Application of a framework to assess the usefulness of alternative sepsis criteria

Christopher W. Seymour[1], **Craig M. Coopersmith**[2], **Clifford S Deutschman**[3], **Foster Gesten**[4], **Michael Klompas**[5], **Mitchell Levy**[6], **Gregory S. Martin**[7], **Tiffany M. Osborn**[8], **Chanu Rhee**[5], **David Warren**[9], **R. Scott Watson**[10], and **Derek C. Angus**[1]

[1]The Clinical Research, Investigation, and Systems Modeling of Acute illness (CRISMA) Center, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA

[2]Department of Surgery, Emory University School of Medicine, Atlanta, GA

[3]Hofstra-North Shore-LIJ School of Medicine, Cohen Children's Medical Center, New Hyde Park, NY

[4]Office of Quality and Patient Safety, New York State Health Department, Albany, NY

[5]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA; Department of Medicine, Brigham and Women's Hospital, Boston, MA

[6]Division of Pulmonary and Critical Care Medicine, Brown University School of Medicine, Providence, RI

[7]Department of Critical Care, Division of Pulmonary, Allergy, and Critical Care Medicine, Emory University School of Medicine, Atlanta, GA

[8]Departments of Surgery and Emergency Medicine, Washington University School of Medicine, St. Louis, MO

[9]Department of Medicine, Infectious Diseases, Washington University School of Medicine, St. Louis, MO

[10]Department of Pediatrics, Pediatric Critical Care Medicine, University of Washington and Center for Child Health Behavior and Development, Seattle Children's Research Institute, Seattle, WA

Abstract

The current definition for sepsis is life-threatening, acute organ dysfunction secondary to a dysregulated host response to infection. Criteria to operationalize this definition can be judged by 6 domains of usefulness (reliability; content, construct and criterion validity, measurement burden, and timeliness). The relative importance of these 6 domains depends on the intended purpose for the criteria (clinical care, basic and clinical research, surveillance, or quality improvement (QI) and audit). For example, criteria for clinical care should have high content and construct validity, timeliness, and low measurement burden to facilitate prompt care. Criteria for surveillance or QI/ audit place greater emphasis on reliability across individuals and sites and lower emphasis on

Corresponding Author: Derek C Angus, MD, MPH, Department of Critical Care Medicine, University of Pittsburgh, 3550 Terrace Street, 614 Scaife Hall, Pittsburgh, PA 15261, USA, angusdc@upmc.edu, Tel: +1 412 647 6965, Fax: +1 412 647 5258.

timeliness. Criteria for clinical trials require timeliness to ensure prompt enrollment and reasonable reliability but can tolerate high measurement burden. Basic research also tolerates high measurement burden and may not need stability over time. In an illustrative case study, we compared examples of criteria designed for clinical care, surveillance and QI/audit among 396,241 patients admitted to 12 academic and community hospitals in an integrated health system. Case rates differed 4-fold and mortality 3-fold. Predictably, clinical care criteria, which emphasized timeliness and low burden and therefore used vital signs and routine laboratory tests, had the highest case identification with lowest mortality. QI /audit criteria, which emphasized reliability and criterion validity, used discharge information and had the lowest case identification with highest mortality. Using this framework to identify the purpose and apply domains of usefulness can help with the evaluation of existing sepsis diagnostic criteria and provide a roadmap for future work.

Keywords

In the accompanying paper,[1] we highlighted several challenges for the definition of sepsis. Although there is currently considerable agreement regarding the overall conceptual definition that sepsis is a life-threatening acute organ dysfunction secondary to a dysregulated host response to infection,[2] there is no 'gold standard' approach to identify cases. Furthermore, the knowledge base that informs our understanding of sepsis is constantly changing, there are different and competing views on what aspects about any potential definition are most important, and many available criteria used to identify patients with sepsis are expressed on a continuum with no zones of rarity. Thus, there is no perfect method to unambiguously categorize patients as having sepsis or not.

We outlined a series of methodological steps to guide development and evaluation of any candidate criteria[1]: first, define the epistemological underpinning and all relevant terms used to frame the exercise; second, state the intended purpose for any proposed set of criteria, and; third, adopt a scientific approach to inform on the usefulness of any proposed criteria with regard to the intended purpose. We identified four broad purposes for sepsis criteria (clinical care, research, surveillance, and quality improvement and audit) and six domains of usefulness by which any proposed criteria might be judged (reliability, content, construct and criterion validity, measurement burden, and timeliness). Of note, the relative importance of the six domains varies by purpose. Here, we discuss these relationships in more detail (Table 1), providing examples under each purpose, and conducting an illustrative case study that compares and contrasts case identification with different criteria in a common dataset. We conclude with a roadmap for future work.

Clinical care

Sepsis is a clinical emergency, and the standard of care for bedside clinicians is prompt diagnosis and early intervention.[3,4] To promote rapid recognition of patients most likely to be septic, a definition and its criteria for clinical care should prioritize both content and construct validity (Table 1). The elements of these criteria should be representative of the

conceptual model of a threat-to-life, dependent on acute organ dysfunction, a dysregulated host response, and infection.[1,5] Placing a high priority on criterion validity means that clinical sepsis criteria should distinguish patients with sepsis from patients without sepsis A similarly high value would be placed on timeliness and low burden in terms of cost, safety, and complexity (Table 1). From a clinicians' perspective, criteria for sepsis should be simple, easy to apply at the bedside, while being as reliable as possible between patients, clinicians, and centers. It is possible that as criterion validity is maximized with more complexity, inter-rater reliability may suffer. Temporal stability (a component of meta-reliability) of criteria may be of moderate value, compared to the need for stability of the criteria across different locations of care (prehospital, emergency department, ward, or intensive care).

Example of clinical criteria

One example of sepsis criteria for clinical care is the European Society of Intensive Care Medicine/Society of Critical Care Medicine (ESICM/SCCM) Third International Consensus Definitions for Sepsis and Septic Shock. These criteria, shown in Table 2, include the presence of infection, accompanied by an acute change in the sepsis-related organ failure assessment (SOFA) score by >2 points from baseline (if available).[6] Derived from expert panel consensus deliberations and empiric data analyses in US and non-US data (including electronic health records (EHRs), administrative data, and prospective cohorts), these criteria sought to balance content validity, construct, and criterion validity with other domains. For example, the SOFA score was chosen to represent acute organ dysfunction due to its superior content and criterion validity,[7] while the alternative logistic organ dysfunction score (LODS)[8] was considered but rejected due to its poor timeliness and higher measurement burden. In addition, a quick prompt to identify most patients likely to be septic (qSOFA) was developed with moderate content validity for acute organ dysfunction,[7] but emphasized timely and low burden recognition of sepsis. At the same time, important elements of the sepsis definition were not given criteria at all, such as infection, a causal link between infection and organ dysfunction, or a dysregulated host response. The Task Force did not deem such elements unimportant. Rather, it deferred to existing guidelines for infection, decided there was no good way currently to operationalize the causal link, other than by relying on clinician judgment, and decided the available criteria for the host response (the systemic inflammatory response syndrome (SIRS) criteria)[9] lacked sufficient criterion or construct validity to be included. Acknowledging these limitations and others, the Task Force advocated a philosophy similar to the American Psychiatric Association *Diagnostic and Statistical Manual of Mental Disorders*, where this iteration of clinical criteria for sepsis would be one step of many.

Research

Research in sepsis is broad, and we highlight the different priorities for sepsis criteria at two ends of the research spectrum: randomized clinical trials (RCTs) and animal models. In RCTs, enrolled patients may be quite heterogeneous,[10] likely contributing to the preponderance of neutral results and frustratingly few advances in the treatment of sepsis.[11] When used for enrollment in RCTs, sepsis criteria are most useful if they have adequate

content validity, are consistently and reliably measured across participating sites. However, temporal stability is only required over the duration of the study and, given the relatively small number of cases, considerable measurement burden, other than patient safety, can often be tolerated. If the trial tests a time-sensitive intervention, the ease of measurement and accessibility of data elements included in the criteria are of paramount importance. In contrast, if the trial enrolls a small population of patients to test the efficacy of a novel agent, cost (a component of measurement burden) and timeliness may not be a priority. Compared to clinical criteria and depending on the intervention under study, criteria in RCTs may tolerate a measurement burden.

To inform basic research, a sepsis definition will be challenged by the differences in the clinical manifestations of sepsis between humans and animals.[12] One such example is the typical hypothermic response to a cecal-ligation and puncture model of sepsis in mice,[13] which contrasts with fever commonly seen in many humans with infection.[14] Sepsis criteria for this purpose may place more value on domains like congruence with specific biologic pathways, offending organisms, or directionality of the immune response, captured within content and construct validity.[15] At the same time, issues such as the measurement burden, cost and complexity, or timeliness may be easier to manage in a laboratory setting. For example, the study of a dysregulated host response in murine sepsis models may rely on flow cytometry data (high cost, not timely) versus a physical examination (low cost, timely). More work is needed to develop congruent sepsis definitions and criteria that translate from human phenotypes to those found in contemporary animal models.

Example in clinical trials

In the ProCESS, ARISE, and ProMISe trials in septic shock, investigators chose enrollment criteria that placed high value on reliability between sites and ease of measurement in the emergency department.[16] They used objective measures like serum lactate level or presence of hypotension to find similar patients, and went to great lengths to harmonize enrollment across three continents over five years,[16] despite different practice patterns, standards of care, and emergency care systems across the participating countries.[17] The success of this strategy can be measured in the similar outcome rate of the usual care arms, direction of the treatment effects, and successful compilation of results in a meta-analysis.[18]

Audit and quality improvement

Audit and quality improvement (QI) are widely used to understand barriers to best practice, analyze gaps in care, and conduct systematic measurement during implementation of new tools. Sepsis is no stranger to quality improvement, and many QI reports document how international clinical practice guidelines impact the outcomes of septic patients.[19,20] A sepsis definition focused on quality improvement would likely place high value on validity, approximating criteria used by clinicians at the bedside, but potentially have even greater emphasis on reliability, timeliness, and measurement burden (Table 1). For example, surveillance criteria may rely on objective documentation of mechanical ventilation as a respiratory organ dysfunction versus a clinical diagnosis of acute respiratory distress syndrome (ARDS),[21] as the latter may suffer from poor inter-rater, meta-, and test-re-test reliability.[22–25] This approach would promote benchmarking of doctors and hospital care of

patients thought to be septic during routine clinical care. Audit may also occur across large populations of hospitals and patients, such that ease of measurement, cost, and burden gain importance. But, audit typically occurs after the care episode, and QI criteria have the benefit of hindsight and may include events at any time during the patient encounter.

Example of audit and QI criteria

New York's recently implemented sepsis reporting legislation and new Centers for Medicare & Medicaid Services (CMS) reporting requirements (National Quality Forum 0500 Severe Sepsis and Septic Shock Management bundle) provide examples of sepsis criteria used for quality improvement.[26] Focusing on the CMS criteria as an example, they use for the denominator specific ICD10 discharge diagnoses in administrative claims, and among those with ICD10 diagnoses present, cases are identified with >2 SIRS criteria[27] and >1 organ dysfunctions, as proposed by the Surviving Sepsis Campaign.[3] This approach uses low cost, routinely collected data for billing, and favors efficiency but may place lower value on reliability as clinicians differ widely in how they document and code for sepsis. The QI criteria they use to determine a numerator of individuals for whom quality of care was good emphasize content validity by employing well-studied, international guideline-recommended criteria for organ dysfunction.[3] Yet, the approach increases measurement burden by requiring manual abstraction to identify a 'time zero' and inspection of the care episode for laboratory values suggestive of acute organ dysfunction. This example of QI criteria will be evaluated going forward, and it is possible that criteria for infection, host response, and organ dysfunction will be modified. Feedback regarding the ease of measurement, reliability across hospitals, and updates to clinical criteria may well inform potential modifications.

Surveillance and epidemiology

A purpose of public health surveillance is to describe the scope and magnitude of disease to inform public health policy and research, prioritize resources, and identify opportunities for prevention and improving care. For sepsis, an ideal surveillance definition would place a high value on reliability across healthcare facilities and moderate-high value on validity. However, it would place lower value on timeliness as surveillance definitions are not intended for use in the clinical management of individual patients and detection of events for surveillance purposes often occurs after the episode of care (Table 1). Surveillance criteria are therefore different to clinical care criteria but similar to audit and QI criteria in that hospitals must be able to measure the criteria at low burden and cost across a large population of patients.

Example of surveillance criteria

Prior work to define the national burden of sepsis has used an administrative claims-based approach.[28,29] These algorithms use either implicit or explicit ICD-9-CM diagnosis and procedure codes. Claims data may have only moderate construct validity and content validity.[30] A growing body of literature base suggests that trends over time sepsis and acute organ dysfunction claims may not track with changes over time, with clinical evidence of sepsis and acute organ dysfunction in the EHR (a construct validity and meta-reliability

concern).[31] In light of these limitations and increasing appreciation of sepsis as a public health problem, investigators from the Centers for Disease Control and Prevention (CDC) Prevention Epicenters developed and validated EHR-based surveillance definitions for sepsis (Table 2).[32] In this work, high value was placed on reliability across hospitals, low measurement burden in the EHR, and stability over time. Like the proposed clinical criteria, surveillance criteria used by these investigators focused on a conceptual framework of infection and acute organ dysfunction but included events apparent only in retrospect (such as the duration of antibiotic treatment), since surveillance sepsis definitions are not used for clinical care and do not need to be applied in real time.

As shown in Table 2, the preliminary surveillance criteria (termed 'EpiCenters Complete Surveillance Definition') included suspected infection, defined as a blood culture order and antibiotics administered for 4 consecutive calendar days. Among infected patients, sepsis was present if there was concurrent organ dysfunction, defined by one of six events (Table 2). A shorter version of these criteria was proposed (EpiCenters Simple Surveillance Definition) to minimize measurement burden (lower data extraction costs). These surveillance criteria were piloted in preliminary studies, but their performance in a broader cohort of hospitals, particularly community hospitals, is under investigation. In addition, their exclusion of valuable clinical data like vital signs has unknown impact on construct and criterion validity.

Case study: a crosstalk between approaches in a regional health system

To illustrate how criteria for different purposes identify potentially different patients, we conducted a case study in the EHRs of 396,241 patients admitted to 12 academic and community hospitals in an integrated health system in southwestern Pennsylvania in 2012. Please see the Supplementary Digital Content for more details. We identified cases using: i) clinical criteria in the Third International Consensus Definitions for Sepsis and Septic Shock (SOFA and qSOFA scores); ii) the EpiCenters Complete and Simple Surveillance Definition, and; iii) the QI criteria found in the CMS Severe Sepsis and Septic Shock: Management Bundle measure (NQF #0500).

The clinical criteria found 27 cases per 1000 encounters, compared to 30 per 1000 encounters for the EpiCenters complete criteria, and 6 per 1000 encounters by the CMS criteria for audit and QI (Table 3). The last likely reflects the need for specific diagnosis codes, SIRS criteria, and organ dysfunction, whereas clinical and surveillance criteria did not begin the algorithm with administrative claims or SIRS criteria. The narrow subset identified by QI criteria is also reflected in the greater in-hospital mortality, intensive care unit admission rate, and organ failures compared to clinical or surveillance criteria (Table 3). We note that the EpiCenters complete criteria and clinical criteria that use a change in SOFA score were similar in terms of need for organ support, case fatality rate and utilization of intensive care. We further illustrate the relationship of these criteria in a modified multi-method matrix (Figure 1). The correlation coefficients between criteria ranged from 0.77 to 1.0, and the associated color maps reveal the distribution of agreement. The figure shows that correlation coefficients were lower, and more patients in off-diagonal cells of the heatmaps, comparing the more restrictive QI criteria with others. In comparison, the

Epicenters complete and simple criteria find very similar populations, as evidenced by the high correlation coefficient and preponderance of patients in similar color distributions on the heatmaps (concordant cells where both simple and complete agreed that sepsis was absent or present).

Interpreting the case study findings

The case study above reveals that, as expected, different criteria for sepsis find different patients in a multicenter health system. This finding should not be troubling since differences reflect the distinct priorities for each purpose. Criteria for QI and audit appear to identify a small subset of patients at particularly higher risk of a bad outcome, and could be used to represent sentinel cases on which to measure performance. Such an approach may not be that far removed from strategies in other conditions, such as cardiac arrest, ventilator associated events, or surgical site infections. In contrast, the clinical criteria find a larger cohort of septic patients as they use a broader set of variables, and are intended to be used for prompt recognition and care. The severity of illness is less compared to the QI cohort, suggesting that cases, on average, may be at an earlier stage of acute organ dysfunction. The proposed EHR-based surveillance criteria for sepsis captured a population with a moderate illness severity, which reflects an emphasis on organ support requiring ICU admission. These markers of organ dysfunction may have reliability across centers and ease of measurement in the EHR. They make use of data available at the end of the patient encounter, and are not intended to support clinical decision-making about prompt care. Although possible, it is less likely that the modest differences in patients identified by surveillance and clinical criteria resulted from the approach to finding patients with infection. Both algorithms used a combination of body fluid cultures and antibiotic administration in the EHR.[7,32]

Roadmap for future directions

In these two papers, we established that it is an elusive and unrealistic goal to have a single perfect 'gold standard' definition of sepsis, in part because of evolving knowledge, differing priorities and values, and a lack of discrete, unambiguous, widely deployable diagnostic criteria. However, a methodologic framework can be used to develop and assess different definitions and criteria with the goal of finding good 'working' criteria, even if not 'gold standards'. Furthermore, these different criteria may be valuable in different ways: one set of criteria might be particularly suitable for a particular purpose. Thus, we propose the methodologic framework first requires setting ground rules regarding underlying philosophy, definition of terms that will frame the exercise, and explicit prioritization of values. Values will depend on the intended purpose. We acknowledge that our framing of purposes for sepsis criteria as falling under four broad areas is somewhat simplistic. In reality, these activities do not occur in silos, but are much more interrelated. For example, we would not expect patients clinically recognized to be septic to be excluded from either a prospective QI initiative or retrospective audit of care. And similarly, we would aim for reliable surveillance strategies to track the public health burden of patients *clinically* thought to be septic. Finally, any proposed criteria can be evaluated across six domains of usefulness. The relative importance of these domains will depend on the purpose. Thus, in

our example, we saw that different criteria each behaved differently, but in so doing were more or less suited to different purposes. They also have predictable relationships to each other. For example, one set of criteria may consistently identify fewer but sicker cases.

So what comes next? The first, and most important step, is building awareness that no single definition for sepsis will satisfy for the four purposes described in this paper. A greater understanding of the different purposes for sepsis criteria and their priorities may relax the expectation for a single answer to the question: 'Is this patient septic?' Second, there is a need for consistent terminology. The clinical criteria proposed by the ESICM/SCCM Third International Consensus Definitions for Sepsis and Septic Shock abandoned the term 'severe sepsis', though it has been a part of the Epicenters surveillance criteria, QI proposed criteria, and billing codes. Similarly, terms such as 'suspected' or 'presumed' are variably used across applications to characterize the presence of infection. Standardization of the terminology used in the various approaches to defining sepsis would reduce confusion. Third, many elements of the conceptual framework for sepsis are not defined at all. Features of sepsis such as the causal link between infection and organ dysfunction and a dysregulated vs. normal host response to infection should continue to be the subject of intense investigation. Fourth, future criteria may attempt to reduce zones of rarity by incorporating molecular markers or novel tests. Although more than 2000 biomarkers of sepsis are currently proposed,[33] no marker or set of markers has a balance of burden, reliability, and validity for sepsis similar those used to identify acute myocardial infarction, for example. Finally, a proposed sepsis criterion for any purpose requires prospective study. With the goal of iterative improvement, these studies should compare within and across physicians, within and across hospitals, and measurement of the stability of criteria over time.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Angus DC, Seymour CW, Klompas M, et al. A framework for the development and interpretation of different sepsis definitions. (submitted).

2. Singer M. 2016 Sepsis Task Force definitions. JAMA. :1–29.

3. Dellinger RP, Levy MM, Rhodes A, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. Crit Care Med. Feb; 2013 41(2):580–637. [PubMed: 23353941]

4. Seymour CW, Rosengart MR. Septic Shock: Advances in Diagnosis and Treatment. JAMA. Aug 18; 2015 314(7):708–717. [PubMed: 26284722]

5. Czura CJ. 'Merinoff symposium 2010: sepsis' – speaking with one voice. Mol Med. Jan-Feb;2011 17(1–2):2–3. [PubMed: 21246163]

6. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. Jul; 1996 22(7):707–710. [PubMed: 8844239]

7. Seymour CW, Liu V, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the Third International Consensus Definitions for Sepsis (Sepsis-3). 2016 (forthcoming).

8. Le Gall JR, Klar J, Lemeshow S, et al. The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. JAMA. Sep 11; 1996 276(10):802–810. [PubMed: 8769590]

9. Kaukonen KM, Bailey M, Pilcher D, Cooper DJ, Bellomo R. Systemic Inflammatory Response Syndrome Criteria in Defining Severe Sepsis. N Engl J Med. Mar 17.2015 372:1629–1638. [PubMed: 25776936]

10. Iwashyna TJ, Burke JF, Sussman JB, Prescott HC, Hayward RA, Angus DC. Implications of Heterogeneity of Treatment Effect for Reporting and Analysis of Randomized Trials in Critical Care. Am J Respir Crit Care Med. Nov 1; 2015 192(9):1045–1051. [PubMed: 26177009]

11. Fink MP, Warren HS. Strategies to improve drug development for sepsis. Nature Rev Drug Discovery. Oct; 2014 13(10):741–758. [PubMed: 25190187]

12. Seok J, Warren HS, Cuenca AG, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci U S A. Feb 26; 2013 110(9):3507–3512. [PubMed: 23401516]

13. Lewis A, Rosengart MR, Seymour CW. Use of biotelemetry to define physiology-based deterioration thresholds in a murine cecal ligation and puncture model of sepsis. Crit Care Med. 2016 (forthcoming).

14. Young PJ, Bellomo R. Fever in sepsis: is it cool to be hot? Crit Care. Feb 13.2014 18(1):109. [PubMed: 24521542]

15. Angus DC, van der Poll T. Severe sepsis and septic shock. N Engl J Med. Nov 21.2013 369(21): 2063. [PubMed: 24256390]

16. Huang DT, Angus DC, Barnato A, et al. Harmonizing international trials of early goal-directed resuscitation for severe sepsis and septic shock: methodology of ProCESS, ARISE, and ProMISe. Intensive Care Med. Oct; 2013 39(10):1760–1775. [PubMed: 23958738]

17. Austin S, Murthy S, Wunsch H, et al. Access to urban acute care services in high- vs. middle-income countries: an analysis of seven cities. Intensive Care Med. Mar; 2014 40(3):342–352. [PubMed: 24337401]

18. Angus DC, Barnato AE, Bell D, et al. A systematic review and meta-analysis of early goal-directed therapy for septic shock: the ARISE, ProCESS and ProMISe Investigators. Intensive Care Med. 2015 Sep; 41(9):1549–60. [PubMed: 25952825]

19. Ferrer R, Artigas A, Levy MM, et al. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. JAMA. May 21; 2008 299(19):2294–2303. [PubMed: 18492971]

20. Ferrer R, Martin-Loeches I, Phillips G, et al. Empiric antibiotic treatment reduces mortality in severe sepsis and septic shock from the first hour: results from a guideline-based performance improvement program. Crit Care Med. Aug; 2014 42(8):1749–1755. [PubMed: 24717459]

21. Ranieri VM, Rubenfeld GD, Thompson BT, et al. Acute respiratory distress syndrome: the Berlin Definition. JAMA. Jun 20; 2012 307(23):2526–2533. [PubMed: 22797452]

22. Fröhlich S, Murphy N, Doolan A, et al. Acute respiratory distress syndrome: under-recognition by clinicians. J Crit Care. 2013 Oct; 28(5):663–8. [PubMed: 23806247]

23. Villar J, Pérez-Méndez L, Kacmarek RM. Current definitions of acute lung injury and the acute respiratory distress syndrome do not reflect their true severity and outcome. Intensive Care Med. 1999 Sep; 25(9):930–5. [PubMed: 10501747]

24. Ferguson ND, Kacmarek RM, Chiche JD, et al. Screening of ARDS patients using standardized ventilator settings: influence on enrollment in a clinical trial. Intensive Care Med. 2004 Jun; 30(6): 1111–6. [PubMed: 14991096]

25. Rubenfeld GD, Caldwell E, Granton J, et al. Interobserver variability in applying a radiographic definition for ARDS. Chest. 1999 Nov; 116(5):1347–53. [PubMed: 10559098]

26. https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage %2FQnetTier3&cid=1228774725171. Accessed December 18, 2015, 2015.

27. Rangel-Frausto MS, Pittet D, Costigan M, Hwang T, Davis CS, Wenzel RP. The natural history of the systemic inflammatory response syndrome (SIRS). A prospective study. JAMA. Jan 11; 1995 273(2):117–123. [PubMed: 7799491]

28. Angus DC, Linde-Zwirble WT, Lidicker J, Clermont G, Carcillo J, Pinsky MR. Epidemiology of severe sepsis in the United States: analysis of incidence, outcome, and associated costs of care. Crit Care Med. Jul; 2001 29(7):1303–1310. [PubMed: 11445675]

29. Gaieski DF, Edwards JM, Kallan MJ, Carr BG. Benchmarking the incidence and mortality of severe sepsis in the United States. Crit Care Med. May; 2013 41(5):1167–1174. [PubMed: 23442987]

30. Seymour CW, Iwashyna TJ, Cooke CR. Managing uncertainty in claims-based sepsis research. Crit Care Med. Apr; 2013 41(4):1134–1136. [PubMed: 23528757]

31. Rhee C, Murphy MV, Li L, et al. Comparison of trends in sepsis incidence and coding using administrative claims versus objective clinical data. Clin Infect Dis. Jan 1; 2015 60(1):88–95. [PubMed: 25258352]

32. Rhee C, Kadri S, Huang SS, et al. Objective Sepsis Surveillance Using Electronic Clinical Data. Infect Control Hosp Epidemiol. Nov 3.2015 :1–9.

33. Pierrakos C, Vincent JL. Sepsis biomarkers: a review. Crit Care. 2010; 14(1):R15. [PubMed: 20144219]

34. Opal SM, Laterre PF, Francois B, et al. Effect of eritoran, an antagonist of MD2-TLR4, on mortality in patients with severe sepsis: the ACCESS randomized trial. JAMA. Mar 20; 2013 309(11):1154–1162. [PubMed: 23512062]

35. Shrum B, Anantha RV, Xu SX, et al. A robust scoring system to evaluate sepsis severity in an animal model. BMC Res Notes. 2014; 7:233. [PubMed: 24725742]
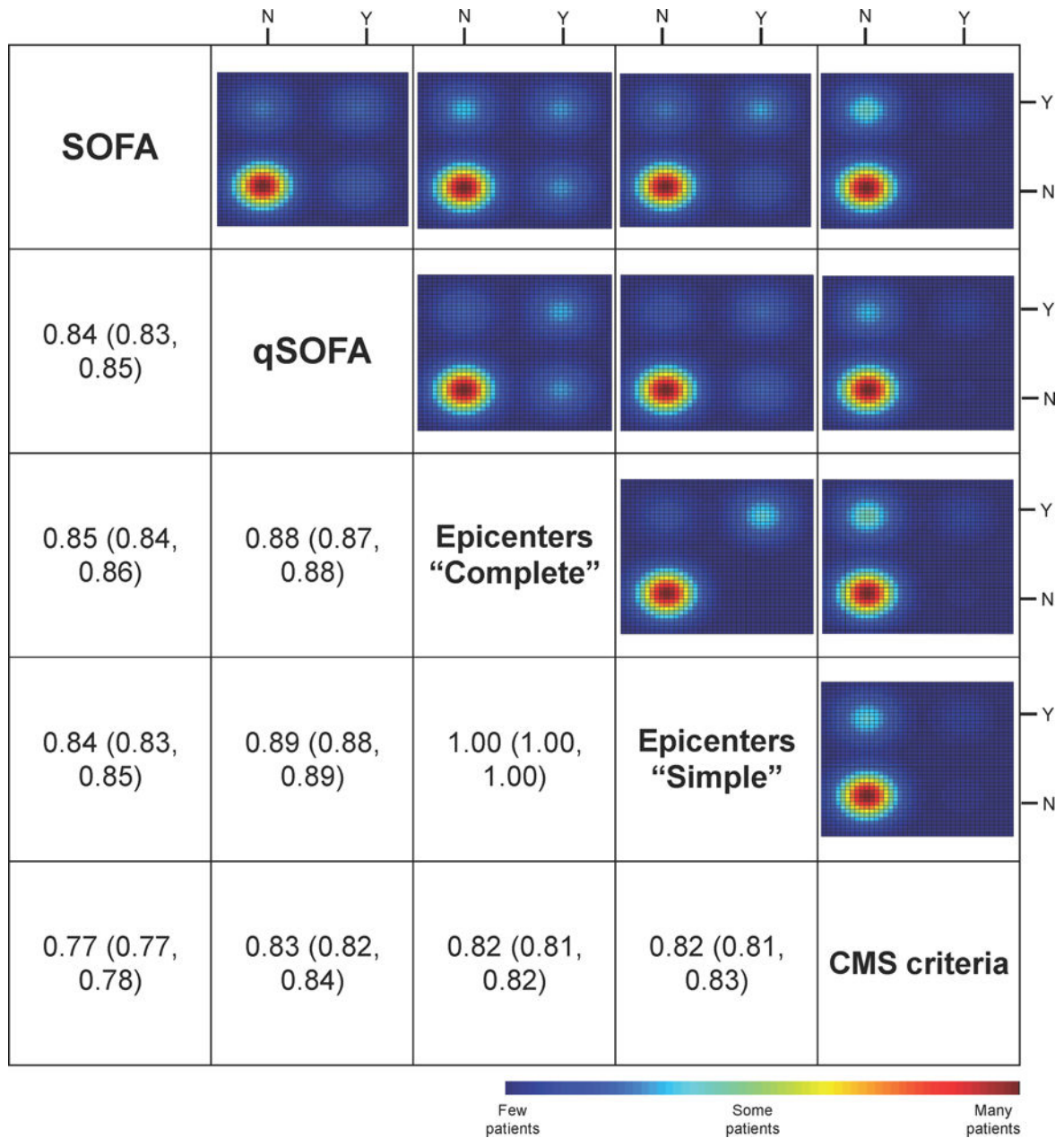
Figure 1.
Modified multi-method matrix for various sepsis criteria.

Below-the-diagonal cells contain the correlation coefficient between dichotomized criteria (with bootstrapped 95%CI). The above diagonal cells illustrate the 2×2 distribution of patients across criteria (either present or absent). Color scale corresponds to the number of patients in each group in the respective 2×2 table (red-many patients in that cell, blue – fewer patients in that cell). SOFA – sepsis-related organ failure assessment; qSOFA – quick SOFA; CMS – Centers for Medicare and Medicaid Services.

Table 1

Domains of usefulness (and subdomains) for potential sepsis diagnostic criteria and their priority by purpose.

Domain	Abbreviated domain definition	Priority by Intended purpose				
		Clinical care	Clinical trials	Basic research	Surveillance	QI and audit
1. Reliability	Criteria yield stable reproducible results					
Test-retest	When tests are repeated	Moderate-high	High	Moderate-high	High	High
Inter-rater	When tests are interpreted	Moderate-high	High	Moderate-high	High	High
Meta-reliability	Resistance of tests or measures to changes unrelated to biology	Moderate-high	High	Very low	High	High
2. Content validity	Criteria fit with current understanding and knowledge (they 'make sense')	High	High	Moderate-high	High	High
3. Construct validity	Criteria measure what they purport to measure					
Convergent	The extent to which 2 or more aspects that should agree do agree	High	High	High	High	High
Discriminant	The extent to which 2 or more aspects that should not agree do not	High	High	High	High	High
Multitrait, multimethod matrix	Multiple assessments of agreement among different measures, comparing across traits (e.g., different infections and organ failures) and measurement methods	Moderate-high	Moderate-high	Moderate	Moderate-high	Moderate-high
4. Criterion validity	New criteria agree with existing 'standard'					
Concurrent	Comparison to a current standard available at the same time	High	High	Moderate	High	High
Predictive	Comparison to a later outcome strongly associated with the disease of interest	High	High	Moderate-high	High	High
5. Measurement burden	Burden to implement criteria					
Lower cost	Financial costs (to patient, provider, or healthcare system)	High	Very low[^^]	Very low[^^]	High	High
Greater safety	Side-effects, complications to patient	High	High[*]	High	High	High
Lower complexity	Difficulties executing the various steps to obtain or interpret the tests and measures	High	Moderate	Very low	Moderate-high	Moderate-high
6. Timeliness	Speed with which criteria are generated with respect to the course of the disease	High	Moderate-high[^]	Low[^]	Low[^']	Low[^']

[*] The safety profile of criteria used in the course of a trial for a patient who provides informed consent may be different than the safety profile for routine clinical care.

[^] The timeliness for randomized clinical trial enrollment criteria will depend on the intervention under study, and may be very low priority depending on the nature of the animal study.

[^^] Greater costs can be tolerated for sepsis criteria deployed for clinical or basic research, often as a result of the selected, smaller group of patients or animals enrolled in research compared to routine clinical practice or surveillance efforts.

*
Timeliness of sepsis criteria used for surveillance or quality improvement/audit may be lower priority as these events often occur after the care episode is complete. For example, surveillance criteria for infection that use the total duration of antibiotics (in days) may identify a group with presumed infection, but is a criteria that is not available "real-time" relative to the course of disease.

Table 2

Examples of alternative sepsis diagnostic criteria by purpose.

Purpose		Objective	Example	Criteria	Caveats
Clinical care		To inform the direct clinical care of sepsis at the bedside	ESICM/SCCM Third International Consensus Definitions for Sepsis and Septic Shock Task Force[*7]	Among patients in whom the clinician suspects infection: • Acute change in SOFA score >2 points (if baseline available) For clinical prompt in infected patients: • >2 qSOFA points outside the ICU	• Sepsis not restricted to confirmed infection, but rather those with infection as determined clinically • Organ dysfunction proposed to be causally related to infection, but often difficult to ascertain • No criteria proposed for infection, left to clinician judgment • SOFA score baseline may not always be available, difficult to assess change • qSOFA score captures may of those most likely to be septic, but requires prospective validation
Research	Clinical	To guide the conduct of clinical research in sepsis, including experimental and observational research	Enrollment criteria for the ACCESS trial[34]	Among patients with evidence of infection, all of: • 3 or more SIRS criteria • 1 major organ dysfunction • High risk of death determined by APACHE II score (21 to 37 points)	• Broad variety in clinical research criteria, and observational studies may not agree with randomized clinical trials • Enrollment in clinical trials will focus on objective criteria to improve reliability • Subsets of patients are often selected to reduce heterogeneity of treatment effects
	Basic	To guide the study of the fundamental principles of sepsis, often in animal models	Murine sepsis score after fecal-induced peritonis[35]	Score ranging from 0 to 28, with 4 points for: • Appearance • Level of consciousness • Activity • Response to stimulus • Eyes • Respiration rate • Respiration quality	• Reported with high inter-rater, test-retest reliability • May be specific to the type of animal model used in experiments • Alternative models using wireless biotelemetry and clinical data are under study

Purpose	Objective	Example	Criteria	Caveats
Surveillance and epidemiology	To track local and national burden, incidence and outcomes of sepsis across hospitals and over time	CDC Prevention Epicenters Preliminary criteria[32]	Among patients with infection, >1 of: • vasopressor use • >2 days of mechanical ventilation • rise in serum creatinine by >0.5	• Avoids data elements not readily available in the electronic health record such as vital signs in order to improve feasibility in wider array of facilities • For similar patients, clinicians may provide organ support differently
QI and audit	To inform quality improvement initiatives and audit performance across hospitals	CMS[22]	• ICD9-10-CM claims based algorithm to find denominator of sepsis patients • Manual chart review for SIRS criteria and organ dysfunction criteria to narrow cases	• Restricts to cohort of patients identified with administrative data • May result in a smaller subset of patients

*Task Force articulated the overarching definition for sepsis as Infection complicated by acute, life-threatening organ dysfunction due to a dysregulated host response

Table 3

Sepsis case identification by alternative criteria in a 12-hospital regional health system (N=396,241).

Variable	Clinical criteria in 2016 Consensus Definitions: Acute change in >=2 SOFA points	Clinical criteria in 2016 Consensus Definitions: Bedside prompt of >=2 qSOFA points	Prevention EpiCenters criteria for surveillance: Simple	Prevention EpiCenters criteria for surveillance: Complete	CMS criteria for QI and audit
Total no.	11,011	9,823	9,176	12,041	2,709
Positive blood cultures, no. (%)	854 (8)	786 (8)	824 (9)	1,034 (9)	n/a
Maximum day 1 SOFA score, mean (SD)	2.9 (3.0)	2.9 (3.5)	3.1 (3.6)	2.6 (3.4)	4.2 (4.2)
>=2 SIRS criteria, no. (%)	7,003 (64)	7,487 (76)	5,903 (64)	7,166 (60)	2,709 (100)
ICU admission, no. (%)	5,402 (49)	6,808 (69)	6,658 (73)	7,288 (61)	2,239 (83)
ICU length of stay, median [IQR] days	5 [3 – 10]	5 [3 – 10]	6 [4 – 12]	6 [4 – 11]	7 [3 – 13]
Hospital length of stay, median [IQR] days	8 [5–13]	9 [6 – 16]	12 [7 – 19]	11 [7 – 18]	12 [7 – 20]
In-hospital mortality, no. (%)	977 (8.9)	1,072 (11)	1,256 (14)	1,319 (11)	663 (24)

Abbreviations: CMS – Centers for Medicaid and Medicare Services; ICU – intensive care medicine; IQR – interquartile range; QI – quality improvement; SIRS – systemic inflammatory response syndrome; SOFA – sepsis related organ failure assessment score