BMJ Open

Validation of discharge diagnosis codes to identify serious infections among older adults

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020857
Article Type:	Research
Date Submitted by the Author:	27-Nov-2017
Complete List of Authors:	Wiese, Andrew; Vanderbilt University Medical Center, Health Policy Griffin, Marie R; Vanderbilt University Medical Center, Health Policy Stein, Michael; Vanderbilt University, Pharmacology Schaffner, William; Vanderbilt University Medical Center, Health Policy Greevy, Robert; Vanderbilt School of Medicine, Biostatistics Mitchel, Jr., Edward; Vanderbilt University Medical Center, Health Policy Grijalva, Carlos; Vanderbilt University, Health Policy
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Infectious diseases, Epidemiology
Keywords:	coding algorithms, Medicaid, older adults, serious infections

Validation of discharge diagnosis codes to identify serious infections among older adults

Running title: Validation of diagnosis codes to identify infections

Authors: Andrew D. Wiese, PhD, MPH[1]; Marie R. Griffin[1,2], MD, MPH; C. Michael Stein, MB, ChB[3]; William Schaffner, MD[1]; Robert Greevy, PhD[4]; Edward F. Mitchel Jr., MS[1]; Carlos G. Grijalva, MD, MPH[1,2]

Affiliations: [1]Department of Health Policy, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; [2]Mid-South Geriatric Research Education and Clinical Center, VA Tennessee Valley Health Care System, Nashville, Tennessee, USA; [3]Departments of Pharmacology and [4]Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Corresponding Author: Andrew D. Wiese, PhD, MPH; Department of Health Policy, Vanderbilt University Medical Center, Suite 2600, Village at Vanderbilt, 1500 21st Avenue South, Nashville, TN 37212; phone: (615) 875-7997; email: andrew.d.wiese@vanderbilt.edu

Key words: coding algorithms; Medicaid; older adults; serious infections

Word Count: 3,411 /4,000

Tables and Figures: (3/5)

ABSTRACT (262/300)

Objectives: Hospitalizations for serious infections are common among older adults and frequently used as study outcomes. Yet few studies have evaluated the performance of diagnosis codes to identify serious infections in this population. We sought to determine the positive predictive value (PPV) of diagnosis codes for identifying hospitalizations due to serious infections among older adults.

Setting and participants: We identified hospitalizations for possible infection among adults ≥50 years enrolled in the Tennessee Medicaid healthcare program (2008-2013) using ICD-9 diagnosis codes for pneumonia, meningitis/encephalitis, bacteremia/sepsis, cellulitis/soft-tissue infections, endocarditis, pyelonephritis and septic arthritis/osteomyelitis.

Design: Medical records were systematically obtained from hospitals randomly selected from a stratified sampling framework based on geographical region and hospital discharge volume.

Measures: Two trained clinical reviewers used a standardized extraction form to abstract information from medical records. Pre-defined algorithms served as reference to adjudicate confirmed infection-specific hospitalizations. We calculated the PPV of diagnosis codes using confirmed hospitalizations as reference. Sensitivity analyses determined the PPV robustness to definitions that required radiological or microbiological confirmation. We also determined interrater reliability between reviewers.

Results: The PPV of diagnosis codes for hospitalizations for infection (n=716) was 90% (95% CI: 88-92). The PPV was highest for pneumonia [97% (95% CI: 95-98)] and cellulitis [91% (95% CI: 86-96)], and lowest for meningitis/encephalitis [50% (95% CI: 19-81)]. The adjudication reliability was excellent [93% agreement; first agreement-coefficient: 0.91]. The overall PPV was lower when requiring microbiological confirmation [45%] and when requiring radiological confirmation for pneumonia [79%].

Conclusions: Discharge diagnosis codes have a high PPV for identifying hospitalizations for serious infections among older adults, especially for common infections.

STRENGTHS AND LIMITATIONS OF THE STUDY

- This study examined the performance of diagnosis coding algorithms to identify hospitalizations due to serious infections among older adults enrolled in a State Medicaid program using a systematic and representative sample of records from hospitals of different sizes and in distinct State regions.
- The reference criteria to identify true infections was based on previous literature and clinical expertise but may be imperfect. Nevertheless, identifying microbiologically-confirmed infections is difficult due to the low sensitivity of culture-based diagnostic methods often used in clinical practice.
- Diagnosis codes were based on the ICD-9-coding system only. These findings will continue to be helpful for retrospective studies that encompass periods of ICD-9 use, yet additional studies evaluating the performance of ICD-10-based codes would be beneficial.
- Our coding algorithms to identify serious infections had a high positive predictive value overall, and will be useful in ongoing and future research using administrative data

INTRODUCTION

Infectious diseases remain a leading cause of morbidity and mortality in the U.S. and elsewhere (1). Older adults, in particular, are at high risk for serious infections and their long-term consequences (2, 3). Among older adults, community-acquired serious infections (including pneumonia, sepsis, and meningitis) often require hospitalization and represent a substantial burden on the U.S. healthcare system (4-7). Therefore, it is important to monitor the incidence of these infections, identify important risk factors, and determine the impact of preventative policies (e.g., vaccination) on these diseases among older adults (8-10).

Large-scale epidemiological studies using administrative data often use serious infections as outcomes (11-15). However, few studies have evaluated the performance of diagnosis codes to identify serious infections among older adults. Most previous studies that have assessed the performance of coded discharge diagnosis codes to identify serious infections have focused mainly on common infections (e.g., pneumonia or sepsis), specific populations (e.g., patients with rheumatoid arthritis), or on healthcare-associated or hospital-acquired infections (16-25). Nevertheless, the performance of coded discharge diagnoses for accurately identifying infections requiring hospitalization among older adults is unclear. Therefore, we sought to determine the positive predictive value (PPV) of specific discharge diagnoses for identifying infections that required hospitalization among older adults.

METHODS

Data sources

TennCare is the managed Medicaid program in the State of Tennessee that provides healthcare insurance to those who are Medicaid eligible (around 20% of the Tennessee population). The adult TennCare population consists of low-income pregnant women and individuals who are elderly or have a disability (over 600,000 annually) (26). We used data from TennCare, supplemented with data from the Tennessee Hospital Discharge System (a registry for all hospitalizations in Tennessee) and pharmacy information from Medicare Part D for those that were dual eligible, to identify a retrospective cohort of

TennCare enrollees ≥50 years of age with pharmacy benefits (2008-2013). Cohort members had at least 180 days of baseline continuous enrollment before cohort entry, and were also required to be free of certain life-threatening conditions known to increase the risk of infection (solid organ transplantation, end-stage renal disease, HIV/AIDS, malignancy and serious kidney, liver and respiratory disease), have evidence of at least one pharmacy prescription fill and evidence of at least one healthcare encounter during baseline (to ensure detection of healthcare usage). Follow-up started on the earliest date the inclusion criteria were met and continued through the earliest of the following: study end date (December 31, 2013), the day prior to diagnosis of a serious life-threatening condition that would have precluded entry to the cohort, loss of enrollment, or date of death. From this retrospective cohort, we identified possible hospitalizations for serious infections (see *Identification of hospitalizations for serious infection)* for our validation study. To avoid including infections that may have originated due to a previous hospital stay, we excluded hospitalizations for infections that occurred in the 30-day period after discharge from a previous hospitalization.

Identification of hospitalizations for serious infection

Clinical knowledge and a literature review were used to identify primary discharge diagnosis codes that have been used previously to identify specific serious infections that require hospitalization (*study infections)*, including pneumonia (alone or with a primary diagnosis of bacteremia/sepsis), bacteremia/sepsis, pyelonephritis, meningitis/encephalitis, osteomyelitis/septic arthritis, endocarditis and cellulitis (25, 27-29). Specific International Classifications of Diseases-Clinical Modification 9th-revision (ICD-9-CM) diagnosis codes used to identify possible hospitalizations for each infection type are presented in Table 1.

Sampling Strategy

We used stratified random sampling to select a representative subset of study infection hospitalizations from among all possible cases identified in the retrospective cohort. Since larger hospitals would be over represented in a purely random sampling, and because there may also be regional variability in coding practices and infection prevalence, we constructed a sampling framework where

hospitals were stratified based on their geographic region in Tennessee (West, Central, and East), and tertiles of reported discharge volume (Low, Medium, and High) during the study period (30-32). From this sampling framework, we randomly selected three hospitals from each of these nine sampling strata, and retrieved their medical records for review and validation (Figure 1). This strategy, relative to a purely random sample, ensured better representation of infections identified in smaller hospitals and those in more rural regions of the State of Tennessee. If a hospital refused to participate, it was replaced by another hospital randomly selected from the same sampling stratum.

The overall goal was to review and validate 675 hospitalizations for serious infection from 27 hospitals (25 hospitalizations for each of the 3 hospitals comprising a stratum, yielding 75 hospitalizations for each of the 9 strata) (Figure 1). We conservatively assumed that up to 80% of records requested would be available for review, and so we requested 32 records per hospital to receive an average of 25 records from each (Figure 1). To ensure that we reviewed sufficient rare infections, we preferentially selected any identified possible hospitalizations for meningitis/encephalitis, osteomyelitis/septic arthritis and endocarditis from each hospital in the sample. We randomly selected the remaining set of possible hospitalizations for other serious infections based on the proportional distribution of common infections at each hospital (pneumonia, bacteremia/sepsis, pyelonephritis and cellulitis) until 32 infections were identified. For hospitals with fewer than 32 infections during the study period, all infections were requested.

Abstraction of Medical Records

Relevant clinical information was abstracted from the medical record (transfer notes, emergency room summary, admission summary, physical/history, pharmacy information, laboratory, microbiology, and radiology information, and discharge summary) of each hospitalization with a primary discharge diagnosis code indicative of infection using a standardized and customized REDCap electronic data capture instrument hosted at Vanderbilt University (33). As we were interested in infections that led to hospitalizations, we focused our reviews on clinical, microbiological and radiological information from the 2 days prior to the admission date through 2 days after admission to limit the possibility of identifying

infections that developed during the hospitalization (i.e. nosocomial infections). In preparation for this study, the case report form was pilot-tested among a separate, convenience sample of 354 possible infections identified in the cohort from 3 hospitals in the same city as Vanderbilt University. This separate sample of hospitalizations was used only for pilot-testing the case report form, and was not included in the current study. One trained medical reviewer abstracted the relevant information for all selected records using the case report form. During the abstraction process, the lack of a particular finding in the medical record was treated as a lack of evidence for that finding, and so no information was considered missing after abstraction.

Adjudication of Medical Records

All records received were abstracted, reviewed and adjudicated. We made the final determination of whether a hospitalization represented a confirmed infection or not using *a priori* definitions of clinical, radiological, and/or microbiological findings compatible with infection for each infection type. Previous validation studies and expert clinical knowledge were used to define these specific *a priori* definitions for each infection type (*Supplementary appendix)* (25, 28, 34).

Statistical analysis

We calculated the PPV of the ICD-9-CM discharge diagnosis codes for identifying hospitalizations for serious infection using the results of the *a priori* definitions applied to the information abstracted from the medical records as the reference. Secondary analyses assessed the PPV for hospitalizations for serious infection across hospitals of different sizes and in different geographical regions of Tennessee.

We also assessed the reliability of the abstraction process. A second trained medical reviewer abstracted relevant information from a subset of selected records, which included all meningitis and endocarditis records, and a random selection of 10% of each of the remaining infection types. Each reviewer conducted the process independently and blinded from one another. For the subset of records abstracted by both reviewers, inter-reviewer agreement for the adjudication of a true or mis-identified hospitalization was assessed using the Gwet's first agreement coefficient (AC_1) (36-38). Since Cohen's

kappa statistic can be unreliable when the prevalence of the event and the level of observer agreement are high in the study sample, we used Gwet's AC_1 as a reliability measure unlikely to be affected by these concerns (38-40). In sensitivity analyses, we assessed the impact of excluding hospitalizations that occurred after the individual was transferred from another healthcare facility, as initial documentation and details of the infection could be missing or incomplete in the receiving hospital (34). We also assessed the impact on the PPV for all infections when requiring microbiological identification of a pathogen (excluding common contaminants) from a sterile site within 2 days before or after the hospitalization admission date. Among hospitalizations for possible pneumonia, we also assessed the PPV when radiological evidence of pneumonia was required [i.e. pneumonia, opacity, or infiltrate mentioned in a chest X-ray or computed tomography scan report] (*Supplementary appendix*). All analyses were performed in Stata-IC, version 15.1 (College Station TX).

RESULTS

Cohort characteristics

Among a retrospective cohort of 129,465 adults ≥50 years of age enrolled in TennCare, 8,322 hospitalizations for serious infection were identified during the study period (2008-2013). Pneumonia, cellulitis and bacteremia/sepsis were the most common infections (54.3%, 20.5% and 18.4%, respectively), followed by pyelonephritis (3.8%) and septic arthritis/osteomyelitis (2.5%). Fewer than 1% of hospitalizations were due to meningitis/encephalitis (n=30) and endocarditis (n=18). Cohort members were primarily female (57.8%) with a median age of 60 years and with residence outside of a nursing home (85.9%).

Collection, review and adjudication of selected medical records

Of the 27 hospitals that were initially selected for the sample, 21 (78%) were able to participate. We selected 7 additional hospitals to replace the 6 non-participants to achieve the desired sample size, including an additional small hospital in the East region due to a large number of unavailable records from a single participating hospital.

We received 716 (89%) of 808 requested records from 28 participating hospitals [Table 2]. Record availability from participating hospitals was lower in medium size hospitals (81.8%) compared to small (93.5%) and large hospitals (91.7%), but did not differ by geographic region. Record availability by infection type was greater than 86% for all infection types, with the exception of hospitalizations for the rare endocarditis cases (57.1%; only 4 of 7 cases).

There was evidence of transfer from a prior healthcare facility for 21.8% of the hospitalizations for serious infection [highest percentage of transfers for bacteremia/sepsis (38.5%) and pneumonia (25.1%)]. The most common healthcare facility source was a nursing home/skilled nursing facility (84.6%), but also included group home sources (7.7%), other sources (4.5%) [assisted living facility, mental health center] and another acute care hospital (3.2%). There was evidence of an emergency department visit within 7 days prior to admission date for the serious infection hospitalization in 4.8% of the records.

Performance of discharge diagnosis codes

A total of 646 [PPV: 90.2% (95% CI: 88.0-92.4)] of the hospitalizations for serious infection identified using ICD-9-CM primary discharge diagnosis codes were confirmed by applying the *a priori* definitions to the abstracted data. The PPV was highest for pneumonia and cellulitis [96.8% (95% CI: 94.5-98.4) and 91.1% (95% CI: 86.0-96.1), respectively], and was ≥75% for bacteremia/sepsis, pyelonephritis, septic arthritis/osteomyelitis, and endocarditis. The PPV was lowest for meningitis/encephalitis [50.0% (95% CI: 19.0-81.0)], although the precision was limited due to a low number of available records for review (Table 2).

When performance was evaluated across stratification sampling parameters, no apparent differences were observed in the PPV for records from hospitals in different geographical regions of Tennessee. Although the PPV was high for all three discharge volume groups, the PPV was significantly lower in large hospitals [84.6% (95% CI: 80.1-89.0)] compared to smaller hospitals [93.9% (95% CI: 90.8-97.0); PPV difference: -9.3% (95% CI: -14.7, -3.9)] and medium hospitals [92.7% (95% CI: 89.4-96.0); PPV difference: -8.1% (95% CI: -13.7, -2.6)] (Table 2). This was likely driven by the different

distributions in the types of infections selected for review in the hospital groups. Large hospitals had a higher proportion of non-pneumonia infections (70.4%) compared to medium and small hospitals (49.4% and 36.1%, respectively). Importantly, the PPV for pneumonia was similar in each discharge volume group (range: 96.0 to 96.6%), whereas the PPV was smaller for non-pneumonia infections in large hospitals (79.8%) compared to medium (88.7%) and small (89.2%) hospitals.

In the 82 records independently abstracted by two reviewers to assess reliability, there was 92.7% (95% CI: 86.9-98.4) agreement for identifying true hospitalizations for serious infection. The inter-rater agreement was also high when assessing reliability, independent of the outcome prevalence, with an AC_1 of 0.91 (95% CI: 0.84-0.99).

Sensitivity analyses

The PPV was virtually unchanged when excluding the 21.8% of hospitalizations that occurred as transfers from another healthcare facility [90.1% (95% CI: 87.7-92.6)]. Microbiological evidence of the specific infection type was found in 47.6% of records, leading to reduced PPVs when requiring microbiological evidence [45.4% (95% CI: 41.7-49.0)]. Microbiological evidence of infection was highest in hospitalizations for suspected pyelonephritis (94.4%), but was $\leq$60% for every other infection type [pneumonia (42.7%); cellulitis/soft tissue infections (58.5%); bacteremia/sepsis (26.1%)]. When requiring radiological confirmation of pneumonia, the PPV for coded diagnoses was 78.8% (95% CI: 74.5-83.2). Approximately 95.6% of possible hospitalizations for pneumonia had at least one documented chest x-ray or CT-scan. Among those patients with a chest x-ray or CT-scan report available (n=325), 83.4% had a finding compatible with pneumonia. The main findings among the 54 patients with possible pneumonia and a radiological report available, but without radiological confirmation of pneumonia included atelectasis (n=6), interstitial pneumonitis (n=3), chronic heart failure with pulmonary edema (n=1), and no radiological findings of any kind (n=44).

DISCUSSION

Discharge diagnoses for identifying hospitalizations due to serious infections among older adults had an overall positive predictive value of 90.2%, with highest values for identification of common serious infections. PPVs were consistently high across different hospital types and regions of Tennessee. Microbiological confirmation was available for fewer than 50% of those admitted with possible hospitalizations for serious infections, and as expected, such a requirement resulted in a lower PPV. Importantly, the PPV for pneumonia hospitalizations remained relatively high even when requiring radiological confirmation. In addition, including hospitalizations for serious infection that were the result of a transfer from another healthcare facility (e.g. acute care hospital, skilled nursing facility) did not change the PPV of hospitalizations for serious infection.

The PPV for hospitalizations for pneumonia in previous smaller validation studies has ranged from 72 to 86% in different healthcare systems, but those studies were not focused on older adults (25, 41-43). In our study of hospitalizations among older adults, we found that coded discharge diagnoses have a higher PPV for pneumonia compared to previous studies. The PPV for bacteremia/sepsis was also on the higher range of previously reported PPVs for diagnosis codes to identify bacteremia/sepsis from administrative data in other populations (reported range from 45% to 97.7%), and for septic arthritis/osteomyelitis compared to a previous study conducted among patients with diabetes (63.9% versus 75.9% in our study) (44-46). Overall, the observed PPV for all infections in our study was comparable to two previous comprehensive validation studies of bacterial infections, one among patients with rheumatoid arthritis in a single hospital system and another among patients in one of the Veteran's Affairs integrated service networks (28, 34). Compared to the these two previous studies of ICD-9 codes, we abstracted and adjudicated a larger number of records while using a more systematic sampling strategy to retrieve and review records for hospitalizations from multiple regions and hospital types as opposed to a single hospital or healthcare system. However, the PPVs for individual infections were less precise and less similar to these previous studies, especially for rare infections, as would be expected due to the low numbers of rare infections across previous studies (28, 34). The results of our study are also

similar to previous validation studies that used corresponding ICD-10 diagnosis codes to identify hospitalizations for serious infection (47, 48).

One limitation to consider in our study was that it was not designed to estimate the sensitivity and specificity of the coding algorithms. This would have required the identification, review and adjudication of a sample of hospitalizations that did not fulfill our algorithm (i.e. presence of the ICD-9 primary discharge diagnosis codes indicative of infection). However, when the prevalence of an outcome is low, the PPV approximates the specificity (49). Importantly, any non-differential outcome misclassification between exposure groups resulting from the use of imperfect but highly-specific measurements would attenuate the impact of the misclassification on the relative risk estimates (50). In addition, we found that the PPV of coded discharge diagnoses for serious infections remained high across hospitals of different sizes and across different geographical areas of Tennessee, which may have different prevalences of hospitalizations for serious infection (51). Although our study applied a systematic sampling strategy to assure the representation of different settings in our population, our population was restricted to older adults enrolled in a State Medicaid program. Therefore, caution is warranted when extrapolating the study findings to other populations.

Another limitation is the use of available clinical information to operationalize definitions for adjudication of true hospitalizations for infections. It is possible that some procedures, laboratory findings and diagnoses that informed the final diagnosis of infection were not fully recorded in the medical records, and thus, were not available for our review and may have contributed to the observed PPV for some infections. Although we used previous validation studies and clinical information to build pre-specified definitions for the adjudication of true infections, our reference criteria may be imperfect, considering the retrospective nature of our determinations and potential variability in clinical practice. Nevertheless, we also assessed how the availability of selected findings (i.e. microbiological and radiological information) in the medical record impacted the overall and infection-specific PPV. We demonstrated that relying on highly specific clinical diagnostics, such as microbiological and radiological information, to confirm true infections would result in lower PPVs for identification of infections in

administrative data. Requiring microbiological confirmation to confirm true infections is challenging because of the known low sensitivity of culture-based diagnostic methods (most commonly used in clinical practice), which may lead to misclassification (52, 53). In addition, requiring radiological evidence compatible with pneumonia within 2 days of hospital admission did lower the observed PPV for pneumonia hospitalizations. Nevertheless, the observed PPV remained close to 80%, which should reduce concerns about using diagnosis codes to identify hospitalizations due to pneumonia. Finally, the coding algorithms were based on the ICD-9-coding system only. Although these findings will be helpful for retrospective studies that encompass periods of ICD-9 use, additional studies evaluating the performance of ICD-10-based codes would be useful to complement our findings.

Our study demonstrated that discharge diagnosis codes can be used to accurately identify hospitalizations for serious infections among older adults. The highest PPVs were observed for the most common infections, and the PPV for pneumonia remained high when requiring radiological confirmation. The PPV was poor when microbiological confirmation of infection was required to identify a true hospitalization for serious infection. This information supports the use of discharge diagnosis codes for infections as outcomes in ongoing and future studies among older adults.

Acknowledgement

We are indebted to the Tennessee Bureau of TennCare of the Department of Finance and Administration, which provided data for the study. We are also indebted to the Tennessee Department of Health for providing data for the study. Statements in the report should not be construed as endorsement by the U.S. Department of Health and Human Services, the Department of Veterans Affairs, or the Tennessee Department of Health.

Funding

This study was funded by the NIH (R03-AG-042981 and R01-AG-043471-01A1) and the TL1 award TL1TR000447.

Contributors

ADW planned the medical record collection and statistical analysis, analyzed and interpreted the data, and drafted and revised the paper. MRG, WS, CMS, and RAG planned the statistical analysis, interpreted the data and revised the paper. EFM prepared the data, and revised the paper. CGG initiated the project, acquired the data from TennCare, planned the medical record collection and statistical analysis, interpreted the data, and revised the paper.

Declaration of interests

CGG has received consulting fees from Pfizer and Merck, and received research support from Sanofi-Pasteur, Campbell Alliance, the Centers for Disease Control and Prevention, National Institutes of Health, and the Agency for Health Care Research and Quality. WS has received personal fees from Pfizer, Merck, Novavax, Dynavax, Sanofi-Pasteur, GSK, Seqirus, and received research support from the Centers for Disease Control and Prevention. ADW, EFM, CMS, MRG and RAG have no conflicts of interest to disclose.

Data sharing

No additional unpublished data are available from the study. The study protocol and statistical code are available from the corresponding author, Andrew Wiese (andrew.d.wiese@vanderbilt.edu).

REFERENCES

1. Khabbaz RF, Moseley RR, Steiner RJ, Levitt AM, Bell BP. Challenges of infectious diseases in the USA. *Lancet* 2014; **384**(9937): 53-63.

2. Grohskopf LA, Sokolow LZ, Broder KR, et al. Prevention and Control of Seasonal Influenza with Vaccines. *MMWR Recommendations and reports : Morbidity and mortality weekly report Recommendations and reports / Centers for Disease Control* 2016; **65**(5): 1-54.

3. Kim DK, Bridges CB, Harriman KH. Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older--United States, 2015. *MMWR Morbidity and mortality weekly report* 2015; **64**(4): 91-2.

4. Jeon CY, Muennig P, Furuya EY, Cohen B, Nash D, Larson EL. Burden of present-on-admission infections and health care-associated infections, by race and ethnicity. *American journal of infection control* 2014; **42**(12): 1296-302.

5. Pilishvili T, Bennett NM. Pneumococcal Disease Prevention Among Adults: Strategies for the Use of Pneumococcal Vaccines. *American journal of preventive medicine* 2015; **49**(6 Suppl 4): S383-90.

6. Sjoding MW, Prescott HC, Wunsch H, Iwashyna TJ, Cooke CR. Longitudinal Changes in ICU Admissions Among Elderly Patients in the United States. *Critical care medicine* 2016; **44**(7): 1353-60.

7. Crotty MP, Meyers S, Hampton N, et al. Epidemiology, Co-Infections, and Outcomes of Viral Pneumonia in Adults: An Observational Cohort Study. *Medicine* 2015; **94**(50): e2332.

8. Simonetti AF, Garcia-Vidal C, Viasus D, et al. Declining mortality among hospitalized patients with community-acquired pneumonia. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2016; **22**(6): 567.e1-7.

9. Rivero-Calle I, Pardo-Seco J, Aldaz P, et al. Incidence and risk factor prevalence of community-acquired pneumonia in adults in primary care in Spain (NEUMO-ES-RISK project). *BMC infectious diseases* 2016; **16**(1): 645.

10. Park JY, Park S, Lee SH, et al. Microorganisms Causing Community-Acquired Acute Bronchitis: The Role of Bacterial Infection. *PloS one* 2016; **11**(10): e0165553.

11. Jackson ML, Walker R, Lee S, Larson E, Dublin S. Predicting 2-Year Risk of Developing Pneumonia in Older Adults without Dementia. *Journal of the American Geriatrics Society* 2016; **64**(7): 1439-47.

12. Clement RC, Haddix KP, Creighton RA, Spang JT, Tennant JN, Kamath GV. Risk Factors for Infection After Knee Arthroscopy: Analysis of 595,083 Cases From 3 United States Databases. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 2016; **32**(12): 2556-61.

13. Weinert BA, Edmonson MB. Hospitalizations at Nonfederal Facilities for Lower Respiratory Tract Infection in American Indian and Alaska Native Children Younger than 5 Years of Age, 1997-2012. *The Journal of pediatrics* 2016; **175**: 33-9.e4.

14. Akiyama T, Chikuda H, Yasunaga H, Horiguchi H, Fushimi K, Saita K. Incidence and risk factors for mortality of vertebral osteomyelitis: a retrospective analysis using the Japanese diagnosis procedure combination database. *BMJ open* 2013; **3**(3).

15. Graversen ME, Dalgaard LS, Jensen-Fangel S, Jespersen B, Ostergaard L, Sogaard OS. Risk and outcome of pyelonephritis among renal transplant recipients. *BMC infectious diseases* 2016; **16**: 264.

16. Abou Zahr Z, Spiegelman A, Cantu M, Ng B. Perioperative use of anti-rheumatic agents does not increase early postoperative infection risks: a Veteran Affairs' administrative database study. *Rheumatology international* 2015; **35**(2): 265-72.

17. Carnahan RM, Moores KG, Perencevich EN. A systematic review of validated methods for identifying infection related to blood products, tissue grafts, or organ transplants using administrative data. *Pharmacoepidemiology and drug safety* 2012; **21 Suppl 1**: 213-21.

18. Condell O, Gubbels S, Nielsen J, et al. Automated surveillance system for hospital-acquired urinary tract infections in Denmark. *The Journal of hospital infection* 2016; **93**(3): 290-6.

19. Gedeborg R, Furebring M, Michaelsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients. *Journal of clinical epidemiology* 2007; **60**(2): 155-62.

20. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of administrative code data for the surveillance of healthcare-associated infections: a systematic review and meta-analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2014; **58**(5): 688-96.

21. Shaklee J, Zerr DM, Elward A, et al. Improving surveillance for pediatric Clostridium difficile infection: derivation and validation of an accurate case-finding tool. *The Pediatric infectious disease journal* 2011; **30**(3): e38-40.

22. Wright SB, Huskins WC, Dokholyan RS, Goldmann DA, Platt R. Administrative databases provide inaccurate data for surveillance of long-term central venous catheter-associated infections. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America* 2003; **24**(12): 946-9.

23. Greenberg JA, Hohmann SF, Hall JB, Kress JP, David MZ. Validation of a Method to Identify Immunocompromised Patients with Severe Sepsis in Administrative Databases. *Annals of the American Thoracic Society* 2016; **13**(2): 253-8.

24. Olsen MA, Young-Xu Y, Stwalley D, et al. The burden of clostridium difficile infection: estimates of the incidence of CDI from U.S. Administrative databases. *BMC infectious diseases* 2016; **16**: 177.

25. Grijalva CG, Chung CP, Stein CM, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. *Pharmacoepidemiology and drug safety* 2008; **17**(9): 890-5.

26. TennCare Fiscal Year 2015-2016 Annual Report. In: Administration DoHCF, ed.2016.

267. Wiese AD, Griffin MR, Stein CM, Mitchel EF, Jr., Grijalva CG. Opioid Analgesics and the Risk of Serious Infections Among Patients With Rheumatoid Arthritis: A Self-Controlled Case Series Study. *Arthritis & rheumatology (Hoboken, NJ)* 2016; **68**(2): 323-31.

28. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. *Journal of clinical epidemiology* 2007; **60**(4): 397-409.

29. Grijalva CG, Kaltenbach L, Arbogast PG, Mitchel EF, Jr., Griffin MR. Initiation of rheumatoid arthritis treatments and the risk of serious infections. *Rheumatology (Oxford, England)* 2010; **49**(1): 82-90.

30. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Statistics in medicine* 1997; **16**(9): 981-91.

31. Kim HM, Smith EG, Stano CM, et al. Validation of key behaviourally based mental health diagnoses in administrative data: suicide attempt, alcohol abuse, illicit drug abuse and tobacco use. *BMC health services research* 2012; **12**: 18.

32. El-Ghitany EM, Farghaly AG, Farag S, Abd El-Wahab EW. Validation of EGCRISC for Chronic Hepatitis C Infection Screening and Risk Assessment in the Egyptian Population. *PloS one* 2016; **11**(12): e0168649.

33. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of biomedical informatics* 2009; **42**(2): 377-81.

34. Patkar NM, Curtis JR, Teng GG, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. *Journal of clinical epidemiology* 2009; **62**(3): 321-7, 7.e1-7.

35. Aronson PL, Williams DJ, Thurm C, et al. Accuracy of diagnosis codes to identify febrile young infants using administrative data. *Journal of hospital medicine : an official publication of the Society of Hospital Medicine* 2015; **10**(12): 787-93.

36. Govatsmark RE, Sneeggen S, Karlsaune H, Slordahl SA, Bonaa KH. Interrater reliability of a national acute myocardial infarction register. *Clinical epidemiology* 2016; **8**: 305-12.

37. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology* 2013; **13**: 61.

38. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *The British journal of mathematical and statistical psychology* 2008; **61**(Pt 1): 29-48.

39. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. *Journal of clinical epidemiology* 1996; **49**(4): 431-4.

40. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology* 1990; **43**(6): 543-9.

41. Drahos J, Vanwormer JJ, Greenlee RT, Landgren O, Koshiol J. Accuracy of ICD-9-CM codes in identifying infections of pneumonia and herpes simplex virus in administrative data. *Annals of epidemiology* 2013; **23**(5): 291-3.

42. Aronsky D, Haug PJ, Lagor C, Dean NC. Accuracy of administrative data for identifying patients with pneumonia. *American journal of medical quality : the official journal of the American College of Medical Quality* 2005; **20**(6): 319-28.

43. Kern DM, Davis J, Williams SA, et al. Validation of an administrative claims-based diagnostic code for pneumonia in a US-based commercially insured COPD population. *International journal of chronic obstructive pulmonary disease* 2015; **10**: 1417-25.

44. Iwashyna TJ, Odden A, Rohde J, et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. *Medical care* 2014; **52**(6): e39-43.

45. Newton KM, Wagner EH, Ramsey SD, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. *Journal of clinical epidemiology* 1999; **52**(3): 199-207.

46. Carnahan RM, Herman RA, Moores KG. A systematic review of validated methods for identifying transfusion-related sepsis using administrative and claims data. *Pharmacoepidemiology and drug safety* 2012; **21 Suppl 1**: 222-9.

47. Sahli L, Lapeyre-Mestre M, Derumeaux H, Moulis G. Positive predictive values of selected hospital discharge diagnoses to identify infections responsible for hospitalization in the French national hospital database. Pharmacoepidemiology and drug safety 2016; 25(7): 785-9.

48. Holland-Bill L, Xu H, Sorensen HT, et al. Positive predictive value of primary inpatient discharge diagnoses of infection among cancer patients in the Danish National Registry of Patients. Annals of epidemiology 2014; 24(8): 593-7, 7.e1-18.

49. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *Journal of clinical epidemiology* 2005; **58**(4): 323-37.

50. Rodgers A, MacMahon S. Systematic underestimation of treatment effects as a result of diagnostic test inaccuracy: implications for the interpretation and design of thromboprophylaxis trials. *Thrombosis and haemostasis* 1995; **73**(2): 167-71.

51. van Walraven C, English S, Austin PC. Administrative database code accuracy did not vary notably with changes in disease prevalence. *Journal of clinical epidemiology* 2016; **79**: 86-9.

52. Mancini N, Carletti S, Ghidoli N, Cichero P, Burioni R, Clementi M. The era of molecular and other non-culture-based methods in diagnosis of sepsis. *Clinical microbiology reviews* 2010; **23**(1): 235-51.

53. Niederman MS. The argument against using quantitative cultures in clinical trials and for the management of ventilator-associated pneumonia. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2010; **51 Suppl 1**: S93-9.

Table 1. Discharge diagnosis code definitions (ICD-9-CM) for hospitalizations for serious infection

Serious Infection	Primary (first listed) discharge diagnosis code
Pneumonia-primary definition	003.22, 480.*‡, 481, 482.*, 483.*, 484.*, 485.*, 486.*, 487.0
Pneumonia-secondary definition (pneumonia diagnosis (above) in any other diagnosis field)	510.*, 038.*, 790.7, 995.91, 995.92
Meningitis/ Encephalitis	003.21, 036.0, 0.47*, 049.*, 053.0, 054.72, 072.1, 091.81, 094.2, 098.82, 100.81, 320.*, 036.1, 054.3, 056.01, 058.21, 058.29, 062.*, 063.*, 064.*, 066.41, 072.2, 094.81, 130.0, 323.*
Bacteremia/ Sepsis†	038.*, 790.7, 995.91, 995.92
Cellulitis/ Soft-tissue infections	035, 040.0, 569.61, 681.*, 682.*, 728.86, 785.4
Endocarditis	036.42, 074.22, 093.2*, 098.84, 421.*, 422.92
Pyelonephritis	590.*
Septic Arthritis/ Osteomyelitis	003.23, 056.71, 098.5*, 711.0, 711.00-711.07, 711.09, 711.9*, 003.24, 376.03, 526.4, 730.0*, 730.1*, 730.2*

† Without a diagnosis of pneumonia in any other diagnosis field

‡ A * indicates all numeric values [0-9]

Table 2. Positive predictive value (PPV) of coded discharge diagnosis definitions for hospitalizations for serious infections among older adults enrolled in Tennessee Medicaid, 2008-2013

Type	Expected Number of Records	Records Received	PPV (95 % CI)
Overall	675	716	90.2 (88.0, 92.4)
Region Specific			
West	225	195	91.3 (87.3, 95.2)
Central	225	225	88.9 (84.8, 93.0)
East	225	296	90.5 (87.2, 93.9)
Bed volume size specific			
Low	225	230	93.9 (90.8, 97.0)
Medium	225	233	92.7 (89.4, 96.0)
High	225	253	84.6 (80.1, 89.0)
Serious Infection			
Pneumonia	305	340	96.8 (94.5, 98.4)
Cellulitis/Soft-tissue infections	125	123	91.1 (86.0, 96.1)
Pyelonephritis	80	89	87.6 (80.8, 94.5)
Bacteremia/Sepsis	100	92	82.6 (74.9, 90.4)
Septic Arthritis/Osteomyelitis	50	58	75.9 (64.8, 86.9)
Meningitis/Encephalitis	10	10	50.0 (19.0, 81.0)
Endocarditis	5	4	75.0 (32.6, 100.0)

Figure legends

Figure 1. Sampling strategy for identifying potential hospitalizations for serious infection

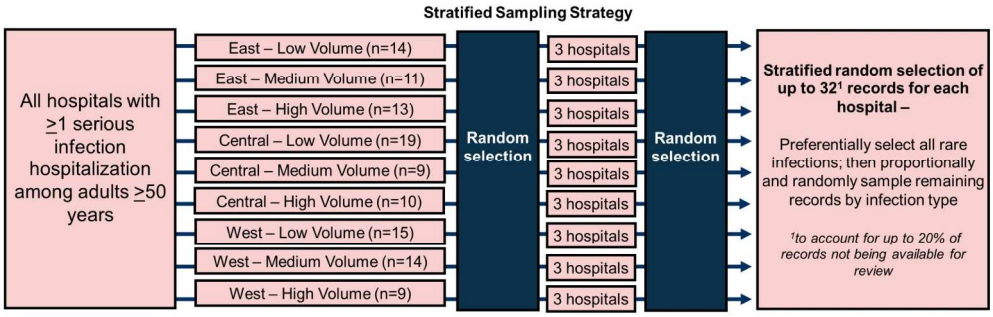

Figure 1. Sampling strategy for identifying potential hospitalizations for serious infection

271x90mm (300 x 300 DPI)

Supplementary Appendix

Infection-Specific Definitions of Hospitalization for Serious Infection

We used a pre-specified adjudication process to determine whether each abstracted medical record corresponded to a true infection or not. Previous validation studies and expert clinical knowledge were used to define specific a priori definitions for each infection type.[1-3] Information abstracted from the medical record was compared to these a priori definitions for each infection type to make the final determination of whether a hospitalization represented a true infection or not.

Outline

I. Sepsis/Septicemia/Bacteremia/Septic Shock/Generalized Infection

Either of the following [1 or 2]:

1. **Positive culture of a non-contaminant pathogen**
    - i. Positive blood culture [any of the following (1-2)]
        1. Any gram-negative organism, except:
            a. No predominant organism
        2. A gram positive organism, except:
            a. Coagulase-negative *Staphylococcus*
            b. *Bacillus spp.* (other than *Bacillus anthracis*)
            c. *Corynebacterium spp.*
            d. *Propionibacterium spp.*
            e. *Micrococcus*
            f. Diptheroids
            g. Viridians Group Streptococci
            h. Enterococci
            i. *Clostridium perfringens*
            j. *Aerococcus*
            k. *Alcaligenes faecalis*
            l. *Citrobacter*
            m. *Neisseria subflava*
            n. *Stomatococcus*
            o. *Streptococcus bovis*
            p. *Veillonella candidemia*
            q. *Mycobacterium tuberculosis*
            r. *S. salivarius*
            s. *"Gram Positive"*
            t. *"No predominant organism"*
            u. *Streptococcus alpha*
2. **At least two of the following, documented at admission +/- 2 days [i-iii]**
    - i. Hypotension
        1. Systolic BP $\leq$ 90 mmHg
        2. Reduction of systolic BP of 40mmHg from earliest measurement collected during the admission of interest
    - ii. Two of the following [1-4]:
        1. Temperature $\geq$ 38^{0}C **or** $\leq$ 36^{0}C
        2. Heart rate $\geq$ 90 beats/minute
        3. Respiratory rate $\geq$ 20 breaths/min or $PaCO_2$ < 32 mmHg
        4. WBC $\geq$ 10,000 cells/mm^3 **or** $\leq$ 4,500 cells/mm^3 **or** WBC with > 10 % immature (band) forms
    - iii. Initiation of antibiotic treatment specifically for sepsis/septicemia/bacteremia/septic shock/generalized infection

II. Pneumonia

1. Pneumonia identified through examination (**all three of the following [a-c]):**
    a. One of the following admission findings indicative of respiratory findings:
        1. New and/or increased cough
        2. Shortness of breath
        3. Pleuritic chest pain
        4. New purulent production
        5. Altered mental status ("agitation" and "lethargy" included)
        6. Crackles
            a. Physical evidence of consolidation such as egophony, whispered pectoriloquy, etc.
    b. One of the following examination findings indicative of systemic infection [1-4]:
        1. Temperature ($T \geq 100.4^0F$ (38^0C) or $\leq 96^0F$) in first 48 hours of admission
        2. Systolic BP $\leq$ 90mmHg
        3. Shock
            a. Volume nonresponsive hypotension
        4. Blood peripheral WBC ($\geq 10.0 \times 10^9/L$ or $\leq 4.5 \times 10^9/L$)
    c. Treatment with antibiotics/antivirals indicated for suspected infection

OR

At least two of the following [1-3]:

1. Two of the following from #1 **([a and b], [a and c], or [b-c])**
2. Any of the following findings listed on chest imaging from radiologic report **documented at admission +/- 2 days**
    a. Pneumonia
    b. Lung abscess
    c. Opacity consistent with pneumonia/lung abscess
    d. Infiltrate consistent with pneumonia/lung abscess
    e. Consolidation consistent with pneumonia/lung abscess
    f. Increased density consistent with pneumonia/lung abscess
    g. Pleural effusion consistent with pneumonia/lung abscess
    h. Interstitial edema consistent with pneumonia/lung abscess
3. Sterile Site Laboratory Findings
    i. Any one of the following [i through v]
        i. Sputum lab findings [any **one** of the following (1, 2)]:
            1. Sputum culture/PCR/serology/gram stain positive for an agent that is not considered a contaminant [see exclusion list below]:
                a. *Aspergillus* species, *Enterococcus* species, viridians group streptococci, and yeast
            2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen
        ii. Blood lab findings [either of the following (1-3)]
            1. Blood culture/PCR/serology positive for an agent that is not considered a contaminant [see exclusion list below]:
                a. Exclusions
                    i. Coagulase-negative *Staphylococcus*

ii. *Bacillus spp.* (other than *Bacillus anthracis*)
iii. *Corynebacterium spp.*
iv. *Propionibacterium spp.*
v. *Micrococcus*
vi. Diptheroids
vii. Viridians Group Streptococci
viii. Enterococci
ix. *Clostridium perfringens*
x. *Aerococcus*
xi. *Alcaligenes faecalis*
xii. *Citrobacter*
xiii. *Neisseria subflava*
xiv. *Stomatococcus*
xv. *Streptococcus bovis*
xvi. *Veillonella candidemia*
xvii. *Mycobacterium tuberculosis*
xviii. *S. salivarius*

2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen

iii. Pleural fluid lab findings [either of the following (1, 2)]
1. Culture/PCR/serology positive for a bacterial pathogen
2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen

iv. Bronchoscopic specimen or deep endotracheal tube aspiration lab findings [either of the following (1, 2)]
1. Culture/PCR/serology positive for a bacterial pathogen
2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen

v. Urine antigen detection testing [either of the following (1, 2)]
1. *Legionella pneumophila*
2. *Streptococcus pneumoniae*

III. Cellulitis/Soft-Tissue Infection

Both of the following:

1. Any mention of the following with recent onset (≤14 days) ***[any of the following]***
    a. Skin erythema
    b. Surgical site infection
    c. Superficial central line infection
    d. Ostomy site infection
    e. Skin infection with associated lymphangitis
2. Antibiotic treatment initiated for suspected infection

IV. Endocarditis

Any one of the following [1-3]:

1. Major Criteria [both of the following]:
    a. Suggestive microbiology [at least one of the following]:
        i. Positive blood culture of an ***endocarditis organism*** **[any of the following]:**
            1. *Streptococcus bovis*
            2. *Viridians streptococci*
            3. *Staphylococcus aureus*
            4. *Enterococcus spp.*
            5. HACEK organisms
            6. Coagulase negative staphylococci
    b. Evidence of endocardial involvement [at least one of the following]:
        i. New regurgiant murmur (a change in a preexisting murmur does not get scored)
        ii. Echocardiogram suspicious for any of the following:
            1. Intracardiac mass with no alternative explanation
            2. Endocardial abscess
            3. New partial prosthesis dehiscence
            4. Vegetation on valve
2. Minor Criteria [at least 4 of the following]:
    a. Predisposing valvular disease or IV drug use
    b. Temperature $\geq$ 100.4°F or 38°C
    c. Vascular phenomena
        i. Janeway lesions, conjunctival hemorrhages, arterial emboli, septic pulmonary infarcts, mycotic aneurysm, intracranial bleed
    d. Immunologic phenomena
        i. Osler nodes, Roth Spots, elevated Rheumatoid factor, hematuria in non-catheter urine, or other evidence of glomerulonephritis
    e. Positive blood cultures
        i. Excluding a single positive culture for coagulase negative staphylococci or a single positive culture for an organism that does not fall into the "reasonable endocarditis organism" (i.e. coagulase-positive and coagulase-negative *S. aureus*, Enterococcus, viridians group Streptococci, *S. bovis*, HACEK organisms)
    f. Positive serology for Brucella, Bartonella, Legionella, or Chlamydia
    g. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
3. At least one Major Criteria **AND** 3 minor criteria.

V. Meningitis/Encephalitis

Any one of the following [1 or 2]:

1. Both of the following [a-b]
    a. Laboratory Findings [any one of the following (i-ix)]
        i. CSF demonstrates any bacterium
            1. Excluding Diptheroids, Propionibacteria, Bacillus, Coagulase Negative *Staphylococcus*
        ii. CSF demonstrates Diptheroids, Propionibacteria, Bacillus, Coagulase Negative *Staphylococcus* in the setting of past neurosurgical intervention **AND** physicians elected to treat with antibacterials
        iii. Blood cultures positive for any of the following:
            1. *S. pneumoniae*
            2. *H. influenza*
            3. *Neisseria meningitidis*
            4. Group B Streptococcus
        iv. Stool cultures positive for enterovirus
        v. Throat or sputum cultures positive for *Neisseria meningitidis* in the setting of a rapid onset, overwhelming infection syndrome, including petechiae
        vi. Serology positive for *Mycoplasma, Leptospira,* measles, mumps, lymphocytic choriomeningitis virus, arboviruses (e.g. St. Louis encephalitis virus), or HIV (if historically consistent with acute seroconversion).
        vii. Brain biopsy demonstrates encephalitis
        viii. Positive CSF culture or PCR detection for any of the following
        ix. Acute or convalescent serology demonstrates positive antibody pattern for any of the following:
            1. Encephalitis arbovirus (La Crosse, St. louis, Eastern Equine, Western Equine, Powassan, Japanese, West Nile)
    b. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected meningitis/encephalitis

2. At least two of the following [a-d]
    a. Clinical meningitis/encephalitis [at least two of the following]:
        i. Petechial rash
        ii. Nuchal rigidity (by history or exam)
        iii. Altered sensorium
        iv. Fever
        v. Altered level of consciousness, including “agitation” or “lethargy”
        vi. Behavioral change
        vii. Diminished level of consciousness (not easily roused)
        viii. History of any of the following: headaches, altered mental status, or recent exposure to patient with known bacterial meningitis
        ix. Reduction in fever within 72 hours of starting anti-bacterial
    b. Inflammatory CSF [at least one of the following i-ii]
        i. Pleocytosis: $\geq$ 15 WBC/mm^3 (after subtracting one WBC for every 1,000 RBC)
        ii. Elevated protein (based on local lab-determined upper limits)
    c. Suggestive Findings [at least one of the following (i-iv)

i. Septic syndrome
ii. Focal neurological deficits documented during examination (such as flaccid paralysis or speech alterations for West Nile Virus)
iii. Abnormal imaging
    1. Computed tomography or magnetic resonance imaging (MRI) demonstrating focal edema or inflammation or hemorrhage
    2. Indicated as “meningitis/encephalitis” or “compatible with meningitis/encephalitis” or “cannot rule out meningitis/encephalitis”
iv. Findings indicating an abnormal electroencephalography (such as focal periodic discharges)

d. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for presumed meningitis/encephalitis

VI. Pyelonephritis

At least two of the following [1-4]:

1. Suggestion of infection [at least one of the following]:
    a. Temperature ≥ 100.4°F (38°C)
    b. Peripheral blood WBC ≥ 10,000/mm^3
    c. Positive blood culture for any of the following:
        i. Gram Negative Rods
        ii. *Enterococcus spp.*
        iii. *Staphylococcus saprophyticus*
    d. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
2. Strong renal localization [at least one of the following]:
    a. CT, MRI, or Ultrasound Suggestive of Renal Inflammation
3. Minor Criteria [at least two of the following]:
    a. Flank pain
    b. Costovertebral angle tenderness
    c. Complaints of dysuria, frequency, or suprapubic pain
    d. Any pyuria
    e. Urine culture positive for a single organism
4. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected pyelonephritis

VII. Septic Arthritis/Osteomyelitis

Any one of the following (1-5):

1. Synovial fluid gram stain or tissue gram stain or special stain demonstrating any organism
2. Joint culture/PCR/serology positive for any organism
3. At least two of the following (a-d):
    a. Positive blood culture/PCR/serology
    b. Joint with acute (≤ 7 days) worsening of inflammatory features (**at least two of the following):**
        i. Pain on history
        ii. ROM
        iii. Warmth
        iv. Effusion
        v. Swelling
        vi. Limited range of motion
    c. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
    d. Any one of the following (i-iv)
        i. Synovial fluid WBC ≥ 30,000/mm^3
        ii. Synovial fluid WBC ≥ 60,000/mm^3 with > 75% PMNs
        iii. Skin lesions, tenosynovitis, or urethral/cervical/rectal Gram stain or culture suggestive of *Neisseria gonorrhoeae*
        iv. Any indication of the following in the synovial fluid: needle-like crystals, CPPD crystals, uric acid.
4. Positive bone biopsy [at least one of the following (a-c)]:
    a. Positive culture for any organism
    b. Positive gram stain
5. Imaging and indirect features **[at least two of the following (a-c)]:**
    a. Consistent imaging [at least one of the following (i-iv)]:
        i. Plain X-ray read by a radiologist as suggestive of osteomyelitis
        ii. CT Scan read by a radiologist as suggestive of osteomyelitis
        iii. MRI read by a radiologist as suggestive of osteomyelitis
        iv. Bone scan or WBC scan read as suggestive of osteomyelitis
    b. Suggestive indirect features[at least one of the following (i-viii)]:
        i. Temperature > 100.4°F (38°C)
        ii. Bony pain or tenderness or erythema over bone suspected to be infected
        iii. Draining soft tissue sinus over bone suspected to be infected
        iv. Positive "probe to bone" (or visible bone in deep ulcer at suspected site)
        v. Blood culture positive for *S. aureus*
        vi. ESR ≥ 75 mm/hour
        vii. Intravenous drug use or indwelling catheter
        viii. Inflammation on imaging associated with an orthopedic prosthesis
    c. Positive culture for any organism form wound sample over the bone suspected of infection
    d. Antibiotic/antiviral/antifungal treatment for suspected infection

VIII. References

1. Grijalva CG, Chung CP, Stein CM, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. Pharmacoepidemiology and drug safety 2008; 17(9): 890-5.
2. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. Journal of clinical epidemiology 2007; 60(4): 397-409.
3. Patkar NM, Curtis JR, Teng GG, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. *Journal of clinical epidemiology* 2009; **62**(3): 321-7, 7.e1-7.

Section & Topic	No	Item	Reported on page #
Title or Abstract			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4,5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4,5
	9	Whether participants formed a consecutive, random or convenience series	4-6
Test methods	**10a**	Index test, in sufficient detail to allow replication	5, Table 1, Supplementary Appendix
	10b	Reference standard, in sufficient detail to allow replication	6, Supplementary Appendix
	11	Rationale for choosing the reference standard (if alternatives exist)	6, Supplementary Appendix
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6, Supplementary Appendix
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6, Supplementary Appendix
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6, Supplementary Appendix
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7,8
	15	How indeterminate index test or reference standard results were handled	7,8
	16	How missing data on the index test and reference standard were handled	7,8
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	7,8
	18	Intended sample size and how it was determined	7,8
RESULTS			
Participants	**19**	Flow of participants, using a diagram	8
	20	Baseline demographic and clinical characteristics of participants	8
	21a	Distribution of severity of disease in those with the target condition	n/a
	21b	Distribution of alternative diagnoses in those without the target condition	n/a
	22	Time interval and any clinical interventions between index test and reference standard	n/a
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	9,10, Table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	9,10 Table 2
	25	Any adverse events from performing the index test or the reference standard	n/a
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	11-13
	27	Implications for practice, including the intended use and clinical role of the index test	11-13

OTHER INFORMATION			
	28	Registration number and name of registry	n/a
	29	Where the full study protocol can be accessed	14
	30	Sources of funding and other support; role of funders	13

BMJ Open

Validation of discharge diagnosis codes to identify serious infections among middle age and older adults

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020857.R1
Article Type:	Research
Date Submitted by the Author:	14-Mar-2018
Complete List of Authors:	Wiese, Andrew; Vanderbilt University Medical Center, Health Policy Griffin, Marie R; Vanderbilt University Medical Center, Health Policy Stein, Michael; Vanderbilt University, Pharmacology Schaffner, William; Vanderbilt University Medical Center, Health Policy Greevy, Robert; Vanderbilt School of Medicine, Biostatistics Mitchel, Jr., Edward; Vanderbilt University Medical Center, Health Policy Grijalva, Carlos; Vanderbilt University, Health Policy
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Infectious diseases, Epidemiology
Keywords:	coding algorithms, Medicaid, older adults, serious infections

Validation of discharge diagnosis codes to identify serious infections among middle age and older adults

Running title: Validation of diagnosis codes to identify infections

Authors: Andrew D. Wiese, PhD, MPH[1]; Marie R. Griffin[1,2], MD, MPH; C. Michael Stein, MB, ChB[3]; William Schaffner, MD[1]; Robert Greevy, PhD[4]; Edward F. Mitchel Jr., MS[1]; Carlos G. Grijalva, MD, MPH[1,2]

Affiliations: [1]Department of Health Policy, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; [2]Mid-South Geriatric Research Education and Clinical Center, VA Tennessee Valley Health Care System, Nashville, Tennessee, USA; [3]Departments of Pharmacology and [4]Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Corresponding Author: Andrew D. Wiese, PhD, MPH; Department of Health Policy, Vanderbilt University Medical Center, Suite 2600, Village at Vanderbilt, 1500 21st Avenue South, Nashville, TN 37212; phone: (615) 875-7997; email: andrew.d.wiese@vanderbilt.edu

Key words: coding algorithms; Medicaid; older adults; serious infections

Word Count: 3,748/4,000

Tables and Figures: (4/5)

ABSTRACT (277/300)

Objectives: Hospitalizations for serious infections are common among middle age and older adults and frequently used as study outcomes. Yet few studies have evaluated the performance of diagnosis codes to identify serious infections in this population. We sought to determine the positive predictive value (PPV) of diagnosis codes for identifying hospitalizations due to serious infections among middle age and older adults.

Setting and participants: We identified hospitalizations for possible infection among adults ≥50 years enrolled in the Tennessee Medicaid healthcare program (2008-2012) using ICD-9 diagnosis codes for pneumonia, meningitis/encephalitis, bacteremia/sepsis, cellulitis/soft-tissue infections, endocarditis, pyelonephritis and septic arthritis/osteomyelitis.

Design: Medical records were systematically obtained from hospitals randomly selected from a stratified sampling framework based on geographical region and hospital discharge volume.

Measures: Two trained clinical reviewers used a standardized extraction form to abstract information from medical records. Pre-defined algorithms served as reference to adjudicate confirmed infection-specific hospitalizations. We calculated the PPV of diagnosis codes using confirmed hospitalizations as reference. Sensitivity analyses determined the robustness of the PPV to definitions that required radiological or microbiological confirmation. We also determined interrater reliability between reviewers.

Results: The PPV of diagnosis codes for hospitalizations for infection (n=716) was 90% (95% CI: 88-92). The PPV was highest for pneumonia [97% (95% CI: 94-98)] and cellulitis [91% (95% CI: 85-95)], and lowest for meningitis/encephalitis [50% (95% CI: 24-76)]. The adjudication reliability was excellent [93% agreement; first agreement-coefficient: 0.91]. The overall PPV was lower when requiring microbiological confirmation [45%] and when requiring radiological confirmation for pneumonia [79%].

Conclusions: Discharge diagnosis codes have a high PPV for identifying hospitalizations for common, serious infections among middle age and older adults. PPV estimates for rare infections were imprecise.

STRENGTHS AND LIMITATIONS OF THE STUDY

- This study examined the performance of diagnosis coding algorithms to identify hospitalizations due to serious infections among middle age and older adults enrolled in a State Medicaid program using a systematic and representative sample of records from hospitals of different sizes and in distinct State regions.
- The reference criteria to identify true infections was based on previous literature and clinical expertise but may be imperfect. Nevertheless, identifying microbiologically-confirmed infections is difficult due to the low sensitivity of culture-based diagnostic methods often used in clinical practice.
- Diagnosis codes were based on the ICD-9-coding system only. These findings will continue to be helpful for retrospective studies that encompass periods of ICD-9 use, yet additional studies evaluating the performance of ICD-10-based codes would be beneficial.
- Our coding algorithms to identify serious infections had a high positive predictive value overall, and will be useful in ongoing and future research using administrative data

INTRODUCTION

Infectious diseases remain a leading cause of morbidity and mortality in the U.S. and elsewhere.(1) Middle age and older adults, in particular, are at high risk for serious infections and their long-term consequences.(2, 3) Among older adults, community-acquired serious infections (including pneumonia, sepsis, and meningitis) often require hospitalization and represent a substantial burden on the U.S. healthcare system.(4-7) The incidence of community-acquired pneumonia is very high among adults ≥ 50 years of age (248 cases per 100,000 adults) with an even higher burden among adults ≥ 80 years of age (1,643 cases per 100,000 adults).(8) Sepsis, cellulitis and pyelonephritis are also very common (sepsis: 100 cases per 100,000 and cellulitis/pyelonephritis: >150 hospitalizations per 100,000 adults) with an increasing incidence of severe sepsis with increased age.(9-11) Meningitis and endocarditis are relatively rare (around 2-3 cases per 100,000), although the case fatality rate is very high.(12, 13) Therefore, it is important to monitor the incidence of these infections, identify important risk factors, and determine the impact of preventative policies (e.g., vaccination) on these diseases among middle age and older adults.(14-16)

Large-scale epidemiological studies using administrative data often use serious infections as outcomes.(17-21) However, few studies have evaluated the performance of diagnosis codes to identify serious infections among middle age and older adults. Most previous studies that have assessed the performance of coded discharge diagnosis codes to identify serious infections have focused mainly on common infections (e.g., pneumonia or sepsis), specific populations (e.g., patients with rheumatoid arthritis), or on healthcare-associated or hospital-acquired infections.(22-31) Nevertheless, the performance of coded discharge diagnoses for accurately identifying infections requiring hospitalization among middle age and older adults is unclear. Therefore, we sought to determine the positive predictive value (PPV) of specific discharge diagnoses for identifying infections that required hospitalization among middle age and older adults.

METHODS

Data sources

TennCare is the managed Medicaid program in the State of Tennessee that provides healthcare insurance to those who are Medicaid eligible (around 20% of the Tennessee population).(32) The adult TennCare population consists of low-income pregnant women and individuals who are elderly or have a disability (over 600,000 annually).(32) We used data from TennCare, supplemented with data from the Tennessee Hospital Discharge Data System (a registry for all hospitalizations in Tennessee) and pharmacy information from Medicare Part D for those that were dual eligible, to identify a retrospective cohort of TennCare enrollees ≥50 years of age with pharmacy benefits (2008-2012). We restricted the hospitalizations for serious infection to those occurring from 2008 through 2012 to only include more recent hospitalizations for which medical records are more likely to be available. Cohort members had at least 180 days of baseline continuous enrollment before cohort entry, and were also required to be free of certain life-threatening conditions known to increase the risk of infection (solid organ transplantation, end-stage renal disease, HIV/AIDS, malignancy and serious kidney, liver and respiratory disease) that may limit longitudinal follow-up and impact the assessments of patients' exposures and their risk of infections. Cohort members were also required to have evidence of at least one pharmacy prescription fill and evidence of at least one healthcare encounter during baseline (to ensure use of benefits so that if a healthcare encounter for an infection occurred, it would be detected). Follow-up started on the earliest date the inclusion criteria were met and continued through the earliest of the following: study end date (December 31, 2012), the day prior to diagnosis of a serious life-threatening condition that would have precluded entry to the cohort, loss of enrollment, or date of death. From this retrospective cohort, we identified possible hospitalizations for serious infections (see *Identification of hospitalizations for serious infection)* for our validation study. To avoid including infections that may have originated due to a previous hospital stay, we excluded hospitalizations for infections that occurred in the 30-day period after discharge from a previous hospitalization. The study was approved by the institutional review boards of Vanderbilt University and the Tennessee Department of Health, and by the Division of TennCare.

Identification of hospitalizations for serious infection

Clinical knowledge and a literature review were used to identify primary discharge diagnosis codes that have been used previously to identify specific serious infections that require hospitalization (*study infections)*, including pneumonia (alone or with a primary diagnosis of bacteremia/sepsis), bacteremia/sepsis, pyelonephritis, meningitis/encephalitis, osteomyelitis/septic arthritis, endocarditis and cellulitis.(31, 33-35) Specific International Classifications of Diseases-Clinical Modification, 9th-revision (ICD-9-CM) diagnosis codes used to identify possible hospitalizations for each infection type are presented in Table 1. As the objective of our study was to determine the PPV of coding algorithms to identify serious infections that required hospitalization, we focused only on primary diagnoses of infection to reduce the possibility of detecting concurrent infections that may not have led to hospitalization or nosocomial infections that developed during the course of the hospitalization.(35)

Sampling Strategy

We used stratified random sampling to select a representative subset of study infection hospitalizations from among all possible cases identified in the retrospective cohort from among hospitals within 200 miles of Vanderbilt University Medical Center (VUMC). Since larger hospitals would be over-represented in a purely random sampling, and because there may also be regional variability in coding practices and infection prevalence, we constructed a sampling framework where hospitals were stratified based on their geographic region in Tennessee (West, Central, and East), and tertiles of reported discharge volume (Low, Medium, and High) during the study period.(36-38) From this sampling framework, we randomly selected three hospitals from each of these nine sampling strata, and retrieved their medical records for review and validation (Figure 1). This strategy, relative to a purely random sample, ensured better representation of infections identified in smaller hospitals and those in more rural regions of the State of Tennessee. If a hospital refused to participate, it was replaced by another hospital randomly selected from the same sampling stratum.

The overall goal was to review and validate 675 hospitalizations for serious infection from 27 hospitals (25 hospitalizations for each of the 3 hospitals comprising a stratum, yielding 75 hospitalizations for each of the 9 strata) (Figure 1). We conservatively assumed that up to 80% of records requested would

be available for review, and so we requested 32 records per hospital to receive an average of 25 records from each (Figure 1). To ensure that we reviewed sufficient rare infections, we preferentially selected any identified possible hospitalizations for meningitis/encephalitis, osteomyelitis/septic arthritis and endocarditis from each hospital in the sample. We randomly selected the remaining set of possible hospitalizations for other serious infections based on the proportional distribution of common infections at each hospital (pneumonia, bacteremia/sepsis, pyelonephritis and cellulitis) until 32 infections were identified. For hospitals with fewer than 32 infections during the study period, all infections were requested.

Abstraction of Medical Records

Relevant clinical information was abstracted from the medical record (transfer notes, emergency room summary, admission summary, physical/history, pharmacy information, laboratory, microbiology, and radiology information, and discharge summary) of each hospitalization with a primary discharge diagnosis code indicative of infection using a standardized and customized REDCap electronic data capture instrument hosted at Vanderbilt University.(39) As we were interested in infections that led to hospitalizations, we focused our reviews on clinical, microbiological and radiological information from the 2 days prior to the admission date through 2 days after admission to limit the possibility of identifying infections that developed during the hospitalization (i.e. nosocomial infections). In preparation for this study, the case report form was pilot-tested among a separate, convenience sample of 354 possible infections identified in the cohort from 3 hospitals in the same city as Vanderbilt University. This separate sample of hospitalizations was used only for pilot-testing the case report form, and was not included in the current study. One trained medical reviewer abstracted the relevant information for all selected records using the case report form. During the abstraction process, the lack of a particular finding in the medical record was treated as a lack of evidence for that finding, and so no information was considered missing after abstraction.

Adjudication of Medical Records

All records received were abstracted, reviewed and adjudicated. We made the final determination of whether a hospitalization represented a confirmed infection or not using *a priori* definitions of clinical, radiological, and/or microbiological findings compatible with infection for each infection type. Previous validation studies and expert clinical knowledge were used to define these specific *a priori* definitions for each infection type (*Supplementary appendix)*.(31, 35, 40)

Statistical analysis

We calculated the PPV of the ICD-9-CM discharge diagnosis codes for identifying hospitalizations for serious infection using the results of the *a priori* definitions applied to the information abstracted from the medical records as the reference (the proportion of cases identified with discharge diagnosis codes that were determined to be true cases after adjudication of the medical record information). We calculated 95% confidence intervals for the PPV using Wilson's formula.(41) Secondary analyses assessed the PPV for hospitalizations for serious infection across hospitals of different sizes and in different geographical regions of Tennessee.

We also assessed the reliability of the abstraction process. A second trained medical reviewer abstracted relevant information from a subset of selected records, which included all meningitis and endocarditis records, and a random selection of 10% of each of the remaining infection types. Each reviewer conducted the process independently and blinded from one another. For the subset of records abstracted by both reviewers, inter-reviewer agreement for the adjudication of a true or mis-identified infection was assessed using the Gwet's first agreement coefficient (AC_1).(42-44) Since Cohen's kappa statistic can be unreliable when the prevalence of the event and the level of observer agreement are high in the study sample, we used Gwet's AC_1 as a reliability measure unlikely to be affected by these concerns.(44-46)

In planned sensitivity analyses, we first assessed the impact of excluding hospitalizations that occurred after the individual was transferred from another healthcare facility, as initial documentation and details of the infection could be missing or incomplete in the receiving hospital.(40) We also assessed the impact on the PPV for all infections when requiring microbiological identification of a pathogen

(excluding common contaminants) from a sterile site within 2 days before or after the hospitalization admission date. A final sensitivity analysis among hospitalizations for possible pneumonia assessed the PPV when radiological evidence of pneumonia was required [i.e. pneumonia, opacity, or infiltrate mentioned in a chest X-ray or computed tomography scan report] (*Supplementary appendix*). All analyses were performed in Stata-IC, version 15.1 (College Station TX).

Patient and Public Involvement

No patients were involved in the development of the research question, the outcome measures, or the design or conduct of the study. As we conducted a retrospective study using administrative data, we have no plans to disseminate the results of the research to study participants.

RESULTS

Cohort characteristics

Among a retrospective cohort of 129,465 adults ≥50 years of age enrolled in TennCare, 9,769 hospitalizations for serious infection were identified during the study period (2008-2012) among 7,770 unique patients (Figure 2). Cohort members were primarily female (57.8%) with a median age of 54 years (mean: 57 years; range: 50-110). Among the 8,322 hospitalizations for serious infection that occurred at a hospital within 200 miles of VUMC, pneumonia, cellulitis and bacteremia/sepsis were the most common infections (54.3%, 20.5% and 18.4%, respectively), followed by pyelonephritis (3.8%) and septic arthritis/osteomyelitis (2.5%). Fewer than 1% of hospitalizations were due to meningitis/encephalitis (n=30) and endocarditis (n=18).

Collection, review and adjudication of selected medical records

Of the 27 hospitals that were initially selected for the sample, 21 (78%) were able to participate. We selected 7 additional hospitals to replace the 6 non-participants to achieve the desired sample size, including an additional small hospital in the East region due to a large number of unavailable records from a single participating hospital.

We received 716 (89%) of 808 requested records from 28 participating hospitals [Table 2]. Record availability from participating hospitals was lower in medium size hospitals (81.8%) compared to small (93.5%) and large hospitals (91.7%), but did not differ by geographic region. Record availability by infection type was greater than 86% for all infection types, with the exception of hospitalizations for the rare endocarditis cases (57.1%; only 4 of 7 cases).

The sample of hospitalizations for serious infection included patients who were primarily female (63.6%), with a median age of 60 years (mean: 64 years; range: 50-101) at the time of hospitalization. There was evidence of transfer from a prior healthcare facility for 21.8% of the hospitalizations for serious infection [highest percentage of transfers for bacteremia/sepsis (38.5%) and pneumonia (25.1%)]. The most common healthcare facility source was a nursing home/skilled nursing facility (84.6%), but also included group home sources (7.7%), other sources (4.5%) [assisted living facility, mental health center] and another acute care hospital (3.2%). There was evidence of an emergency department visit within 7 days prior to admission date for the serious infection hospitalization in 4.8% of the records.

Performance of discharge diagnosis codes

A total of 646 [PPV: 90.2% (95% CI: 87.8-92.2)] of the hospitalizations for serious infection identified using ICD-9-CM primary discharge diagnosis codes were confirmed by applying the *a priori* definitions to the abstracted data. The PPV was highest for pneumonia and cellulitis [96.5% (95% CI: 93.9-98.0) and 91.1% (95% CI: 84.7-94.9), respectively], and was ≥75% for bacteremia/sepsis, pyelonephritis, septic arthritis/osteomyelitis, and endocarditis. The PPV was lowest for meningitis/encephalitis [50.0% (95% CI: 23.7-76.3)], although the precision was limited due to a low number of available records for review (Table 2). Among the 10 potential cases of meningitis/encephalitis, 7 cases were meningitis/meningoencephalitis and 3 were encephalitis. The respective PPVs for meningitis/meningoencephalitis and encephalitis were 71.4% (95% CI: 35.9-91.8) and 0%, respectively.

When performance was evaluated across stratification sampling parameters, no apparent differences were observed in the PPV for records from hospitals in different geographical regions of

Tennessee. Although the PPV was high for all three discharge volume groups, the PPV was significantly lower in large hospitals [84.6% (95% CI: 79.6-88.5)] compared to smaller hospitals [93.9% (95% CI: 90.0-96.3); PPV difference: -9.3% (95% CI: -14.7, -3.9)] and medium hospitals [92.7% (95% CI: 88.6-95.4); PPV difference: -8.1% (95% CI: -13.7, -2.6)] (Table 2). This was likely driven by the different distributions in the types of infections selected for review in the hospital groups. Large hospitals had a higher proportion of non-pneumonia infections (70.4%) compared to medium and small hospitals (49.4% and 36.1%, respectively). Importantly, the PPV for pneumonia was similar in each discharge volume group (range: 96.0 to 96.6%), whereas the PPV was smaller for non-pneumonia infections in large hospitals (79.8%) compared to medium (88.7%) and small (89.2%) hospitals.

In the 82 records independently abstracted by two reviewers to assess reliability, there was 92.7% (95% CI: 86.9-98.4) agreement for identifying true hospitalizations for serious infection. The inter-rater agreement was also high when assessing reliability, independent of the outcome prevalence, with an AC_1 of 0.91 (95% CI: 0.84-0.99). Of the 6 discordant cases, 3 were meningitis/encephalitis (1 meningitis/meningoencephalitis and 2 encephalitis), with one each of bacteremia/sepsis, pyelonephritis and septic arthritis. The main reason for a discrepancy between reviewers was whether or not treatment for the infection of interest occurred within 2 days of the admission date, which was one of the major criteria for adjudication (see *Supplementary appendix*).

Sensitivity analyses

The PPV was virtually unchanged when excluding the 21.8% of hospitalizations that occurred as transfers from another healthcare facility [90.1% (95% CI: 87.4-92.3)]. Microbiological evidence of the specific infection type was found in 47.6% of records, leading to reduced PPVs when requiring microbiological evidence [45.4% (95% CI: 41.8-49.1)]. Microbiological evidence of infection was highest in hospitalizations for suspected pyelonephritis (94.4%), but was ≤60% for every other infection type [pneumonia (42.7%); cellulitis/soft tissue infections (58.5%); bacteremia/sepsis (26.1%)]. When requiring radiological confirmation of pneumonia, the PPV for coded diagnoses was 78.8% (95% CI: 74.2-82.8). Approximately 95.6% of possible hospitalizations for pneumonia had at least one documented

chest x-ray or CT-scan. Among those patients with a chest x-ray or CT-scan report available (n=325), 83.4% had a finding compatible with pneumonia. The main findings among the 54 patients with possible pneumonia and a radiological report available, but without radiological confirmation of pneumonia included atelectasis (n=6), interstitial pneumonitis (n=3), chronic heart failure with pulmonary edema (n=1), and no radiological findings of any kind (n=44).

DISCUSSION

Discharge diagnoses for identifying hospitalizations due to serious infections among middle age and older adults had an overall PPV of 90.2%, with highest values for the identification of common serious infections. PPVs were consistently high across different hospital types and regions of Tennessee. Furthermore, the PPV was similar after exclusion of hospitalizations for serious infection that were the result of a transfer from another healthcare facility (e.g. acute care hospital, skilled nursing facility). Microbiological confirmation was available for fewer than 50% of patients admitted with possible hospitalizations for serious infections, and as expected, the requirement resulted in a low PPV for all infections, with the exception of pyelonephritis. Importantly, the PPV for pneumonia hospitalizations remained relatively high even when requiring radiological confirmation.

The PPV for hospitalizations for pneumonia in previous smaller validation studies has ranged from 72 to 86% in different healthcare systems, but those studies were not focused on middle age and older adults.(31, 47-49) In our study of hospitalizations among middle age and older adults, we found that coded discharge diagnoses have a higher PPV for pneumonia compared to previous studies. The PPV for bacteremia/sepsis was also on the higher range of previously reported PPVs for diagnosis codes to identify bacteremia/sepsis from administrative data in other populations (reported range from 45% to 97.7%), and for septic arthritis/osteomyelitis compared to a previous study conducted among patients with diabetes (63.9% versus 75.9% in our study).(23, 50, 51) Overall, the observed PPV for all infections in our study was comparable to two previous comprehensive validation studies of bacterial infections, one among patients with rheumatoid arthritis in a single hospital system and another among patients in one of

the Veteran's Affairs integrated service networks.(35, 40) Compared to these two previous studies of ICD-9 codes, we abstracted and adjudicated a larger number of records while using a more systematic sampling strategy to retrieve and review records for hospitalizations from multiple regions and hospital types as opposed to a single hospital or healthcare system. However, some of the PPVs for individual infections were less precise and less similar to these previous studies. This was especially true for rare infections, as would be expected due to the low numbers of rare infections in our study and across previous studies.(35, 40) The results of our study are also similar to previous validation studies that used corresponding ICD-10 diagnosis codes to identify hospitalizations for serious infection.(52, 53)

One limitation to consider in our study was that it was not designed to estimate the sensitivity and specificity of the coding algorithms. This would have required the identification, review and adjudication of a sample of hospitalizations that did not fulfill our algorithm (i.e. presence of the ICD-9 primary discharge diagnosis codes indicative of infection). However, when the prevalence of an outcome is low, the PPV approximates the specificity.(54) Importantly, any non-differential outcome misclassification between exposure groups resulting from the use of imperfect but highly-specific measurements would attenuate the impact of the misclassification on the relative risk estimates.(55) In addition, we found that the PPV of coded discharge diagnoses for serious infections remained high across hospitals of different sizes and across different geographical areas of Tennessee, which may have different rates of hospitalizations for serious infection.(56) Although our study applied a systematic sampling strategy to assure the representation of different settings in our population, our population was restricted to middle age and older adults enrolled in a State Medicaid program. Therefore, caution is warranted when extrapolating the study findings to other populations.

Another limitation is the use of available clinical information to operationalize definitions for adjudication of true hospitalizations for infections. It is possible that some procedures, laboratory findings and diagnoses that informed the final diagnosis of infection were not fully recorded in the medical records, and thus, were not available for our review and may have contributed to the observed PPV for some infections. Although we used previous validation studies and clinical information to build pre-

specified definitions for the adjudication of true infections, our reference criteria may be imperfect, considering the retrospective nature of our determinations and potential variability in clinical practice. Nevertheless, we also assessed how the availability of selected findings (i.e. microbiological and radiological information) in the medical record impacted the overall and infection-specific PPV. We demonstrated that relying on highly specific clinical diagnostics, such as microbiological and radiological information, to confirm true infections would result in lower PPVs for identification of infections in administrative data. Requiring microbiological confirmation to confirm true infections is challenging because of the known low sensitivity of culture-based diagnostic methods (most commonly used in clinical practice), which may lead to misclassification.(57, 58) In addition, requiring radiological evidence compatible with pneumonia within 2 days of hospital admission did lower the observed PPV for pneumonia hospitalizations. Nevertheless, the observed PPV remained close to 80%, which should reduce concerns about using diagnosis codes to identify hospitalizations due to pneumonia. Finally, the coding algorithms were based on the ICD-9-coding system only. Although these findings will be helpful for retrospective studies that encompass periods of ICD-9 use, additional studies evaluating the performance of ICD-10-based codes would be useful to complement our findings.

Our study demonstrated that discharge diagnosis codes can be used to accurately identify hospitalizations for serious infections among middle age and older adults. The highest PPVs were observed for the most common infections, and the PPV for pneumonia remained high when requiring radiological confirmation. Importantly, consistently high PPVs were observed across different hospital sizes and regions. However, the estimated PPV was lower and less precise for very rare infections (e.g. encephalitis). This should be an important consideration for studies specifically focused on those less frequent outcomes, especially when strict microbiological confirmation is required. Taken together, these findings support the use of discharge diagnosis codes for infections to identify outcomes in ongoing and future epidemiological studies among middle age and older adults.

Acknowledgement

We are indebted to the Tennessee Bureau of TennCare of the Tennessee Department of Finance and Administration, which provided data for the study. We are also indebted to the Tennessee Department of Health for providing data for the study. Statements in the report should not be construed as endorsement by the U.S. Department of Health and Human Services, the Department of Veterans Affairs, or the Tennessee Department of Health.

Funding

This study was funded by the NIH (R03-AG-042981 and R01-AG-043471-01A1) and the TL1 award TL1TR000447.

Contributors

ADW planned the medical record collection and statistical analysis, analyzed and interpreted the data, and drafted and revised the paper. MRG, WS, CMS, and RAG planned the statistical analysis, interpreted the data and revised the paper. EFM prepared the data, and revised the paper. CGG designed the project, acquired the data from TennCare, planned the medical record collection and statistical analysis, interpreted the data, and revised the paper.

Declaration of interests

CGG has received consulting fees from Pfizer and Merck, and received research support from Sanofi-Pasteur, Campbell Alliance, the Centers for Disease Control and Prevention, National Institutes of Health, The Food and Drug Administration, and the Agency for Health Care Research and Quality. WS has received personal fees from Pfizer, Merck, Novavax, Dynavax, Sanofi-Pasteur, GSK, Seqirus, and received research support from the Centers for Disease Control and Prevention. ADW, EFM, CMS, MRG and RAG have no conflicts of interest to disclose.

Data sharing

No additional unpublished data are available from the study. The study protocol and statistical code are available from the corresponding author, Andrew Wiese (andrew.d.wiese@vanderbilt.edu).

REFERENCES

1. Khabbaz RF, Moseley RR, Steiner RJ, Levitt AM, Bell BP**.** Challenges of infectious diseases in the USA. Lancet. 2014;384(9937):53-63.

2. Grohskopf LA, Sokolow LZ, Broder KR, Olsen SJ, Karron RA, Jernigan DB, et al. Prevention and Control of Seasonal Influenza with Vaccines. MMWR Recomm Rep. 2016;65(5):1-54.

3. Kim DK, Bridges CB, Harriman KH**.** Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older--United States, 2015. MMWR Morb Mortal Wkly Rep. 2015;64(4):91-2.

4. Jeon MH, Chung JW, Choi SH, Kim TH, Lee EJ, Choo EJ**.** Pneumonia risk factors and clinical features of hospitalized patients older than 15 years with pandemic influenza A (H1N1) in South Korea: a multicenter study. Diagn Microbiol Infect Dis. 2011;70(2):230-5.

5. Pilishvili T, Bennett NM**.** Pneumococcal Disease Prevention Among Adults: Strategies for the Use of Pneumococcal Vaccines. Am J Prev Med. 2015;49(6 Suppl 4):S383-90.

6. Sjoding MW, Prescott HC, Wunsch H, Iwashyna TJ, Cooke CR**.** Longitudinal Changes in ICU Admissions Among Elderly Patients in the United States. Crit Care Med. 2016;44(7):1353-60.

7. Crotty MP, Meyers S, Hampton N, Bledsoe S, Ritchie DJ, Buller RS, et al. Epidemiology, Co-Infections, and Outcomes of Viral Pneumonia in Adults: An Observational Cohort Study. Medicine (Baltimore). 2015;94(50):e2332.

8. Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. N Engl J Med. 2015;373(5):415-27.

9. Filbin MR, Arias SA, Camargo CA, Jr., Barche A, Pallin DJ**.** Sepsis visits and antibiotic utilization in U.S. emergency departments*. Crit Care Med. 2014;42(3):528-35.

10. Christensen KL, Holman RC, Steiner CA, Sejvar JJ, Stoll BJ, Schonberger LB**.** Infectious disease hospitalizations in the United States. Clin Infect Dis. 2009;49(7):1025-35.

11. Ginde AA, Moss M, Shapiro NI, Schwartz RS. Impact of older age and nursing home residence on clinical outcomes of US emergency department visits for severe sepsis. J Crit Care. 2013;28(5):606-11.

12. Iung B, Vahanian A. Epidemiology of acquired valvular heart disease. Can J Cardiol. 2014;30(9):962-70.

13. Thigpen MC, Whitney CG, Messonnier NE, Zell ER, Lynfield R, Hadler JL, et al. Bacterial meningitis in the United States, 1998-2007. N Engl J Med. 2011;364(21):2016-25.

14. Simonetti AF, Garcia-Vidal C, Viasus D, Garcia-Somoza D, Dorca J, Gudiol F, et al. Declining mortality among hospitalized patients with community-acquired pneumonia. Clin Microbiol Infect. 2016;22(6):567.e1-7.

15. Rivero-Calle I, Pardo-Seco J, Aldaz P, Vargas DA, Mascaros E, Redondo E, et al. Incidence and risk factor prevalence of community-acquired pneumonia in adults in primary care in Spain (NEUMO-ES-RISK project). BMC Infect Dis. 2016;16(1):645.

16. Park JY, Park S, Lee SH, Lee MG, Park YB, Oh KC, et al. Microorganisms Causing Community-Acquired Acute Bronchitis: The Role of Bacterial Infection. PLoS One. 2016;11(10):e0165553.

17. Jackson ML, Walker R, Lee S, Larson E, Dublin S. Predicting 2-Year Risk of Developing Pneumonia in Older Adults without Dementia. J Am Geriatr Soc. 2016;64(7):1439-47.

18. Clement RC, Haddix KP, Creighton RA, Spang JT, Tennant JN, Kamath GV. Risk Factors for Infection After Knee Arthroscopy: Analysis of 595,083 Cases From 3 United States Databases. Arthroscopy. 2016;32(12):2556-61.

19. Weinert BA, Edmonson MB. Hospitalizations at Nonfederal Facilities for Lower Respiratory Tract Infection in American Indian and Alaska Native Children Younger than 5 Years of Age, 1997-2012. J Pediatr. 2016;175:33-9.e4.

20. Akiyama T, Chikuda H, Yasunaga H, Horiguchi H, Fushimi K, Saita K. Incidence and risk factors for mortality of vertebral osteomyelitis: a retrospective analysis using the Japanese diagnosis procedure combination database. BMJ Open. 2013;3(3).

21. Graversen ME, Dalgaard LS, Jensen-Fangel S, Jespersen B, Ostergaard L, Sogaard OS. Risk and outcome of pyelonephritis among renal transplant recipients. BMC Infect Dis. 2016;16:264.

22. Abou Zahr Z, Spiegelman A, Cantu M, Ng B. Perioperative use of anti-rheumatic agents does not increase early postoperative infection risks: a Veteran Affairs' administrative database study. Rheumatol Int. 2015;35(2):265-72.

23. Carnahan RM, Herman RA, Moores KG. A systematic review of validated methods for identifying transfusion-related sepsis using administrative and claims data. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 1:222-9.

24. Condell O, Gubbels S, Nielsen J, Espenhain L, Frimodt-Moller N, Engberg J, et al. Automated surveillance system for hospital-acquired urinary tract infections in Denmark. J Hosp Infect. 2016;93(3):290-6.

25. Gedeborg R, Furebring M, Michaelsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients. J Clin Epidemiol. 2007;60(2):155-62.

26. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of administrative code data for the surveillance of healthcare-associated infections: a systematic review and meta-analysis. Clin Infect Dis. 2014;58(5):688-96.

27. Shaklee J, Zerr DM, Elward A, Newland J, Leckerman K, Asti L, et al. Improving surveillance for pediatric Clostridium difficile infection: derivation and validation of an accurate case-finding tool. Pediatr Infect Dis J. 2011;30(3):e38-40.

28. Wright SB, Huskins WC, Dokholyan RS, Goldmann DA, Platt R. Administrative databases provide inaccurate data for surveillance of long-term central venous catheter-associated infections. Infect Control Hosp Epidemiol. 2003;24(12):946-9.

29. Greenberg JA, Hohmann SF, Hall JB, Kress JP, David MZ. Validation of a Method to Identify Immunocompromised Patients with Severe Sepsis in Administrative Databases. Ann Am Thorac Soc. 2016;13(2):253-8.

30. Olsen MA, Young-Xu Y, Stwalley D, Kelly CP, Gerding DN, Saeed MJ, et al. The burden of clostridium difficile infection: estimates of the incidence of CDI from U.S. Administrative databases. BMC Infect Dis. 2016;16:177.

31. Grijalva CG, Chung CP, Stein CM, Gideon PS, Dyer SM, Mitchel EF, Jr., et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. Pharmacoepidemiol Drug Saf. 2008;17(9):890-5.

32. TennCare Fiscal Year 2015-2016 Annual Report. In: Administration DoHCF, ed; 2016.

33. Wiese AD, Griffin MR, Stein CM, Mitchel EF, Jr., Grijalva CG**.** Opioid Analgesics and the Risk of Serious Infections Among Patients With Rheumatoid Arthritis: A Self-Controlled Case Series Study. Arthritis Rheumatol. 2016;68(2):323-31.

34. Grijalva CG, Kaltenbach L, Arbogast PG, Mitchel EF, Jr., Griffin MR**.** Initiation of rheumatoid arthritis treatments and the risk of serious infections. Rheumatology (Oxford). 2010;49(1):82-90.

35. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH**.** Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. J Clin Epidemiol. 2007;60(4):397-409.

36. Brenner H, Gefeller O**.** Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med. 1997;16(9):981-91.

37. Kim HM, Smith EG, Stano CM, Ganoczy D, Zivin K, Walters H, et al. Validation of key behaviourally based mental health diagnoses in administrative data: suicide attempt, alcohol abuse, illicit drug abuse and tobacco use. BMC Health Serv Res. 2012;12:18.

38. El-Ghitany EM, Farghaly AG, Farag S, Abd El-Wahab EW**.** Validation of EGCRISC for Chronic Hepatitis C Infection Screening and Risk Assessment in the Egyptian Population. PLoS One. 2016;11(12):e0168649.

39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-81.

40. Patkar NM, Curtis JR, Teng GG, Allison JJ, Saag M, Martin C, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. J Clin Epidemiol. 2009;62(3):321-7, 7.e1-7.

41. Niesner K, Murff HJ, Griffin MR, Wasserman B, Greevy R, Grijalva CG, et al. Validation of VA administrative data algorithms for identifying cardiovascular disease hospitalization. Epidemiology. 2013;24(2):334-5.

42. Govatsmark RE, Sneeggen S, Karlsaune H, Slordahl SA, Bonaa KH. Interrater reliability of a national acute myocardial infarction register. Clin Epidemiol. 2016;8:305-12.

43. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol. 2013;13:61.

44. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;61(Pt 1):29-48.

45. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. J Clin Epidemiol. 1996;49(4):431-4.

46. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543-9.

47. Drahos J, Vanwormer JJ, Greenlee RT, Landgren O, Koshiol J. Accuracy of ICD-9-CM codes in identifying infections of pneumonia and herpes simplex virus in administrative data. Ann Epidemiol. 2013;23(5):291-3.

48. Aronsky D, Haug PJ, Lagor C, Dean NC. Accuracy of administrative data for identifying patients with pneumonia. Am J Med Qual. 2005;20(6):319-28.

49. Kern DM, Davis J, Williams SA, Tunceli O, Wu B, Hollis S, et al. Validation of an administrative claims-based diagnostic code for pneumonia in a US-based commercially insured COPD population. Int J Chron Obstruct Pulmon Dis. 2015;10:1417-25.

50. Iwashyna TJ, Odden A, Rohde J, Bonham C, Kuhn L, Malani P, et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. Med Care. 2014;52(6):e39-43.

51. Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. J Clin Epidemiol. 1999;52(3):199-207.

52. Sahli L, Lapeyre-Mestre M, Derumeaux H, Moulis G**.** Positive predictive values of selected hospital discharge diagnoses to identify infections responsible for hospitalization in the French national hospital database. Pharmacoepidemiol Drug Saf. 2016;25(7):785-9.

53. Holland-Bill L, Xu H, Sorensen HT, Acquavella J, Svaerke C, Gammelager H, et al. Positive predictive value of primary inpatient discharge diagnoses of infection among cancer patients in the Danish National Registry of Patients. Ann Epidemiol. 2014;24(8):593-7, 7.e1-18.

54. Schneeweiss S, Avorn J**.** A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323-37.

55. Rodgers A, MacMahon S**.** Systematic underestimation of treatment effects as a result of diagnostic test inaccuracy: implications for the interpretation and design of thromboprophylaxis trials. Thromb Haemost. 1995;73(2):167-71.

56. van Walraven C, English S, Austin PC**.** Administrative database code accuracy did not vary notably with changes in disease prevalence. J Clin Epidemiol. 2016;79:86-9.

57. Mancini N, Carletti S, Ghidoli N, Cichero P, Burioni R, Clementi M**.** The era of molecular and other non-culture-based methods in diagnosis of sepsis. Clin Microbiol Rev. 2010;23(1):235-51.

58. Niederman MS**.** The argument against using quantitative cultures in clinical trials and for the management of ventilator-associated pneumonia. Clin Infect Dis. 2010;51 Suppl 1:S93-9.

Table 1. Discharge diagnosis code definitions (ICD-9-CM) for hospitalizations for serious infection

Serious Infection	Primary (first listed) discharge diagnosis code
Pneumonia-primary definition	003.22, 480.*‡, 481, 482.*, 483.*, 484.*, 485.*, 486.*, 487.0
Pneumonia-secondary definition (pneumonia diagnosis (above) in any other diagnosis field)	510.*, 038.*, 790.7, 995.91, 995.92
Meningitis/ Encephalitis	003.21, 036.0, 0.47*, 049.*, 053.0, 054.72, 072.1, 091.81, 094.2, 098.82, 100.81, 320.*, 036.1, 054.3, 056.01, 058.21, 058.29, 062.*, 063.*, 064.*, 066.41, 072.2, 094.81, 130.0, 323.*
Bacteremia/ Sepsis†	038.*, 790.7, 995.91, 995.92
Cellulitis/ Soft-tissue infections	035, 040.0, 569.61, 681.*, 682.*, 728.86, 785.4
Endocarditis	036.42, 074.22, 093.2*, 098.84, 421.*, 422.92
Pyelonephritis	590.*
Septic Arthritis/ Osteomyelitis	003.23, 056.71, 098.5*, 711.0, 711.00-711.07, 711.09, 711.9*, 003.24, 376.03, 526.4, 730.0*, 730.1*, 730.2*

† Without a diagnosis of pneumonia in any other diagnosis field

‡ A * indicates all numeric values [0-9]

Table 2. Positive predictive value (PPV) of coded discharge diagnosis definitions for hospitalizations for serious infections among adults ≥ 50 years of age enrolled in Tennessee Medicaid, 2008-2012

Type	Expected Number of Records	Records Received	PPV (95 % CI)
Overall	675	716	90.2 (87.8, 92.2)
Region Specific			
West	225	195	91.3 (86.5, 94.5)
Central	225	225	88.9 (84.1, 92.4)
East	225	296	90.5 (86.7, 93.4)
Bed volume size specific			
Low	225	230	93.9 (90.0, 96.3)
Medium	225	233	92.7 (88.6, 95.4)
High	225	253	84.6 (79.6, 88.5)
Serious Infection			
Pneumonia	305	340	96.5 (93.9, 98.0)
Cellulitis/Soft-tissue infections	125	123	91.1 (84.7, 94.9)
Pyelonephritis	80	89	87.6 (79.2, 93.0)
Bacteremia/Sepsis	100	92	82.6 (73.6, 89.0)
Septic Arthritis/Osteomyelitis	50	58	75.9 (63.5, 85.0)
Meningitis/Encephalitis	10	10	50.0 (23.7, 76.3)
Endocarditis	5	4	75.0 (30.1, 95.4)

Figure legends

Figure 1. Sampling strategy for identifying potential hospitalizations for serious infection

Figure 2. Identifying a retrospective cohort of patients ≥50 years of age without serious/life-threatening conditions, Tennessee Medicaid (2008-2012)

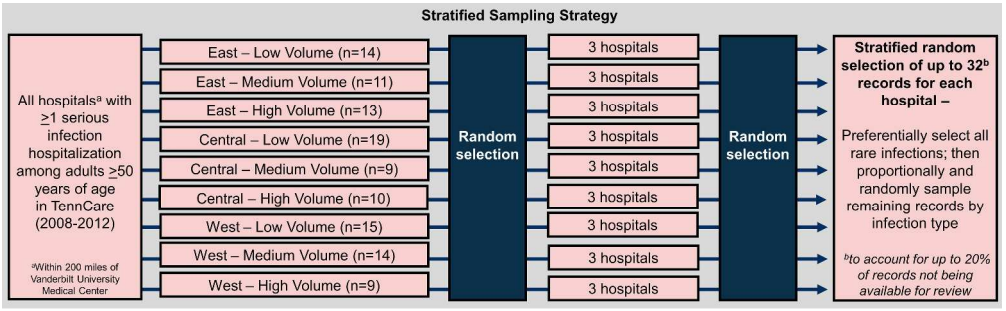

Figure 1. Sampling strategy for identifying potential hospitalizations for serious infection

457x139mm (300 x 300 DPI)

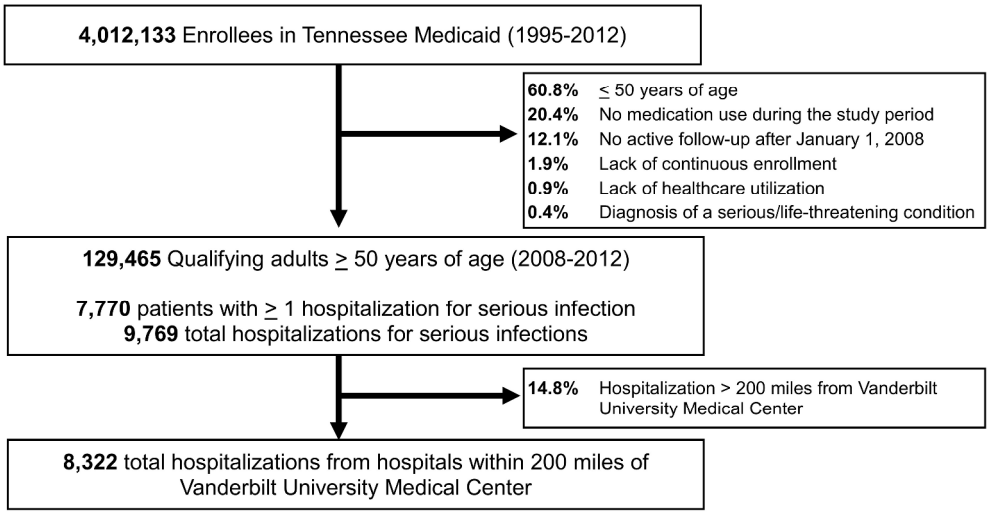

Figure 2. Identifying a retrospective cohort of patients ≥50 years of age without serious/life-threatening conditions, Tennessee Medicaid (2008-2012)

355x190mm (300 x 300 DPI)

Supplementary Appendix

Infection-Specific Definitions of Hospitalization for Serious Infection

We used a pre-specified adjudication process to determine whether each abstracted medical record corresponded to a true infection or not. Previous validation studies and expert clinical knowledge were used to define specific a priori definitions for each infection type.[1-3] Information abstracted from the medical record was compared to these *a priori* definitions for each infection type to make the final determination of whether a hospitalization represented a true infection or not.

Outline

I. Sepsis/Septicemia/Bacteremia/Septic Shock/Generalized Infection

Either of the following [1 or 2]:

1. **Positive culture of a non-contaminant pathogen**
    - i. Positive blood culture [any of the following (1-2)]
        1. Any gram-negative organism, except:
            a. No predominant organism
        2. A gram positive organism, except:
            a. Coagulase-negative *Staphylococcus*
            b. *Bacillus spp.* (other than *Bacillus anthracis*)
            c. *Corynebacterium spp.*
            d. *Propionibacterium spp.*
            e. *Micrococcus*
            f. Diptheroids
            g. Viridians Group Streptococci
            h. Enterococci
            i. *Clostridium perfringens*
            j. *Aerococcus*
            k. *Alcaligenes faecalis*
            l. *Citrobacter*
            m. *Neisseria subflava*
            n. *Stomatococcus*
            o. *Streptococcus bovis*
            p. *Veillonella candidemia*
            q. *Mycobacterium tuberculosis*
            r. *S. salivarius*
            s. *"Gram Positive"*
            t. *"No predominant organism"*
            u. *Streptococcus alpha*
2. **At least two of the following, documented at admission +/- 2 days [i-iii]**
    - i. Hypotension
        1. Systolic BP $\leq$ 90 mmHg
        2. Reduction of systolic BP of 40mmHg from earliest measurement collected during the admission of interest
    - ii. Two of the following [1-4]:
        1. Temperature $\geq$ 38^0C **or** $\leq$ 36^0C
        2. Heart rate $\geq$ 90 beats/minute
        3. Respiratory rate $\geq$ 20 breaths/min or $PaCO_2$ < 32 mmHg
        4. WBC $\geq$ 10,000 cells/mm^3 **or** $\leq$ 4,500 cells/mm^3 **or** WBC with > 10 % immature (band) forms
    - iii. Initiation of antibiotic treatment specifically for sepsis/septicemia/bacteremia/septic shock/generalized infection

II. Pneumonia

1. Pneumonia identified through examination (**all three of the following [a-c]):**
    a. One of the following admission findings indicative of respiratory findings:
        1. New and/or increased cough
        2. Shortness of breath
        3. Pleuritic chest pain
        4. New purulent production
        5. Altered mental status ("agitation" and "lethargy" included)
        6. Crackles
            a. Physical evidence of consolidation such as egophony, whispered pectoriloquy, etc.
    b. One of the following examination findings indicative of systemic infection [1-4]:
        1. Temperature (T $\geq$ 100.4^{0}F (38^{0}C) or $\leq$ 96^{0}F) in first 48 hours of admission
        2. Systolic BP $\leq$ 90mmHg
        3. Shock
            a. Volume nonresponsive hypotension
        4. Blood peripheral WBC ($\geq$ 10.0 x 10^9/L or $\leq$ 4.5 x 10^9/L)
    c. Treatment with antibiotics/antivirals indicated for suspected infection

OR

At least two of the following [1-3]:

1. Two of the following from #1 **([a and b], [a and c], or [b-c])**
2. Any of the following findings listed on chest imaging from radiologic report **documented at admission +/- 2 days**
    a. Pneumonia
    b. Lung abscess
    c. Opacity consistent with pneumonia/lung abscess
    d. Infiltrate consistent with pneumonia/lung abscess
    e. Consolidation consistent with pneumonia/lung abscess
    f. Increased density consistent with pneumonia/lung abscess
    g. Pleural effusion consistent with pneumonia/lung abscess
    h. Interstitial edema consistent with pneumonia/lung abscess
3. Sterile Site Laboratory Findings
    i. Any one of the following [i through v]
        i. Sputum lab findings [any **one** of the following (1, 2)]:
            1. Sputum culture/PCR/serology/gram stain positive for an agent that is not considered a contaminant [see exclusion list below]:
                a. *Aspergillus* species, *Enterococcus* species, viridians group streptococci, and yeast
            2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen
        ii. Blood lab findings [either of the following (1-3)]
            1. Blood culture/PCR/serology positive for an agent that is not considered a contaminant [see exclusion list below]:
                a. Exclusions
                    i. Coagulase-negative *Staphylococcus*

ii. *Bacillus spp.* (other than *Bacillus anthracis*)
iii. *Corynebacterium spp.*
iv. *Propionibacterium spp.*
v. *Micrococcus*
vi. Diptheroids
vii. Viridians Group Streptococci
viii. Enterococci
ix. *Clostridium perfringens*
x. *Aerococcus*
xi. *Alcaligenes faecalis*
xii. *Citrobacter*
xiii. *Neisseria subflava*
xiv. *Stomatococcus*
xv. *Streptococcus bovis*
xvi. *Veillonella candidemia*
xvii. *Mycobacterium tuberculosis*
xviii. *S. salivarius*

2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen

iii. Pleural fluid lab findings [either of the following (1, 2)]
   1. Culture/PCR/serology positive for a bacterial pathogen
   2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen
iv. Bronchoscopic specimen or deep endotracheal tube aspiration lab findings [either of the following (1, 2)]
   1. Culture/PCR/serology positive for a bacterial pathogen
   2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen
v. Urine antigen detection testing [either of the following (1, 2)]
   1. *Legionella pneumophila*
   2. *Streptococcus pneumoniae*

III. Cellulitis/Soft-Tissue Infection

Both of the following:

1. Any mention of the following with recent onset (≤14 days) ***[any of the following]***
   a. Skin erythema
   b. Surgical site infection
   c. Superficial central line infection
   d. Ostomy site infection
   e. Skin infection with associated lymphangitis
2. Antibiotic treatment initiated for suspected infection

IV. Endocarditis

Any one of the following [1-3]:

1. Major Criteria [both of the following]:
    a. Suggestive microbiology [at least one of the following]:
        i. Positive blood culture of an ***endocarditis organism*** **[any of the following]:**
            1. *Streptococcus bovis*
            2. *Viridians streptococci*
            3. *Staphylococcus aureus*
            4. *Enterococcus spp.*
            5. HACEK organisms
            6. Coagulase negative staphylococci
    b. Evidence of endocardial involvement [at least one of the following]:
        i. New regurgiant murmur (a change in a preexisting murmur does not get scored)
        ii. Echocardiogram suspicious for any of the following:
            1. Intracardiac mass with no alternative explanation
            2. Endocardial abscess
            3. New partial prosthesis dehiscence
            4. Vegetation on valve
2. Minor Criteria [at least 4 of the following]:
    a. Predisposing valvular disease or IV drug use
    b. Temperature $\geq$ 100.4°F or 38°C
    c. Vascular phenomena
        i. Janeway lesions, conjunctival hemorrhages, arterial emboli, septic pulmonary infarcts, mycotic aneurysm, intracranial bleed
    d. Immunologic phenomena
        i. Osler nodes, Roth Spots, elevated Rheumatoid factor, hematuria in non-catheter urine, or other evidence of glomerulonephritis
    e. Positive blood cultures
        i. Excluding a single positive culture for coagulase negative staphylococci or a single positive culture for an organism that does not fall into the "reasonable endocarditis organism" (i.e. coagulase-positive and coagulase-negative *S. aureus*, Enterococcus, viridians group Streptococci, *S. bovis*, HACEK organisms)
    f. Positive serology for Brucella, Bartonella, Legionella, or Chlamydia
    g. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
3. At least one Major Criteria **AND** 3 minor criteria.

V. Meningitis/Encephalitis

Any one of the following [1 or 2]:

1. Both of the following [a-b]
    a. Laboratory Findings [any one of the following (i-ix)]
        i. CSF demonstrates any bacterium
            1. Excluding Diptheroids, Propionibacteria, Bacillus, Coagulase Negative *Staphylococcus*
        ii. CSF demonstrates Diptheroids, Propionibacteria, Bacillus, Coagulase Negative *Staphylococcus* in the setting of past neurosurgical intervention **AND** physicians elected to treat with antibacterials
        iii. Blood cultures positive for any of the following:
            1. *S. pneumoniae*
            2. *H. influenza*
            3. *Neisseria meningitidis*
            4. Group B Streptococcus
        iv. Stool cultures positive for enterovirus
        v. Throat or sputum cultures positive for *Neisseria meningitidis* in the setting of a rapid onset, overwhelming infection syndrome, including petechiae
        vi. Serology positive for *Mycoplasma*, *Leptospira*, measles, mumps, lymphocytic choriomeningitis virus, arboviruses (e.g. St. Louis encephalitis virus), or HIV (if historically consistent with acute seroconversion).
        vii. Brain biopsy demonstrates encephalitis
        viii. Positive CSF culture or PCR detection for any of the following
        ix. Acute or convalescent serology demonstrates positive antibody pattern for any of the following:
            1. Encephalitis arbovirus (La Crosse, St. louis, Eastern Equine, Western Equine, Powassan, Japanese, West Nile)
    b. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected meningitis/encephalitis

2. At least two of the following [a-d]
    a. Clinical meningitis/encephalitis [at least two of the following]:
        i. Petechial rash
        ii. Nuchal rigidity (by history or exam)
        iii. Altered sensorium
        iv. Fever
        v. Altered level of consciousness, including “agitation” or “lethargy”
        vi. Behavioral change
        vii. Diminished level of consciousness (not easily roused)
        viii. History of any of the following: headaches, altered mental status, or recent exposure to patient with known bacterial meningitis
        ix. Reduction in fever within 72 hours of starting anti-bacterial
    b. Inflammatory CSF [at least one of the following i-ii]
        i. Pleocytosis: $\geq$ 15 WBC/mm^3 (after subtracting one WBC for every 1,000 RBC)
        ii. Elevated protein (based on local lab-determined upper limits)
    c. Suggestive Findings [at least one of the following (i-iv)

i. Septic syndrome
ii. Focal neurological deficits documented during examination (such as flaccid paralysis or speech alterations for West Nile Virus)
iii. Abnormal imaging
    1. Computed tomography or magnetic resonance imaging (MRI) demonstrating focal edema or inflammation or hemorrhage
    2. Indicated as “meningitis/encephalitis” or “compatible with meningitis/encephalitis” or “cannot rule out meningitis/encephalitis”
iv. Findings indicating an abnormal electroencephalography (such as focal periodic discharges)

d. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for presumed meningitis/encephalitis

VI. Pyelonephritis

At least two of the following [1-4]:

1. Suggestion of infection [at least one of the following]:
    a. Temperature ≥ 100.4°F (38°C)
    b. Peripheral blood WBC ≥ 10,000/mm^3
    c. Positive blood culture for any of the following:
        i. Gram Negative Rods
        ii. *Enterococcus spp.*
        iii. *Staphylococcus saprophyticus*
    d. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
2. Strong renal localization [at least one of the following]:
    a. CT, MRI, or Ultrasound Suggestive of Renal Inflammation
3. Minor Criteria [at least two of the following]:
    a. Flank pain
    b. Costovertebral angle tenderness
    c. Complaints of dysuria, frequency, or suprapubic pain
    d. Any pyuria
    e. Urine culture positive for a single organism
4. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected pyelonephritis

VII. Septic Arthritis/Osteomyelitis

Any one of the following (1-5):

1. Synovial fluid gram stain or tissue gram stain or special stain demonstrating any organism
2. Joint culture/PCR/serology positive for any organism
3. At least two of the following (a-d):
    a. Positive blood culture/PCR/serology
    b. Joint with acute (≤ 7 days) worsening of inflammatory features (**at least two of the following):**
        i. Pain on history
        ii. ROM
        iii. Warmth
        iv. Effusion
        v. Swelling
        vi. Limited range of motion
    c. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
    d. Any one of the following (i-iv)
        i. Synovial fluid WBC ≥ 30,000/mm^3
        ii. Synovial fluid WBC ≥ 60,000/mm^3 with > 75% PMNs
        iii. Skin lesions, tenosynovitis, or urethral/cervical/rectal Gram stain or culture suggestive of *Neisseria gonorrhoeae*
        iv. Any indication of the following in the synovial fluid: needle-like crystals, CPPD crystals, uric acid.
4. Positive bone biopsy [at least one of the following (a-c)]:
    a. Positive culture for any organism
    b. Positive gram stain
5. Imaging and indirect features **[at least two of the following (a-c)]:**
    a. Consistent imaging [at least one of the following (i-iv)]:
        i. Plain X-ray read by a radiologist as suggestive of osteomyelitis
        ii. CT Scan read by a radiologist as suggestive of osteomyelitis
        iii. MRI read by a radiologist as suggestive of osteomyelitis
        iv. Bone scan or WBC scan read as suggestive of osteomyelitis
    b. Suggestive indirect features[at least one of the following (i-viii)]:
        i. Temperature > 100.4°F (38°C)
        ii. Bony pain or tenderness or erythema over bone suspected to be infected
        iii. Draining soft tissue sinus over bone suspected to be infected
        iv. Positive "probe to bone" (or visible bone in deep ulcer at suspected site)
        v. Blood culture positive for *S. aureus*
        vi. ESR ≥ 75 mm/hour
        vii. Intravenous drug use or indwelling catheter
        viii. Inflammation on imaging associated with an orthopedic prosthesis
    c. Positive culture for any organism form wound sample over the bone suspected of infection
    d. Antibiotic/antiviral/antifungal treatment for suspected infection

VIII. References

1. Grijalva CG, Chung CP, Stein CM, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. Pharmacoepidemiology and drug safety 2008; 17(9): 890-5.
2. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. Journal of clinical epidemiology 2007; 60(4): 397-409.
3. Patkar NM, Curtis JR, Teng GG, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. *Journal of clinical epidemiology* 2009; **62**(3): 321-7, 7.e1-7.

Section & Topic	No	Item	Reported on page #
Title or Abstract			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4,5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4,5
	9	Whether participants formed a consecutive, random or convenience series	4-6
Test methods	**10a**	Index test, in sufficient detail to allow replication	5, Table 1, Supplementary Appendix
	10b	Reference standard, in sufficient detail to allow replication	6, Supplementary Appendix
	11	Rationale for choosing the reference standard (if alternatives exist)	6, Supplementary Appendix
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6, Supplementary Appendix
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6, Supplementary Appendix
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6, Supplementary Appendix
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7,8
	15	How indeterminate index test or reference standard results were handled	7,8
	16	How missing data on the index test and reference standard were handled	7,8
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	7,8
	18	Intended sample size and how it was determined	7,8
RESULTS			
Participants	**19**	Flow of participants, using a diagram	8
	20	Baseline demographic and clinical characteristics of participants	8
	21a	Distribution of severity of disease in those with the target condition	n/a
	21b	Distribution of alternative diagnoses in those without the target condition	n/a
	22	Time interval and any clinical interventions between index test and reference standard	n/a
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	9,10, Table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	9,10 Table 2
	25	Any adverse events from performing the index test or the reference standard	n/a
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	11-13
	27	Implications for practice, including the intended use and clinical role of the index test	11-13

OTHER INFORMATION			
	28	Registration number and name of registry	n/a
	29	Where the full study protocol can be accessed	14
	30	Sources of funding and other support; role of funders	13

BMJ Open

Validation of discharge diagnosis codes to identify serious infections among middle age and older adults

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020857.R2
Article Type:	Research
Date Submitted by the Author:	02-May-2018
Complete List of Authors:	Wiese, Andrew; Vanderbilt University Medical Center, Health Policy Griffin, Marie R; Vanderbilt University Medical Center, Health Policy Stein, Michael; Vanderbilt University, Pharmacology Schaffner, William; Vanderbilt University Medical Center, Health Policy Greevy, Robert; Vanderbilt School of Medicine, Biostatistics Mitchel, Jr., Edward; Vanderbilt University Medical Center, Health Policy Grijalva, Carlos; Vanderbilt University, Health Policy
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Infectious diseases, Epidemiology
Keywords:	coding algorithms, Medicaid, older adults, serious infections

Validation of discharge diagnosis codes to identify serious infections among middle age and older adults

Running title: Validation of diagnosis codes to identify infections

Authors: Andrew D. Wiese, PhD, MPH[1]; Marie R. Griffin[1,2], MD, MPH; C. Michael Stein, MB, ChB[3]; William Schaffner, MD[1]; Robert Greevy, PhD[4]; Edward F. Mitchel Jr., MS[1]; Carlos G. Grijalva, MD, MPH[1,2]

Affiliations: [1]Department of Health Policy, Vanderbilt University School of Medicine, Nashville, Tennessee, USA; [2]Mid-South Geriatric Research Education and Clinical Center, VA Tennessee Valley Health Care System, Nashville, Tennessee, USA; [3]Departments of Pharmacology and [4]Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

Corresponding Author: Andrew D. Wiese, PhD, MPH; Department of Health Policy, Vanderbilt University Medical Center, Suite 2600, Village at Vanderbilt, 1500 21st Avenue South, Nashville, TN 37212; phone: (615) 875-7997; email: andrew.d.wiese.1@vumc.org

Key words: coding algorithms; Medicaid; older adults; serious infections

Word Count: 3,777/4,000

Tables and Figures: (4/5)

ABSTRACT (277/300)

Objectives: Hospitalizations for serious infections are common among middle age and older adults and frequently used as study outcomes. Yet few studies have evaluated the performance of diagnosis codes to identify serious infections in this population. We sought to determine the positive predictive value (PPV) of diagnosis codes for identifying hospitalizations due to serious infections among middle age and older adults.

Setting and participants: We identified hospitalizations for possible infection among adults ≥50 years enrolled in the Tennessee Medicaid healthcare program (2008-2012) using ICD-9 diagnosis codes for pneumonia, meningitis/encephalitis, bacteremia/sepsis, cellulitis/soft-tissue infections, endocarditis, pyelonephritis and septic arthritis/osteomyelitis.

Design: Medical records were systematically obtained from hospitals randomly selected from a stratified sampling framework based on geographical region and hospital discharge volume.

Measures: Two trained clinical reviewers used a standardized extraction form to abstract information from medical records. Pre-defined algorithms served as reference to adjudicate confirmed infection-specific hospitalizations. We calculated the PPV of diagnosis codes using confirmed hospitalizations as reference. Sensitivity analyses determined the robustness of the PPV to definitions that required radiological or microbiological confirmation. We also determined interrater reliability between reviewers.

Results: The PPV of diagnosis codes for hospitalizations for infection (n=716) was 90% (95% CI: 88-92). The PPV was highest for pneumonia [97% (95% CI: 94-98)] and cellulitis [91% (95% CI: 85-95)], and lowest for meningitis/encephalitis [50% (95% CI: 24-76)]. The adjudication reliability was excellent [93% agreement; first agreement-coefficient: 0.91]. The overall PPV was lower when requiring microbiological confirmation [45%] and when requiring radiological confirmation for pneumonia [79%].

Conclusions: Discharge diagnosis codes have a high PPV for identifying hospitalizations for common, serious infections among middle age and older adults. PPV estimates for rare infections were imprecise.

STRENGTHS AND LIMITATIONS OF THE STUDY

- This study examined the performance of diagnosis coding algorithms to identify hospitalizations due to serious infections among middle age and older adults enrolled in a State Medicaid program using a systematic and representative sample of records from hospitals of different sizes and in distinct State regions.
- The reference criteria to identify true infections was based on the previous literature and clinical expertise but may be imperfect. Nevertheless, identifying microbiologically-confirmed infections is difficult due to the low sensitivity of culture-based diagnostic methods often used in clinical practice.
- Diagnosis codes were based on the ICD-9-coding system only. These findings will continue to be helpful for retrospective studies that encompass periods of ICD-9 use, yet additional studies evaluating the performance of ICD-10-based codes would be beneficial.
- Our coding algorithms to identify serious infections had a high positive predictive value overall, and will be useful in ongoing and future research using administrative data

INTRODUCTION

Infectious diseases remain a leading cause of morbidity and mortality in the U.S. and elsewhere.(1) Middle age and older adults, in particular, are at high risk for serious infections and their long-term consequences.(2, 3) Among older adults, community-acquired serious infections (including pneumonia, sepsis, and meningitis) often require hospitalization and represent a substantial burden on the U.S. healthcare system.(4-7) The incidence of community-acquired pneumonia is very high among adults ≥ 50 years of age (248 cases per 100,000 adults) with an even higher burden among adults ≥ 80 years of age (1,643 cases per 100,000 adults).(8) Sepsis, cellulitis and pyelonephritis are also very common (sepsis: 100 cases per 100,000 and cellulitis/pyelonephritis: >150 hospitalizations per 100,000 adults) with an increasing incidence of severe sepsis with increased age.(9-11) Meningitis and endocarditis are relatively rare (around 2-3 cases per 100,000), although the case fatality rate is very high.(12, 13) Therefore, it is important to monitor the incidence of these infections, identify important risk factors, and determine the impact of preventative policies (e.g., vaccination) on these diseases among middle age and older adults.(14-16)

Large-scale epidemiological studies using administrative data often use serious infections as outcomes.(17-21) However, few studies have evaluated the performance of diagnosis codes to identify serious infections among middle age and older adults. Most previous studies that have assessed the performance of coded discharge diagnosis codes to identify serious infections have focused mainly on common infections (e.g., pneumonia or sepsis), specific populations (e.g., patients with rheumatoid arthritis), or on healthcare-associated or hospital-acquired infections.(22-31) Nevertheless, the performance of coded discharge diagnoses for accurately identifying infections requiring hospitalization among middle age and older adults is unclear. Therefore, we sought to determine the positive predictive value (PPV) of specific discharge diagnoses for identifying infections that required hospitalization among middle age and older adults.

METHODS

Data sources

TennCare is the managed Medicaid program in the State of Tennessee that provides healthcare insurance to those who are Medicaid eligible (around 20% of the Tennessee population).(32) The adult TennCare population consists of low-income pregnant women and individuals who are elderly or have a disability (over 600,000 annually).(32) We used data from TennCare, supplemented with data from the Tennessee Hospital Discharge Data System (a registry for all hospitalizations in Tennessee) and pharmacy information from Medicare Part D for those that were dual eligible, to identify a retrospective cohort of TennCare enrollees ≥50 years of age with pharmacy benefits (2008-2012). We restricted the hospitalizations for serious infection to those occurring from 2008 through 2012 to only include more recent hospitalizations for which medical records are more likely to be available. Cohort members had at least 180 days of baseline continuous enrollment before cohort entry, and were also required to be free of certain life-threatening conditions known to increase the risk of infection (solid organ transplantation, end-stage renal disease, HIV/AIDS, malignancy and serious kidney, liver and respiratory disease) that may limit longitudinal follow-up and impact the assessments of patients' exposures and their risk of infections. Cohort members were also required to have evidence of at least one pharmacy prescription fill and evidence of at least one healthcare encounter during baseline (to ensure use of benefits so that if a healthcare encounter for an infection occurred, it would be detected). Follow-up started on the earliest date the inclusion criteria were met and continued through the earliest of the following: study end date (December 31, 2012), the day prior to diagnosis of a serious life-threatening condition that would have precluded entry to the cohort, loss of enrollment, or date of death. From this retrospective cohort, we identified possible hospitalizations for serious infections (see *Identification of hospitalizations for serious infection)* for our validation study. To avoid including infections that may have originated due to a previous hospital stay, we excluded hospitalizations for infections that occurred in the 30-day period after discharge from a previous hospitalization. The study was approved by the institutional review boards of Vanderbilt University and the Tennessee Department of Health, and by the Division of TennCare.

Identification of hospitalizations for serious infection

Clinical knowledge and a literature review were used to identify primary discharge diagnosis codes that have been used previously to identify specific serious infections that require hospitalization (*study infections)*, including pneumonia (alone or with a primary diagnosis of bacteremia/sepsis), bacteremia/sepsis, pyelonephritis, meningitis/encephalitis, osteomyelitis/septic arthritis, endocarditis and cellulitis.(31, 33-35) Specific International Classifications of Diseases-Clinical Modification, 9th-revision (ICD-9-CM) diagnosis codes used to identify possible hospitalizations for each infection type are presented in Table 1. As the objective of our study was to determine the PPV of coding algorithms to identify serious infections that required hospitalization, we focused only on primary diagnoses of infection to reduce the possibility of detecting concurrent infections that may not have led to hospitalization or nosocomial infections that developed during the course of the hospitalization.(35)

Sampling Strategy

We used stratified random sampling to select a representative subset of study infection hospitalizations from among all possible cases identified in the retrospective cohort from among hospitals within 200 miles of Vanderbilt University Medical Center (VUMC). Since larger hospitals would be over-represented in a purely random sampling, and because there may also be regional variability in coding practices and infection prevalence, we constructed a sampling framework where hospitals were stratified based on their geographic region in Tennessee (West, Central, and East), and tertiles of reported discharge volume (Low, Medium, and High) during the study period.(36-38) From this sampling framework, we randomly selected three hospitals from each of these nine sampling strata, and retrieved their medical records for review and validation (Figure 1). This strategy, relative to a purely random sample, ensured better representation of infections identified in smaller hospitals and those in more rural regions of the State of Tennessee. If a hospital refused to participate, it was replaced by another hospital randomly selected from the same sampling stratum.

The overall goal was to review and validate 675 hospitalizations for serious infection from 27 hospitals (25 hospitalizations for each of the 3 hospitals comprising a stratum, yielding 75 hospitalizations for each of the 9 strata) (Figure 1). We conservatively assumed that up to 80% of records requested would

be available for review, and so we requested 32 records per hospital to receive an average of 25 records from each (Figure 1). To ensure that we reviewed sufficient rare infections, we preferentially selected any identified possible hospitalizations for meningitis/encephalitis, osteomyelitis/septic arthritis and endocarditis from each hospital in the sample. We randomly selected the remaining set of possible hospitalizations for other serious infections based on the proportional distribution of common infections at each hospital (pneumonia, bacteremia/sepsis, pyelonephritis and cellulitis) until 32 infections were identified. For hospitals with fewer than 32 infections during the study period, all infections were requested.

Abstraction of Medical Records

Relevant clinical information was abstracted from the medical record (transfer notes, emergency room summary, admission summary, physical/history, pharmacy, laboratory, microbiology, and radiology information, and discharge summary) of each hospitalization with a primary discharge diagnosis code indicative of infection using a standardized and customized REDCap electronic data capture instrument hosted at Vanderbilt University.(39) As we were interested in infections that led to hospitalizations, we focused our reviews on clinical, microbiological and radiological information from the 2 days prior to the admission date through 2 days after admission to limit the possibility of identifying infections that developed during the hospitalization (i.e. nosocomial infections). In preparation for this study, the case report form was pilot-tested among a separate, convenience sample of 354 possible infections identified in the cohort from 3 hospitals in the same city as Vanderbilt University. This separate sample of hospitalizations was used only for pilot-testing the case report form, and was not included in the current study. One trained medical reviewer abstracted the relevant information for all selected records using the case report form. During the abstraction process, the lack of a particular finding in the medical record was treated as a lack of evidence for that finding, and so no information was considered missing after abstraction.

Adjudication of Medical Records

All records received were reviewed, abstracted and adjudicated. We made the final determination of whether a hospitalization represented a confirmed infection or not using *a priori* definitions of clinical, radiological, and/or microbiological findings compatible with infection for each infection type. Previous validation studies and expert clinical knowledge were used to define these specific *a priori* definitions for each infection type (*Supplementary appendix)*.(31, 35, 40)

Statistical analysis

We calculated the PPV of the ICD-9-CM discharge diagnosis codes for identifying hospitalizations for serious infection using the results of the *a priori* definitions applied to the information abstracted from the medical records as the reference (the proportion of cases identified with discharge diagnosis codes that were determined to be true cases after adjudication of the medical record information). We calculated 95% confidence intervals for the PPV using Wilson's formula.(41) Secondary analyses assessed the PPV for hospitalizations for serious infection across hospitals of different sizes and in different geographical regions of Tennessee.

We also assessed the reliability of the abstraction process. A second trained medical reviewer abstracted relevant information from a subset of selected records, which included all meningitis and endocarditis records, and a random selection of 10% of each of the remaining infection types. Each reviewer conducted the process independently and blinded from one another. For the subset of records abstracted by both reviewers, inter-reviewer agreement for the adjudication of a true or mis-identified infection was assessed using the Gwet's first agreement coefficient (AC_1).(42-44) Since Cohen's kappa statistic can be unreliable when the prevalence of the event and the level of observer agreement are high in the study sample, we used Gwet's AC_1 as a reliability measure unlikely to be affected by these concerns.(44-46)

In planned sensitivity analyses, we first assessed the impact of excluding hospitalizations that occurred after the individual was transferred from another healthcare facility, as initial documentation and details of the infection could be missing or incomplete in the receiving hospital.(40) We also assessed the impact on the PPV for all infections when requiring microbiological identification of a pathogen

(excluding common contaminants) from a sterile site within 2 days before or after the hospitalization admission date. A final sensitivity analysis among hospitalizations for possible pneumonia assessed the PPV when radiological evidence of pneumonia was required [i.e. pneumonia, opacity, or infiltrate mentioned in a chest X-ray or computed tomography scan report] (*Supplementary appendix*). All analyses were performed in Stata-IC, version 15.1 (College Station TX).

Patient and Public Involvement

No patients were involved in the development of the research question, the outcome measures, or the design or conduct of the study. As we conducted a retrospective study using administrative data, we have no plans to disseminate the results of the research to study participants.

RESULTS

Cohort characteristics

Among a retrospective cohort of 129,465 adults ≥50 years of age enrolled in TennCare, 9,769 hospitalizations for serious infection were identified during the study period (2008-2012) among 7,770 unique patients (Figure 2). Cohort members were primarily female (57.8%) with a median age of 54 years (mean: 57 years; range: 50-110). For efficiency considerations, our medical chart review activities then focused on hospitalizations for serious infection (n=8,322) that occurred at hospitals within 200 miles of VUMC. Pneumonia, cellulitis and bacteremia/sepsis were the most common infections identified using discharge diagnosis codes (54.3%, 20.5% and 18.4%, respectively), followed by pyelonephritis (3.8%) and septic arthritis/osteomyelitis (2.5%). Fewer than 1% of hospitalizations were due to meningitis/encephalitis (n=30) and endocarditis (n=18).

Collection, review and adjudication of selected medical records

Of the 27 hospitals that were initially selected for the sample, 21 (78%) were able to participate. We selected 7 additional hospitals to replace the 6 non-participants to achieve the desired sample size, including an additional small hospital in the East region due to a large number of unavailable records from a single participating hospital.

We received 716 (88.6%) of 808 requested records from 28 participating hospitals [Table 2]. Record availability from participating hospitals was lower in medium size hospitals (81.8%) compared to small (93.5%) and large hospitals (91.7%), but did not differ by geographic region. Record availability by infection type was greater than 86% for all infection types, with the exception of hospitalizations for the rare endocarditis cases (57.1%; only 4 of 7 cases).

The sample of hospitalizations for serious infection included patients who were primarily female (63.6%), with a median age of 60 years (mean: 64 years; range: 50-101) at the time of hospitalization. There was evidence of transfer from a prior healthcare facility for 21.8% of the hospitalizations for serious infection [highest percentage of transfers for bacteremia/sepsis (38.5%) and pneumonia (25.1%)]. The most common healthcare facility source was a nursing home/skilled nursing facility (84.6%), but also included group home sources (7.7%), other sources (4.5%) [assisted living facility, mental health center] and another acute care hospital (3.2%). There was evidence of an emergency department visit within 7 days prior to admission date for the serious infection hospitalization in 4.8% of the records.

Performance of discharge diagnosis codes

A total of 646 [PPV: 90.2% (95% CI: 87.8-92.2)] of the hospitalizations for serious infection identified using ICD-9-CM primary discharge diagnosis codes were confirmed by applying the *a priori* definitions to the abstracted data. The PPV was highest for pneumonia and cellulitis [96.5% (95% CI: 93.9-98.0) and 91.1% (95% CI: 84.7-94.9), respectively], and was ≥75% for bacteremia/sepsis, pyelonephritis, septic arthritis/osteomyelitis, and endocarditis. The PPV was lowest for meningitis/encephalitis [50.0% (95% CI: 23.7-76.3)], although the precision was limited due to a low number of available records for review (Table 2). Among the 10 potential cases of meningitis/encephalitis, 7 cases were meningitis/meningoencephalitis and 3 were encephalitis. The respective PPVs for meningitis/meningoencephalitis and encephalitis were 71.4% (95% CI: 35.9-91.8) and 0%, respectively.

When performance was evaluated across stratification sampling parameters, no apparent differences were observed in the PPV for records from hospitals in different geographical regions of

Tennessee. Although the PPV was high for all three discharge volume groups, the PPV was significantly lower in large hospitals [84.6% (95% CI: 79.6-88.5)] compared to smaller hospitals [93.9% (95% CI: 90.0-96.3); PPV difference: -9.3% (95% CI: -14.7, -3.9)] and medium hospitals [92.7% (95% CI: 88.6-95.4); PPV difference: -8.1% (95% CI: -13.7, -2.6)] (Table 2). This was likely driven by the different distributions in the types of infections selected for review in the hospital groups. Large hospitals had a higher proportion of non-pneumonia infections (70.4%) compared to medium and small hospitals (49.4% and 36.1%, respectively). Importantly, the PPV for pneumonia was similar in each discharge volume group (range: 96.0 to 96.6%), whereas the PPV was smaller for non-pneumonia infections in large hospitals (79.8%) compared to medium (88.7%) and small (89.2%) hospitals.

In the 82 records independently abstracted by two reviewers to assess reliability, there was 92.7% (95% CI: 86.9-98.4) agreement for identifying true hospitalizations for serious infection. The inter-rater agreement was also high when assessing reliability, independent of the outcome prevalence, with an AC_1 of 0.91 (95% CI: 0.84-0.99). Of the 6 discordant cases, 3 were meningitis/encephalitis (1 meningitis/meningoencephalitis and 2 encephalitis), with one each of bacteremia/sepsis, pyelonephritis and septic arthritis. The main reason for a discrepancy between reviewers was whether or not treatment for the infection of interest occurred within 2 days of the admission date, which was one of the major criteria for adjudication (see *Supplementary appendix*).

Sensitivity analyses

The PPV was virtually unchanged when excluding the 21.8% of hospitalizations that occurred as transfers from another healthcare facility [90.1% (95% CI: 87.4-92.3)]. Microbiological evidence of the specific infection type was found in 47.6% of records, leading to reduced PPVs when requiring microbiological evidence [45.4% (95% CI: 41.8-49.1)]. Microbiological evidence of infection was highest in hospitalizations for suspected pyelonephritis (94.4%), but was ≤60% for every other infection type [pneumonia (42.7%); cellulitis/soft tissue infections (58.5%); bacteremia/sepsis (26.1%)]. When requiring radiological confirmation of pneumonia, the PPV for coded diagnoses was 78.8% (95% CI: 74.2-82.8). Approximately 95.6% of possible hospitalizations for pneumonia had at least one documented

chest x-ray or CT-scan. Among those patients with a chest x-ray or CT-scan report available (n=325), 83.4% had a finding compatible with pneumonia. The main findings among the 54 patients with possible pneumonia and a radiological report available, but without radiological confirmation of pneumonia included atelectasis (n=6), interstitial pneumonitis (n=3), chronic heart failure with pulmonary edema (n=1), and no radiological findings of any kind (n=44).

DISCUSSION

Discharge diagnoses for identifying hospitalizations due to serious infections among middle age and older adults had an overall PPV of 90.2%, with the highest values for the identification of common serious infections. PPVs were consistently high across different hospital types and regions of Tennessee. Furthermore, the PPV was similar after exclusion of hospitalizations for serious infection that were the result of a transfer from another healthcare facility (e.g. acute care hospital, skilled nursing facility). Microbiological confirmation was available for fewer than 50% of patients admitted with possible hospitalizations for serious infections, and as expected, the requirement resulted in a low PPV for all infections, with the exception of pyelonephritis. Importantly, the PPV for pneumonia hospitalizations remained relatively high even when requiring radiological confirmation.

The PPV for hospitalizations for pneumonia in previous smaller validation studies has ranged from 72 to 86% in different healthcare systems, but those studies were not focused on middle age and older adults.(31, 47-49) In our study of hospitalizations among middle age and older adults, we found that coded discharge diagnoses have a higher PPV for pneumonia compared to previous studies. The PPV for bacteremia/sepsis was also on the higher range of previously reported PPVs for diagnosis codes to identify bacteremia/sepsis from administrative data in other populations (reported range from 45% to 97.7%), and for septic arthritis/osteomyelitis compared to a previous study conducted among patients with diabetes (63.9% versus 75.9% in our study).(23, 50, 51) Overall, the observed PPV for all infections in our study was comparable to two previous comprehensive validation studies of bacterial infections, one among patients with rheumatoid arthritis in a single hospital system and another among patients in one of

the Veteran's Affairs integrated service networks.(35, 40) Compared to these two previous studies of ICD-9 codes, we abstracted and adjudicated a larger number of records while using a more systematic sampling strategy to retrieve and review records for hospitalizations from multiple regions and hospital types as opposed to a single hospital or healthcare system. However, some of the PPVs for individual infections were less precise and less similar to these previous studies. This was especially true for rare infections, as would be expected due to the low numbers of rare infections in our study and across previous studies.(35, 40) The results of our study are also similar to previous validation studies that used corresponding ICD-10 diagnosis codes to identify hospitalizations for serious infection.(52, 53)

One limitation to consider in our study was that it was not designed to estimate the sensitivity and specificity of the coding algorithms. This would have required the identification, review and adjudication of a sample of hospitalizations that did not fulfill our algorithm (i.e. presence of the ICD-9 primary discharge diagnosis codes indicative of infection). However, when the prevalence of an outcome is low, the PPV approximates the specificity.(54) Importantly, any non-differential outcome misclassification between exposure groups resulting from the use of imperfect but highly-specific measurements would attenuate the impact of the misclassification on the relative risk estimates.(55) In addition, we found that the PPV of coded discharge diagnoses for serious infections remained high across hospitals of different sizes and across different geographical areas of Tennessee, which may have different rates of hospitalizations for serious infection.(56) Although our study applied a systematic sampling strategy to assure the representation of different settings in our population, our population was restricted to middle age and older adults enrolled in a State Medicaid program. Therefore, caution is warranted when extrapolating the study findings to other populations.

Another limitation is the use of available clinical information to operationalize definitions for adjudication of true hospitalizations for infections. It is possible that some procedures, laboratory findings and diagnoses that informed the final diagnosis of infection were not fully recorded in the medical records, and thus, were not available for our review and may have contributed to the observed PPV for some infections. Although we used previous validation studies and clinical information to build pre-

specified definitions for the adjudication of true infections, our reference criteria may be imperfect, considering the retrospective nature of our determinations and potential variability in clinical practice. Nevertheless, we also assessed how the availability of selected findings (i.e. microbiological and radiological information) in the medical record impacted the overall and infection-specific PPV. We demonstrated that relying on highly specific clinical diagnostics, such as microbiological and radiological information, to confirm true infections would result in lower PPVs for identification of infections in administrative data. Requiring microbiological confirmation to confirm true infections is challenging because of the known low sensitivity of culture-based diagnostic methods (most commonly used in clinical practice), which may lead to misclassification.(57, 58) In addition, requiring radiological evidence compatible with pneumonia within 2 days of hospital admission did lower the observed PPV for pneumonia hospitalizations. Nevertheless, the observed PPV remained close to 80%, which should reduce concerns about using diagnosis codes to identify hospitalizations due to pneumonia. Finally, the coding algorithms were based on the ICD-9-coding system only. Although these findings will be helpful for retrospective studies that encompass periods of ICD-9 use, additional studies evaluating the performance of ICD-10-based codes would be useful to complement our findings.

Our study demonstrated that discharge diagnosis codes can be used to accurately identify hospitalizations for serious infections among middle age and older adults. The highest PPVs were observed for the most common infections, and the PPV for pneumonia remained high when requiring radiological confirmation. Importantly, consistently high PPVs were observed across different hospital sizes and regions. However, the estimated PPV was lower and less precise for very rare infections (e.g. encephalitis). This should be an important consideration for studies specifically focused on those less frequent outcomes, especially when strict microbiological confirmation is required. Taken together, these findings support the use of discharge diagnosis codes for infections to identify outcomes in ongoing and future epidemiological studies among middle age and older adults.

Acknowledgement

We are indebted to the Tennessee Bureau of TennCare of the Tennessee Department of Finance and Administration, which provided data for the study. We are also indebted to the Tennessee Department of Health for providing data for the study. Statements in the report should not be construed as endorsement by the U.S. Department of Health and Human Services, the Department of Veterans Affairs, or the Tennessee Department of Health.

Funding

This study was funded by the NIH (R03-AG-042981 and R01-AG-043471-01A1) and the TL1 award TL1TR000447.

Contributors

ADW planned the medical record collection and statistical analysis, analyzed and interpreted the data, and drafted and revised the paper. MRG, WS, CMS, and RAG planned the statistical analysis, interpreted the data and revised the paper. EFM prepared the data, and revised the paper. CGG designed the project, acquired the data from TennCare, planned the medical record collection and statistical analysis, interpreted the data, and revised the paper.

Declaration of interests

CGG has received consulting fees from Pfizer and Merck, and received research support from Sanofi-Pasteur, Campbell Alliance, the Centers for Disease Control and Prevention, National Institutes of Health, The Food and Drug Administration, and the Agency for Health Care Research and Quality. WS has received personal fees from Pfizer, Merck, Novavax, Dynavax, Sanofi-Pasteur, GSK, Seqirus, and received research support from the Centers for Disease Control and Prevention. ADW, EFM, CMS, MRG and RAG have no conflicts of interest to disclose.

Data sharing

No additional unpublished data are available from the study. The study protocol and statistical code are available from the corresponding author, Andrew Wiese (andrew.d.wiese.1@vumc.org).

REFERENCES

1. Khabbaz RF, Moseley RR, Steiner RJ, Levitt AM, Bell BP**.** Challenges of infectious diseases in the USA. Lancet. 2014;384(9937):53-63.

2. Grohskopf LA, Sokolow LZ, Broder KR, Olsen SJ, Karron RA, Jernigan DB, et al. Prevention and Control of Seasonal Influenza with Vaccines. MMWR Recomm Rep. 2016;65(5):1-54.

3. Kim DK, Bridges CB, Harriman KH**.** Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older--United States, 2015. MMWR Morb Mortal Wkly Rep. 2015;64(4):91-2.

4. Jeon MH, Chung JW, Choi SH, Kim TH, Lee EJ, Choo EJ**.** Pneumonia risk factors and clinical features of hospitalized patients older than 15 years with pandemic influenza A (H1N1) in South Korea: a multicenter study. Diagn Microbiol Infect Dis. 2011;70(2):230-5.

5. Pilishvili T, Bennett NM**.** Pneumococcal Disease Prevention Among Adults: Strategies for the Use of Pneumococcal Vaccines. Am J Prev Med. 2015;49(6 Suppl 4):S383-90.

6. Sjoding MW, Prescott HC, Wunsch H, Iwashyna TJ, Cooke CR**.** Longitudinal Changes in ICU Admissions Among Elderly Patients in the United States. Crit Care Med. 2016;44(7):1353-60.

7. Crotty MP, Meyers S, Hampton N, Bledsoe S, Ritchie DJ, Buller RS, et al. Epidemiology, Co-Infections, and Outcomes of Viral Pneumonia in Adults: An Observational Cohort Study. Medicine (Baltimore). 2015;94(50):e2332.

8. Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, et al. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. N Engl J Med. 2015;373(5):415-27.

9. Filbin MR, Arias SA, Camargo CA, Jr., Barche A, Pallin DJ**.** Sepsis visits and antibiotic utilization in U.S. emergency departments*. Crit Care Med. 2014;42(3):528-35.

10. Christensen KL, Holman RC, Steiner CA, Sejvar JJ, Stoll BJ, Schonberger LB**.** Infectious disease hospitalizations in the United States. Clin Infect Dis. 2009;49(7):1025-35.

11. Ginde AA, Moss M, Shapiro NI, Schwartz RS. Impact of older age and nursing home residence on clinical outcomes of US emergency department visits for severe sepsis. J Crit Care. 2013;28(5):606-11.

12. Iung B, Vahanian A. Epidemiology of acquired valvular heart disease. Can J Cardiol. 2014;30(9):962-70.

13. Thigpen MC, Whitney CG, Messonnier NE, Zell ER, Lynfield R, Hadler JL, et al. Bacterial meningitis in the United States, 1998-2007. N Engl J Med. 2011;364(21):2016-25.

14. Simonetti AF, Garcia-Vidal C, Viasus D, Garcia-Somoza D, Dorca J, Gudiol F, et al. Declining mortality among hospitalized patients with community-acquired pneumonia. Clin Microbiol Infect. 2016;22(6):567.e1-7.

15. Rivero-Calle I, Pardo-Seco J, Aldaz P, Vargas DA, Mascaros E, Redondo E, et al. Incidence and risk factor prevalence of community-acquired pneumonia in adults in primary care in Spain (NEUMO-ES-RISK project). BMC Infect Dis. 2016;16(1):645.

16. Park JY, Park S, Lee SH, Lee MG, Park YB, Oh KC, et al. Microorganisms Causing Community-Acquired Acute Bronchitis: The Role of Bacterial Infection. PLoS One. 2016;11(10):e0165553.

17. Jackson ML, Walker R, Lee S, Larson E, Dublin S. Predicting 2-Year Risk of Developing Pneumonia in Older Adults without Dementia. J Am Geriatr Soc. 2016;64(7):1439-47.

18. Clement RC, Haddix KP, Creighton RA, Spang JT, Tennant JN, Kamath GV. Risk Factors for Infection After Knee Arthroscopy: Analysis of 595,083 Cases From 3 United States Databases. Arthroscopy. 2016;32(12):2556-61.

19. Weinert BA, Edmonson MB. Hospitalizations at Nonfederal Facilities for Lower Respiratory Tract Infection in American Indian and Alaska Native Children Younger than 5 Years of Age, 1997-2012. J Pediatr. 2016;175:33-9.e4.

20. Akiyama T, Chikuda H, Yasunaga H, Horiguchi H, Fushimi K, Saita K. Incidence and risk factors for mortality of vertebral osteomyelitis: a retrospective analysis using the Japanese diagnosis procedure combination database. BMJ Open. 2013;3(3).

21. Graversen ME, Dalgaard LS, Jensen-Fangel S, Jespersen B, Ostergaard L, Sogaard OS. Risk and outcome of pyelonephritis among renal transplant recipients. BMC Infect Dis. 2016;16:264.

22. Abou Zahr Z, Spiegelman A, Cantu M, Ng B. Perioperative use of anti-rheumatic agents does not increase early postoperative infection risks: a Veteran Affairs' administrative database study. Rheumatol Int. 2015;35(2):265-72.

23. Carnahan RM, Herman RA, Moores KG. A systematic review of validated methods for identifying transfusion-related sepsis using administrative and claims data. Pharmacoepidemiol Drug Saf. 2012;21 Suppl 1:222-9.

24. Condell O, Gubbels S, Nielsen J, Espenhain L, Frimodt-Moller N, Engberg J, et al. Automated surveillance system for hospital-acquired urinary tract infections in Denmark. J Hosp Infect. 2016;93(3):290-6.

25. Gedeborg R, Furebring M, Michaelsson K. Diagnosis-dependent misclassification of infections using administrative data variably affected incidence and mortality estimates in ICU patients. J Clin Epidemiol. 2007;60(2):155-62.

26. Goto M, Ohl ME, Schweizer ML, Perencevich EN. Accuracy of administrative code data for the surveillance of healthcare-associated infections: a systematic review and meta-analysis. Clin Infect Dis. 2014;58(5):688-96.

27. Shaklee J, Zerr DM, Elward A, Newland J, Leckerman K, Asti L, et al. Improving surveillance for pediatric Clostridium difficile infection: derivation and validation of an accurate case-finding tool. Pediatr Infect Dis J. 2011;30(3):e38-40.

28. Wright SB, Huskins WC, Dokholyan RS, Goldmann DA, Platt R. Administrative databases provide inaccurate data for surveillance of long-term central venous catheter-associated infections. Infect Control Hosp Epidemiol. 2003;24(12):946-9.

29. Greenberg JA, Hohmann SF, Hall JB, Kress JP, David MZ. Validation of a Method to Identify Immunocompromised Patients with Severe Sepsis in Administrative Databases. Ann Am Thorac Soc. 2016;13(2):253-8.

30. Olsen MA, Young-Xu Y, Stwalley D, Kelly CP, Gerding DN, Saeed MJ, et al. The burden of clostridium difficile infection: estimates of the incidence of CDI from U.S. Administrative databases. BMC Infect Dis. 2016;16:177.

31. Grijalva CG, Chung CP, Stein CM, Gideon PS, Dyer SM, Mitchel EF, Jr., et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. Pharmacoepidemiol Drug Saf. 2008;17(9):890-5.

32. TennCare Fiscal Year 2015-2016 Annual Report. In: Administration DoHCF, ed; 2016.

33. Wiese AD, Griffin MR, Stein CM, Mitchel EF, Jr., Grijalva CG**.** Opioid Analgesics and the Risk of Serious Infections Among Patients With Rheumatoid Arthritis: A Self-Controlled Case Series Study. Arthritis Rheumatol. 2016;68(2):323-31.

34. Grijalva CG, Kaltenbach L, Arbogast PG, Mitchel EF, Jr., Griffin MR**.** Initiation of rheumatoid arthritis treatments and the risk of serious infections. Rheumatology (Oxford). 2010;49(1):82-90.

35. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH**.** Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. J Clin Epidemiol. 2007;60(4):397-409.

36. Brenner H, Gefeller O**.** Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Stat Med. 1997;16(9):981-91.

37. Kim HM, Smith EG, Stano CM, Ganoczy D, Zivin K, Walters H, et al. Validation of key behaviourally based mental health diagnoses in administrative data: suicide attempt, alcohol abuse, illicit drug abuse and tobacco use. BMC Health Serv Res. 2012;12:18.

38. El-Ghitany EM, Farghaly AG, Farag S, Abd El-Wahab EW**.** Validation of EGCRISC for Chronic Hepatitis C Infection Screening and Risk Assessment in the Egyptian Population. PLoS One. 2016;11(12):e0168649.

39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-81.

40. Patkar NM, Curtis JR, Teng GG, Allison JJ, Saag M, Martin C, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. J Clin Epidemiol. 2009;62(3):321-7, 7.e1-7.

41. Niesner K, Murff HJ, Griffin MR, Wasserman B, Greevy R, Grijalva CG, et al. Validation of VA administrative data algorithms for identifying cardiovascular disease hospitalization. Epidemiology. 2013;24(2):334-5.

42. Govatsmark RE, Sneeggen S, Karlsaune H, Slordahl SA, Bonaa KH. Interrater reliability of a national acute myocardial infarction register. Clin Epidemiol. 2016;8:305-12.

43. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol. 2013;13:61.

44. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;61(Pt 1):29-48.

45. Lantz CA, Nebenzahl E. Behavior and interpretation of the kappa statistic: resolution of the two paradoxes. J Clin Epidemiol. 1996;49(4):431-4.

46. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol. 1990;43(6):543-9.

47. Drahos J, Vanwormer JJ, Greenlee RT, Landgren O, Koshiol J. Accuracy of ICD-9-CM codes in identifying infections of pneumonia and herpes simplex virus in administrative data. Ann Epidemiol. 2013;23(5):291-3.

48. Aronsky D, Haug PJ, Lagor C, Dean NC. Accuracy of administrative data for identifying patients with pneumonia. Am J Med Qual. 2005;20(6):319-28.

49. Kern DM, Davis J, Williams SA, Tunceli O, Wu B, Hollis S, et al. Validation of an administrative claims-based diagnostic code for pneumonia in a US-based commercially insured COPD population. Int J Chron Obstruct Pulmon Dis. 2015;10:1417-25.

50. Iwashyna TJ, Odden A, Rohde J, Bonham C, Kuhn L, Malani P, et al. Identifying patients with severe sepsis using administrative claims: patient-level validation of the angus implementation of the international consensus conference definition of severe sepsis. Med Care. 2014;52(6):e39-43.

51. Newton KM, Wagner EH, Ramsey SD, McCulloch D, Evans R, Sandhu N, et al. The use of automated data to identify complications and comorbidities of diabetes: a validation study. J Clin Epidemiol. 1999;52(3):199-207.

52. Sahli L, Lapeyre-Mestre M, Derumeaux H, Moulis G**.** Positive predictive values of selected hospital discharge diagnoses to identify infections responsible for hospitalization in the French national hospital database. Pharmacoepidemiol Drug Saf. 2016;25(7):785-9.

53. Holland-Bill L, Xu H, Sorensen HT, Acquavella J, Svaerke C, Gammelager H, et al. Positive predictive value of primary inpatient discharge diagnoses of infection among cancer patients in the Danish National Registry of Patients. Ann Epidemiol. 2014;24(8):593-7, 7.e1-18.

54. Schneeweiss S, Avorn J**.** A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005;58(4):323-37.

55. Rodgers A, MacMahon S**.** Systematic underestimation of treatment effects as a result of diagnostic test inaccuracy: implications for the interpretation and design of thromboprophylaxis trials. Thromb Haemost. 1995;73(2):167-71.

56. van Walraven C, English S, Austin PC**.** Administrative database code accuracy did not vary notably with changes in disease prevalence. J Clin Epidemiol. 2016;79:86-9.

57. Mancini N, Carletti S, Ghidoli N, Cichero P, Burioni R, Clementi M**.** The era of molecular and other non-culture-based methods in diagnosis of sepsis. Clin Microbiol Rev. 2010;23(1):235-51.

58. Niederman MS**.** The argument against using quantitative cultures in clinical trials and for the management of ventilator-associated pneumonia. Clin Infect Dis. 2010;51 Suppl 1:S93-9.

Table 1. Discharge diagnosis code definitions (ICD-9-CM) for hospitalizations for serious infection

Serious Infection	Primary (first listed) discharge diagnosis code
Pneumonia-primary definition	003.22, 480.*‡, 481, 482.*, 483.*, 484.*, 485.*, 486.*, 487.0
Pneumonia-secondary definition (primary diagnosis code with pneumonia diagnosis (above) in any other diagnosis field)	510.*, 038.*, 790.7, 995.91, 995.92
Meningitis/ Encephalitis	003.21, 036.0, 0.47*, 049.*, 053.0, 054.72, 072.1, 091.81, 094.2, 098.82, 100.81, 320.*, 036.1, 054.3, 056.01, 058.21, 058.29, 062.*, 063.*, 064.*, 066.41, 072.2, 094.81, 130.0, 323.*
Bacteremia/ Sepsis†	038.*, 790.7, 995.91, 995.92
Cellulitis/ Soft-tissue infections	035, 040.0, 569.61, 681.*, 682.*, 728.86, 785.4
Endocarditis	036.42, 074.22, 093.2*, 098.84, 421.*, 422.92
Pyelonephritis	590.*
Septic Arthritis/ Osteomyelitis	003.23, 056.71, 098.5*, 711.0, 711.00-711.07, 711.09, 711.9*, 003.24, 376.03, 526.4, 730.0*, 730.1*, 730.2*

‡ A * indicates all numeric values [0-9]

† Without a diagnosis of pneumonia in any other diagnosis field

Table 2. Positive predictive value (PPV) of coded discharge diagnosis definitions for hospitalizations for serious infections among adults ≥ 50 years of age enrolled in Tennessee Medicaid, 2008-2012

Type	Expected Number of Records	Records Received	PPV (95 % CI)
Overall	675	716	90.2 (87.8, 92.2)
Region Specific			
West	225	195	91.3 (86.5, 94.5)
Central	225	225	88.9 (84.1, 92.4)
East	225	296	90.5 (86.7, 93.4)
Bed volume size specific			
Low	225	230	93.9 (90.0, 96.3)
Medium	225	233	92.7 (88.6, 95.4)
High	225	253	84.6 (79.6, 88.5)
Serious Infection			
Pneumonia	305	340	96.5 (93.9, 98.0)
Cellulitis/Soft-tissue infections	125	123	91.1 (84.7, 94.9)
Pyelonephritis	80	89	87.6 (79.2, 93.0)
Bacteremia/Sepsis	100	92	82.6 (73.6, 89.0)
Septic Arthritis/Osteomyelitis	50	58	75.9 (63.5, 85.0)
Endocarditis	5	4	75.0 (30.1, 95.4)
Meningitis/Encephalitis	10	10	50.0 (23.7, 76.3)

Figure legends

Figure 1. Sampling strategy for identifying potential hospitalizations for serious infection

Figure 2. Identifying a retrospective cohort of patients ≥50 years of age without serious/life-threatening conditions, Tennessee Medicaid (2008-2012)

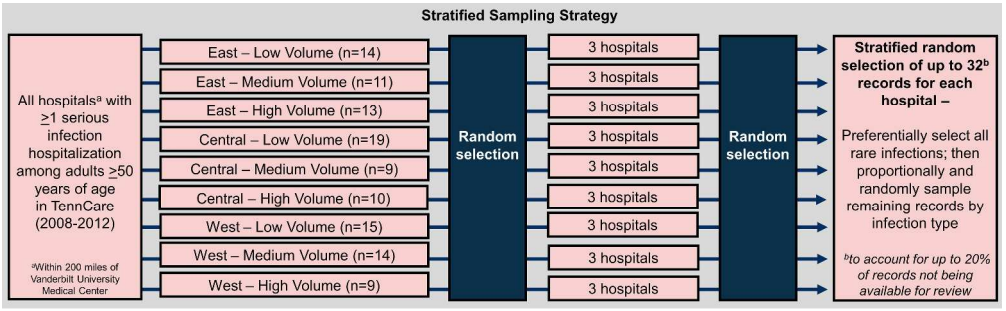

Figure 1. Sampling strategy for identifying potential hospitalizations for serious infection

457x139mm (300 x 300 DPI)

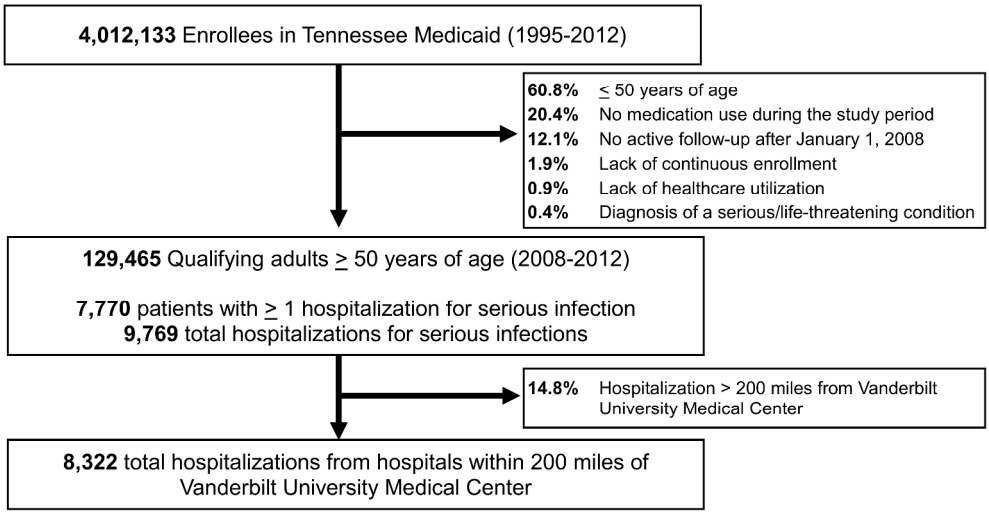

Figure 2. Identifying a retrospective cohort of patients ≥50 years of age without serious/life-threatening conditions, Tennessee Medicaid (2008-2012)

355x190mm (300 x 300 DPI)

Supplementary Appendix

Infection-Specific Definitions of Hospitalization for Serious Infection

We used a pre-specified adjudication process to determine whether each abstracted medical record corresponded to a true infection or not. Previous validation studies and expert clinical knowledge were used to define specific a priori definitions for each infection type.[1-3] Information abstracted from the medical record was compared to these *a priori* definitions for each infection type to make the final determination of whether a hospitalization represented a true infection or not.

Outline

I. Sepsis/Septicemia/Bacteremia/Septic Shock/Generalized Infection

Either of the following [1 or 2]:

1. **Positive culture of a non-contaminant pathogen**
    - i. Positive blood culture [any of the following (1-2)]
        1. Any gram-negative organism, except:
            a. No predominant organism
        2. A gram positive organism, except:
            a. Coagulase-negative *Staphylococcus*
            b. *Bacillus spp.* (other than *Bacillus anthracis*)
            c. *Corynebacterium spp.*
            d. *Propionibacterium spp.*
            e. *Micrococcus*
            f. Diptheroids
            g. Viridians Group Streptococci
            h. Enterococci
            i. *Clostridium perfringens*
            j. *Aerococcus*
            k. *Alcaligenes faecalis*
            l. *Citrobacter*
            m. *Neisseria subflava*
            n. *Stomatococcus*
            o. *Streptococcus bovis*
            p. *Veillonella candidemia*
            q. *Mycobacterium tuberculosis*
            r. *S. salivarius*
            s. *"Gram Positive"*
            t. *"No predominant organism"*
            u. *Streptococcus alpha*
2. **At least two of the following, documented at admission +/- 2 days [i-iii]**
    - i. Hypotension
        1. Systolic BP $\leq$ 90 mmHg
        2. Reduction of systolic BP of 40mmHg from earliest measurement collected during the admission of interest
    - ii. Two of the following [1-4]:
        1. Temperature $\geq$ 38^0C **or** $\leq$ 36^0C
        2. Heart rate $\geq$ 90 beats/minute
        3. Respiratory rate $\geq$ 20 breaths/min or $PaCO_2$ < 32 mmHg
        4. WBC $\geq$ 10,000 cells/mm^3 **or** $\leq$ 4,500 cells/mm^3 **or** WBC with > 10 % immature (band) forms
    - iii. Initiation of antibiotic treatment specifically for sepsis/septicemia/bacteremia/septic shock/generalized infection

II. Pneumonia

1. Pneumonia identified through examination (**all three of the following [a-c]):**
    a. One of the following admission findings indicative of respiratory findings:
        1. New and/or increased cough
        2. Shortness of breath
        3. Pleuritic chest pain
        4. New purulent production
        5. Altered mental status ("agitation" and "lethargy" included)
        6. Crackles
            a. Physical evidence of consolidation such as egophony, whispered pectoriloquy, etc.
    b. One of the following examination findings indicative of systemic infection [1-4]:
        1. Temperature (T $\geq$ 100.4^{0}F (38^{0}C) or $\leq$ 96^{0}F) in first 48 hours of admission
        2. Systolic BP $\leq$ 90mmHg
        3. Shock
            a. Volume nonresponsive hypotension
        4. Blood peripheral WBC ($\geq$ 10.0 x 10^9/L or $\leq$ 4.5 x 10^9/L)
    c. Treatment with antibiotics/antivirals indicated for suspected infection

OR

At least two of the following [1-3]:

1. Two of the following from #1 **([a and b], [a and c], or [b-c])**
2. Any of the following findings listed on chest imaging from radiologic report **documented at admission +/- 2 days**
    a. Pneumonia
    b. Lung abscess
    c. Opacity consistent with pneumonia/lung abscess
    d. Infiltrate consistent with pneumonia/lung abscess
    e. Consolidation consistent with pneumonia/lung abscess
    f. Increased density consistent with pneumonia/lung abscess
    g. Pleural effusion consistent with pneumonia/lung abscess
    h. Interstitial edema consistent with pneumonia/lung abscess
3. Sterile Site Laboratory Findings
    i. Any one of the following [i through v]
        i. Sputum lab findings [any **one** of the following (1, 2)]:
            1. Sputum culture/PCR/serology/gram stain positive for an agent that is not considered a contaminant [see exclusion list below]:
                a. *Aspergillus* species, *Enterococcus* species, viridians group streptococci, and yeast
            2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen
        ii. Blood lab findings [either of the following (1-3)]
            1. Blood culture/PCR/serology positive for an agent that is not considered a contaminant [see exclusion list below]:
                a. Exclusions
                    i. Coagulase-negative *Staphylococcus*

ii. *Bacillus spp.* (other than *Bacillus anthracis*)
iii. *Corynebacterium spp.*
iv. *Propionibacterium spp.*
v. *Micrococcus*
vi. Diptheroids
vii. Viridians Group Streptococci
viii. Enterococci
ix. *Clostridium perfringens*
x. *Aerococcus*
xi. *Alcaligenes faecalis*
xii. *Citrobacter*
xiii. *Neisseria subflava*
xiv. *Stomatococcus*
xv. *Streptococcus bovis*
xvi. *Veillonella candidemia*
xvii. *Mycobacterium tuberculosis*
xviii. *S. salivarius*

2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen

iii. Pleural fluid lab findings [either of the following (1, 2)]
   1. Culture/PCR/serology positive for a bacterial pathogen
   2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen
iv. Bronchoscopic specimen or deep endotracheal tube aspiration lab findings [either of the following (1, 2)]
   1. Culture/PCR/serology positive for a bacterial pathogen
   2. Positive viral study (culture/PCR/antigen screen) for a viral pathogen
v. Urine antigen detection testing [either of the following (1, 2)]
   1. *Legionella pneumophila*
   2. *Streptococcus pneumoniae*

III. Cellulitis/Soft-Tissue Infection

Both of the following:

1. Any mention of the following with recent onset (≤14 days) ***[any of the following]***
   a. Skin erythema
   b. Surgical site infection
   c. Superficial central line infection
   d. Ostomy site infection
   e. Skin infection with associated lymphangitis
2. Antibiotic treatment initiated for suspected infection

IV. Endocarditis

Any one of the following [1-3]:

1. Major Criteria [both of the following]:
    a. Suggestive microbiology [at least one of the following]:
        i. Positive blood culture of an ***endocarditis organism*** **[any of the following]:**
            1. *Streptococcus bovis*
            2. *Viridians streptococci*
            3. *Staphylococcus aureus*
            4. *Enterococcus spp.*
            5. HACEK organisms
            6. Coagulase negative staphylococci
    b. Evidence of endocardial involvement [at least one of the following]:
        i. New regurgiant murmur (a change in a preexisting murmur does not get scored)
        ii. Echocardiogram suspicious for any of the following:
            1. Intracardiac mass with no alternative explanation
            2. Endocardial abscess
            3. New partial prosthesis dehiscence
            4. Vegetation on valve
2. Minor Criteria [at least 4 of the following]:
    a. Predisposing valvular disease or IV drug use
    b. Temperature $\geq$ 100.4°F or 38°C
    c. Vascular phenomena
        i. Janeway lesions, conjunctival hemorrhages, arterial emboli, septic pulmonary infarcts, mycotic aneurysm, intracranial bleed
    d. Immunologic phenomena
        i. Osler nodes, Roth Spots, elevated Rheumatoid factor, hematuria in non-catheter urine, or other evidence of glomerulonephritis
    e. Positive blood cultures
        i. Excluding a single positive culture for coagulase negative staphylococci or a single positive culture for an organism that does not fall into the "reasonable endocarditis organism" (i.e. coagulase-positive and coagulase-negative *S. aureus*, Enterococcus, viridians group Streptococci, *S. bovis*, HACEK organisms)
    f. Positive serology for Brucella, Bartonella, Legionella, or Chlamydia
    g. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
3. At least one Major Criteria **AND** 3 minor criteria.

V. Meningitis/Encephalitis

Any one of the following [1 or 2]:

1. Both of the following [a-b]
    a. Laboratory Findings [any one of the following (i-ix)]
        i. CSF demonstrates any bacterium
            1. Excluding Diptheroids, Propionibacteria, Bacillus, Coagulase Negative *Staphylococcus*
        ii. CSF demonstrates Diptheroids, Propionibacteria, Bacillus, Coagulase Negative *Staphylococcus* in the setting of past neurosurgical intervention **AND** physicians elected to treat with antibacterials
        iii. Blood cultures positive for any of the following:
            1. *S. pneumoniae*
            2. *H. influenza*
            3. *Neisseria meningitidis*
            4. Group B Streptococcus
        iv. Stool cultures positive for enterovirus
        v. Throat or sputum cultures positive for *Neisseria meningitidis* in the setting of a rapid onset, overwhelming infection syndrome, including petechiae
        vi. Serology positive for *Mycoplasma*, *Leptospira*, measles, mumps, lymphocytic choriomeningitis virus, arboviruses (e.g. St. Louis encephalitis virus), or HIV (if historically consistent with acute seroconversion).
        vii. Brain biopsy demonstrates encephalitis
        viii. Positive CSF culture or PCR detection for any of the following
        ix. Acute or convalescent serology demonstrates positive antibody pattern for any of the following:
            1. Encephalitis arbovirus (La Crosse, St. louis, Eastern Equine, Western Equine, Powassan, Japanese, West Nile)
    b. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected meningitis/encephalitis

2. At least two of the following [a-d]
    a. Clinical meningitis/encephalitis [at least two of the following]:
        i. Petechial rash
        ii. Nuchal rigidity (by history or exam)
        iii. Altered sensorium
        iv. Fever
        v. Altered level of consciousness, including “agitation” or “lethargy”
        vi. Behavioral change
        vii. Diminished level of consciousness (not easily roused)
        viii. History of any of the following: headaches, altered mental status, or recent exposure to patient with known bacterial meningitis
        ix. Reduction in fever within 72 hours of starting anti-bacterial
    b. Inflammatory CSF [at least one of the following i-ii]
        i. Pleocytosis: $\geq$ 15 WBC/mm^3 (after subtracting one WBC for every 1,000 RBC)
        ii. Elevated protein (based on local lab-determined upper limits)
    c. Suggestive Findings [at least one of the following (i-iv)

i. Septic syndrome
ii. Focal neurological deficits documented during examination (such as flaccid paralysis or speech alterations for West Nile Virus)
iii. Abnormal imaging
    1. Computed tomography or magnetic resonance imaging (MRI) demonstrating focal edema or inflammation or hemorrhage
    2. Indicated as “meningitis/encephalitis” or “compatible with meningitis/encephalitis” or “cannot rule out meningitis/encephalitis”
iv. Findings indicating an abnormal electroencephalography (such as focal periodic discharges)

d. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for presumed meningitis/encephalitis

VI. Pyelonephritis

At least two of the following [1-4]:

1. Suggestion of infection [at least one of the following]:
    a. Temperature ≥ 100.4°F (38°C)
    b. Peripheral blood WBC ≥ 10,000/mm^3
    c. Positive blood culture for any of the following:
        i. Gram Negative Rods
        ii. *Enterococcus spp.*
        iii. *Staphylococcus saprophyticus*
    d. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
2. Strong renal localization [at least one of the following]:
    a. CT, MRI, or Ultrasound Suggestive of Renal Inflammation
3. Minor Criteria [at least two of the following]:
    a. Flank pain
    b. Costovertebral angle tenderness
    c. Complaints of dysuria, frequency, or suprapubic pain
    d. Any pyuria
    e. Urine culture positive for a single organism
4. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected pyelonephritis

VII. Septic Arthritis/Osteomyelitis

Any one of the following (1-5):

1. Synovial fluid gram stain or tissue gram stain or special stain demonstrating any organism
2. Joint culture/PCR/serology positive for any organism
3. At least two of the following (a-d):
    a. Positive blood culture/PCR/serology
    b. Joint with acute (≤ 7 days) worsening of inflammatory features (**at least two of the following):**
        i. Pain on history
        ii. ROM
        iii. Warmth
        iv. Effusion
        v. Swelling
        vi. Limited range of motion
    c. Antibiotic/antiviral/antifungal/antifungal treatment initiated/recommended for suspected infection
    d. Any one of the following (i-iv)
        i. Synovial fluid WBC ≥ 30,000/mm^3
        ii. Synovial fluid WBC ≥ 60,000/mm^3 with > 75% PMNs
        iii. Skin lesions, tenosynovitis, or urethral/cervical/rectal Gram stain or culture suggestive of *Neisseria gonorrhoeae*
        iv. Any indication of the following in the synovial fluid: needle-like crystals, CPPD crystals, uric acid.
4. Positive bone biopsy [at least one of the following (a-c)]:
    a. Positive culture for any organism
    b. Positive gram stain
5. Imaging and indirect features **[at least two of the following (a-c)]:**
    a. Consistent imaging [at least one of the following (i-iv)]:
        i. Plain X-ray read by a radiologist as suggestive of osteomyelitis
        ii. CT Scan read by a radiologist as suggestive of osteomyelitis
        iii. MRI read by a radiologist as suggestive of osteomyelitis
        iv. Bone scan or WBC scan read as suggestive of osteomyelitis
    b. Suggestive indirect features[at least one of the following (i-viii)]:
        i. Temperature > 100.4°F (38°C)
        ii. Bony pain or tenderness or erythema over bone suspected to be infected
        iii. Draining soft tissue sinus over bone suspected to be infected
        iv. Positive "probe to bone" (or visible bone in deep ulcer at suspected site)
        v. Blood culture positive for *S. aureus*
        vi. ESR ≥ 75 mm/hour
        vii. Intravenous drug use or indwelling catheter
        viii. Inflammation on imaging associated with an orthopedic prosthesis
    c. Positive culture for any organism form wound sample over the bone suspected of infection
    d. Antibiotic/antiviral/antifungal treatment for suspected infection

VIII. References

1. Grijalva CG, Chung CP, Stein CM, et al. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. Pharmacoepidemiology and drug safety 2008; 17(9): 890-5.
2. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. Journal of clinical epidemiology 2007; 60(4): 397-409.
3. Patkar NM, Curtis JR, Teng GG, et al. Administrative codes combined with medical records based criteria accurately identified bacterial infections among rheumatoid arthritis patients. *Journal of clinical epidemiology* 2009; **62**(3): 321-7, 7.e1-7.

Section & Topic	No	Item	Reported on page #
Title or Abstract			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	4
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4,5
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	5
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4,5
	9	Whether participants formed a consecutive, random or convenience series	4-6
Test methods	**10a**	Index test, in sufficient detail to allow replication	5, Table 1, Supplementary Appendix
	10b	Reference standard, in sufficient detail to allow replication	6, Supplementary Appendix
	11	Rationale for choosing the reference standard (if alternatives exist)	6, Supplementary Appendix
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	6, Supplementary Appendix
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	6, Supplementary Appendix
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	6, Supplementary Appendix
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	7,8
	15	How indeterminate index test or reference standard results were handled	7,8
	16	How missing data on the index test and reference standard were handled	7,8
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	7,8
	18	Intended sample size and how it was determined	7,8
RESULTS			
Participants	**19**	Flow of participants, using a diagram	8
	20	Baseline demographic and clinical characteristics of participants	8
	21a	Distribution of severity of disease in those with the target condition	n/a
	21b	Distribution of alternative diagnoses in those without the target condition	n/a
	22	Time interval and any clinical interventions between index test and reference standard	n/a
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	9,10, Table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	9,10 Table 2
	25	Any adverse events from performing the index test or the reference standard	n/a
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	11-13
	27	Implications for practice, including the intended use and clinical role of the index test	11-13

OTHER INFORMATION			
	28	Registration number and name of registry	n/a
	29	Where the full study protocol can be accessed	14
	30	Sources of funding and other support; role of funders	13