth weight of newborns was documented in several studies [25, 86, 93]. This information is used to determine the DRG of a hospitalization, with hospitalizations of newborns documented with low birth weight yielding substantially higher payments. Unlike other upcoding situations that can be mitigated by instituting routine and systematic coding audits, birthweight is virtually impossible to verify [86].

A large number of root causes identified are related to the coding activity. Besides upcoding, several studies reported issues in the selection of codes as a root cause of data quality issues of hospital administrative data ("Miscoding/misclassification"). This encompasses coding unsupported by the clinical record, the assignment of generic codes when information exists to assign more specific codes, assignment of incorrect codes according to the governing rules, or assign codes without the physician attesting to their accuracy [44, 77, 89]. Other data issues associated with coding arise from coders' subjectivity, bias, learning curve, different experience and training, which are accompanied by variability of coding practices across different specialties and across hospitals. The adequate guidelines and terminologies could help reducing the impact of some of these root causes. However, issues in these tools have also been identified as root causes of data quality problems of the DRG-based administrative hospital databases.

4.1 Potential solutions to the root causes

To achieve the greatest value, it is essential that data has the highest possible levels of quality. Through the Ishikawa diagram, a widely used framework to identify and analyze root causes of problems, it was possible to discern categories with causal links to data quality issues of the DRG-based hospital databases, capable of capturing the different aspects of the process of information generation in these databases and allowing a better representation and analysis of which causes to address in efforts to improve their data quality. For example, root causes linked to "Man" could

be mitigated through training programs and incentives aimed at raising awareness of the purpose of the data and the implications of data quality issues, as well as improving skills with existing tools and the communication between the different actors involved in the data generation [41, 60, 77]. Moreover, root causes linked with “Method” and “Machine” could be solved with the creation or improvement of tools used to generate/collect data, like automated coding processing applications [95], or text processing instruments to omit repeated text generated through copy and paste. Furthermore, within the “Management” branch, a leadership that is involved with the coding team, that manages to improve communication and education among the different professionals and that negotiates incentive schemes (bilaterally), will have a positive impact in the motivation, dedication and in the meeting of deadlines of the collaborators [59, 72]. Finally, within the “Mission” branch increasing information technology adoption with, for example, systems that can alert the coding professional on-the-fly for unusual combination of codes could improve the quality of data for the envisioned purpose by minimizing upcoding [37, 91, 94].

4.2 Limitations

This review has limitations. We have only searched studies in two databases and have not included grey literature. This can question whether we have reached a saturation point, although we have usually found duplicates for each root cause.

Some root causes are very specific of certain contexts such procedural codes that are used in the United States only. On the other hand, some root causes may no longer apply in some countries. For example, “ICD terminology limitations / lack of discriminatory detail” has been addressed with the transition from the ICD-9-CM to the ICD-10-CM, that resolved many of these issues. Nevertheless, awareness of this issue is important when comparing/combining data from different time periods or when using different classification systems.

Additionally, in the presented Ishikawa diagram we assigned a root cause to one and only one branch / category and subcategory. Despite the fact that the root causes would be the same, the diagram and respective table of results would have been very different if this were not the case, since this process of mutually excluding causes was demanding and most decisions were made via consensus.

4.3 Implications for research

The quality of DRG-based administrative hospital databases is paramount to hospital financing but also for its reutilization in research, holding potentially valuable insights that can improve health care practice and planning. Future research can consider improving the methodology for identifying root causes of data quality problems. It can also explore the relevance of the impact these root causes have on data quality of DRG-based administrative hospital databases, by using, for example, a Delphi panel. In addition, further research could investigate possible solutions to some of the discovered root cause of problems.

5. Conclusions

The presented systematic review sets forward a catalogue of 105 individual root causes of data quality problems affecting DRG-based administrative hospital databases. Through the Ishikawa categorization this catalogue of root causes is represented in a meaningful way, generalizable to other countries collecting this type of data, and in terms that allow for the tackling of these issues, with a potentially significant impact on improving hospital funding and health care research.

Authors' contributions

All authors have made a substantial, direct, intellectual contribution to this systematic review.

Acknowledgements

The authors would like to thank the support given by the Project "POCI-01-0145-FEDER-030766" (1st.IndiQare - Quality indicators in primary health care: validation and implementation of quality indicators as an assessment and comparison tool), funded by Fundação para a Ciência e a Tecnologia (FCT) and co-funded by Fundo de Desenvolvimento Regional (FEDER) through Operacional Competitividade e Internacionalização (COMPETE 2020).The authors would also like to thank the project GEMA: Generation and Evaluation of Models for Data Quality (Ref.:SBPLY/17/180501/000293)

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors report no declarations of competing interest.

Summary table

What was already known on the topic?

- DRG-based administrative hospital data is important for hospital financing and for clinical, epidemiological and health services research.
- Fallible information is included in these types of databases, eliciting data quality problems

What this study added to our knowledge?

- This is the first systematic review of root causes of data quality problems affecting DRG-based databases, identifying 105 different root causes
- The Ishikawa framework allows for a useful categorization of root causes that improve the attainment of solutions to these problems

References

1. Busse, R., et al., *Diagnosis related groups in Europe: moving towards transparency, efficiency, and quality in hospitals?* BMJ : British Medical Journal, 2013. **346**: p. f3197.
2. Fetter, R.B., et al., *Case mix definition by diagnosis-related groups*. Med Care, 1980. **18**(2 Suppl): p. iii, 1-53.
3. Freitas, J.A., et al. *Implications of Data Quality Problems within Hospital Administrative Databases*. 2010. Berlin, Heidelberg: Springer Berlin Heidelberg.
4. Miguel, A., et al., *Detection of adverse drug reactions using hospital databases-a nationwide study in Portugal*. Pharmacoepidemiol Drug Saf, 2013. **22**(8): p. 907-13.
5. Santos, J.V., et al., *Hospitalisations with Hidradenitis Suppurativa: An Increasing Problem That Deserves Closer Attention*. Dermatology, 2016. **232**(5): p. 613-618.
6. Santos, J.V., et al., *Atrial Fibrillation as an Ischemic Stroke Clinical and Economic Burden Modifier: A 15-Year Nationwide Study*. Value Health, 2017. **20**(8): p. 1083-1091.
7. Freitas, A., et al., *Quality in hospital administrative databases*. Applied Mathematics & Information Sciences, 2014. **8**: p. 1-6.
8. Wang, R.Y. and D.M. Strong, *Beyond Accuracy: What Data Quality Means to Data Consumers*. Journal of Management Information Systems, 1996. **12**(4): p. 5-33.
9. Assareh, H., H.M. Achat, and J.F. Levesque, *Accuracy of inter-hospital transfer information in Australian hospital administrative databases*. Health Informatics J, 2019. **25**(3): p. 960-972.
10. Gaspar, J., F. Lopes, and A. Freitas, *Detection of inconsistencies in hospital data coding*. HEALTHINF 2012 - Proceedings of the International Conference on Health Informatics, 2012: p. 189-194.
11. Hsia, D.C., et al., *Accuracy of Diagnostic Coding for Medicare Patients under the Prospective-Payment System*. New England Journal of Medicine, 1988. **318**(6): p. 352-355.
12. Lobo, M.F., et al. *Protocol for Analysis of Root Causes of Problems Affecting the Quality of the Diagnosis Related Group-Based Hospital Data: A Rapid Review and Delphi Process*. 2020. Cham: Springer International Publishing.
13. Moher, D., et al., *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement*. Systematic Reviews, 2015. **4**(1): p. 1.
14. Aiello, F.A., et al., *A physician-led initiative to improve clinical documentation results in improved health care documentation, case mix index, and increased contribution margin*. J Vasc Surg, 2018. **68**(5): p. 1524-1532.
15. Bastani, H., J. Goh, and M. Bayati, *Evidence of Upcoding in Pay-for-Performance Programs*. Management Science, 2019. **65**(3): p. 1042-1060.
16. Carter, G.M., J.P. Newhouse, and D.A. Relles, *How much change in the case mix index is DRG creep?* J Health Econ, 1990. **9**(4): p. 411-28.
17. Ouzzani, M., et al., *Rayyan-a web and mobile app for systematic reviews*. Syst Rev, 2016. **5**(1): p. 210.
18. Ishikawa, K., *Introduction to Quality Control*. 1990: Taylor & Francis.
19. Nouraei, S.A., et al., *A Study of Clinical Coding Accuracy in Surgery: Implications for the Use of Administrative Big Data for Outcomes Management*. Ann Surg, 2015. **261**(6): p. 1096-107.
20. Alonso, V., et al., *Problems and barriers in the transition to ICD-10-CM/PCS: A qualitative study of medical coders' perceptions*. Advances in Intelligent Systems and Computing, 2019. **932**: p. 72-82.
21. Butz, J., et al., *Differences in coder and physician perspectives on the transition to ICD-10-CM/PCS: A survey study*. Health Policy and Technology, 2016. **5**(3): p. 251-259.

22. Stausberg, J., et al., *Reliability of diagnoses coding with ICD-10*. International Journal of Medical Informatics, 2008. **77**(1): p. 50-57.
23. Khan, A., et al., *Limited Accuracy of Administrative Data for the Identification and Classification of Adult Congenital Heart Disease*. J Am Heart Assoc, 2018. **7**(2).
24. Cunningham, J., et al., *The quality of medical record documentation and External cause of fall injury coding in a tertiary teaching hospital*. Health Inf Manag, 2014. **43**(1): p. 6-15.
25. Freitas, A., et al., *Quality in hospital administrative databases*. Applied Mathematics and Information Sciences, 2014. **8**(1): p. 1-6.
26. Prins, H. and A. Hasman, *Appropriateness of ICD-coded diagnostic inpatient hospital discharge data for medical practice assessment. A systematic review*. Methods Inf Med, 2013. **52**(1): p. 3-17.
27. Farzandipour, M. and A. Sheikhtaheri, *Evaluation of factors influencing accuracy of principal procedure coding based on ICD-9-CM: an Iranian study*. Perspect Health Inf Manag, 2009. **6**: p. 5.
28. Hay, P., et al., *The importance of clinical documentation improvement for Australian hospitals*. Health Inf Manag, 2019: p. 1833358319854185.
29. Gaspar, J., F. Lopes, and A. Freitas, *Detection of inconsistencies in hospital data coding*. 2012, CIDES, Department of Health Information and Decision Sciences, Porto, Portugal CINTESIS - Center for Research in Health Technologies and Information Systems, Porto, Portugal Faculty of Medicine, University of Porto, Porto, Portugal. 189-194.
30. Nilson, F., et al., *The effect of the transition from the ninth to the tenth revision of the International Classification of Diseases on external cause registration of injury morbidity in Sweden*. Inj Prev, 2015. **21**(3): p. 189-94.
31. Fermat, M., et al., *Detection of complicated ectopic pregnancies in the hospital discharge database: A validation study*. PLoS One, 2019. **14**(6): p. e0217674.
32. Souza, J., et al., *Importance of coding co-morbidities for APR-DRG assignment: Focus on cardiovascular and respiratory diseases*. Health Inf Manag, 2019: p. 1833358319840575.
33. Ressler, M.A., et al., *Sources of variability in hospital administrative data: Clinical coding of postoperative ileus*. Health Inf Manag, 2019. **48**(2): p. 101-108.
34. Doktorchik, C., et al., *A qualitative evaluation of clinically coded data quality from health information manager perspectives*. Health Inf Manag, 2019: p. 1833358319855031.
35. Alonso, V., et al., *Health records as the basis of clinical coding: Is the quality adequate? A qualitative study of medical coders' perceptions*. Health Inf Manag, 2019: p. 1833358319826351.
36. Roberts, L., et al., *Capturing causative agents in clinical coding*. British Journal of Health Care Management, 2016. **22**(9): p. 461-468.
37. Park, Y.T., J. Lee, and J. Lee, *Association between Health Information Technology and Case Mix Index*. Healthc Inform Res, 2017. **23**(4): p. 322-327.
38. Carlson, K.F., et al., *Accuracy of external cause-of-injury coding in VA polytrauma patient discharge records*. J Rehabil Res Dev, 2010. **47**(8): p. 689-97.
39. Bramley, M. and B.A. Reid, *Morbidity data quality initiatives in Ireland*. Health Inf Manag, 2005. **34**(2): p. 47-53.
40. Kessler, B.A., M.P. Catalino, and J.D. Jordan, *Reducing the reported mortality index within a neurocritical care unit through documentation and coding accuracy*. World Neurosurg, 2019.
41. Dyers, R., et al., *Training and support to improve ICD coding quality: A controlled before-and-after impact evaluation*. S Afr Med J, 2017. **107**(6): p. 501-506.
42. Farzandipour, M., A. Sheikhtaheri, and F. Sadoughi, *Effective factors on accuracy of principal diagnosis coding based on International Classification of Diseases, the 10th*

- revision (ICD-10). *International Journal of Information Management*, 2010. **30**(1): p. 78-84.
43. Lungeanu, D., et al., *Exploratory analysis of medical coding practices: the relevance of reported data quality in obstetrics-gynaecology*. *Stud Health Technol Inform*, 2008. **136**: p. 839-44.
 44. O'Malley, K.J., et al., *Measuring diagnoses: ICD code accuracy*. *Health Serv Res*, 2005. **40**(5): p. 1620-39.
 45. Muzerengi, S., et al., *Coding accuracy for Parkinson's disease hospital admissions: implications for healthcare planning in the UK*. *Public Health*, 2017. **146**: p. 4-9.
 46. Surján, G., *Questions on validity of International Classification of Diseases-coded diagnoses*. *International Journal of Medical Informatics*, 1999. **54**(2): p. 77-95.
 47. Cohen, S., et al., *Accuracy of claim data in the identification and classification of adults with congenital heart diseases in electronic medical records*. *Arch Cardiovasc Dis*, 2019. **112**(1): p. 31-43.
 48. Li, J., et al., *Coding accuracy for endophthalmitis diagnosis and cataract procedures in Western Australia: The Endophthalmitis Population Study of Western Australia (EPSWA): Second Report*. *Ophthalmic Epidemiology*, 2003. **10**(2): p. 133-145.
 49. Joos, C., et al., *Accuracy of ICD-10 codes for identifying hospitalizations for acute anticoagulation therapy-related bleeding events*. *Thromb Res*, 2019. **181**: p. 71-76.
 50. Alechna, N., J. Westbrook, and R. Roberts, *The quality of burns coding. Compliance with standards and the effects on clinical data*. *Health information management : journal of the Health Information Management Association of Australia*, 1998. **28**(4): p. 181-185.
 51. Januel, J.M., et al., *Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data*. *BMC Health Serv Res*, 2011. **11**: p. 194.
 52. Lucyk, K., K. Tang, and H. Quan, *Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study*. *BMC Health Serv Res*, 2017. **17**(1): p. 766.
 53. Aiello, F.A., et al., *A multidisciplinary approach to vascular surgery procedure coding improves coding accuracy, work relative value unit assignment, and reimbursement*. *J Vasc Surg*, 2016. **64**(2): p. 465-470.
 54. Nouraei, S.A., 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)*, 2016. **38**(2): p. 352-62.
 55. Rodenberg, H., et al., *The Expanding Role of Clinical Documentation Improvement Programs in Research and Analytics*. *Perspect Health Inf Manag*, 2019. **16**: p. 1d.
 56. Filler, B.C., *Coding basics for orthopaedic surgeons*. *Clin Orthop Relat Res*, 2007. **457**: p. 105-13.
 57. Mercier, F., et al., *Unexpected discrepancies in hospital administrative databases can impact the accuracy of monitoring thyroid surgery outcomes in France*. *PLoS One*, 2018. **13**(12): p. e0208416.
 58. Michel, J.L., D. Cheng, and T.J. Jackson, *Comparing the coding of complications in Queensland and Victorian admitted patient data*. *Aust Health Rev*, 2011. **35**(3): p. 245-52.
 59. Rangachari, P., *The strategic management of organizational knowledge exchange related to hospital quality measurement and reporting*. *Qual Manag Health Care*, 2008. **17**(3): p. 252-69.
 60. Lorenzoni, L., R. Da Cas, and U.L. Aparo, *Continuous training as a key to increase the accuracy of administrative data*. *J Eval Clin Pract*, 2000. **6**(4): p. 371-7.
 61. Lorenzoni, L., R. Da Cas, and U.L. Aparo, *The quality of abstracting medical information from the medical record: the impact of training programmes*. *Int J Qual Health Care*, 1999. **11**(3): p. 209-13.

62. Macaulay, E.M., et al., *Prospective audit of discharge summary errors*. Br J Surg, 1996. **83**(6): p. 788-90.
63. Naran, S., et al., *Audit of accuracy of clinical coding in oral surgery*. Br J Oral Maxillofac Surg, 2014. **52**(8): p. 735-9.
64. Burns, E.M., et al., *Systematic review of discharge coding accuracy*. J Public Health (Oxf), 2012. **34**(1): p. 138-48.
65. Dalal, S. and B. Roy, *Reliability of clinical coding of hip fracture surgery: implications for payment by results?* Injury, 2009. **40**(7): p. 738-41.
66. Soo, I.H., et al., *Do we have enough information? How ICD-10-AM Activity codes measure up*. Health Inf Manag, 2009. **38**(1): p. 22-34.
67. Aelvoet, W., et al., *Do inter-hospital comparisons of in-hospital, acute myocardial infarction case-fatality rates serve the purpose of fostering quality improvement? An evaluative study*. BMC Health Serv Res, 2010. **10**: p. 334.
68. Silverman, E. and J. Skinner, *Medicare upcoding and hospital ownership*. J Health Econ, 2004. **23**(2): p. 369-89.
69. Howard, R. and R.M. Reddy, *Coding Discrepancies Between Medical Student and Physician Documentation*. J Surg Educ, 2018. **75**(5): p. 1230-1235.
70. Nouraei, S.A., et al., *Impact of developing a multidisciplinary coded dataset standard on administrative data accuracy for septoplasty, septorhinoplasty and nasal trauma surgery*. J Laryngol Otol, 2017. **131**(4): p. 341-346.
71. Farhan, J., et al., *Documentation and coding of medical records in a tertiary care center: a pilot study*. Ann Saudi Med, 2005. **25**(1): p. 46-9.
72. Rangachari, P., *Knowledge sharing networks related to hospital quality measurement and reporting*. Health Care Management Review, 2008. **33**(3): p. 253-263.
73. Georgescu, I. and F.G.H. Hartmann, *Sources of financial pressure and up coding behavior in French public hospitals*. Health Policy, 2013. **110**(2): p. 156-163.
74. Ayub, S., et al., *Financial implications of coding inaccuracies in patients undergoing elective endovascular abdominal aortic aneurysm repair*. J Vasc Surg, 2019. **69**(1): p. 210-218.
75. Reid, B.A., et al., *Best practice in the management of clinical coding services: Insights from a project in the republic of Ireland, Part 1*. Health Information Management Journal, 2017. **46**(2): p. 69-77.
76. Prasad, A., et al., *Understanding Differences in Administrative and Audited Patient Data in Cardiac Surgery: Comparison of the University HealthSystem Consortium and Society of Thoracic Surgeons Databases*. J Am Coll Surg, 2016. **223**(4): p. 551-557.e4.
77. Cheng, P., et al., *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, 2009. **38**(1): p. 35-46.
78. Rabia, L., I.A. Amarouche, and K. Beghdad Bey, *Rule-based approach for detecting dirty data in discharge summaries*. Proceedings of the 2018 13th International Symposium on Programming and Systems, ISPS 2018, 2018: p. 1-6.
79. Botsis, T., et al., *Secondary Use of EHR: Data Quality Issues and Informatics Opportunities*. Summit Transl Bioinform, 2010. **2010**: p. 1-5.
80. Ceratti, A., F.H. Roger France, and C. Beguin, *Health data quality improvement by comparing administrative medical data and billing data*. Int J Med Inform, 2008. **77**(8): p. 527-33.
81. Marques, B., et al., *Quality of data from central and departmental inpatient databases: A comparative study*. 2012, Department of Health Information and Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal CINTESIS - Center for Research in Health Technologies and Information Systems, University of Porto, Porto, Portugal. 195-200.

82. Freitas, J.A., et al., *Implications of data quality problems within hospital administrative databases*. Vol. 29. 2010, Department of Biostatistics and Medical Informatics, Faculdade de Medicina, Universidade Do Porto, Alameda Hernani Monteiro, 4200 Porto, Portugal CINTESIS, Center for Research in Health Technologies and Information Systems, Universidade Do Porto, Portugal. 823-826.
83. Pellegrin, K.L., et al., *Improving Accuracy and Relevance of Race/Ethnicity Data: Results of a Statewide Collaboration in Hawaii*. J Healthc Qual, 2016. **38**(5): p. 314-21.
84. Andrews, R.M., *Race and ethnicity reporting in statewide hospital data: Progress and future challenges in a key resource for local and state monitoring of health disparities*. Journal of Public Health Management and Practice, 2011. **17**(2): p. 167-173.
85. Hajibandeh, S., et al., *Accuracy of routinely collected comorbidity data in patients undergoing colectomy: a retrospective study*. Int J Colorectal Dis, 2017. **32**(9): p. 1341-1344.
86. Jorges, H. and J. Koberlein, *What explains DRG upcoding in neonatology? The roles of financial incentives and infant health*. J Health Econ, 2015. **43**: p. 13-26.
87. Steinbusch, P.J., et al., *The risk of upcoding in casemix systems: a comparative study*. Health Policy, 2007. **81**(2): p. 289-99.
88. Tsopra, R., et al., *The impact of three discharge coding methods on the accuracy of diagnostic coding and hospital reimbursement for inpatient medical care*. Int J Med Inform, 2018. **115**: p. 35-42.
89. Hsia, D.C., et al., *Medicare reimbursement accuracy under the prospective payment system, 1985 to 1988*. Jama, 1992. **268**(7): p. 896-9.
90. Hsia, D.C., *Accuracy of Medicare reimbursement for cardiac arrest*. Jama, 1990. **264**(1): p. 59-62.
91. Barros, P. and G. Braun, *Upcoding in a National Health Service: the evidence from Portugal*. Health Econ, 2017. **26**(5): p. 600-618.
92. Januleviciute, J., et al., *How do Hospitals Respond to Price Changes? Evidence from Norway*. Health Econ, 2016. **25**(5): p. 620-36.
93. Abler, S., et al., *Effect of the introduction of diagnosis related group systems on the distribution of admission weights in very low birthweight infants*. Arch Dis Child Fetal Neonatal Ed, 2011. **96**(3): p. F186-9.
94. Harrington, K., A. Allen, and L. Ruchala, *Restraining medicare abuse: The case of upcoding*. Research in Healthcare Financial Management, 2007. **11**(1): p. 1-25.
95. Ramalho, A., J. Souza, and A. Freitas. *The Use of Artificial Intelligence for Clinical Coding Automation: A Bibliometric Analysis*. in *Distributed Computing and Artificial Intelligence, 17th International Conference*. 2021. Cham: Springer International Publishing.

Appendix A

Table A1- Medline search strategy

Search item	Concept	Query	Number of hits
#1	Administrative data	(administrative*[All Fields] OR "claims"[All Fields] OR "billing"[All Fields] OR "Diagnosis-related group"[All Fields] OR "DRG"[All Fields] OR hospitalization*[All Fields] OR hospitalisation*[All Fields] OR discharge*[All Fields] OR admission*[All Fields] OR "inpatients"[MeSH Terms] OR inpatient*[All Fields] OR "clinical coding"[MeSH Terms] OR "clinical coding"[All Fields] OR "medical coding"[All Fields] OR "international classification of diseases"[All Fields] OR "ICD"[All Fields]) AND (Data*[All Fields] OR record*[All Fields])	316093
#2	Outcome	("data accuracy"[MeSH Terms] OR "data accuracy"[All Fields] OR "data quality"[All Fields] OR "information quality"[All Fields] OR "data reliability"[All Fields] OR "data consistency"[All Fields] OR "data timeliness"[All Fields] OR "data completeness"[All Fields] OR "data credibility"[All Fields] OR "data validity"[All Fields] OR "data error"[All Fields] OR "data errors"[All Fields] OR "upcoding"[All Fields] OR "downcoding"[All Fields] OR "coding compliance"[All Fields] OR "coding accuracy")	11349
#1 AND #2		(((((("data accuracy"[MeSH Terms] OR "data accuracy"[All Fields] OR "data quality"[All Fields] OR "information quality"[All Fields] OR "data reliability"[All Fields] OR "data consistency"[All Fields] OR "data timeliness"[All Fields] OR "data completeness"[All Fields] OR "data credibility"[All Fields] OR "data validity"[All Fields] OR "data error"[All Fields] OR "data errors"[All Fields] OR "upcoding"[All Fields] OR "downcoding"[All Fields] OR "coding compliance"[All Fields] OR "coding accuracy")))) AND (((administrative*[All Fields] OR "claims"[All Fields] OR "billing"[All Fields] OR "Diagnosis-related group"[All Fields] OR "DRG"[All Fields] OR hospitalization*[All Fields] OR hospitalisation*[All Fields] OR discharge*[All Fields] OR admission*[All Fields] OR "inpatients"[MeSH Terms] OR inpatient*[All Fields] OR "clinical coding"[MeSH Terms] OR "clinical coding"[All Fields] OR "medical coding"[All Fields] OR "international classification of diseases"[All Fields] OR "ICD"[All Fields]) AND (Data*[All Fields] OR Database* [All Fields] OR record*[All Fields]))))	1213

Appendix B

Table B1- Study characteristics

Study ID	Study design	Number of root causes extracted	Data reporting year	Country of origin of data	Study description / Study aims	Reference
Abler S	Retrospective observational study	2	1999–2001 and 2004–2006	Germany	This study aimed specifically to examine the border areas between weight categories for distribution patterns of admission weights and identify potential changes associated with the introduction of the DRG system.	[93]
Aelvoet W	Retrospective observational study	5	2002–2005	Belgium	By determining the existence of inter-hospital differences in AMI-CFR, to evaluate to which extent Belgian discharge records allow the assessment of quality of care in the field of AMI, and to identify starting points for quality improvement.	[67]
Aiello FA (a)	Retrospective observational study	1	2013–2016	USA	The study compared documentation of the inpatient hospital stay for all patients at a single institution undergoing carotid endarterectomy (CEA) and open infra-inguinal procedures (OIPs). A physician-led coding initiative to educate clinicians on inpatient documentation was undertaken and clinical documentation and billing was stratified into the pre-initiative (January 2013 to June 2014) and post-initiative (July 2014 to July 2016) groups.	[14]
Aiello FA (b)	Prospective observational study	3	2014	USA	This study aimed to study hypothesizes that direct physician involvement in procedural coding will lead to improved coding accuracy, increased work relative value unit (wRVU) assignment, and increased physician reimbursement.	[53]
Alechna N	Retrospective observational study	3	1994–1996	Australia	This study aimed to determine 1. the degree of compliance with the Australian Coding Standards (ACS) relating to burns separations, 2. the relationship between coding errors and poor-quality documentation, and 3. the impact of coding errors on ANDRG allocation.	[50]
Alonso V (a)	Qualitative study	8	N.A.	Portugal	The study investigated the perceptions of medical coders regarding possible problems with health records that may affect the quality of the coded data.	[35]

Alonso V (b)	Qualitative study	4	N.A.	Portugal	The study explored the perceptions of medical coders (medical doctors) regarding problems related to the transition from ICD-9-CM to ICD-10-CM/PCS that may affect the quality of coded data.	[20]
Andrews RM	Retrospective observational study	3	1991-2008	USA	This study aimed to examine the extent of race-ethnicity coding in statewide hospital discharge data systems, compare the collection to national standards, and to assess the completeness and accuracy of race-ethnicity data collected in these systems.	[84]
Ayub S	Retrospective observational study	3	2010-2015	USA	This study aimed to evaluate coding accuracy and its effect on hospital cost for patients undergoing EVAR (endovascular aneurysm repair).	[74]
Barros P	Retrospective observational study	1	2001-2008	Portugal	This study aimed to understand how hospitals respond to price increases or decreases using the 2006 DRGs price change.	[91]
Botsis T	Retrospective observational study	4	1999-2009	USA	This study aimed to report first-hand experience with some data quality issues in a survival analysis study for pancreatic cancer.	[79]
Bramley M	Qualitative study	4	2004	Ireland	In this paper, we focus on the results, key findings and recommendations for the Unit's data quality initiatives.	[39]
Burns EM	Systematic review	4	1989-2011	UK	This study aimed to assess the published accuracy of routinely collected data sets in Great Britain.	[64]
Butz J	Qualitative study	2	2014 and 2015	USA	This study aimed to assess similarities and differences in perspectives between coders and physicians on the use of ICD-10.	[21]
Carlson KF	Retrospective observational study	4	2001-2006	USA	This study aimed to conduct a preliminary examination of E-coding practice and accuracy with use of a population of Polytrauma Rehabilitation Center's inpatients.	[38]
Ceratti A	Retrospective observational study	1	2000	Belgium	This study aimed to compare diagnoses found in the medical database to treatments mentioned in the billing database and to identify hospital stays with discrepancies.	[80]
Cheng P	Retrospective observational study	4	2004-2005	Australia	This study aimed to examine the clinical coding and DRG processes in a large tertiary teaching hospital in Melbourne, in relation to the status of the hospital's revenues.	[77]
Cohen S	Retrospective observational study	3	2000-2014	France	The aim of this study was to assess the accuracy of ICD-10 codes retrieved from administrative claims data in the identification and classification of adult congenital heart disease (ACHD).	[47]

Cunningham J	Retrospective observational study	4	2007-2008	Australia	This paper reviews the documentation and coding of External causes of admitted fall cases in a major hospital.	[24]
Dalal S	Retrospective observational study	3	2003-2006	UK	This study aimed at evaluating how well hip fracture operative activity is recorded.	[65]
Doktorchik C	Qualitative study	3	N.A.	Canada	This study assessed qualitatively the strengths and barriers regarding clinical coding quality from the perspective of health information managers.	[34]
Dyers R	Quasi-experimental study	2	2014	South Africa	This study aimed to evaluate the impact of the ICD coding support package by comparing data quality before and after the introduction of the package at the intervention site and a control site.	[41]
Farhan J	Retrospective observational study	2	2001	Saudi Arabia	This study examined the frequency and sources of errors in the medical records at King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia (KFSHRC).	[71]
Farzandipour M (a)	Retrospective observational study	6	2007-2008	Iran	This study aimed to assess the accuracy of principal diagnosis coding and its effective factors based on International Classification of Diseases and Related Health Problems, the 10th revision (ICD-10).	[42]
Farzandipour M (b)	Retrospective observational study	7	2007-2008	Iran	This study aimed to evaluate the accuracy of principal procedure coding and its effective factors in teaching hospitals affiliated with Kashan University of Medical Sciences (KAUMS) in Kashan, Iran.	[27]
Fermaut M	Retrospective observational study	3	2012	France	The objective of this study was to propose a method to measure the validity of the hospital discharge database (HDD) to detect CEPSB (Complicated ectopic pregnancies with severe bleeding) among hospital stays in two gynecological units.	[31]
Filler BC	Narrative review	3	N.A.	USA	This study describes the current CPT coding process, stressing areas where physicians have incorrectly submitted or omitted codes.	[56]
Freitas A (a)	Retrospective observational study	2	2000-2010	Portugal	This study aimed to discuss the use and the quality of administrative data, specifically data from hospital discharges.	[25]
Freitas A (b)	Retrospective observational study	4	2000-2007	Portugal	This study aimed to present some issues about quality problems, to expose some examples of poor data in administrative databases and to discuss some potential implications of these problems.	[82]

Gaspar J	Retrospective observational study	3	2000-2009	Portugal	This study aimed to develop a prototype for the detection of data quality problems and inconsistencies in inpatient and ambulatory episodes in hospital databases.	[29]
Georgescu I	Qualitative study	1	2009	France	This study investigates the effects of health care decision pressure from the hospital's administration and from the professional peer group on physician's inclination to engage in up coding.	[73]
Hajibandeh S	Retrospective observational study	1	2015-2017	UK	This study aimed to determine the baseline accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of HES codes for comorbidities in patients undergoing any types of colectomy.	[85]
Harrington K	Retrospective observational study	3	1993-2001	USA	This paper investigates three issues arising from HIPAA regulation and enforcement. First, it investigates the effectiveness of HIPAA and the ensuing enforcement actions in combating upcoding. Second, it investigates the causes of upcoding by examining the hospital characteristics associated with upcoding. Finally, it investigates the determinants of the reduction in upcoding after HIPAA.	[94]
Hay P	Article commentary	9	N.A.	Australia	This article has focused on validating the relevancy of clinical documentation improvement (CDI) for the Australian healthcare environment.	[28]
Howard R	Retrospective observational study	1	2014-2016	USA	This study aimed to assess the coding accuracy of medical student documentation.	[69]
Hsia DC (a)	Retrospective observational study	3	1988	USA	From a simple random sample of 2451 hospital charts for Medicare discharges from 1988, the American Medical Record Association reabstracted the ICD-9-CM codes on a blinded basis, grouped them to DRGs, and determined the reasons for discrepancies.	[89]
Hsia DC (b)	Retrospective observational study	3	1985	USA	This article investigates the coding accuracy of Medicare discharges billed as DRG 129.	[90]
Januel JM	Retrospective observational study	3	1999,2001 and 2003	Switzerland	This study aimed to assess data accuracy in the five years following the implementation of ICD-10 in the French part of Switzerland, with assessments of administrative data accuracy relative to charts in 1999, 2001, and 2003.	[51]

Januleviciute J	Retrospective observational study	2	2003-2007	Norway	This study aimed to investigate whether price changes for different treatments affect the number of patients treated and the mix of activity provided by hospitals.	[92]
Joos C	Retrospective observational study	3	2017	USA	The study determines the sensitivity, specificity, PPV, and NPV of ICD-10 codes for bleeding events in hospitalized patients previously prescribed any type of oral anticoagulant therapy who were admitted to the University of Utah Hospital from October 1, 2017 to December 31, 2017 for any reason were included.	[49]
Jurges H	Retrospective observational study	2	1996–2010 and 2006-2011	Germany	This study aimed to study the determinants of upcoding, using the introduction of diagnosis related groups (DRGs) in German neonatology.	[86]
Kessler BA	Experimental study	5	2016	USA	The authors aimed to evaluate whether a reduction in the reported mortality index could be achieved by the incorporation of documentation and coding changes using mortality modelling.	[40]
Khan A	Retrospective observational study	5	2010	USA	The authors sought to examine the validity of ICD used across a large tertiary referral center for both identifying and classifying patients with coronary heart disease (CHD) and to examine major sources of error.	[23]
Li J	Retrospective observational study	3	1980-1999	Australia	This study aimed for the validation of all endophthalmitis cases with ocular surgery identified for the period from 1980 to June 1999 from the Hospital Morbidity Data System (HMDS) of the WA Record Linkage Project and compared with external sources.	[48]
Lorenzoni L (a)	Retrospective observational study	2	1994,1995,1997 and 1999	Italy	This study aimed to evaluate the impact of a program of training, education and awareness on the accuracy of the data collected from hospital discharge abstracts.	[60]
Lorenzoni L (b)	Retrospective observational study	2	1994,1995 and 1997	Italy	This study aimed to evaluate the impact of a program of training, education and awareness on the quality of the data collected through discharge abstracts.	[61]
Lucyk K	Qualitative study	5	2015-2016	Canada	This study aimed to explore the potential barriers that exist for the high-quality coding of administrative health data, through qualitative inquiry into the roles and responsibilities of Canadian medical chart coders.	[52]
Lungeanu D	Retrospective observational study	3	2006	Romania	This study aimed at identifying a suitable data analysis approach to investigate potential patterns in the current medical coding in obstetrics and perinatal care.	[43]

Macaulay EM	Prospective observational study	4	1994	UK	In this study, the accuracy of 637 discharge summaries was audited prospectively over an 8-month interval in a single surgical unit.	[62]
Marques B	Retrospective observational study	3	2004-2011	Portugal	This work should allow the comparison between the quality of the data collected by medical departments for clinical purposes, and the data collected in central hospital databases for administrative and epidemiological purposes.	[81]
Mercier F	Prospective observational study	4	2008-2009	France	The purpose of this multicenter study was to develop and assess the validity of novel patient safety indicators in thyroid surgery.	[57]
Michel JL	Retrospective observational study	3	2006-2007	Australia	This study aimed to examine differences between Queensland and Victorian coding of hospital-acquired conditions and suggest ways to improve the usefulness of these data in the monitoring of patient safety events.	[58]
Muzerengi S	Retrospective observational study	3	2009-2013	UK	This study aimed to estimate the number of Parkinson's disease admissions to a large Birmingham hospital and to assess the coding accuracy of the admissions.	[45]
Naran S	Retrospective observational study	3	2011-2012	UK	This study aimed to study the accuracy of clinical coding within oral surgery and to identify ways in which it can be improved.	[63]
Nilson F	Retrospective observational study	1	1987-2012	Sweden	This study aimed to investigate the impact of implementing the tenth revision of the ICD on external cause injury trends in hospitalizations (ICD chapter XX) in Sweden by time series intervention analysis of the proportion of injuries registered with missing external cause codes.	[30]
Nouraei (a)	SA Retrospective observational study	2	2010-2013	UK	This study aimed to develop a multidisciplinary coded dataset standard for nasal surgery and to assess its impact on data accuracy.	[70]
Nouraei (b)	SA Retrospective observational study	4	N.A.	UK	This study aimed to examine the accuracy of clinical coding in a large sample of emergency medical admissions to better characterize the scope and limitations of the administrative dataset as a clinical tool in emergency medicine.	[54]
Nouraei (c)	SA Retrospective observational study	1	2010-2013	UK	This study aimed to provide an account of the current scope and limitations of the administrative data and examine some of the ways in which its accuracy and clinical relevance could be improved by reporting the results of a 2-center study of clinical coding subjectivity, variability, and error (SVE) across surgery.	[19]

O'Malley KJ	Narrative review	11	N.A.	USA	This study aimed to examine potential sources of errors at each step of the described inpatient International Classification of Diseases (ICD) coding process.	[44]
Park YT	Retrospective observational study	1	2006-2007	USA	This study aimed to clarify the relationship between health information technology and Case Mix Index.	[37]
Pellegrin KL	Retrospective observational study	1	2008-2012	USA	This study aimed to determine the impact of statewide hospitals collaboration on key measures of race/ethnicity documentation.	[83]
Prasad A	Retrospective observational study	3	2012-2013	USA	This study aimed to determine the accuracy of cardiac surgical data in the administrative UHC database vs data in the clinical Society of Thoracic Surgeons database.	[76]
Prins H	Systematic review	4	N.A.	The Netherlands	This study aimed to investigate the quality of diagnostic inpatient hospital discharge data as reported in scientific journals.	[26]
Rabia L	Methodological study	3	N.A.	Algeria	The authors address the definition, the categorization and the automatic detection of dirty data quality that could be revealed in discharge summaries (DS).	[78]
Rangachari P (a)	Retrospective observational study	5	2000-2004	USA	This study aimed to explore the relationship between the organizational knowledge exchange network structure related to quality and hospital coding accuracy for quality measurement.	[59]
Rangachari P (b)	Retrospective observational study	4	2000-2004	USA	This study aimed to identify strategies not only for improving hospital coding performance but also for the hospital organization to adapt to the changing environment of quality reporting.	[72]
Reid BA	Combination of study designs (Review, Qualitative and retrospective observational)	4	2015-2016	Ireland	This study aimed to identify best practice in the management of the coding service.	[75]
Ressler MA	Experimental and qualitative study	10	N.A.	USA	The authors aimed to analyze sources of coding variations to understand how clinical coding professionals arrive at postoperative ileus (POI) coding decisions and to verify existing knowledge that current clinical coding practices lack standardized applications of regulatory guidelines.	[33]

Roberts L	Retrospective observational study	4	2014-2015	UK	This study aimed to assess the quality of causative agent coding and quantify whether its impact on secondary uses— notably financial reimbursement—is of significance.	[36]
Rodenberg H	Retrospective observational study	3	N.A.	USA	In this article, the authors share four examples of studies that might be included within the expanding role of clinical documentation improvement (CDI) programs. The authors explored the use of claims data to determine whether a clinical condition meets the criteria for a secondary diagnosis and to evaluate whether a clinical problem should be elevated to the status of a comorbid or complicating condition.	[55]
Silverman E	Retrospective observational study	2	1989-1997	USA	This study aimed to examine at the national level whether and how hospital ownership and market structure affected upcoding behavior during 1989–1997.	[68]
Soo IH	Retrospective observational study	3	2001/02-2005/06	Australia	This study aimed to examine the effects of the additional activity codes on data quality within the NHMD.	[66]
Souza J	Retrospective observational study	3	2011-2016	Portugal	The authors present a machine learning approach to explore the technical structure of the APR-DRG classification in order to characterize the impact of underreporting secondary diagnoses representing Elixhauser and Charlson co-morbidities on APR-DRG classification and hospital funding.	[32]
Stausberg J	Retrospective observational study	2	N.A.	Germany	This study aimed to compare the reliability of diagnoses coding with ICD-10 between three groups of coding subjects.	[22]
Steinbusch PJ	Narrative review	4	N.A.	The Netherlands	By comparing the US, Australian and Dutch casemix systems, this article analyses which systems are more open to upcoding than other systems and why.	[87]
Surján G	Narrative review	3	N.A.	Hungary	This study aimed to investigate—based on the available literature—how serious the problem of validity of coded data is.	[46]
Tsopra R	Prospective observational study	3	2015	UK	This study investigates the impact of three alternative coding methods on the inaccuracy of diagnosis codes and hospital reimbursement: using (i) the discharge summary alone, (ii) case notes and discharge summary, and (iii) discharge summary with the addition of medical input.	[88]

Table B2 – Ishikawa categorization and description of the root causes

Root Cause #	Ishikawa category	Ishikawa subcategory	Root cause name	Root cause description / Example	Reference
1	Machine	Clinical coding classification	Complexity of coding classification / rules	Coding in ICD-10-CM/PCS is difficult. The new classification system is more complex and coders cannot memorize the codes, making coding more laborious. / The ICD-10 entails increased complexity and specificity that makes harder to select the correct code and increase the need for training. / Classifications and coding rules have to be radically simplified. It is also erroneous to think that detailed classifications are a prerequisite of a valid image of the clinical situation. Previous studies have shown that detailed data are neither necessary for grouping in the DRG-system nor a realistic performance measure of surgical work. / Stausberg et al. reveal weak results concerning the reliability of diagnoses coding with ICD-10	[20-22]
2			ICD terminology limitations / lack of discriminatory detail	Descriptors that accompany the codes at the time of selection are often more comprehensive but map to the same numeric code, meaning that differentiation is lost when the descriptor disappears. ICD codes, for instance, indicate the underlying diagnosis but do not provide further information about prior repairs or anatomical features that may be associated with outcomes. Despite its comprehensive nature, ICD 9 was never designed to extensively differentiate the congenital heart disease spectrum. / Several authors criticize the limitations of the ICD: the vague definitions of ICD codes, Guillain-Barré syndrome is included in code 357.0 and does not have a separate diagnostic code, borderline tumors could not be excluded from the code for ovarian cancer, in general, in ICD there is no one single code for a disorder, the codes are not mutually exclusive and definitions are more based on pathological than on clinical information, lack of detail limited by the restrictive nature of the numbering system, the strict hierarchical structure and changes of meaning of terms when the terminology is updated. / Farzandipour et al. reported a nonsignificant improvement in coding accuracy in their study that was thought to be related to the structure of ICD-9-CM. Particularly, the weaknesses of ICD-9-CM for classifying specialized procedures.	[23-27]
3		Guidelines and consensus	Inefficient management of reviewed / discontinued ICD codes	Clinical classification systems are updated regularly with new, expired codes and or description presenting difficulties for a Clinical Documentation Improvement program to maintain currency. / The process associated to the management of ICD codes modifications is not performed efficiently, leading to the utilization of discontinued diagnostic and procedure codes.	[28, 29]
4			Transition between coding classification systems	Experienced ICD-9-CM coders stated that coding is easier to those who began already with ICD-10-CM/PCS. The participant who worked only with ICD-10-CM/PCS said that he/she also has many difficulties. / The changes in ICD coding between ICD-9 and ICD-10 have caused problems in the registration of hospitalized injuries in Sweden. The two coding systems differ in the number of codes available with almost five times as many diagnosis codes in ICD-10 compared with ICD-9, causing in the long-term problems with compatibility and conversion between the two coding systems. While some external code categories are convertible from the primary code, for example, fall-related and transport-related injuries, for many other external causes this is not possible.	[20, 30]
5			Use of two classifications	Some coding errors could also be due to the use of two classifications, one for the diagnosis and one for the procedure.	[31]
6		Guidelines and consensus	Deficient guidelines for coding comorbidities	Sousa et al. we recommended that future guidelines to be used by medical coders should include specific rules concerning coding of co-morbidities.	[32]
7			Lack of guidelines and consensus	Coding procedures arise uncertainty for participants, due to the lack of guidelines and consensus. There is variation in coding due to lack of normalization/consensus.	[20]
8			Multiple sources of advice	The official ICD-9-CM conventions, the CMS Coding Guidelines, and the AHA Coding Clinic advice. (To be applied in this order)	[33]

9	Man	Tools related with source information	Unstandardized / ambiguous coding guidelines	Clinical coding practices lack standardized applications of regulatory guidelines. / Clarity of clinical coding standards, access to reference materials and inconsistent clinical coding standards across provinces (i.e. use of terminology or mandatory comorbidity codes) were recurring issues related to clinical coding guidelines.	[33, 34]
10			Impossibility of making explanatory drawings in EHR	EHR do not allow for making explanatory drawings that could help in the surgical areas.	[35]
11			Inadequate discharge summaries standards	Roberts et al. reported that there is no current standard for ensuring discharge summaries are fit for coding purposes.	[36]
12			Increased information technology adoption	Park et al. found that IT was positively associated with CMI, indicating that increased IT adoption could lead to a higher CMI or billing though DRG up-coding. The authors concluded that the positive association between IT and CMI implies that IT adoption itself could lead to higher patient billings, and measures to prevent upcoding through IT systems should be implemented.	[37]
13			Limitations of the computer system	According to Alonso et al. one coder reported that the computer system on the eighth day after discharge automatically closes [and blocks] the record. Blocking the inclusion of a delayed discharge note. / Carlson et al considered systems issues, such as insufficiencies in the electronic system in which coders enter diagnosis and E-codes, a likely reason for the observed incomplete E-coding. Coders have only 13 fields in which they can enter ICD-9-CM diagnosis codes other than the principal diagnosis code. / One pertinent example came from two hospitals where coders did not have online access to the pathology results and X-ray reports necessary to allocate the correct code.	[35, 38, 39]
14	Man	Coders and health care providers	Limited awareness of the impact of inaccurate complication coding	There is a need to increase clinical coders' and physicians' awareness of how inaccurate complication coding affects quality reporting metrics.	[33]
15			Mismatch of coding terminology and clinical language	Differences between descriptors of the ICD coding system and everyday clinical terminology also contribute to inaccurate and incomplete coding. / Variable definitions between coding professionals and clinicians. Coding guidelines differ from clinical criteria.	[33, 40, 41]
16			Polarized perspectives of coders and physicians	Front-line coder and physician respondents share the polarized perspectives of their respective national organizations, AHIMA and AMA, regarding when and how to implement ICD-10, desired levels of specificity for codes, and whether to integrate documentation for clinical and billing and objectives. Physicians felt that selecting more specific codes and producing more detailed documentation to support the codes was not beneficial for clinical care and reduced their excitement about their work, potentially leading to increased burnout and desire to stop practicing.	[21]
17		Clinical coders	Coders' experience	The clinical coders valued experience more than education or credentials. According to Ressler et al. experience offered a frame of reference for interpreting the guidelines: with experience, and with payer denials, a clinical coder understands that just because a diagnosis is documented, does not necessarily mean it gets coded; the clinical parameters must be present.	[27, 33, 42-44]
18			Coders' insufficient clinical knowledge	Coding inaccuracies for inpatient spells may be due to poor recording of secondary diagnosis codes by coders, who are not medically trained. / Roberts et al. considered that insufficient clinical knowledge, misinterpretation of medical terminology, and coder competency level could explain the assignment of the incorrect code, J21.9: Acute bronchiolitis, unspecified.	[36, 45]
19			Coders' lack of education	It frequently happens however that the indexing is done routinely by uneducated clerks, who cannot do more than select one item from a prepared short list of commonly used diagnoses.	[46]
20			Coders' lack of knowledge on a specific condition	Misclassification can occur because of lack of understanding on the part of those entering the codes. For example, support staff without specific expertise in congenital heart disease (CHD) may theoretically be more apt to select a generic code such as congenital heart anomaly in place of a more specific code. Physicians or administrative personnel in charge of coding may lack expertise in CHD terminology (especially in adult hospitals that are less familiar with CHD coding)	[23, 27, 47, 48]
21			Coding bias	Certain acute medical conditions or complications tend to be regarded by medical coders as more important than others, thereby originating this coding bias.	[32]

22	Health care providers	Coding infrequent events	Joos et al findings indicate that unverified ICD-10 codes erroneously identify a large number of bleeding events (false positives). The observed limited utility of using ICD-10 codes to identify hospitalizations for acute bleeding in this population may be associated with the relatively low rates of bleeding resulting in hospitalization. / It is difficult to establish systematic and consistent coding practices for cases that are infrequently seen by coders.	[49, 50]
23		Coding learning curve	Coders' knowledge and skills in using ICD-10 coding methods and ICD-10 guidelines may have improved with time (i.e., a coding 'learning curve'), contributing to an improved adherence to coding guidelines.	[51]
24		Coding subjectivity and variability	To the extent that variation and uncertainty are intrinsic to the practice of medicine, they are ineliminable from the process of reflecting medical care in codes, for even if every source of documentation ambiguity was to be eliminated in every patient every time, diagnostic uncertainty and competition will always require interpretation and judgment and as such, abstraction of codes will always remain subject to legitimate variation.	[19, 24, 49, 52-54]
25		Definition of secondary diagnosis is subject to interpretation	Rodenberg et al. state that while the definition of secondary diagnosis may seem clear, the interpretation involves a plethora of nuances. What constitutes treatment? How do you measure increased nursing care and monitoring? Might some conditions, simply by their presence and without specific mention of diagnostic tests or focused treatment, meet the definition of a secondary diagnosis, and if so, how could we assess this?	[55]
26		Inadequate procedural coding education	There is no single source for learning CPT coding and courses and literature on coding are often inaccurate or plain dishonest.	[56]
27		Inter-coder variation	Mercier et al. assumed that the difference in correlation performance from the one hospital is related to an individual coder performance. / Januel et al. considered that it is hard to avoid inter-coder variation in the quality of coding in Switzerland since there are few training programs for coders and lay persons or clinically trained nurses and physicians are coding charts after only a short training period.	[38, 51, 57, 58]
28		Lack of knowledge of coding guidelines	Coders lacked knowledge of coding guidelines and/or the linkages among documentation, coding, and quality measurement. / Some coders do not know whether the guidelines are mandatory and some admit to not reading them.	[39, 59]
29		Lack of understanding of procedure coding	During the review of period 1 at both academic institutions, there was a clear lack of understanding of how to appropriately code these procedures despite being among the more common procedures performed by vascular surgeons.	[53]
30		Medical record administrators' lack of education	This study underpins the need for formal education programs for medical record administrators (e.g. university programs) to prepare professionals in charge with supervising medical coding activities in hospitals.	[43]
31		Poor training and support interventions to coders	Exposure to the training and support intervention (e.g. training in fundamentals of ICD coding, senior review of discharge summaries prepared by junior staff, access to an in-house-developed online ICD coding training course, and on-site support of a case manager designated to support users of software assisting ICD coding) had the strongest association with ICD coding completeness.	[41, 60, 61]
32	Health care providers	Reluctance to record more than one code	Inaccuracy in coding secondary diagnosis might result from the reluctance of a coder to record more than one code.	[42]
33		Clinicians' experience	In this unit the vast majority of discharge summaries are prepared by junior staff. With decreasing hours, cross-cover and holidays, junior doctors often perform summaries on patients with whom they are not familiar.	[62]
34		Clinicians' lack of awareness / education on the coding process and purpose	Doctors may not fully understand that their documentation has many uses that extend well beyond patient care. / Lack of knowledge on the coding protocols and HRG generation, as well as, the clinical experience, laboratory logistics and professional cultural differences may lead to the paucity of coding-relevant clinical information. Incentivizing coding-focused documentation might address these issues. / Providers are rarely, if ever, trained in the utility of administrative coding.	[23, 28, 36, 45, 63, 64]

35	Management	Leadership involvement	Clinicians' lack of time and motivation to document records	Recording structured data in the consultation takes time, can be distracting and depends on the level of motivation of the clinician.	[28, 65]
36			Clinicians' unawareness of documentation needs	Clinicians may be unaware of the level of documentation necessary to detail the complete circumstances surrounding the injury event in the medical records because their priorities lie in the immediate clinical care of the patient especially if the patient was admitted for an injury.	[66]
37			Difficulties / uncertainty in establishing a diagnosis	Clinical difficulties in establishing an accurate diagnosis could explain the misuse of ICD-10 codes. / For example, AMI is characterized by diagnostic uncertainty, manifesting itself in marked differences in sensitivity and specificity of the measurement of myocardial proteins, ECG recordings and imaging modalities. The absence of national guidelines leaves the choice of the diagnostic criteria to the individual clinicians. / The inherent uncertainty in diagnosis provided fertile ground for consultants advising hospitals on how to maximize Medicare billing.	[23, 26, 31, 67, 68]
38			Lack of assertiveness in diagnostic documentation	The lack of assertiveness of some physicians who, instead of interpreting exam reports and making a diagnosis, only transcribed exam reports. In some services, there are quite extensive notes, in which exams are transcribed but diagnosis may not be specified. / It requires a clinician to commit to documenting what is a significant clinical diagnosis. There are diagnoses that are based on the presence of objective criteria, but presence of a positive diagnostic test does not unequivocally imply presence of a clinical diagnosis.	[35, 54]
39			Student documentation	Student documentation is more likely to be coded at a lower level of service and be assigned a different ICD-10 diagnostic code compared to faculty. These errors in level of service are primarily due to students failing to document a sufficient number of elements in the evaluation, failure to specify studies ordered, and low medical decision making.	[69]
40			Undercoding of diagnoses and morbidities	Physicians tend to under evaluate the extra time and services provided to other secondary problems dealt with during hospitalization, which results in undercoding of such problems. / Muzerengi et al. showed a significant under-coding of Parkinson's disease on discharge from hospital by identifying episodes in which Parkinson's disease medications were prescribed during hospitalization, but Parkinson's disease diagnosis was not coded on discharge. Since Parkinson's disease was a secondary diagnosis during these admissions, the diagnosis was possibly less likely to be recorded on discharge. / Nouraei et al. observed under-coding of many nasal diagnoses and morbidities.	[45, 70, 71]
41			Lack of feedback	The monthly feedback of information to professionals has proven to have a major impact on the improvement in the quality of data. Charts and discharge abstract reviews with professionals, at department level, to highlight and discuss errors in abstracting appear to be the most effective tool with which to correct systematic errors (...)	[61]
42			Lack of knowledge of health information management systems	Poor coding facilities were found to possess minimal knowledge and awareness of HIM systems and measurement science. / Quality administrators in poor-coding facilities showed minimal knowledge of HIM systems and aggregate data management, in general.	[59, 72]
43			Lack of senior leadership involvement	In poor-coding facilities, interviews noted negligible involvement from senior management in quality measurement and quality related priority setting. / None of the 3 subgroup administrators was structurally integrated with the senior hospital administration. As such, Figure 6 shows a lack of senior leadership involvement in knowledge exchange related to quality.	[59, 72]
44			Lack of support and respect for the coding function	Poor coding facilities expressed a lack of support and/or respect for the HIM/coding function. They attributed coding issues to coders' technical incompetence and did not provide for the role of documentation or system limitations in coding issues.	[59, 72]
45			Pressure from administration and health care actors	Georgescu et al. show that upcoding is a consequence of financial pressure, and that the kind of upcoding depends on the source of pressure: pressure from actors in the administrative core and pressure from health care actors.	[73]
46			Unsatisfactory education and engagement of	Education and engagement of hospital administrators, information technologist and researchers within the health service domain for improvement in data quality should be encouraged.	[74]

			hospital administrator, information technologists and researchers		
47		Limited resources	Coders' quotas and expectations	The presence of a quota added pressure to coding that may influence data quality. Where coders face pressures of benchmarking, there exists the potential to focus on productivity and quantity over quality, which may result in a lack of attention to detail among coders	[52]
48			Coders' time constraints and workload	Shorter work periods in the coding unit may result in fewer interactions with physicians and hence little knowledge about procedures, which might lead to major errors. / Coders reported to the auditors that they did not have time to read all the allied health and nursing notes because of the turnaround time requirements. / Coders generally monitor their own workloads. There are very few incentive schemes at a local level for meeting productivity targets or data reporting timelines, even though coding deadlines are linked to financial rewards.	[27, 39, 40, 42, 75]
49			Lack of budget and time	Limited resources, mainly budget and time, create team dissatisfaction and compromise data quality. Lack of continuing education, staffing support, fixed and tight budgets and quick turnaround times impact clinical coders' ability to produce good quality clinically coded data.	[34]
50			Poor investment in documentation / education of physicians	The study found that good coding facilities made substantial investments in documentation improvement and education and training of physicians. / This suggests that hospital leaders may need to invest in documentation improvement programs and the education/training necessary to communicate linkages between documentation, coding, and quality measurement.	[59, 72]
51			Priority of coding	The priority of clinical coding within an organization represents an important barrier for timely coding admissions.	[65]
52		Other	Condition not formally acknowledged as to complicate the clinical care of the patient	Some clinical conditions and scenarios are thought by providers to clinically complicate patient care but are not reflected in the list of complications or comorbidities issued each year by the ICD-10 Coordination and Maintenance Committee of the Centers for Medicare and Medicaid Services (CMS) and the Centers for Disease Control and Prevention. (e.g. Endotracheal Intubation for Airway Protection)	[55]
53			Lack of incentives for comprehensive coding	No financial benefit to the institution for further effort on the coder's part to continue to capture a patient's comorbidities.	[40]
54			Lack of routine and systematic internal coding audits	Administrative databases are fraught with potential for error because information is aggregated from different systems and is often unaudited. / In a focus group study by Alonso et al. participants stated that audits in ICD-10-CM/PCS do not exist and advanced some causes: lack of auditors, lack of time, the impossibility to use the software previously developed in Portugal to perform audits as it available only for ICD-9-CM coded episodes.	[20, 28, 76, 77]
55	Material	Incomplete / missing information	Incomplete / missing documentation of diagnostic information	Incomplete or missing discharge note or surgical report. Doctors delay in presenting the discharge note. / The documentation that healthcare providers complete to describe a patient's encounter with the healthcare facility is perhaps the resource most influential to data quality, as it constitutes the information that is interpreted by coders, abstracted, and sent to institutional repositories. As such, issues that occur upstream in the process of data generation have the potential to adversely influence data quality downstream for coders, researchers and analysts, and decision-makers. / Poor documentation often resulted from missing or miscommunicated information.	[28, 32, 35, 36, 38, 40, 44, 46, 52, 66, 78-80]
56			Lack of contextual information	Lack of contextual information in structured disease diagnoses – e.g. for some patients, the pancreatic tumor was not primary but metastatic, while pathology reports only captured information for the primary tumors.	[79]
57			Lack of meaningful codes	In the absence of a "present on admission" qualifier, it is difficult to distinguish postoperative complications that occurred during a hospital stay from patient comorbidities. / The interviews revealed that senior health information managers do not appear to have been made aware of the benefits and uses of the condition onset flag, claiming they collect the item only because corporate office requires it.	[57, 58]

58		Limited clinical and socio-economic detail	Administrative databases often do not contain sufficient detail on a disease or procedure to aid in analysis of outcomes. / In the context of AMI, important information, such as symptom-onset to needle-time, the time lag between the onset of symptoms and the initiation of the treatment, is lacking in this type of data, as well as pharmacological treatment details, which are essential with respect to quality assessment, especially in primary hospitals. Also, although low socio-economic status (SES) is related with lesser AMI-outcomes, there is no information regarding SES.	[67, 76]
59		Missing / incorrect codes or values	Missing additional diagnosis codes and/or missing procedure codes. / Missing and incorrect values occur when a required attribute is not filled or filled incorrectly, that is, there is absence of value or an error in a mandatory attribute. Freitas et al. observed missing values describing type of care, incorrect values for patient age, and negative length of stay times.	[50, 65, 67, 77, 81, 82]
60	Inconsistent information	Inconsistency	Inconsistencies are detected through violations of integrity constraint (e.g. 'Principal diagnosis: childbirth' and 'sex: male'). This discharge summary will not be grouped into a DRG because it contains two inconsistent information or data. / Freitas et al. detected different representations for the variable that represents the hospital department code. For example, the department of Pneumology had been represented as "Pneum", "Pneumo" and the code "210"	[78, 81, 82]
61		Inconsistent / conflicting information in the record	Inconsistent or conflicting information in the record was noted in a Canadian study as the cause of most coding difference between the CC and the auditor. / Differences in coding may also be attributable to conflicting documentary descriptions of the injury event, for example, it was observed that the clinical documentation for one case incorporated three different versions for mechanism of injury, place of occurrence and activity.	[24, 75]
62		Inconsistent or incomplete racial / ethnic classification	Information about the racial/ethnic subpopulations is often inconsistent, incomplete, or missing. Individuals often do not identify with broad categories, reducing data reliability and broad categories alone mask intragroup diversity, reducing data validity. For example, federal categories are inadequate for most of the population in Hawaii.	[83, 84]
63	Information fragmentation	Information fragmentation / lack of a standard organization of EHR	Information fragmentation is a big cause for information incompleteness. Uncoordinated or redundant data entries into different data sources in EHR could not only cause information discrepancies but also form big barriers to selecting reliable data sources for secondary use of EHR data.	[79]
64		Information fragmentation / lack of unique patient identifier	However, as most countries do not have unique patient identifiers and lack the ability to track patients after hospital discharge, the indicator is based on unique hospital admissions and restricted to mortality within the hospital. Thus, differences in practices in discharging and transferring patients may influence the findings.	[67]
65	Medical record factors	Availability of notes	Dictated and typed operating notes may not have found their way to case notes by the time of summary preparation.	[62]
66		Medical record organization	The turnaround time adversely affected how charts were organized, which tended to delay the coding process. Some coders explained that the charts were no longer re-organized by records management clerks once received from the unit to save costs and improve efficiency. Consequently, this resulted in frustration and inefficiencies for coders. / Lack of consistent and orderly organization of the medical records at a number of hospitals had a significant impact on the ability to code from the medical records.	[52, 62, 75]
67		Poor-quality documentation	Poor documentation in medical notes can affect the quality of the source document for coding which in turn can lead to inaccurate coding even when the coding practice is efficient. / Cheng et al. reported that the 57% of all audited episodes showing a DRG change, the change was very clearly due to issues relating to the quality of the medical record documentation, specifically the lack of proper medical record documentation at the time of the original coding. / The quality of chart documentation in hospital charts, including clarity, completeness, organization and legibility, and hybrid charts, which made it difficult to find information.	[14, 24, 33, 34, 39, 47, 48, 50, 62, 71, 75, 77, 85]
68	Other information issues	Ambiguous documentation	The likelihood of clinical coding subjectivity, variability and error was clearly associated with the ambiguity of the informatics handover' between clinicians and coders and this was in turn a function of the nature of the diagnosis being captured	[54]
69		Duplicated episodes	Patients have different registries in the same date, at the same hospital, and matching all clinical information (e.g. diagnosis and procedures).	[82]

70			Imprecise discharge summary values for DRG grouping	Imprecise discharge summary values do not contain sufficient information of the actual value and will (e.g. the main diagnosis has the value "J40" which represents bronchitis, but it is not specified as acute or chronic (without precision)	[78]
71			Incorrect or absent documentation of a procedure	The cause for most of these errors was physician error in describing catheter placement (instead of wire placement), understanding the different regions of fistula intervention (peripheral, central, and perianastomotic), and documenting secondary accesses. Incorrect documentation within the operative report and completely absent documentation with the procedure code discovered only after reviewing the images.	[53]
72			Lack of standardized recognition and documentation of transfers	Significant inaccuracy and variability in identification of transfers in the administrative dataset could lead to imprecise conclusions and potential biases.	[74]
73			Non-verifiable ex post information	False documentation of birth weights (within reasonable limits) is practically non-verifiable ex post because preterm infants generally lose some 10% of their initial weight in the first few days after birth.	[86]
74			Updates to definitions of clinical diagnoses	Rodenberg et al. analyzed the impact of changes in definitions of clinical diagnoses (e.g. Sepsis-2 or Sepsis-3). 35% of all cases meeting the Sepsis-2 criteria for the diagnosis of sepsis fulfilled the Sepsis-3 criteria. The mortality rates were 12% and 28.6% for the Sepsis-2 and Sepsis-3 respectively.	[55]
75		Readability	Illegibility	In paper health records, coders need to decipher doctors handwriting. / There were times when coders were required to apply their 'best guess' to deciphering documentation, which could result in the erroneous or imprecise capture of a diagnosis or procedure	[35, 52]
76			Unreadability	Coders believe that unreadability is a cause of coding errors. / Records with less readability had significantly more errors.	[27, 42]
77			Use of abbreviations / acronyms / synonyms / nonstandard terminology	Inconsistent use of abbreviations or nonstandard terminology may affect how the information is interpreted. / Clinicians frequently use generalized clinical terms, signs and symbols and abbreviations. While this is meaningful for communication between treating healthcare professionals for managing patients, these terms are not always able to be translated into ICD-10-AM/ACHI codes or coded to the required specificity that reflects the complexity of the patient. / Some errors are due to the shortening or lengthening wording as previously described.	[23, 28, 31, 35, 44, 63]
78	Method	Basis for coding	Ambiguous classification criteria	Classification criteria can be considered ambiguous, if they are not medically meaningful and not aligned with clinical practice and therefore open to upcoding or fraud.	[87]
79			Code sequencing rule	The apparent greater depth of coding may be attributable to Queensland's external cause code sequencing rule. In cases with multiple injuries, this rule requires coders to add additional codes to both the principal injury diagnosis and to the set of other (secondary) injury codes.	[58]
80			Coding based on case notes only	A case note is often voluminous, disorganized and contains multiple abbreviations, making it difficult for coders to extract the information they need.	[88]
81			Coding based on codebook	The findings suggest that using the ICD-9-CM book decreases coding accuracy.	[27, 42]
82			Coding based on discharge summary	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. Tsopra et al. observed that diagnostic inaccuracy was decreased by the addition of a doctor's input (point-of-care coding) or when coders used a combination of the case notes and discharge summaries.	[88]
83			Coding based on face sheet and memory - based coding	Most codes assigned by more experienced coders were mainly based on only face sheets and memory-based coding.	[42]

84		Coding team	The use of a clinician coding multi-disciplinary team resulted in a change to 24.1% of records and an increase in departmental revenue of £443 371. This suggests that clinician involvement may be a cost-effective means of improving data quality and hospital reimbursement.	[64]
85		No checks with physicians for clinical relevance	Coders must check with physician on clinical relevance of diagnoses; not all coders do this.	[33]
86		Poor usage of coding materials	(...) there is a poor usage of the medical coding instruments.	[43]
87	Coding errors	Errors at the point of attestation	Errors at the point of attestation include reviewing too quickly or not reviewing the face sheet and supporting documents, poor recall of the details of the patient's conditions, and incorrect recording on the attestation sheet.	[44]
88		Induction errors / Misspecification	The diagnosis represents a lower level of abstraction than the disease, and the number of diagnoses does not necessarily match the number of diseases. The disease is confused with one of its more or less characteristic symptoms. A single disease is described by more diagnoses. The disease is over-specified. / The attending physician attested to a narrative principal diagnosis, secondary diagnosis, or procedure not supported by the medical record.	[44, 46, 89, 90]
89		Miscoding / Misclassification	Coding unsupported by the clinical record. / Miscoding includes assignment of generic codes when information exists for assigning more specific codes, assignment of incorrect codes according to the governing rules, or assignment of codes without the physician attesting to their accuracy.	[29, 33, 38, 44, 48, 66, 70, 74, 77, 82, 84, 89]
90		Resequencing	Resequencing codes, or changing the order of them, comprises another potential error source. The hospital substituted a secondary diagnosis for the correctly attested and coded principal diagnosis. / The size and proportion of overpayments suggest an economic motivation for the 34.2% of errors attributable to resequencing.	[44, 89, 90]
91		Unbundling	When coders assign codes for all the separate parts of a diagnosis instead of assigning a code for the overall diagnosis, the practice is called unbundling. / The submission of bundled codes for payment is defined as unbundling and is considered as fraudulent (e.g. coding for a diagnostic arthroscopy when another arthroscopic surgical procedure is performed in the same joint.	[44, 56]
92	Quality of filling	Errors assigning information	Marques et al. attributed the root cause to administrative staff filling wrongly the post code as part of the address variable.	[81]
93		Transcription errors	Transcription or scanning errors are additional threats to code accuracy.	[44]
94		Use of copy and paste	Copy and paste is the origin of large transcriptions to the records without relevant information and leading to repeating errors from the first to the last daily progress note, complicating and slowing coding.	[35, 79]
95	Retrospective information	Point in time of the initial classification	It is much easier to have a complete overview of the opportunities for upcoding after the care process has finished than at the beginning of the process,	[87]
96		Retrospective coding / Information collected at discharge	All procedures that a patient undergoes during their hospital stay are systematically coded by surgeons at the time of surgery, while diagnosis related to postoperative complications may be forgotten because they are usually collected at discharge.	[57]
97		Retrospective queries for coding	Queries are generated retrospectively, after the patient has been discharged, which will be negatively affected by the passing of time and will delay the coding process.	[28]
98		Retrospective writing of discharge summary	While discharge summaries are concise documents, they can be written retrospectively due to time constraints and the information in summaries can be inaccurate.	[88]

99	Mission	Coding purpose	Coding purpose	For some uses, three-digit accuracy may be sufficient. One study included in the review by Burns et al. reported an accuracy of 51% at the four-digit level but 90% at the three-digit level suggesting that many inaccuracies occur at four-digit level.	[64]
100			Misclassification of severity / risk	Misclassification of low-risk procedures as high-risk procedures led to lower calculated mortality rate, and high-risk procedures coded as low-risk procedures led to a higher calculated mortality rate. These results could translate into inaccurate assessment of the quality of institutions and individual surgeons and could affect the ability to conduct accurate and reliable research.	[76]
101		Financial incentives or disincentives	Financial incentives or disincentives	For example, aggressive reporting could also be the result of an effort to increase payments in a volume-based payment system. / Underlying data collection is strongly influenced by economic issues, as data coders are mostly instructed to invoice patient hospitalization rather than accurately reporting adverse events. / Upcoding to maximize reimbursement is a well-known phenomenon. / The primary purpose of these data is billing; coding may be financially driven. / Increased prices per treatment with a prospective payment system give incentives to increase the number of those treatments.	[25, 29, 33, 44, 47, 49, 51, 56, 57, 60, 64, 67, 68, 86, 87, 91-94]
102			Reimbursement system	Data quality may be higher for patients over 63 or 64 years old due to the influence of the Medicare reimbursement system for patients over 64 years old where diagnostic hospital discharge data are necessary for payment and the on average higher disease severity in patients over 64 years old.	[26]
103			Fast turnover and day-case nature	The fast turnover and day-case nature of some specialties means that many patients' existing morbidities are not recorded comprehensively on admissions documents.	[63]
104	Variability of practices	Variability in coding practices	Variability in coding practices	Differences in data quality across hospitals may have to do with the local coding practices and settings: differences in diagnostic practices, differences in case-mix, clarity and structure of documentation of diagnoses in the medical record, training of medical record coders, access to patient data, time per case to code and coding instructions. / Previous literature, including our own, has demonstrated a wide variability in clinical coding accuracy across specialties from ,20% to .95%.	[26, 27, 54]
105			Variability of the quality of health records	The quality of health records is affected by great variability in content, depending on the hospital, the specialty or, even, the professional. In some hospitals, coders do not access all clinical documentation (e.g. no access to the anesthesia sheet)	[35]

Annex

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page and paragraph/ table #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both. - MANDATÓRIO	p.5 “Analysis of Root Causes of Problems Affecting the Quality of the Diagnosis Related Group-Based Hospital Data: A Systematic Review and Ishikawa Diagram”.
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number. – SEGUIR RECOMENDAÇÕES DA REVISTA	p.5-6 “Introduction: Administrative hospital databases built around Diagnosis-Related Groups (DRGs) represent an important tool for hospital financing in many national health systems and are also an important data source for clinical, epidemiological and health services research.”
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. – MANDATÓRIO <i>O rationale corresponde à justificação da importância da revisão sistemática</i>	p. 6-7 “For a number of reasons (errors, lack of standards, financial interests) fallible data can be included in these types of databases, eliciting data quality problems. This can potentially compromise hospital financing/reimbursement and their reutilization for epidemiological or health services research”
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS). - MANDATÓRIO	p.7 “We felt motivated to conduct an investigation to bridge this gap, aiming to systematically identify causes of problems affecting the quality of DRG-based administrative hospital data, whether these relate with hospital financing, epidemiological or health services research, and to propose a framework to derive and analyze a list of root causes.”
METHODS			

Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number. – FACULTATIVO	p.8 “This study implemented a systematic review, following a previously published PRISMA-based protocol”
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. – MANDATÓRIO <i>É altamente recomendado, de acordo com as boas práticas da Cochrane, que não sejam aplicados critérios de exclusão baseados na língua e/ou data de publicação dos estudos.</i>	p.8 “Studies were subsequently screen according to the following inclusion criteria: Peer-reviewed papers (including conference papers), written in English.(...)”
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched. – MANDATÓRIO <i>Em consonância com as boas práticas da Cochrane, é mandatório que se verifique pesquisa em pelo menos duas bases de pesquisa bibliográfica (idealmente, deverão ser pesquisadas duas bases generalistas e uma específica da área). No caso de revisões sistemáticas de estudos experimentais/ensaiois clínicos aleatorizados, é altamente recomendado que uma das bases pesquisadas corresponda à CENTRAL ou a bases de ensaios clínicos como a ClinicalTrials.gov.</i> <i>Estudos de revisão da literatura em que a pesquisa decorra numa única base de dados não serão classificados como revisões sistemáticas.</i>	p.8 “The MEDLINE and Scopus scientific research databases have been searched to identify potentially relevant studies reporting causes of data quality problems of DRG-based administrative hospital databases.”
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. – MANDATÓRIO <i>A query de pesquisa deve ser obrigatoriamente disponibilizada. A utilização de filtros de pesquisa da InterTASC é altamente recomendada (https://sites.google.com/a/york.ac.uk/lisssg-search-filters-resource/home)</i>	p.26 “Table A1- Medline search strategy”
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis). – MANDATÓRIO <i>As fases de selecção dos estudos primários devem ser descritas. Em consonância com as boas práticas da Cochrane, é mandatório que o processo de selecção envolva duas fases (fase de rastreio, em que os registos são seleccionados por título e abstract, e fase de inclusão, na qual se procede à leitura integral dos full texts). Em cada uma destas fases, o processo de selecção deve mandatoriamente envolver dois investigadores actuando de forma independente.</i>	p.9 “Three reviewers (ARO, ML, MO) independently screened the titles and abstracts, two for each study. Disagreements were solved by consensus with a third reviewer.”
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators. – MANDATÓRIO <i>Trata-se de descrever de que forma se procedeu à extracção de dados dos estudos primários. Em consonância com as boas práticas da Cochrane, tal processo deverá envolver dois investigadores de forma independente.</i>	p.9-10 “Data-charting consisted in extracting information relevant for the study of root causes of data quality problems.”

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made. – MANDATÓRIO <i>Trata-se de descrever as variáveis para as quais foi obtida informação.</i>	p.9 “This was accomplished using a predefined form including several items: 1) study title, 2) type of study (i.e. observational, experimental, qualitative), 3) country of origin of the data, 4) data reporting years, 5) study description, 6) the name of the root cause, 7) the description of root causes identified, including, when possible, an example of the root cause, 8) a sequential identification number, and 9) the study where the root cause had been extracted from.”
Risk of bias in individual studies / Risk of bias across studies	12/15	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis. – MANDATÓRIO <i>Em todas as revisões sistemáticas, deverá existir um processo de avaliação da qualidade dos estudos primários. No caso de revisões sistemáticas de estudos experimentais/ensaio clínicos aleatorizados, a aplicação dos critérios de risco de viés (Risk of Bias) da Cochrane é altamente recomendada. No caso de revisões sistemáticas de estudos observacionais, poderão ser seguidos os critérios ROBINS ou os critérios dos National Institutes of Health (https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools).</i>	p.9 “Evaluation of the quality of the studies was not considered as an exclusion criterion as we privileged comprehensiveness over study quality”
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means). – FACULTATIVO. APENAS NECESSÁRIO SE FOR FEITA META-ANÁLISE	Não aplicável, uma vez que esta revisão sistemática não se acompanhou de meta-análise.
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis. – FACULTATIVO. APENAS NECESSÁRIO SE FOR FEITA META-ANÁLISE	Não aplicável, uma vez que esta revisão sistemática não se acompanhou de meta-análise.
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified. – FACULTATIVO. APLICÁVEL APENAS SE FOR FEITA META-ANÁLISE	Não aplicável, uma vez que esta revisão sistemática não se acompanhou de meta-análise.
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram. – MANDATÓRIO	p.11 “ From the total of 2009 retrieved studies, 238 studies were assessed for eligibility based on their full-texts and 77 studies were included for extraction of root causes (Figure 1). Figure 1 – PRISMA flow diagram for the work conducted in this investigation
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations. – MANDATÓRIO	p.27-34 Table B1- Study characteristics

Risk of bias within and across studies	19/22	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12). – MANDATÓRIO	p.9 “Evaluation of the quality of the studies was not considered as an exclusion criterion as we privileged comprehensiveness over study quality
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot. – FACULTATIVO. APLICÁVEL APENAS SE FOR FEITA META-ANÁLISE	Não aplicável, uma vez que esta revisão sistemática não se acompanhou de meta-análise.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency. – FACULTATIVO. MANDATÓRIO APENAS SE FOR FEITA META-ANÁLISE	Não aplicável, uma vez que esta revisão sistemática não se acompanhou de meta-análise.
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]). – FACULTATIVO. APLICÁVEL APENAS SE FOR FEITA META-ANÁLISE	Não aplicável, uma vez que esta revisão sistemática não se acompanhou de meta-análise.
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers). – MANDATÓRIO	p.16-17“This study provides a comprehensive list of root causes of data quality problems affecting DRG-based administrative hospital databases and proposes a framework (the Ishikawa classification) for deriving and analyzing such list.”
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias). – MANDATÓRIO	p.18 “This review has limitations. We have only searched studies in two databases and have not included grey literature.”
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research. – MANDATÓRIO	p.19 “The presented systematic review sets forward a catalogue of 105 individual root causes of data quality problems affecting DRG-based administrative hospital databases.”
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. – SEGUIR RECOMENDAÇÕES DA REVISTA	p.19 “This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.”

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Annex

GUIDE FOR AUTHORS

Your Paper Your Way

We now differentiate between the requirements for new and revised submissions. You may choose to submit your manuscript as a single Word or PDF file to be used in the refereeing process. Only when your paper is at the revision stage, will you be requested to put your paper in to a 'correct format' for acceptance and provide the items required for the publication of your article.

To find out more, please visit the Preparation section below.

Aims and Scope

International Journal of Medical Informatics provides an international medium for dissemination of original results and interpretative reviews concerning the field of medical informatics. The Journal emphasizes the evaluation of systems in healthcare settings. The scope of journal covers: Information systems, including national or international registration systems, hospital information systems, departmental and/or physician's office systems, document handling systems, electronic medical record systems, standardization, systems integration etc.;

Computer-aided medical decision support systems using heuristic, algorithmic and/or statistical methods as exemplified in decision theory, protocol development, artificial intelligence, etc. Educational computer based programs pertaining to medical informatics or medicine in general; Organizational, economic, social, clinical impact, ethical and cost-benefit aspects of IT applications in health care.

Short technical communications concerning (solved) problems in implementing or using existing information systems are welcome. Review articles concerning subjects falling in the scope of the journal are also invited.

General Considerations

IJMEDI has adopted the guidelines of the International Committee of Medical Journal Editors (ICMJE). Some of the important issues are noted below. Visit <http://www.icmje.org> for more details.

Contact details

All submissions should be made through Editorial Manager at <https://www.editorialmanager.com/IJMEDI>). If you have any questions or problems relating to the preparation or submission of your manuscript, please visit our Support Centre.

Submission checklist

You can use this list to carry out a final check of your submission before you send it to the journal for review. Please check the relevant section in this Guide for Authors for more details.

Ensure that the following items are present:

One author has been designated as the corresponding author with contact details:

- E-mail address
- Full postal address

All necessary files have been uploaded:

Manuscript:

- Include keywords
- All figures (include relevant captions)
- All tables (including titles, description, footnotes)
- Ensure all figure and table citations in the text match the files provided
- Indicate clearly if color should be used for any figures in print

Graphical Abstracts / Highlights files (where applicable)

Supplemental files (where applicable)

Further considerations

- Manuscript has been 'spell checked' and 'grammar checked'
- All references mentioned in the Reference List are cited in the text, and vice versa
- Permission has been obtained for use of copyrighted material from other sources (including the Internet)

- A competing interests statement is provided, even if the authors have no competing interests to declare
- Journal policies detailed in this guide have been reviewed
- Referee suggestions and contact details provided, based on journal requirements

For further information, visit our Support Center.

BEFORE YOU BEGIN

Ethics in publishing

Please see our information pages on Ethics in publishing and Ethical guidelines for journal publication.

Studies in humans and animals

If the work involves the use of human subjects, the author should ensure that the work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The manuscript should be in line with the Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals and aim for the inclusion of representative human populations (sex, age and ethnicity) as per those recommendations. The terms sex and gender should be used correctly. Authors should include a statement in the manuscript that informed consent was obtained for experimentation with human subjects. The privacy rights of human subjects must always be observed. All animal experiments should comply with the ARRIVE guidelines and should be carried out in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and associated guidelines, EU Directive 2010/63/EU for animal experiments, or the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978) and the authors should clearly indicate in the manuscript that such guidelines have been followed. The sex of animals must be indicated, and where appropriate, the influence (or association) of sex on the results of the study.

Informed consent and patient details

Studies on patients or volunteers require ethics committee approval and informed consent, which should be documented in the paper. Appropriate consents, permissions and releases must be obtained where an author wishes to include case details or other personal information or images of patients and any other individuals in an Elsevier publication. Written consents must be retained by the author

but copies should not be provided to the journal. Only if specifically requested by the journal in exceptional circumstances (for example if a legal issue arises) the author must provide copies of the consents or evidence that such consents have been obtained. For more information, please review the Elsevier Policy on the Use of Images or Personal Information of Patients or other Individuals. Unless you have written permission from the patient (or, where applicable, the next of kin), the personal details of any patient included in any part of the article and in any supplementary materials (including all illustrations and videos) must be removed before submission.

Declaration of interest

All authors must disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work. Examples of potential competing interests include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding. Authors must disclose any interests in two places: 1. A summary declaration of interest statement in the title page file (if double anonymized) or the manuscript file (if single anonymized). If there are no interests to declare then please state this: 'Declarations of interest: none'. This summary statement will be ultimately published if the article is accepted. 2. Detailed disclosures as part of a separate Declaration of Interest form, which forms part of the journal's official records. It is important for potential interests to be declared in both places and that the information matches.

Submission declaration and verification

Submission of an article implies that the work described has not been published previously (except in the form of an abstract, a published lecture or academic thesis, see 'Multiple, redundant or concurrent publication' for more information), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright holder. To verify originality, your article may be checked by the originality detection service Crossref Similarity Check.

Preprints

Please note that preprints can be shared anywhere at any time, in line with Elsevier's sharing policy.

Sharing your preprints e.g. on a preprint server will not count as prior publication (see 'Multiple, redundant or concurrent publication' for more information).

Use of inclusive language

Inclusive language acknowledges diversity, conveys respect to all people, is sensitive to differences, and promotes equal opportunities. Content should make no assumptions about the beliefs or commitments of any reader; contain nothing which might imply that one individual is superior to another on the grounds of age, gender, race, ethnicity, culture, sexual orientation, disability or health condition; and use inclusive language throughout. Authors should ensure that writing is free from bias, stereotypes, slang, reference to dominant culture and/or cultural assumptions. We advise to seek gender neutrality by using plural nouns ("clinicians, patients/clients") as default/wherever possible to avoid using "he, she," or "he/she." We recommend avoiding the use of descriptors that refer to personal attributes such as age, gender, race, ethnicity, culture, sexual orientation, disability or health condition unless they are relevant and valid. These guidelines are meant as a point of reference to help identify appropriate language but are by no means exhaustive or definitive.

Authorship

All authors should have made substantial contributions to all of the following: (1) the conception and design of the study, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for important intellectual content, (3) final approval of the version to be submitted.

Acknowledgements

All contributors who do not meet the criteria for authorship as defined above should be listed in an acknowledgements section. Examples of those who might be acknowledged include a person who provided purely technical help, writing assistance, or a department chair who provided only general support. Authors should disclose whether they had any writing assistance and identify the entity that paid for this assistance. Financial support like grants should also be mentioned in the acknowledgements section.

Changes to authorship

Authors are expected to consider carefully the list and order of authors **before** submitting their manuscript and provide the definitive list of authors at the time of the original submission. Any

addition, deletion or rearrangement of author names in the authorship list should be made only **before** the manuscript has been accepted and only if approved by the journal Editor. To request such a change, the Editor must receive the following from the **corresponding author**: (a) the reason for the change in author list and (b) written confirmation (e-mail, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed. Only in exceptional circumstances will the Editor consider the addition, deletion or rearrangement of authors **after** the manuscript has been accepted. While the Editor considers the request, publication of the manuscript will be suspended. If the manuscript has already been published in an online issue, any requests approved by the Editor will result in a corrigendum.

Clinical trial results

In line with the position of the International Committee of Medical Journal Editors, the journal will not consider results posted in the same clinical trials registry in which primary registration resides to be prior publication if the results posted are presented in the form of a brief structured (less than 500 words) abstract or table. However, divulging results in other circumstances (e.g., investors' meetings) is discouraged and may jeopardise consideration of the manuscript. Authors should fully disclose all posting in registries of results of the same or closely related work.

Reporting clinical trials

Randomized controlled trials should be presented according to the CONSORT guidelines. At manuscript submission, authors must provide the CONSORT checklist accompanied by a flow diagram that illustrates the progress of patients through the trial, including recruitment, enrollment, randomization, withdrawal and completion, and a detailed description of the randomization procedure. The CONSORT checklist and template flow diagram are available online.

Registration of clinical trials

Registration in a public trials registry is a condition for publication of clinical trials in this journal in accordance with International Committee of Medical Journal Editors recommendations. Trials must register at or before the onset of patient enrolment. The clinical trial registration number should be included at the end of the abstract of the article. A clinical trial is defined as any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects of health outcomes. Health-related interventions

include any intervention used to modify a biomedical or health-related outcome (for example drugs, surgical procedures, devices, behavioural treatments, dietary interventions, and process-of-care changes). Health outcomes include any biomedical or health-related measures obtained in patients or participants, including pharmacokinetic measures and adverse events. Purely observational studies (those in which the assignment of the medical intervention is not at the discretion of the investigator) will not require registration.

Copyright

Upon acceptance of an article, authors will be asked to complete a 'Journal Publishing Agreement' (see more information on this). An e-mail will be sent to the corresponding author confirming receipt of the manuscript together with a 'Journal Publishing Agreement' form or a link to the online version of this agreement.

Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution and for all other derivative works, including compilations and translations. If excerpts from other copyrighted works are included, the author(s) must obtain written permission from the copyright owners and credit the source(s) in the article. Elsevier has preprinted forms for use by authors in these cases.

For gold open access articles: Upon acceptance of an article, authors will be asked to complete a 'License Agreement' (more information). Permitted third party reuse of gold open access articles is determined by the author's choice of user license.

Author rights

As an author you (or your employer or institution) have certain rights to reuse your work.

Elsevier supports responsible sharing

Find out how you can share your research published in Elsevier journals.

Role of the funding source

You are requested to identify who provided financial support for the conduct of the research and/or preparation of the article and to briefly describe the role of the sponsor(s), if any, in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to

submit the article for publication. If the funding source(s) had no such involvement then this should be stated.

Open access

Please visit our Open Access page for more information.

Elsevier Researcher Academy

Researcher Academy is a free e-learning platform designed to support early and mid-career researchers throughout their research journey. The "Learn" environment at Researcher Academy offers several interactive modules, webinars, downloadable guides and resources to guide you through the process of writing for research and going through peer review. Feel free to use these free resources to improve your submission and navigate the publication process with ease.

Language (usage and editing services)

Please write your text in good English (American or British usage is accepted, but not a mixture of these). Authors who feel their English language manuscript may require editing to eliminate possible grammatical or spelling errors and to conform to correct scientific English may wish to use the English Language Editing service available from Elsevier's Author Services.

Submission

Our online submission system guides you stepwise through the process of entering your article details and uploading your files. The system converts your article files to a single PDF file used in the peer-review process. Editable files (e.g., Word, LaTeX) are required to typeset your article for final publication. All correspondence, including notification of the Editor's decision and requests for revision, is sent by e-mail.

Structure of manuscripts

The following types of contributions will be published: (i) Papers reporting original work; (ii) Interpretative reviews; (iii) Technical notes; (iv) Letters to the Editor. All manuscripts, except Letters to the Editor, should have the following structure:

- Title page, including keywords
- Structured abstract
- Body of the manuscript

- Authors' contributions
- Acknowledgements
- Statement on conflicts of interest
- Summary table
- References
- Appendices (if applicable)
- Maximum word count for research articles: 3000
- Maximum word count for reviews: 4000

Manuscripts not conforming to this structure may be returned to author without prejudice but without review.

Submit your article

Please submit your article via <https://www.editorialmanager.com/IJMEDI>.

PREPARATION

Queries

For questions about the editorial process (including the status of manuscripts under review) or for technical support on submissions, please visit our Support Center.

NEW SUBMISSIONS

Submission to this journal proceeds totally online and you will be guided stepwise through the creation and uploading of your files. The system automatically converts your files to a single PDF file, which is used in the peer-review process.

As part of the Your Paper Your Way service, you may choose to submit your manuscript as a single file to be used in the refereeing process. This can be a PDF file or a Word document, in any format or layout that can be used by referees to evaluate your manuscript. It should contain high enough quality figures for refereeing. If you prefer to do so, you may still provide all or some of the source files at the initial submission. Please note that individual figure files larger than 10 MB must be uploaded separately.

References

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the article number or pagination must be present. Use of DOI is highly encouraged. The reference style used by the journal will be applied to the accepted article by Elsevier at the proof stage. Note that missing data will be highlighted at proof stage for the author to correct.

Formatting requirements

There are no strict formatting requirements but all manuscripts must contain the essential elements needed to convey your manuscript, for example Abstract, Keywords, Introduction, Materials and Methods, Results, Conclusions, Artwork and Tables with Captions.

If your article includes any Videos and/or other Supplementary material, this should be included in your initial submission for peer review purposes.

Divide the article into clearly defined sections.

Please ensure the text of your paper is double-spaced this is an essential peer review requirement.

Figures and tables embedded in text

Please ensure the figures and the tables included in the single file are placed next to the relevant text in the manuscript, rather than at the bottom or the top of the file. The corresponding caption should be placed directly below the figure or table.

Peer review

This journal operates a single anonymized review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final. Editors are not involved in decisions about papers which they have written themselves or have been written by family members or colleagues or which relate to products or services in which the editor has an interest. Any such submission is subject to all of the journal's usual procedures, with peer review handled independently of the relevant editor and their research groups. More information on types of peer review.

REVISED SUBMISSIONS

Use of word processing software

Regardless of the file format of the original submission, at revision you must provide us with an editable file of the entire article. Keep the layout of the text as simple as possible. Most formatting codes will be removed and replaced on processing the article. The electronic text should be prepared in a way very similar to that of conventional manuscripts (see also the Guide to Publishing with Elsevier). See also the section on Electronic artwork.

To avoid unnecessary errors you are strongly advised to use the 'spell-check' and 'grammar-check' functions of your word processor.

Article structure

Subdivision - numbered sections

Divide your article into clearly defined and numbered sections. Subsections should be numbered 1.1 (then 1.1.1, 1.1.2, ...), 1.2, etc. (the abstract is not included in section numbering). Use this numbering also for internal cross-referencing: do not just refer to 'the text'. Any subsection may be given a brief heading. Each heading should appear on its own separate line.

Introduction

State the objectives of the work and provide an adequate background, avoiding a detailed literature survey or a summary of the results.

Material and methods

Provide sufficient details to allow the work to be reproduced by an independent researcher. Methods that are already published should be summarized, and indicated by a reference. If quoting directly from a previously published method, use quotation marks and also cite the source. Any modifications to existing methods should also be described.

Results

Results should be clear and concise.

Discussion

This should explore the significance of the results of the work, not repeat them. A combined Results and Discussion section is often appropriate. Avoid extensive citations and discussion of published literature.

Summary table

The authors shall provide a table with in 2-4 bullets statements on 'what was already known on the topic' and also in 2-4 bullets statements on 'what this study added to our knowledge'. Note that the second part of the table should not list the results of the study as such. It should address what this study has proven and what insights have been gained.

Appendices

If there is more than one appendix, they should be identified as A, B, etc. Formulae and equations in appendices should be given separate numbering: Eq. (A.1), Eq. (A.2), etc.; in a subsequent appendix, Eq. (B.1) and so on. Similarly for tables and figures: Table A.1; Fig. A.1, etc.

Essential title page information

- **Title.** Concise and informative. Titles are often used in information-retrieval systems. Avoid abbreviations and formulae where possible.
- **Author names and affiliations.** Please clearly indicate the given name(s) and family name(s) of each author and check that all names are accurately spelled. You can add your name between parentheses in your own script behind the English transliteration. Present the authors' affiliation addresses (where the actual work was done) below the names. Indicate all affiliations with a lowercase superscript letter immediately after the author's name and in front of the appropriate address. Provide the full postal address of each affiliation, including the country name and, if available, the e-mail address of each author.
- **Corresponding author.** Clearly indicate who will handle correspondence at all stages of refereeing and publication, also post-publication. This responsibility includes answering any future queries about Methodology and Materials. **Ensure that the e-mail address is given and that contact details are kept up to date by the corresponding author.**
- **Present/permanent address.** If an author has moved since the work described in the article was done, or was visiting at the time, a 'Present address' (or 'Permanent address') may be indicated as

a footnote to that author's name. The address at which the author actually did the work must be retained as the main, affiliation address. Superscript Arabic numerals are used for such footnotes.

Highlights

Highlights are optional yet highly encouraged for this journal, as they increase the discoverability of your article via search engines. They consist of a short collection of bullet points that capture the novel results of your research as well as new methods that were used during the study (if any). Highlights should be submitted in a separate editable file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point).

Structured abstract

A structured abstract, by means of appropriate headings, should provide the context or background for the research and should state its purpose, basic procedures (selection of study subjects or laboratory animals, observational and analytical methods), main findings (giving specific effect sizes and their statistical significance, if possible), and principal conclusions. It should emphasize new and important aspects of the study or observations.

Graphical abstract

Although a graphical abstract is optional, its use is encouraged as it draws more attention to the online article. The graphical abstract should summarize the contents of the article in a concise, pictorial form designed to capture the attention of a wide readership. Graphical abstracts should be submitted as a separate file in the online submission system. Image size: Please provide an image with a minimum of 531 × 1328 pixels (h × w) or proportionally more. The image should be readable at a size of 5 × 13 cm using a regular screen resolution of 96 dpi. Preferred file types: TIFF, EPS, PDF or MS Office files. You can view [Example Graphical Abstracts](#) on our information site. Authors can make use of Elsevier's Illustration Services to ensure the best presentation of their images and in accordance with all technical requirements.

Keywords

Immediately after the abstract, provide a maximum of 6 keywords, using American spelling and avoiding general and plural terms and multiple concepts (avoid, for example, 'and', 'of'). Be sparing

with abbreviations: only abbreviations firmly established in the field may be eligible. These keywords will be used for indexing purposes.

Abbreviations

Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at their first mention there, as well as in the footnote. Ensure consistency of abbreviations throughout the article. The abbreviation should be used, except in section headings and subheadings where full text is preferred. Abbreviation should not contain periods or intervening space between letters. Universally known abbreviations (USA for United States of America) need not be defined. Avoid using abbreviations in the abstract of the manuscript.

Acknowledgements

Collate acknowledgements in a separate section at the end of the article before the references and do not, therefore, include them on the title page, as a footnote to the title or otherwise. List here those individuals who provided help during the research (e.g., providing language help, writing assistance or proof reading the article, etc.).

Formatting of funding sources

List funding sources in this standard way to facilitate compliance to funder's requirements: Funding: This work was supported by the National Institutes of Health [grant numbers xxxx, yyyy]; the Bill & Melinda Gates Foundation, Seattle, WA [grant number zzzz]; and the United States Institutes of Peace [grant number aaaa].

It is not necessary to include detailed descriptions on the program or type of grants and awards. When funding is from a block grant or other resources available to a university, college, or other research institution, submit the name of the institute or organization that provided the funding. If no funding has been provided for the research, please include the following sentence: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Units

Follow internationally accepted rules and conventions: use the international system of units (SI). If other units are mentioned, please give their equivalent in SI.

Footnotes

Footnotes should be used sparingly. Number them consecutively throughout the article. Many word processors build footnotes into the text, and this feature may be used. Should this not be the case, indicate the position of footnotes in the text and present the footnotes themselves separately at the end of the article.

Artwork

Electronic artwork

General points

- Make sure you use uniform lettering and sizing of your original artwork.
- Preferred fonts: Arial (or Helvetica), Times New Roman (or Times), Symbol, Courier.
- Number the illustrations according to their sequence in the text.
- Use a logical naming convention for your artwork files.
- Indicate per figure if it is a single, 1.5 or 2-column fitting image.
- For Word submissions only, you may still provide figures and their captions, and tables within a single file at the revision stage.
- Please note that individual figure files larger than 10 MB must be provided in separate source files.

A detailed guide on electronic artwork is available.

Formats

Regardless of the application used, when your electronic artwork is finalized, please 'save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):
EPS (or PDF): Vector drawings. Embed the font or save the text as 'graphics'.
TIFF (or JPG): Color or grayscale photographs (halftones): always use a minimum of 300 dpi.
TIFF (or JPG): Bitmapped line drawings: use a minimum of 1000 dpi.
TIFF (or JPG): Combinations bitmapped line/half-tone (color or grayscale): a minimum of 500 dpi is required.

Please do not:

- Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); the resolution is too low.
- Supply files that are too low in resolution.
- Submit graphics that are disproportionately large for the content.

Color artwork

Please make sure that artwork files are in an acceptable format (TIFF (or JPEG), EPS (or PDF), or MS Office files) and with the correct resolution. If, together with your accepted article, you submit usable color figures then Elsevier will ensure, at no additional charge, that these figures will appear in color online (e.g., ScienceDirect and other sites) regardless of whether or not these illustrations are reproduced in color in the printed version. **For color reproduction in print, you will receive information regarding the costs from Elsevier after receipt of your accepted article.** Please indicate your preference for color: in print or online only. Further information on the preparation of electronic artwork.

Illustration services

Elsevier's Author Services offers Illustration Services to authors preparing to submit a manuscript but concerned about the quality of the images accompanying their article. Elsevier's expert illustrators can produce scientific, technical and medical-style images, as well as a full range of charts, tables and graphs. Image 'polishing' is also available, where our illustrators take your image(s) and improve them to a professional standard. Please visit the website to find out more.

Figure captions

Ensure that each illustration has a caption. A caption should comprise a brief title (**not** on the figure itself) and a description of the illustration. Keep text in the illustrations themselves to a minimum but explain all symbols and abbreviations used.

Tables

Please submit tables as editable text and not as images. Tables can be placed either next to the relevant text in the article, or on separate page(s) at the end. Number tables consecutively in accordance with their appearance in the text and place any table notes below the table body. Be

sparing in the use of tables and ensure that the data presented in them do not duplicate results described elsewhere in the article. Please avoid using vertical rules and shading in table cells.

References

Citation in text

Please ensure that every reference cited in the text is also present in the reference list (and vice versa). Any references cited in the abstract must be given in full. Unpublished results and personal communications are not recommended in the reference list, but may be mentioned in the text. If these references are included in the reference list they should follow the standard reference style of the journal and should include a substitution of the publication date with either 'Unpublished results' or 'Personal communication'. Citation of a reference as 'in press' implies that the item has been accepted for publication.

Reference links

Increased discoverability of research and high quality peer review are ensured by online links to the sources cited. In order to allow us to create links to abstracting and indexing services, such as Scopus, CrossRef and PubMed, please ensure that data provided in the references are correct. Please note that incorrect surnames, journal/book titles, publication year and pagination may prevent link creation. When copying references, please be careful as they may already contain errors. Use of the DOI is highly encouraged.

A DOI is guaranteed never to change, so you can use it as a permanent link to any electronic article. An example of a citation using DOI for an article not yet in an issue is: VanDecar J.C., Russo R.M., James D.E., Ambeh W.B., Franke M. (2003). Aseismic continuation of the Lesser Antilles slab beneath northeastern Venezuela. *Journal of Geophysical Research*, <https://doi.org/10.1029/2001JB000884>. Please note the format of such citations should be in the same style as all other references in the paper.

Web references

As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list.

Data references

This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

References in a special issue

Please ensure that the words 'this issue' are added to any references in the list (and any citations in the text) to other articles in the same Special Issue.

Reference management software

Most Elsevier journals have their reference template available in many of the most popular reference management software products. These include all products that support Citation Style Language styles, such as Mendeley. Using citation plug-ins from these products, authors only need to select the appropriate journal template when preparing their article, after which citations and bibliographies will be automatically formatted in the journal's style. If no template is yet available for this journal, please follow the format of the sample references and citations as shown in this Guide. If you use reference management software, please ensure that you remove all field codes before submitting the electronic manuscript.

Users of Mendeley Desktop can easily install the reference style for this journal by clicking the following link: <http://open.mendeley.com/use-citation-style/international-journal-of-medical-informatics>

When preparing your manuscript, you will then be able to select this style using the Mendeley plugins for Microsoft Word or LibreOffice.

Reference formatting

There are no strict requirements on reference formatting at submission. References can be in any style or format as long as the style is consistent. Where applicable, author(s) name(s), journal title/book title, chapter title/article title, year of publication, volume number/book chapter and the article number or pagination must be present. Use of DOI is highly encouraged. The reference style used by the journal will be applied to the accepted article by Elsevier at the proof stage. Note that missing data

will be highlighted at proof stage for the author to correct. If you do wish to format the references yourself they should be arranged according to the following examples:

Reference style

Text: Indicate references by number(s) in square brackets in line with the text. The actual authors can be referred to, but the reference number(s) must always be given.

Example: '..... as demonstrated [3,6]. Barnaby and Jones [8] obtained a different result'

List: Number the references (numbers in square brackets) in the list in the order in which they appear in the text.

Examples:

Reference to a journal publication:

[1] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, The art of writing a scientific article, J. Sci. Commun. 163 (2010) 51–59. <https://doi.org/10.1016/j.Sc.2010.00372>.

Reference to a journal publication with an article number:

[2] J. van der Geer, J.A.J. Hanraads, R.A. Lupton, 2018. The art of writing a scientific article. Heliyon. 19, e00205. <https://doi.org/10.1016/j.heliyon.2018.e00205>.

Reference to a book:

[3] W. Strunk Jr., E.B. White, The Elements of Style, fourth ed., Longman, New York, 2000.

Reference to a chapter in an edited book:

[4] G.R. Mettam, L.B. Adams, How to prepare an electronic version of your article, in: B.S. Jones, R.Z. Smith (Eds.), Introduction to the Electronic Age, E-Publishing Inc., New York, 2009, pp. 281–304.

Reference to a website:

[5] Cancer Research UK, Cancer statistics reports for the UK. <http://www.cancerresearchuk.org/aboutcancer/statistics/cancerstatsreport/>, 2003 (accessed 13 March 2003).

Reference to a dataset:

[dataset] [6] M. Oguro, S. Imahiro, S. Saito, T. Nakashizuka, Mortality data for Japanese oak wilt disease and surrounding forest compositions, Mendeley Data, v1, 2015. <https://doi.org/10.17632/xwj98nb39r.1>.

Reference to software:

[7] E. Coon, M. Berndt, A. Jan, D. Svyatsky, A. Atchley, E. Kikinzon, D. Harp, G. Manzini, E. Shelef, K. Lipnikov, R. Garimella, C. Xu, D. Moulton, S. Karra, S. Painter, E. Jafarov, S. Molins, Advanced Terrestrial Simulator (ATS) v0.88 (Version 0.88), Zenodo, March 25, 2020. <https://doi.org/10.5281/zenodo.3727209>.

Journal abbreviations source

Journal names should be abbreviated according to the List of Title Word Abbreviations. The referencing style as used by the NLM are preferred (see http://www.nlm.nih.gov/bsd/uniform_requirements.html for examples) Abbreviations for journals are those used in MeSH published by the US National Library of Medicine <http://www.ncbi.nlm.nih.gov/nlmcatalog/journals>.

Video

Elsevier accepts video material and animation sequences to support and enhance your scientific research. Authors who have video or animation files that they wish to submit with their article are strongly encouraged to include links to these within the body of the article. This can be done in the same way as a figure or table by referring to the video or animation content and noting in the body text where it should be placed. All submitted files should be properly labeled so that they directly relate to the video file's content. In order to ensure that your video or animation material is directly usable, please provide the file in one of our recommended file formats with a preferred maximum size of 150 MB per file, 1 GB in total. Video and animation files supplied will be published online in the electronic version of your article in Elsevier Web products, including ScienceDirect. Please supply 'stills' with your files: you can choose any frame from the video or animation or make a separate image. These will be used instead of standard icons and will personalize the link to your video data. For more detailed instructions please visit our video instruction pages. Note: since video and animation cannot be embedded in the print version of the journal, please provide text for both the electronic and the print version for the portions of the article that refer to this content.

Data visualization

Include interactive data visualizations in your publication and let your readers interact and engage more closely with your research. Follow the instructions here to find out about available data visualization options and how to include them with your article.

Supplementary material

Supplementary material such as applications, images and sound clips, can be published with your article to enhance it. Submitted supplementary items are published exactly as they are received (Excel or PowerPoint files will appear as such online). Please submit your material together with the article and supply a concise, descriptive caption for each supplementary file. If you wish to make changes to supplementary material during any stage of the process, please make sure to provide an updated file. Do not annotate any corrections on a previous version. Please switch off the 'Track Changes' option in Microsoft Office files as these will appear in the published version.

Research data

This journal encourages and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles. Research data refers to the results of observations or experimentation that validate research findings. To facilitate reproducibility and data reuse, this journal also encourages you to share your software, code, models, algorithms, protocols, methods and other useful materials related to the project. Below are a number of ways in which you can associate data with your article or make a statement about the availability of your data when submitting your manuscript. If you are sharing data in one of these ways, you are encouraged to cite the data in your manuscript and reference list. Please refer to the "References" section for more information about data citation. For more information on depositing, sharing and using research data and other relevant research materials, visit the research data page.

Data linking

If you have made your research data available in a data repository, you can link your article directly to the dataset. Elsevier collaborates with a number of repositories to link articles on ScienceDirect with relevant repositories, giving readers access to underlying data that gives them a better understanding of the research described.

There are different ways to link your datasets to your article. When available, you can directly link your dataset to your article by providing the relevant information in the submission system. For more information, visit the database linking page.

For supported data repositories a repository banner will automatically appear next to your published article on ScienceDirect.

In addition, you can link to relevant data or entities through identifiers within the text of your manuscript, using the following format: Database: xxxx (e.g., TAIR: AT1G01020; CCDC: 734053; PDB: 1XFN).

Mendeley Data

This journal supports Mendeley Data, enabling you to deposit any research data (including raw and processed data, video, code, software, algorithms, protocols, and methods) associated with your manuscript in a free-to-use, open access repository. During the submission process, after uploading your manuscript, you will have the opportunity to upload your relevant datasets directly to *Mendeley Data*. The datasets will be listed and directly accessible to readers next to your published article online. For more information, visit the Mendeley Data for journals page.

Data statement

To foster transparency, we encourage you to state the availability of your data in your submission. This may be a requirement of your funding body or institution. If your data is unavailable to access or unsuitable to post, you will have the opportunity to indicate why during the submission process, for example by stating that the research data is confidential. The statement will appear with your published article on ScienceDirect. For more information, visit the Data Statement page.

AFTER ACCEPTANCE

Accepted Manuscripts

As a service to the community, this journal makes available online the accepted manuscripts as soon as possible after acceptance. At this stage, the author's accepted manuscript (in both full-text and PDF) is given a Digital Object Identifier (DOI) and is fully citable, and searchable by title, author(s) name and the full-text. The article also carries a disclaimer noting that it is an unedited manuscript which has not yet been copyedited, typeset or proofread. When the fully copyedited version is ready for publication, it simply replaces the author accepted manuscript version.

Online proof correction

To ensure a fast publication process of the article, we kindly ask authors to provide us with their proof corrections within two days. Corresponding authors will receive an e-mail with a link to our online proofing system, allowing annotation and correction of proofs online. The environment is similar to MS Word: in addition to editing text, you can also comment on figures/tables and answer questions

from the Copy Editor. Web-based proofing provides a faster and less error-prone process by allowing you to directly type your corrections, eliminating the potential introduction of errors. If preferred, you can still choose to annotate and upload your edits on the PDF version. All instructions for proofing will be given in the e-mail we send to authors, including alternative methods to the online version and PDF.

We will do everything possible to get your article published quickly and accurately. Please use this proof only for checking the typesetting, editing, completeness and correctness of the text, tables and figures. Significant changes to the article as accepted for publication will only be considered at this stage with permission from the Editor. It is important to ensure that all corrections are sent back to us in one communication. Please check carefully before replying, as inclusion of any subsequent corrections cannot be guaranteed. Proofreading is solely your responsibility.

Offprints

The corresponding author will, at no cost, receive a customized Share Link providing 50 days free access to the final published version of the article on ScienceDirect. The Share Link can be used for sharing the article via any communication channel, including email and social media. For an extra charge, paper offprints can be ordered via the offprint order form which is sent once the article is accepted for publication. Both corresponding and co-authors may order offprints at any time via Elsevier's Author Services. Corresponding authors who have published their article gold open access do not receive a Share Link as their final published version of the article is available open access on ScienceDirect and can be shared through the article DOI link.

AUTHOR INQUIRIES

Visit the Elsevier Support Center to find the answers you need. Here you will find everything from Frequently Asked Questions to ways to get in touch.

You can also check the status of your submitted article or find out when your accepted article will be published.

© Copyright 2018 Elsevier | <https://www.elsevier.com>