



U.S. Department of Veterans Affairs

Public Access Author manuscript

Med Care. Author manuscript; available in PMC 2019 July 01.

Published in final edited form as:

Med Care. 2018 July ; 56(7): 626–633. doi:10.1097/MLR.0000000000000916.

In Data We Trust? Comparison of Electronic versus Manual Abstraction of Antimicrobial Prescribing Quality Metrics for Hospitalized Veterans with Pneumonia

Barbara E Jones, MD, MS.,

Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS 2.0) Center of Innovation, VA SLC Health System; and Division of Pulmonary & Critical Care Medicine, University of Utah

Candace Haroldsen, MS.,

Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS 2.0) Center of Innovation, VA SLC Health System. Division of Epidemiology, University of Utah

Karl Madaras-Kelly, PharmD, MPH.,

Boise VA Medical Center and Idaho State University College of Pharmacy

Matthew Bidwell Goetz, MD.,

VA Greater Los Angeles Healthcare System and David Geffen School of Medicine at UCLA

Jian Ying, PhD.,

Division of Epidemiology, University of Utah

Brian Sauer, PhD.,

Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS 2.0) Center of Innovation, VA SLC Health System and Division of Epidemiology, University of Utah

Makoto M Jones, MD, MS.,

Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS 2.0) Center of Innovation, VA SLC Health System and Division of Epidemiology, University of Utah

Molly Leecaster, PhD.,

Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS 2.0) Center of Innovation, VA SLC Health System and Division of Epidemiology, University of Utah

Tom Greene, PhD.,

Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS 2.0) Center of Innovation, VA SLC Health System and Division of Epidemiology, University of Utah

Scott K Fridkin, MD.,

Emory University School of Medicine, Department of Medicine, Atlanta, GA

Melinda M. Neuhauser, PharmD, MPH, and

VA Center for Medication Safety (MedSAFE), Pharmacy Benefits Management Services, Department of Veterans Affairs, Hines, IL

Address: SLC VA Medical Center 500 Foothill Blvd, Building 2, Salt Lake City, UT 84148, Fax 801-584-5640, Telephone 406-581-1930, Barbara.jones@hsc.utah.edu.

All authors declare that they have no conflicts of interest for the past 3 years.

Matthew Samore, MD

Informatics, Decision-Enhancement, and Analytic Sciences (IDEAS 2.0) Center of Innovation, VA SLC Health System

Abstract

BACKGROUND—Electronic health records (EHR) provide the opportunity to assess system-wide quality measures. Veterans Affairs (VA) Pharmacy Benefits Management Center for Medication Safety employs Medication Use Evaluation (MUE) through manual review of the EHR.

OBJECTIVE—To compare an electronic MUE approach versus human/manual review for extraction of antibiotic use (choice and duration) and severity metrics.

RESEARCH DESIGN—Retrospective.

SUBJECTS—Hospitalizations for uncomplicated pneumonia occurring during 2013 at 30 VA facilities.

MEASURES—We compared summary statistics, individual hospitalization-level agreement, facility-level consistency, and patterns of variation between electronic and manual MUE for initial severity, antibiotic choice, daily clinical stability, and antibiotic duration.

RESULTS—Among 2,004 hospitalizations, electronic and manual abstraction methods demonstrated high individual hospitalization-level agreement for initial severity measures (agreement=86–98%, kappa=0.5–0.82), antibiotic choice (agreement=89–100%, kappa=0.70–0.94), and facility-level consistency for empiric antibiotic choice (anti-MRSA $r=0.97$, $p<0.001$; anti-pseudomonal $r=0.95$, $p<0.001$) and therapy duration ($r=0.77$, $p<0.001$) but lower facility-level consistency for days to clinical stability ($r=0.52$, $p=0.006$) or excessive duration of therapy ($r=0.55$, $p=0.005$). Both methods identified widespread facility-level variation in antibiotic choice, but we found additional variation in manual estimation of excessive antibiotic duration and initial illness severity.

CONCLUSIONS—Electronic and manual MUE agreed well for illness severity, antibiotic choice, and duration of therapy in pneumonia at both the individual and facility levels. Manual MUE demonstrated additional reviewer-level variation in estimation of initial illness severity and excessive antibiotic use. Electronic MUE allows for reliable, scalable tracking of national patterns of antimicrobial use, enabling the examination of system-wide interventions to improve quality.

Keywords

Pneumonia; Quality Measurement; Electronic Health Records; Antibiotic Stewardship; Antimicrobial Stewardship Taskforce

INTRODUCTION

Pneumonia is the leading infectious cause of hospitalization and death in the United States. (1) Overtreatment with excessive spectrum and duration of antibiotics are common. Thus, pneumonia is a chief target of both quality improvement efforts and antibiotic stewardship programs.

A core element of antibiotic stewardship is the tracking and reporting of antibiotic use, often with manual case review.(2) In 2011 the Veterans Health Administration (VHA) created the Antimicrobial Stewardship Taskforce (ASTF) to optimize care by developing, deploying, and monitoring a national-level strategic plan to improve antibiotic use. In January 2014, the VA issued a directive requiring all facilities to create local Antimicrobial Stewardship Programs.(3) In collaboration with the Pharmacy Benefits Management Services Center for Medication Safety, the ASTF has conducted several multicenter Medication Use Evaluations (MUE) to evaluate antimicrobial prescribing for common infectious syndromes, including pneumonia,(4) urinary tract infections, asymptomatic bacteruria,(5) and acute respiratory infections. Clinicians (primarily pharmacists) manually reviewed electronic health records (EHR) using a standardized chart abstraction system to examine antibiotic choice and duration. MUEs provide information regarding patterns of antibiotic use that is crucial to quality improvement. However, manual reviews are labor-intensive and thus neither scalable nor reproducible on a regular basis for most healthcare systems. The mMUE of pneumonia conducted by the VA required over 1,000 hours of manual chart review and 111 clinicians from 30 facilities.

Recent advances in the EHR provide opportunity for automated monitoring of antimicrobial prescribing at the national level,(3) offering potential savings in human time and rapid feedback to antibiotic stewards and policy-makers. The quality of system-wide electronic data is crucial to reliable reporting. We developed and evaluated an electronic data extraction tool, the “eMUE”, that extracts the same information found in the mMUE and can be applied to the national VA population for ongoing monitoring of pneumonia severity and antibiotic use metrics. Our aim was to compare electronic (eMUE) to manual (mMUE) abstraction for consistency in extraction of antibiotic use and severity metrics at the aggregate, facility, and individual hospitalization levels.

METHODS

Setting & Subjects

The study was performed with approval from the University of Utah Institutional Review Board (IRB#00068717) and Salt Lake City VA Human Research Protection Program. We included data from all patients in an existing dataset of the multicenter mMUE mentioned above, which has been previously described in the literature.(4) In brief, the mMUE included records from patients at 30 VA Medical Centers (VAMCs) who were hospitalized to medical wards or intensive care units during 2013 with a primary *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)* principal diagnosis for pneumonia, or a secondary diagnosis of pneumonia with a principal diagnosis of sepsis. One hundred-eleven clinicians reviewed random samples of 200 hospitalizations at each of the 30 facilities. Hospitalizations were included if patients had a length of stay 48 hours, >24 hours of inpatient antibiotic therapy, survived to hospital discharge, and reviewers confirmed a diagnosis of pneumonia in the discharge summary. Hospitalizations were excluded if manual review identified direct transfer from another facility, a previous hospitalization within the past 28 days, diagnosis of complicated pneumonia (e.g. lung abscess, necrotizing pneumonia, effusion requiring thoracentesis or thoracostomy),

immunosuppression, extra-pulmonary infection (e.g. endocarditis or meningitis), gram-negative enteric bacteremia, hospitalization greater than 14 days, or evidence that a 5-day course of antibiotic therapy was not given. The prior mMUE study also excluded cases of patients who were not found to achieve clinical stability during hospitalization. However, as these patients had complete data, they were included in the present study for comparative purposes.

To test the feasibility of a national eMUE, we also applied the eMUE to 128,757 hospitalizations occurring at 120 VA medical centers during the years 2008–2013 with diagnosis codes for pneumonia listed above. As our purpose of the national eMUE was to include all pneumonia cases and not only uncomplicated pneumonia cases, we did not apply any other inclusion criteria.

Measurements

Manual reviewers extracted information using the VA Computerized Patient Record System (CPRS) interface according to a standardized manual chart abstraction procedure previously described.(4) The electronic Medication Use Evaluation (eMUE) targeted the same elements as the manual extraction, with identical time windows and definitions as the mMUE, using data extracted from VA's Clinical Data Warehouse, accessed through the Veterans Informatics and Computing Infrastructure (VINCI). (6) To characterize patient outcomes, we extracted 30-day mortality and readmission within 28 days using electronic abstraction.

Both manual and electronic approaches extracted patient factors comprising healthcare-associated pneumonia criteria according to IDSA/ATS guidelines (7, 8) and the following physiologic variables contributing to the pneumonia severity index(9): minimum systolic blood pressure, maximum heart rate, minimum and maximum temperature, maximum respiratory rate, minimum pH, maximum blood urea nitrogen, minimum sodium, maximum glucose, and minimum hematocrit between 12 hours prior and 6 hours after hospital admission. Subsequent vital signs contributing to clinical stability criteria defined by IDSA/ATS guidelines(7) were extracted for each calendar day the patient was hospitalized. The severity information extracted by the eMUE differed from the mMUE in three important ways. First, because mental status is not routinely captured in structured data, the eMUE lacked assessments of mental status. Second, use of supplemental oxygen was not included in the mMUE but was extracted by the eMUE. Third, mMUE classified whether the patient's values met criteria for instability but did not record numeric values; eMUE extracted numeric values, then applied rules according to the criteria.

Antibiotic use was extracted by both methods following identical guidelines. Manual reviewers identified administration of each antibiotic class for each calendar day of hospitalization, and the name and duration of all outpatient prescriptions of antibiotics prescribed at hospital discharge from medication records and discharge summaries. eMUE extracted antibiotics administered for each hospital day using bar code medication administration (BCMA), which records all medications administered to hospitalized patients.(10) Antibiotic prescribing after hospitalization was extracted from outpatient prescription data upon discharge. Similar to the mMUE, the eMUE determined total duration

of antibiotic therapy by adding the number of inpatient treatment days plus days of therapy prescribed at discharge.

Manual reviewers identified antibiotic names; they did not classify antibiotic coverage, guideline concordance, or appropriateness of antibiotic selection or duration. Rules to define antibiotic class, coverage, and excessive duration were applied after extraction. We defined initial antibiotic choice as the systemic administration of that antibiotic during the first two calendar days of each hospitalization. In order to report antibiotics of interest to stewards, we further classified antibiotic coverage against two pathogen types: methicillin-resistant *Staphylococcus aureus* (MRSA: vancomycin or linezolid), and single coverage for *Pseudomonas aeruginosa* (PAER: piperacillin-tazobactam, ceftazidime, cefepime, meropenem, doripenem, imipenem, aztreonam). We defined excessive duration of antibiotic therapy as greater than the guideline-recommended duration of 5 days, plus 3 days after the first *single* day that the patient met all criteria for clinical stability(7): maximum temperature 100F, maximum respiratory rate 24 breaths per minute, minimum systolic blood pressure 90 mmHg, minimum pulse oximetry 90%, and baseline mental status. For example, if a patient met all stability criteria by days 1–5, duration of therapy >8 days was classified as excessive; however, if a patient became stable on day 7, duration >10 days was deemed excessive. Missing stability data were treated as normal/stable for both abstraction methods.

Analysis

Overall and Individual-level comparisons between eMUE and mMUE—We compared the proportions of hospitalizations identified by eMUE versus mMUE with unstable vital signs or laboratory results, initial administration of broad-spectrum antibiotic therapy, or excessive antibiotic duration, to compare how closely each of the methods might report these occurrences across the entire VA system. We generated 2-by-2 contingency tables and reported the percent agreement and kappa statistics(11) to characterize individual-level consistency between eMUE and mMUE for these binary variables. We reported the mean, standard deviation, median and interquartile range of the duration of antibiotic therapy for eMUE and mMUE to summarize overall levels and variation of this metric for eMUE and mMUE, and used paired t-tests for a formal comparison of the duration of antibiotic therapy.(12)

Facility-level comparisons between eMUE and mMUE—We assessed the implications of applying eMUE in place of mMUE for assessing facility-level quality metrics by computing, for each hospital, the median antibiotic duration and the percent of hospitalizations with a) initial anti-MRSA therapy, b) initial anti-PAER therapy, and c) excessive antibiotic duration using both abstraction methods. We graphically displayed regressions relating the resulting facility-level summary measures between eMUE and mMUE and reported the associated Pearson correlations.

Impact of reviewer on differences in variability between eMUE and mMUE—To assess the degree to which variation may arise in the manual MUE due to differences in abstraction methods among reviewers, we fit mixed effect logistic models with facility and reviewer as random effects to initial clinical stability (defined by the PSI physiologic

characteristics), excessive antibiotic duration, anti-MRSA coverage, and anti-PAER coverage data for both abstracted datasets. eMUE algorithms were applied systematically, so no reviewer was involved in the eMUE. However, we included an artificially-defined random effect that corresponded to the mMUE reviewers when modeling eMUE to account for any “true” variation among the hospitalizations assigned to different reviewers. The objective of this strategy was to ascertain whether non-random assignment of hospitalizations to reviewers may have inflated the variability in measures between reviewers beyond that attributable to the review process itself. We estimated the variances of the random effects corresponding to facility and to the individual reviewers using eMUE and mMUE. When the observed variation was less than the expected variation due to chance alone, the estimated variance was reported as 0. Because our logistic model provided random effect variance estimates on the logit scale, which is difficult to interpret clinically, we visually displayed the distributions in probabilities of initial clinical instability and excessive antibiotic use based on the sum of the facility and reviewer random effects variances. The variation in these probabilities reflects the total variation in these metrics between facilities using either eMUE or mMUE. All statistical analyses were performed using STATA 14 MP (StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP) or R (R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Of 3,558 initial hospitalizations identified for mMUE, 2,004 underwent complete manual review after exclusions. Median age was 71 years; 97% of the patients were male, and 31% had at least 1 HCAP criterion identified by manual review.(8) Among those 72% patients with HCAP received an anti-MRSA antibiotic, and 69% received an anti-pseudomonal, compared to 18% and 19% of CAP patients, respectively. Median length of stay was 6 days (IQR 5–9), and 3.1% died within 30 days (Table 1).

For the 2004 MUE hospitalizations, both extraction methods identified similar but statistically significantly different proportions of patients with initial clinical instability and initial antibiotic selection for each of the measures extracted (Table 2). eMUE identified initial vital signs in 99.5% of all hospitalizations. The most common unstable initial vital sign was pulse oximetry <90% (19% identified by mMUE, 21% by eMUE); the most common unstable laboratory value was BUN > 29 (25% for both methods). Individual-level agreement ranged from 86% to 98.5%; kappas ranged from 0.50 to 0.86. Individual-level agreement for initial antibiotic selection ranged from 89% to 99.7%, with kappas ranging between 0.57 and 0.93 (Appendix A).

Clinical stability on subsequent hospital days demonstrated high overall consistency, but lower agreement between abstraction approaches at the individual hospitalization level (Table 2). eMUE identified values in 94% of all hospitalizations for pulse oximetry data, and in over 96% for all other vital signs. The mMUE determined 65% of all hospitalizations to be clinically stable by day 3; 72% by day 5, and 72% by day 7. eMUE proportions were but slightly higher (68%, 73%, 75%). When examining individual contributors to daily clinical stability, eMUE identified more patients with unstable pulse oximetry, heart rate, and

temperature on hospital days 1 and 2 but fewer in subsequent hospital days; overall, both abstraction methods were tightly aligned at the aggregate level (Appendix B). Time to clinical stability was slightly longer for mMUE than eMUE (2.6 versus 1.8 days), as was total duration of antibiotic therapy (10.5 versus 10.1 days, Table 2). The manual MUE classified 82% of hospitalizations as receiving excessive antibiotic duration, while eMUE identified 84% (agreement 83%, kappa 0.35).

eMUE's lack of mental status extraction contributed to some but not all of the discordance with the mMUE observed: removal of mental status classification increased agreement between eMUE and mMUE for initial clinical stability (agreement 86.7%, kappa 0.72), daily clinical stability (2.4 versus 1.8 days), and excessive antibiotic use (agreement 85%, kappa 0.38).

Facility-level estimations by both methods demonstrated widespread facility-level variation in antibiotic selection (Figure 1a,1b) and treatment duration (Figure 1c). Metrics with the greatest facility-level consistency were the percentage of patients receiving initial anti-MRSA coverage ($r=0.97$), percent receiving initial anti-PAER coverage ($r=0.95$), and total treatment duration (median days- $r=0.77$), while correlation was poorer for days to clinical stability ($r=0.52$) and percent with excessive duration of therapy ($r=0.55$). All correlation coefficients were statistically significant.

Both mMUE and eMUE identified significant variation in broad-spectrum antibiotic choice at the facility level (Table 3, rows 1 & 2). The eMUE found no significant variation in initial clinical stability and excessive antibiotic duration, while manual MUE demonstrated greater variation in assessment of initial clinical stability and excessive antibiotic duration at the reviewer level (Table 3, rows 3 & 4). This reviewer-level variation contributed to a greater variation in clinical stability and excessive antibiotic duration estimated by the manual MUE (Figure 2).

Application of the eMUE to the national population (Appendix C) demonstrated greater severity of illness, comorbidities, and outcomes than the comparison cohort but similar overall rates of anti-MRSA therapy (34%), anti-pseudomonal therapy (39%), and median duration of antibiotic therapy (10 days).

CONCLUSIONS

Our study demonstrated consistency between electronic extraction of data and manual chart abstraction for evaluating initial illness severity, antibiotic choice, and total duration of antibiotic therapy in patients hospitalized with pneumonia. We found high agreement at the aggregate, facility, and individual hospitalization levels for these metrics, but lower individual-level agreement for excessive antibiotic duration based upon daily clinical stability criteria. Our application of the electronic abstraction to national VA data suggests that standardized electronic abstraction can be applied across a large system for accurate, replicable, clinically relevant tracking and reporting.

Development of trustworthy metrics is a challenging but crucial component to quality improvement.(13) Reliable antibiotic use metrics allow us to study and report patterns to

stewards and clinicians, target interventions, and examine changes after their implementation, thus promoting a learning health system by leveraging existing clinical data for continuous improvement.(14) (15) Advancement of the EHR provides the opportunity to develop more flexible, nimble approaches to quality measurement.(16) However, unreliable metrics can result in a misunderstanding of clinical reality, misdirected implementation efforts, or erosion of trust between stewards, clinicians, and patients, especially if combined with punitive or incentivizing approaches to quality improvement.(17) (18)

Using the EHR to examine medical care presents an important epistemological challenge: while the data left behind from an episode of care can provide an important glimpse into the nature of that care, how can we know whether the data reflect what actually occurred? Quality metrics using the EHR have shown promise for both inpatient(19) and outpatient (20) care, but limitations to both electronic and manual approaches have been identified, specifically for pneumonia.(21) Either approach can be subject to errors in collection or interpretation, and it is important to understand these limitations. We adopted the convergent validity approach from the social sciences, seeking those metrics with high consistency across both measurements as the most accurate.(22)

While manual case reviews are time-consuming, they are generally accepted as a gold standard due to humans' ability to incorporate multiple types of information to generate a more comprehensive view of clinical reality. (23) We found a large degree of reviewer-level variation in metrics that required estimations of clinical stability. Qualitative chart reviews of discordant cases revealed that misclassification (i.e., temperature > 100F rather than 103.8F) misidentification (i.e., urine pH rather than serum pH), and failure to capture all numeric values within time windows were common sources of human error. Classifying clinical stability required multiple integrative steps: each reviewer had to identify numeric data within a time window and determine whether that data met a certain criteria. Further complicating this task, *initial* clinical stability criteria in pneumonia are different from stability on subsequent hospital days for temperature, heart rate, and respiratory rate. Human errors can be expected to vary by reviewer, as we observed.

The problem of human variability in manual chart review is important to consider, especially when comparing facilities. Both MUE approaches found facility-level variation in antibiotic choice and duration, and understanding the forces driving this variation, such as differences in microbiology,(24) practice cultures, or stewardship efforts,(25, 26) is an important step to improvement. However, the mMUE was organized such that clinicians performed reviews in cases from their own facilities. When reviewers are nested within facilities, reviewer-level variation may be misinterpreted as facility differences, hindering the ability to compare facilities. If the aim is to examine facilities, reliability may be more important than overall accuracy. In a comparison of definitions of catheter-related bloodstream infections (CRBSI), a simple definition based upon microbiology results was found to be less accurate than clinical definitions determined by clinician reviewers; however, the objective criteria were more reliable at identifying differences in CRBSI rates across institutions.(27) When the goal is to comparing facilities, consistent electronic abstraction of simple, objective metrics may be more trustworthy than metrics interpreted by humans.

In contrast to the mMUE, the greatest limitation to the eMUE was missing data, which we identified in three notable areas. First, because we used BCMA to identify antibiotic administration during hospitalization, the eMUE missed antibiotics administered in the emergency department, contributing to inaccuracy if the patient's antibiotics were changed upon hospital admission. We found this phenomenon to be rare; however, discordance between ED and hospital treatment is important information. Extraction of reliable antimicrobial data from the ED is the subject of future work. Similarly, the eMUE missed post-hospitalization antibiotics for patients discharged to nursing homes, lacking outpatient prescriptions. This may explain the lower duration of therapy identified by eMUE. Second, the current eMUE extraction missed vital sign data from some VA intensive care units, as charting software are used that do not deposit into the data warehouse. Missing daily vital signs in the intensive care units are a recognized limitation of VA data. Natural language processing extracting vital signs from text provides a promising approach to this problem. (28) Third, the eMUE had no extraction of mental status assessment. Extracting these data using natural language processing are also the subject of ongoing work.

Our study has some limitations. The study population was defined by the existing manual review, which excluded many patients with poorer clinical outcomes (such as inpatient deaths or empyema). Our eMUE does not identify these exclusion criteria, as we aimed to develop an approach that can be applied to a large population of patients hospitalized for pneumonia. Avoiding exclusion criteria based upon clinical outcomes reduces the problem of conditioning on future events, allowing the tracking of quality measures beyond antibiotic use, such as mortality or hospital readmission. However, omission of these exclusion criteria makes defining "appropriate" duration of therapy more difficult. At an individual level, defining appropriate antibiotic use remains a challenge. The complexity of individualized care likely exceeds the ability of any electronic abstraction method to be 100% accurate in identifying appropriate deviations from standard care. While the eMUE holds promise in more accurately and reliably identifying those deviations, we feel that human review plays an essential role in identifying the factors contributing to them. Further, periodic manual review can improve local credibility of eMUE findings and identify errors in data reporting or changes in data infrastructure. Neither the current mMUE nor the eMUE were designed to capture implicit measures of quality,(29) and future work is required to operationalize concepts of appropriateness or good clinical judgment. However, evaluating simple explicit metrics, especially at the facility level, appear to be a good place to start. For example, the median duration of antibiotic therapy was 10 days, far greater than the recommended duration for uncomplicated pneumonia, and high proportions of patients received broad-spectrum antibiotics despite a low prevalence of resistant organisms.(24) Thus, examining the simple metrics of duration and empiric use of broad-spectrum antibiotics may be a reliable approach to identifying opportunities for improvement.

Measuring practice is a key feature of a learning health system, and increasingly advanced computational capabilities provide the opportunity for replicable system-wide evaluations, (23) decision support for stewardship,(30) and quality improvement. However, clinical data must not be taken at face value. The eMUE described in this study can extract antibiotic choice, duration, and initial severity of illness data from the entire VA system in hours, allowing for frequent, timely feedback. Consistency, reliability, and speed are important

strengths of electronic tracking which, when accurate, can produce trustworthy data that we can use to improve.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Francesca Cunningham, PharmD, of the VA Pharmacy Benefits Management Program for collaboration and input on the final manuscript. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

Disclosure of Funding:

This work was supported by a grant from the Centers for Disease Control and Prevention (Grant # 14FED1408985). B. Jones is supported by a career development grant from the Veterans' Affairs Health Service Research & Development (150HX001240). M. Jones is supported by a career development grant from the Veterans' Affairs Health Service Research & Development (IK2HX001165).

References

1. QuickStats: Number of Deaths from 10 Leading Causes,* by Sex - National Vital Statistics System, United States, 2015. *MMWR Morb Mortal Wkly Rep.* 2017; 66(15):413.doi: 10.15585/mmwr.mm6615a8 [PubMed: 28426642]
2. Pollack LA, Srinivasan A. Core elements of hospital antibiotic stewardship programs from the Centers for Disease Control and Prevention. *Clin Infect Dis.* 2014; 59(Suppl 3):S97–100. DOI: 10.1093/cid/ciu542 [PubMed: 25261548]
3. Kelly AA, Jones MM, Echevarria KL, Kralovic SM, Samore MH, Goetz MB, et al. A Report of the Efforts of the Veterans Health Administration National Antimicrobial Stewardship Initiative. *Infect Control Hosp Epidemiol.* 2017; 38(5):513–20. DOI: 10.1017/ice.2016.328 [PubMed: 28118861]
4. Madaras-Kelly KJ, Burk M, Caplinger C, Bohan JG, Neuhauser MM, Goetz MB, et al. Total duration of antimicrobial therapy in veterans hospitalized with uncomplicated pneumonia: Results of a national medication utilization evaluation. *J Hosp Med.* 2016; 11(12):832–9. DOI: 10.1002/jhm.2648 [PubMed: 27527659]
5. Spivak ES, Burk M, Zhang R, Jones MM, Neuhauser MM, Goetz MB, et al. Management of Bacteriuria in Veterans Affairs Hospitals. *Clin Infect Dis.* 2017; doi: 10.1093/cid/cix474
6. [cited 2017 July 27]. Available from: https://www.hsrd.research.va.gov/for_researchers/vinci/.
7. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis.* 2007; 44(Suppl 2):S27–72. DOI: 10.1086/511159 [PubMed: 17278083]
8. American Thoracic S, Infectious Diseases Society of A. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. *Am J Respir Crit Care Med.* 2005; 171(4):388–416. DOI: 10.1164/rccm.200405-644ST [PubMed: 15699079]
9. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med.* 1997; 336(4):243–50. DOI: 10.1056/NEJM199701233360402 [PubMed: 8995086]
10. Schneider R, Bagby J, Carlson R. Bar-code medication administration: a systems perspective. *Am J Health Syst Pharm.* 2008; 65(23):2216, 8–9. DOI: 10.2146/ajhp080163 [PubMed: 19020188]
11. Berry CC. The kappa statistic. *JAMA.* 1992; 268(18):2513–4. [PubMed: 1404812]
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986; 1(8476):307–10. [PubMed: 2868172]

13. Roth CP, Lim YW, Pevnick JM, Asch SM, McGlynn EA. The challenge of measuring quality of care from the electronic health record. *Am J Med Qual.* 2009; 24(5):385–94. DOI: 10.1177/1062860609336627 [PubMed: 19482968]
14. Olsen, LA, Aisner, D., McGinnis, JM., editors. *The National Academies Collection: Reports funded by National Institutes of Health.* Washington (DC): 2007. *The Learning Healthcare System: Workshop Summary.*
15. Rubenstein LV, Mittman BS, Yano EM, Mulrow CD. From understanding health care provider behavior to improving health care: the QUERI framework for quality improvement. *Quality Enhancement Research Initiative. Med Care.* 2000; 38(6 Suppl 1):I129–41. [PubMed: 10843277]
16. McGlynn EA, Schneider EC, Kerr EA. Reimagining quality measurement. *N Engl J Med.* 2014; 371(23):2150–3. DOI: 10.1056/NEJMp1407883 [PubMed: 25470693]
17. Pronovost PJ, Miller M, Wachter RM. The GAAP in quality measurement and reporting. *JAMA.* 2007; 298(15):1800–2. DOI: 10.1001/jama.298.15.1800 [PubMed: 17940238]
18. Werner RM, Asch DA. The unintended consequences of publicly reporting quality information. *JAMA.* 2005; 293(10):1239–44. DOI: 10.1001/jama.293.10.1239 [PubMed: 15755946]
19. Davies, SM., Geppert, J., McClellan, M., McDonald, KM., Romano, PS., Shojania, KG. *AHRQ Technical Reviews.* Rockville (MD): 2001. *Refinement of the HCUP Quality Indicators.*
20. Kerr EA, Hofer TP, Hayward RA, Adams JL, Hogan MM, McGlynn EA, et al. Quality by any other name?: a comparison of three profiling systems for assessing health care quality. *Health Serv Res.* 2007; 42(5):2070–87. DOI: 10.1111/j.1475-6773.2007.00730.x [PubMed: 17850534]
21. Linder JA, Kaleba EO, Kmetik KS. Using electronic health records to measure physician performance for acute conditions in primary care: empirical evaluation of the community-acquired pneumonia clinical quality measure set. *Med Care.* 2009; 47(2):208–16. DOI: 10.1097/MLR.0b013e318189375f [PubMed: 19169122]
22. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull.* 1955; 52(4):281–302. [PubMed: 13245896]
23. Goulet JL, Erdos J, Kancir S, Levin FL, Wright SM, Daniels SM, et al. Measuring performance directly using the veterans health administration electronic medical record: a comparison with external peer review. *Med Care.* 2007; 45(1):73–9. DOI: 10.1097/01.mlr.0000244510.09001.e5 [PubMed: 17279023]
24. Jones BE, Brown KA, Jones MM, Huttner BD, Greene T, Sauer BC, et al. Variation in Empiric Coverage Versus Detection of Methicillin-Resistant *Staphylococcus aureus* and *Pseudomonas aeruginosa* in Hospitalizations for Community-Onset Pneumonia Across 128 US Veterans Affairs Medical Centers. *Infect Control Hosp Epidemiol.* 2017; 38(8):937–44. DOI: 10.1017/ice.2017.98 [PubMed: 28633678]
25. Chou AF, Graber CJ, Jones M, Zhang Y, Goetz MB, Madaras-Kelly K, et al. Characteristics of Antimicrobial Stewardship Programs at Veterans Affairs Hospitals: Results of a Nationwide Survey. *Infect Control Hosp Epidemiol.* 2016; 37(6):647–54. DOI: 10.1017/ice.2016.26 [PubMed: 26905338]
26. Graber CJ, Jones MM, Chou AF, Zhang Y, Goetz MB, Madaras-Kelly K, et al. Association of Inpatient Antimicrobial Utilization Measures with Antimicrobial Stewardship Activities and Facility Characteristics of Veterans Affairs Medical Centers. *J Hosp Med.* 2017; 12(5):301–9. DOI: 10.12788/jhm.2730 [PubMed: 28459897]
27. Rubin MA, Mayer J, Greene T, Sauer BC, Hota B, Trick W, et al. An agent-based model for evaluating surveillance methods for catheter-related bloodstream infection. *AMIA Annu Symp Proc.* 2008:631–5. [PubMed: 18999291]
28. Patterson OV, Jones M, Yao Y, Viernes B, Alba PR, Iwashyna TJ, et al. Extraction of Vital Signs from Clinical Notes. *Stud Health Technol Inform.* 2015; 216:1035. [PubMed: 26262334]
29. Berwick DM, Knapp MG. Theory and practice for measuring health care quality. *Health Care Financ Rev.* 1987; (Spec No):49–55. [PubMed: 10312319]
30. Evans RS, Olson JA, Stenehjem E, Buckel WR, Thorell EA, Howe S, et al. Use of computer decision support in an antimicrobial stewardship program (ASP). *Appl Clin Inform.* 2015; 6(1):120–35. DOI: 10.4338/ACI-2014-11-RA-0102 [PubMed: 25848418]

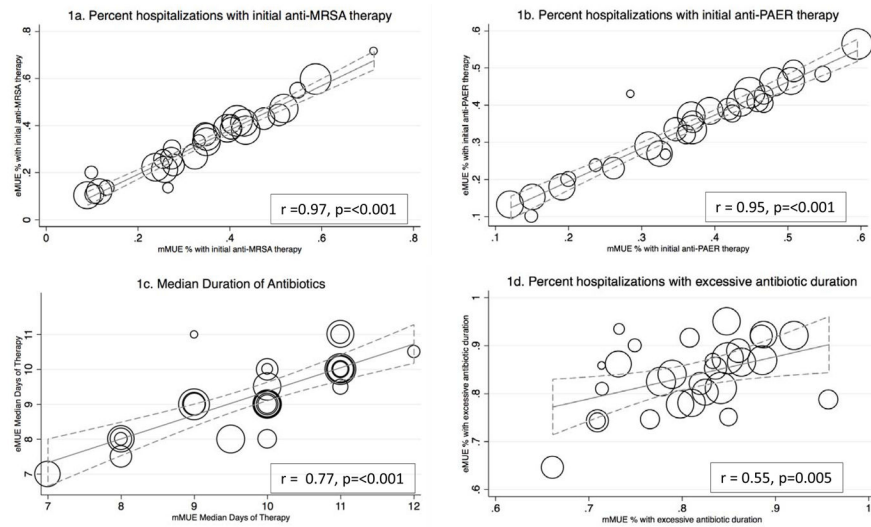


Figure 1. Facility-level consistency between eMUE and mMUE for antibiotic use metrics N=2004 patients at 30 facilities. Circles represent facilities; size of circle is in proportion to number of cases per facility. Metrics are plotted for each facility by eMUE abstraction (Y-axis) versus mMUE abstraction (x-axis). Solid lines represent best fit linear regression; dotted lines represent 95% confidence intervals.

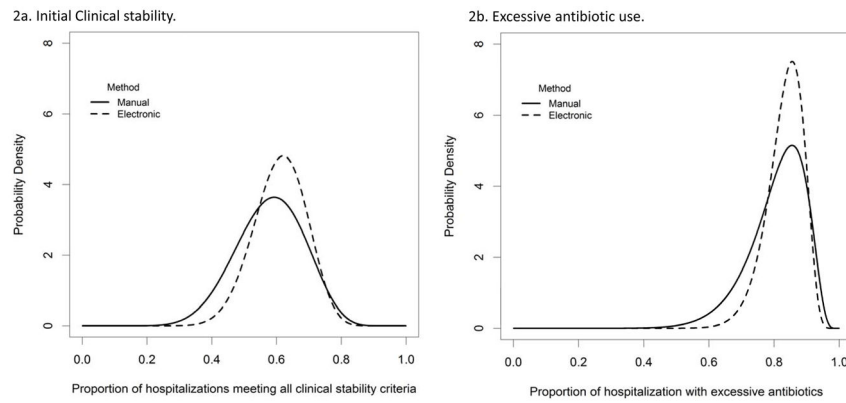


Figure 2. Variation in electronic versus manual extraction of initial clinical stability (2a) and excessive antibiotic duration (2b)

Model-based distributions across facilities were plotted for the proportions of hospitalizations meeting all initial clinical stability criteria (a) and with excessive antibiotics (b). Distributions obtained from eMUE (dashed lines) are more concentrated, implying less variation, than the corresponding distributions obtained from mMUE (solid lines). Additional facility-level variation in mMUE was attributable primarily to reviewer-level variation (Table 3).

Table 1

Patient characteristics and outcomes.

Number of cases	2,004	
Number of facilities	30	
	N	%
Male gender	1941	96.9%
Age (Median, StdDev)	71	64–82
HCAP risk factors	617	30.8%
Renal Disease	589	29.4%
Liver Disease	57	2.8%
Congestive Heart Failure	660	32.9%
Cerebrovascular Disease	397	19.8%
Neoplastic Disease	450	22.5%
ICU admission	224	11.2%
30-day mortality	63	3.1%
Length of Stay (Median + IQR Days)	6	5–9
28-day readmission rate	42	17.4%

Table 2

Comparison between eMUE and mMUE for severity, antibiotic choice, clinical stability and antibiotic duration metrics. N=2,004 hospitalizations across 30 facilities.

	mMUE %	eMUE %	p-value*	Agreement %**	Kappa**
Initial Severity of Illness					
Temperature >103.8 or <95F	1.9%	1.3%	0.05	98.5%	0.5
Pulse > 125 bpm	9.8%	10.9%	0.01	96.5%	0.81
Respiratory rate > 29	9.6%	7.8%	<0.001	96.0%	0.75
SBP < 90 mmHg	6.8%	8.8%	<0.001	86.0%	0.66
Pulse Oximetry <90%	18.8%	21.4%	0.001	91.0%	0.72
On supplemental O2 at time of lowest SpO2 value	NA	45.5%	NA	NA	NA
Altered mental status	12.4%	NA	NA	NA	NA
pH < 7.35	4.5%	3.8%	0.05	97.5%	0.69
BUN > 29 mg/dL	24.8%	25.1%	0.56	94.7%	0.86
Sodium < 130 mEq/L	5.7%	6.6%	0.03	97.0%	0.75
Glucose > 249 mg/dL	7.8%	10.2%	<0.001	94.0%	0.69
Hematocrit <30%	10.0%	11.3%	0.02	94.2%	0.70
Overall Clinical Stability (patient met no instability criteria)	58.4%	61.4%	0.003	80.7%	0.60
Initial Antibiotic Choice & Excessive Duration					
Antipseudomonal Coverage	37.2%	35.3%	<0.001	96.6%	0.93
Anti-MRSA Coverage	33.30%	35.10%	<0.001	95.20%	0.89
Excessive antibiotic duration based on clinical stability	82.3%	84.0%	<0.001	81.7%	0.35
Days to Stability and Duration					
	mMUE days	eMUE days		Mean ± S.D. of individual differences	
Days to clinical stability (the first single day the patient met all stability criteria) Mean, median (IQR)	2.6, 2 (1-3)	1.8, 1 (1-2)	0.01	-0.08 ± 0.119	
Number of days of antibiotic therapy	10.5 ± 4.5 (10, 8-12)	10.1±5.3 (9, 7-12)	<0.001	0.39 ± 4.1	

* p-values calculated using McNemar's for comparison of overall proportions.

** Agreement % and Kappa statistics were calculated using 2 by 2 contingency tables at the individual hospitalization level.

Differences were calculated at the individual hospitalization level ((mMUE - eMUE) / S.D. = standard deviation.

VA Author Manuscript

VA Author Manuscript

VA Author Manuscript

Table 3
Sources of variation in initial clinical stability and antibiotic use metrics by manual versus electronic extraction

Variance \pm standard error for each random effect (facility and reviewer) were calculated from mixed effect regression models and are shown on a logit scale. Estimated variances of zero reflect variables whose variation is less than that expected by chance.

Metric	Manual MUE		Electronic MUE	
	Facility	Reviewer	Facility	Reviewer
Initial Anti-MRSA coverage	0.388 \pm 0.138 p<0.001	0.016 \pm 0.049 p=0.36	0.335 \pm 0.127 p<0.001	0.044 \pm 0.060 p=0.20
Initial Anti-PAER coverage	0.258 \pm 0.09 p<0.001	0	0.222 \pm 0.081 p<0.001	0
Initial clinical stability	0.066 \pm 0.059 p=0.13	0.147 \pm 0.074 p=0.002	0.038 \pm 0.057 p=0.25	0.085 \pm 0.068 p=0.06
Excessive antibiotic duration (based upon clinical stability)	0	0.338 \pm 0.180 p=0.002	0.095 \pm 0.103 p=0.16	0.074 \pm 0.096 p=0.19