

# Validity and Reliability of Administrative Coded Data for the Identification of Hospital-Acquired Infections: An Updated Systematic Review with Meta-Analysis and Meta-Regression Analysis

Olga Redondo-González , José María Tenías , Ángel Arias ,  
and Alfredo J. Lucendo 

**Objective.** To conduct an updated assessment of the validity and reliability of administrative coded data (ACD) in identifying hospital-acquired infections (HAIs).

**Methods.** We systematically searched three libraries for studies on ACD detecting HAIs compared to manual chart review. Meta-analyses were conducted for prosthetic and nonprosthetic surgical site infections (SSIs), *Clostridium difficile* infections (CDIs), ventilator-associated pneumonias/events (VAPs/VAEs) and non-VAPs/VAEs, catheter-associated urinary tract infections (CAUTIs), and central venous catheter-related bloodstream infections (CLABSIs). A random-effects meta-regression model was constructed.

**Results.** Of 1,906 references found, we retrieved 38 documents, of which 33 provided *meta-analyzable data* ( $N = 567,826$  patients). ACD identified HAI incidence with high specificity ( $\geq 93$  percent), prosthetic SSIs with high sensitivity (95 percent), and both CDIs and nonprosthetic SSIs with moderate sensitivity (65 percent). ACD exhibited substantial agreement with traditional surveillance methods for CDI ( $\kappa = 0.70$ ) and provided strong diagnostic odds ratios (DORs) for the identification of CDIs (DOR = 772.07) and SSIs (DOR = 78.20). ACD performance in identifying nosocomial pneumonia depended on the ICD coding system (DOR<sub>ICD-10/ICD-9-CM</sub> = 0.05;  $p = .036$ ). Algorithmic coding improved ACD's sensitivity for SSIs up to 22 percent. Overall, high heterogeneity was observed, without significant publication bias.

**Conclusions.** Administrative coded data may not be sufficiently accurate or reliable for the majority of HAIs. Still, subgrouping and algorithmic coding as tools for improving ACD validity deserve further investigation, specifically for prosthetic SSIs. Analyzing a potential lower discriminative ability of ICD-10 coding system is also a pending issue.

**Key Words.** Hospital infections, International Classification of Diseases, surveillance, systematic review, meta-analysis

---

At the beginning of this century, the WHO Regional Office for Europe (WHO EURO) gave key recommendations for designing an international collaboration strategy to identify a reliable and effective surveillance model for “hospital-acquired” infections (HAIs). In particular, the recommendations promoted research on adjustment tools and standardized surveillance methods (Pittet et al. 2005).

“Hospital-acquired” infections represent the most frequent adverse event during care delivery and result in prolonged hospital stays, increased resistance to antimicrobials, additional costs for health systems and societies, and unnecessary deaths. As an example, the mortality rate for central venous catheter-related bloodstream infections (CR-BSIs) is as high as 25 percent. Approximately 10 percent of patients admitted to hospital acquire at least one infection during their hospital stay (Burke 2003; Allegranzi et al. 2011; Ling, Apisarnthanarak, and Madriaga 2015). Urinary tract infection (UTI) is the most common, accounting for up to 40 percent of all HAIs. It is noteworthy that up to 25 percent of hospitalized patients receive urinary catheters during their hospital stay, and over 80 percent of UTIs are associated with catheters (CAUTIs; Saint et al. 2008). Surgical site infections (SSIs) are the second most common HAI, leading to complications in 2–5 percent of inpatient surgical procedures. Finally, *Clostridium difficile* infection (CDI) is the most common cause of nosocomial infectious diarrhea (Schmiedeskamp et al. 2009). Despite their relevance, the real burden of HAIs remains unknown due to the complexity of the various surveillance systems and the lack of uniform criteria from country to country.

Most infection prevention programs in hospitals use objective definitions established by the Centers for Disease Control and Prevention’s

---

Address correspondence to Olga Redondo-González, M.D., M.P.H., Research Support Unit, Hospital General La Mancha Centro, Avda. Constitución n° 3, 13600 Alcázar de San Juan, Ciudad Real, Spain; e-mail: oredgon@gmail.es. José María Tenías, M.D., Ph.D., is with the Preventive Medicine Unit, Pare Jofré Hospital, Valencia, Spain. Ángel Arias, B.Sc., M.Sc., is with the Research Support Unit, Hospital General La Mancha Centro, Ciudad Real, Spain, and Centro de Investigación Biomédica En Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Madrid, Spain. Alfredo J. Lucendo, M.D., Ph.D., F.E.B.G.H., is with the Centro de Investigación Biomédica En Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), Madrid, Spain, and Department of Gastroenterology, Hospital General de Tomelloso, Ciudad Real, Spain.

(CDC's) National Nosocomial Infection Surveillance system (now the National Healthcare Safety Network or NHSN; CDC's National Healthcare Safety Network [NHSN], 2016). Still, although the NHSN disposes outlined criteria for identifying HAIs, both the methods and the data sources used for surveillance vary substantially among institutions. Additionally, conventional surveillance requires time-intensive medical record review by trained infection preventionists and is subject to interobserver variability (Klompas and Yokoe 2009). Electronic health data systems have been proposed as an automated alternative to improve the efficiency and accuracy of the surveillance process. Specifically, there is a growing interest in the evaluation of administrative coded data (ACD) as a tool for diagnosing HAIs (Jhung and Banerjee 2009). ACD were designed to be gathered from hospital discharge data (HDD) and are internationally recognized as they were created with the standardized coding format of the International Classification of Diseases (ICD). Notwithstanding, the precision and accuracy of ACD are affected by the subjectivity of the coding process and the variability of distinct coding versions (Jhung and Banerjee 2009; Moher et al. 2009; Schmiedeskamp et al. 2009; Cevasco et al. 2011).

The validity of ACD in estimating certain HAIs has been assessed in two previous systematic reviews conducted until March 2013 (Goto et al. 2014; Van Mourik et al. 2015). Nevertheless, the degree of agreement with respect to traditional surveillance remains unknown, and disparities in the accuracy of ACD in identifying subgroups of HAIs deserve further analysis, with special attention to the independent attributable influence of factors such as the reference standard used, the methodological quality of the study, coding formats and systems employed, and the country in which the study was conducted. An estimate of the heterogeneity found in this type of research must also be made. According to Booth et al. (Booth, Sutton, and Papaioannou 2012), a systematic review should be enhanced under the coverage of the following criteria: more participants than all previously included studies combined or larger than the previous; stimulation of both uptake and research by policy initiatives; a time gap for volatile topic areas since completion of the previous review; and appearance of new conceptual thinking/theory to supply an alternative framework.

By meeting the criteria set by Booth et al., the aim of this study was to conduct an updated systematic review with meta-analysis and meta-regression analysis on both the validity and reliability of ACD in identifying HAIs, with a particular emphasis on the subgroup analysis.

## METHODS

This systematic review has been registered in the PROSPERO International Prospective Register of Systematic Reviews (<http://www.crd.york.ac.uk/PROSPERO>), with registration number CRD42015023933, and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (Urrútia and Bonfill 2010).

### *Data Sources and Study Selection*

A systematic literature search was performed independently by three researchers (OR-G, AA, and JMT) in three major bibliographic databases (PubMed, EMBASE, and Scopus) for the period up to March 31, 2015. The search was not restricted with regard to language of publication.

Comprehensive search criteria were used to identify articles dealing with both HAIs and ACD in children and adults. In addition, we consulted the thesauri for MEDLINE (MESH) and EMBASE (EMTREE). The search strategy was as follows: (“Cross Infection” OR “Infection Control” OR “hospital infection” OR “intra-hospital infection” OR “nosocomial infection” OR “nosocomial” OR “hospital-acquired infections” OR “healthcare-associated infections” OR “healthcare associated infections”) AND (“Medical Records” OR “Administrative code data” OR “administrative coding data” OR “specific administrative codes” OR “ICD-9-CM administrative data” OR “ICD-10 administrative data” OR “coded-based algorithms” OR “coding-based algorithms” OR “coding based algorithms” OR “mbds” OR “Minimum Basic Data Set” OR “international classification of diseases” OR “Surveillance System Data” OR “billing codes” OR “billing data” OR “payment data”) AND (“Validation Studies” OR “Validity” OR “accuracy” OR “Validation” OR “predictive values” OR “diagnostic performance” OR “Sensitivity and Specificity” OR “reproducibility of results” OR “reliability” OR “diagnostic concordance” OR “Kappa index” OR “agreement” OR “interrater reliability” OR “interrater agreement” OR “Cohen’s Kappa”). For the Scopus database, only free-text searches with truncations were carried out. We also checked the reference lists from all retrieved articles in order to identify and examine additional relevant studies.

In the first stage, four reviewers (OR-G, AA, JMT, and AJL) independently screened the database search of all the articles retrieved for titles and abstracts by peer review. Afterward, the two epidemiologists, OR-G and JMT,

reviewed the preliminary selection, again independently. If any of the four reviewers felt that a title or abstract met the study eligibility criteria, the full text of the study was retrieved.

### *Inclusion Criteria*

1. Original research papers conducted in humans of any age.
2. Reports explicitly or implicitly on true positives, true negatives, false positives, and false negatives of ACD in detecting any type of HAI. Code-based algorithms were also considered.
3. Refers to electronic medical records (EMR; administrative data) exclusively using administrative coded data (ICD-9-CM or ICD-10 administrative coding systems).
4. Provides the list of code(s) for the detection of HAIs, or failing that, a bibliographic reference to the codes used.
5. Provides a comparison with the current gold standard based on manual chart review by trained personnel, which constitutes the traditional surveillance method, using either CDC or other standardized criteria (manual chart review and/or microbiological data in case of MRSA or *Clostridium difficile* infection).

### *Exclusion Criteria*

1. Long-stay institutions other than hospitals (e.g., nursing homes).
2. Evaluations of surveillance systems that do not use ACD.
3. Does not exclusively assess ACD, but also microbiological (except for CDI and MRSA), pharmacological, or radiological data or any other complementary diagnostic dataset.
4. Review articles, clinical guidelines, books, and consensus documents.
5. Letters or editorials not providing original data.
6. Studies providing duplicated information.

### *Quality Assessment*

We used the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) statement (Whiting et al. 2011) to evaluate the methodological and reporting quality of all the retrieved studies with respect to risk of bias and applicability concerns. Four researchers (OR-G, AA, JMT, and AJL)

independently gave each eligible study a rating of high, low, or unclear risk of bias. Studies were considered to have a low risk of bias if each of the bias items (patient selection, index test, reference test, flow, and timing) could be categorized as low risk; studies were classified as having a high risk of bias if even one of these items was valued as “unclear” or “high risk.”

### *Data Extraction*

Four researchers (OR-G, AA, JMT, and AJL) independently extracted the following data from each eligible study using a standardized data extraction sheet: last name of the first author; publication year; sample size; type of study population (child vs. adult); type of nosocomial infection; specific ICD-9 or ICD-10 codes; identified reference standard (CDC criteria vs. Other standardized criteria); number of true positives, true negatives, false positives, and false negatives; methodological design; study period; and all items for quality assessment, whenever possible. After cross-checking the results, discrepancies between reviewers were resolved through discussion. The authors were contacted by e-mail when additional information was needed.

The reviewers OR-G and JMT reorganized the extracted data by HAI type and made subgroups for SSIs (prosthetic vs. nonprosthetic SSIs) and nosocomial pneumonia (ventilator-associated pneumonia/event [VAP/VAE] vs. nosocomial pneumonias not caused by a ventilator [Non-VAP/VAE]). Recognizing the recent transition of criteria in NHSN surveillance, we reviewed VAE studies in addition to VAP ones, though only considering VAP in VAE criteria.

### *Statistical Analysis*

The standard methods recommended for the diagnostic accuracy of meta-analyses were used (Jones et al. 2010). The following measures of test accuracy and concordance were calculated: pooled sensitivity, specificity, positive likelihood ratio (LHR), negative LHR, diagnostic odds ratio (DOR), and  $\kappa$  coefficients (reliability) of all included studies, including 95 percent confidence intervals (CI). Sensitivity, specificity, and  $\kappa$  values were meta-analyzed by HAI type, as long as data on a specific HAI type were provided by at least three different studies. The threshold effect is caused by differences of sensitivity and specificity. To detect cutoff threshold effects, the relationship between sensitivity and specificity was evaluated for all meta-analyzable studies by the Spearman correlation coefficient.

Heterogeneity between studies was assessed by means of the chi-square test (Cochran  $Q$  statistic) and quantified with the  $I^2$  statistic. The various heterogeneity levels were as follows: 25–49 percent (low), 50–74 percent (moderate), and  $\geq 75$  percent (high; Higgins et al. 2003). When a significant  $Q$  test indicated heterogeneity among studies ( $p < .05$  or  $I^2 > 50$  percent), the random-effect model (Der Simonian–Laird method) was conducted for the meta-analysis to calculate the pooled sensitivity, specificity, and other related indexes. Funnel plots were designed to check for the existence of publication bias in articles on the sensitivity and specificity of ACD in identifying HAIs; we also used the Begg and Mazumdar correlation rank test (Begg and Mazumdar 1994) and the Egger (Egger et al. 1997) and Harbord tests (Harbord, Egger, and Sterne 2006) to further assess publication bias.

We estimated the discriminative capacity of ACDs to identify each type of meta-analyzable HAI by simultaneously considering the methodological quality (low vs. high risk of bias), the country where the study was conducted (USA vs. Non-USA), the type of ICD coding system used (ICD-10 vs. ICD-9-CM), the coding format employed (algorithmic coding vs. a single code), the reference standard (CDC vs. other standardized criteria), and the consideration or not by the different studies of the absence of the “Present on admission” (POA) code. To this end, a random-effects meta-regression model was constructed with the DOR as a measure of global discriminative capacity of ACD, since DOR represents the ratio of positive and negative likelihood ratios.

All statistical analyses were carried out with *StatsDirect* statistical software version 2.7.9 (StatsDirect Ltd, Cheshire, UK) and the freeware software *MetaDiSc* 1.4 for Windows (XI Cochrane Colloquium, Barcelona, Spain; Zamora et al. 2006). In every test, a two-sided  $p$ -value of  $<.05$  was considered statistically significant.

## RESULTS

Our search strategy yielded 1,906 documents; 1,790 references were excluded after examining the title and abstract because they did not fulfill the inclusion criteria. The full texts of the remaining 116 documents were retrieved for detailed evaluation; of these, 83 were excluded either because they did not fully meet the inclusion criteria or they met at least one of the exclusion criteria. Five additional references were found after tracking the reference lists of

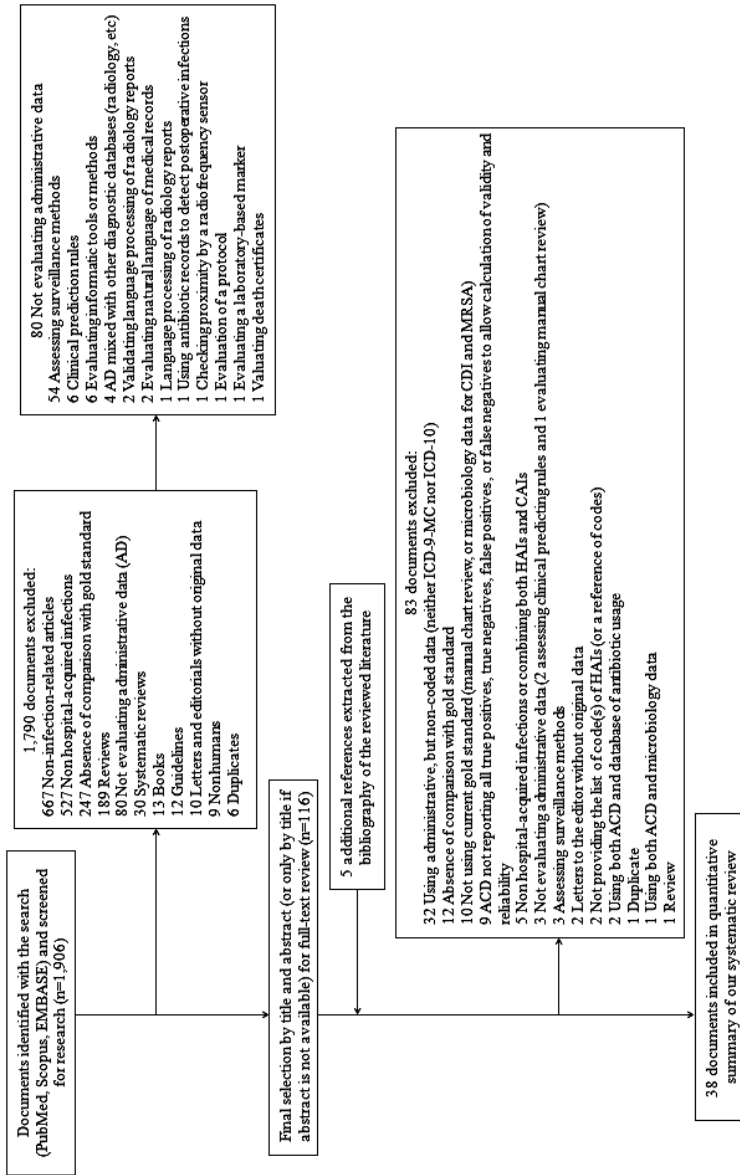
the reviewed literature; these were added to the selected documents. In the end, 38 documents were included in our systematic review (Figure 1). As the number of HAI types assessed in each individual document ranged from one to five, the 38 papers accounted for a total of 53 “*HAI-related included studies*.” *Meta-analyzable results* were provided by 33 of the 38 selected documents, overall providing data from 45 source evaluations across different HAIs that were then included in quantitative summaries (Table 1; Landis and Koch 1977; Hirschhorn, Currier, and Platt 1993; Baker et al. 1995; Hebden 2000; Cadwallader et al. 2001; Romano, Schembri, and Rainwater 2002; Curtis et al. 2004; Dubberke et al. 2006; Scheurer et al. 2007; Azaouagh and Stausberg 2008; Stausberg and Azaouagh 2008; Stevenson et al. 2008; Chang et al. 2008; Bolon et al. 2009; Schmiedeskamp et al. 2009; Zhan et al. 2009; Olsen and Fraser 2010; Schaefer et al. 2010; Verelst et al. 2010; Gerbier et al. 2011; Hollenbeak et al. 2011; Inacio et al. 2011; Schweizer et al. 2011; Shaklee et al. 2011; Chan et al. 2011; Gerbier-Colomban et al. 2012; Jones et al. 2012; Welker and Bertumen 2012; Yokoe et al. 2012; Calderwood et al. 2012; Knepper et al. 2013; Patrick et al. 2013; Van Mourik et al. 2013; Cass et al. 2013; Gardner et al. 2014; Grammatico-Guillon et al. 2014; Leclère et al. 2014; Redondo-González 2015; Pakyz et al. 2015). Inter-reviewer reliability for data extraction ranged from 0.64 to 0.78, indicating substantial agreement on the manuscripts selected for full reading.

The main characteristics of all the included studies, along with the validity and reliability estimators classified by HAI type, are summarized in Table 1. Overall, the 38 documents retrieved provided data from 1,071,935 reported cases of any kind of HAI, with study populations ranging from 130 to 466,819 subjects. When only meta-analyzable studies were considered, data from 567,826 patients (with study sizes ranging from 130 to 317,033 subjects) were retrieved. Table S1 shows contingency tables with the raw data of the included studies.

The pooled sensitivity and specificity of all the included studies were 0.46 (95 percent CI = 0.46–0.47;  $I^2 = 98.7$ ) and 0.99 (95 percent CI = 0.99–0.99;  $I^2 = 99.8$  percent), respectively (Figure S1, Table 2). The pooled sensitivity and specificity of the meta-analyzable studies were 0.65 (95 percent CI = 0.64–0.67;  $I^2 = 97.4$  percent) and 0.98 (95 percent CI = 0.98–0.98;  $I^2 = 99.8$  percent), respectively (Figure S2, Table 2). The pooled  $\kappa$  coefficients for all studies included were 0.35 (95 percent CI = 0.35–0.35;  $I^2 = 99.9$  percent) and for meta-analyzable studies was 0.34 (95 percent CI = 0.33–0.34;  $I^2 = 99.8$  percent; Table 2). The correlation found for all



Figure 1: Flow Chart



Note. ACD, administrative coded data; CAI, community-acquired infections; CDI, *Clostridium difficile* infection; MRSA, methicillin-resistant *Staphylococcus aureus*.

Table 1: Characteristics, Validity, and Reliability of All Included Studies

<i>HAI</i>	<i>First Author, Year</i>	<i>Country</i>	<i>SS (N)</i>	<i>Study Design</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>LHR+</i>	<i>LHR−</i>	<i>DOR</i>	<i>Kappa</i>
Prosth. SSI	Curtis M. 2004	Australia	380	Retrospective cohort	0.60 (0.39–0.79)	0.99 (0.97–1.00)	53.25 (19.09–148.51)	0.41 (0.25–0.65)	131.63 (86.98–468.45)	0.66 (0.56–0.76)
	Bolon MK. 2009	USA	2,128	Retrospective cohort	0.89 (0.75–0.96)	0.98 (0.98–0.99)	48.61 (34.87–67.77)	0.12 (0.05–0.26)	419.97 (156.87–1,124.3)	0.64 (0.60–0.68)
	Bolon MK. 2009	USA	4,194	Retrospective cohort	0.81 (0.70–0.88)	0.99 (0.98–0.99)	65.00 (48.44–87.22)	0.20 (0.13–0.31)	329.53 (175.87–617.47)	0.64 (0.61–0.67)
	Inacio MCS. 2011	USA	42,173	Retrospective cohort	0.97 (0.95–0.98)	0.92 (0.91–0.92)	11.71 (11.29–12.14)	0.04 (0.02–0.06)	322.33 (192.42–539.94)	0.19 (0.18–0.20)
	Calderwood MS. 2012	USA	576	Retrospective cohort	1.00 (0.29–1.00)	0.93 (0.91–0.95)	12.72 (7.89–20.49)	0.13 (0.01–1.80)	94.72 (4.81–1,866.3)	0.12 (0.08–0.16)
Non- Prosth. SSI	Calderwood MS. 2012	USA	724	Retrospective cohort	1.00 (0.40–1.00)	0.96 (0.94–0.97)	21.28 (13.51–33.50)	0.10 (0.01–1.45)	203.75 (10.73–3,870.3)	0.20 (0.16–0.24)
	Grammatico-Guillon L. 2014	France	1,010	Case–control	0.98 (0.96–0.99)	0.71 (0.67–0.74)	3.35 (2.97–3.77)	0.03 (0.01–0.06)	133.54 (58.56–304.49)	0.60 (0.54–0.66)
	Hirschhorn LR. 1993	USA	2,197	Retrospective cohort	0.65 (0.58–0.72)	0.98 (0.97–0.98)	26.95 (20.03–36.26)	0.36 (0.30–0.43)	75.30 (50.29–112.73)	0.66 (0.62–0.70)
	Baker C. 1995	USA	145	Retrospective	0.89 (0.52–1.00)	0.88 (0.81–0.93)	7.11 (4.31–11.74)	0.13 (0.02–0.81)	56.00 (6.59–476.00)	0.42 (0.28–0.56)
	Hebden J. 2000	USA	423	Cross-sectional	1.00 (0.75–1.00)	0.98 (0.96–0.99)	46.63 (23.79–91.38)	0.04 (0.01–0.56)	1,278.5 (70.12–23,312.4)	0.76 (0.67–0.85)
	Cadwalader HL. 2001	Australia	510	Retrospective cohort	0.81 (0.58–0.95)	0.99 (0.98–1.00)	79.17 (32.31–194.03)	0.19 (0.08–0.47)	411.40 (101.36–1,669.8)	0.78 (0.69–0.87)
	Romano PS. 2002	USA	991	Cross-sectional	0.63 (0.24–0.91)	0.99 (0.99–1.00)	307.19 (69.39–1,356.1)	0.38 (0.15–0.92)	817.50 (111.37–6,000.5)	0.66 (0.60–0.72)
	Stevenson KB. 2008	USA	3,882	Retrospective cohort	0.65 (0.57–0.73)	0.90 (0.88–0.91)	6.72 (5.76–7.84)	0.39 (0.31–0.48)	17.48 (12.20–25.05)	0.27 (0.24–0.30)
	Olsen MA. 2010	USA	12,00	Nested case–control	0.88 (0.77–0.94)	0.99 (0.99–1.00)	124.25 (61.91–249.38)	0.13 (0.07–0.24)	987.00 (357.31–2,726.4)	0.87 (0.81–0.93)
	Verelst S. 2010	Belgium	763	Retrospective cohort	0.79 (0.70–0.86)	0.95 (0.93–0.96)	15.65 (11.07–22.12)	0.22 (0.15–0.32)	71.07 (39.88–126.64)	0.71 (0.64–0.78)
	Gerbier S. 2011	France	446	Cross-sectional	0.79 (0.63–0.90)	0.66 (0.61–0.70)	2.30 (1.86–2.84)	0.32 (0.17–0.60)	71.8 (3.21–16.08)	0.17 (0.11–0.24)

Continued

Table 1. *Continued*

<i>HAI</i>	<i>First Author, Year</i>	<i>Country</i>	<i>SS (N)</i>	<i>Study Design</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>LHR+</i>	<i>LHR-</i>	<i>DOR</i>	<i>Kappa</i>
CDI	Hollenbeak CS, 2011	USA	1,066	Retrospective cohort	0.20 (0.13–0.30)	0.96 (0.94–0.97)	4.57 (2.78–7.51)	0.84 (0.75–0.93)	5.47 (3.04–9.86)	0.19 (0.13–0.25)
	Calderwood MS, 2012	USA	366	Retrospective cohort	1.00 (0.79–1.00)	0.89 (0.85–0.92)	8.85 (6.49–12.06)	0.03 (0.01–0.51)	267.86 (15.75–4,554.2)	0.42 (0.34–0.50)
	Griebier- Colomban S, 2012	France	446	Retrospective cohort	0.79 (0.63–0.90)	0.66 (0.61–0.70)	2.30 (1.86–2.84)	0.32 (0.17–0.60)	7.18 (3.21–16.08)	0.17 (0.11–0.23)
	Yokoe DS, 2012	USA	3,332	Retrospective	0.72 (0.67–0.77)	0.81 (0.80–0.83)	3.85 (3.49–4.26)	0.34 (0.29–0.41)	11.21 (8.71–14.42)	0.34 (0.31–0.37)
	Knipper BC, 2013	USA	2,449	Cross-sectional	0.66 (0.54–0.77)	0.96 (0.95–0.96)	15.15 (11.76–19.52)	0.35 (0.25–0.49)	42.84 (24.98–73.47)	0.39 (0.35–0.43)
	Leclerc B, 2014	France	4,400	Cross-sectional	0.25 (0.17–0.34)	0.98 (0.98–0.98)	12.47 (8.50–18.28)	0.77 (0.69–0.85)	16.28 (10.10–26.27)	0.23 (0.20–0.26)
	Dubberke ER, 2006	USA	3,630	Retrospective cohort	0.76 (0.73–0.80)	0.92 (0.91–0.93)	9.49 (8.35–10.80)	0.26 (0.22–0.29)	37.04 (29.64–46.29)	0.65 (0.62–0.68)
	Scheurer DB, 2007	USA	3,003	Retrospective cohort	0.71 (0.65–0.75)	0.99 (0.98–0.99)	53.92 (38.52–75.49)	0.30 (0.25–0.35)	181.05 (120.27–272.54)	0.76 (0.72–0.80)
	Schmiedeskamp M, 2009	USA	23,920	Retrospective cohort	0.97 (0.89–1.00)	1.00 (1.00–1.00)	265.38 (214.13–328.91)	0.03 (0.01–0.13)	8,196.9 (1972.5–34,063.1)	0.57 (0.56–0.58)
	Chan M, 2011	Singapore	2,212	Cross-sectional	0.50 (0.43–0.56)	1.00 (0.99–1.00)	120.59 (59.77–243.32)	0.51 (0.45–0.57)	238.41 (114.38–496.97)	0.62 (0.58–0.66)
	Shaklee J, 2011	USA	27,122	Retrospective cohort	0.81 (0.72–0.88)	1.00 (1.00–1.00)	703.51 (489.07–1,012.0)	0.19 (0.13–0.28)	3,647.3 (2,017.5–6,593.9)	0.77 (0.76–0.78)
	Jones G, 2012	France	31,7033	Retrospective cohort	0.36 (0.32–0.40)	1.00 (1.00–1.00)	1,909.2 (1,448.9–2,515.7)	0.64 (0.61–0.68)	2,964.1 (2,189.9–4,012.0)	0.49 (0.45–0.53)
	Welker JA, 2012	USA	23,495	Retrospective cohort	0.87 (0.83–0.90)	1.00 (0.99–1.00)	179.48 (148.61–216.77)	0.13 (0.10–0.17)	1,371.7 (966.23–1,947.4)	0.80 (0.79–0.81)
	Pakyz AL, 2015	USA	353	Retrospective cohort	0.68 (0.60–0.75)	0.93 (0.89–0.96)	10.19 (6.06–17.12)	0.35 (0.27–0.44)	29.37 (15.39–56.03)	0.63 (0.55–0.72)
	Romano PS, 2002	USA	991	Retrospective cohort	1.00 (0.03–1.00)	1.00 (0.99–1.00)	212.36 (56.91–792.39)	0.25 (0.02–2.77)	846.43 (29.16–24,571.0)	0.40 (0.35–0.45)
Non-VA Nos. Preu.— VA Event	Azounagh A, 2008	Germany	130	Retrospective cohort	0.43 (0.24–0.63)	0.99 (0.95–1.00)	43.714 (5.94–321.94)	0.58 (0.42–0.80)	75,750 (9.21–622,97)	0.52 (0.33–0.71)

*Continued*

Table 1. Continued

HA1	First Author, Year	Country	SS (N)	Study Design	Sensitivity	Specificity	LHR+	LHR-	DOR	Kappa
VAP/ VAE	Slausberg J, 2008	Germany	23,356	Retrospective cohort	0.43 (0.37-0.49)	1.00 (1.00-1.00)	125.03 (96.88-161.37)	0.57 (0.52-0.63)	218.06 (159.01-299.03)	0.50 (0.49-0.51)
	Gerbier S, 2011	France	1,499	Cross-sectional	0.40 (0.33-0.48)	0.92 (0.90-0.93)	4.73 (3.65-6.13)	0.65 (0.57-0.74)	725 (5.00-10.49)	0.30 (0.25-0.35)
	Stevenson KB, 2008	USA	193	Retrospective cohort	0.58 (0.37-0.78)	0.84 (0.78-0.89)	3.65 (2.25-5.92)	0.50 (0.31-0.80)	736 (2.96-18.29)	0.32 (0.19-0.45)
	Verelst S, 2010	Belgium	763	Cross-sectional	0.72 (0.53-0.87)	0.92 (0.90-0.94)	9.16 (6.563-12.80)	0.30 (0.17-0.54)	30.60 (12.98-72.11)	0.35 (0.29-0.41)
	Cass AL, 2013	USA	6,104	Retrospective cohort	0.61 (0.41-0.78)	0.93 (0.93-0.94)	9.00 (6.58-12.30)	0.42 (0.27-0.67)	21.36 (9.94-45.90)	0.07 (0.06-0.08)
	Romano PS, 2002	USA	991	Retrospective cohort	0.50 (0.32-0.68)	0.99 (0.99-1.00)	95.70 (37.51-244.16)	0.50 (0.36-0.70)	190.40 (62.97-575.70)	0.60 (0.45-0.75)
CAUTI	Zhan C, 2009	USA	17,059	Retrospective cohort	0.68 (0.64-0.72)	0.91 (0.91-0.92)	7.76 (7.195-8.36)	0.35 (0.31-0.39)	22.44 (18.59-27.08)	0.28 (0.27-0.29)
	Gerbier S, 2011	France	1,499	Cross-sectional	0.00 (0.00-0.03)	1.00 (0.99-1.00)	0.83 (0.05-14.56)	1.00 (0.99-1.01)	0.82 (0.05-14.72)	-0.01 (-0.03-0.01)
	Cass AL, 2013	USA	6,283	Retrospective cohort	0.70 (0.50-0.86)	0.95 (0.94-0.95)	13.84 (10.60-18.08)	0.31 (0.18-0.56)	44.34 (19.27-102.08)	0.10 (0.09-0.11)
	Gardner A, 2014	Australia	1,109	Retrospective cohort	0.60 (0.26-0.88)	0.93 (0.91-0.94)	8.24 (4.76-14.26)	0.43 (0.20-0.92)	19.11 (5.28-69.09)	0.11 (0.08-0.14)
CLABSI	Stevenson KB, 2008	USA	1,599	Retrospective cohort	0.48 (0.40-0.56)	0.66 (0.63-0.68)	1.41 (1.19-1.67)	0.79 (0.68-0.91)	1.79 (1.31-2.45)	0.07 (0.03-0.11)
	Cass AL, 2013	USA	28,761	Retrospective cohort	0.62 (0.51-0.72)	0.99 (0.99-0.99)	63.01 (51.59-76.96)	0.39 (0.30-0.50)	162.23 (104.65-251.50)	0.26 (0.25-0.27)
	Patrick SW, 2013	USA	2,920	Retrospective cohort	0.07 (0.01-0.18)	1.00 (0.99-1.00)	27.37 (7.31-102.47)	0.94 (0.87-1.01)	29.26 (7.31-117.03)	0.10 (0.07-0.13)
	Schaefer MK, 2010*	USA	22,172	Retrospective cohort	0.59 (0.54-0.64)	1.00 (1.00-1.00)	753.07 (464.98-1,219.6)	0.41 (0.37-0.46)	1,826.8 (1,091.2-3,058.2)	0.72 (0.70-0.73)
MRSA		USA	46,6819		0.24 (0.23-0.25)	0.99 (0.99-0.99)	35.47 (33.48-37.58)	0.77 (0.75-0.78)	46.36 (43.23-49.72)	0.26 (0.25-0.27)

Continued

Table 1. Continued

HA1	First Author, Year	Country	SS (N)	Study Design	Sensitivity	Specificity	LHR+	LHR-	DOR	Kappa
NARGE	Schweizer ML, 2011*	Spain	9,602	Retrospective cohort	0.67 (0.52-0.80)	1.00 (1.00-1.00)	201.05 (135.17-299.05)	0.33 (0.22-0.49)	613.66 (30760-1,224.23)	0.58 (0.46-0.70)
	Redondo- Gonzalez O, 2015*									
Nos. Bact.	Gerbier S. 2011*	France	1,498	Cross-sectional	0.59 (0.44-0.73)	0.84 (0.82-0.86)	3.76 (2.90-4.88)	0.48 (0.34-0.68)	777 (4.32-13.97)	0.14 (0.04-0.24)
NOI	Gerbier S. 2011*	France	902	Cross-sectional	0.43 (0.24-0.63)	0.85 (0.83-0.88)	2.95 (1.87-4.66)	0.67 (0.48-0.93)	4.41 (2.04-9.54)	0.10 (-0.04;0.23)
NIA	Chang DC. 2008*	USA	1,736	Cross-sectional	0.84 (0.60-0.97)	0.97 (0.96-0.98)	30.12 (21.44-42.33)	0.16 (0.06-0.46)	185.44 (52.28-657.76)	0.37 (0.21-0.54)
DRM	Van Mourik MSM, 2013*	Netherl.	617	Cross-sectional	0.32 (0.23-0.42)	0.89 (0.86-0.91)	2.87 (1.96-4.18)	0.77 (0.66-0.89)	3.74 (2.25-6.21)	0.22 (0.09-0.34)
PO	Verdet S. 2010*	Belgium	763	Cross-sectional	0.69 (0.50-0.84)	0.95 (0.93-0.96)	13.58 (9.18-20.09)	0.33 (0.20-0.55)	41.26 (18.22-93.45)	0.45 (0.30-0.61)
Sepsis										

\*Non-meta-analyzable studies.

CAUTI, catheter-associated urinary tract infection; CDI, *Clostridium difficile* infection; CLABSI, central line-associated bloodstream infection; DOR, diagnostic odds ratio; DRM, drain-related meningitis; HAI, hospital-associated infections; LHR, likelihood ratio; MRSA, methicillin-resistant *Staphylococcus aureus*; NARGE, nosocomial acute rotavirus gastroenteritis; Nos. Bact., nosocomial bacteremia; Nos. Pneu., nosocomial pneumonia; NIA, nosocomial invasive aspergillosis; Netherl., Netherlands; NOI, nosocomial obstetric infection; PO Sepsis, postoperative sepsis; Prosth. SSI, prosthetic surgical site infection; SS, sample size; VA, ventilator-associated; VAE, ventilator-associated event; VAP, ventilator-associated pneumonia.

meta-analyzable studies was weak (Spearman correlation coefficient =  $-0.241$ ;  $p = .111$ ).

The pooled likelihood ratios and overall DOR are summarized in Table 2. An analysis of the validity and reliability of each type of HAI is detailed below, along with their corresponding LHRs and DORs. Table S2 gives the list of the managed codes and reference standard type used in every study included.

### *Surgical Site Infections*

Overall, 19 documents evaluating 22 SSIs ( $N = 73,801$  patients) were analyzed. The number of diagnostic codes used varied greatly between studies, ranging from 1 to 104, with only four studies employing ICD-10 codes. One study assessed the single diagnostic code 996.62 (“Infectious and inflammatory reaction due to devices, implants and vascular grafts”; Yokoe et al. 2012) and three studies assessed the single diagnostic code 998.59 (“Other postoperative infection”; Hebden 2000; Hollenbeak et al. 2011; Knepper et al. 2013). The pooled sensitivity, specificity, and  $\kappa$  coefficients among the 19 studies were 0.79 (95 percent CI = 0.77–0.80), 0.92 (95 percent CI = 0.92–0.92), and 0.25 (95 percent CI = 0.24–0.25), respectively. The pooled positive LHR, negative LHR, and DOR were 15.20 (95 percent CI = 10.11–22.86), 0.21 (95 percent CI = 0.13–0.34), and 78.20 (95 percent CI = 40.41–151.33), respectively.

The pooled sensitivity of the four studies assessing a single diagnostic code was 0.62 (95 percent CI = 0.58–0.67); the pooled specificity was 0.89 (95 percent CI = 0.89–0.91), and the pooled DOR was 21.11 (95 percent CI = 7.34–60.77). The pooled sensitivity of the remaining SSIs identified by algorithmic ACD surveillance (combination of codes) resulted in 0.82 (95 percent CI = 0.80–0.84), whereas the pooled specificity was 0.93 (95 percent CI = 0.93–0.93) and the pooled DOR was 100.43 (95 percent CI = 48.52–207.91). Significant differences were observed between the pooled sensitivity and specificity of SSIs identified by a single ICD code and SSIs identified with algorithmic coding ( $p < .0001$ ).

The subgroup analysis included 5 documents assessing seven prosthetic SSIs ( $N = 51,185$  patients) and 15 documents assessing nonprosthetic SSIs ( $N = 22,616$  patients). A significant difference was observed between the pooled sensitivities of prosthetic and nonprosthetic SSIs, with values of 0.95 (95 percent CI = 0.93–0.96) and 0.65 (95 percent CI = 0.62–0.67),

respectively ( $p < .0001$ ). Nevertheless, no significant differences were found for the pooled specificities or pooled  $\kappa$  coefficients of these subgroups ( $p = .746$ , and  $p = .738$ ; respectively; Table 2, Figures 2 and 3a).

#### *Clostridium Difficile Infection*

Eight documents evaluating CDIs ( $N = 400,768$  patients) were analyzed. A single diagnostic code, either ICD-9-CM 008.45 ( $N = 7$ ) or ICD-10 A04.7 ( $N = 1$ ), was applied in all of them. The pooled sensitivity, specificity, and  $\kappa$  coefficient values of these studies were 0.65 (95 percent CI = 0.63–0.67), 1.00 (95 percent CI = 1.00–1.00), and 0.70 (95 percent CI = 0.69–0.70), respectively (Table 2, Figures 2 and 3a). The pooled positive LHR, negative LHR, and DOR were 119.20 (95 percent CI = 22.99–618.22), 0.26 (95 percent CI = 0.16–0.42), and 512.58 (95 percent CI = 90.44–2,905.0), respectively.

#### *Nosocomial Pneumonia*

Seven documents evaluating nosocomial pneumonia ( $N = 33,036$  patients) were analyzed. All of the studies on VAP/VAE employed ICD-9-CM codes while 75 percent of those dealing with non-VAP/VAE used ICD-10 codes. The pooled sensitivity, specificity, and  $\kappa$  coefficient of these studies were 0.45 (95 percent CI = 0.41–0.49), 0.98 (95 percent CI = 0.98–0.98), and 0.28 (95 percent CI = 0.27–0.29), respectively. Pooled positive LHR, negative LHR, and DOR were 19.19 (95 percent CI = 5.61–65.58), 0.56 (95 percent CI = 0.48–0.64), and 38.94 (95 percent CI = 7.64–198.58), respectively.

The subgroup analysis included four documents evaluating non-VAP/VAE ( $N = 25,976$  patients) and three documents assessing VAP/VAE ( $N = 7,060$  patients). No significant differences were observed between the pooled sensitivity, pooled specificity, and pooled  $\kappa$  coefficient of non-VAP/VAE cases with respect to VAP/VAE:  $p = .393$ ;  $p = .186$ , and  $p = .082$ , respectively (Table 2, Figures 2 and 3b).

#### *Catheter-Associated Urinary Tract Infections*

Five documents evaluating CAUTIs ( $N = 26,941$  patients) were analyzed. The pooled sensitivity, specificity, and  $\kappa$  coefficient of these studies were 0.56 (95 percent CI = 0.52–0.60), 0.93 (95 percent CI = 0.93–0.93), and 0.16 (95 percent CI = 0.15–0.17), respectively. The pooled positive LHR, negative

Table 2: Meta-Analysis Results of Meta-Analyzable HAIs by Groups and All Included (Meta-Analyzable and Non-Meta-Analyzable) HAIs

Groups of HAIs	Total Sample Size (N)	Pooled Sensitivity (CI 95%)	Pooled Specificity (CI 95%)	Pooled LHR+ (CI 95%)	Pooled LHR- (CI 95%)	Pooled DOR (CI 95%)	Pooled $\kappa$ (CI 95%)
Prosth. SSI	51,185	0.95 (0.93–0.96)	0.93 (0.92–0.93)	20.54 (7.81–54.03)	0.10 (0.04–0.29)	263.66 (183.68–378.48)	0.22 (0.21–0.22)
Non-Prosth. SSI	22,616	0.65 (0.62–0.67)	0.93 (0.92–0.93)	13.13 (8.21–21.00)	0.32 (0.22–0.46)	45.98 (23.87–88.55)	0.39 (0.38–0.40)
All SSI	73,801	0.79 (0.77–0.80)	0.92 (0.92–0.92)	15.20 (10.11–22.86)	0.21 (0.13–0.34)	78.20 (40.41–151.33)	0.25 (0.24–0.25)
CDI	400,768	0.65 (0.63–0.67)	1.00 (1.00–1.00)	119.20 (22.99–618.22)	0.26 (0.16–0.42)	512.58 (90.44–2,905.0)	0.70 (0.69–0.70)
Non-VA Nos. Pneu.-VA	25,976	0.42 (0.38–0.47)	0.99 (0.99–0.99)	47.28 (4.27–523.73)	0.60 (0.55–0.65)	84.65 (5.72–1,253.9)	0.48 (0.47–0.50)
Event							
VAP/VAE	7,060	0.64 (0.53–0.75)	0.93 (0.92–0.94)	6.89 (3.99–11.89)	0.41 (0.31–0.55)	17.20 (7.69–38.49)	0.11 (0.10–0.12)
All Nosocomial Pneumonia	33,036	0.45 (0.41–0.49)	0.98 (0.98–0.98)	19.19 (5.61–65.58)	0.56 (0.48–0.64)	38.94 (7.64–198.58)	0.28 (0.27–0.29)
CAUTI	26,941	0.56 (0.52–0.60)	0.93 (0.93–0.93)	12.94 (7.15–23.43)	0.47 (0.01–49.85)	30.69 (11.80–79.78)	0.16 (0.15–0.17)
CLABSI	33,280	0.46 (0.40–0.52)	0.98 (0.97–0.98)	13.30 (0.42–425.63)	0.67 (0.37–1.19)	20.27 (0.50–823.55)	0.23 (0.23–0.24)
All meta-analyzable studies	567,826	0.65 (0.64–0.67)	0.98 (0.98–0.98)	22.53 (15.50–32.74)	0.28 (0.16–0.48)	82.92 (44.40–154.88)	0.34 (0.33–0.34)
All included studies	1,071,935	0.46 (0.46–0.47)	0.99 (0.99–0.99)	22.52 (16.07–31.55)	0.34 (0.28–0.41)	75.77 (47.06–122.00)	0.35 (0.35–0.35)

Mohering RW 2013 (CLABSI). Stamm AM 2012 (CLABSI, CAUTI, AND VAP) and Drees M 2010 (VAP) were not included because these studies only showed sensitivity. When these studies were added to meta-analyze sensitivity by meta-analyzable groups, pooled sensitivities changed down in some cases: VAP/VAE 51.3 (43.2–59.3), CLABSI 22.3 (19.9–24.8), and CAUTI 54.3 (50.7–57.9).

CAUTI, catheter-associated urinary tract infection; CDI, *Clostridium difficile* infection; CLABSI, central line-associated bloodstream infection; DOR, diagnostic odds ratio; HAI, Hospital-associated infections; LHR, likelihood ratio; Nos. Pneu., nosocomial pneumonia; Prosth. SSI, prosthetic surgical site infection; VA, ventilator-associated; VAE, ventilator-associated event; VAP, ventilator-associated pneumonia.



Figure 2: Forest Plots of Sensibility and Specificity by Groups of All Meta-Analyzable Studies [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

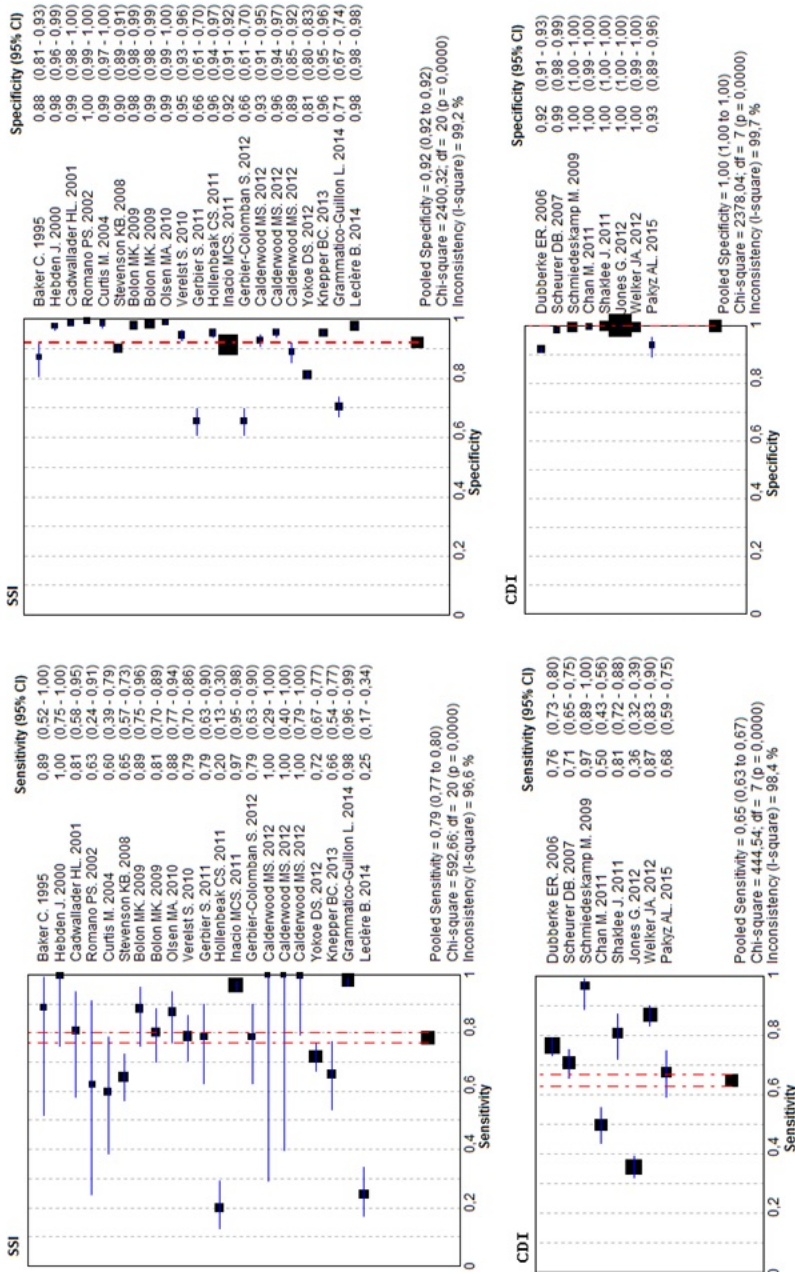


Figure 2: *Continued*

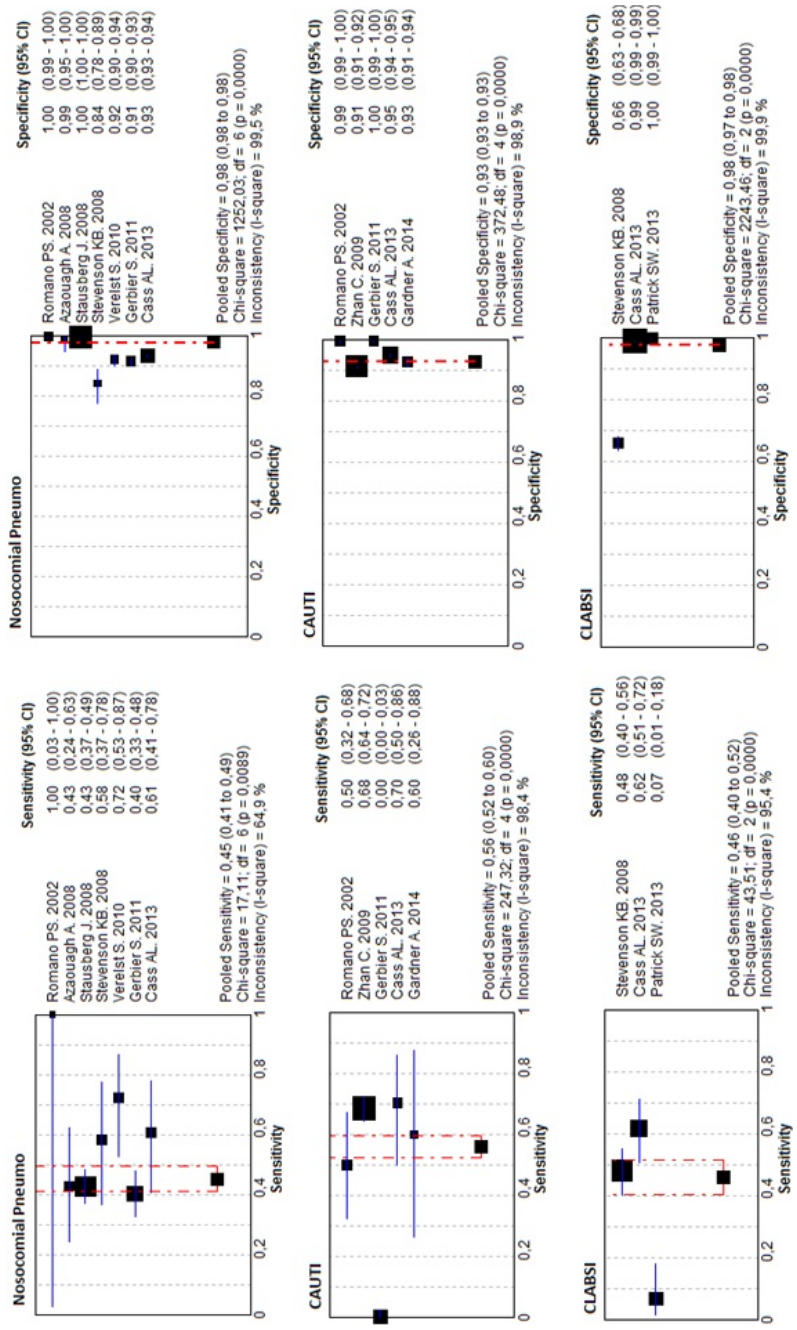


Figure 3: Forest Plots of  $\kappa$  Coefficients by Groups of Meta-Analyzable Studies: (a) SSI, CDI, and CLABSI. (b) Nosocomial Pneumonia and CAUTI

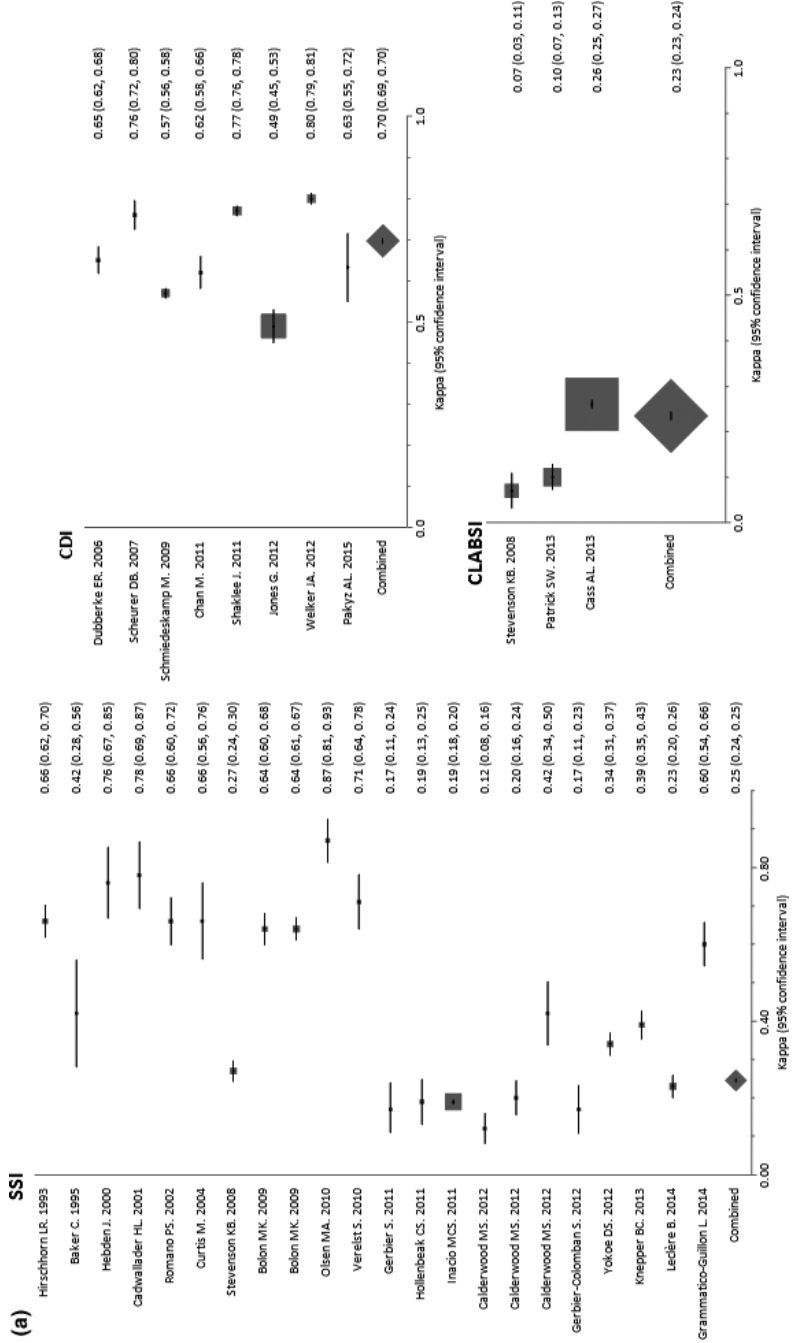
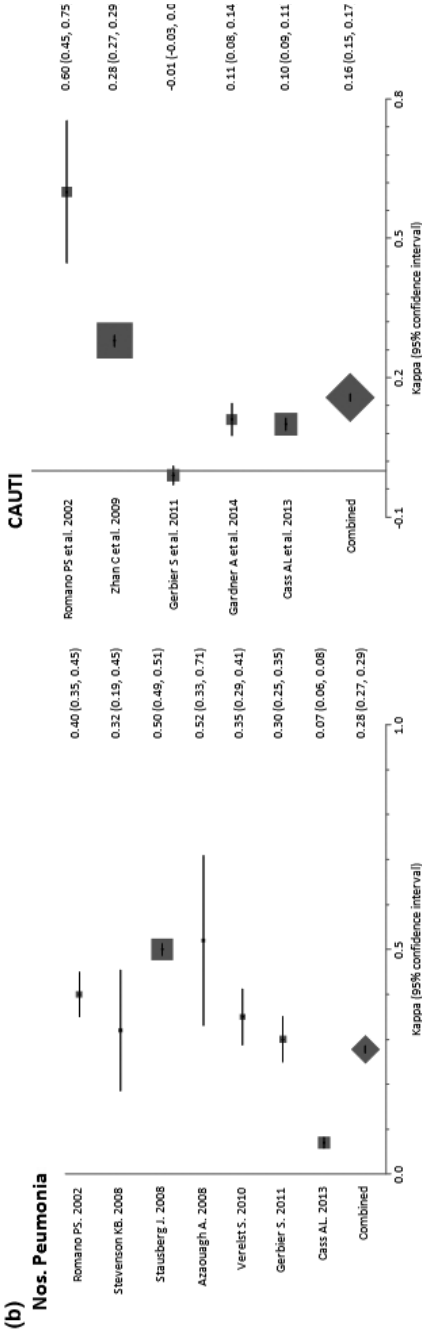


Figure 3: Continued



LHR, and DOR were 12.94 (95 percent CI = 7.15–23.43), 0.47 (95 percent CI = 0.01–49.85), and 30.69 (95 percent CI = 11.80–79.78), respectively (Table 2, Figures 2 and 3b).

### *Central Line-Associated Bloodstream Infections*

Three documents evaluating CLABSI ( $N = 33,280$  patients) were analyzed. No study used specific code for CLABSI (999.32), which was introduced to ICD-9-CM in October 2011. One study applied a procedure code (38.93) with whatever bloodstream infection code (Stevenson et al. 2008). The pooled sensitivity, specificity, and  $\kappa$  coefficient of these studies were 0.46 (95 percent CI = 0.40–0.52), 0.98 (95 percent CI = 0.97–0.98), and 0.23 (95 percent CI = 0.23–0.24), respectively. The pooled positive LHR, negative LHR, and DOR were 13.30 (95 percent CI = 0.42–425.63), 0.67 (95 percent CI = 0.37–1.19), and 20.27 (95 percent CI = 0.50–823.55), respectively (Table 2, Figures 2 and 3a).

The pooled sensitivity of two studies each assessing a single diagnostic code [38.93 “Venous catheterization, not elsewhere classified” (Stevenson et al. 2008) and 999.31 “Other and unspecified infection due to central venous catheter” (Patrick et al. 2013)] was 0.40 (95 percent CI 0.33–0.46), the pooled specificity was 0.89 (95 percent CI = 0.88–0.90), and the pooled DOR was 6.66 (95 percent CI = 0.42–105.5). The sensitivity, specificity, and DOR of the only CLABSI study to use algorithmic coding were 0.62 (95 percent CI = 0.51–0.72), 0.99 (95 percent CI = 0.99–0.99), and 162.23 (95 percent CI = 104.65–251.50), respectively.

### *Quality Assessment*

Table 3 summarizes the overall study quality as assessed with the aid of the QUADAS-2 checklist. Thirty-four papers were retrospective cohort studies, fifteen were cross-sectional, two were retrospective, one was a nested case-control, and another was a case-control study. Most of the studies were categorized as low quality due to their risk of bias as well as applicability concerns.

### *Meta-Regression Analysis*

Finally, both an overall analysis and analysis of subgroups categorized according to methodological quality, country of origin, type of ICD coding system,

Table 3: Summary of Quality Assessment of Diagnostic Accuracy of All Included Studies

HAI	First Author, Year	Risk of Bias			Applicability Concerns		
		Patient Selection	Index Tests	Reference Standard	Flow and Timing	Patient Selection	Index Tests
Prosth. SSI	Curtis M. 2004	Low	Low	Low	Low	Low	Low
	Bolon MK. 2009	Low	Low	Low	Low	Low	Low
	Bolon MK. 2009	Low	Low	Low	Low	Low	Low
	Inacio MCS. 2011	Low	Low	Low	Low	Low	Low
	Calderwood MS. 2012	Low	Low	Low	Low	Low	Low
	Calderwood MS. 2012	Low	Low	Low	Low	Low	Low
	Grammatico-Guillon L. 2014	Unclear	Low	Unclear	Low	Low	Unclear
	Hirschhorn LR. 1993	Low	Low	Low	Unclear	Low	Low
	Baker C. 1995	Low	Low	Low	Unclear	Low	Low
	Hebden J. 2000	Unclear	Low	Unclear	Unclear	Low	Unclear
Non-Prosth. SSI	Cadwallader HL. 2001	Low	Low	Low	Low	Low	Low
	Romano PS. 2002	Low	Low	Low	Low	Low	Low
	Stevenson KB. 2008	Low	Low	Low	Low	Low	Low
	Olsen MA. 2010	Low	Low	Low	Low	Low	Low
	Vereist S. 2010	Low	Unclear	Unclear	Unclear	Low	Unclear
	Gerbier S. 2011	Low	Low	Unclear	Low	Low	Unclear
	Hollenbeak CS. 2011	Low	Low	Low	Low	Low	Low
	Calderwood MS. 2012	Low	Low	Low	Low	Low	Low
	Gerbier-Colomban S. 2012	Low	Low	Low	Low	Low	Low
	Yokoe DS. 2012	Low	Low	Low	Low	Low	Low
CDI	Knepper BC. 2013	Low	Low	Low	Low	Low	Low
	Leclère B. 2014	Low	Low	Low	Low	Low	Low
	Dubberke ER. 2006	Unclear	Low	Low	Unclear	Unclear	Low
	Scheurer DB. 2007	Low	Low	Low	Low	Low	Low

Continued

Table 3. Continued

HAI	First Author, Year	Risk of Bias			Applicability Concerns			
		Patient Selection	Index Tests	Reference Standard	Flow and Timing	Patient Selection	Index Tests	Reference Standard
Non-VA Nos. Pneu.- VA Event	Schmiedeskamp M. 2009	Low	Low	Low	Low	Low	Low	Low
	Chan M. 2011	Low	Low	Low	Low	Low	Low	Low
	Shaklee J. 2011	Low	Low	Low	Low	Low	Low	Low
	Jones G. 2012	Low	Low	Unclear	Unclear	Low	Low	Unclear
	Welker JA. 2012	Low	Low	Low	Low	Low	Low	Low
	Pakyz AL. 2015	Low	Low	Low	Low	Low	Low	Low
	Romano PS. 2002	Low	Low	Low	Low	Low	Low	Low
	Azaouagh A. 2008	Low	Unclear	Low	Low	Low	Unclear	Low
	Stausberg J. 2008	High	Low	Low	Low	High	Low	Low
	Gerbier S. 2011	Low	Low	Unclear	Low	Low	Low	Unclear
VAP/VAE	Stevenson KB. 2008	Low	Low	Low	Low	Low	Low	Low
	Verelst S. 2010	Low	Unclear	Unclear	Unclear	Low	Unclear	Unclear
	Cass AL. 2013	Low	High	Low	Low	Low	High	Low
	Romano PS. 2002	Low	Low	Low	Low	Low	Low	Low
CAUTI	Zhan C. 2009	Low	Low	Unclear	Low	Low	Low	Unclear
	Gerbier S. 2011	Low	Low	Unclear	Low	Low	Low	Unclear
	Cass AL. 2013	Low	High	Low	Low	Low	High	Low
	Gardner A. 2014	Low	Low	Low	Low	Low	Low	Low
CLABSI	Stevenson KB. 2008	Low	Low	Low	Low	Low	Low	Low
	Cass AL. 2013	Low	High	Low	Low	Low	High	Low
	Patrick SW. 2013	Low	Low	Low	Low	Low	Low	Low
	Schaefer MK. 2010*	Low	Low	Low	Low	Unclear	Low	Low
MRSA	Schweizer ML. 2011*	Unclear	Low	Low	Unclear	Unclear	Low	Low
	Redondo-González O. 2015*	Low	Low	Low	Low	Low	Low	Low

Continued

Table 3. Continued

HAI	First Author, Year	Risk of Bias			Applicability Concerns		
		Patient Selection	Index Tests	Reference Standard	Flow and Timing	Patient Selection	Index Tests
Nos. Bact.	Gerbier S. 2011*	Low	Low	Unclear	Low	Low	Low
NOI	Gerbier S. 2011*	Low	Low	Unclear	Low	Low	Unclear
NIA	Chang DC. 2008*	Low	Unclear	Unclear	Low	Low	Low
DRM	Van Mourik MSM. 2013*	Low	Low	Low	Low	Low	Low
PO Sepsis	Verelst S. 2010*	Low	Unclear	Unclear	Unclear	Low	Unclear

\*Non-meta-analyzable studies.  
CAUTI, catheter-associated urinary tract infection; CDI, *Clostridium difficile* infection; CLABSI, central line-associated bloodstream infection; DRM, drain-related meningitis; HAI, Hospital-associated infections; MRSA, methicillin-resistant *Staphylococcus aureus*; NARGE, nosocomial acute rotavirus gastroenteritis; NIA, nosocomial invasive aspergillosis; NOI, nosocomial obstetric infection; Nos. Bact., nosocomial bacteremia; Nos. Pneu., nosocomial pneumonia; PO Sepsis, postoperative sepsis; Prosth. SSI, prosthetic surgical site infection; VA, ventilator-associated; VAE, ventilator-associated event; VAP, ventilator-associated pneumonia.



coding format, type of reference standard used, and the consideration or not of a negative POA code were carried out. No differences in the discriminative ability of ACD for the different types of HAIs were observed with regard to methodological quality, country of origin, coding format, type of reference standard used, or POA code. As for the coding system employed, studies using ICD-10 showed a significantly lower discriminative ability of ACD in identifying nosocomial pneumonias compared with studies that used ICD-9 ( $\text{DOR}_{\text{ICD-10/ICD-9-CM}} = 0.05$ ; 95 percent CI = 0.00–0.62;  $p = .0364$ ).

### *Publication Bias*

The funnel plots for sensitivity, specificity, and DOR by HAI group revealed no obvious asymmetry (Figures S3 and S4). The Begg-Mazumdar, Egger, and Harbord tests likewise indicated no evidence of publication bias ( $p > .05$  in all three cases).

## DISCUSSION

The present systematic review and meta-analysis of 33 diagnostic studies comprising 567,890 patients found that ACD identifies any given HAI with high specificity ( $\geq 0.93$ ) and prosthetic SSIs with high sensitivity (95 percent CI = 0.93–0.96). In contrast, both CDI and nonprosthetic SSIs were detected with the same moderate sensitivity (95 percent CI = 0.63–0.67 and 0.62–0.67), while nosocomial pneumonia, CAUTI, and CLABSI were detected with low sensitivity ( $< 0.60$ ). According to Landis and Koch criteria (Landis and Koch 1977), ACD-based surveillance showed a substantial agreement with traditional surveillance for CDI ( $\kappa > 0.60$ ); a discreet concordance for SSI, nosocomial pneumonia, and CLABSI ( $\kappa_s = 0.21$ –0.40); and an insignificant agreement for CAUTI ( $\kappa < 0.21$ ).

The Spearman correlation coefficient of sensitivity and 1-specificity was weak and not significant, suggesting the absence of threshold effect; in which case separate pooling of sensitivity and specificity is promoted, as we did (Reitsma et al. 2005). This assertion is reinforced by Simel and Bossuyt (2009), who suggest univariate meta-analysis for studies showing “good” likelihood ratios like ours. Better than predictive values, we evaluated the yield of LHR and DOR (positive LHR/negative LHR) results to assess ACD-based surveillance in each type of HAI. As pointed out by Glas et al., the DOR is closely linked to existing indicators of diagnostic performance, it facilitates formal

meta-analysis of studies on diagnostic test performance, and it is derived from logistic models, which allow for the inclusion of additional variables to correct for heterogeneity. DOR combines the strengths of sensitivity and specificity, as prevalence independent indicators, with the advantage of accuracy as a single indicator, characteristics that are also highly convenient in systematic reviews and meta-analyses (Glas et al. 2003). Anyway, raw data of all included studies for estimating any indicator of diagnostic performance are available in Table S1. In our study, ACD provided strong diagnostic evidence for identifying SSIs (DOR 40.41–151.33) and CDIs (DOR 15.39–56.03), moderate-to-strong evidence for nosocomial pneumonias (DOR 7.64–198.58) and CAUTIs (DOR 11.80–79.78), and poor evidence in the case of CLABSI (DOR 0.50–823.55). In the subgroup analyses, ACD performance in identifying nosocomial pneumonia was found to be significantly worse when ICD-10 codes were used. Moreover, CDC criteria seemed to be a stricter reference for checking ACD performance in identifying SSIs compared to other standardized criteria, but still fell just short of significance ( $p = .0728$ ). With respect to the coding format, while Goto affirmed that ACD employing algorithmic coding can only serve as a supplemental surveillance component, van Mourik went further, highlighting the need of improvement and validation for existing algorithms identifying HAIs (Goto et al. 2014; Van Mourik et al. 2015). We did also find significant improvements in the accuracy of retrospective surveillance of HAIs associated with certain processes (SSIs or CLABSI) when assessing algorithmic coding with respect to single codes. Although the performance of a single code in the case of SSIs did not differ from that of algorithmic coding ( $p = .0728$ ), significant differences were found between sensitivities ( $p < .0001$ ). In fact, the improvements in sensitivity and specificity for SSIs when using algorithmic coding reached a relevance of 22 and 4 percent, respectively. This observation would not be valid for CDI, as its diagnosis depends on a single code.

Although ACD from hospital discharge records were originally conceived as management tools, they now constitute a hopeful way to optimize surveillance, albeit not exempt from drawbacks. Previous comparisons of the accuracy of ACD with that of NNIS/NHSN methodology as the reference standard have yielded a wide range of results, depending on the infections identified and the codes used to indicate them. An overall sensitivity of 61 percent and a PPV of only 20 percent have been described (Sherman et al. 2006), with the lowest values being reported for CLABSI (Stone et al. 2007). Compared to traditional prospective surveillance, inconsistent documentation of an infection or the status of such infection upon admission may introduce

additional inaccuracy. But even prospective surveillance is not error-free; indeed, it may miss genuine infections. In some cases, patients may develop an infection and be discharged in the period between surveillance visits, as infection control preventionists are not able to visit every patient on a daily basis. In general, the published literature reports an overestimation or false-positive misclassification of HAIs with retrospective electronic surveillance as compared with prospective manual surveillance (Klompas and Yokoe 2009; Moher et al. 2009). Specifically, ACD has been proved to detect less than one-fifth of all HAI types, even though some studies assert that more than three-fourths of the cases labeled as HAIs by ACD were most likely false positives (Julian et al. 2006; Stevenson et al. 2008; Drees et al. 2010; Stamm and Bettacchi 2012). It is well known that 30 percent of the false positives produced by retrospective surveillance are actually cases that have been coded incorrectly (Cevasco et al. 2011). The two previous systematic reviews carried out prior to ours revealed an overall high specificity, but moderate-to-low sensitivity in identifying HAIs (Goto et al. 2014; Van Mourik et al. 2015). Similarly, our research has estimated a maximum 2 percent probability of ACD identifying a false HAI, but a maximum 36 percent probability for failing to diagnose a true HAI. As for SSIs, antimicrobial and/or diagnosis code-based screening criteria have been demonstrated to be more sensitive than routine surveillance, detecting many cases missed by this method (Yokoe et al. 2012).

A relevant drawback of codes in identifying HAIs is that they are assigned independently from the moment at which the infectious event begins; it is thus hard to ascertain whether the infection was already present at the time of admission or whether it developed over the course of hospitalization (Schmiedeskamp et al. 2009). In this regard, the introduction of a “Present on admission” (POA) flag has seemed to generate a better predictive ability (Bahl et al. 2008). However, both overall and making subgroups in the specific case of CDI and CAUTIs according to POA code (Table S2), we have not found differences between studies considering a negative POA code versus those not considering such code ( $p = .1$  for CAUTIs and  $p = .615$  for CDIs). Anyway, coding itself is slow and labor intensive, with expected ascertainment bias, as the results of automated ACD are influenced by local coding practices.

Just as Goto and van Mourik had pointed out, the moderate-to-low sensitivity we observed for all HAIs reinforces the hypothesis that ACD-based surveillance methods miss a relevant number of HAIs (Goto et al. 2014; Van Mourik et al. 2015). However, we have now found an exception for ACD displaying a high sensitivity in identifying prosthetic implants in SSIs. Generally,

improvements in accuracy were noted when a specific device or technique was involved in HAIs associated with procedures. In fact, Van Mourik et al. (2015) had already concluded that the identification of device-associated infections by ACD is the most challenging finding. In this line, our results revealed a 35 percent probability that ACD would not identify a true, nonprosthetic SSI, but only a 5 percent probability that ACD would fail to diagnose a prosthetic SSI. There is, however, a probable reason why differences in the accuracy of ACD-aided diagnosis of VAP/VAE compared to that of non-VAP/VAE did not reach statistical significance, namely the scant number of published studies dealing with nosocomial pneumonias. Further research is necessary to clarify this point.

With respect to other non-meta-analyzable infections, it is remarkable the substantial improvement in accuracy and concordance of the V09 code ("Infection by drug-resistant microorganisms") for MRSA diagnosis when it is combined with *Staphylococcus aureus* infection codes, reaching a maximum sensitivity of 64 percent and a  $\kappa$  index about 0.72 (Table 1). Such values are similar to those of both nosocomial rotavirus acute gastroenteritis and CDI summarized above, suggesting a probable good accuracy of ACD in identifying specific nosocomial pathogens.

The transition from ICD-9-CM to ICD-10 has recently arisen as a matter of concern for global health care systems. Despite ICD-10 providing a more detailed description of clinical situations, researchers have not identified significant improvements in the quality of ICD-10 data over that of ICD-9-CM (Quan et al. 2008; Topaz, Shafran-Topaz, and Bowles 2013). In this context, Quan et al. (2008) found that the validity of ICD-9-CM and ICD-10 data in recording 32 clinical conditions was generally similar, although the validity differed between coding versions for some events, none of which were an infectious condition. Similarly, we found no differences in the ability of ICD-9-CM and ICD-10 data to record HAIs, except for nosocomial pneumonias. Otherwise, while non-meta-analyzable analyzed studies employing the ICD-10 system display a sensibility under 0.60, those managing the ICD-9-CM system globally exceed 60 percent sensitivity. It is important to note, however, that while the existing evidence is limited, it is expected that the validity and generalizability of ICD-10 data will improve as coders gain experience with this new system.

One of the key questions when we consider CDI diagnosis is the choice of microbiologic diagnosis methods, which can greatly affect the reference standard. In this sense, we extracted all diagnostic methods used for CDI studies. No PCR was used for diagnosing *C. difficile* in any case, which might have

increased the prevalence of positive result greatly, making sensitivity of ACD decrease. However, we tested a stratified analysis of documents referring to EIA methods versus culture tests with or without other combinations; and no statistical differences were found ( $p = .910$ ); thus, we decided not to show these data disaggregated. The main disadvantage of detecting toxins by EIA methods is their relative lack of sensitivity with values that are in the range of 40–60 percent when compared to toxigenic culture, which has high sensitivity (Alcalá Hernández et al. 2015). But remarkably, the lowest sensitivity and  $\kappa$  coefficient (0.49; 95 percent CI 0.45–0.53) are for Jones G. 2012, which has a double combination of culture-based microbiologic techniques.

This study has several limitations. To begin with, coding inaccuracies are an important element that in and of themselves limit the value of ACD, as pointed out above. Furthermore, ACD cannot unequivocally determine the hospital-acquired condition of an infection, although it would seem logical that infections coded as a secondary diagnosis are most likely nosocomial (Iezzoni 2013). Even the onset of a HAI cannot be specifically clarified, as ICD codes can be assigned at any time during hospitalization. As mentioned before, HAIs can be present at the time of admission or can develop after an antimicrobial treatment or during hospitalization. Apart from that, definitions from the CDC's guidelines can vary and are used differently across countries, hospitals, and clinicians. It has been suggested that there are areas of residual subjectivity in discerning and applying the standard infection criteria from the NHSN definitions (Hebden 2012). Both previous systematic reviews conducted by Van Mourik and by Goto on the accuracy of ACD found that about 79 percent of the selected studies accomplishing the included criteria were from the United States Subgroups limited to U.S. studies for SSI and CDI in the Goto study showed that pooled sensitivity increased with minimal decrease on pooled specificity (Goto et al. 2014). We hypothesize that this finding might be due to the fact that surveillance practices tend to be more aggressive in countries with a health care system having more important implications in pay-for-performance. To all of that must be added that ICD codes are not specific enough to differentiate between the CDC definitions of laboratory-confirmed HAIs (Schmiedeskamp et al. 2009). Finally, all studies rely on the completeness and accuracy of their own chart review process where subjectivity is itself implicit. In this sense, traditional surveillance also has its limitations, as previously discussed. A consequence of all this variability due to the diversity of criteria or the coding processes, among others, is the heterogeneity in all estimates of validity and reliability (all  $I^2 \geq 95.4$ ), with

the exception of the pooled sensitivity of nosocomial pneumonias ( $I^2 = 64.9$  percent).

The main strength of our study is that it is the first one to simultaneously evaluate the accuracy and reliability of ACD in identifying HAIs; specifically, it is the first to explore subgroups of the same type of HAI and to adjust the findings by the type of ICD coding system, coding format, or reference standard criteria used. Additionally, we found no evidence of publication bias, neither graphically nor statistically, which strengthens the pooled quality of our results. Following the criteria of Booth et al. (Booth, Sutton, and Papaioannou 2012), we have appraised and reanalyzed both existing studies evaluated in two previous systematic reviews (Goto et al. 2014; Van Mourik et al. 2015) and the new covering 2 years after, through a thorough research that includes more participants than all previously included studies combined. We have contributed not only to formal, but also to methodological and statistical assessment of both heterogeneity and publication bias. New and different from the two previous reviews, this research received a protocol registration in Prospero, had no limits to language, and the search covered a more updated period of time until March 31, 2015 (Goto and van Mourik's studies accomplished their search until March 2013). Due to the complexity of the review process and unlike the other two, in our review we involved four reviewers who conducted a peer review, all of them with extensive experience and vast publications on systematic reviews and meta-analysis. Also new with our research, two epidemiologists checked independently the preliminary selection in a subsequent two-step review just to reinforce the quality of the search. In addition, we have conducted supplementary analysis looking at specific subgroups and certain study characteristics. We have also provided supplementary qualitative and quantitative material to enhance the interpretation of the review. We have unpicked, isolated, and proposed subgroups of HAIs as prosthetic SSIs for their potential implementation. We have enhanced the scope and extent of qualitative discussions of review limitations, implications for practice, and implications for further research; and also, we have re-examined the applicability of findings.

As requirements for continuous HAI reporting expand in scope and become more widely needed for hospital comparisons and reimbursement purposes, the labor-intensive nature of manual review by trained preventionists becomes an increasing problem. In the last two decades, automated ACD surveillance has been brought to the fore as a way to improve objectivity more rapidly and efficiently. However, we have provided evidence that ACD only displays discrete to poor reliability with manual surveillance in identifying

HAIs, except a good ability for CDI. Our findings reinforce the assessment that ACD are not sufficiently accurate for the surveillance of the majority of HAIs; furthermore, the high sensitivity found in the case of prosthetic SSIs arises as an encouraging result that deserves further evaluation. Its validation might contribute to enhance surveillance of the prosthetic joint infection, one of the leading causes of arthroplasty failure. Our results also suggest that subgrouping of HAIs and improving algorithmic coding might further optimize the ACD validity in identifying HAIs, specifically and again in the case of SSIs. Finally, albeit we have found no overall differences between ICD-9-CM and ICD-10 in recording HAIs, the potential lower discriminative ability of ICD-10 system remains an issue, especially for nosocomial pneumonias and other HAIs that could not be meta-analyzed.

## FUNDING

This research was not funded.

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* We would like to thank María del Carmen San Andrés Redondo for her contribution in translating available literature written in German.

*Disclosures:* None.

*Disclaimer:* None.

## REFERENCES

- Alcalá Hernández, L., M. Martín Arriaza, A. Mena Ribas, and J. Niubó Bosh. 2015. "Procedures in Clinical Microbiology," edited by E. Cercenado Mansilla, R. Cantón Moreno. Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica (SEIMC) [accessed on December 23, 2016]. Available at <https://www.seimc.org/contenidos/documentoscientificos/procedimientosmicrobiologia/seimc-procedimientomicrobiologia53.pdf>
- Allegranzi, B., S. BagheriNejad, C. Combescure, W. Graafmans, H. Attar, L. Donaldson, and D. Pittet. 2011. "Burden of Endemic Health-Care-Associated Infection in Developing Countries: Systematic Review and Meta-Analysis." *Lancet* 377 (9761): 228–41.



- Azaouagh, A., and J. Stausberg. 2008. "Frequency of Hospital-Acquired Pneumonia - Comparison between Electronic and Paper-Based Patient Records." *Pneumologie* 62 (5): 273-8.
- Bahl, V., M. A. Thompson, T.-Y. Kau, H. M. Hu, and D. A. Campbell Jr. 2008. "Do the AHRQ Patient Safety Indicators Flag Conditions That Are Present at the Time of Hospital Admission?" *Medical Care* 46 (5): 516-22.
- Baker, C., J. Luce, C. Chenoweth, and C. Friedman. 1995. "Comparison of Case Finding Methodologies for Endometritis after Cesarean Section." *American Journal of Infection Control* 23 (1): 27-33.
- Begg, C. B., and M. Mazumdar. 1994. "Operating Characteristics of a Rank Correlation Test for Publication Bias." *Biometrics* 50 (4): 1088-101.
- Bolon, M. K., D. Hooper, K. B. Stevenson, M. Greenbaum, M. A. Olsen, L. Herwaldt, G. A. Noskin, V. J. Fraser, M. Climo, Y. Khan, J. Vostok, and D. S. Yokoe. 2009. "Improved Surveillance for Surgical Site Infections after Orthopedic Implantation Procedures: Extending Applications for Automated Data." *Clinical Infectious Diseases* 48 (9): 1223-9.
- Booth, A., A. Sutton, and D. Papaioannou. 2012. *Systematic Approaches to a Successful Literature Review*, 2d Edition, edited by M. Steele. Thousand Oaks, CA: Sage.
- Burke, J. P. 2003. "Infection Control — A Problem for Patient Safety." *The New England Journal of Medicine* 348 (7): 651-6.
- Cadwallader, H. L., M. Toohey, S. Linton, A. Dyson, and T. V. Riley. 2001. "A Comparison of Two Methods for Identifying Surgical Site Infections Following Orthopaedic Surgery." *Journal of Hospital Infection* 48 (4): 261-6.
- Calderwood, M. S., A. Ma, Y. M. Khan, M. A. Olsen, D. W. Bratzler, D. S. Yokoe, D. C. Hooper, K. Stevenson, V. J. Fraser, R. Platt, and S. S. Huang. 2012. "Use of Medicare Diagnosis and Procedure Codes to Improve Detection of Surgical Site Infections Following Hip Arthroplasty, Knee Arthroplasty, and Vascular Surgery." *Infection Control and Hospital Epidemiology* 33 (1): 40-9.
- Cass, A. L., J. W. Kelly, J. C. Probst, C. L. Addy, and R. E. McKeown. 2013. "Identification of Device-Associated Infections Utilizing Administrative Data." *American Journal of Infection Control* 41 (12): 1195-9.
- CDC's National Healthcare Safety Network (NHSN). 2016. CDC/NHSN Surveillance Definitions for Specific Types of Infections [accessed on August 25, 2016]. Available at [http://www.cdc.gov/nhsn/PDFs/pscManual/17pscNosInfDef\\_current.pdf](http://www.cdc.gov/nhsn/PDFs/pscManual/17pscNosInfDef_current.pdf)
- Cevasco, M., A. M. Borzecki, D. A. McClusky, Q. Chen, M. H. Shin, K. M. F. Itani, and A. K. Rosen. 2011. "Positive Predictive Value of the AHRQ Patient Safety Indicator 'Postoperative Wound Dehiscence.'" *Journal of the American College of Surgeons* 212 (6): 962-7.
- Chan, M., P. L. Lim, A. Chow, M. K. Win, and T. M. Barkha. 2011. "Surveillance for Clostridium Difficile Infection: ICD-9 Coding Has Poor Sensitivity Compared to Laboratory Diagnosis in Hospital Patients, Singapore." *PLoS ONE* 6 (1): e15603.
- Chang, D. C., L. A. Burwell, G. M. Lyon, P. G. Pappas, T. M. Chiller, K. A. Wannermuehler, S. K. Fridkin, and B. J. Park. 2008. "Comparison of the Use of



- Administrative Data and an Active System for Surveillance of Invasive Aspergillosis." *Infection Control and Hospital Epidemiology* 29 (1): 25–30.
- Curtis, M., N. Graves, F. Birrell, S. Walker, B. Henderson, M. Shaw, and M. Whitby. 2004. "A Comparison of Competing Methods for the Detection of Surgical-Site Infections in Patients Undergoing Total Arthroplasty of the Knee, Partial and Total Arthroplasty of Hip and Femoral or Similar Vascular Bypass." *The Journal of Hospital Infection* 57 (3): 189–93.
- Drees, M., S. Hausman, A. Rogers, L. Freeman, and K. Wroten. 2010. "Underestimating the Impact of Ventilator-Associated Pneumonia by Use of Surveillance Data." *Infection Control and Hospital Epidemiology* 31 (6): 650–2.
- Dubberke, E. R., K. A. Reske, L. C. McDonald, and V. J. Fraser. 2006. "ICD-9 Codes and Surveillance for Clostridium Difficile-Associated Disease." *Emerging Infectious Diseases* 12 (10): 1576–9.
- Egger, M., G. Davey Smith, M. Schneider, and C. Minder. 1997. "Bias in Meta-Analysis Detected by a Simple, Graphical Test." *British Medical Journal* 315 (7109): 629–34.
- Gardner, A., B. Mitchell, W. Beckingham, and O. Fasugba. 2014. "A Point Prevalence Cross-Sectional Study of Healthcare-Associated Urinary Tract Infections in Six Australian Hospitals." *British Medical Journal Open* 4 (7): e005099.
- Gerbier, S., S. Bouzbid, E. Pradat, J. Baulieux, A. Lepape, M. Berland, J. Fabry, and M.-H. Metzger. 2011. "Use of the French Medico-Administrative Database (PMSI) to Detect Nosocomial Infections in the University Hospital of Lyon." *Revue d'épidémiologie et de Santé Publique* 59 (1): 3–14.
- Gerbier-Colomban, S., M. Bourjault, J.-C. Cêtre, J. Baulieux, and M.-H. Metzger. 2012. "Evaluation Study of Different Strategies for Detecting Surgical Site Infections Using the Hospital Information System at Lyon University Hospital, France." *Annals of Surgery* 255 (5): 896–900.
- Glas, A. S., J. G. Lijmer, M. H. Prins, G. J. Bonsel, and P. M. Bossuyt. 2003. "The Diagnostic Odds Ratio: A Single Indicator of Test Performance." *Journal of Clinical Epidemiology* 56 (11): 1129–35.
- Goto, M., M. E. Ohl, M. L. Schweizer, and E. N. Perencevich. 2014. "Accuracy of Administrative Code Data for the Surveillance of Healthcare-Associated Infections: A Systematic Review and Meta-Analysis." *Clinical Infectious Diseases* 58 (5): 688–96.
- Grammatico-Guillon, L., S. Baron, C. Gaborit, E. Rusch, and P. Astagneau. 2014. "Quality Assessment of Hospital Discharge Database for Routine Surveillance of Hip and Knee Arthroplasty-Related Infections." *Infection Control and Hospital Epidemiology* 35 (6): 646–51.
- Harbord, R. M., M. Egger, and J. A. C. Sterne. 2006. "A Modified Test for Small-Study Effects in Meta-Analyses of Controlled Trials with Binary Endpoints." *Statistics in Medicine* 25 (20): 3443–57.
- Hebden, J. 2000. "Use of ICD-9-CM Coding as a Case-Finding Method for Sternal Wound Infections after CABG Procedures." *American Journal of Infection Control* 28 (2): 202–3.

- Hebden, J. N. 2012. "Rationale for Accuracy and Consistency in Applying Standardized Definitions for Surveillance of Health Care-Associated Infections." *American Journal of Infection Control* 40 (5 suppl): S29–31.
- Higgins, J. P. T., S. G. Thompson, J. J. Deeks, and D. G. Altman. 2003. "Measuring Inconsistency in Meta-Analyses." *British Medical Journal* 327 (7414): 557–60.
- Hirschhorn, L. R., J. S. Currier, and R. Platt. 1993. "Electronic Surveillance of Antibiotic Exposure and Coded Discharge Diagnoses as Indicators of Postoperative Infection and Other Quality Assurance Measures." *Infection Control and Hospital Epidemiology* 14 (1): 21–8.
- Hollenbeak, C. S., M. M. Boltz, L. E. Nikkel, E. Schaefer, G. Ortenzi, and P. W. Dillon. 2011. "Electronic Measures of Surgical Site Infection: Implications for Estimating Risks and Costs." *Infection Control and Hospital Epidemiology* 32 (8): 784–90.
- Iezzoni, L. I. 2013. "Reasons for Risk Adjustment." In *Risk Adjustment for Measuring Health Care Outcomes*, 4th Edition, edited by L. I. Iezzoni, pp. 139–162. Chicago, IL: Health Administration Press.
- Inacio, M. C. S., E. W. Paxton, Y. Chen, J. Harris, E. Eck, S. Barnes, R. S. Namba, and C. F. Ake. 2011. "Leveraging Electronic Medical Records for Surveillance of Surgical Site Infection in a Total Joint Replacement Population." *Infection Control and Hospital Epidemiology* 32 (4): 351–9.
- Jhung, M. A., and S. N. Banerjee. 2009. "Administrative Coding Data and Health Care Associated Infections." *Clinical Infectious Diseases* 49 (6): 949–55.
- Jones, C. M., H. Ashrafian, P. Skapinakis, S. Arora, A. Darzi, K. Dimopoulos, and T. Athanasiou. 2010. "Diagnostic Accuracy Meta-Analysis: A Review of the Basic Principles of Interpretation and Application." *International Journal of Cardiology* 140: 138–44.
- Jones, G., N. Taright, P. Y. Boelle, J. Marty, V. Lalande, C. Eckert, and F. Barbut. 2012. "Accuracy of ICD-10 Codes for Surveillance of Clostridium Difficile Infections, France." *Emerging Infectious Diseases* 18 (6): 979–81.
- Julian, K. G., A. M. Brumbach, M. K. Chicora, C. Houlihan, A. M. Riddle, T. Umberger, and C. J. Whitener. 2006. "First Year of Mandatory Reporting of Healthcare-Associated Infections, Pennsylvania: An Infection Control-Chart Abstractor Collaboration." *Infection Control and Hospital Epidemiology* 27 (9): 926–30.
- Klompas, M., and D. S. Yokoe. 2009. "Automated Surveillance of Health Care Associated Infections." *Clinical Infectious Diseases* 48 (9): 1268–75.
- Knepper, B. C., H. Young, T. C. Jenkins, and C. S. Price. 2013. "Time-Saving Impact of an Algorithm to Identify Potential Surgical Site Infections." *Infection Control and Hospital Epidemiology* 34 (10): 1094–8.
- Landis, J. R., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–74.
- Leclère, B., C. Lasserre, C. Bourigault, M.-E. Juvin, M.-P. Chaillet, N. Mauduit, J. Cailion, M. Hanf, D. Lepelletier, and SSI Study Group. 2014. "Matching Bacteriological and Medico-Administrative Databases Is Efficient for a Computer-Enhanced Surveillance of Surgical Site Infections: Retrospective Analysis of

- 4,400 Surgical Procedures in a French University Hospital." *Infection Control and Hospital Epidemiology* 35 (11): 1330–5.
- Ling, M. L., A. Apisarnthanarak, and G. Madriaga. 2015. "The Burden of Healthcare Associated Infections in Southeast Asia: A Systematic Literature Review and Meta-Analysis." *Clinical Infectious Diseases* 60 (11): 1690–9.
- Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group. 2009. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *British Medical Journal* 339: b2535.
- Olsen, M. A., and V. J. Fraser. 2010. "Use of Diagnosis Codes and/or Wound Culture Results for Surveillance of Surgical Site Infection after Mastectomy and Breast Reconstruction." *Infection Control and Hospital Epidemiology* 31 (5): 544–7.
- Pakyz, A. L., J. A. Patterson, C. Motzkus-Feagans, S. F. Hohmann, M. B. Edmond, and K. L. Lapane. 2015. "Performance of the Present-on-Admission Indicator for Clostridium Difficile Infection." *Infection Control and Hospital Epidemiology* 36 (7): 838–40.
- Patrick, S. W., M. M. Davis, A. B. Sedman, J. A. Meddings, S. Hieber, G. M. Lee, T. L. Stillwell, C. E. Chenoweth, C. Espinosa, and R. E. Schumacher. 2013. "Accuracy of Hospital Administrative Data in Reporting Central Line-Associated Bloodstream Infections in Newborns." *Pediatrics* 131 (suppl): S75–80.
- Pittet, D., B. Allegranzi, H. Sax, L. Bertinato, E. Concia, B. Cookson, J. Fabry, H. Richet, P. Philip, R. C. Spencer, B. W. Ganter, and S. Lazzari. 2005. "Considerations for a WHO European Strategy on Health-Care-Associated Infection, Surveillance, and Control." *The Lancet Infectious Diseases* 5 (4): 242–50.
- Quan, H., B. Li, L. D. Saunders, G. A. Parsons, C. I. Nilsson, A. Alibhai, W. A. Ghali, and IMECCHI Investigators. 2008. "Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database." *Health Services Research* 43 (4): 1424–41.
- Redondo-González, O. 2015. "Validity and Reliability of the Minimum Basic Data set in Estimating Nosocomial Acute Gastroenteritis Caused by Rotavirus." *Revista española de Enfermedades digestivas* 107 (3): 152–61.
- Reitsma, J. B., A. S. Glas, A. W. S. Rutjes, R. J. Scholten, P. M. Bossuyt, P. M. Bossuyt, and A. H. Zwinderman. 2005. "Bivariate Analysis of Sensitivity and Specificity Produces Informative Summary Measures in Diagnostic Reviews." *Journal of Clinical Epidemiology* 58 (10): 982–90.
- Romano, P. S., M. E. Schembri, and J. A. Rainwater. 2002. "Can Administrative Data be Used to Ascertain Clinically Significant Postoperative Complications?" *American Journal of Medical Quality* 17 (4): 145–54.
- Saint, S., C. P. Kowalski, S. R. Kaufman, T. P. Hofer, C. A. Kauffman, R. N. Olmsted, J. Forman, J. Banaszak-Holl, L. Damschroder, and S. L. Krein. 2008. "Preventing Hospital-Acquired Urinary Tract Infection in the United States: A National Study." *Clinical Infectious Diseases* 46 (2): 243–50.
- Schaefer, M. K., K. Ellingson, C. Conover, A. E. Genisca, D. Currie, T. Esposito, L. Panttila, P. Ruestow, K. Martin, D. Cronin, M. Costello, S. Sokalski, S. Fridkin, and A. Srinivasan. 2010. "Evaluation of International Classification of Diseases,

- Ninth Revision, Clinical Modification Codes for Reporting Methicillin-Resistant Staphylococcus Aureus Infections at a Hospital in Illinois." *Infection Control and Hospital Epidemiology* 31 (5): 463–8.
- Scheurer, D. B., L. S. Hicks, E. F. Cook, and J. L. Schnipper. 2007. "Accuracy of ICD 9 Coding for Clostridium Difficile Infections: A Retrospective Cohort." *Epidemiology and Infection* 135 (6): 1010–3.
- Schmiedeskamp, M., S. Harpe, R. Polk, M. Oinonen, and A. Pakyz. 2009. "Use of International Classification of Diseases, Ninth Revision, Clinical Modification Codes and Medication Use Data to Identify Nosocomial Clostridium Difficile Infection." *Infection Control and Hospital Epidemiology* 30 (11): 1070–6.
- Schweizer, M. L., M. R. Eber, R. Laxminarayan, J. P. Furuno, K. J. Popovich, B. Hota, M. A. Rubin, and E. N. Perencevich. 2011. "Validity of ICD-9-CM Coding for Identifying Incident Methicillin-Resistant Staphylococcus Aureus (MRSA) Infections: Is MRSA Infection Coded as a Chronic Disease?" *Infection Control and Hospital Epidemiology* 32 (2): 148–54.
- Shaklee, J., D. M. Zerr, A. Elward, J. Newland, K. Leckerman, L. Asti, R. Guth, J. Bass, R. Selvarangan, S. Coffin, and T. Zaoutis. 2011. "Improving Surveillance for Pediatric Clostridium Difficile Infection: Derivation and Validation of an Accurate Case-Finding Tool." *The Pediatric Infectious Disease Journal* 30 (3): e38–40.
- Sherman, E. R., K. H. Heydon, K. H. St John, E. Teshner, S. L. Rettig, S. K. Alexander, T. Z. Zaoutis, and S. E. Coffin. 2006. "Administrative Data Fail to Accurately Identify Cases of Healthcare-Associated Infection." *Infection Control and Hospital Epidemiology* 27 (4): 332–7.
- Simel, D. L., and P. M. M. Bossuyt. 2009. "Differences between Univariate and Bivariate Models for Summarizing Diagnostic Accuracy May Not Be Large." *Journal of Clinical Epidemiology* 62 (12): 1292–300.
- Stamm, A. M., and C. J. Bettacchi. 2012. "A Comparison of 3 Metrics to Identify Health Care-Associated Infections." *American Journal of Infection Control* 40 (8): 688–91.
- Stausberg, J., and A. Azaouagh. 2008. "Frequency of Hospital-Acquired Pneumonia in Electronic and Paper-Based Patient Record." *Studies in Health Technology and Informatics* 136: 479–83.
- Stevenson, K. B., Y. Khan, J. Dickman, T. Gillenwater, P. Kulich, C. Myers, D. Taylor, J. Santangelo, J. Lundy, D. Jarjoura, X. Li, J. Shook, and J. E. Mangino. 2008. "Administrative Coding Data, Compared with CDC/NHSN Criteria, Are Poor Indicators of Health Care-Associated Infections." *American Journal of Infection Control* 36 (3): 155–64.
- Stone, P. W., T. C. Horan, H.-C. Shih, C. Mooney-Kane, and E. Larson. 2007. "Comparisons of Health Care-Associated Infections Identification Using Two Mechanisms for Public Reporting." *American Journal of Infection Control* 35 (3): 145–9.
- Topaz, M., L. Shafraan-Topaz, and K. H. Bowles. 2013. "ICD-9 to ICD-10: Evolution, Revolution, and Current Debates in the United States." *Perspectives in Health Information Management* 10: 1d.
- Urrútia, G., and X. Bonfill. 2010. "PRISMA Declaration: A Proposal to Improve the Publication of Systematic Reviews and Meta-Analyses." *Medicina Clínica* 135 (11): 507–11.

- Van Mourik, M. S. M., A. Troelstra, K. G. M. Moons, and M. J. M. Bonten. 2013. "Accuracy of Hospital Discharge Coding Data for the Surveillance of Drain-Related Meningitis." *Infection Control and Hospital Epidemiology* 34 (4): 433–6.
- Van Mourik, M. S. M., P. J. van Duijn, K. G. M. Moons, M. J. M. Bonten, and G. M. Lee. 2015. "Accuracy of Administrative Data for Surveillance of Healthcare-Associated Infections: A Systematic Review." *British Medical Journal Open* 5 (8): e008424.
- Verelst, S., J. Jacques, K. Van den Heede, P. Gillet, P. Kolh, A. Vleugels, and W. Sermeus. 2010. "Validation of Hospital Administrative Dataset for Adverse Event Screening." *Quality & Safety in Health Care* 19 (5): e25.
- Welker, J. A., and J. B. Bertumen. 2012. "Toxin Assay is More Reliable Than ICD-9 Data and Less Time-Consuming Than Chart Review for Public Reporting of Clostridium Difficile Hospital Case Rates." *Journal of Hospital Medicine* 7 (3): 170–5.
- Whiting, P. F., A. W. S. Rutjes, M. E. Westwood, S. Mallett, J. J. Deeks, J. B. Reitsma, M. M. G. Leeflang, J. A. C. Sterne, P. M. M. Bossuyt, and QUADAS-2 Group. 2011. "QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies." *Annals of Internal Medicine* 155 (8): 529–36.
- Yokoe, D. S., Y. Khan, M. A. Olsen, D. C. Hooper, M. Greenbaum, J. Vostok, J. Lankiewicz, V. J. Fraser, and K. B. Stevenson. 2012. "Enhanced Surgical Site Infection Surveillance Following Hysterectomy, Vascular, and Colorectal Surgery." *Infection Control and Hospital Epidemiology* 33 (8): 768–73.
- Zamora, J., V. Abraira, A. Muriel, K. Khan, and A. Coomarasamy. 2006. "Meta-DiSc: A Software for Meta-Analysis of Test Accuracy Data." *BMC Medical Research Methodology* 6: 31.
- Zhan, C., A. Elixhauser, C. L. Richards, Y. Wang, W. B. Baine, M. Pineau, N. Verzier, R. Kliman, and D. Hunt. 2009. "Identification of Hospital-Acquired Catheter-Associated Urinary Tract Infections from Medicare Claims: Sensitivity and Positive Predictive Value." *Medical Care* 47 (3): 364–9.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.

Figure S1. Forest Plots of Sensibility and Specificity for All Included Studies.

Figure S2. Forest Plots of Sensibility and Specificity for All Meta-Analizable Studies.

Figure S3. Begg Funnel Plots of Studies Evaluating Publication Bias of Articles on the Sensitivity and Specificity of Administrative Coded Data in Evaluating Hospital-Acquired Infections.

Figure S4. Begg Funnel Plots of Studies Evaluating Publication Bias of Articles on the Diagnosis Odds Ratio of Administrative Coded Data in Evaluating Hospital-Acquired Infections.

Table S1. Contingency Tables Including True and False Positives and Negatives of All Included Studies.

Table S2. ICD-9-CM and ICD-10 Codes of All Included Studies.