

The Diagnosis-Wide Landscape of Hospital-Acquired AKI

Anne-Sophie Jannot,^{*,†‡§} Anita Burgun,^{*,†‡§} Eric Thervet,^{†‡||} and Nicolas Pallet^{†‡||¶}

Abstract

Background and objectives The exploration of electronic hospital records offers a unique opportunity to describe in-depth the prevalence of conditions associated with diagnoses at an unprecedented level of comprehensiveness. We used a diagnosis-wide approach, adapted from phenome-wide association studies (PheWAS), to perform an exhaustive analysis of all diagnoses associated with hospital-acquired AKI (HA-AKI) in a French urban tertiary academic hospital over a period of 10 years.

Design, setting, participants, & measurements We retrospectively extracted all diagnoses from an i2b2 (Informatics for Integrating Biology and the Bedside) clinical data warehouse for patients who stayed in this hospital between 2006 and 2015 and had at least two plasma creatinine measurements performed during the first week of their stay. We then analyzed the association between HA-AKI and each International Classification of Diseases (ICD)–10 diagnostic category to draw a comprehensive picture of diagnoses associated with AKI. Hospital stays for 126,736 unique individuals were extracted.

Results Hemodynamic impairment and surgical procedures are the main factors associated with HA-AKI and five clusters of diagnoses were identified: sepsis, heart diseases, polytrauma, liver disease, and cardiovascular surgery. The ICD-10 code corresponding to AKI (N17) was recorded in 30% of the cases with HA-AKI identified, and in this situation, 20% of the diagnoses associated with HA-AKI corresponded to kidney diseases such as tubulointerstitial nephritis, necrotizing vasculitis, or myeloma cast nephropathy. Codes associated with HA-AKI that demonstrated the greatest increase in prevalence with time were related to influenza, polytrauma, and surgery of neoplasms of the genitourinary system.

Conclusions Our approach, derived from PheWAS, is a valuable way to comprehensively identify and classify all of the diagnoses and clusters of diagnoses associated with HA-AKI. Our analysis delivers insights into how diagnoses associated with HA-AKI evolved over time. On the basis of ICD-10 codes, HA-AKI appears largely underestimated in this academic hospital.

Clin J Am Soc Nephrol 12: 874–884, 2017. doi: <https://doi.org/10.2215/CJN.10981016>

Introduction

AKI is a sudden episode of renal dysfunction, and is a major risk factor for the development of CKD, the progression of CKD, and the incidence of ESRD, regardless of the cause of the injury (1). Hospital-acquired AKI (HA-AKI) is a frequent disorder that significantly affects morbidity, length of stay, and residual renal function, and, as such, is a major health care concern. The epidemiology and outcomes of HA-AKI have been described in retrospective studies of various sizes, which, in general, yielded similar results in terms of prevalence and risk factors (2–4). Informatics methods and multiple hypothesis testing techniques help turn data provided by real-life health care data into information, and information into knowledge (5). In this study, we applied this paradigm to address the issue of HA-AKI. Although most of the previous studies explored known risk factors of HA-AKI occurrence, the selected list of risk factors might not be exhaustive. To our knowledge, a

comprehensive exploration of all possible diagnoses associated with AKI has never been performed; however, this objective may be achieved by mining electronic health records (6) and employing strategies, such as a phenome-wide association study (PheWAS) and its derivatives (7,8), to analyze the associations between HA-AKI and different diagnoses. PheWAS, initially developed to unveil unknown associations between a genetic trait and phenotypic parameters (a genotype-to-phenotype approach) using methods that are the reverse of those employed by genome-wide association studies (which are on the basis of a phenotype-to-genotype approach), has emerged as a valuable tool for uncovering unexpected associations between factors, such as diagnoses, and a qualitative trait (the outcome variable), as opposed to a genotypic trait, using high-dimension matrices (9–11). This holistic approach may be made possible through the implementation of a clinical data warehouse (CDW) that integrates electronic medical records (EMR) data

Departments of
*Medical Informatics, Biostatistics and Public Health, Nephrology, and Clinical Chemistry, Hôpital Européen Georges Pompidou, Paris, France; †Assistance Publique Hôpitaux de Paris, Paris, France; ‡Paris Descartes University, Paris, France; and §National Institute for Health and Research (INSERM) U1138, Centre de Recherche des Cordeliers, Paris, France

Correspondence: Dr. Nicolas Pallet, Clinical Chemistry Department, Hôpital Européen Georges Pompidou, 20, Rue Leblanc, 75015 Paris, France. Email: npallet@yahoo.fr

from all individuals admitted to a hospital and includes International Classification of Diseases (ICD) diagnostic codes from the billing system. These exhaustive methods can generate a hierarchic classification of the diagnoses found to be associated with an outcome variable, and several visualization tools have been developed to make these hierarchies easily understandable (12). In the study presented here, we adopted an approach derived from PheWAS to perform a comprehensive analysis of the diagnoses associated with HA-AKI in a French urban tertiary academic hospital (Georges Pompidou European Hospital, GPEH) over a 10-year period. We extracted data from the **GPEH i2b2** (Informatics for Integrating Biology and the Bedside) **CDW**, and, on the basis of a large inpatient AKI population (126,736 unique patients), we identified all diagnoses significantly associated with HA-AKI. In addition, we identified subpopulations at high risk of AKI and analyzed how physicians identify (or fail to identify) AKI.

Materials and Methods

Study and CDW

We performed an *in silico* comparative **cross-sectional and retrospective** period prevalence survey using data from an academic hospital (GPEH) in **Paris, France**. We extracted data from the GPEH CDW, an i2b2 CDW containing the **EMRs** of the >750,000 individual patients who have attended the hospital since its opening in 2000 (13,14). This CDW contains EMRs made of routine care data and is divided into several categories including **diagnostic codes (ICD-10)**.

Study Population

All hospital stays between 2006 and 2015, for which data were available regarding plasma creatinine measurements

during the first 24 hours of the admission and subsequently during the following 7 days, were extracted for inclusion in this study. Therefore, the baseline creatinine level was the value measured at admission. For each patient, only the first hospital stay was included in this study and, consequently, only one hospital stay was analyzed for each unique patient. During this period, 225,471 unique patients were hospitalized, and 98,735 of them who had no or one plasma creatinine measurement were excluded because we were unable to evaluate changes in their plasma creatinine levels. Patients under chronic hemodialysis were excluded on the basis of the N18.5 code (ESRD). HA-AKI was diagnosed on the basis of two plasma creatinine measurements (in the 24 hours and within the 7 days after admission), and was defined as an increase in plasma creatinine to ≥ 1.5 times baseline. Furthermore, HA-AKI has been staged in severity according to the Acute Kidney Injury Network (AKIN) criteria (15,16). Urine output was not used a criterion for AKI identification. Plasma creatinine levels were measured using a colorimetric assay (modified kinetic Jaffe method) on a Beckman Coulter DXC analyzer.

Statistical Analyses

Our analysis is not *stricto sensu* a PheWAS, as was proposed by Denny *et al.* (17). PheWAS groups are custom case groups that comprise ICD-9 codes developed to perform phenome-wide association analyses with a genetic marker. Because our outcome was not genetic, the diagnosis classification system used in France is the ICD-10, and because there is no validated mapping table between ICD-9 codes and ICD-10 codes used in France, we used the ICD-10 categories as phenotypic groups, as was validated by the World Health Organization.

Groups of patients were divided according to the quantitative trait studied (presence or absence of HA-AKI),

Table 1. Characteristics of patients

Characteristic	No HA-AKI	HA-AKI	AKIN1	AKIN2	AKIN3
Individuals, <i>n</i> (%)	118,257 (92)	8479 (8)	5961 (70)	806 (10)	1712 (20)
Age, yr	61 (47–74)	71 (58–83)	73 (59–84)	69 (57–81)	66 (54–78)
Sex ratio, M/F	1.18	1.66	1.57	1.44	2.23
eGFR at admission, <i>n</i> (%)					
<15 ml/min per 1.73 m ²	1558 (1.5)	1118 (13)	192 (3)	0	926 (54)
15–29 ml/min per 1.73 m ²	4311 (3.5)	1180 (14)	830 (14)	7 (0.8)	343 (34)
30–59 ml/min per 1.73 m ²	20,842 (17.5)	2668 (31)	2223 (37)	260 (32)	185 (11)
60–89 ml/min per 1.73 m ²	45,246 (38)	2134 (25)	1689 (28)	315 (40)	130 (7.5)
>90 ml/min per 1.73 m ²	45,873 (39)	1379 (16)	1027 (17)	224 (28)	128 (7.5)
Intensive care unit, <i>n</i> (%)	12,072 (10)	3976 (46)	2641 (44)	548 (67)	787 (45)
Length of stay, d	2 (0–7)	10 (5–18)	10 (5–15)	12 (6–23)	8 (3–18)
In-hospital mortality, <i>n</i> (%)					
In ICU	2800 (2.4)	2062 (25)	1212 (20)	334 (41)	516 (30)
Not in ICU	1528 (13)	1573 (40)	875 (33)	279 (51)	419 (53)
	1272 (1.2)	489 (11)	337 (10)	55 (21)	97 (10)

Categoric variables are presented as *n* (%), and continuous variables are presented as the mean and interquartile range. AKI was diagnosed on the basis of two plasma creatinine measurements (in the 24 h and 7 d after admission), and the severity was classified according to the AKIN stages. AKIN1: 1.5–2-fold plasma creatinine increase from baseline; AKIN2: >2–3-fold plasma creatinine increase from baseline; AKIN3: >3-fold plasma creatinine increase from baseline, or plasma creatinine >4 mg/dl with an acute increase of at least 0.5 mg/dl, or need for RRT. HA-AKI, hospital-acquired AKI; AKIN, Acute Kidney Injury Network; M/F, Male/Female; ICU, Intensive Care Unit.

Downloaded from <http://journals.lww.com/cjn> by BMDM5epHkav1zEoun1tQIN4a+KJLhEZgpbSHo4XM0i0nCjwCX1A
WNYQp/llQrHD3iD00dOr/y77VtSF14C3VC1y0abgqZXdtnfKZEBtws= on 10/08/2024

Table 2. Evolution of HA-AKI over time										
Variable	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
No HA-AKI, <i>n</i>	11,270	10,448	10,575	10,684	11,960	12,270	13,485	14,185	11,545	11,795
HA-AKI, <i>n</i>	960	725	739	835	897	836	962	917	766	829
Total, <i>n</i>	12,230	11,173	11,314	11,519	12,857	13,106	14,447	15,102	12,311	12,624
% AKI	7.85	6.49	6.53	7.25	6.98	6.38	6.66	6.07	6.22	6.57

For each year, the number of patients with and without HA-AKI has been accounted for, and the % of AKI calculated. HA-AKI, hospital-acquired AKI.

defined as the presence or absence of the criteria required for having HA-AKI. We measured the strength of the association between each ICD-10 category and HA-AKI by calculating odds ratios (ORs) and *P* values using Fisher exact tests. A total of 1730 ICD-10 diagnoses were observed in the whole cohort, and 217 were present in only one or two patients; for these infrequent diagnoses, the corresponding association tests were not performed. Multiple comparisons were performed on the remaining 1513 diagnoses, leading to a threshold of significance of $(0.05/1513)=3.3\times 10^{-5}$ after Bonferroni correction. We performed these analyses on the whole period, and for 2006–2010 and 2011–2015, to assess whether the diagnoses associated with HA-AKI changed over time. In the sample of patients with HA-AKI, we also analyzed the association between the ICD-10 diagnostic code corresponding to AKI (N17 [“Acute kidney failure”]) in calculating ORs and *P* values using Fisher exact tests: in the group of individuals with HA-AKI, $[(\text{ICD-10+}/\text{ICD-10-})]$ in individuals with N17 code) divided by $[(\text{ICD-10+}/\text{ICD-10-})]$ in individuals without N17 code). Multiple comparisons were performed on the 817 associated diagnoses, leading to a *P* value considered statistically significant of $(0.05/817)=6\times 10^{-5}$ after Bonferroni correction.

Dissimilarities between diagnostic categories in the sample of patients with HA-AKI were estimated using the Jaccard distance (18). A hierarchic cluster analysis was performed on these estimated dissimilarities to identify clusters of associated diagnoses using the Ward method, and a resulting heatmap was generated. Only diagnoses for which the Jaccard distance was lower than four on the heatmap were represented.

Data Management

An open database connection linking an Oracle database (11 g Enterprise Edition Release 11.2.0.1.0) of i2b2 CDW (version 1.6) to R software (version 3.1.0) was set up. The dataset containing data from the cohort was imported into R and analyzed further using the pheatmap and epitools packages. This work adheres to the Reporting of studies Conducted using Observational Routinely-collected health Data guidelines (19).

Ethics Statement

This study was approved by the Institutional Review Board of GPEH.

Results

HA-AKI Prevalence and Risks Factors

In the population of 126,736 unique individuals admitted to the hospital between 2006 and 2015 (see Materials and Methods), the prevalence of AKI was 8% (Table 1). Men were more prone to having HA-AKI, as were older individuals. As expected, HA-AKI prevalence increased with reduced renal function at admission: 13% and 14% of patients with $\text{eGFR}<15\text{ ml/min per }1.73\text{ m}^2$ and $\text{eGFR }15\text{--}29\text{ ml/min per }1.73\text{ m}^2$, respectively, were affected by HA-AKI, whereas HA-AKI was identified in $<5\%$ of patients with $\text{eGFR}>60\text{ ml/min per }1.73\text{ m}^2$ at admission, confirming that CKD is a risk factor for AKI, especially HA-AKI (Supplemental Table 1). HA-AKI appeared to be a critical

medical situation, because 46% of patients with HA-AKI were hospitalized in intensive care units versus 10% of patients without HA-AKI; additionally, 25% of patients with HA-AKI died during their hospital stay, corresponding to a mortality OR of 13.2 (95% confidence interval, 12.4 to 14.1). Notably, there was no obvious change in the prevalence of HA-AKI over the follow-up period (Table 2).

Diagnosis-Wide Association with HA-AKI

To define the medical landscape within which HA-AKI develops, we analyzed the association between each ICD-10 diagnostic code and HA-AKI, in estimating ORs (see Materials and Methods) and their corresponding *P* values (the complete list of items associated with HA-AKI with a *P* value $<3.3\times10^{-5}$ [*n*=251] is in Supplemental Table 2). Figure 1 illustrates the distribution of the 208 ICD-10 diagnoses positively (*OR*>1) and significantly (*P*< 3.3×10^{-5}) associated with HA-AKI according to the diagnosis family. Some categories of ICD-10 diagnoses were overrepresented compared with others among the HA-AKI-associated diagnoses. For example, there were many diagnoses in the ICD-10 categories I (“Diseases of the circulatory system”), J (“Diseases of the respiratory system”), and S (“Injury, poisoning, and certain other consequences of external causes”), but very few diagnoses were identified in the ICD-10 categories L (“Diseases of the skin

and subcutaneous tissues”) or Q (“Congenital malformations, deformations, and chromosomal abnormalities”), as would be expected. The diagnoses with large effect size could be roughly divided into two clinically relevant categories (Supplemental Figure 1, Table 3): surgical procedures, severe disorders of the circulatory system, and hemodynamic impairment were pre-eminent factors associated with or promoting AKI and may be causal factors for AKI or markers of the severity of the medical situation leading to AKI, whereas diagnoses describing the AKI condition, such as N17 (“Acute kidney failure”), Z49 (“Care involving dialysis”), and R34 (“Anuria/oliguria”), may be interpreted as “internal controls,” as they were obligatorily associated with HA-AKI. On the other hand, most of the codes negatively associated with AKI were related to routine examinations (Supplemental Table 3).

Despite a lack of information regarding the temporal sequence of ICD-10 code assignment, the association between ICD-10 diagnoses and HA-AKI may be interpreted as follows: (1) “likely causative” (for example, R57 [“Shock”], A41 [“Sepsis”], and T86 [“Failure and rejection of transplanted organs”]), in which the diagnosis is likely the cause of AKI; (2) “reverse causality,” for example, R34 (“Anuria-oliguria”) and Z49 (“Care involving dialysis”), in which AKI is responsible for the assigned ICD-10 diagnosis code; (3) “possible confounder,” such as R40 (“Somnolence,

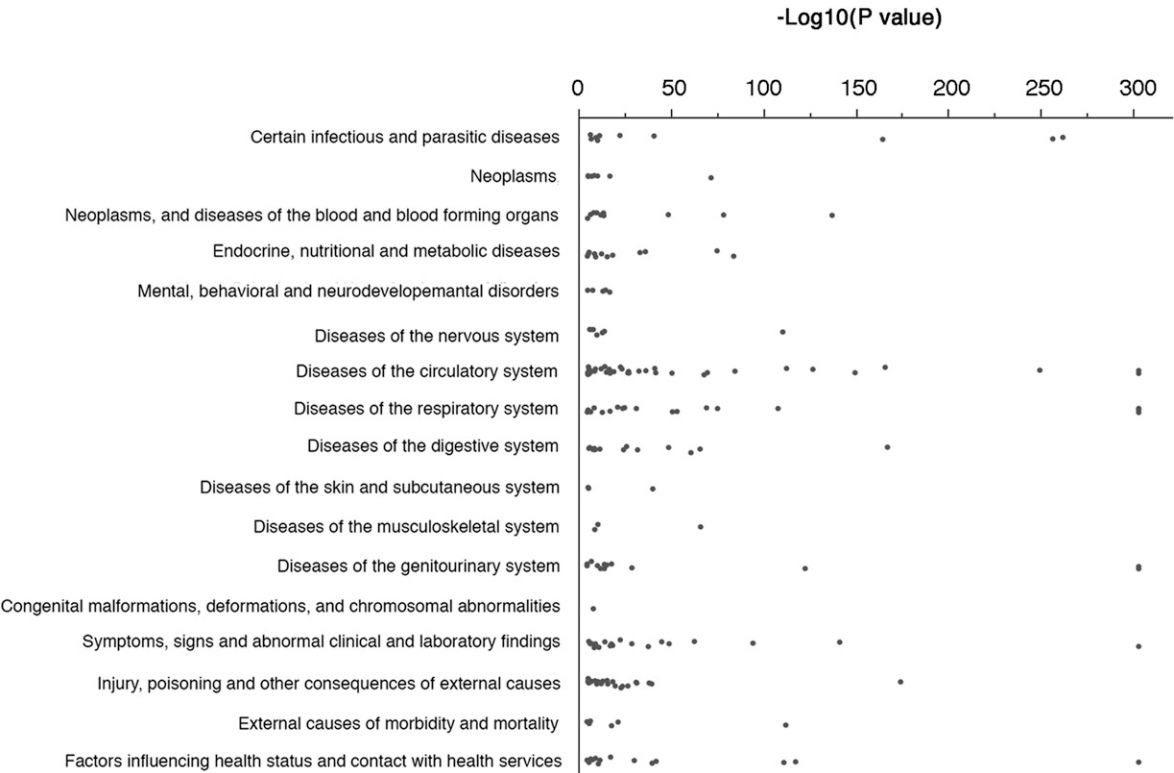


Figure 1. | ICD-10 codes positively and significantly associated with HA-AKI have a specific distribution pattern. A total of 1730 ICD-10 diagnoses were reported within the whole cohort; of these diagnoses, 217 were present in only one or two patients and were, therefore, excluded from the analysis. Multiple comparisons were performed on 1513 patients (1730–217), leading to a *P* value considered statistically significant of $(0.05/1513)=3.3\times10^{-5}$ after Bonferroni correction. The scatter plot represents the 208 ICD-10 codes significantly (*P*< 3.3×10^{-5}) and positively (*OR*>1) associated with HA-AKI, and classified according to diagnosis family. HA-AKI, hospital-acquired AKI; ICD-10, International Classification of Diseases–10.

Downloaded from http://ajcn.ajcn.org/ by guest on June 12, 2017

Table 3. ICD-10 diagnoses positively associated with HA-AKI

ICD-10	Diagnosis	Occurrences	P Value	OR (95% CI)
J80	Adult respiratory distress syndrome	1281	5×10^{-330}	16.8 (15 to 18.8)
N17	Acute kidney failure	5568	5×10^{-330}	15.9 (15 to 16.8)
R57	Shock	4777	5×10^{-330}	14.4 (13.5 to 15.3)
I46	Cardiac arrest	1633	5×10^{-330}	12.8 (11.5 to 14.1)
Z99	Dependence on enabling machines and devices	1800	5×10^{-330}	9.1 (8.2 to 10)
R40	Somnolence, stupor, and coma	1955	5×10^{-330}	9 (8.2 to 9.9)
J96	Respiratory failure	7672	5×10^{-330}	6 (5.7 to 6.4)
I50	Heart failure	8858	5×10^{-330}	4.8 (4.6 to 5.1)
N18	CKD	8085	5×10^{-330}	5.3 (4.9 to 5.7)
J15	Bacterial pneumonia	4081	4.4×10^{-323}	4.1 (3.9 to 4.4)
B96	Other specified bacterial agents	2871	4.8×10^{-262}	5.6 (5.2 to 6.1)
A41	Other sepsis	1948	1.4×10^{-256}	7.1 (6.5 to 7.9)
T81	Complications of procedures, not elsewhere classified	2681	2.7×10^{-174}	4.5 (4.1 to 5)
K72	Hepatic failure	685	3.4×10^{-167}	11.4 (9.8 to 13.3)
B95	<i>Streptococcus</i> and <i>Staphylococcus</i> as the cause of diseases	1552	1.2×10^{-164}	6 (5.4 to 6.8)
I97	Postprocedural disorders of the circulatory system	937	1.1×10^{-149}	8 (7 to 9.2)
R58	Hemorrhage, not elsewhere classified	1031	2.2×10^{-141}	7.1 (6.3 to 8.1)
D62	Acute posthemorrhagic anemia	1797	2.9×10^{-137}	4.9 (4.4 to 5.5)
I21	Acute myocardial infarction	2028	7.3×10^{-127}	4.4 (3.9 to 4.9)
N99	Postprocedural disorders of the genitourinary system	295	1.1×10^{-122}	20.2 (16 to 25.4)
Y95	Nosocomial condition	512	3.2×10^{-112}	10.3 (8.6 to 12.2)
Z49	Care involving dialysis	386	3.4×10^{-111}	13.2 (10.8 to 16.2)
G93	Other disorders of the brain	520	1.3×10^{-110}	10 (8.4 to 11.9)
J95	Postprocedural respiratory disorders	791	4.9×10^{-108}	7.1 (6.1 to 8.2)
R34	Anuria/oliguria	360	1.9×10^{-94}	12 (9.8 to 14.8)
D69	Purpura and other hemorrhagic conditions	1013	1.3×10^{-78}	4.9 (4.2 to 5.6)
C64	Malignant neoplasm of kidney, except renal pelvis	824	8.2×10^{-72}	5.2 (4.4 to 6.1)
J69	Pneumonitis due to solids and liquids	918	2.4×10^{-69}	4.8 (4.1 to 5.5)
I31	Other diseases of pericardium	939	5×10^{-68}	4.7 (4 to 5.4)
K65	Peritonitis	738	7.5×10^{-66}	5.3 (4.5 to 6.2)
R41	Other symptoms and signs involving cognitive functions and awareness	1096	8.6×10^{-63}	4.1 (3.5 to 4.7)
K55	Vascular disorders of intestine	717	6.4×10^{-61}	5.1 (4.3 to 6)
J81	Complications of procedures	205	2.1×10^{-53}	11.7 (8.9 to 15.4)
D65	Disseminated intravascular coagulation	217	1.1×10^{-48}	10.2 (7.8 to 13.3)
R39	Other symptoms and signs involving the urinary system	205	4.1×10^{-45}	10 (7.5 to 13.2)
A40	Streptococcal sepsis	277	5.8×10^{-41}	7.2 (5.6 to 9.3)
R65	Systemic inflammatory response syndrome	441	7.2×10^{-38}	5 (4.1 to 6.2)
I33	Endocarditis	406	1.4×10^{-36}	5.2 (4.2 to 6.5)
J13	Pneumonia due to <i>Streptococcus pneumoniae</i>	416	2.1×10^{-31}	4.6 (3.7 to 5.8)
R45	Symptoms and signs involving emotional state	364	6.1×10^{-29}	4.8 (3.8 to 6)
I08	Multiple valves disease	424	8.2×10^{-27}	4.2 (3.3 to 5.2)
K26	Duodenal ulcer	373	4.1×10^{-26}	4.4 (3.5 to 5.6)
J85	Abscess of lung and mediastinum	245	4.8×10^{-25}	5.5 (4.2 to 7.3)
T79	Certain early complications of trauma, not elsewhere classified	290	7.3×10^{-24}	4.8 (3.7 to 6.3)
S35	Injury of blood vessels at abdomen, lower back, and pelvis level	138	4.2×10^{-23}	7.7 (5.4 to 11)

Table 3. (Continued)

ICD-10	Diagnosis	Occurrences	P Value	OR (95% CI)
B37	Candidiasis	297	1.6×10^{-22}	4.6 (3.5 to 6)
J14	Pneumonia due to <i>Haemophilus influenzae</i>	226	2.7×10^{-21}	5.2 (3.9 to 7)
T68	Hypothermia	87	5.8×10^{-20}	9.9 (6.5 to 15.2)
R68	Other general symptoms and signs	79	1.4×10^{-18}	10.1 (6.4 to 15.7)
D70	Agranulocytosis	162	8.6×10^{-14}	4.8 (3.3 to 6.8)
T86	Failure and rejection of transplanted organs and tissues	120	3.2×10^{-13}	5.5 (3.7 to 8.3)
G41	Status epilepticus	192	4.7×10^{-13}	4.2 (3 to 5.8)
I23	Certain current complications following acute myocardial infarction	77	2.2×10^{-12}	7.2 (4.5 to 11.5)
N19	Unspecified kidney failure	138	5×10^{-12}	4.8 (3.2 to 7)
A48	Other bacterial diseases, not elsewhere classified	102	1.5×10^{-11}	5.6 (3.6 to 8.6)
R99	Other ill-defined and unspecified causes of mortality	99	4×10^{-11}	5.5 (3.6 to 8.6)
S26	Injury of heart	90	1.1×10^{-10}	5.7 (3.6 to 9)
M96	Postprocedural musculoskeletal disorders, not elsewhere classified	84	1.3×10^{-10}	5.9 (3.7 to 9.5)
T46	Poisoning by agents primarily affecting the cardiovascular system	42	5.3×10^{-10}	9.5 (5.1 to 17.7)
R98	Unattended death	76	2.9×10^{-9}	5.7 (3.5 to 9.4)
M72	Fibroblastic disorders	79	6.5×10^{-9}	5.4 (3.3 to 8.9)
J10	Influenza with pneumonia, seasonal influenza virus identified	89	1.4×10^{-8}	4.9 (3 to 7.9)
G97	Postprocedural disorders of nervous system, not elsewhere classified	80	4.4×10^{-8}	5 (3 to 8.2)
F99	Mental disorder, not otherwise specified	56	8.2×10^{-8}	6.1 (3.5 to 10.8)
C65	Malignant neoplasm of renal pelvis	31	3.3×10^{-7}	8.9 (4.3 to 18.3)
S33	Dislocation, sprain, and strain of joints and ligaments of lumbar spine	37	4.1×10^{-7}	7.6 (3.9 to 14.9)
S25	Injury of blood vessels of thorax	88	1.1×10^{-6}	4.1 (2.5 to 6.8)
X70	Intentional self-harm by hanging, strangulation, and suffocation	19	1.3×10^{-6}	12.6 (5.1 to 31)
I22	Subsequent myocardial infarction	60	1.3×10^{-6}	5.1 (2.9 to 9.1)
Y69	Unspecified misadventure during surgical and medical care	35	1.5×10^{-6}	7.3 (3.6 to 14.7)
Y57	Other and unspecified drugs and medicaments	46	6.6×10^{-6}	5.5 (2.9 to 10.5)
C66	Malignant neoplasm of ureter	58	2.1×10^{-5}	4.5 (2.5 to 8.2)
S48	Traumatic amputation of shoulder and upper arm	11	3.3×10^{-5}	16.7 (5.1 to 54.6)

Diagnoses significantly associated after correction for multiple comparisons (P value $< 3.3 \times 10^{-5}$), and with an OR > 4 are shown. Occurrence indicates the number of times an ICD-10 code was identified in the entire cohort ($n=126,736$ individuals). ORs were calculated as follows: in the whole population, $([ICD-10+]/ICD-10-]$ in patients with HA-AKI) divided by $([ICD-10+]/ICD-10-]$ in patients without HA-AKI). ICD-10, International Classification of Diseases–10; HA-AKI, hospital-acquired AKI; OR, odds ratio; 95% CI, 95% confidence interval..

stupor, and coma”) and J80 (“Adult respiratory distress syndrome”), which likely reflects the severity of a medical condition rather than the cause or the consequence of the AKI; (4) “risk factor,” such as N18 (“CKD”) which is a known factor that may have facilitated the occurrence but was not a cause of AKI; and (5) “truism,” such as N17 (“Acute kidney failure”), in which AKI is obligatorily associated with AKI.

Because we excluded from our association analysis 98,735 patients who had no or only one plasma creatinine measurement during the first week after admission (and for who AKI could not be diagnosed), we evaluated the effect of this exclusion on the prevalence of ICD-10 diagnoses. We

made the assumption that these excluded patients had no HA-AKI, and the results of this analysis indicated that our inclusion criteria are not responsible for a bias in the prevalence of the associations between HA-AKI and ICD-10 diagnoses (Supplemental Figure 2).

Changes of Association Patterns over Time

We next assessed if the diagnoses associated with HA-AKI changed over time. To do this, we split the observation period into 2006–2010 and 2011–2015, and we plotted the corresponding log-transformed P values of each of the 251 ICD-10 diagnoses identified in the whole period (Figure

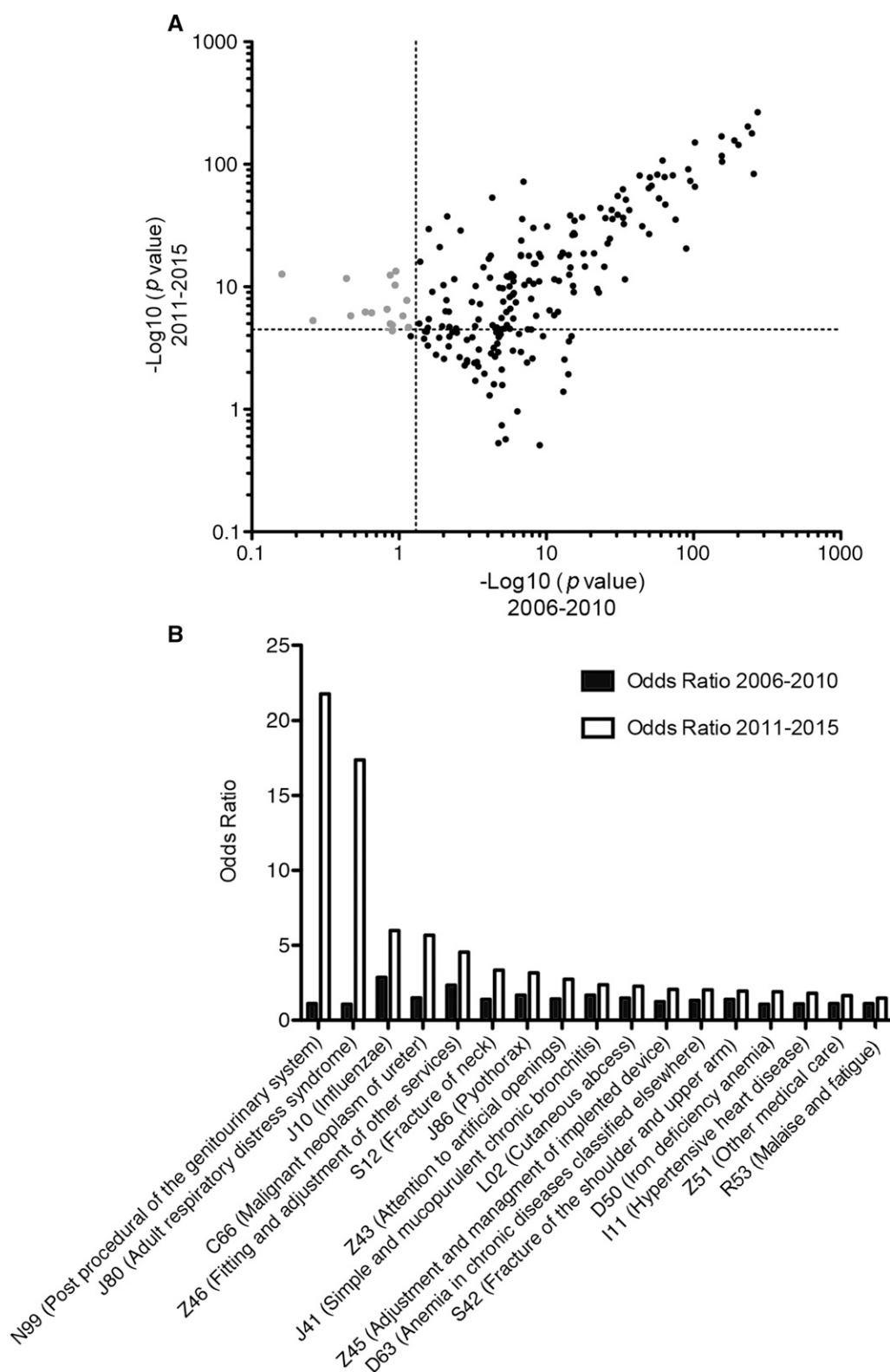


Figure 2. | Some diagnoses have a specific evolution of association pattern over time. (A) Scatter plots of the $-\log_{10}(P\text{value})$ of the 251 ICD-10 diagnosis codes significantly associated with HA-AKI according to the observation period (2006–2010 and 2011–2016). Gray dots denote diagnoses with a $P\text{value} > 0.05$ in 2006–2010 and $< 3.3 \times 10^{-5}$ in 2011–2015. (B) Histogram representing the ORs of the 17 ICD-10 diagnoses significantly associated with HA-AKI in 2011–2015 but not in 2006–2010. HA-AKI, hospital-acquired AKI; ICD-10, International Classification of Diseases–10; OR, odds ratio.

2A). To appreciate the variations in association patterns, we looked at the diagnoses not significantly associated with HA-AKI between 2006 and 2010 ($P>0.05$), and which became significantly associated between 2011 and 2015 ($P<3.3\times10^{-5}$). Seventeen diagnoses were selected in this way, with some of them presumably reflecting changes in the severity of common diseases, such as J10 (“Influenzae”), for which ORs nearly doubled; or markers of polytrauma (S12 and S42) (Figure 2B). Codes associated with HA-AKI that demonstrated the greatest increase in prevalence with time were related to surgery of neoplasms of the genitourinary system, possibly reflecting changes in indication and/or severity of the cases. Finally, acute respiratory distress syndrome was the diagnosis for which

the change in OR was the most striking (OR=1 during 2006–2010 versus 17.3 during 2011–2015), and in this case, changes in policies for ICD-10 encoding can be involved. Notably, these observations were not related to changes in the prevalence of diagnoses, which remained stable over time (not shown), but rather to changes in the prevalence of HA-AKI.

Clusters of Diagnoses Associated with AKI

To highlight discrete subpopulations of high-risk patients with distinctive diagnosis patterns, we generated clusters of ICD-10 diagnoses in the population of patients with HA-AKI. Additionally, to enhance cluster interpretation and visualization, we plotted HA-AKI-associated

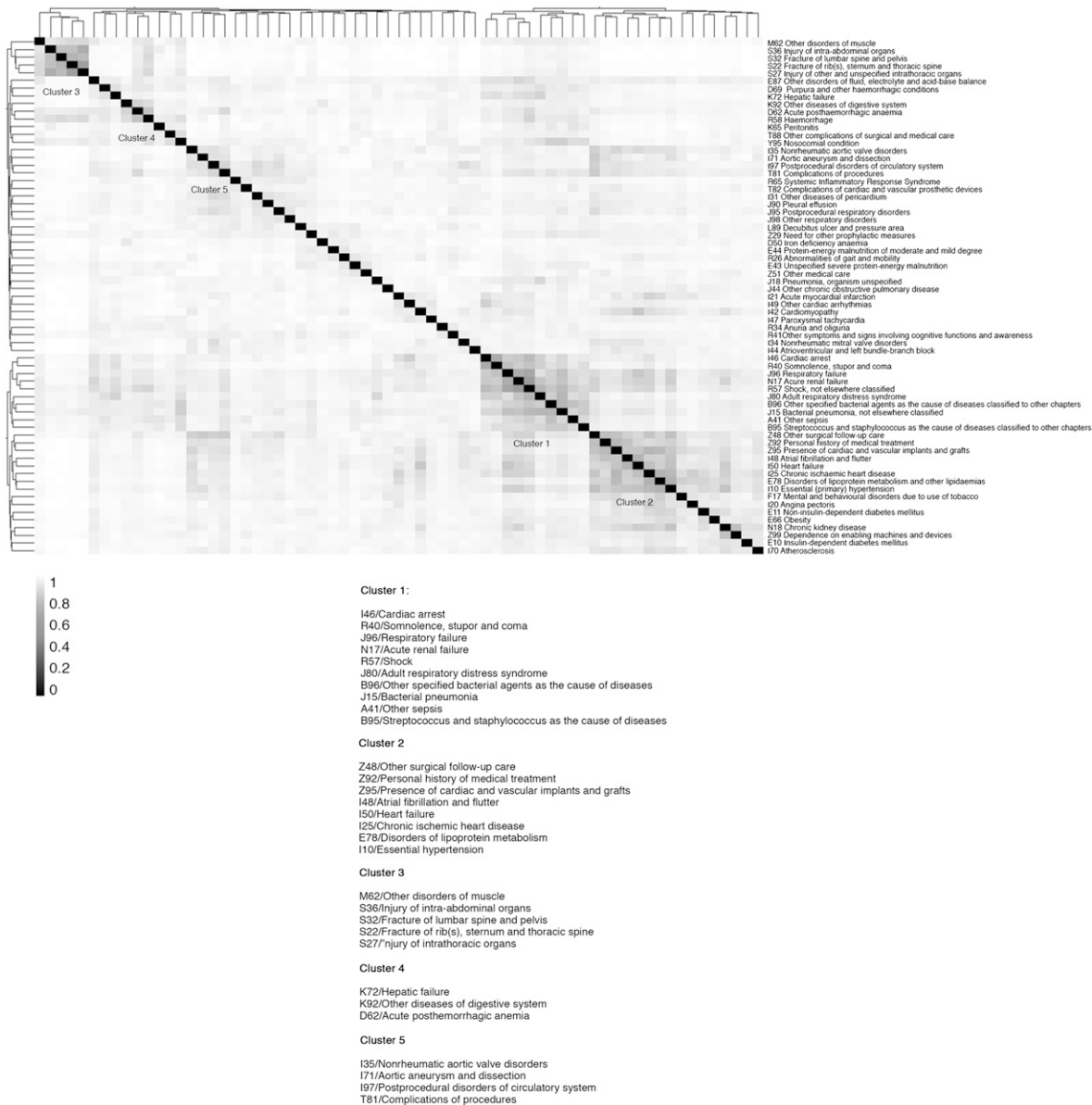


Figure 3. | Clusters of diagnoses associated with AKI can be identified. Heatmap representation of the hierarchic cluster analysis of the Jaccard distances of diagnoses in the HA-AKI population. Diagnoses for which the sum of the dissimilarity coefficients was lower than four are represented on the heatmap. Five major clusters are highlighted. HA-AKI, hospital-acquired AKI.

Table 4. Contingency table of N17 distribution according to HA-AKI status

Variable	N17 Positive	N17 Negative	Total
HA-AKI positive, <i>n</i> (%)	2513 (30)	5966 (70)	8479 (100)
HA-AKI negative, <i>n</i> (%)	3055 (2.5)	115,202 (97.5)	118,257 (100)
Total, <i>n</i>	5568	121,168	126,736
HA-AKI, hospital-acquired AKI.			

diagnoses exhibiting strong similarity with each other (see Materials and Methods), resulting in the representation of 67 diagnoses (Figure 3). Overall, five clusters of diagnoses were identified. Cluster 1, corresponding to 66% (*n*=5493) of the patients with HA-AKI, was defined by a severe septic situation complicated with hemodynamic instability and organ failure. Cluster 2, corresponding to 44% (*n*=3815) of the patients with HA-AKI, was likely representative of individuals with heart diseases that increased their risk of developing HA-AKI. Cluster 3, corresponding to 19% (*n*=1600) of the patients with HA-AKI, corresponded to patients with severe polytrauma complicated with hemodynamic instability and rhabdomyolysis and, therefore, AKI. A fourth cluster, corresponding to 10% (*n*=814) of the patients with HA-AKI, likely highlighted patients with severe liver disease that may be associated with portal hypertension and/or hepatorenal syndrome. Finally, a fifth cluster, corresponding to 6% (*n*=522) of patients, corresponded to individuals who developed HA-AKI after cardiovascular surgery.

Distribution of the ICD-10 N17 Code (“Acute Kidney Failure”) among Patients with HA-AKI

The ICD-10 code corresponding to AKI, N17 (“Acute kidney failure”), was recorded in the EMRs of 30% of the patients with HA-AKI identified using the AKIN classification (Table 4); conversely, HA-AKI “negative” patients were assigned the N17 code in nearly 2.5% of the cases. It is possible that these patients were anuric at admission (and therefore coded N17) but without plasma creatinine elevation, which may have contributed to this observation. In fact, ICD-10 coding seemed to depend on the AKIN class: the majority of patients with HA-AKI but without the N17 ICD-10 code had AKIN stage 1, whereas patients with more severe HA-AKI (AKIN stages 2 and 3) were more often coded as N17 (Table 5), indicating that HA-AKI of low severity (AKIN1) was likely overlooked. Among patients with HA-AKI, we identified the ICD-10 diagnostic codes

associated with the corresponding N17 code to highlight medical conditions associated with N17 coding (Figure 4). Fifty-one diagnoses were significantly associated with N17 coding with a $P < 6 \times 10^{-5}$ after Bonferroni correction (see Materials and Methods); these diagnoses included numerous renal-related conditions, including (1) disorders associated with hypovolemia, (2) tubulointerstitial nephropathies, (3) necrotizing vasculopathies, and (4) multiple myeloma and malignant plasma cell neoplasms. These results indicate that patients with HA-AKI coded with N17 (“Acute kidney failure”) suffered from specific renal diseases and might have been managed by nephrologists, increasing the possibility of a code reporting-bias mediated by the medical specialty of the physician.

Discussion

We provide a proof of concept of the applicability of large-scale diagnosis-wide analyses without prior hypotheses in addressing important issues related to renal diseases, and we demonstrate how these diseases may be associated with complex medical situations. Using this method, we obtained a description of all of the medical situations associated with HA-AKI, as defined using validated criteria, at a high level of comprehensiveness. Conducted without an *a priori* hypothesis, this systematic search for associations yielded a definitive picture of the phenotypes of HA-AKI. Our findings highlight the frequency and the severity of the medical situations associated with HA-AKI. Hemodynamic impairment and surgical procedures are the main parameters associated with HA-AKI and five clusters of diagnoses were identified: sepsis, heart diseases, polytrauma, liver disease, and cardiovascular surgery. AKI was recorded in the EMRs as the ICD-10 N17 diagnostic code in 30% of the cases with HA-AKI. Among patients with HA-AKI, ICD-10 N17 (“Acute kidney failure”) coding was highly associated with renal diseases, including myeloma, vasculitis, and acute interstitial nephropathies. Codes associated with HA-AKI that

Table 5. Contingency table of N17 distribution according to the severity of HA-AKI

Variable	N17 Positive	N17 Negative	Total
HA-AKI positive, <i>n</i>	2513	5966	8479
AKIN1, <i>n</i> (%)	1414 (56)	4547 (76)	
AKIN2, <i>n</i> (%)	335 (13)	471 (8)	
AKIN3, <i>n</i> (%)	764 (30)	948 (16)	
HA-AKI, hospital-acquired AKI; AKIN, Acute Kidney Injury Network.			

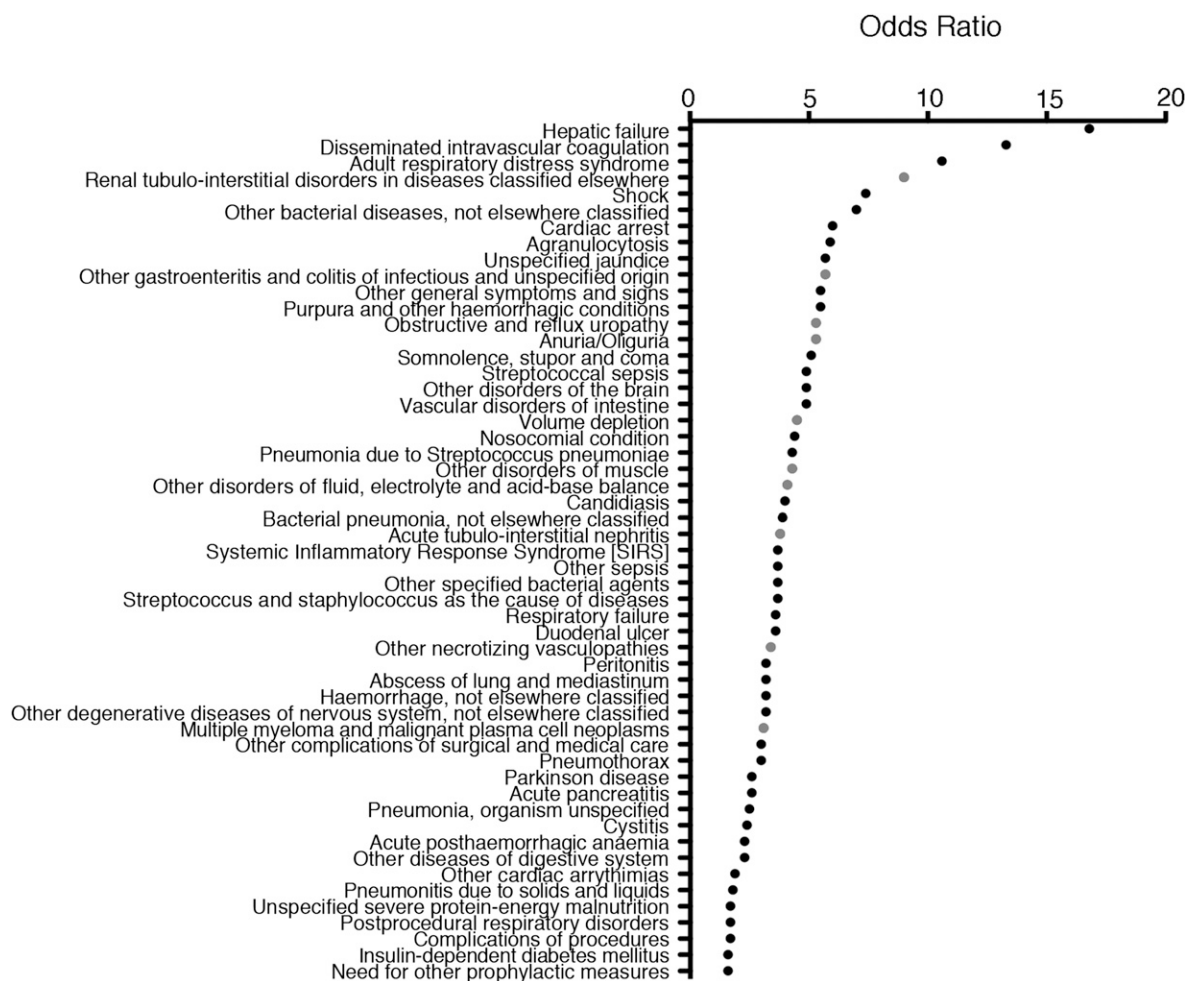


Figure 4. | ICD10 codes associated with N17 in patients with HA-AKI are enriched in diagnoses related to kidney diseases. Plot of the ORs for the 51 ICD-10 codes associated with N17 with a P value $< 6 \times 10^{-5}$ in the cohort of 8479 individuals with HA-AKI. ORs were calculated as follows: in the group of individuals with HA-AKI, $(\text{ICD-10+}/\text{ICD-10-})$ in individuals with N17 code divided by $(\text{ICD-10+}/\text{ICD-10-})$ in individuals without N17 code. A total of 1201 ICD-10 codes were identified in this cohort, of which 384 were present in one or two patients and excluded. Multiple comparisons were performed on 817 patients ($1201 - 384$), leading to a P value considered statistically significant of $(0.05/817) = 6 \times 10^{-5}$ after Bonferroni correction. Gray dots correspond to a diagnosis related to a kidney disease. HA-AKI, hospital-acquired AKI; ICD-10, International Classification of Diseases–10; OR, odds ratio

demonstrated the greatest increase in prevalence with time were related to surgery of neoplasms of the genitourinary system, possibly reflecting changes in indication and/or severity of the cases.

Thirty percent of patients with HA-AKI had the corresponding N17 code assigned. Bias in N17 coding may occur, favoring distortions between the occurrence of AKI defined by objective criteria and assigning of the N17 (“Acute kidney failure”) code. Nevertheless, this example underscores the gap that exists between the occurrence of a medical situation (namely, AKI) and its translation into a corresponding ICD-10 code: either AKI was not identified and diagnosed by the physician and, consequently, not coded, or AKI was diagnosed but considered a medical issue of minor importance and, therefore, not relevant enough to be encoded. In addition, coding may be influenced by payment incentives and local policies. The fact that AKI coding yields less reimbursement compared

with other codes might have influenced the prevalence of N17 coding. It is widely acknowledged that a large amount of medical information is not encoded but yet present in hospital stay reports, and appealing advancements in automatized methods of textual analysis have been recently made, allowing for the retrieval of diagnoses and other medical information of importance embedded in text (20–22). Taking into account the major effect of HA-AKI in terms of mortality and length of hospital stay (and its consequent medico-economic weight), our findings support the urgent need for efforts to be implemented to ensure more accurate identification of HA-AKI.

Our results should be interpreted while taking into account the limitations inherent to the design of our study. A single-center analysis is a limitation because diagnosis billing may differ between centers, and changes in EMR software and clinical decision support may influence some of the findings. The use of ICD-10 codes as explanatory

variables is also a limitation because their sensitivity and specificity can vary significantly, and their report is highly dependent of numerous factors, including the reporting physician and the influence of the billing system. In addition, the number of plasma creatinine measurements could be positively correlated with the false discovery of AKI, especially in patients with CKD. This association could be a consequence of the technical and biologic variability of plasma creatinine measurement results over time in an individual patient (23). We excluded patients who only had one plasma creatinine measurement performed during the first week after admission because AKI could not be identified in these patients, and their exclusion likely resulted in this study identifying AKI as more common than it is in actuality. The choice to use plasma creatinine at admission as the baseline for identifying AKI (because out-of-hospital baseline creatinine results are not recorded in our EMR) was likely to be associated with some bias, because this criterion may have missed cases of AKI that occurred on admission or overestimated the prevalence of CKD at admission. Along similar lines, a limitation of the utilization of an approach on the basis of a hospital CDW was the inability to retrieve information on long-term follow-up in cases with outpatient after-care, such as CKD occurrence. In line with this limitation, biologic parameters, including estimation of renal function, if performed outside the hospital clinical chemistry department, would not be included in the EMR.

In conclusion, we provide evidence that an EMR-based exploration of diagnosis-wide associations for a renal condition increased our ability to efficiently draw a comprehensive picture of the disease in real-life settings. We anticipate that researchers will be able to use the approach we present here to address complex issues in renal medicine.

Disclosures

None.

References

- Chawla LS, Eggers PW, Star RA, Kimmel PL: Acute kidney injury and chronic kidney disease as interconnected syndromes. *N Engl J Med* 371: 58–66, 2014
- Chertow GM, Burdick E, Honour M, Bonventre JV, Bates DW: Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. *J Am Soc Nephrol* 16: 3365–3370, 2005
- Xu X, Nie S, Liu Z, Chen C, Xu G, Zha Y, Qian J, Liu B, Han S, Xu A, Xu X, Hou FF: Epidemiology and clinical correlates of AKI in Chinese hospitalized adults. *Clin J Am Soc Nephrol* 10: 1510–1518, 2015
- Selby NM, Crowley L, Fluck RJ, McIntyre CW, Monaghan J, Lawson N, Kolhe NV: Use of electronic results reporting to diagnose and monitor AKI in hospitalized patients. *Clin J Am Soc Nephrol* 7: 533–540, 2012
- Tenenbaum JD, Avillach P, Benham-Hutchins M, Breitenstein MK, Crowgey EL, Hoffman MA, Jiang X, Madhavan S, Mattison JE, Nagarajan R, Ray B, Shin D, Visweswaran S, Zhao Z, Freimuth RR: An informatics research agenda to support precision medicine: Seven key areas. *J Am Med Inform Assoc* 23: 791–795, 2016
- Jensen PB, Jensen LJ, Brunak S: Mining electronic health records: Towards better research applications and clinical care. *Nat Rev Genet* 13: 395–405, 2012
- Warner JL, Alterovitz G, Bodio K, Joyce RM: External phenome analysis enables a rational federated query strategy to detect changing rates of treatment-related complications associated with multiple myeloma. *J Am Med Inform Assoc* 20: 696–699, 2013
- Neuraz A, Chouchana L, Malamut G, Le Beller C, Roche D, Beaune P, Degoulet P, Burgun A, Lorient MA, Avillach P: Phenome-wide association studies on a quantitative trait: Application to TPMT enzyme activity and thiopurine therapy in pharmacogenomics. *PLoS Comput Biol* 9: e1003405, 2013
- Roque FS, Jensen PB, Schmock H, Dalgaard M, Andreatta M, Hansen T, Søbey K, Bredkjær S, Juul A, Werge T, Jensen LJ, Brunak S: Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLoS Comput Biol* 7: e1002141, 2011
- Boland MR, Hripsak G, Albers DJ, Wei Y, Wilcox AB, Wei J, Li J, Lin S, Breene M, Myers R, Zimmerman J, Papapanou PN, Weng C: Discovering medical conditions associated with periodontitis using linked electronic health records. *J Clin Periodontol* 40: 474–482, 2013
- Lyalina S, Percha B, LePendu P, Iyer SV, Altman RB, Shah NH: Identifying phenotypic signatures of neuropsychiatric disorders from electronic medical records. *J Am Med Inform Assoc* 20(e2): e297–e305, 2013
- Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD: Visually integrating and exploring high throughput Phenome-Wide Association Study (PheWAS) results using PheWAS-View. *BioData Min* 5: 5, 2012
- Weber GM, Murphy SN, McMurphy AJ, Macfadden D, Nigrin DJ, Churchill S, Kohane IS: The Shared Health Research Information Network (SHRINE): A prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* 16: 624–630, 2009
- Zapletal E, Rodon N, Grabar N, Degoulet P: Methodology of integration of a clinical data warehouse with a clinical information system: The HEGP case. *Stud Health Technol Inform* 160: 193–197, 2010
- Ricci Z, Cruz DN, Ronco C: Classification and staging of acute kidney injury: Beyond the RIFLE and AKIN criteria. *Nat Rev Nephrol* 7: 201–208, 2011
- Lopes JA, Jorge S: The RIFLE and AKIN classifications for acute kidney injury: A critical and comprehensive review. *Clin Kidney J* 6: 8–14, 2013
- Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC: PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26: 1205–1210, 2010
- Real R, Vargas JM: The probabilistic basis of Jaccard's index of similarity. *Syst Biol* 45: 380–385, 1996
- Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM; RECORD Working Committee: The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med* 12: e1001885, 2015
- Scheurwegs E, Luyckx K, Luyten L, Daelemans W, Van den Bulcke T: Data integration of structured and unstructured sources for assigning clinical codes to patient stays. *J Am Med Inform Assoc* 23(e1): e11–e19, 2016
- DeLisle S, South B, Anthony JA, Kalp E, Gundlapalli A, Curriero FC, Glass GE, Samore M, Perl TM: Combining free text and structured electronic medical record entries to detect acute respiratory infections. *PLoS One* 5: e13377, 2010
- Ford E, Carroll JA, Smith HE, Scott D, Cassell JA: Extracting information from the text of electronic medical records to improve case detection: A systematic review. *J Am Med Inform Assoc* 23: 1007–1015, 2016
- Lin J, Fernandez H, Shashaty MG, Negoianu D, Testani JM, Berns JS, Parikh CR, Wilson FP: False-positive rate of AKI using consensus creatinine-based criteria. *Clin J Am Soc Nephrol* 10: 1723–1731, 2015

Received: October 20, 2016 **Accepted:** March 1, 2017

Published online ahead of print. Publication date available at www.cjasn.org.

This article contains supplemental material online at <http://cjasn.asnjournals.org/lookup/suppl/doi:10.2215/CJN.10981016/-/DCSupplemental>.