# Supplemental Content

### Tables

**Supplemental Table 1** provides the percentage of patients who satisfied various sepsis (and non-sepsis) criteria, grouped by fiscal year of hospital admission (October - October). These percentages were used to create Supplemental Figure 2.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | Angus | Martin | Explicit | CMS | CDC | Sepsis-3 | Blood culture | Antibiotic | Suspicion of Infection |
| 2001 | 22.8 | 9.3 | 0 | 4.2 | 0 | 0 | 52.3 | 0 | 0 |
| 2002 | 22.3 | 11.2 | 0 | 8.3 | 0 | 0 | 51.1 | 0 | 0 |
| 2003 | 19.2 | 9.6 | 0.4 | 7.1 | 23.9 | 32.2 | 54.4 | 61.1 | 37.6 |
| 2004 | 22.6 | 14.6 | 6.9 | 11.9 | 27.3 | 36.5 | 56.3 | 67.1 | 42.9 |
| 2005 | 27.8 | 17.9 | 12.6 | 14.8 | 30.4 | 41.2 | 59.1 | 67.7 | 46.6 |
| 2006 | 29.5 | 17.6 | 12.7 | 14 | 31 | 43.2 | 67.5 | 69.6 | 52.6 |
| 2007 | 27.5 | 15.4 | 9.4 | 11.6 | 30.1 | 45.5 | 72 | 68.4 | 55.1 |
| 2008 | 26.6 | 14.9 | 10.1 | 12.1 | 27.1 | 42.1 | 71.2 | 63.9 | 51.6 |
| 2009 | 27.4 | 13.2 | 7.8 | 9.4 | 30.7 | 49.5 | 89.8 | 65.5 | 59.7 |
| 2010 | 28.3 | 14 | 8.9 | 10.4 | 33 | 49.5 | 94.2 | 64.9 | 60.9 |
| 2011 | 28.3 | 14.9 | 8.6 | 11.2 | 33.7 | 50.4 | 94.8 | 64.9 | 61 |
| 2012 | 29.8 | 15.7 | 10 | 12.2 | 31.1 | 47.3 | 95.6 | 62.1 | 58.7 |

**Supplemental Table 2** provides the agreement between each sepsis definition using Cronbach’s alpha. Confidence intervals are calculated using the 5th and 95th percentile across 1,000 bootstrap samples.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Martin | Explicit | CDC | CMS | Sepsis-3 |
| Angus | 0.69 [0.68-0.70] | 0.62 [0.61-0.63] | 0.62 [0.60-0.63] | 0.63 [0.62-0.65] | 0.62 [0.61-0.64] |
| Martin |  | 0.85 [0.84-0.85] | 0.54 [0.52-0.55] | 0.91 [0.90-0.92] | 0.49 [0.48-0.50] |
| Explicit |  |  | 0.49 [0.48-0.51] | 0.89 [0.89-0.90] | 0.40 [0.39-0.41] |
| CDC |  |  |  | 0.53 [0.52-0.55] | 0.76 [0.75-0.76] |
| CMS |  |  |  |  | 0.45 [0.43-0.46] |

###

###

### Figures

**Supplemental Figure 1.** Proportion of patients over time who: satisfied one of the sepsis criteria, had a blood culture, or a prescription of antibiotics. The provider order entry (POE) system was deployed in the early 2000s, and antibiotic information is only available from 2003 onward. As MetaVision patients are admitted from 2008 onward, this exclusion isolates to the population of MIMIC which have antibiotic prescription measured.



**Supplemental Figure 2.** A three-set Venn diagram showing the overlap amongst the Martin, Angus, and Explicit coding criteria. Note that the Explicit criteria are entirely subsumed by the Angus and almost entirely subsumed by the Martin criteria. It is also worth noting that the Martin criteria were proposed in 2003, before the introduction of the explicit ICD-9 codes in this hospital (see Supplemental Figure 2).



###

###

### Detailed discussion of the usefulness of the criteria within a recently proposed framework

Angus et al. [1] provide a framework for evaluating sepsis criteria, and we adopt this framework for assessing the Sepsis-3 criteria for secondary analysis of electronic health records. This framework involves assessing the criteria in six domains, and determining whether these domains are important for the application of the criteria. For example, a criteria for retrospective research does not need information to be available early in the patient’s stay, but a criteria for clinical care does require timely delivery of information. Seymour et al. [2] apply the framework to assess the usefulness of alternative criteria. There are six domains assessed for sepsis criteria within the framework:

* Reliability – criteria are stable and reproducible
* Content validity – criteria make clinical sense
* Construct validity – criteria measure what they purport to measure
* Criterion validity – criteria agree with the existing standard
* Measurement burden – criteria have low cost, low risk to patients, and low complexity
* Timeliness – criteria are generated in a timely manner in relation to illness progression

We now iterate through these domains, highlighting the importance of these domains in electronic health record (EHR) research and discussing the presented criteria in the context of that domain.

*Reliability – Importance:* ***High****.* In order to leverage the increasingly large EHR databases, criteria derived must reliably produce the same results across a variety of settings. Sepsis criteria which are based on administrative coding may be susceptible to changes in coding practices unrelated to patient physiology, e.g. the introduction of new ICD-9 codes may cause apparent spikes in disease. The criteria of Angus et al., Martin et al., and the explicit criteria are most susceptible to this as they are purely administrative criteria. The CMS criteria requires administrative coding, but refines this subgroup with physiologic criteria. The Sepsis-3 and CDC criteria avoid this issue by omitting all administrative coding. However, both these latter criteria rely on treatments as surrogates for organ failure. For example, the cardiovascular component of SOFA is 2 if a patient is administered low-dose dopamine, though this is less frequently done in contemporary clinical practice. As a result, the test lacks *meta-reliability*, that is, it is susceptible to changes unrelated to the biology of the patient. Issues such as this can be avoided by lessening the dependence of the criteria on specific treatments, i.e. instead of quantifying the level of organ dysfunction based on the type of vasopressor (as is done in SOFA), criteria could be simplified to use of any vasopressor (such as in the CDC definition).

*Content validity – Importance:* ***High***. Key to translation of research results into clinical practice relies is acceptance of definitions used in the study by the practicing clinician, i.e. a high content validity of the criteria. The Sepsis-3 task force’s new definition of sepsis is that of a life-threatening organ failure subsequent to infection. Under this predisposition, the Sepsis-3 criteria of suspected infection with quantified organ failure has high content validity. Similarly, the CDC and CMS criteria both exhibit high content validity. The criterion of Angus et al. requires confirmed infection which deviates from modern clinical intuition. Finally, the criterion of Martin et al. has low content validity, focusing on the concept of septicemia and blood borne infection, of which sepsis is a superset.

*Construct validity – Importance:* ***High****.* Construction of decision support tools, association analysis of novel computational biomarkers, and phenotyping of disease states all rely on high construct validity of the criteria utilized. We quantitatively evaluate the criteria for construct validity using a modified form of the multitrait multimethod matrix. Unsurprisingly the explicit, CMS, and Martin et al. criteria all had excellent agreement as they share a large number of ICD-9 codes. Agreement amongst the various criteria was adequate (> 0.4) in all cases and good (> 0.6) in most cases. While construct validity is difficult to concretely measure, these results are reassuring.

For the Sepsis-3 criteria, there are two components: suspicion of infection and organ failure. The algorithm for suspicion of infection involves the acquisition of a blood culture (test for infection) contemporaneous to administration of antibiotics (a treatment for infection), and is an intuitively sensitive marker for suspicion of infection with high construct validity. Organ failure is captured by SOFA which has weaker construct validity, particularly in regard to the neurological, respiratory, and cardiovascular components. The neurological component utilizes the Glasgow Coma Scale, which has known issues regarding inter-rater reliability [?]. In the absence of advances in automated quantification of neurological function, e.g. via the EEG, simplifications such as only requiring altered mentation (perhaps using GCS < 15) may improve construct validity. The respiratory component uses a low PaO2:FiO2 ratio as a marker of severity of illness, and while low PaO2:FiO2 is an indicator of severity, it is confounded by dependence on the current treatment regimen. Finally, the cardiovascular component is primarily determined by the type and rate of vasopressor administration, and not on the degree of organ failure. The component is thus susceptible to the clinician’s propensity for fluid resuscitation versus pharmacological intervention.

*Criterion validity – Importance:* ***Moderate***. As there is no easily available gold standard of sepsis, it is difficult to ascribe high importance to criterion validity. We evaluated the criteria in a predictive fashion using the surrogate outcomes of mortality and long length of stay as these outcomes are strongly related to sepsis (and were those used by Seymour et al.). Mortality, and arguably criterion validity, was highest among the most selective criteria which focused on septicemia. As a broader cohort was selected, the mortality rate and average length of stay decreased. All criteria identified a population at high risk (>10% mortality according to the ESICM/SCCM task force), indicating good predictive criterion validity.

*Measurement burden – Importance:* ***Moderate****.* Research on EHRs has low cost and low patient risk, but suffers from high complexity. The biggest challenge involves identifying clinical concepts from surrogate measures, e.g. intubation and extubation from ventilator settings, total fluid balance from hourly measurements, and so on. An example of this complexity is the PaO2:FiO2 ratio, which can be extracted (i) only if explicitly documented or (ii) by retrospectively matching PaO2 values with FiO2 values. The former approach may result in large amounts of missing data, while the latter approach requires complex coding and definition of an acceptable lead time of FiO2 compared to PaO2. As the number of physiologic parameters increase, so do the number of decisions made by the researcher. One solution is open sourcing code used in the analysis and consequently providing other researchers full insight into the modelling decisions made, as was done in this paper.

 *Timeliness – Importance:* ***Low****.* As analysis of EHRs is inherently retrospective, the timeliness of the criteria is of low importance. If optimal criteria are determined with low timeliness, future research can focus on approximating the criteria with more timely measures which could then be directly translated into decision support tools. The Sepsis-3 criteria are timely, but they need not be so. One challenge is the need for a “baseline” measurement of organ function, which is rarely available. The CDC and CMS definitions circumvent this issue by identifying organ failure in relation to the most normal measurement during the patient’s entire hospitalization. While this is a fundamentally acausal calculation, it likely provides a better estimate of baseline organ function for the individual patient, and allows for definitions of organ failure which use relative changes (e.g. creatinine increase of 50%).

### References

[1] Angus DC, Seymour CW, Coopersmith CM, Deutschman C, Klompas M, Levy MM, Martin GS, Osborn TM, Rhee C, Watson RS. A framework for the development and interpretation of different sepsis definitions and clinical criteria. Critical care medicine. 2016 Mar;44(3):e113.

[2] Seymour CW, Coopersmith CM, Deutschman CS, Gesten F, Klompas M, Levy M, Martin GS, Osborn TM, Rhee C, Warren D, Watson RS. Application of a framework to assess the usefulness of alternative sepsis criteria. Critical care medicine. 2016 Mar;44(3):e122.