Can we use automated analyses of electronic health record data to improve surveillance for healthcare-associated infections?

Seminar on Application of Artificial Intelligence (AI) on Infectious Diseases and Infection Control
November 13, 2024

Michael Klompas MD, MPH, FIDSA, FSHEA
Professor, Harvard Medical School
Hospital Epidemiologist, Brigham and Women's Hospital

1

Disclosures

- **Grant funding**
  - Centers for Disease Control and Prevention
  - Massachusetts Department of Public Health
  - Agency for Healthcare Research and Quality
- **Royalties**
  - UpToDate

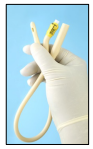
2

Current State

We focus the bulk of our effort on:

CMS CENTERS FOR MEDICARE & MEDICAID SERVICES

CAUTI

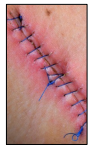


Catheter-associated urinary tract infection

CLABSI

Central-line associated bloodstream infection

SSI (HYST & COLO)

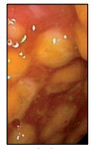


Surgical site infection (hysterectomy & colectomy)

HO-MRSA Bacteremia

Hospital-onset MRSA bacteremia

HO-CDI



Hospital-onset *C. difficile* infection

3

THE SENIC PROJECT

STUDY ON THE EFFICACY OF NOSOCOMIAL INFECTION CONTROL (SENIC PROJECT)

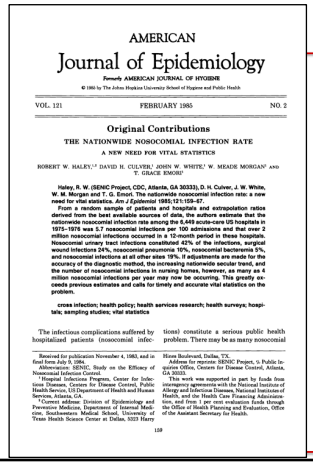
SUMMARY OF STUDY DESIGN

An Enduring Legacy of SENIC

- Landmark study designed by CDC in 1974 to describe US infection control programs, measure nosocomial infection rates, and determine if implementing infection surveillance and control programs lead to lower HAI rates
- Resounding success
- Set the framework for infection control for the next 50 years

Haley, *Am J Epi* 1980;111:472-485

4

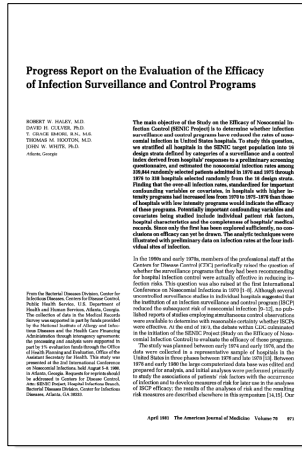

AMERICAN
Journal of Epidemiology
Formerly AMERICAN JOURNAL OF HYGIENE
© 1985 by The Johns Hopkins University School of Hygiene and Public Health
VOL. 121 FEBRUARY 1985 NO. 2
Original Contributions
THE NATIONWIDE NOSOCOMIAL INFECTION RATE
A NEW NEED FOR VITAL STATISTICS
ROBERT W. HALEY,[1] DAVID H. CULVER,[2] JOHN W. WHITE,[3] W. MEADE MORGAN[4] AND T. GRACE KNOBS[5]
Haley, R. W. (SENIC Project, CDC, Atlanta, GA 30333), D. H. Culver, J. W. White, W. M. Morgan and T. G. Emori. The nationwide nosocomial infection rate: a new need for vital statistics. *Am J Epidemiol* 1985;121:159–67.
From a random sample of patients and hospitals and extrapolation ratios derived from the best available sources of data, the authors estimate that the nationwide nosocomial infection rate among the 6,449 acute-care US hospitals in 1975–1976 was 5.7 nosocomial infections per 100 admissions and that over 2 million nosocomial infections occurred in a 12-month period in these hospitals. Nosocomial urinary tract infections constituted 42% of the infections, surgical wound infections 24%, nosocomial pneumonia 16%, nosocomial bacteremia 5%, and nosocomial infections at all other sites 19%. If adjustments are made for the accuracy of the diagnostic method, the increasing nationwide secular trend, and the number of nosocomial infections in nursing homes, however, as many as 4 million nosocomial infections per year may now be occurring. This greatly exceeds previous estimates and calls for timely and accurate vital statistics on the problem.
cross infection; health policy; health services research; health surveys; hospitals; sampling studies; vital statistics
The infectious complications suffered by hospitalized patients (nosocomial infections) constitute a serious public health problem. There may be as many nosocomial
Received for publication November 4, 1983, and in final form July 9, 1984.
[1] Study Officer, Study on the Efficacy of Nosocomial Infection Control.
[2] Hospital Infections Program, Center for Infectious Diseases, Centers for Disease Control, Public Health Service, US Department of Health and Human Services, Atlanta, GA.
[3] Department of Internal Medicine, Division of Epidemiology and Preventive Medicine, Department of Internal Medicine, Southwestern Medical School, University of Texas Health Science Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX.
[4] Address for reprints: SENIC Project, 5 Public Information Officer, Centers for Disease Control, Atlanta, GA 30333.
[5] This work was supported in part by funds from interagency agreements with the National Institute of Allergy and Infectious Diseases, National Institutes of Health, and the Health Care Financing Administration, and from a research evaluation funds through the Office of Health Planning and Evaluation, Office of the Assistant Secretary for Health.
159

An Enduring Legacy of SENIC

- Used detailed chart reviews to estimate an HAI rate of 5.7 events per 100 admissions
  - UTI 42%
  - SSI 24%
  - HAP 10%
  - Bacteremia 5%
  - Other sites 19%
- Documented that robust infection control programs were associated with substantial reductions in HAIs
- Parallel data demonstrated that devices, procedures, and intensive care accounted for the majority of nosocomial infections

Haley, *Am J Epi* 1985;121:159-167
Stamm, *Ann Intern Med* 1978;89:764-769

5


Progress Report on the Evaluation of the Efficacy of Infection Surveillance and Control Programs
ROBERT W. HALEY, M.D.
DAVID H. CULVER, Ph.D.
T. GRACE KNOBS, R.N., M.S.
THOMAS M. HOOTON, M.D.
JOHN W. WHITE, Ph.D.
Atlanta, Georgia
The main objective of the Study on the Efficacy of Nosocomial Infection Control (SENIC Project) is to determine whether infection control programs have contributed significantly to the decline in nosocomial infection in United States hospitals. To study this question, we stratified all hospitals in the SENIC target population into 16 design strata defined by categories of a surveillance and a control index derived from hospitals' responses to a preliminary screening questionnaire, and estimated the nosocomial infection rates among 338,044 randomly selected patients admitted in 1970 and 1975 through 1976 to 338 hospitals selected randomly from the 16 design strata. Finding that the over-all infection rates, standardized for important risk factors, had not declined from 1970 to 1975–1976 and that the majority programs had increased rate from 1970 to 1975–1976 than those of hospitals with low intensity programs would indicate the efficacy of these programs. Potentially important confounding variables and covariates being studied include individual patient risk factors, hospital characteristics and the completeness of hospitals' medical records. Since only the first has been explored sufficiently, no conclusions on efficacy can yet be drawn. The analytic techniques were illustrated with preliminary data on infection rates at the four individual sites of infection.
In the 1980s and early 1970s, members of the professional staff at the Centers for Disease Control (CDC) periodically visited the operation of whether the surveillance programs that they had been recommending for hospital infection control were actually effective in reducing infection risks. This question was also raised at the first International Conference on Nosocomial Infections in 1970 [1–3]. Although several retrospective surveillance studies in individual hospitals suggested that the institution of surveillance and infection control programs had reduced the subsequent risk of nosocomial infection [4–12], no published reports of studies employing simultaneous control observations were available to determine with reasonable certainty whether ISCPs were effective. At the end of 1973, the debate within CDC culminated in the initiation of the SENIC Project (Study on the Efficacy of Nosocomial Infection Control) to evaluate the efficacy of these programs.
The study was planned between early 1974 and early 1976, and the data were collected in a representative sample of hospitals in the United States between mid-1975 and mid-1976. The delay between 1970 and early 1980 the large computerized data base was edited and prepared for analysis, and initial analyses were performed primarily to study the associations of patients' risk factors with the occurrence of infection and to develop measures of risk for later use in the analyses of ISCP efficacy; the results of the analyses of risk and the resulting risk measures are described elsewhere in this symposium [14,15]. Our
From the Bacterial Diseases Division, Center for Infectious Diseases, Centers for Disease Control, Public Health Service, U.S. Department of Health and Human Services, Atlanta, Georgia.
April 1981 The American Journal of Medicine Volume 70 871

An Enduring Legacy of SENIC

- SENIC set multiple precedents:
  1. Infection control programs can be very impactful; a required program for hospitals
  2. We should focus our attention on a small set of high yield infections (SSI, CAUTI, CLABSI, ±HAP)
  3. Surveillance requires detailed chart reviews by trained experts applying clinically nuanced definitions

Haley, *Am J Med* 1981;70:971-978

6

Is this still the right approach for the current era?

- Are we capturing the right events (completeness, morbidity)?
- Are we conducting surveillance in an efficient and objective fashion?
- Is our approach commensurate with current data sources and documentation practices (i.e. EHRs)?

7

Difficulties with our current surveillance targets

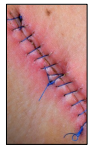
CAUTI

Catheter-associated urinary tract infection

CLABSI

Central-line associated bloodstream infection

SSI (HYST & COLO)



Surgical site infection (hysterectomy & colectomy)

HO-MRSA Bacteremia

Hospital-onset MRSA bacteremia

HO-CDI

Hospital-onset *C. difficile* infection

8

Complicated

Labor Intensive

Subjective

Non-Specific

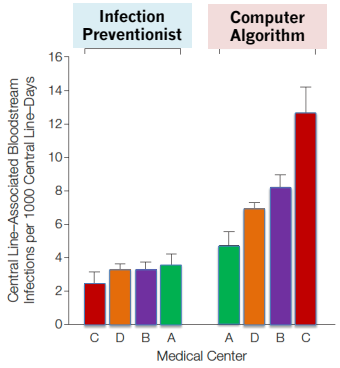
9

It's Hard!

- Is this patient's fever due to infection or some other cause?
- Is this bacteremia due to abdominal infection or line-infection?
- Is this coag-negative Staph a contaminant or a true pathogen?
- Is this positive *C.diff* test colonization or infection?
- Was this present on admission or was it hospital-acquired?
- Does the surgeon's prescription for antibiotics for "erythema" count as an attending physician's diagnosis of infection?
- Is cholecystitis after colectomy a surgical site infection?
- etc.

10

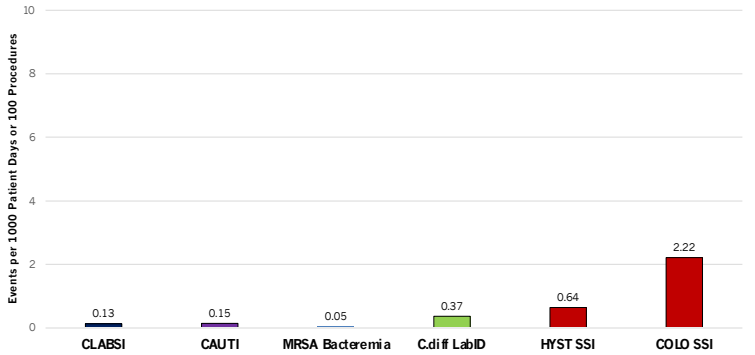
Manual versus Automated Surveillance on CLABSI Rank



Lin, *JAMA* 2010;304:2035-41

11

They Are Rare!

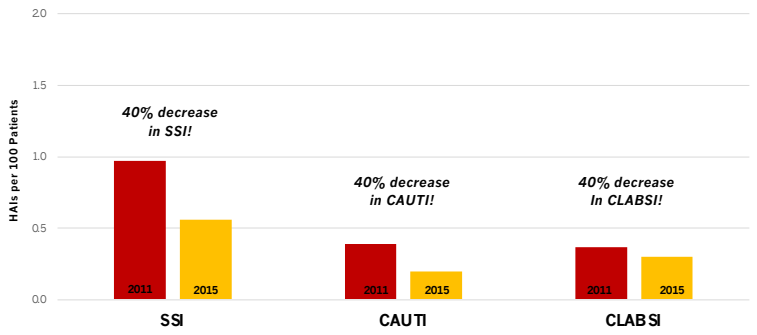


CDC HAI Progress Report for 2019

12

And Becoming Rarer

CDC point-prevalence surveys for HAIs in 2011 (183 hospitals) vs 2015 (199 hospitals)



Magill, *N Engl J Med* 2018;379:1732-1744

13

They Miss the Most Common HAI

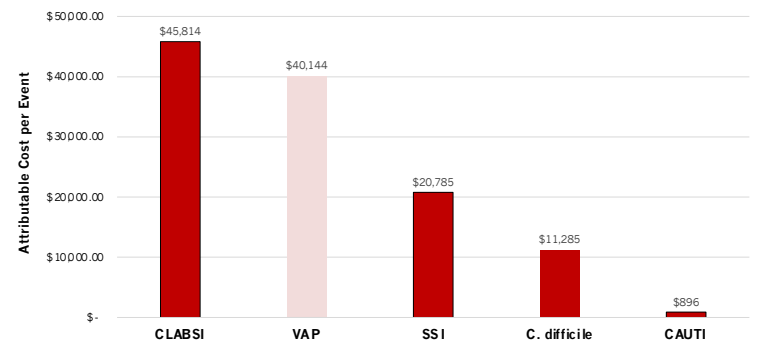
CDC one day point-prevalence survey of HAIs in 199 hospitals in 10 U.S. states

1. **Hospital-acquired pneumonia**
2. GI infections including *C. difficile*
3. Surgical site infection
4. Primary bloodstream infections
5. Urinary tract infections

Magill, *N Engl J Med* 2018;379:1732-1744

14

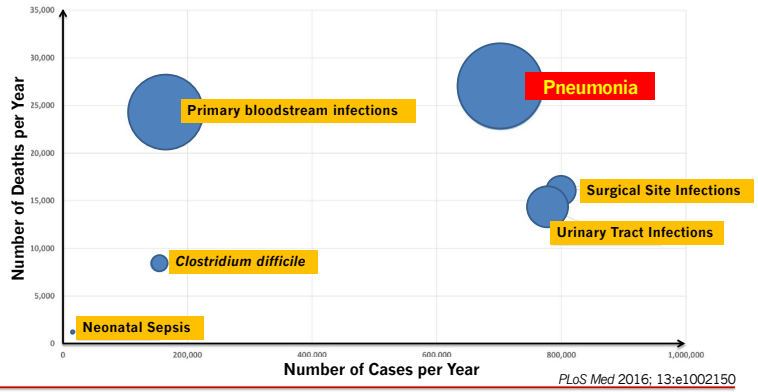
They vary widely in their impact



Zimlichman, JAMA Internal Med 2013;173:2039-46

15

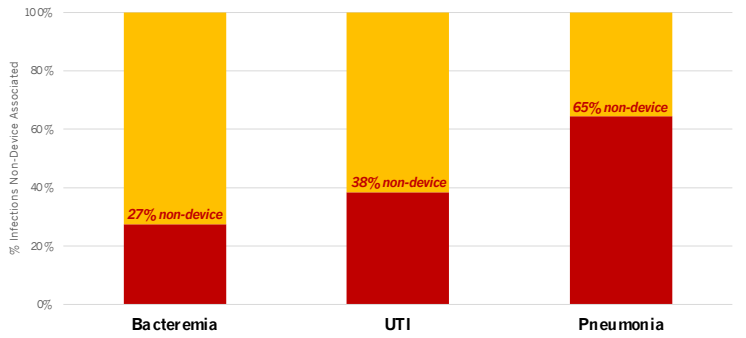
Burden & Mortality of HAIs in Europe



PLoS Med 2016; 13:e1002150

16

They miss non-device-associated infections



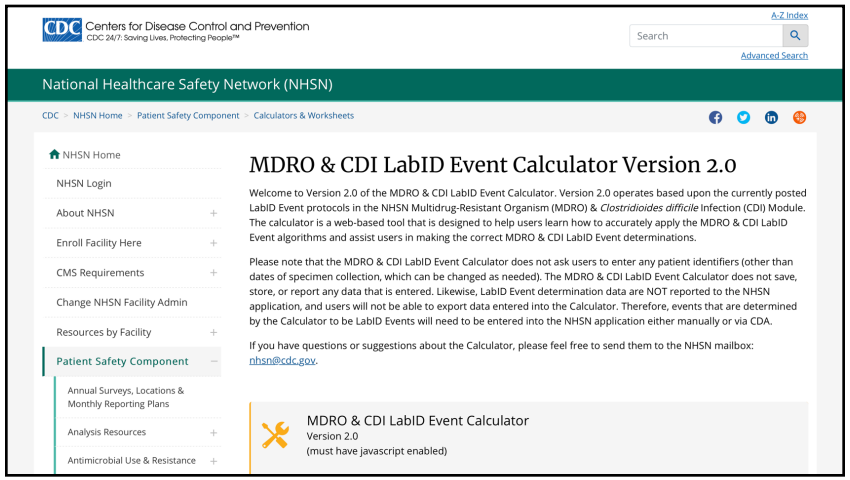
Magill, *N Engl J Med* 2018;379:1732-1744

17

What's the alternative?

18

19



20

Special Point of View Article

Developing a New, National Approach to Surveillance for Ventilator-Associated Events*

Shelley S. Magill, MD, PhD[1]; Michael Klompas, MD, MPH[2,3,4]; Robert Balk, MD[5,6]; Suzanne M. Burns, RN, ACNP, MSN, RRT[6,7]; Clifford S. Deutschman, MS, MD[6,8]; Daniel Diekema, MD[9,10]; Scott Fridkin, MD[1]; Linda Greene, RN, MPS[11,12]; Alice Guh, MD, MPH[1]; David Gutterman, MD[6,13]; Beth Hammer, RN, MSN, ANP-BC[6,14]; David Henderson, MD[15]; Dean Hess, PhD, RRT[16,17,18]; Nicholas S. Hill, MD[6,19]; Teresa Horan, MPH[1]; Marin Kollef, MD[6,20]; Mitchell Levy, MD[6,21]; Edward Septimus, MD[22,23]; Carole VanAntwerpen, RN, BSN[24,25]; Don Wright, MD, MPH[26]; Pamela Lipsett, MD, MHPE[6,27]

Critical Care Medicine 2013;41:2467-2475

21

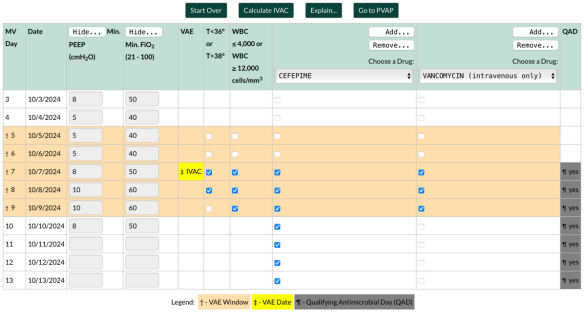
Centers for Disease Control and Prevention
CDC 24/7: Saving Lives, Protecting People™

nhsn.cdc.gov/VAECalculator/vaecalc.html

National Healthcare Safety Network (NHSN)

CDC > NHSN > Materials for Enrolled Facilities

NHSN Ventilator-Associated Event (VAE) Calculator Ver. 10.0



22

Brief report

Assessment of an automated surveillance system for detection of initial ventilator-associated events

Major Article

Development and validation of an automated ventilator-associated event electronic surveillance system: A report of a successful implementation

Electronic Implementation of a Novel Surveillance Paradigm for Ventilator-associated Events

Feasibility and Validation

Building and Validating a Computerized Algorithm for Surveillance of Ventilator-Associated Events

Development, Implementation and Use of Electronic Surveillance for Ventilator-Associated Events (VAE) in Adults

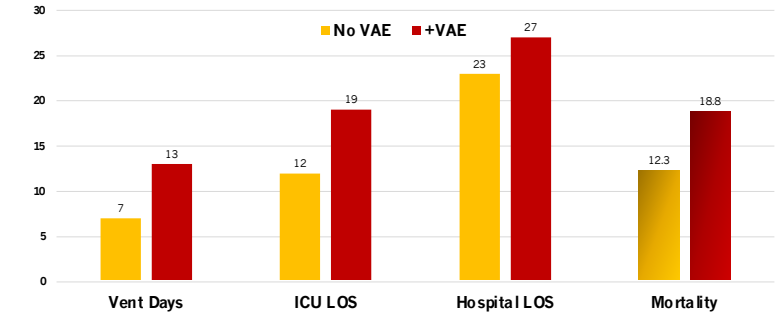
An automated retrospective VAE-surveillance tool for future quality improvement studies

Oliver Wolffers[1,2], Martin Faltys[1], Janos Thomann[1], Stephan M. Jakob[1], Jonas Marschall[3], Tobias M. Merz[1,4] & Rami Sommerstein[1,5]

23

VAE Associated with Poor Outomes

Propensity matched analysis of 1803 VAEs vs 2,319 patients without VAEs, West China Hospital, 2015-2018*



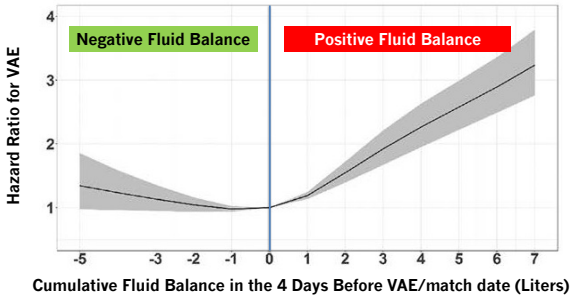
*Variables in propensity score included age, APACHE II, comorbidities, pneumonia, organ failure, surgery, transfusions, immunosuppressives, central lines, IMV after ICU admission

Zhu, *Infect Control Hospital Epidemiol* 2022;1:48-55

24

Strong Association between Fluid Balance and VAEs

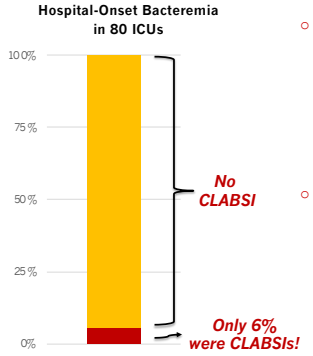
Cumulative fluid balance amongst 1,528 VAE patients matched to 3,038 non-VAE patients on basis of age, time to VAE, and time from ICU admission until initiation of mechanical ventilation, West China Hospital, 2015-2018. Adjusted for demographics, ICU type, comorbidities, ICU diagnosis, APACHE II, meds, procedures, and others.



Wang, *Critical Care Medicine* 2022;50:307-316

25

Hospital Onset Bacteremia (HOB)



- Prospective surveillance for hospital-onset bacteremia in 80 ICUs in 16 hospitals from 2012-2013
  - 11,280 episodes of hospital-onset bacteremia
  - Only 663 were classified as CLABSIs (6%)
- HOB allowed for much more discrimination between ICUs versus using CLABSI rates
  - 75% of ICUs' CLABSI SIRs included 1 versus 25% of ICUs' projected HOB SIRs

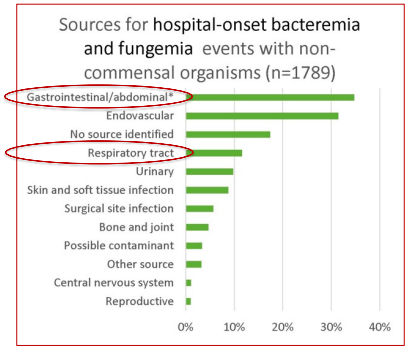
Rock, *ICHE* 2016;37:143-148
Dante, *ICHE* 2019;40:358-361

26

Characteristics of Hospital-Onset Bacteremia Patients

Cross-sectional analysis of 2109 hospital-onset bacteremia, 13 US hospitals, 2016-2019

- Median age: 56
  - ICU: 32%
  - Active malignancy: 22%
  - Any device present: 92%
- Median hospital length-of-stay
  - 24 days
- Hospital mortality: 23%
- 40% of events not captured through existing surveillance definitions

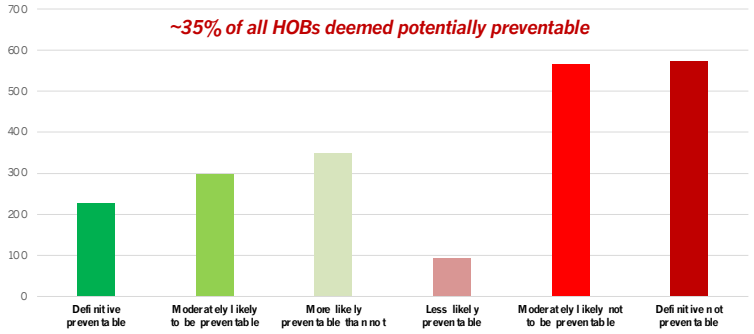


Leekha, *BMJ Quality Safety* 2024;33:487-498

27

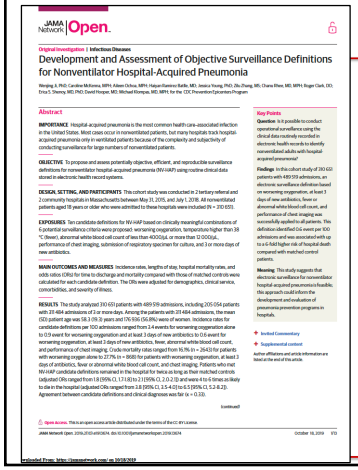
Estimated Preventability of Hospital Onset Bacteremia Events

Cross-sectional analysis of 2109 hospital-onset bacteremia, 13 US hospitals, 2016-2019



Leekha, *BMJ Quality Safety* 2024;33:487-498

28


JAMA Network Open.

Original Investigation | Infectious Diseases

Development and Assessment of Objective Surveillance Definitions for Nonventilator Hospital-Acquired Pneumonia

Abstract

Key Points

Non-Ventilator HAP

- Pilot surveillance strategy using EHR data including vital signs, oxygen devices, oxygen flow rates, lab tests, antibiotics…
- Flags patients on hospital day ≥3 with:
  - Sustained deterioration in oxygenation, *and*
  - Abnormal temperature or WBC count, *and*
  - Order for chest imaging, *and*
  - ≥3 days of new antibiotics
- Applied to 2 years of data from 4 hospitals (489,519 admissions)
  - 0.6 events per 100 hospitalizations
  - 28% mortality
  - Moderate correlation with clinical dx and PNEU

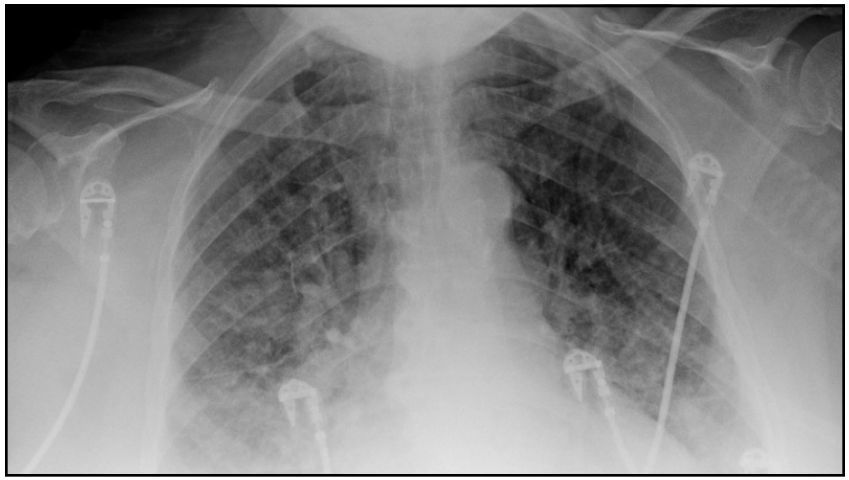
Ji, *JAMA Network Open* 2019;2(10):e1913674
Ramirez Battle, *ICHE* 2020;41:219-221

29

The Challenge of NV-HAP Diagnosis & Surveillance

Imaging	***Systemic Signs***	***Pulmonary Signs***
New or progressive and persistent Infiltrate or Consolidation or Cavitation	Fever >38.0°C WBC <4K or >12K Altered mental status	New onset purulent sputum, change in character of sputum, increased secretions New onset or worsening cough, dyspnea, or tachypnea Rales of bronchial breath sounds Worsening gas exchange, increased oxygen requirement

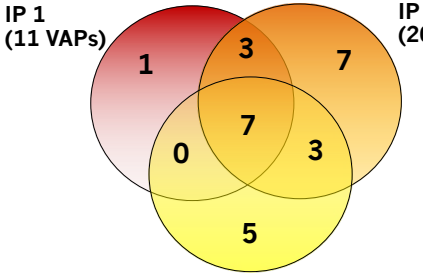
30



31

Interobserver Agreement on PNEU Definitions

50 ventilated patients with respiratory deterioration



Kappa = 0.40

Am J Infect Control 2010:38:237

32

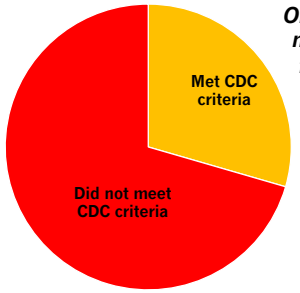
Accuracy of Diagnosis Codes

4,451 patients with pneumonia diagnosis codes, not present-on-admission, 21 hospitals


American Journal of Infection Control

Hospital Acquired Pneumonia Prevention Initiative-2: Incidence of nonventilator hospital-acquired pneumonia in the United States

Dian Baker PhD, RN [1]; Barbara Quinn MS, RN [2]

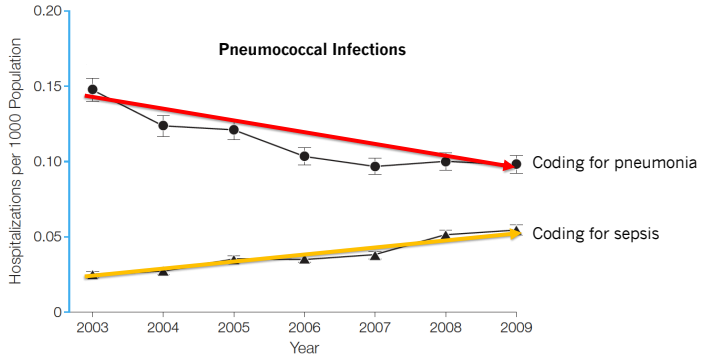


Only 30% of cases met CDC criteria for pneumonia!

Baker, *Am J Infection Control* 2018;46:2-7

33

Coding for Pneumonia is in Flux



Lindenauer, JAMA 2012;307:1405-1413

34

JAMA Network Open

Original Investigation | Infectious Diseases

Development and Assessment of Objective Surveillance Definitions for Nonventilator Hospital-Acquired Pneumonia

Wenjing Ji, PhD; Caroline McKenna, MPH; Aileen Ochoa, MPH; Haiyan Ramirez Batlle, MD; Jessica Young, PhD; Zilu Zhang, MS; Chanu Rhee, MD, MPH; Roger Clark, DO; Erica S. Shenoy, MD, PhD; David Hooper, MD; Michael Klompas, MD, MPH; for the CDC Prevention Epicenters Program

Abstract

IMPORTANCE Hospital-acquired pneumonia is the most common health care–associated infection in the United States. Most cases occur in nonventilated patients, but many hospitals track hospital-acquired pneumonia only in ventilated patients because of the complexity and subjectivity of conducting surveillance for large numbers of nonventilated patients.

OBJECTIVE To propose and assess potentially objective, efficient, and reproducible surveillance definitions for nonventilator hospital-acquired pneumonia (NV-HAP) using routine clinical data stored in electronic health record systems.

DESIGN, SETTING, AND PARTICIPANTS This cohort study was conducted in 2 tertiary referral and

Key Points

Question Is it possible to conduct operational surveillance using the clinical data routinely recorded in electronic health records to identify nonventilated adults with hospital-acquired pneumonia?

Findings In this cohort study of 310 651 patients with 489 519 admissions, an electronic surveillance definition based on worsening oxygenation, at least 3

35

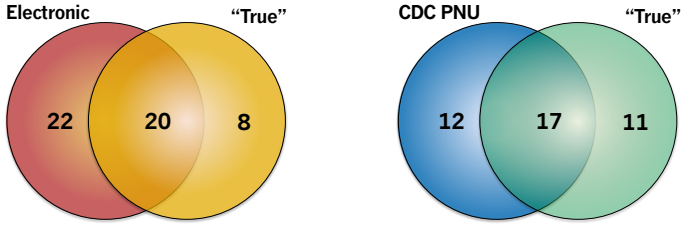
NV-HAP Electronic Surveillance Definition

Worsening Oxygenation	Fever or Leukocytosis	Performance of Chest Imaging	≥3 days of New Antibiotics
Drop in SpO2 to <95% on room air, initiation of supplemental oxygen, or escalation of supplemental oxygen	*WBC <4 or >12 or Temp <36° or >38° C*	*Chest X-ray or CT Scan*	*New = not given in previous 2 days; can be different antibiotics for each day so long as each one is new*

36

Correlation with "True" Pneumonia per Reviewer

Chart review exercise to assess correlation between the electronic NV-HAP surveillance definition vs "truth" (N=120)



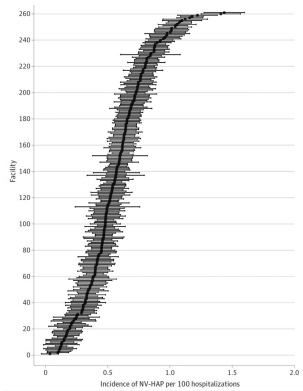
sensitivity 71%
positive predictive value 48%
kappa 0.40

sensitivity 61%
positive predictive value 59%
kappa 0.47

Ramirez Batlle, *ICHE* 2020;41:219-221

37

NV-HAP Incidence and Characteristics



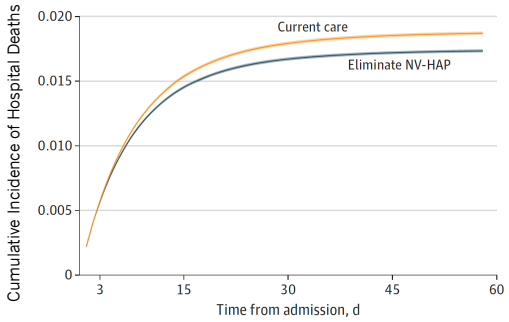
- Electronic NV-HAP criteria applied to 284 US hospitals, 2015-2020
- Median 0.55 NV-HAP events per 100 admissions
- 75% of cases in non-ICU settings
- Crude mortality 22.4%
- Median length-of-stay 16 days (IQR 11-26)
- 250 charts reviewed for accuracy: pneumonia confirmed in 81%

Jones, *JAMA Network Open* 2023;6(5):e2314185

38

What if we could eliminate NV-HAP?

Analysis of 6.02 million admissions to 284 US hospitals. Modeled impact of eliminating NV-HAP accounting for hospital factors and patients' demographics, comorbidities, service, daily vital signs, & daily laboratory test results



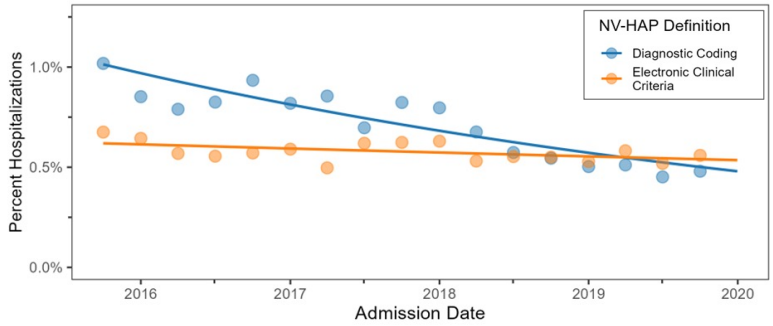
Eliminating NV-HAP would reduce overall hospital mortality by ~7.3%

Jones, *JAMA Network Open* 2023;6(5):e2314185

39

NV-HAP Trends with Codes vs Electronic Clinical Criteria

Trends in NV-HAP rates detected with diagnosis codes vs electronic clinical criteria, 17 VA hospitals, 2015-2019



Jones 2024, unpublished data

40

41

Infection Control & Hospital Epidemiology (2024), **45**, 96–99
doi:10.1017/ice.2023.189

SHEA

Concise Communication

Application of OpenAI GPT-4 for the retrospective detection of catheter-associated urinary tract infections in a fictitious and curated patient data set

Jasmin Perret MSc[a] and Adrian Schmid MD[a]
Infectious Diseases and Hospital Epidemiology, Department of General Internal Medicine, Cantonal Hospital Winterthur, Winterthur, Switzerland

Abstract
The use of the OpenAI GPT-4 model in detecting catheter-associated urinary tract infection (CAUTI) cases in small fictitious and curated patient data sets was investigated. Final analysis of 50 patients including 11 CAUTI cases yielded sensitivity, specificity and positive and negative predictive values of 91%, 92%, 83%, and 96%, respectively.

(Received 4 May 2023; accepted 12 July 2023; electronically published 7 September 2023)

42

Large Language Models for CAUTI Surveillance

- Researchers in Switzerland created two fictitious sets of patients to feed to GPT-4. Case descriptions brief due to token limits (max 8,192)
- Dataset 1 (structured data only, 79 patients)
  - GPT identified 6 CAUTIs (sensitivity 80%, positive predictive value 67%)
    - 4 confirmed
    - 2 false positives (one had 3 species in the urine, one did not meet all criteria in infection window)
    - 1 false negative (catheter removed day before infection)
  - Performance improved with training on the same patients
- Dataset 2 (structured data + symptoms, 50 patients)
  - GPT identified 12 CAUTIs
    - 10 confirmed
    - 2 false positives (catheter only present one day, not all infection criteria in infection window)
    - 1 false negative (catheter removed the day before infection)

Perret, *ICHE* 2024;45:96-99

43

Infection Control & Hospital Epidemiology (2024), 1–4
doi:10.1017/ice.2024.164

SHEA

Concise Communication

Performance of a large language model for identifying central line-associated bloodstream infections (CLABSI) using real clinical notes

Guillermo Rodriguez-Nava MD[1], Goar Egoryan MD[2], Katherine E. Goodman PhD, JD[3,4], Daniel J. Morgan MD, MS[3,5] and Jorge L. Salinas MD[1]
[1]Division of Infectious Diseases & Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA, [2]Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA, [3]Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA, [4]University of Maryland Institute for Health Computing, Bethesda, MD, USA and [5]VA Maryland Healthcare System, Baltimore, MD, USA

Abstract
We evaluated one of the first secure large language models approved for protected health information, for identifying central line-associated bloodstream infections (CLABSIs) using real clinical notes. Despite no pretraining, the model demonstrated rapid assessment and high sensitivity for CLABSI identification. Performance would improve with access to more patient data.

(Received 7 June 2024; accepted 30 August 2024)

44

Large language models for CLABSI surveillance

- Pilot study in Stanford of CLABSI surveillance using GPT 4.0
- 40 patients already reviewed by IPs for possible CLABSI
  - 20 with CLABSI, 20 without CLABSI
- Provided GPT with just the index positive blood culture result & the last two progress notes in the infection window period
- GPT detected 16/20 CLABSIs and 7/20 non-CLABSIs
  - Sensitivity 80%, Specificity 35%
- Investigators attributed 11/17 incorrect classifications to missing information elsewhere in the chart (i.e. not in the last 2 notes)
  - e.g. BSI present on admission, another source for BSI, additional clinical signs, etc.
- Providing this additional information to GPT improved performance
  - Sensitivity 90%, Specificity 75%
- Residual errors due to misclassification of commensals and mucosal barrier injury organisms as pathogens, misclassification of CLABSI organism as commensal, misattribution of BSI to pancreatitis.

Rodriguez-Nava, *ICHE* 2024;ePub; 10.1017/ice.2024.164

45

NEJM AI 2024;1(11)
DOI: 10.1056/AIcs2400420

CASE STUDY

Large Language Models for More Efficient Reporting of Hospital Quality Measures

Aaron Boussina, Ph.D.,[1] Rishivardhan Krishnamoorthy, M.S.,[1] Kimberly Quintero, R.N., M.S.,[2] Shreyansh Joshi, Gabriel Wardi, M.D.,[1,3,4] Hayden Pour, M.S.,[1] Nicholas Hilbert, R.N., M.S.N.,[2] Atul Malhotra, M.D.,[3] Michael Hogarth, M.D.,[3] Arny M. Sitapati, M.D.,[1] Chad VanDenBerg, M.P.H.,[2] Karandeep Singh, M.D., M.M.Sc.,[3] Christopher A. Longhurst, M.D., M.S.,[2] and Shamim Nemati, Ph.D.[1]

Received: April 25, 2024; Revised: August 9, 2025; Accepted: August 16, 2024; Published: October 21, 2024

Abstract

Hospital quality measures are a vital component of a learning health system, yet they can be costly to report, statistically underpowered, and inconsistent due to poor interrater reliability. Large language models (LLMs) have recently demonstrated impressive performance on health care–related tasks and offer a promising way to provide accurate abstraction of complete charts at scale. To evaluate this approach, we deployed an LLM-based system that ingests Fast Healthcare Interoperability Resources data and outputs a completed Severe Sepsis and Septic Shock Management Bundle (SEP-1) abstraction. We tested the system on a sample of 100 manual SEP-1 abstractions that University of California San Diego Health reported to the Centers for Medicare & Medicaid Services in 2022. The LLM system achieved agreement with manual abstractors on the measure category assignment in 90 of the abstractions (90%; κ=0.82; 95% confidence interval, 0.71 to 0.92). Expert review of the 10 discordant cases identified four that were mistakes introduced by manual abstraction. This pilot study suggests that LLMs using interoperable electronic health record data may perform accurate abstractions for complex quality measures. (Funded by the National Institute of Allergy and Infectious Diseases [1R42AI177108-1] and others.)

46

CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES

SEP-1 Mandatory Sepsis Bundle

47

SEP-1 Mandatory Sepsis Bundle

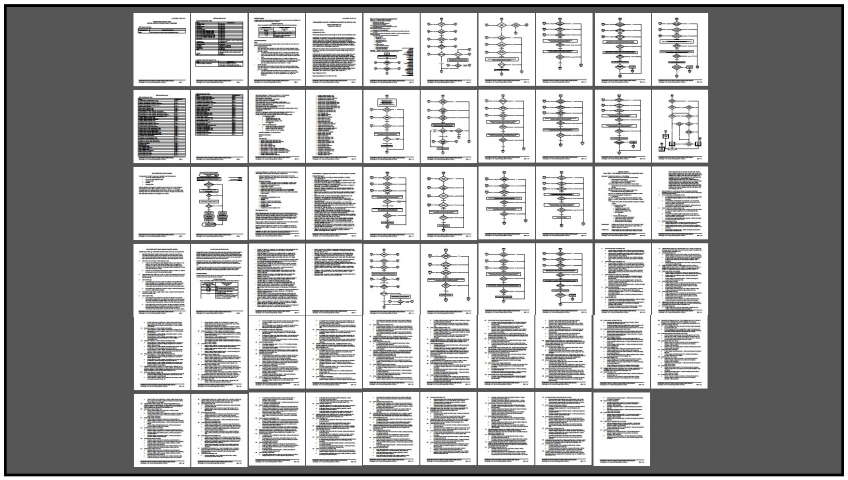
3 Hours

- Measure lactate
- Draw blood cultures
- Administer broad spectrum antibiotics
- Shock? 30cc/kg IV crystalloids

6 Hours

- Repeat lactate if initial level was elevated
- Administer vasopressors if still hypotensive
- Reassess volume status & tissue perfusion

48



49

Large language models for SEP-1 compliance

100 patients with sepsis discharge codes fed into LLM model for SEP-1 assessment, UCSD, 2022

SEP-1 Question	No. (%) of Abstractions Where System Answer Resulted in Agreement with Manual Category Assignment†
Was the patient received as a transfer from an inpatient, outpatient, or emergency/observation department of an outside hospital or from an ambulatory surgery center?‡	99/100 (99)
During this hospital stay, was the patient enrolled in a clinical trial in which patients with the same condition as the measure set were being studied?‡	100/100 (100)
Is there documentation the patient is at least 20 weeks pregnant or within 3 days after delivery at the time severe sepsis is identified?‡	100/100 (100)
Was severe sepsis present?‡	98/100 (98)
When was the last criterion met to establish the presence of severe sepsis?‡	97/100 (97)
Is there documentation that the patient or authorized patient advocate refused either a blood draw, IV fluid administration, or IV antibiotic administration within the specified time frame?‡	100/100 (100)

63 questions in all…

Boussina, NEJM AI 2024; DOI: 10.1056/Alcs2400420

50

Large language models for SEP-1 compliance

100 patients with sepsis discharge codes fed into LLM model for SEP-1 assessment, UCSD, 2022

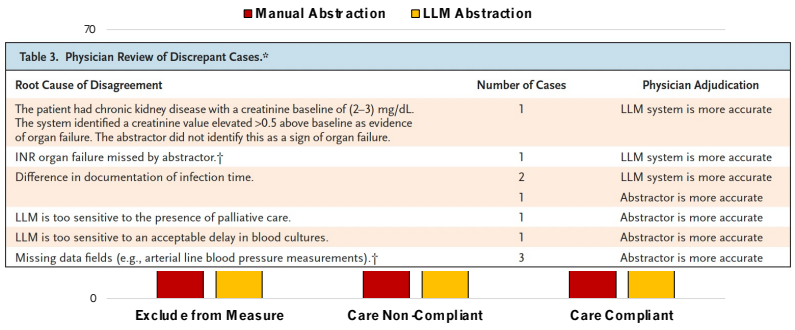


Table 3. Physician Review of Discrepant Cases.*

Root Cause of Disagreement	Number of Cases	Physician Adjudication
The patient had chronic kidney disease with a creatinine baseline of (2–3) mg/dL. The system identified a creatinine value elevated >0.5 above baseline as evidence of organ failure. The abstractor did not identify this as a sign of organ failure.	1	LLM system is more accurate
INR organ failure missed by abstractor.†	1	LLM system is more accurate
Difference in documentation of infection time.	2	LLM system is more accurate
	1	Abstractor is more accurate
LLM is too sensitive to the presence of palliative care.	1	Abstractor is more accurate
LLM is too sensitive to an acceptable delay in blood cultures.	1	Abstractor is more accurate
Missing data fields (e.g., arterial line blood pressure measurements).†	3	Abstractor is more accurate

Boussina, NEJM AI 2024; DOI: 10.1056/Alcs2400420

51

Promising!

Can we also use sepsis as a surveillance target to get a more complete picture of serious infections in hospitals?

52

GET AHEAD OF SEPSIS

KNOW THE RISKS. SPOT THE SIGNS. ACT FAST.

CDC

53

Sepsis Surveillance Methods Using Diagnosis Codes

Angus Method:

1286 infection codes, **13** organ dysfunction codes

Martin Method:

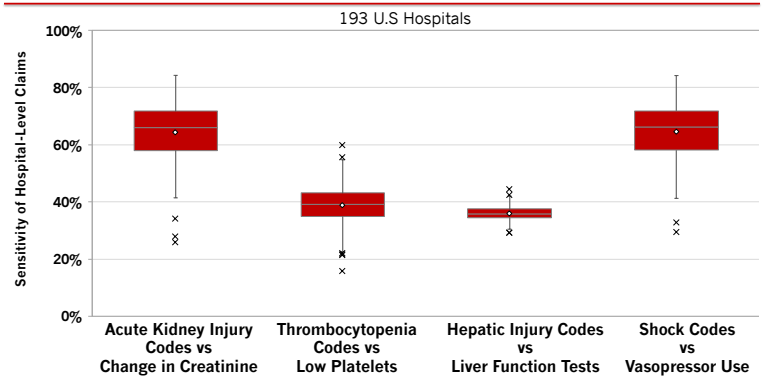
6 infection codes, **13** organ dysfunction codes

Dombrovskiy Method:

18 infection codes, **22** organ dysfunction codes

54

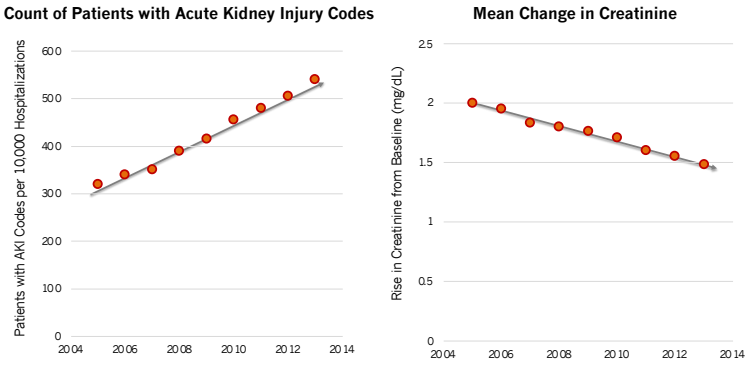
Sensitivity of Diagnosis Codes vs Clinical Criteria



Rhee, *Crit Care Med* 2019;47:493-500

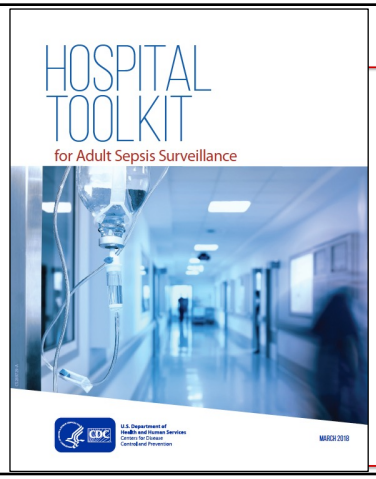
55

Changes in Coding Thresholds



Rhee et al, Critical Care 2015;19:338

56



CDC to the rescue!

- CDC created an EHR-based surveillance definition for sepsis called "Adult Sepsis Events"
- Uses detailed EHR-data to identify patients with suspected infection + organ dysfunction
- Implementation toolkit available that describes the required data elements, how to organize them, and provides analytic code for event detection

cdc.gov/sepsis/clinicaltools

57

Clinical Criteria for Sepsis

Suspected Infection	**Order for blood cultures** *and* **At least four days of new antibiotics**
Concurrent Organ Dysfunction	**Vasopressors** *or* **Initiation of mechanical ventilation** *or* **Doubling in creatinine (exclude dialysis pts)** *or* **Rise in bilirubin to ≥2.0mg/dL** *or* **Fall in platelets to ≤100 cells/μL** *or* **Serum lactate ≥2.0 mmol/L**

If all criteria are met on hospital day ≥3 than classify as

Hospital-Onset Adult Sepsis Event

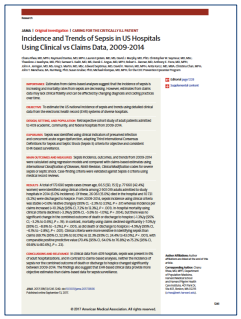
(HO-ASE)

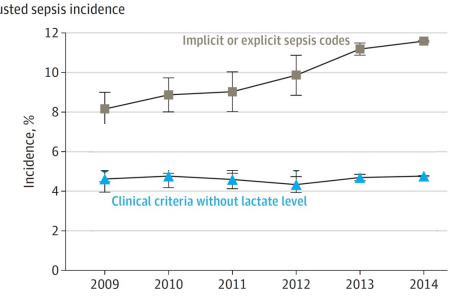
Rhee, *JAMA 2017;318:1241-1249*

58

Estimated Burden of Sepsis, USA, 2014

Estimated US sepsis incidence & mortality using EHR data, 409 hospitals, 2.9 million encounters


Incidence and Trends of Sepsis in US Hospitals Using Clinical vs Claims Data, 2009-2014



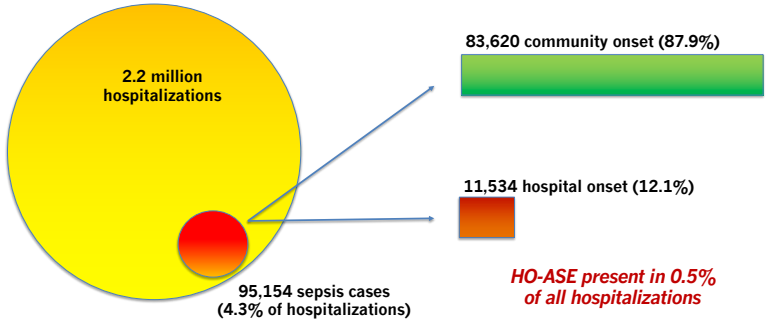
~1.7 million cases, 6.0% of hospital admissions

Rhee et al, *JAMA* 2017;318:1241-1249

59

Hospital-Onset Adult Sepsis Events

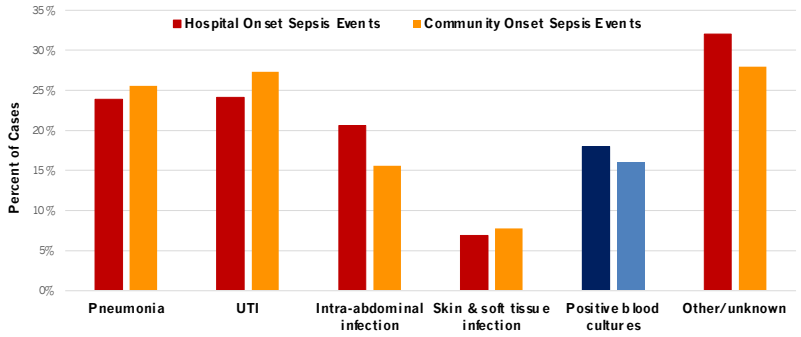
We retrospectively applied Adult Sepsis Event criteria to EHR data from 136 U.S. hospitals, 2009-2015



Rhee *Crit Care Med* 2019;47:1169-1176

60

Syndromes



Rhee *Crit Care Med* 2019;47:1169-1176

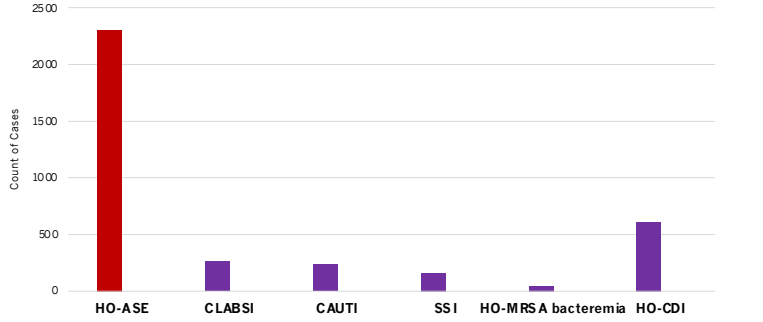
61

How does Hospital-Onset Adult Sepsis Event surveillance compare to current surveillance targets?

62

HO-ASE vs Current Reportable Conditions

Retrospectively applied HO-ASE criteria to EHR data from 282,441 hospitalizations in 3 hospitals and compared case counts, overlap, and mortality for CLABSI, CAUTI, *C. difficile, MRSA bacteremia, and SSI*

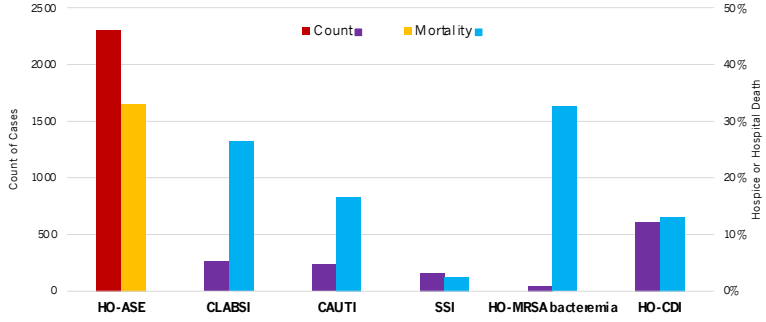


Page, *Clin Infect Dis* 2021;73:1013-1019

63

HO-ASE vs Current Reportable Conditions

Retrospectively applied HO-ASE criteria to EHR data from 282,441 hospitalizations in 3 hospitals and compared case counts, overlap, and mortality for CLABSI, CAUTI, *C. difficile, MRSA bacteremia, and SSI*

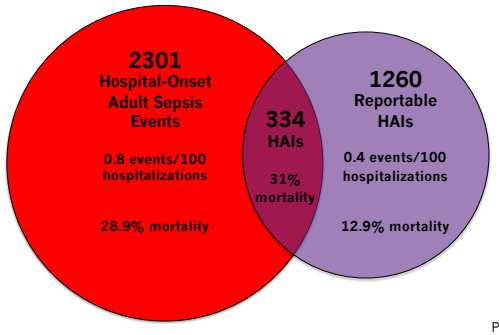


Page, *Clin Infect Dis* 2021;73:1013-1019

64

HO-ASE vs Current Reportable Conditions

Retrospectively applied HO-ASE criteria to EHR data from 282,441 hospitalizations in 3 hospitals and compared case counts, overlap, and mortality for CLABSI, CAUTI, *C. difficile, MRSA bacteremia, and SSI*



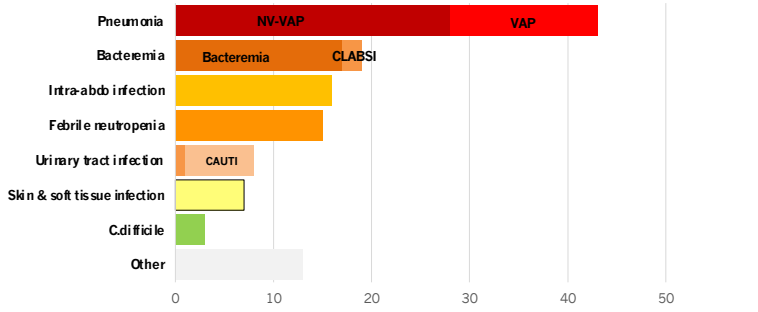
HO-ASE identified a large number of high mortality hospital-onset infections missed by current reportables

Page, *Clin Infect Dis* 2021;73:1013-1019

65

Distribution of Infections Identified by HO-ASE

Retrospectively applied HO-ASE criteria to EHR data from 282,441 hospitalizations in 3 hospitals and compared case counts, overlap, and mortality for CLABSI, CAUTI, *C. difficile, MRSA bacteremia, and SSI*

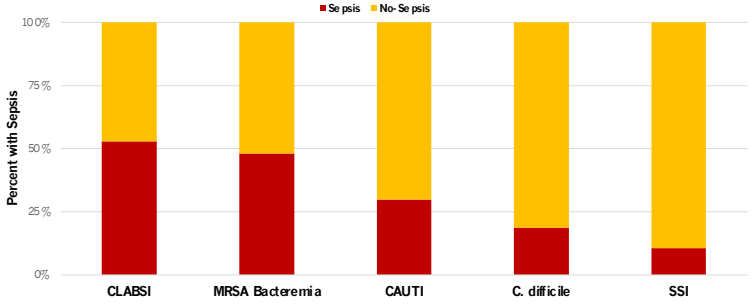


Page, *Clin Infect Dis* 2021;73:1013-1019

66

Proportion of Reportables that Qualified as Sepsis

Retrospectively applied HO-ASE criteria to EHR data from 282,441 hospitalizations in 3 hospitals and compared case counts, overlap, and mortality for CLABSI, CAUTI, *C. difficile, MRSA bacteremia, and SSI*



Page, *Clin Infect Dis* 2021;73:1013-1019

67

Summary

- **Our current surveillance strategy was developed before electronic clinical data became ubiquitous**
  - Focuses on relatively few infections, uses clinically nuanced definitions, requires manual chart review
  - Mandatory reporting requirements have intensified our limited focus
  - Our current approach is labor intensive yet subjective – undermines benchmarking
  - We are missing many high stakes hospital-acquired infections
- **Promising new surveillance strategies emerging (HOB, NV-HAP, HO-ASE)**
  - Use detailed electronic health record data rather than discharge codes or chart reviews
  - Can be efficiently applied to large populations
  - Makes surveillance objective and reproducible
  - Increases the number and breadth of nosocomial infections under surveillance (ICU and non-ICU, device and non-device)

68


Thank You!
For all the lives we touch
Clean hands protect our patients.
Always perform hand hygiene and help others do the same.
BRIGHAM HEALTH
BRIGHAM AND WOMEN'S HOSPITAL
mklompas@bwh.harvard.edu
LIVES WE TOUCH

69