NIH-PA Author Manuscript

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

**Wayne D. Rosamond, PhD, MS**[1], **Patricia P. Chang, MD**[2], **Chris Baggett, PhD**[3], **Anna Johnson, PhD**[1], **Alain G. Bertoni, MD**[4], **Eyal Shahar, MD**[5], **Anita Deswal, MD, MPH**[6], **Gerardo Heiss, MD, PhD**[1], and **Lloyd E. Chambless, PhD**[3]

[1]Department of Epidemiology, University of North Carolina, Chapel Hill, NC

[2]Department of Medicine, University of North Carolina, Chapel Hill, NC

[3]Department of Biostatistics, University of North Carolina, Chapel Hill, NC

[4]Department of Epidemiology and Prevention, Wake Forest University School of Medicine, Winston-Salem, NC

[5]Division of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ

[6]Section of Cardiology, Michael E. DeBakey VA Medical Center & Baylor College of Medicine, Houston, TX

## 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.

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, wayne_rosamond@unc.edu.

**Disclosures**
None.

### Keywords

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 [1–2]. Signs and symptoms may differ depending on the level of systolic or diastolic dysfunction and further complicate disease classification [3–4]. Various diagnostic criteria are published, and comparisons between these criteria report mixed results [5–14]. 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 [15–18]. 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 were 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 [19] and chance-corrected agreement by kappa coefficients [20]. 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[9, 15–17, 21–26]. 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[27]. 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%)[4, 28]. 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 [29].

Remes and colleague reported on cases of clinically suspected HF[30] 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 [13]. 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[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.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

1. Jessup M. Heart failure. N Engl J Med. 2003; 348:2007–2018. [PubMed: 12748317]

2. Mudd JO, Kass DA. Tackling heart failure in the twenty-first century. Nature. 2008; 451:919–928. [PubMed: 18288181]

3. Yturralde FR, Gaasch WH. Diagnostic criteria for diastolic heart failure. Prog Cardiovascular Dis. 2005; 47:314–319.

4. Maestre A, Gil V, Gallego J, Aznar J, Mora A, Martin-Hidalgo A. Diagnostic accuracy of clinical criteria for identifying systolic and diastolic heart failure: cross-sectional study. J Eval Clin Pract. 2009; 15:55–61. [PubMed: 19239582]

5. Ho KK, Anderson KM, Kannel WB, Grossman W, Levy D. Survival after the onset of congestive heart failure in Framingham Heart study subjects. Circulation. 1993; 88:107–115. [PubMed: 8319323]

6. Carlson KJ, Lee DC, Goroll AH, Leahy M, Johnson RA. An analysis of physicians' reasons for presecribing long-term digitalis therapy in outpatients. J Chron Dis. 1985; 38:733–739. [PubMed: 4030999]

7. Eriksson H, Caidahl K, Larsson B, Ohlson L, Welin L, Wilhelmsen L, Svardsudd K. Cardiac and pulmonary causes of dyspnoea-validation of a scoring test for clinical-epidemiological use: The study of men born in 1913. Eur Heart J. 1987; 8:1007–1014. [PubMed: 3665952]

8. Schocken DD, MIA, Leaverton PE, Ross EA. Prevalence and mortality rate of congestive heart failure in the United States. J Am Coll Cardiol. 1992; 20:301–306. [PubMed: 1634664]

9. Alqaisi F, Williams LK, Peterson EL, Lanfear DE. Comparing methods for identifying patients with heart failure using electronic data sources. BMC Health Serv Res. 2009; 9:237–241. [PubMed: 20021648]

10. Clerico A, Fontana M, Zyw L, Passino C, Emdin M. Comparison of the diagnostic accuracy of brain natriurectic peptide (BNP) and the N-terminal part of the propeptide of BNP immunoassays in chronic and acute heart failure: A systematic review. Clin Chemisty. 2007; 53:813–822.

11. Di Bari MPC, Cavaillini MC, Innocenti F, Baldereschi G, De Alfieri W, Antonini E, Pini R, Masotti G, Marchionni N. The diagnosis of heart failure in the community-Comparative validation of four sets of criteria in unselected older adults: The ICARe Dicomano Study. J Am Coll Cardiol. 2004; 44:1601–1608. [PubMed: 15489092]

12. Kim J, Jacobs DR, Luepker RV, Shahar E, Margolis KL, Becher MP. Prognostic value of a novel classification scheme for heart failure: The Minnesota Heart Failure Criteria. Am J Epidemiol. 2006; 164:184–193. [PubMed: 16707656]

13. Mosterd A, Deckers JW, Hoes AW, Nederpel A, Smeets A, Linker DT, Grobbee DE. Classification of heart failure in population based research: An assessment of six heart failure scores. Eur J Epidemiol. 1997; 13:491–502. [PubMed: 9258559]

14. Zannad F, Stough WG, Pitt B, Cleland JG, Adams KF, Geller NL, Torp-Pedersen C, Kirwan BA, Follath F. Heart failure as an endpoint in heart failure and non-heart failure cardiovascular clinical trials: the need for a consensus definition. Eur Heart J. 2008; 29:413–421. [PubMed: 18245122]

15. Goldberg R. Assessing the population burden from heart failure: Need for sentinel population-based surveillance systems. Arch Intern Med. 1999; 159:15–17. [PubMed: 9892324]

16. Goldberg RJ, Spencer FA, Farmer C, Meyer T, Pezzella S. Incidence and hospital death rate associated with heart failure: A community-wide perspective. Am J Med. 2005; 118:728–734. [PubMed: 15989906]

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 population. JAMA. 2004; 292:344–350. [PubMed: 15265849]

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 Engl J Med. 2002; 347:1397–1402. [PubMed: 12409541]

19. Kelsey, J. Methods in Observational Epidemiology. New York: Oxford University 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 congestive heart failure: The Framingham Study. N Engl J Med. 1971; 285:1441–1446. [PubMed: 5122894]

22. Kannel WB, Belanger AJ. Epidemiology of heart failure. Am Heart J. 1991; 121:951–957. [PubMed: 2000773]

23. Ho KK, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. J Am Coll Cardiol. 1993; 22:6A–13A. [PubMed: 8509564]

24. The Task Force on Heart Failure of the European Society of Cardiology. Guidelines for the diagnosis of heart failure. Eur Heart J. 1995; 16:741–751. [PubMed: 7588917]

25. Cowie MR, Wood DA, Coats AJ, Thompson SG, Poole-Wilson PA, Suresh V, Sutton GC. Incidence and aetiology of heart failure: A population-based study. Eur Heart J. 1999; 20:421–428. [PubMed: 10213345]

26. Eriksson H, Svarsudd K, Larsson B, Ohlson LO, Tibblin G, Welin L, Wilhelmsen L. Risk factors for heart failure in the general population: the study of men born in 1913. Eur Heart J. 1989; 10:647–656. [PubMed: 2788575]

27. Fonseca C, Oliveira AG, Mota T, Matias F, Morais H, Costa C, Ceia F. Evaluation of the performance and concordance of clinical questionnaires for the diagnosis of heart failure in primary care. Eur J Heart Fail. 2004; 6:813–820. [PubMed: 15542422]

28. Jimeno Sainz A, Gil V, Merino J, García M, Jordán A, Guerrero L. Validity of Framingham criteria as a clinical test for systolic heart failure. Rev Clin Esp. 2006; 206:495–498. [PubMed: 17203567]

29. Marantz PR, Tobin JN, Wassertheil-Smoller S, Steingart RM, Wexler JP, Budner N, Lense L, Wachspress J. The relationship between left ventricular systolic function and congestive heart failure diagnosed by clinical criteria. Circulation. 1998; 77:607–612. [PubMed: 3342491]

30. Remes J, Reunanen A, Aromaa A, Pyorala K. Incidence of heart failure in eastern Finland: a population-based surveillance study. Eur Heart J. 1992; 13:588–593. [PubMed: 1618198]

31. Schellenbaum GD, Rea T, Heckbert SR, Smith NL, Lumley T, Roger VL, Kitzman DW, Taylor HA, Levy D, Psaty BM. Survival associated with two sets of diagnostic criteria for congestive heart failure. Am J Epidemiol. 2004; 160:628–635. [PubMed: 15383406]

32. Morrison A. The effects of early treatment, lead time and length bias on the mortality experienced by cases detected by screening. Int J Epidemiol. 1982; 11:261–267. [PubMed: 7129740]

33. Dieplinger B, Gegenhuber A, Haltmayer M, Meuller T. Evaluation of novel biomarkers for the diagnosis of acute destabilised heart failure in patients with shortness of breath. Heart. 2009; 95:1508–1513. [PubMed: 19525245]

34. Mant J, Doust J, Roalfe A, Barton P, Cowie MR, Glasziou P, Mant D, McManus RJ, Holder R, Deeks J, Fletcher K, Qume M, Sohanpal S, Sanders S, Hobbs FD. Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care. Health Technol Assess. 2009; 13:1–207. iii. [PubMed: 19586584]

35. Allen L. Use of multiple biomarkers in heart failure. Curr Cardiol Rep. 2010; 12:230–236. [PubMed: 20424966]

36. Lee DS, Gona P, Vasan RS, Larson MG, Benjamin EJ, Wang TJ, Tu JV, Levy D. Relation of Disease Pathogenesis and Risk Factors to Heart Failure With Preserved or Reduced Ejection Fraction Insights From the Framingham Heart Study of the National Heart, Lung, and Blood Institute. Circulation. 2009; 119:3070–3077. [PubMed: 19506115]

37. Bibbins-Domingo K, Pletcher MJ, Lin F, Vittinghoff E, Gardin JM, Arynchyn A, Lewis CE, Williams OD, Hulley SB. Racial differences in incident heart failure among young adults. N Engl J Med. 2009; 360:1179–1190. [PubMed: 19297571]

38. Nieminen MS, Harjola VP, Hochadel M, Drexler H, Komajda M, Brutsaert D, Dickstein K, Ponikowski P, Tavazzi L, Follath F, Lopez-Sendon JL. Gender relatd differences in patients presenting with acute heart failure. Results from EuroHeart Failure Survey II. Eur J Heart Fail. 2008; 10:140–148. [PubMed: 18279769]

39. Krumholz HF, Merrill AR, Schone EM, Schreiner GC, Chen J, Bradley EH, Wang Y, Lin Z, Straube BM, Rapp M, Normand Sl, Drye E. Patterns of hospital performance in acute myocardial infarction and heart failure 30-day mortality and readmission. Circ Cardiovasc Qual Outcomes. 2009; 2:407–413. [PubMed: 20031870]

40. Rathore SS, Foody JM, Wang Y, Smith GL, Herrin J, Masoudi FA, Wolfe P, Havranek EP, Ordin DL, Krumholz HM. Race, quality of care, and outcomes of elderly patients hospitalized with heart failure. JAMA. 2003; 289:2517–2524. [PubMed: 12759323]
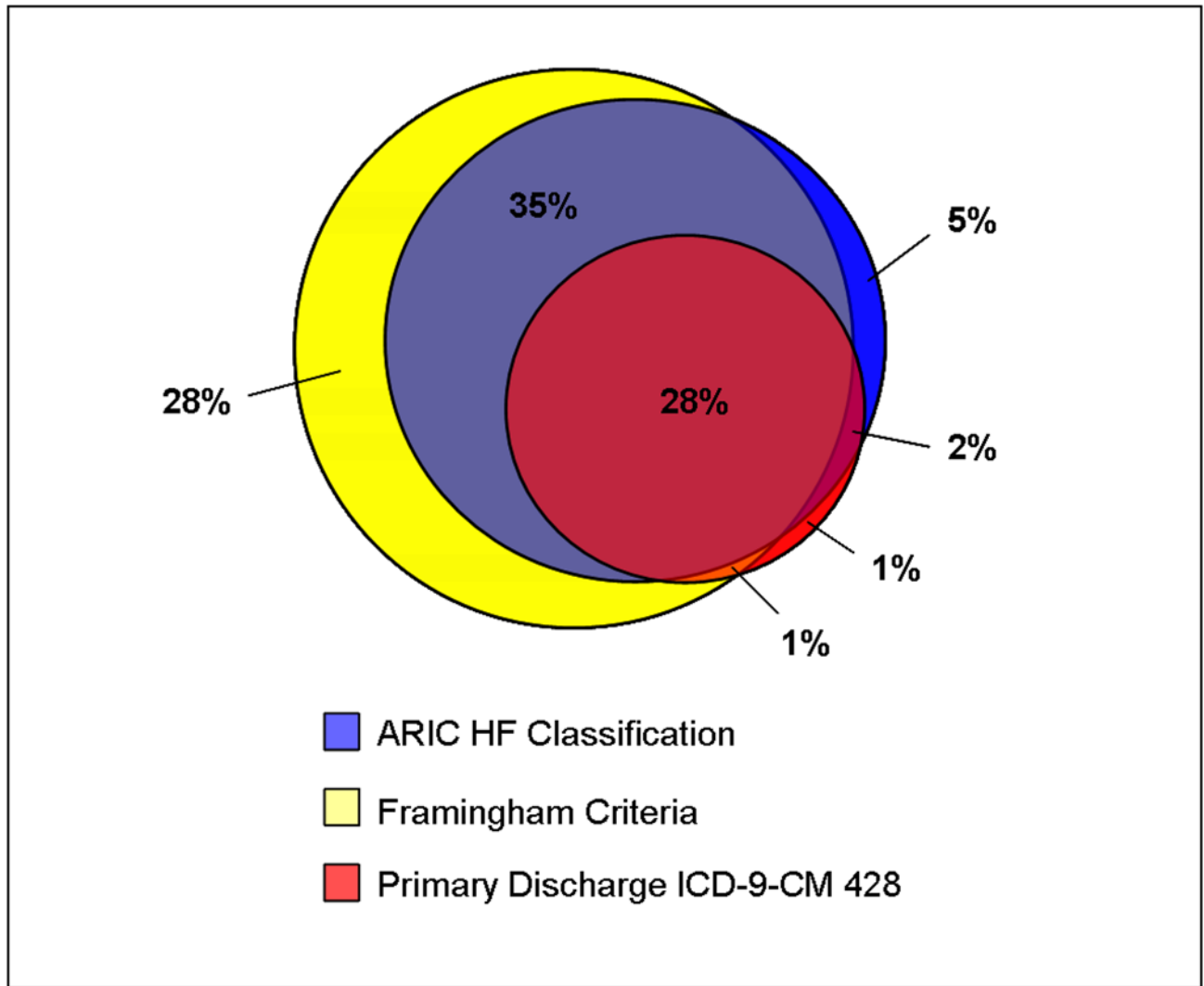
**Figure 1.**
Overlap between Framingham criteria, ARIC HF Classification Committee review, and ICD-9-CM 428 code as primary discharge code for the classification of a HF hospitalization.

**Table 1**

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

| | Heart Failure classification by ARIC Classification Committee* | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | Acute decompensated heart failure | Chronic stable heart failure | No heart failure | Unclassifiable | Did not meet screening criteria† | Total | p-value |
| | | | Percent | | | | |
| All | 36.0 | 8.5 | 10.2 | 6.3 | 38.9 | 100 | |
| **Sex** | | | | | | | |
| Men | 39.0 | 10.7 | 8.4 | 6.6 | 35.4 | 100 | 0.016 |
| Women | 33.6 | 6.7 | 11.7 | 6.2 | 41.8 | 100 | |
| **Race** | | | | | | | |
| Black | 42.9 | 8.6 | 6.6 | 5.3 | 36.7 | 100 | 0.019 |
| White | 33.8 | 8.4 | 11.4 | 6.7 | 39.7 | 100 | |
| **Race-Sex** | | | | | | | |
| Black men | 51.5 | 8.5 | 2.7 | 2.8 | 34.5 | 100 | <0.0001 |
| White men | 34.5 | 11.4 | 10.5 | 7.9 | 35.7 | 100 | |
| Black women | 35.1 | 8.6 | 10.1 | 7.6 | 38.6 | 100 | |
| White women | 33.2 | 6.1 | 12.2 | 5.7 | 42.8 | 100 | |
| **Community** | | | | | | | |
| Forsyth Co., NC | 38.7 | 6.4 | 7.2 | 3.9 | 43.7 | 100 | 0.083 |
| Jackson, MS | 44.4 | 8.8 | 5.0 | 6.9 | 35.0 | 100 | |
| Minneapolis, MN | 32.1 | 9.0 | 6.5 | 7.2 | 45.2 | 100 | |
| Washington Co. MD | 31.7 | 10.3 | 20.2 | 8.3 | 29.5 | 100 | |
| *ICD-9-CM* 428 code | | | | | | | |
| Any listed | | | | | | | |
| Present | 38.5 | 9.1 | 7.9 | 6.5 | 38.0 | 100 | <0.0001 |
| Not present | 16.8 | 3.8 | 28.0 | 5.4 | 46.1 | 100 | |

**Heart Failure classification by ARIC Classification Committee[*]**

| | Acute decompensated heart failure | Chronic stable heart failure | No heart failure | Unclassifiable | Did not meet screening criteria[†] | Total | p-value |
|---|---|---|---|---|---|---|---|
| | | | **Percent** | | | | |
| Primary discharge code | | | | | | | |
| Present | 88.3 | 3.9 | 1.4 | 1.3 | 5.1 | 100 | <0.0001 |
| Not present | 25.0 | 9.4 | 12.1 | 7.4 | 46.1 | 100 | |

[*] 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.

| | Heart Failure Classification by ARIC Classification Committee | | | |
| | Acute decompensated HF | Chronic stable HF | No HF | Total |
| | **Percent** | | | |
| **All events reviewed** | 59.0 | 13.9 | 27.1 | 100 |
| **Classification criteria** | | | | |
| Framingham | | | | |
| HF present | 68.4 | 9.6 | 21.9 | 100 |
| HF not present | 26.8 | 28.3 | 44.9 | 100 |
| Modified Boston | | | | |
| Definite or probable HF | 63.9 | 11.2 | 24.9 | 100 |
| HF unlikely | 37.2 | 25.7 | 37.0 | 100 |
| NHANES | | | | |
| HF present | 61.6 | 11.8 | 26.6 | 100 |
| HF not present | 43.2 | 26.4 | 30.4 | 100 |
| Gothenburg† | | | | |
| HF present | 62.2 | 14.0 | 23.9 | 100 |
| HF not present | 48.9 | 13.6 | 37.5 | 100 |
| Any listed *ICD-9-CM* 428 code | | | | |
| Present | 62.2 | 14.6 | 23.2 | 100 |
| Not present | 31.0 | 7.0 | 62.0 | 100 |
| Primary discharge *ICD-9-CM* 428 code | | | | |
| Present | 93.0 | 4.1 | 2.9 | 100 |
| Not present | 46.3 | 17.5 | 36.2 | 100 |

*
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.[*]

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

[*]
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.

| Comparison Classifications | ARIC Acute decompensated HF[‡] | | | | ARIC Acute decompensated HF + chronic stable HF[**] | | | |
|---|---|---|---|---|---|---|---|---|
| | Sensitivity | Positive predictive value | Specificity | Comp. ratio | Sensitivity | Positive predictive value | Specificity | Comp. ratio |
| **Classification criteria**[†] | | | | | | | | |
| Framingham | 0.90 | 0.68 | 0.40 | 1.31 | 0.83 | 0.78 | 0.37 | 1.06 |
| Modified Boston | 0.88 | 0.64 | 0.28 | 1.38 | 0.84 | 0.75 | 0.25 | 1.12 |
| NHANES | 0.90 | 0.62 | 0.19 | 1.46 | 0.87 | 0.73 | 0.16 | 1.18 |
| Gothenburg | 0.80 | 0.62 | 0.30 | 1.29 | 0.80 | 0.76 | 0.33 | 1.04 |
| Any listed *ICD-9-CM* 428 code | 0.95 | 0.62 | 0.17 | 1.52 | 0.95 | 0.77 | 0.23 | 1.23 |
| Primary discharge *ICD-9-CM* 428 code | 0.43 | 0.93 | 0.95 | 0.46 | 0.36 | 0.97 | 0.97 | 0.37 |

[*]
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.