VIZIENT CONNECTIONS SUMMIT

imagine

THE POSSIBILITIES

vizient

UC San Diego Health

AI-Driven Quality Improvement: Reducing Burden, Increasing Impact

Kimberly Quintero, MS, RN, CEN, Quality Improvement Specialist

Aaron Boussina, PhD, Assistant Professor

Chad VanDenBerg, MPH, FACHE, Chief Quality and Patient Safety Officer

Disclosure of relevant financial relationship

Aaron Boussina, PhD, speaker for this educational activity, is a former employee of Clairyon, Inc.

All relevant financial relationships listed for these individual(s) have been mitigated.

All others in a position to control content for this educational activity have no relevant financial relationship(s) to disclose with ineligible companies whose primary business is producing, marketing, selling, re-selling or distributing healthcare products used by or on patients.

Learning Objectives

- Describe how a human-in-the-loop AI approach improves accuracy and efficiency in quality abstraction.
- Explain how integrating AI, automation and dashboards into a structured leadership framework improves efficiency and enhances data-driven decision-making.

imagine
THE POSSIBILITIES

UC San Diego Health

AI-Driven Quality Improvement: Reducing Burden, Increasing Impact

Kimberly Quintero, MS, RN, CEN, Quality Improvement Specialist
Aaron Boussina, PhD, Assistant Professor
Chad VanDenBerg, MPH, FACHE, Chief Quality and Patient Safety Officer

vizient.

UC San Diego Health – World-Class Care

~13,500
Employees

1,101
Licensed Beds

36,000+
Annual Hospital Admissions

>1.5M
Annual Outpatient Visits & Surgeries

UC San Diego Health
The region's *only* academic health system

MISSION

Deliver outstanding patient care through commitment to the community, groundbreaking research and inspired teaching.

VISION

Create a healthier world – one life at a time – through new science, new medicine and new cures.

THE LEAPFROG GROUP TOP TEACHING HOSPITAL 2023

MAGNET RECOGNIZED AMERICAN NURSES CREDENTIALING CENTER

NCI Designated Comprehensive Cancer Center

2024 LGBTQ+ HEALTHCARE EQUALITY HIGH PERFORMER Healthcare Equality Index

UC San Diego Health BEST HOSPITALS USNews Leading the Way *in San Diego*

Bernard A Birnbaum, MD Quality Leadership vizient.

Patty Maysent, MPH
Chief Executive Officer
UC San Diego Health

Christopher Longhurst, MD, MS
Chief Clinical & Innovation Officer
UC San Diego Health

Journey to High Reliability

A highly reliable, learning health system

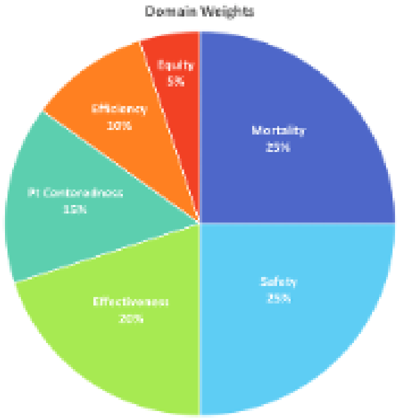
Our Journey to High Reliability blends the principles of a learning health system with a high reliability model for healthcare.

- High Reliability Organization
- Zero Preventable Harm
- Leading Academic Healthcare System
- Safety Culture
- People

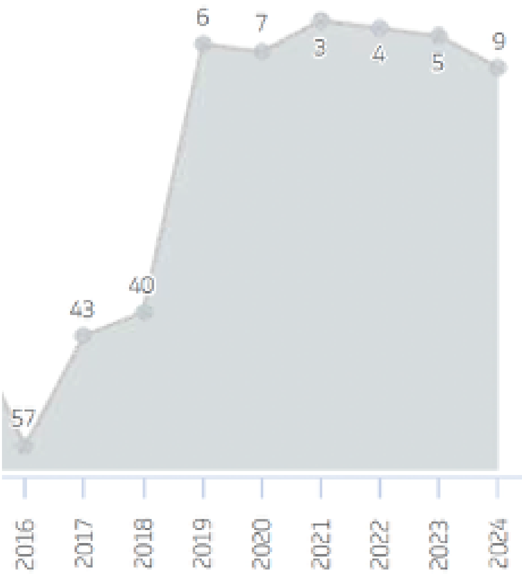
SAFETY

Experience - Leading the Way

Vizient Quality & Accountability (Q&A) Scorecard



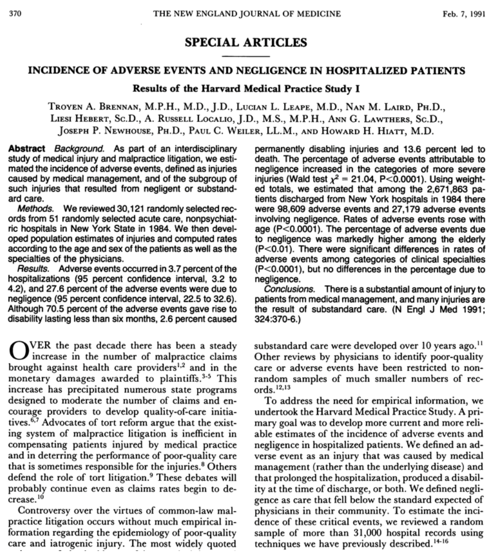
Vizient Quality & Accountability Scorecard Rank (n = 115)



Traditional patient safety efforts are not enough…

Not nearly enough progress

3.7% of inpatients had
adverse events with harm



370 THE NEW ENGLAND JOURNAL OF MEDICINE Feb. 7, 1991

SPECIAL ARTICLES

INCIDENCE OF ADVERSE EVENTS AND NEGLIGENCE IN HOSPITALIZED PATIENTS

Results of the Harvard Medical Practice Study I

Troyen A. Brennan, M.P.H., M.D., J.D., Lucian L. Leape, M.D., Nan M. Laird, Ph.D.,
Liesi Hebert, Sc.D., A. Russell Localio, J.D., M.S., M.P.H., Ann G. Lawthers, Sc.D.,
Joseph P. Newhouse, Ph.D., Paul C. Weiler, LL.M., and Howard H. Hiatt, M.D.

Abstract *Background.* As part of an interdisciplinary study of medical injury and malpractice litigation, we estimated the incidence of adverse events, defined as injuries caused by medical management, and of the subgroup of such injuries that resulted from negligent or substandard care.

Methods. We reviewed 30,121 randomly selected records from 51 randomly selected acute care, nonpsychiatric hospitals in New York State in 1984. We then developed population estimates of injuries and computed rates according to the age and sex of the patients as well as the specialties of the physicians.

Results. Adverse events occurred in 3.7 percent of the hospitalizations (95 percent confidence interval, 3.2 to 4.2), and 27.6 percent of the adverse events were due to negligence (95 percent confidence interval, 22.5 to 32.6). Although 70.5 percent of the adverse events gave rise to disability lasting less than six months, 2.6 percent caused permanently disabling injuries and 13.6 percent led to death. The percentage of adverse events attributable to negligence increased in the categories of more severe injuries (Wald test $\chi^2 = 21.04$, $P<0.0001$). Using weighted totals, we estimated that among the 2,671,863 patients discharged from New York hospitals in 1984 there were 98,609 adverse events and 27,179 adverse events involving negligence. Rates of adverse events rose with age ($P<0.0001$). The percentage of adverse events due to negligence was markedly higher among the elderly ($P<0.01$). There were significant differences in rates of adverse events among categories of clinical specialties ($P<0.0001$), but no differences in the percentage due to negligence.

Conclusions. There is a substantial amount of injury to patients from medical management, and many injuries are the result of substandard care. (N Engl J Med 1991; 324:370-6.)

OVER the past decade there has been a steady increase in the number of malpractice claims brought against health care providers[1,2] and in the monetary damages awarded to plaintiffs.[3-5] This increase has precipitated numerous state programs designed to moderate the number of claims and encourage providers to develop quality-of-care initiatives.[6,7] Advocates of tort reform argue that the existing system of malpractice litigation is inefficient in compensating patients injured by medical practice and in deterring the performance of poor-quality care that is sometimes responsible for the injuries.[8] Others defend the role of tort litigation.[9] These debates will probably continue even as claims rates begin to decrease.[10]

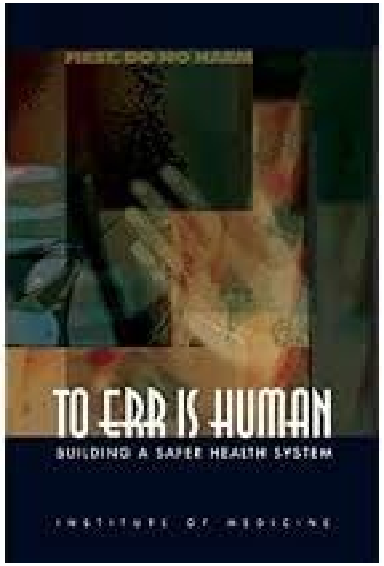
Controversy over the virtues of common-law malpractice litigation occurs without much empirical information regarding the epidemiology of poor-quality care and iatrogenic injury. The most widely quoted

substandard care were developed over 10 years ago.[11] Other reviews by physicians to identify poor-quality care or adverse events have been restricted to nonrandom samples of much smaller numbers of records.[12,13]

To address the need for empirical information, we undertook the Harvard Medical Practice Study. A primary goal was to develop more current and more reliable estimates of the incidence of adverse events and negligence in hospitalized patients. We defined an adverse event as an injury that was caused by medical management (rather than the underlying disease) and that prolonged the hospitalization, produced a disability at the time of discharge, or both. We defined negligence as care that fell below the standard expected of physicians in their community. To estimate the incidence of these critical events, we reviewed a random sample of more than 31,000 hospital records using techniques we have previously described.[14-16]

Harvard Medical Practice Study
NEJM
1991

~3% of inpatients had
adverse events and
~50% preventable



FIRST, DO NO HARM

TO ERR IS HUMAN

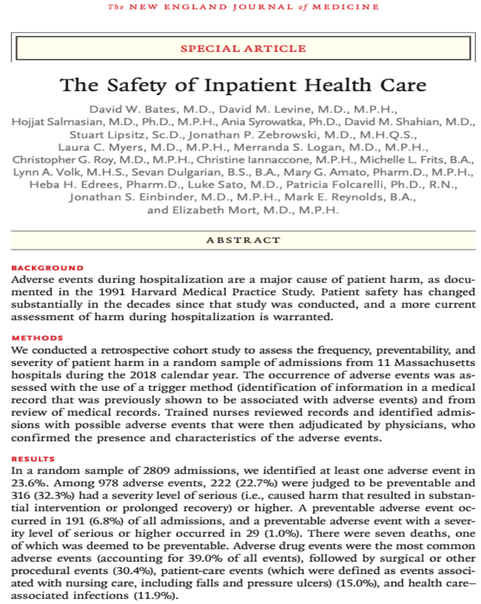
BUILDING A SAFER HEALTH SYSTEM

INSTITUTE OF MEDICINE

To Err is Human
IOM
1999

Traditional patient safety efforts

~25% of hospitalizations had
adverse events and
~25% preventable



The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

The Safety of Inpatient Health Care

David W. Bates, M.D., David M. Levine, M.D., M.P.H.,
Hojjat Salmasian, M.D., Ph.D., M.P.H., Ania Syrowatka, Ph.D., David M. Shahian, M.D.,
Stuart Lipsitz, Sc.D., Jonathan P. Zebrowski, M.D., M.H.Q.S.,
Laura C. Myers, M.D., M.P.H., Merranda S. Logan, M.D., M.P.H.,
Christopher G. Roy, M.D., M.P.H., Christine Iannaccone, M.P.H., Michelle L. Frits, B.A.,
Lynn A. Volk, M.H.S., Sevan Dulgarian, B.S., B.A., Mary G. Amato, Pharm.D., M.P.H.,
Heba H. Edrees, Pharm.D., Luke Sato, M.D., Patricia Folcarelli, Ph.D., R.N.,
Jonathan S. Einbinder, M.D., M.P.H., Mark E. Reynolds, B.A.,
and Elizabeth Mort, M.D., M.P.H.

ABSTRACT

BACKGROUND

Adverse events during hospitalization are a major cause of patient harm, as documented in the 1991 Harvard Medical Practice Study. Patient safety has changed substantially in the decades since that study was conducted, and a more current assessment of harm during hospitalization is warranted.

METHODS

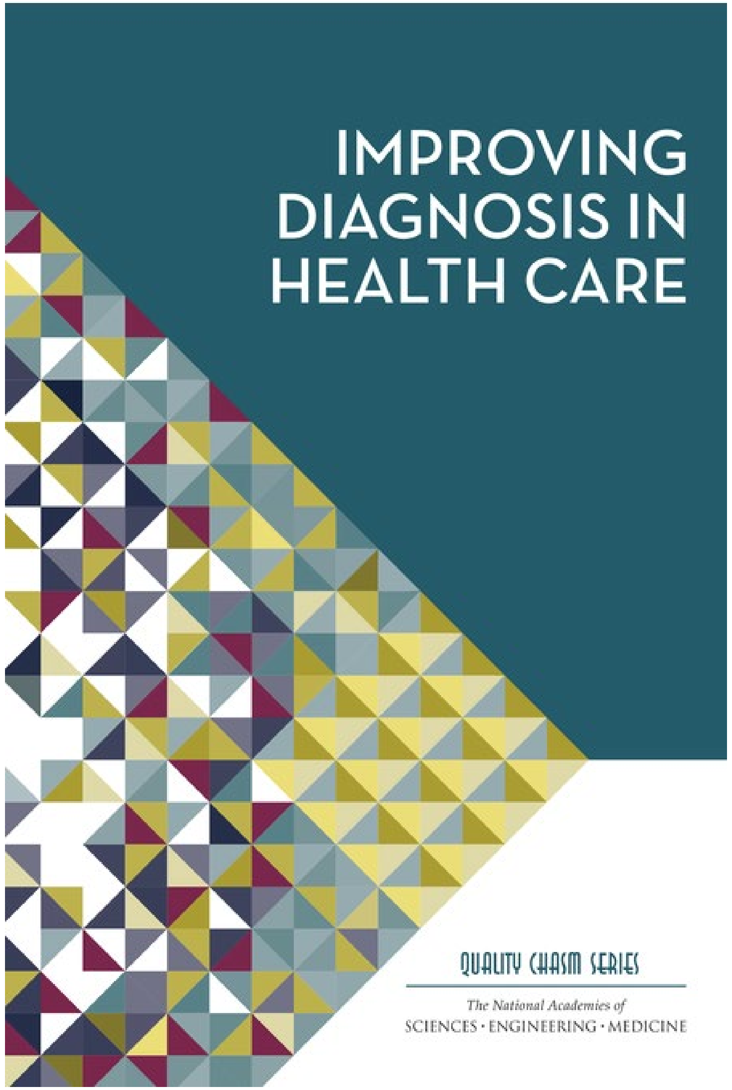
We conducted a retrospective cohort study to assess the frequency, preventability, and severity of patient harm in a random sample of admissions from 11 Massachusetts hospitals during the 2018 calendar year. The occurrence of adverse events was assessed with the use of a trigger method (identification of information in a medical record that was previously shown to be associated with adverse events) and from review of medical records. Trained nurses reviewed records and identified admissions with possible adverse events that were then adjudicated by physicians, who confirmed the presence and characteristics of the adverse events.

RESULTS

In a random sample of 2809 admissions, we identified at least one adverse event in 23.6%. Among 978 adverse events, 222 (22.7%) were judged to be preventable and 316 (32.3%) had a severity level of serious (i.e., caused harm that resulted in substantial intervention or prolonged recovery) or higher. A preventable adverse event occurred in 191 (6.8%) of all admissions, and a preventable adverse event with a severity level of serious or higher occurred in 29 (1.0%). There were seven deaths, one of which was deemed to be preventable. Adverse drug events were the most common adverse events (accounting for 39.0% of all events), followed by surgical or other procedural events (30.4%), patient-care events (which were defined as events associated with nursing care, including falls and pressure ulcers) (15.0%), and health care–associated infections (11.9%).

The Safety of Inpatient Health Care
NEJM
2023

…And there's more below the surface



Improving Diagnosis in Health Care

Committee on Diagnostic Error in Health Care; Board on Health Care Services; Institute of Medicine; The National Academies of Sciences, Engineering, and Medicine
Erin P. Balogh, Bryan T. Miller, John R. Ball, editors.

Washington (DC): National Academies Press (US); 2015 Dec 29.

PMID: 26803862 Bookshelf ID: NBK338596 DOI: 10.17226/21794

Free Books & Documents

Diagnostic Errors occur in an estimated:
- ✓ 5% of outpatient visits
- ✓ 10% of hospital inpatient deaths
- ✓ 6-17% of hospital adverse events

"… *the committee concluded that most people will experience at least one diagnostic error in their lifetime*…"

Key Points

- AI ~~will~~ **is fundamentally changing** approach to Quality and Safety
- Framing matters:
  - Administrative Burden
  - Detection and Deeper Learning
  - Prediction and Prevention
- AI efforts are reviewed by our governance to ensure each algorithm is fair, appropriate, valid, effective, and safe (FAVES)

The San Diego Union-Tribune

A $22 million donation to UC San Diego Health will establish a mission control center to manage emerging AI

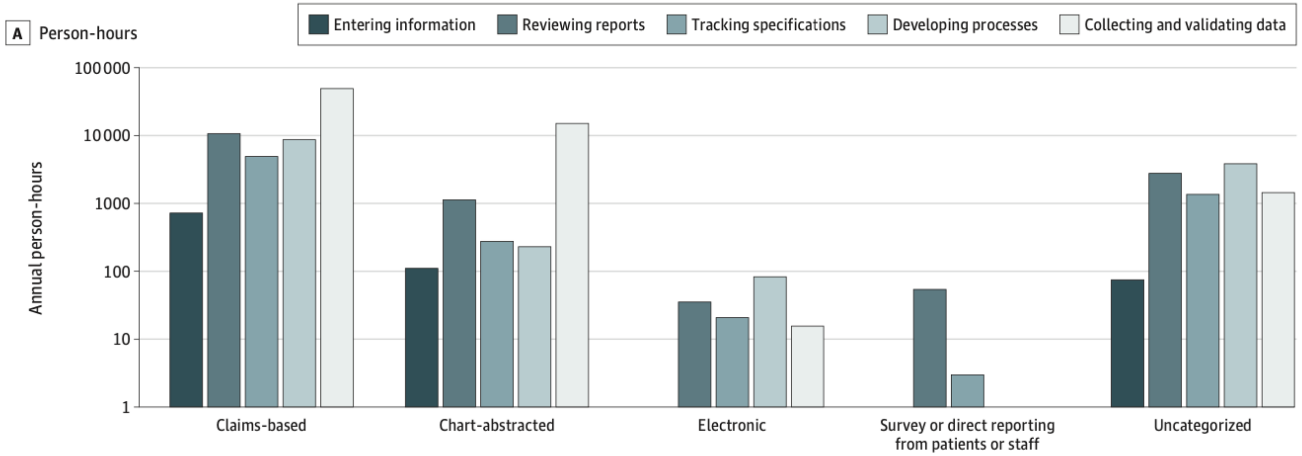
Dr. Christopher Longhurst, medical and digital director at UC San Diego Health, explains plans for a digital command center at the Jacobs Medical Center in La Jolla during a symposium on Friday. (Paul Sisson)

https://www.sandiegouniontribune.com/2023/05/05/a-22-million-donation-to-uc-san-diego-health-will-establish-a-mission-control-center-to-manage-emerging-ai-4/

Leveraging AI to Lower the Administrative Burden of Quality Reporting

Preparing and reporting inpatient quality metrics required > $5 Million in personnel costs plus more than $600 Thousand in vendor fees

Figure 2. Estimated Person-Hour and Personnel Cost by Quality Reporting Activity at a Large Academic Medical Center in Maryland



JAMA. 2023;329(21):1840-1847. doi:10.1001/jama.2023.7271

JAMA | Original Investigation

The Volume and Cost of Quality Metric Reporting

Anirudh Saraswathula, MD, MS; Samantha J. Merck, MD, MHS; Ge Bai, PhD, CPA; Christine M. Weston, PhD; Elizabeth Ann Skinner, MSW; April Taylor, MHA, MS; Allen Kachalia, MD, JD; Renee Demski, MSW, MBA; Albert W. Wu, MD, MPH; Stephen A. Berry, MD, PhD

IMPORTANCE US hospitals report data on many health care quality metrics to government and independent health care rating organizations, but the annual cost to acute care hospitals of measuring and reporting quality metric data, independent of resources spent on quality interventions, is not well known.

OBJECTIVE To evaluate externally reported inpatient quality metrics for adult patients and estimate the cost of data collection and reporting, independent of quality-improvement efforts.

DESIGN, SETTING, AND PARTICIPANTS Retrospective time-driven activity-based costing study at the Johns Hopkins Hospital (Baltimore, Maryland) with hospital personnel involved in quality metric reporting processes interviewed between January 1, 2019, and June 30, 2019, about quality reporting activities in the 2018 calendar year.

MAIN OUTCOMES AND MEASURES Outcomes included the number of metrics, annual person-hours per metric type, and annual personnel cost per metric type.

RESULTS A total of 162 unique metrics were identified, of which 96 (59.3%) were claims-based, 107 (66.0%) were outcome metrics, and 101 (62.3%) were related to patient safety. Preparing and reporting data for these metrics required an estimated 108 478 person-hours, with an estimated personnel cost of $5 038 218.28 (2022 USD) plus an additional $602 730.66 in vendor fees. Claims-based (96 metrics; $37 553.58 per metric per year) and chart-abstracted (26 metrics; $33 871.30 per metric per year) metrics used the most resources per metric, while electronic metrics consumed far less (4 metrics; $1901.58

Leveraging AI to Lower the Administrative Burden of Healthcare

imagine THE POSSIBILITIES

Randomized Controlled Trial > JAMA Netw Open. 2024 Apr 1;7(4):e246565.
doi: 10.1001/jamanetworkopen.2024.6565.

AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication

Ming Tai-Seale[1,2], Sally L Baxter[2,3], Florin Vaida[4], Amanda Walker[1], Amy M Sitapati[2], Chad Osborne[1], Joseph Diaz[2], Nimit Desai[5], Sophie Webb[1], Gregory Polston[6], Teresa Helsten[2], Erin Gross[7], Jessica Thackaberry[8], Ammar Mandvi[1], Dustin Lillie[1], Steve Li[2], Geneen Gin[1], Suraj Achar[1], Heather Hofflich[2], Christopher Sharp[9], Marlene Millen[2], Christopher A Longhurst[2,10]

PMID: 38619840 PMCID: PMC11019394 DOI: 10.1001/jamanetworkopen.2024.6565

JAMIA Open. 2025 Feb 21;8(1):ooaf013. doi: 10.1093/jamiaopen/ooaf013

Enhancing clinical documentation with ambient artificial intelligence: a quality improvement survey assessing clinician perspectives on work burden, burnout, and job satisfaction

Michael Albrecht[1,✉], Denton Shanks[2], Tina Shah[3], Taina Hudson[4], Jeffrey Thompson[5], Tanya Filardi[6], Kelli Wright[7], Gregory A Ator[8], Timothy Ryan Smith[9]

PMCID: PMC11843214 PMID: 39991073

COMMENTARY

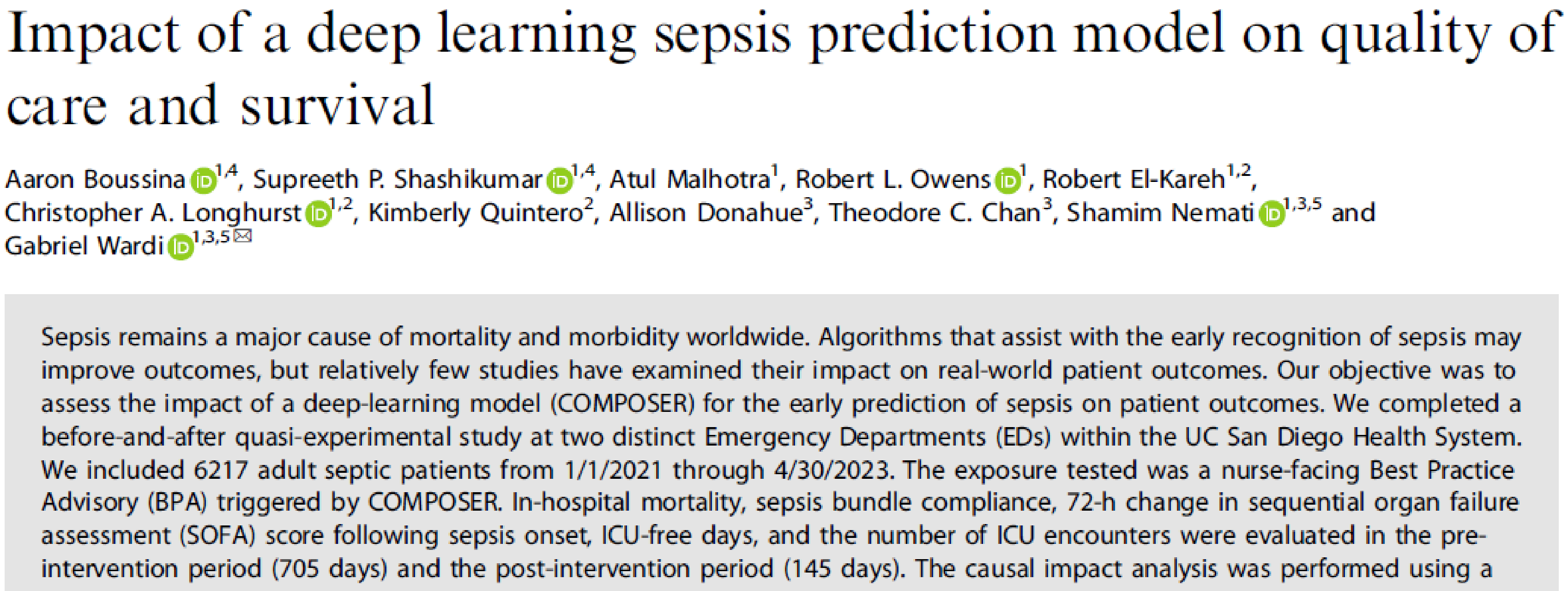
Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation

Authors: Aaron A. Tierney, PhD, Gregg Gayre, MD, Brian Hoberman, MD, MBA, Britt Mattern, MBA, Manuel Ballesca, MD, Patricia Kipnis, PhD, Vincent Liu, MD, MS, and Kristine Lee, MD Author Info & Affiliations

Published February 21, 2024 | NEJM Catal Innov Care Deliv 2024;5(3) | DOI: 10.1056/CAT.23.0404 | VOL. 5 NO. 3
Copyright © 2024

From “automated” responses to ambient documentation and more…

Using AI to Predict Sepsis



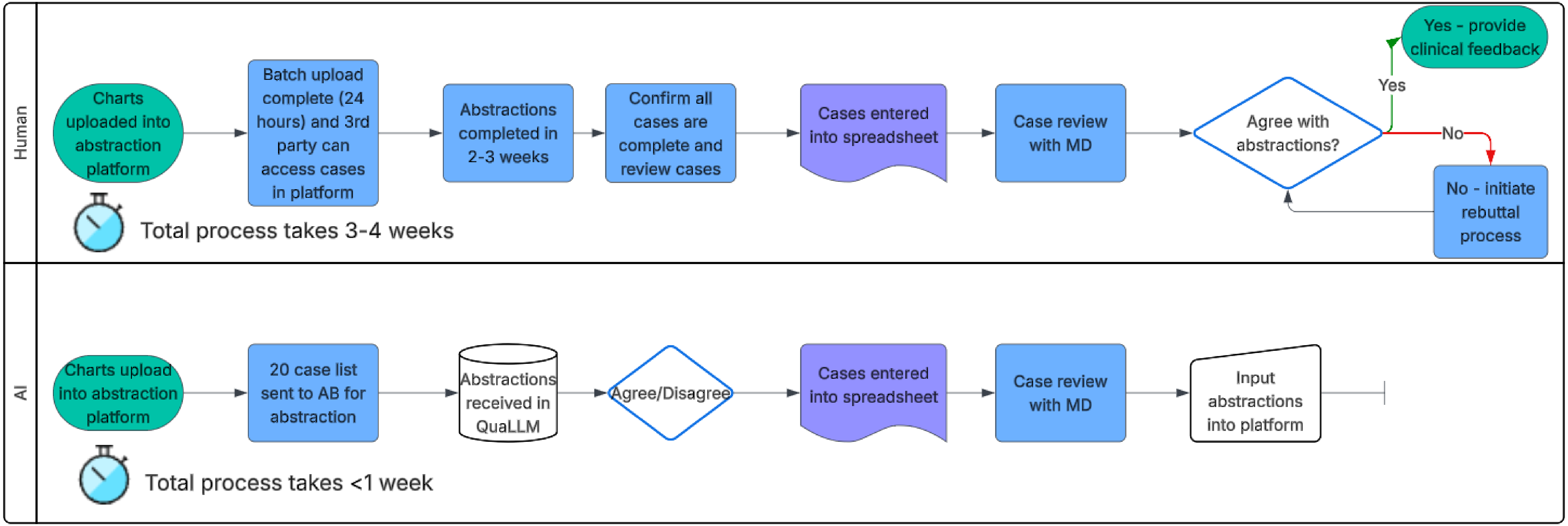
Impact of a deep learning sepsis prediction model on quality of care and survival

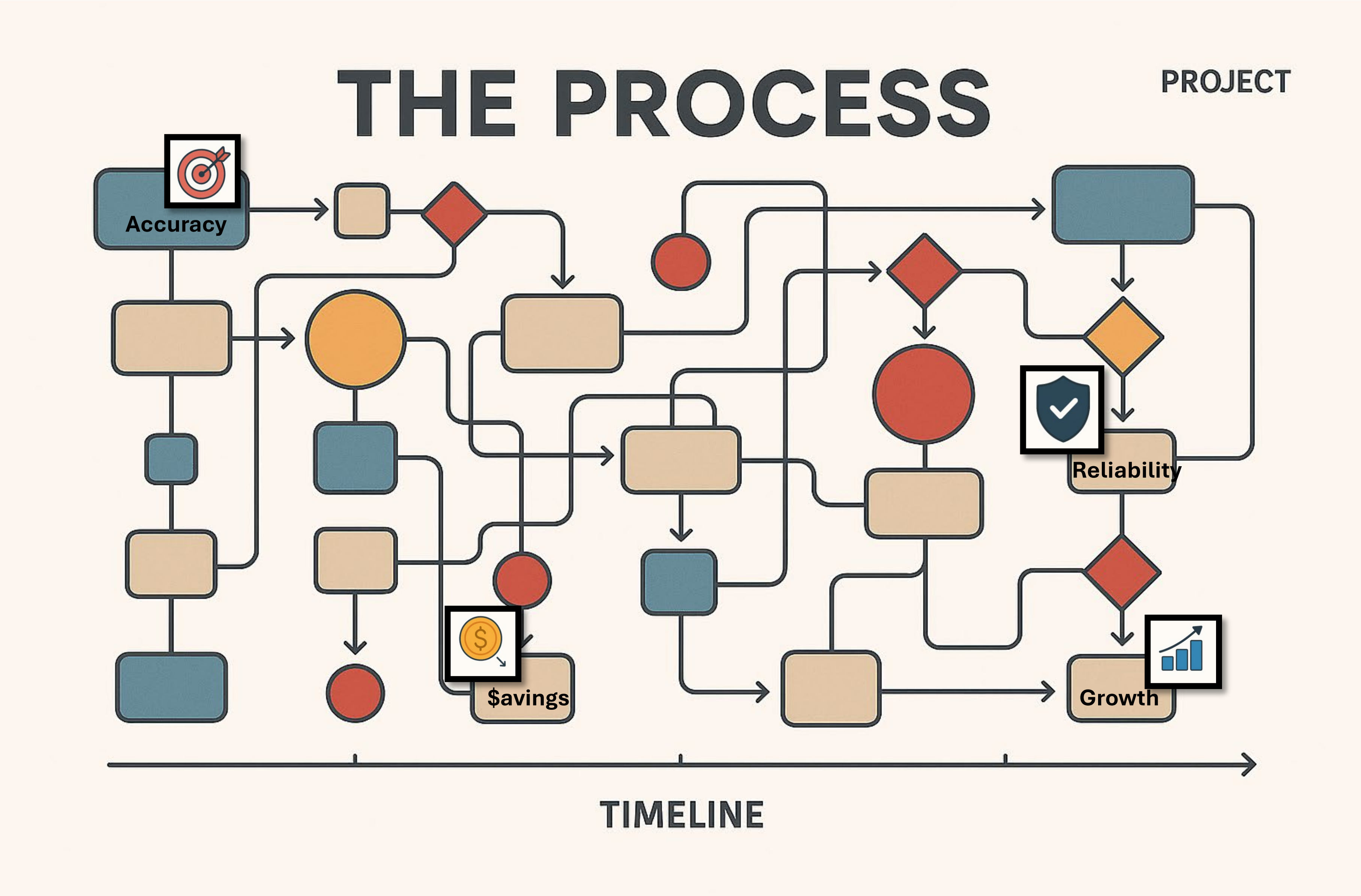
Aaron Boussina[1,4], Supreeth P. Shashikumar[1,4], Atul Malhotra[1], Robert L. Owens[1], Robert El-Kareh[1,2], Christopher A. Longhurst[1,2], Kimberly Quintero[2], Allison Donahue[3], Theodore C. Chan[3], Shamim Nemati[1,3,5] and Gabriel Wardi[1,3,5]✉

Sepsis remains a major cause of mortality and morbidity worldwide. Algorithms that assist with the early recognition of sepsis may improve outcomes, but relatively few studies have examined their impact on real-world patient outcomes. Our objective was to assess the impact of a deep-learning model (COMPOSER) for the early prediction of sepsis on patient outcomes. We completed a before-and-after quasi-experimental study at two distinct Emergency Departments (EDs) within the UC San Diego Health System. We included 6217 adult septic patients from 1/1/2021 through 4/30/2023. The exposure tested was a nurse-facing Best Practice Advisory (BPA) triggered by COMPOSER. In-hospital mortality, sepsis bundle compliance, 72-h change in sequential organ failure assessment (SOFA) score following sepsis onset, ICU-free days, and the number of ICU encounters were evaluated in the pre-intervention period (705 days) and the post-intervention period (145 days). The causal impact analysis was performed using a Bayesian structural time-series approach with confounder adjustments to assess the significance of the exposure at the 95% confidence level. The deployment of COMPOSER was significantly associated with a 1.9% absolute reduction (17% relative

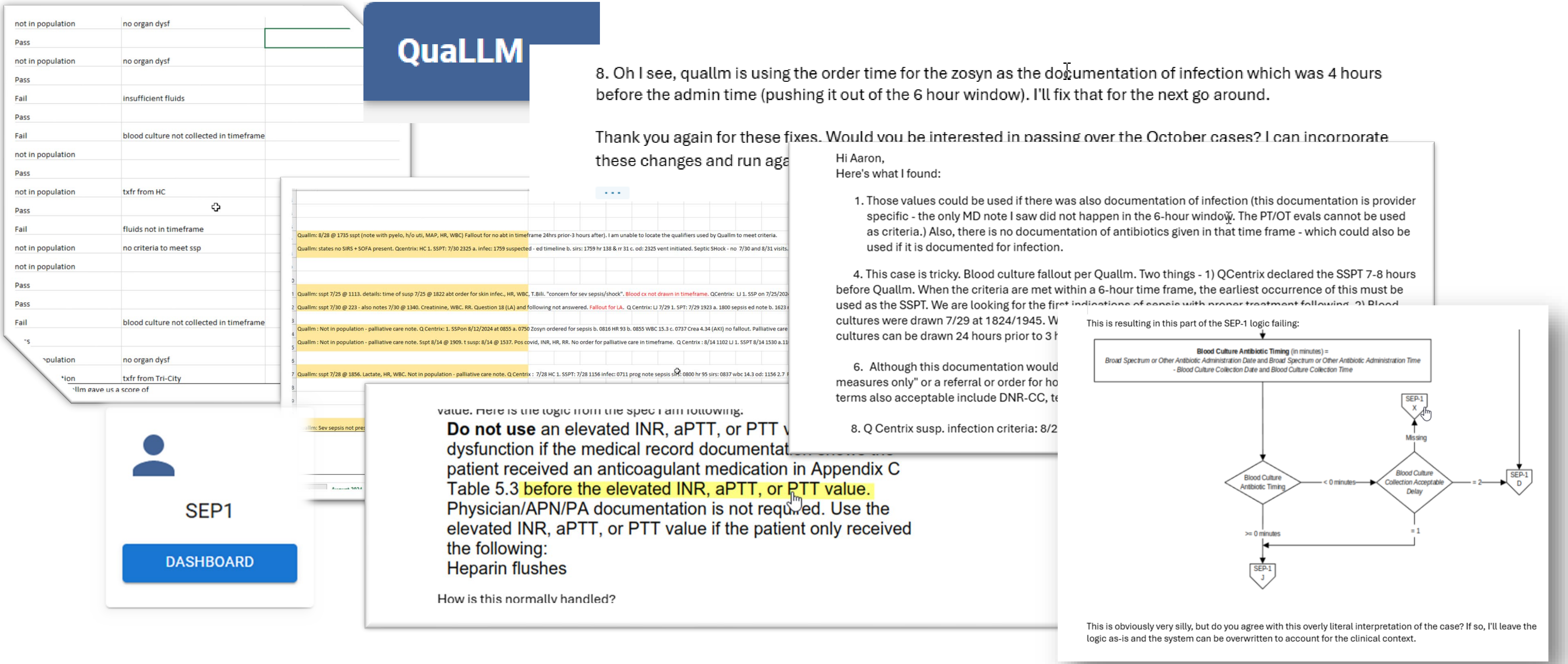
The deployment of COMPOSER was significantly associated with a 1.9% absolute reduction (17% relative decrease) in in-hospital sepsis mortality (95% CI, 0.3%–3.5%), a 5.0% absolute increase (10% relative increase) in sepsis bundle compliance

Abstraction Process Comparison




THE PROCESS
PROJECT
Accuracy
$avings
Reliability
Growth
TIMELINE

Learning SEP-1



not in population	no organ dysf
Pass	
not in population	no organ dysf
Pass	
Fail	insufficient fluids
Pass	
Fail	blood culture not collected in timeframe
not in population	
Pass	
not in population	txfr from HC
Pass	
Fail	fluids not in timeframe
not in population	no criteria to meet ssp
not in population	
Pass	
Pass	
Fail	blood culture not collected in timeframe
s	
pulation	no organ dysf
tion	txfr from Tri-City

llm gave us a score of

QuaLLM

Quallm: 8/28 @ 1735 sspt (note with pyelo, h/o uti, MAP, HR, WBC) Fallout for no abt in timeframe 24hrs prior-3 hours after). I am unable to locate the qualifiers used by Quallm to meet criteria.

Quallm: states no SIRS + SOFA present. Qcentrix: HC 1. SSPT: 7/30 2325 a. infec: 1759 suspected - ed timeline b. sirs: 1759 hr 138 & rr 31 c. od: 2325 vent initiated. Septic SHock - no 7/30 and 8/31 visits.

Quallm: sspt 7/25 @ 1113. details: time of susp 7/25 @ 1822 abt order for skin infec., HR, WBC, T.Bili. "concern for sev sepsis/shock". Blood cx not drawn in timeframe. QCentrix: LJ 1. SSP on 7/25/202

Quallm: sspt 7/30 @ 223 - also notes 7/30 @ 1340. Creatinine, WBC. RR. Question 18 (LA) and following not answered. Fallout for LA. Q Centrix: LJ 7/29 1. SPT: 7/29 1923 a. 1800 sepsis ed note b. 1623

Quallm : Not in population - palliative care note. Q Centrix: 1. SSPon 8/12/2024 at 0855 a. 0750 Zosyn ordered for sepsis b. 0816 HR 93 b. 0855 WBC 15.3 c. 0737 Crea 4.34 (AKI) no fallout. Palliative care

Quallm : Not in population - palliative care note. Sspt 8/14 @ 1909. t susp: 8/14 @ 1537. Pos covid, INR, HR, RR. No order for palliative care in timeframe. Q Centrix : 8/14 1102 LJ 1. SSPT 8/14 1530 a.11

Quallm: sspt 7/28 @ 1856. Lactate, HR, WBC. Not in population - palliative care note. Q Centrix : 7/28 HC 1. SSPT: 7/28 1156 infec: 0711 prog note sepsis sirs: 0800 hr 95 sirs: 0837 wbc 14.3 od: 1156 2.7

llm: Sev sepsis not pres

8. Oh I see, quallm is using the order time for the zosyn as the documentation of infection which was 4 hours before the admin time (pushing it out of the 6 hour window). I'll fix that for the next go around.

Thank you again for these fixes. Would you be interested in passing over the October cases? I can incorporate these changes and run aga

Hi Aaron,
Here's what I found:

1. Those values could be used if there was also documentation of infection (this documentation is provider specific - the only MD note I saw did not happen in the 6-hour window. The PT/OT evals cannot be used as criteria.) Also, there is no documentation of antibiotics given in that time frame - which could also be used if it is documented for infection.

4. This case is tricky. Blood culture fallout per Quallm. Two things - 1) QCentrix declared the SSPT 7-8 hours before Quallm. When the criteria are met within a 6-hour time frame, the earliest occurrence of this must be used as the SSPT. We are looking for the first indications of sepsis with proper treatment following. 2) Blood cultures were drawn 7/29 at 1824/1945. W cultures can be drawn 24 hours prior to 3 h

6. Although this documentation would measures only" or a referral or order for ho terms also acceptable include DNR-CC, te

8. Q Centrix susp. infection criteria: 8/2

This is resulting in this part of the SEP-1 logic failing:

This is obviously very silly, but do you agree with this overly literal interpretation of the case? If so, I'll leave the logic as-is and the system can be overwritten to account for the clinical context.

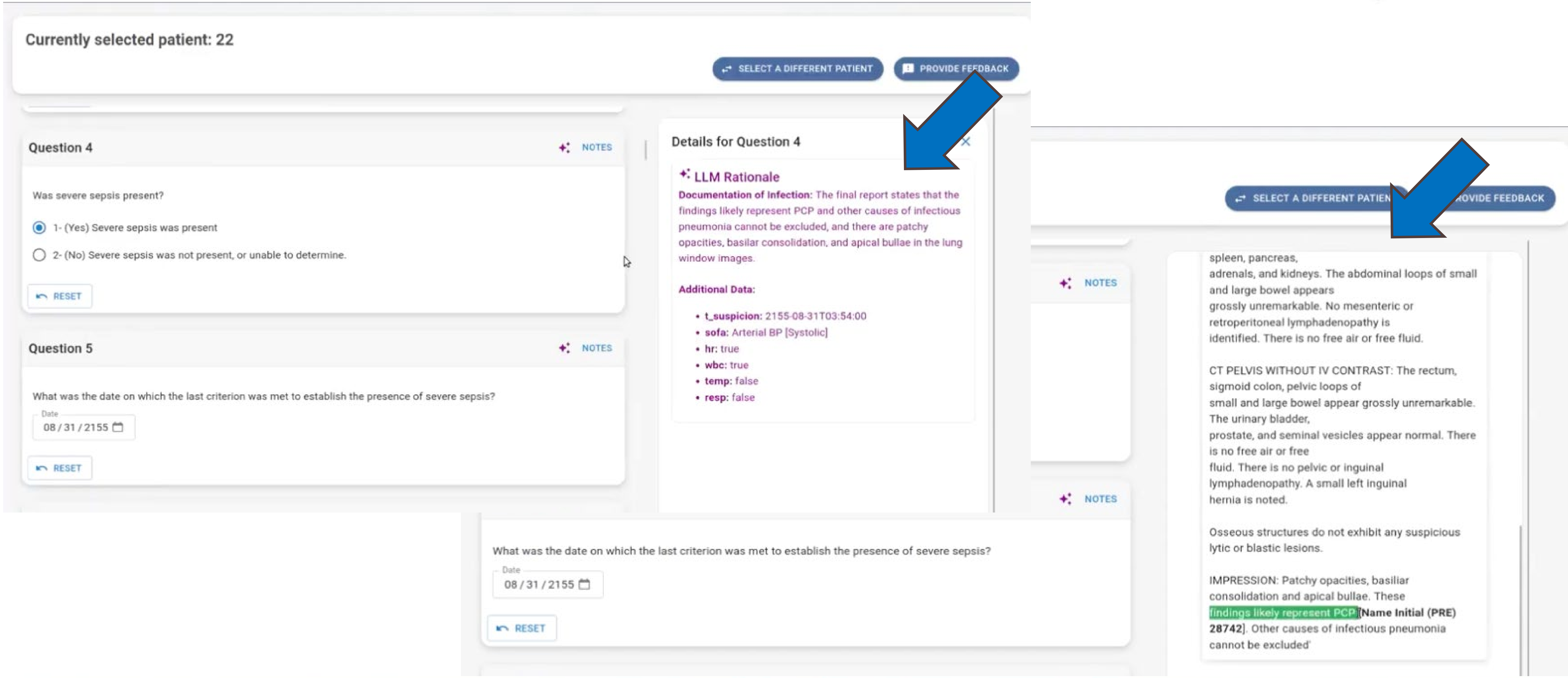
value. Here is the logic from the spec I am following:

Do not use an elevated INR, aPTT, or PTT v dysfunction if the medical record documentat patient received an anticoagulant medication in Appendix C Table 5.3 before the elevated INR, aPTT, or PTT value. Physician/APN/PA documentation is not required. Use the elevated INR, aPTT, or PTT value if the patient only received the following:
Heparin flushes

How is this normally handled?

SEP1

Abstraction Process



Currently selected patient: 22

Question 4

Was severe sepsis present?

● 1- (Yes) Severe sepsis was present

○ 2- (No) Severe sepsis was not present, or unable to determine.

Question 5

What was the date on which the last criterion was met to establish the presence of severe sepsis?

Date: 08 / 31 / 2155

Details for Question 4

LLM Rationale

Documentation of Infection: The final report states that the findings likely represent PCP and other causes of infectious pneumonia cannot be excluded, and there are patchy opacities, basilar consolidation, and apical bullae in the lung window images.

Additional Data:

- **t_suspicion:** 2155-08-31T03:54:00
- **sofa:** Arterial BP [Systolic]
- **hr:** true
- **wbc:** true
- **temp:** false
- **resp:** false

What was the date on which the last criterion was met to establish the presence of severe sepsis?

Date: 08 / 31 / 2155

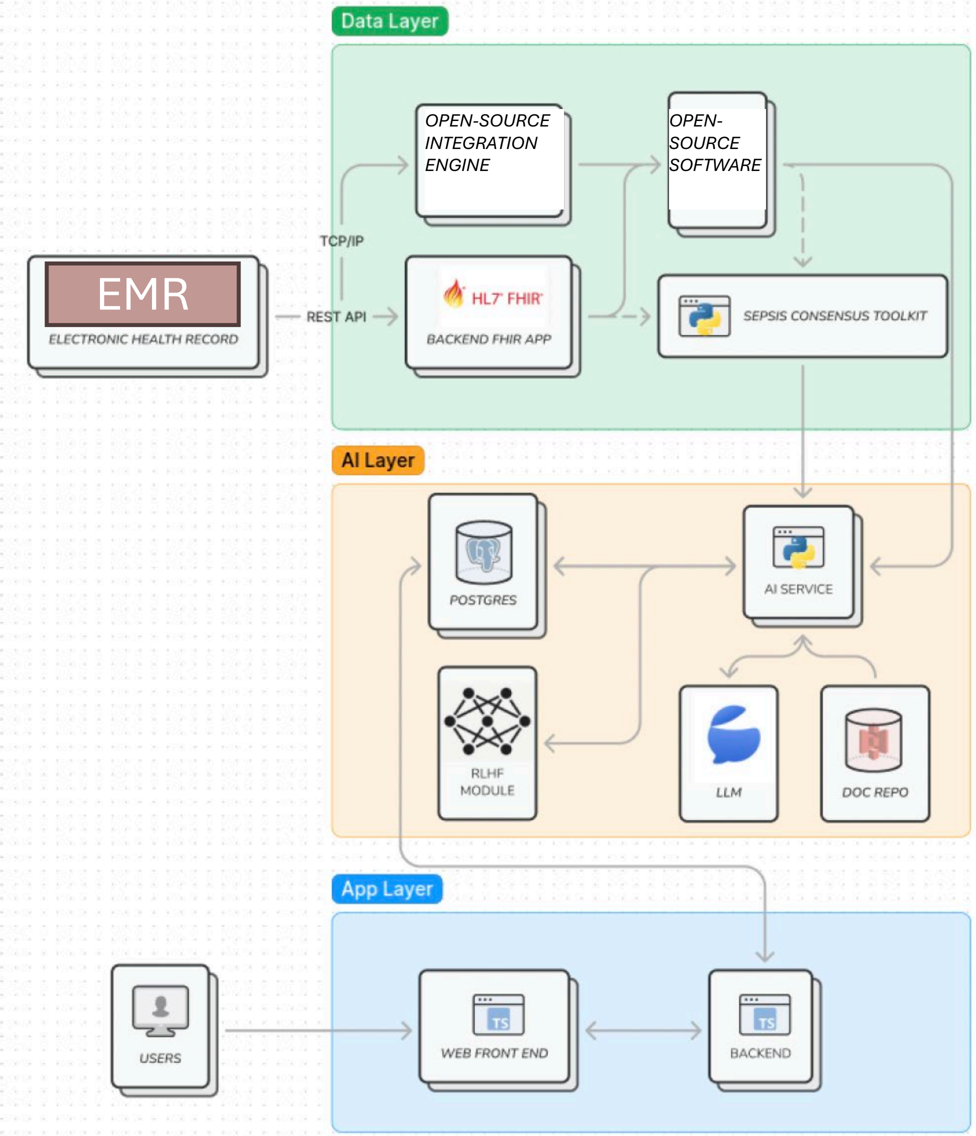
spleen, pancreas,
adrenals, and kidneys. The abdominal loops of small and large bowel appears
grossly unremarkable. No mesenteric or retroperitoneal lymphadenopathy is
identified. There is no free air or free fluid.

CT PELVIS WITHOUT IV CONTRAST: The rectum, sigmoid colon, pelvic loops of
small and large bowel appear grossly unremarkable. The urinary bladder,
prostate, and seminal vesicles appear normal. There is no free air or free
fluid. There is no pelvic or inguinal lymphadenopathy. A small left inguinal
hernia is noted.

Osseous structures do not exhibit any suspicious lytic or blastic lesions.

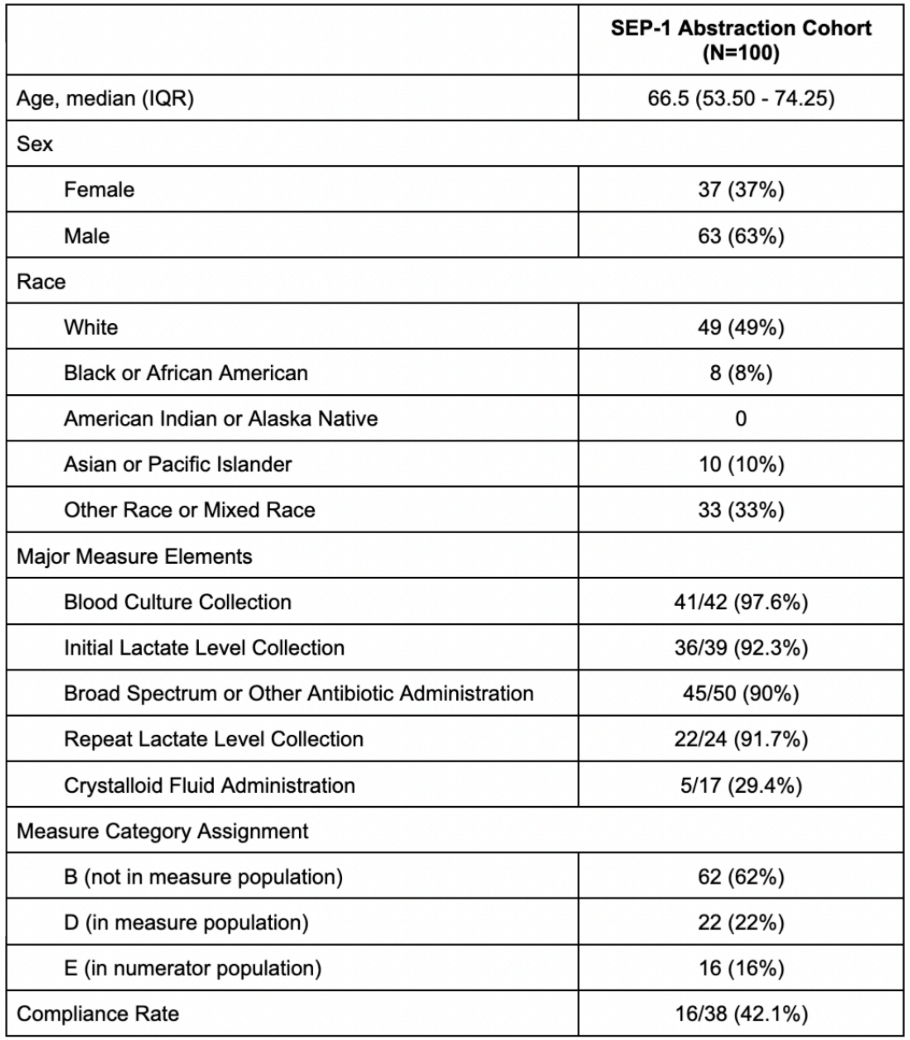
IMPRESSION: Patchy opacities, basilar consolidation and apical bullae. These findings likely represent PCP [**Name Initial (PRE) 28742**]. Other causes of infectious pneumonia cannot be excluded'

Architecture



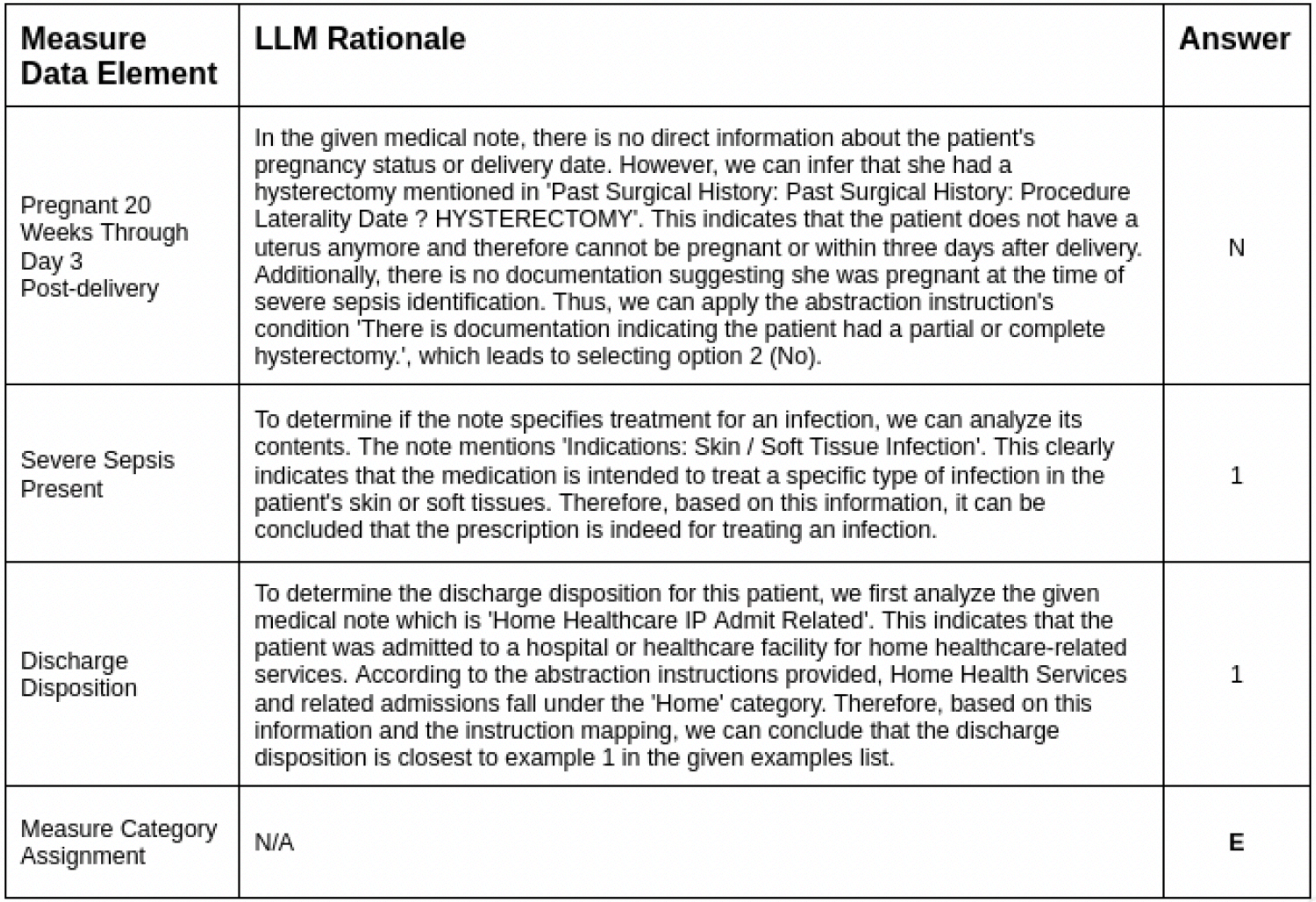
Methods

- **Cohort:** 5 months (100 patients) of SEP-1 abstractions reported to CMS in 2022.
- **Outcome:** Measure category agreement (pass, fail, or out-of-measure) between the LLM system and manual abstraction.
- **LLM Implementation:**
  - SOLAR 10.7B model with 8-bit quantization.
  - No fine-tuning.
  - Chain-of-thoughts and few-shot prompting with a temperature of 0.1.



	SEP-1 Abstraction Cohort (N=100)
Age, median (IQR)	66.5 (53.50 - 74.25)
Sex	
Female	37 (37%)
Male	63 (63%)
Race	
White	49 (49%)
Black or African American	8 (8%)
American Indian or Alaska Native	0
Asian or Pacific Islander	10 (10%)
Other Race or Mixed Race	33 (33%)
Major Measure Elements	
Blood Culture Collection	41/42 (97.6%)
Initial Lactate Level Collection	36/39 (92.3%)
Broad Spectrum or Other Antibiotic Administration	45/50 (90%)
Repeat Lactate Level Collection	22/24 (91.7%)
Crystalloid Fluid Administration	5/17 (29.4%)
Measure Category Assignment	
B (not in measure population)	62 (62%)
D (in measure population)	22 (22%)
E (in numerator population)	16 (16%)
Compliance Rate	16/38 (42.1%)

Sample Case



Measure Data Element	LLM Rationale	Answer
Pregnant 20 Weeks Through Day 3 Post-delivery	In the given medical note, there is no direct information about the patient's pregnancy status or delivery date. However, we can infer that she had a hysterectomy mentioned in 'Past Surgical History: Past Surgical History: Procedure Laterality Date ? HYSTERECTOMY'. This indicates that the patient does not have a uterus anymore and therefore cannot be pregnant or within three days after delivery. Additionally, there is no documentation suggesting she was pregnant at the time of severe sepsis identification. Thus, we can apply the abstraction instruction's condition 'There is documentation indicating the patient had a partial or complete hysterectomy.', which leads to selecting option 2 (No).	N
Severe Sepsis Present	To determine if the note specifies treatment for an infection, we can analyze its contents. The note mentions 'Indications: Skin / Soft Tissue Infection'. This clearly indicates that the medication is intended to treat a specific type of infection in the patient's skin or soft tissues. Therefore, based on this information, it can be concluded that the prescription is indeed for treating an infection.	1
Discharge Disposition	To determine the discharge disposition for this patient, we first analyze the given medical note which is 'Home Healthcare IP Admit Related'. This indicates that the patient was admitted to a hospital or healthcare facility for home healthcare-related services. According to the abstraction instructions provided, Home Health Services and related admissions fall under the 'Home' category. Therefore, based on this information and the instruction mapping, we can conclude that the discharge disposition is closest to example 1 in the given examples list.	1
Measure Category Assignment	N/A	**E**

Results

SEP-1 Question	**N (%) of Abstractions where System Answer Resulted in Agreement with Manual Category Assignment**	**Data Element Distribution (%) from Manual Abstraction**
Was severe sepsis present?	98/100 (98%)	Y (53.8%)
		N (46.2%)
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%	Y (5.4%)
		N (94.6%)
Is there physician/APN/PA documentation of comfort measures only, palliative care, or another inclusion term before or within six hours after the presentation of severe sepsis?	99/100 (99%)	Y (8.9%)
		N (91.1%)
Is there documentation supporting an acceptable delay in collecting a blood culture?	99/100 (99%)	Y (33.3%)
		N (66.6%)
Final Measure Category Assignment	**90/100 (90%)***	B (62%)
		D (22%)
		E (16%)

K = 0.82; 95% CI, 0.71-0.92)

Root Cause of Disagreement	**Number of Cases**	**Physician Adjudication**
The patient had Chronic Kidney Disease with a creatinine baseline of (2-3). 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

Using LLMs to Abstract Records for Sepsis

NEJM AI

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.,[1] Amy M. Sitapati, M.D.,[1] Chad VanDenBerg, M.P.H.,[2] Karandeep Singh, M.D., M.M.Sc.,[5] Christopher A. Longhurst, M.D., M.S.,[3] and Shamim Nemati, Ph.D.[1]

Lessons Learned

- Start with measures that have strong physician and clinical support
- Simply – it can be done
- Pace will be based on resources allocated
- Validation is ongoing process
- Retrieval augmented generation has limits and more inference calls on smaller sections of the clinical chart can improve performance
- Integration (w/ eye on automation) in operations is essential

Key Takeaways

- AI is fundamentally changing approach to Quality and Patient Safety
- Eliminating chart abstraction has cascading implication and application
- Test time scaling is important for mitigating hallucinations and achieving consistent results with smaller models
- Less administrative burden returns highly-skilled team members to improvement

Questions?

Contact:

Kimberly Quintero: kquintero@health.ucsd.edu

Aaron Boussina: aboussina@health.ucsd.edu

Chad VanDenBerg: cevandenberg@health.ucsd.edu