Classification of Heart Failure in the Atherosclerosis Risk in Communities (ARIC) Study: A Comparison of Diagnostic Criteria

Rosamond et al: Comparison of Heart Failure Classification Methods

Wayne D. Rosamond, PhD, MS; Patricia P. Chang, MD; Chris Baggett, PhD; Anna Johnson, PhD; Alain G. Berton, MD; Eyal Shahar, MD; Anita Deswal, MD, MPH; Gerardo Heiss, MD, PhD; Lloyd E. Chambless, PhD

Departments of Epidemiology, Medicine and Biostatistics, University of North Carolina, Chapel Hill, NC; Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC; Division of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ; Section of Cardiology, Michael E. DeBakey VA Medical Center & Baylor College of Medicine, Houston, TX

Correspondence to
Wayne Rosamond, PhD, MS,
Department of Epidemiology
University of North Carolina at Chapel Hill
137 E. Franklin Street, Suite 203
Chapel Hill, NC 27514
Tel: 919-962-3230
Fax: 919-966-9800
Email: wayne_rosamond@unc.edu

Journal Subject Codes: [110], [8]
Abstract

Background—Population-based research on heart failure (HF) is hindered by lack of consensus on diagnostic criteria. Framingham (FRM), National Health and Nutrition Examination Survey (NHANES), Modified Boston (MBS), Gothenburg (GTH), and International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM) code criteria do not differentiate acute decompensated heart failure (ADHF) from chronic stable HF. We developed a new classification protocol for identifying ADHF in the Atherosclerosis Risk in Communities (ARIC) Study and compared it with these other schemes.

Methods and Results—A sample of 1180 hospitalizations with a patient address in four study communities and eligible discharge codes were selected. After assessing whether the chart contained evidence of possible HF signs, 705 were fully abstracted. Two independent reviewers classified each case as ADHF, chronic stable HF or no HF using ARIC classification guidelines. Fifty-nine percent of cases met ARIC criteria for ADHF and 13.9% and 27.1% were classified as chronic stable HF or no HF, respectively. Among events classified as HF by FRM criteria, 68.4% were validated as ADHF, 9.6% as chronic stable HF and 21.9% as no HF. However, 92.5% of hospitalizations with a primary ICD-9-CM 428 “heart failure” code were validated as ADHF. Sensitivities of comparison criteria to classify ADHF ranged from 38 to 95%, positive predictive values from 62 to 92%, and specificities from 19 to 96%.

Conclusions—Although comparison criteria for classifying HF were moderately sensitive in identifying ADHF, specificity varied when applied to a randomly selected set of suspected HF hospitalizations in the community.

Key Words: heart failure, epidemiology
Heart failure (HF) is a complex clinical syndrome resulting from a structural or functional cardiac disorder that impairs the ability of one or both ventricles to fill with or eject blood sufficiently to meet the needs of the body. There is no universally accepted definition of HF. Signs and symptoms may differ depending on the level of systolic or diastolic dysfunction and further complicate disease classification. Various diagnostic criteria are published, and comparisons between these criteria report mixed results. Population-based studies in HF are challenged by the lack of clear diagnostic consensus, making estimates of prevalence and incidence difficult to interpret and compare. Furthermore, currently available classification criteria do not differentiate acute decompensated HF (ADHF) episodes from other clinical events accompanied with chronic stable HF. Separating acute from chronic HF in population based studies would enhance our understanding of prediction and prevention of HF as well as provide better estimates of trends of the HF burden in the general population. In 2005, The Atherosclerosis Risk in Communities (ARIC) Study began surveillance of HF and developed a process to classify hospitalizations for ADHF and chronic stable HF. The purpose of this report is to describe the ARIC HF classification guidelines and compare its classification of ADHF and chronic stable HF with five established diagnostic schemes for HF.

Methods

Beginning in 2005, the ARIC Study conducted continuous, retrospective surveillance of hospital discharges for HF for all residents age 55 years and older in four US communities: Forsyth County, North Carolina; the city of Jackson, Mississippi; eight northwest suburbs
of Minneapolis, Minnesota; and Washington County, Maryland. In 2005, there were 31 hospitals serving the four ARIC communities. The combined population in 2005 for these regions was approximately 177,000 men and women 55 years of age or older. Because of the small number of hospitalizations in the sample among race/ethnic groups other than black or white (n=55), we categorized these as white for the purposes of these analyses.

Annual electronic discharge indices were obtained from all hospitals admitting residents from the four ARIC communities. Discharges meeting eligibility criteria were sampled from these files. A hospitalization was considered eligible for validation as a HF event based on its International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM) code, age, gender, race, and residence in the community surveillance area. Target primary or secondary hospital discharge diagnoses codes included: heart failure (428), rheumatic heart disease (398.91), hypertensive heart disease- with congestive heart failure (402.01, 402.11, 402.91), hypertensive heart disease and renal failure- with CHF (404.01, 404.03, 404.13, 404.91, 404.93), acute cor pulmonale (415.0), chronic pulmonary heart disease, unspecified (416.9), other primary cardiomyopathies (425.4), acute edema of lung, unspecified (518.4), dyspnea and respiratory abnormalities (786.0). Sampling probabilities were created to optimize variance estimates around event rate estimates with a pre-set maximum number of cases to be abstracted in 2005 of 1499 (See Supplemental Methods). This fixed number of abstractions was estimated and set based on a target number (n=500) of hospitalized events that could be investigated and validated considering available resources and time constraints. All analyses were weighted to account for the sampling probabilities.
Diagnostic Methods

Centrally trained and certified staff abstracted data from eligible medical records in two steps. First, the record was reviewed for any evidence of relevant HF symptoms (i.e. new onset or worsening of shortness of breath, edema, paroxysmal nocturnal dyspnea, orthopnea, or hypoxia) or any mention by the treating physicians of HF as the reason for the hospitalization. If the hospitalization included such evidence, a second more detailed abstraction of the medical record was completed. Detailed abstraction included recording: evidence of new onset of symptoms, history of HF, general medical history, physical exam signs and symptoms, diagnostic tests (chest X-ray, echocardiogram, cardiac catheterization, coronary angiography, cardiac radionuclide ventriculogram, cardiac magnetic resonance imaging, cardiac CT scan, stress test), biomarkers (brain natriuretic peptide (BNP), N-terminal pro-hormone brain natriuretic peptide (pro-BNP)), and medications. Data abstracted included required elements of four diagnostic criteria commonly used in comparative studies: Framingham (FRM) 5, modified Boston (MBS) 6, National Health and Nutrition Examination Survey (NHANES) 8, and Gothenburg (GTH) 7. See Supplemental Methods. A fifth HF diagnostic scheme using ICD-9-CM coding was also used. Abstractors made copies of sections of the medical record (discharge summary, history and physical report, admission note, and imaging reports) for use by the ARIC HF Classification Committee. The inter-abstractor agreement rate for determining whether or not to conduct detailed abstraction in a quality control sample was 99%.
In addition to a HF classification based on the five comparison schemes mentioned above (FRM, MBS, NHANES, GTH, and, *ICD-9-CM*) each hospitalization eligible for full abstraction was independently reviewed by two centrally trained and certified physicians on the ARIC HF Classification Committee. The reviewers were provided a report of the abstracted data as well as the copied materials noted above. Each reviewer was provided a summary of the abstracted data noted above (including measurement of ejection fraction and biomarkers) and the copied portions of medical records, and in light of the guidelines below classified each hospitalization into one of five categories: definite ADHF, possible ADHF, chronic stable HF, HF unlikely, or unclassifiable. Physicians were trained and certified to follow the ARIC classification guidelines were randomly assigned cases to review. A single physician adjudicator (Chair of the ARIC HF Classification Committee) resolved disagreements. For the purpose of this report we combined cases classified as either definite or possible ADHF into one category designated as ADHF.

**ARIC Heart Failure Event Classification Guidelines**

**Acute Decompensated Heart Failure**

Definite ADHF required clear evidence either from symptoms, signs, imaging or treatment of an acute exacerbation, worsening or new onset of symptoms or other decompensated circulatory state. Evidence of a decompensated state included augmentation of therapy for worsening HF signs or symptoms, documentation of subsequent in-hospital control of symptoms by therapy, documentation of the specificity of HF for decompensated state as opposed to other co-morbidities (e.g. chronic obstructive pulmonary disease (COPD), end-stage renal disease). For a classification of definite ADHF, evidence that the HF treatment
(e.g. diuresis) was the main treatment that resulted in improvement is required. For example, control of symptoms by therapy would include diuresis followed by relevant weight loss, clinical improvement in symptoms or of pulmonary edema on chest x-ray, or evidence that the patient no longer requires oxygen. A case was considered possible ADHF if the presence of co-morbidity could also account for the acute symptoms or if there was not enough information to classify as definite ADHF. For example, in cases in which renal failure, chronic COPD, or pneumonia may also be the etiology of the presentation, or where multiple treatments are provided that result in clinical improvement of symptoms (e.g. antibiotics for possible pneumonia, nebulizers for possible COPD, and diuretics for possible HF), then a classification of possible ADHF is preferred.

Chronic Stable Heart Failure

Chronic stable HF required evidence of compensated HF signs and symptoms controlled by therapy with no evidence of therapy augmentation or symptom worsening during the hospitalization. Evidence of left ventricular systolic dysfunction (ejection fraction < 50%) with no HF symptoms was sufficient for classification as chronic stable HF. Asymptomatic diastolic dysfunction was not sufficient for a classification as chronic stable HF.

Heart failure unlikely and unclassifiable events

Hospitalizations were classified as no HF if the available documentation in the medical record indicated directly or indirectly that heart function was normal. A designation of unclassifiable was usually used in cases where medical records were insufficient to
differentiate between a classification of chronic stable HF and no HF or in the infrequent case of missing medical records. For the purposes of these analyses, cases classified as HF unlikely or determined to be unclassifiable were combined as no HF.

**Data Analysis**

We computed reliability and validity metrics comparing ARIC classification and the five comparison diagnostic schemes using two rubrics. First, we compared a three-level ARIC HF category (ADHF, chronic stable HF, no HF) with results of the algorithms using FRM, MBS, NHANES, GTH, and *ICD-9-CM* heart failure schemes. Second, we created a more general two-level ARIC HF classification combining ADHF together with chronic stable HF and compared this two-level ARIC category (i.e., ADHF or chronic stable HF, no HF) with the above criteria.

Calculations of percent agreement and kappa coefficients transformed the 3-level MBS and 5-level GTH classifications into dichotomous groups (HF, no HF). We combined GTH criteria levels 2 and 3 together as a positive classification for HF. For MBS criteria, the categories of definite and probable HF were combined. NHANES and FRM criteria were retained as their original two-level categories. We also created two *ICD-9-CM* code-based criteria for comparison purposes. One considered the presence of an *ICD-9-CM* 428 code in any position on the discharge list as sufficient to be classified as HF and the other required a 428 code as the primary discharge diagnosis.
Sensitivity, positive predictive value, and specificity using the ARIC HF classification as the gold standard were computed in the standard fashion. The comparability ratio reported was computed as the ratio of the number of HF events defined by established criteria to the number of HF events validated as determined by the ARIC HF guidelines. We defined specificity as the proportion of sampled and reviewed hospitalizations that were classified by ARIC HF review as non-HF events that were classified as non-HF by the comparison criteria.

We assessed percent agreement between ARIC HF classification and the comparison criteria using standard methods and chance-corrected agreement by kappa coefficients. Chi-square tests on the weighted proportions were used to determine statistical significance of differences in percent of events validated by ARIC classification.

**Results**

In 2005, residents age 55 years or older in the four ARIC communities had 11,544 hospital discharges with ICD-9-CM diagnosis codes within our target list. We selected a random sample of 1499 hospitalizations for investigation. After exclusion of hospitalizations where medical records were unavailable (n=16), that contained ineligible patient addresses (n=303) or that lacked relevant HF symptoms needed for full abstraction (n=475), we conducted detail abstraction and validation of 705 hospitalizations. The agreement rate between two physician reviewers for classifying an event was 75% for hospitalization with an ICD-9-CM 428 code and 86% for hospitalizations without an ICD-9-CM 428 code.
Table 1 shows the classification of all sampled HF hospitalizations combining those fully abstracted and reviewed by ARIC (n=705) with those hospitalizations not eligible for full abstraction and not reviewed by the committee (n=475). For the purposes of this analysis, we categorized this latter group as non-HF hospitalizations. For all hospitalizations, 36% were classified as ADHF, 8.5% as chronic stable HF, and 10.2% as no HF. A small percent (6.3%) were not classifiable by the classification committee and 38.9% did not meet initial screening to merit full review and are considered hospitalizations for reasons other than for HF. Men, blacks, and hospitalizations with an ICD-9-CM 428 code were more likely to be validated as ADHF. Differences in the percent of events validated as ADHF across the four communities were not statistically significant. Among all sampled hospitalizations (including those not meeting the full abstraction criteria) with an ICD-9-CM 428 code in any position, 38.5% were validated as acute decompensated HF and 9.1% as chronic stable HF. Of note, 16.8% of cases without an ICD-9-CM 428 code were validated as ADHF by ARIC review. The majority (88.3%) of hospitalizations with an ICD-9-CM 428 code as the primary discharged diagnosis were validated as ADHF.

The percent of hospitalization meriting full abstraction and review that were validated as HF using each of the comparison classification criteria is shown in Table 2. Of the hospitalizations meeting FRM criteria for HF, 68.4% were classified as ADHF by ARIC review. An additional 9.6% were classified as chronic stable HF and 21.9% were determined to be hospitalization for conditions other than HF. Approximately one-quarter of hospitalizations determined not to be HF by FRM criteria were actually classified as
ADHF by ARIC review. A similar pattern was seen when comparing ARIC review with MBS, NHANES, and GTH criteria or to presence of an *ICD-9-CM* 428 code in any position. However, among hospitalizations with a primary discharge diagnosis of HF (*ICD-9-CM* 428), 93.0% were validated as ADHF.

The crude agreement between the various classifications schema was moderate (Table 3). As expected, the agreement between ARIC review and the comparison classification criteria increased when the ARIC review endpoints of ADHF plus chronic stable HF were combined. Chance corrected estimates of agreement (kappa) between criteria were generally poor.

Framingham criteria were 90% sensitive and 40% specific for classifying ADHF (Table 4). These combined with a positive predictive value of 68% resulted in a comparability ratio of 1.31. The sensitivity and specificity of FRM criteria was slightly reduced to 83% and 37%, respectively and the positive predictive value increased from 68% to 78% when compared to the combined ARIC endpoint of either (ADHF plus chronic stable HF). As a result, the comparability ratio improved to 1.06 when FRM criteria are used to estimate the presence of either ADHF or chronic stable HF. Similar results were seen for MBS, NHANES and GTH criteria. Although the sensitivity of an *ICD-9-CM* 428 code in any position was slightly higher (sensitivity = 95%) compared to the other established criteria, its comparability ratio for classifying either AHDF deviated from unity more than the other criteria (comparability ratio =1.52). Presence of a primary *ICD-9-CM* 428 discharge code had high specificity (95%), but poor sensitivity (43%) for ADHF.
The concordance of hospitalizations classified as HF by FRM, ARIC, or primary discharge diagnosis code is shown on the Figure. Only 28% of cases meet all three criteria and an equal proportion (28%) met FRM criteria but not ARIC or discharge code criteria. A small proportion of cases (5%) were called HF by ARIC when FRM or discharge code criteria indicated a non-HF event. The percent overlap between these three classifications increased to 52% when expanding the discharge code definition to include a 428 code listed in any position (data not shown).

**Discussion**

The ARIC HF classification guidelines described in this paper provide a more detailed categorization of HF hospitalizations than currently available criteria. The ARIC classification was specifically designed to differentiate ADHF from hospitalizations associated with chronic or stable HF, a feature not possible with the other commonly used criteria. Thus, the ARIC HF classification protocol is likely to result in improved accuracy of the rates of ADHF hospitalizations (although, because many people with chronic HF are not hospitalized at the time of diagnosis, it will not necessarily result in improved accuracy of total HF incidence). Although the other criteria were not designed for this level of granularity in classification, evaluating their validity in classifying ADHF as well as total HF may help inform interpretation of previous work as well as shape future studies of HF.
We found that the five comparison diagnostic criteria were highly sensitive in identifying ADHF but had poor specificity. Comparison criteria had similar levels of accuracy with one another in identifying any HF (decompensated or chronic stable HF). In contrast, a primary ICD-9-CM 428 discharge code had poor sensitivity in identifying either decompensated or any HF but was highly specific for both. These measures of validity combined with the moderate to poor agreement among all classification schemes underscore the lack of consensus on epidemiologic definitions of HF.

The limited population-based data available on the incidence of HF use varying criteria. Although the FRM criteria has emerged as the standard for the identification of HF in many epidemiologic studies; studies disagree about which is diagnostically superior. In studies using echocardiographic evidence of left ventricular dysfunction as a gold standard, FRM criteria were found to have high sensitivity (92%) but moderate specificity (79%). In contrast, FRM criteria were reported to have lower sensitivity and specificity in a study of suspected HF patients who were referred for radioisotopic assessment of systolic ventricular function.

Remes and colleague reported on cases of clinically suspected HF in comparison with Boston criteria’s HF diagnosis and found a relatively high sensitivity and specificity (80% and 92%, respectively). In contrast, Mosterd and colleagues (1997) report that the sensitivity of FRM, MBS, GTH, NHANES classifications schemes relative to clinical cardiologist’s diagnosis vary considerably. In a large community-based study of seven
HF criteria using clinical physician review as the gold standard, all criteria investigated had low sensitivity (range 46% to 84%) yet high specificity (range 81% to 96%) \(^{27}\).

Studies that assess agreement among the criteria are equally as varied in their conclusions. Substantial concordance among the MBS FRM, NHANES schemes (kappa coefficients generally > 0.60) have been reported \(^{27,30}\), but agreement between GTH and the others are poor. Our findings of poor agreement among criteria are supported by previous work comparing FRM criteria, European Society of Cardiology (ESC) criteria and independent physician review \(^{11}\).

When the FRM criteria were compared with those developed by the Cardiovascular Health Study (CHS), the FRM criteria resulted in an incidence estimate approximately 23 percent greater than the estimate calculated using the CHS criteria \(^{31}\). In another setting, HF incidence varied using Boston, FRM, GTH and ESC criteria (12%, 11%, 21%, and 9%, respectively) \(^{11}\). Boston criteria more accurately identified HF cases (using physician review as gold standard) than GTH, FRM, and ESC criteria and also were better at predicting cardiovascular death, incident disability, and hospitalizations \(^{11}\). The incorporation of echocardiography or biomarkers evidence in physician review, neither of which are included in most established diagnostic criteria may result in earlier detection of less advanced cases and may result in lead-time bias. \(^{18,32}\)

Elements included in each of the comparison HF criteria differ. While all four HF incorporate patient’s medical history and physical examination, a chest X-ray is not
required in the GTH criteria, and the FRM score is the only one to incorporate vital capacity. Many of the current criteria rely on elements frequently missing in routine medical records. The FRM and Boston criteria rely heavily upon the presence of pulmonary congestion to diagnose HF; however, this may limit the ability to adequately classify HF in the presence of preserved systolic function. 3, 29 Key differences in the ARIC HF classification guidelines compared to HF criteria scores are that it incorporates more current diagnostic tests, which have been shown to improve prognostic ability 10, 17, 33-34. These diagnostic methods are becoming increasingly available for clinical use 35 as well as increasingly required in HF definitions used in clinical trials 14.

The differentiation between ADHF and chronic stable HF is crucial for epidemiologic studies of HF etiology. HF mortality rates differ based on the underlying cause of HF 36 and can vary by the population studied and the differential criteria used making proper categorization of HF essential. 25 While some studies exclude patients who developed HF secondary to admission for another illness 16, others include them but do not adequately determine these underlying conditions 25. Given that substantial race/ethnic 37 and gender 38 differences in HF etiology exist, accurate classification is also critical in measuring and preventing HF in different populations. Clinicians and policy makers are concerned with reducing HF rehospitalizations. Improved epidemiologic methods for differentiating between ADHF and chronic stable HF would improve our accuracy in defining rehospitalizations due to ADHF and aid the examination of outcomes and how they related to clinical practice, therapy advances and policies.
Studies employing *ICD-9-CM* diagnosis codes to define HF are not consistent. Some define HF as a primary discharge code of 428; others include patients with 428 listed in any position. In our study, 39% of hospitalizations with a 428 code in any position were categorized as ADHF and 9% were categorized as chronic stable HF. Further, 12% of cases with a primary discharge diagnosis of HF were determined to not have HF. We found that 17% of cases without an 428 code were validated as ADHF, suggesting that limiting the definition of HF to a code of 428 may result in inaccurately low estimates of HF. Indeed, studies using claims databases often use the presence of primary diagnosis codes 402 or 404 in addition to 428 to define HF events\(^{39-40}\). In our study, adding these additional code groups did not appreciably change the validation estimates (data not shown).

**Strengths and Limitations**

Diagnostic accuracy was rigorously tested, with each case being subject to review by two independent physician reviewers. However, a number of study limitations must be considered. In some chronic HF cases, it may be difficult to determine whether the patient’s status matches the baseline HF status or indicates some deterioration. In these cases, the totality of the evidence provided was taken into consideration. A potential limitation in all studies of this type is that the diagnostic accuracy of criteria depends on the population characteristics including the prevalence of HF.
Conclusions

An improved method of diagnosis of HF is critical if primary and secondary prevention efforts are to target individuals at risk for HF\textsuperscript{15}. The ARIC HF classification guidelines, created for use in ongoing community and cohort surveillance, provide a methodology for the diagnosis of hospitalized HF and differentiate ADHF and chronic stable HF. These methods could be applied in other study populations where access to medical records is feasible and some training of reviewers to follow the guidelines presented is practical. We found that a principle ICD-9-CM code for heart failure (code 428) was highly specific but had poor sensitivity for ADHF using the ARIC classification. A next step in the assessment of this classification system is to investigate its accuracy for the prediction of outcomes such as mortality, disability and future HF-related hospitalizations and ultimately to evaluate disease trends.
Acknowledgements

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022. The authors thank the staff and participants of the ARIC Study for their important contributions.

Sources of Funding

The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022.

Disclosures

None.
References


2005; 118:728-734.

17. Roger VL, Weston SA, Redfield MM, Hellermann-Homan, JP, Killian J, Yawn BP,
Jacobsen SJ. Trend in heart failure incidence and survival in a community-based

18. Levy D, Kenchaiah S, Larson MG, Benjamin EJ, Kupka MJ, Ho KL, Murabito JM,
Vasan RS. Long-term trends in the incidence of and survival with heart failure. N

Press; 1986.

20. Fleiss J. *Statistical methods for rates and proportions*. New York: John Wiley and
Sons’ 1981.

21. McKee PA, Castelli WP, McNamara PM, Kannel BW. The natural history of
1446.

121:951–957.


24. The Task Force on Heart Failure of the European Society of Cardiology.


Table 1. Proportion (weighted) of hospitalizations classified according to ARIC HF Criteria by demographic characteristics of events, The Atherosclerosis Risk in Communities (ARIC) Study.

<table>
<thead>
<tr>
<th>Heart Failure classification by ARIC Classification Committee *</th>
<th>Acute decompensated heart failure</th>
<th>Chronic stable heart failure</th>
<th>No heart failure</th>
<th>Unclassifiable Did not meet screening criteria †</th>
<th>Total</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>36.0</td>
<td>8.5</td>
<td>10.2</td>
<td>6.3</td>
<td>38.9</td>
<td>100</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>39.0</td>
<td>10.7</td>
<td>8.4</td>
<td>6.6</td>
<td>35.4</td>
<td>100</td>
</tr>
<tr>
<td>Women</td>
<td>33.6</td>
<td>6.7</td>
<td>11.7</td>
<td>6.2</td>
<td>41.8</td>
<td>100</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>42.9</td>
<td>8.6</td>
<td>6.6</td>
<td>5.3</td>
<td>36.7</td>
<td>100</td>
</tr>
<tr>
<td>White</td>
<td>33.8</td>
<td>8.4</td>
<td>11.4</td>
<td>6.7</td>
<td>39.7</td>
<td>100</td>
</tr>
<tr>
<td>Race-Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black men</td>
<td>51.5</td>
<td>8.5</td>
<td>2.7</td>
<td>2.8</td>
<td>34.5</td>
<td>100</td>
</tr>
<tr>
<td>White men</td>
<td>34.5</td>
<td>11.4</td>
<td>10.5</td>
<td>7.9</td>
<td>35.7</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>Not present</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black women</td>
<td>35.1</td>
<td>8.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White women</td>
<td>33.2</td>
<td>6.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Community</th>
<th>Present</th>
<th>Not present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forsyth Co., NC</td>
<td>38.7</td>
<td>6.4</td>
</tr>
<tr>
<td>Jackson, MS</td>
<td>44.4</td>
<td>8.8</td>
</tr>
<tr>
<td>Minneapolis, MN</td>
<td>32.1</td>
<td>9.0</td>
</tr>
<tr>
<td>Washington Co. MD</td>
<td>31.7</td>
<td>10.3</td>
</tr>
</tbody>
</table>

**ICD-9-CM 428 code**

<table>
<thead>
<tr>
<th></th>
<th>Present</th>
<th>Not present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any listed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>38.5</td>
<td>38.0</td>
</tr>
<tr>
<td>Not present</td>
<td>16.8</td>
<td>46.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Present</th>
<th>Not present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary discharge code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>88.3</td>
<td>5.1</td>
</tr>
<tr>
<td>Not present</td>
<td>25.0</td>
<td>46.1</td>
</tr>
</tbody>
</table>

* Independent review and classification by two physician members of ARIC HF review committee with differences adjudicated by committee chair.
† Hospitalizations triaged during medical records review for not having sufficient evidence of HF to warrant complete abstraction or review by ARIC HF committee. The percentages shown in the table estimated from sampled events (n=1180) and account for sampling fractions.
Table 2. Various HF classification criteria and percent validated as HF by ARIC HF Classification Committee review.*

The Atherosclerosis Risk in Communities (ARIC) Study.

<table>
<thead>
<tr>
<th>Classification criteria</th>
<th>Acute decompensated</th>
<th>Chronic stable</th>
<th>No HF</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>68.4</td>
<td>9.6</td>
<td>21.9</td>
<td>100</td>
</tr>
<tr>
<td>HF not present</td>
<td>26.8</td>
<td>28.3</td>
<td>44.9</td>
<td>100</td>
</tr>
</tbody>
</table>

Percent

| All events reviewed        | 59.0 | 13.9 | 27.1 | 100   |

*Percent validated as HF by ARIC HF Classification Committee review.
<table>
<thead>
<tr>
<th></th>
<th>Present</th>
<th>Not present</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HF present</td>
<td>62.2</td>
<td>14.0</td>
<td>23.9</td>
<td>100</td>
</tr>
<tr>
<td>HF not present</td>
<td>48.9</td>
<td>13.6</td>
<td>37.5</td>
<td>100</td>
</tr>
<tr>
<td>Any listed ICD-9-CM 428 code</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>62.2</td>
<td>14.6</td>
<td>23.2</td>
<td>100</td>
</tr>
<tr>
<td>Not present</td>
<td>31.0</td>
<td>7.0</td>
<td>62.0</td>
<td>100</td>
</tr>
<tr>
<td>Primary discharge ICD-9-CM 428 code</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>93.0</td>
<td>4.1</td>
<td>2.9</td>
<td>100</td>
</tr>
<tr>
<td>Not present</td>
<td>46.3</td>
<td>17.5</td>
<td>36.2</td>
<td>100</td>
</tr>
</tbody>
</table>

* Among hospitalized events eligible for review by the ARIC HF Classification Committee. The percentages are weighted to account for the sampling probabilities (705 sampled events yielding a weighted number of 5011).

† HF present includes Gothenburg level 2 and 3; HF not present includes Gothenburg level 0,1,4,5.
### Table 3. Percent agreement (kappa coefficient) between HF classification criteria. The Atherosclerosis Risk in Communities (ARIC) Study. *

<table>
<thead>
<tr>
<th>Heart failure diagnostic classification criteria</th>
<th>Framingham</th>
<th>Modified</th>
<th>NHANES</th>
<th>Gothenburg</th>
<th>ICD-9-CM 428</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC † (acute decompensated HF)</td>
<td>69.5 (0.32)</td>
<td>63.7 (0.18)</td>
<td>60.9 (0.10)</td>
<td>59.5 (0.11)</td>
<td>62.9 (0.13)</td>
</tr>
<tr>
<td>ARIC ‡ (any HF)</td>
<td>70.5 (0.21)</td>
<td>68.1 (0.10)</td>
<td>67.4 (0.03)</td>
<td>66.9 (0.13)</td>
<td>75.3 (0.22)</td>
</tr>
<tr>
<td>Framingham</td>
<td>87.2 (0.61)</td>
<td>81.1 (0.38)</td>
<td>72.9 (0.24)</td>
<td>75.7 (0.14)</td>
<td></td>
</tr>
<tr>
<td>Modified Boston</td>
<td>89.5 (0.62)</td>
<td>74.2 (0.23)</td>
<td>77.4 (0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NHANES</td>
<td></td>
<td></td>
<td>73.3 (0.14)</td>
<td>79.3 (0.03)</td>
<td></td>
</tr>
<tr>
<td>Gothenburg</td>
<td></td>
<td></td>
<td></td>
<td>74.4 (0.12)</td>
<td></td>
</tr>
</tbody>
</table>

* Percent agreement (kappa coefficient) between HF classification criteria. The Atherosclerosis Risk in Communities (ARIC) Study.
* Among hospitalized events eligible for review by the ARIC HF Classification Committee. The data are weighted to account for the sampling probabilities (705 sampled events yielding a weighted number of 5011).

† combines cases classified as definite and possible decompensated HF together as HF = yes; chronic stable HF, or HF unlikely combined as heart failure = no

‡ combines cases classified as definite and possible decompensated HF, and chronic stable HF together as HF = yes; HF unlikely classified as HF = no
Table 4. Sensitivity, positive predictive value, specificity, and comparability ratio for various HF classification criteria according to ARIC HF classification criteria *. The Atherosclerosis Risk in Communities (ARIC) Study.

<table>
<thead>
<tr>
<th>Comparison Classifications</th>
<th>Sensitivity</th>
<th>Positive predictive value</th>
<th>Specificity</th>
<th>Comp. ratio</th>
<th>Sensitivity</th>
<th>Positive predictive value</th>
<th>Specificity</th>
<th>Comp. ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Framingham</td>
<td>0.90</td>
<td>0.68</td>
<td>0.40</td>
<td>1.31</td>
<td>0.83</td>
<td>0.78</td>
<td>0.37</td>
<td>1.06</td>
</tr>
<tr>
<td>Modified Boston</td>
<td>0.88</td>
<td>0.64</td>
<td>0.28</td>
<td>1.38</td>
<td>0.84</td>
<td>0.75</td>
<td>0.25</td>
<td>1.12</td>
</tr>
<tr>
<td>NHANES</td>
<td>0.90</td>
<td>0.62</td>
<td>0.19</td>
<td>1.46</td>
<td>0.87</td>
<td>0.73</td>
<td>0.16</td>
<td>1.18</td>
</tr>
<tr>
<td>Gothenburg</td>
<td>0.80</td>
<td>0.62</td>
<td>0.30</td>
<td>1.29</td>
<td>0.80</td>
<td>0.76</td>
<td>0.33</td>
<td>1.04</td>
</tr>
<tr>
<td>Any listed ICD-9-CM 428</td>
<td>0.95</td>
<td>0.62</td>
<td>0.17</td>
<td>1.52</td>
<td>0.95</td>
<td>0.77</td>
<td>0.23</td>
<td>1.23</td>
</tr>
<tr>
<td>Primary discharge ICD-9-CM 428 code</td>
<td>0.43</td>
<td>0.93</td>
<td>0.95</td>
<td>0.46</td>
<td>0.36</td>
<td>0.97</td>
<td>0.97</td>
<td>0.37</td>
</tr>
</tbody>
</table>
* Among hospitalized events eligible for review by the ARIC HF Classification Committee. Data shown in the table are weighted to account for the sampling probabilities (705 sampled events yielding a weighted number of 5011).

† Based on transformation of the all criteria into dichotomous groups (heart failure yes or no) as follows: Gothenburg criteria combines levels 2 and 3 together as heart failure = yes, otherwise heart failure = no; Modified Boston criteria combines definite and probable heart failure as heart failure = yes, otherwise heart failure = no; NHANES criteria class heart failure “present” defined as heart failure = yes and heart failure “not present” as heart failure = no; ICD-9-CM 428 defines heart failure = yes if a 428 code is present otherwise heart failure = no; Primary Discharge ICD-9-CM defines heart failure = yes if a 428 code is given as the primary discharge code otherwise heart failure = no.

‡ Combines definite and possible decompensated HF together as HF = yes. Classes chronic stable, or HF unlikely combined as HF = no

** Combines cases classified as definite and possible decompensated HF, and chronic stable HF together HF = yes and HF unlikely as HF = no.
Figure Legend

Figure. Overlap between Framingham criteria, ARIC HF Classification Committee review, and ICD-9-CM 428 code as primary discharge code for the classification of a HF hospitalization.
### Table 1. Eligible ICD-9-CM diagnosis codes and sampling probabilities for HF related hospital discharges, The ARIC Study.

<table>
<thead>
<tr>
<th>ICD-9-CM code</th>
<th>Forsyth County, NC Men</th>
<th>Forsyth County, NC Women</th>
<th>Jackson, MS Men</th>
<th>Jackson, MS Women</th>
<th>Minneapolis, MN Men</th>
<th>Minneapolis, MN Women</th>
<th>Washington County, MD Men</th>
<th>Washington County, MD Women</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>428 (55-84 years)</td>
<td>0.26 0.10</td>
<td>0.23 0.10</td>
<td>0.10 0.26</td>
<td>0.16 0.07</td>
<td>0.20 0.07</td>
<td>0.10 0.10</td>
<td>Heart Failure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>428 (&gt; 85 years)</td>
<td>0.23 0.10</td>
<td>0.19 0.16</td>
<td>0.06 0.23</td>
<td>0.10 0.06</td>
<td>0.19 0.03</td>
<td>0.10 0.06</td>
<td>Heart Failure</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (&gt; 55 years)</td>
<td>0.74 0.36</td>
<td>0.30 0.36</td>
<td>0.61 1.0</td>
<td>0.42 0.81</td>
<td>0.16 0.16</td>
<td>0.70 0.58</td>
<td>See footnote</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ICD-9-CM, *International Classification of Disease, Ninth Revision, Clinical Modification*

* Other diagnosis groups (*ICD-9-CM code*) included 398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.13, 404.91, 404.93, 415.0, 416.9, 425.4, 518.4, 786.0.
Table 2. Framingham, Modified Boston, National Health and Nutrition Examination Survey (NHANES), Gothenburg, and ARIC Criteria for classifying heart failure.

<table>
<thead>
<tr>
<th>Criteria name</th>
<th>Classification Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Framingham Criteria (^{(1)})</td>
<td>Heart failure classification based on evidence of major or minor symptoms. Heart failure is considered present if 2 major or 1 major plus 2 minor criteria are met.</td>
</tr>
<tr>
<td></td>
<td><strong>Major:</strong> Paroxysmal nocturnal dyspnea or orthopnea, neck vein distension, rales, cardiomegaly, acute pulmonary edema, S3 gallop, increase venous pressure (\geq 16\ \text{cm}\ \text{H}_2\text{O}), circulation time (\geq 25\ \text{seconds}), hepatojugular reflux.</td>
</tr>
<tr>
<td></td>
<td><strong>Minor:</strong> ankle edema, night cough, dyspnea on exertion, hepatomegaly, pleural effusion, vital capacity decreased one third from maximum, tachycardial rate (\geq 120/\text{min}). Weight loss (\geq 4.5\ \text{kg}) in 5 days in response to treatment, major criterion if weight loss occurred during therapy, otherwise minor.</td>
</tr>
<tr>
<td>Modified Boston (^{(2)})</td>
<td>A points system where heart failure is considered definite (8-12 points), possible (5-7 points), or unlikely (&lt; 5 points).</td>
</tr>
<tr>
<td>Category I: Medical history</td>
<td>No dyspnea (0 pts), leg fatigue on walking on level (1 pt), dyspnea walking on level (2 pts), paroxysmal nocturnal dyspnea (3 pts), orthopnea (4 pts), dyspnea at rest (4 pts).</td>
</tr>
<tr>
<td>Category II: Physical findings:</td>
<td>Heart rate &lt; 90 (0 pts), 91-110 (1 pt), &gt; 110 (2 pts)</td>
</tr>
<tr>
<td></td>
<td>Jugular venous pressure: &lt; 6 mm H(_2)O (0 pts), &gt; 6 mm H(_2)O (2 pts), &gt; 6 mm H(_2)O plus liver enlargement or pitting edema (3 pts)</td>
</tr>
<tr>
<td></td>
<td>Pulmonary rales: No (0 pts), at the bases only (1pt), more than basilar (2 pts)</td>
</tr>
<tr>
<td></td>
<td>Wheezes: No (0 pts), yes (3 pts)</td>
</tr>
<tr>
<td></td>
<td>S3 gallop: No (0 pts), yes (3 pts)</td>
</tr>
<tr>
<td>Category III: Chest X-ray evidence</td>
<td>Chest X-ray - normal (0 pts), upper flow redistribution (2 pts), cardiac enlargement (relative heart volume&gt;540 ml.m(^{-2}) in men and &gt; 490 ml.m(^{-2}) in women) (3 pt), interstitial pulmonary edema (3 pts), bilateral pleural effusion (3 pts), alveolar pulmonary edema (4 pts)</td>
</tr>
</tbody>
</table>
No more than 4 points allowed for each of three categories

NHANES (3) A points system where heart failure is considered present with a total score greater than or equal to 3.

History:
- Shortness of breath when hurrying on the level or up slight hill (1 pt), shortness of breath when walking at ordinary pace on the level (1 pt), stops for breath when walking at own pace (2 pts), stops for breath after 100 yards on the level (2 pts)

Physical exam:
- Heart rate 91-110 (1 pt), > 110 (2 pts), basal rales (1 pt), > basal rates (2 pts), neck vein distension (1 pt), neck vein distention and edema or hepatomegaly (2 pts)

Chest x-ray:
- Cephalization of pulmonary veins (1 pt), interstitial edema (2 pts), alveolar fluid and pleural fluid (3 pts), interstitial edema and pleural fluid (3 pts)

Gothenburg Criteria (4) A points system that assigns heart failure grades depending on medical history, physical findings, and drug treatment. The following grades are assigned: Grade of 0 (HF absent) if all 3 scores are 0; Grade 1 (latent) if cardiac score > 0 and pulmonary and therapy score = 0; Grade 2 (manifest heart failure) if cardiac score > 0 and either pulmonary or therapy score > 0; Grade 3 if cardiac score > 0 and both pulmonary and therapy score > 0; and Grade 4 if the person died in heart failure.

Cardiac score:
- Coronary heart disease present in past (1 pt), present within last year (2 pts); angina pectoris present in the past (1 pt), present within last year (2 pts); swollen legs at end of day (1 pt); pulmonary rales at physical exam (1 pt); atrial fibrillation on ECG (1 pt). Note: heart disease and angina can only contribute 2 points together.

Pulmonary disease score:
- History of chronic bronchitis (1 pt), history of chronic bronchitis within last year (2 pts); history of asthma (1 pt), history of asthma within last year (2 pts); history of coughing, phlegm or wheezing (1 pt), presence of rhonchi at physical examination (1 pt).

Therapy score:
- History of digitalis administration (1 pt), history of diuretic administration (1 pt).
A physician reviewer based system aided by comprehensive abstraction of medical records. Abstracted items include: evidence of new onset of symptoms, history of heart failure, general medical history, physical exam signs and symptoms, chest X-ray findings, measures of ejection fraction, biomarkers (brain natriuretic peptide (BNP), N-terminal pro-hormone brain natriuretic peptide (pro-BNP)), and medications.

Reviewers also provided event’s discharge summary, history and physical report, admission note, and imaging reports. Hospitalized events classified into one of five categories: definite acute decompensated heart failure, possible acute decompensated heart failure, chronic stable heart failure, heart failure unlikely, or unclassifiable.

Acute decompensated heart failure:
Clear evidence either from symptoms, signs, imaging or treatment of an acute exacerbation, worsening or new onset of symptoms or other decompensated circulatory state.

Evidence of augmentation of therapy for worsening heart failure signs or symptoms; in-hospital control of symptoms by therapy; specificity of heart failure for decompensated state as opposed to other co-morbidities (e.g. chronic obstructive pulmonary disease, end-stage renal disease).

A case was considered possible acute decompensated heart failure if the presence of co-morbidity could also account for the acute symptoms or if there was not enough information to classify as definite.

Chronic heart failure:
Evidence of compensated heart failure signs and symptoms controlled by therapy with no evidence of therapy augmentation or symptom worsening during the hospitalization.

Evidence of left ventricular systolic dysfunction (ejection fraction < 50%) with no HF symptoms was sufficient for classification. Asymptomatic diastolic dysfunction was not sufficient for a classification as chronic stable heart failure.

Heart failure unlikely and unclassifiable:
Documentation directly or indirectly that heart function was normal.
Hospitalizations with Insufficient data or missing medical records designated as unclassifiable.
Figure 1. Sampling of heart failure hospitalizations and relation to analyses presented

Supplemental Figure 1. Sampling of heart failure hospitalizations and relation to analyses presented.

Total HF hospitalizations* (n=11,544)

Random sample investigated (n=1499)

Exclusions
- Medical record not found (n=16)
- Patient address out of catchment area (n=303)

Reviewed for any evidence of relevant HF symptoms? † (n=1180)

Not present
- Does not meet screening criteria, auto-classified as no heart failure (n=475)

Present

Data used for Table 2-4 and Figure 1 comparing heart failure multiple criteria (n=705, n= 5011 weighted)

Full abstraction of medical record and physician review (n=705)

Data used for Table 1 showing ARIC classification by demographic characteristics (n=1180, n=8205 weighted)

* From electronic discharge lists from hospitals serving the surveillance communities for patient 55 years or older with diagnoses suggestive of heart failure: ICD-9-CM codes 398.91, 402, 404, 415, 416.9, 425.4, 428, 518.4, 786.0

† Medical record reviewed for any evidence of increasing or new onset of shortness of breath, edema, paroxysmal nocturnal dyspnea, orthopnea, hypoxia, or mention of heart failure in doctor's notes.
Supplemental Methods - References


