Critical Review Form
Clinical Prediction or Decision Rule

Collins SP, Jenkins CA, Harrell FE Jr, et al. Identification of Emergency Department Patients With Acute Heart Failure at Low Risk for 30-Day Adverse Events: The STRATIFY Decision Tool. JACC Heart Fail. 2015 Oct;3(10):737-47.

Objectives: **"to develop an AHF [acute heart failure] decision tool to identify ED patients at low risk of death or serious complications who could therefore be considered for ED discharge and subsequent outpatient management." (p. 738)**

Methods: **This prospective, observation cohort study was conducted at 2 academic tertiary care EDs and 2 community EDs in Nashville, TN (n = 1) and Cincinnati, OH (n = 3) between July 20, 2007 and February 4, 2011. A convenience sample of patients being treated for AHF were screened for enrollment when study associates were available (16 hours a day during the week and 12 hours on the weekends). All treatment and disposition decisions were made by the treating clinician without influence from study personnel.**

Using a modified set of Framingham criteria for the diagnosis of heart failure, patients aged 18 years or older with the presence of at least 2 major, or 1 major and 2 minor criteria were eligible for inclusion. Data was collected by patient and physician interview and by review of the electronic medical record during the initial 3 hours of ED management. Two investigators, blinded to the inpatient medical record, reviewed the ED medical record independently to confirm that the ultimate ED diagnosis with AHF. Adjudication by a third reviewer occurred in cases of disagreement.

The primary outcome was "the most severe adverse event experienced within 30 days of ED evaluation." (p. 740) A five-point ordinal scale created *a priori* by both emergency physicians and cardiologists was used to rank the severity of adverse events, with more serious events receiving higher weighting than less serious events. Adverse events were determined by both chart review and telephone follow-up at 5 days and 30 days after enrollment, and by Social Security Death Index search at 30 and 90 days if no contact had been made.

Once the STRATIFY decision rule was derived by statistical analysis of a predetermined set of predictors, the rule was then validated using bootstrap resampling. A total of 2074 subjects were recruited, of whom 63 withdrew and 18 were lost to follow-up. Of the remaining cohort, 1033 were determined to have had AHF in the ED. The median age was 64, 57% were male, and 44% were African American. Adverse events occurred within 5 days in 7% of patients, and within 30 days in 12%.

Guide		Comments
I.	***Is this a newly derived instrument* (Level IV)?**	
A.	Was validation restricted to the retrospective use of statistical techniques on the original database? (If so, this is a Level IV rule & is not ready for clinical application).	Yes. While the authors used a bootstrapping method to "validate" their clinical decision rule, such a method does not replace prospective validation of the rule on an independent sample of patients. *Caveat: some would argue this raises the level of the rule to 3c.
II.	**Has the instrument been validated? (Level II or III). If so, consider the following:**	
1a	Were all important predictors included in the derivation process?	Seemingly yes. The authors report that, "a large number of ED candidate predictor variables were considered based on established risk factors for AHF." (p. 740) However, they do not provide a comprehensive list of all 57 predictors included in the process, either in the body of the paper, or in an appendix or supplement.
1b	Were all important predictors present in significant proportion of the study population?	Difficult to quantify, as most of the predictors were continuous variables (such as age, lab values, and vital signs) and the authors do not provide a range a range for these variables. For dichotomous predictors, the authors do not provide information regarding the number of patients with those predictors.
1c	Does the rule make clinical sense?	Yes. The rule is comprised of 13 variables, including laboratory values, vital signs, and demographic information that seems clinically relevant to risk of deterioration in patients with AHF.
2	Did validation include prospective studies on several different populations from that used to derive it (II) or was it restricted to a single population (III)?	No. Validation was only conducted using bootstrap resampling with the original patient cohort.
3	*How well did the validation study meet the following criteria?*	
3a	Did the patients represent a wide spectrum of severity of disease?	Uncertain, but likely so. The authors provide limited information regarding the spectrum of disease, including quartiles for certain vital signs (SBP, pulse, RR) and lab values (BNP, sodium, BUN, hemoglobin). They do not provide such information or ranges for most of the criteria used in the final rule. Given the

		method of enrollment and the participation of four different clinical sites, it seems likely that a wide array of patients with a wide spectrum of disease was included. The authors did not specifically exclude patients requiring invasive or noninvasive positive pressure ventilation, or those with mild disease whom the treating clinician already planned to discharge.
3b	Was there a blinded assessment of the gold standard?	Somewhat. There was no specific gold standard in this study, but rather a composite of 5 outcomes was used. Most of these outcomes are fairly objective (death, need for emergent dialysis, intubation, PCI, CABG, or mechanical cardiac support) the diagnosis of ACS is can be somewhat subjective. The assessment was blinded, as the authors note that "the investigators assessing outcomes were masked to the predictor variables and vice versa." (p. 740)
3c	Was there an explicit and accurate interpretation of the predictor variables & the actual rule without knowledge of the outcome?	Yes. As noted above, "the investigators assessing outcomes were masked to the predictor variables and vice versa." (p. 740)
3d	Did the results of the assessment of the variables or of the rule influence the decision to perform the gold standard?	N/A. Again, there was no gold standard test, but rather an assessment was made to see if one of several adverse events occurred. This assessment was made for all patients, regardless of the results of predictor variable assessment.
4	How powerful is the rule (in terms of sensitivity & specificity; likelihood ratios; proportions with alternative outcomes; or relative risks or absolute outcome rates)?	• Using a low-risk cut-threshold of 3%, 5%, and 10% risk of adverse outcomes, the rule had negative predictive values of 100%, 96%, and 93% respectively, sensitivity of 100%, 95%, and 71% respectively, and specificity of 2%, 14%, and 52%, respectively. • 1.4% of patients were found to have a risk of 3% or less for adverse events, with an LR+ of 1.0 2(95% CI 1.02-1.03) and LR - of 0.00 (95% CI 0.01-4.11). • 13% of patients were found to have a risk of 5% or less for adverse events, with an LR+ of 1.11 (95% CI 1.06-1.16) and LR - of 0.34 (95% CI 0.15-0.74). • 49.5% of patients were found to have a risk of 10% or less for adverse events, with an LR+ of 1.50 (95% CI 132-1.71) and LR - of 0.55 (95% CI 0.41-0.72).

III.	**Has an impact analysis demonstrated change in clinical behavior or patient outcomes as a result of using the instrument? (Level I). If so, consider the following:**	
1	How well did the study guard against bias in terms of differences at the start (concealed randomization, adjustment in analysis) or as the study proceeded (blinding, co-intervention, loss to follow-up)?	Fairly well. The authors enrolled only a convenience sample of patients. While they did make multiple attempts to contact patients for follow-up, they do not mention blinding of outcome assessors. Only 18 patients were lost to follow-up, but it is unclear if the remaining patients were all contacted, or if their follow-up was only through chart review.
2	What was the impact on clinician behavior and patient-important outcomes?	This was not addressed in this study. Given that for a risk threshold of 3%—which was the only cutoff found to have a sufficiently low LR- to be safely used to identify low-risk patients without a high false positive rate—only 1.4% of the study population would have been deemed low risk, it is unlikely that this rule would have any meaningful clinical utility.

Limitations:

1. **Out of nearly 2000 patients recruited and diagnosed with AHF, only 1033 were determined to actually have had AHF in the ED, making it nearly impossible to use this rule prospectively.**

2. **Bootstrapping was used to validate the rule, rather than independent validation in a new subset of patients. This is therefore a level 4 rule and cannot be used until it is validated in multiple environments.**

3. **The authors did not provide comprehensive list of predictor variables used to create the decision rule either in the body of the paper or in an appendix or supplement.**

4. **The authors enrolled only a convenience sample of patients and provide no information regarding patients that were eligible but not enrolled.**

5. **The rule and the nomogram provided are very complicated and not very intuitive. Use of this rule in clinical practice with EHR support to calculate risk would likely be cumbersome and distracting.**

6. **A sufficient LR- was seen only using a risk threshold cut-off of 3%. Unfortunately, only 1.4% of the patients were found to have such a low risk, making the clinical impact of the rule negligible.**

<u>Bottom Line:</u>

This prospective, multicenter trial sought to derive a clinical decision rule to identify patients treated in the ED for acute heart failure who are at low risk of adverse outcomes. They were able to derive a rule that identified patients with a < 3% chance of adverse events with a negative likelihood ratio of 0.00 (95% CI 0.01-4.11). Unfortunately, only 1.4% of patients would be identified by using such a cutoff, making the clinical impact of the rule negligible. In addition, the rule itself requires the use of a complicated nomogram that would make its use in the ED overly cumbersome.