

The “gray zone” for the heart to mediastinum MIBG uptake ratio

Mario Petretta, MD,^a Teresa Pellegrino, MD, PhD,^b and Alberto Cuocolo, MD^c

^a Department of Translational Medicine, University Federico II, Naples, Italy

^b Institute of Biostructure and Bioimaging, National Council of Research, Naples, Italy

^c Department of Advanced Biomedical Sciences, University Federico II, Naples, Italy

Received Mar 14, 2014; accepted Mar 15, 2014

doi:10.1007/s12350-014-9894-4

See related article, pp. 913–920

Increased myocardial sympathetic activity is an important feature of heart failure (HF) and is associated with progressive myocardial remodeling, worsening in left ventricular (LV) function, and poor outcome.¹⁻³ The AdreView Myocardial Imaging for Risk Evaluation in Heart Failure (ADMIRE-HF) study prospectively evaluated the value of quantitation of sympathetic innervation of the myocardium, measured by the heart to mediastinum (H/M) uptake ratio on iodine-123-metaiodobenzylguanidine (MIBG) scintigraphy, for predicting prognosis in subjects with HF and significant LV dysfunction.⁴ The study showed a highly significant relationship between time to HF-related events and the H/M ratio, which was independent of LV ejection fraction and B-type natriuretic peptide. The authors used a protocol-defined binary division of patients (with late H/M ratio ≥ 1.6 and with late H/M ratio < 1.6).⁴ This choice was based on published data on 182 control subjects with a mean late H/M ratio of 2.2 ± 0.3 .⁵ Thus, assuming a normal distribution of H/M ratio, 95% of controls will have a late H/M ratio between 1.6 and 2.8 (i.e., mean ± 2 standard deviation) so that late H/M ratio values < 1.6 should be considered abnormally low. Using this cutoff, the risk of the primary end point was significantly lower for subjects with late H/M ratio ≥ 1.60 , with a hazard ratio of 0.40 (97.5% confidence interval 0.25-0.64; $P < .001$).

Continuous variables, such as age, weight, blood pressure, and many others, including laboratory and imaging data, are commonly measured in clinical trials. Such measurements are often used for diagnostic purpose, for risk stratification, or to select patient management strategies. Appealing, continuous variables are often converted into binary or categorical variables by grouping values into two or more categories. This approach avoids to assuming a linear effect of the variable on outcome, but categorization discards information and raises several critical issues.^{6,7} The cutoff of continuous variables is also selected from receiver-operating characteristic (ROC) analysis.⁸ The “ideal” cutoff value is almost always a trade-off between sensitivity (true positives) and specificity (true negatives). As both change with each “cutoff” value it becomes difficult for the reader to imagine which cut-off is ideal. The ROC curve offers a graphical illustration of these trade-offs at each “cutoff” for any diagnostic test that uses a continuous variable. Ideally, the best “cutoff” value provides both the highest sensitivity and the highest specificity, easily located on the ROC curve by finding the highest point on the vertical axis and the furthest to the left on the horizontal axis (upper left corner). However, it is rare that this ideal value can be achieved, so that, for example, one may opt to choose a higher sensitivity at the cost of lower specificity. Youden index is a frequently used summary measure of the ROC curve, and a simple formula is: sensitivity + specificity - 1; however, other estimation procedures of this index exist.⁹ Probably there is not a “best cutoff” for any prediction model, and it depends on the needs of the diagnostic or prognostic stratification. In fact, sensitivity may be more important than specificity (but how much?). Nevertheless, one may need to have at least a threshold level of specificity, and so on.^{10,11}

Nakata et al¹² recently performed a patient-level analysis of six prospective multicenter cohort studies of

Reprint requests: Alberto Cuocolo, MD, Department of Advanced Biomedical Sciences, University Federico II, Via Pansini 5, 80131 Naples, Italy; cuocolo@umina.it

J Nucl Cardiol 2014;21:921–4.

1071-3581/\$34.00

Copyright © 2014 American Society of Nuclear Cardiology.

MIBG imaging of sympathetic innervation for assessment of long-term prognosis in HF. The pooled database consisted of 1,322 patients with HF followed up for a mean of 78 months. Lethal events were observed in 326 patients, with a 5-year mortality rate of 19.7%. Multivariate Cox proportional hazard model analysis for all-cause mortality identified age, New York Heart Association functional class, late H/M ratio, and LV ejection fraction significant independent predictors. ROC analysis was performed to determine the optimal cutoff value of independent significant parameters. This analysis identified 1.68 as optimal thresholds of late H/M ratio for dichotomizing the population into higher- and lower-risk patients for lethal outcomes. The threshold of late H/M ratio <1.68 identified patients at significantly increased risk in any LV ejection fraction category. Survival rates decreased progressively with decreasing late H/M ratio, with 5-year all-cause mortality rates >7% annually for late H/M ratio <1.25, and <2% annually for late H/M ratio \geq 1.95. Likewise, late H/M ratio differentiated the high- from the low-risk patients both for sudden cardiac death and for pump failure death at 5 years. Thus, the results of this pooled analysis suggest a threshold higher than that of ADMIRE,⁴ despite this latter study was included in the analysis. The study of Nakata et al¹² also demonstrates that, although it has been reported that late H/M ratio has a threshold for identifying patients at increased risk for fatal outcomes, the patient survival rate decreases linearly depending on impaired cardiac MIBG activity.^{4,13}

Another important problem in dichotomizing a continuous variable is related to the measurement error. Reproducibility is particularly relevant when repeated testing is used to evaluate changes induced by disease progression and/or by interventions.^{14,15} Intra- and inter-observer reproducibility and test-retest reliability may be evaluated with several statistical methods,¹⁶⁻¹⁸ but due to both biological variability and measurement uncertainty, all have weakness and limitations.^{19,20} An ongoing study was designed to assess the reproducibility of quantitative measurements of myocardial uptake of MIBG on planar and SPECT imaging.²¹ Efficacy will be assessed based upon the absolute differences between quantitative analyses of imaging data on two scans performed 5-14 days apart. The estimated study completion date is June 2014.

Matsui et al²² found that the H/M ratio on planar MIBG imaging before and after optimized treatment provides independent prognostic information in patients with dilated cardiomyopathy. Kasama et al²³ reported that serial MIBG studies can be useful for predicting cardiac death and sudden death in stabilized patients with HF. Drakos et al²⁴ found that clinical, functional, and hemodynamic improvements induced by ventricular

unloading are accompanied by enhancements in sympathetic innervation in the failing heart. Yet, in these studies the reproducibility of MIBG cardiac imaging was not addressed. Veltman et al²⁵ showed a high reproducibility of planar MIBG myocardial scintigraphy in HF patients. However, in that study regional myocardial sympathetic activity on MIBG SPECT images was not assessed. We recently demonstrated a high observer reproducibility of planar H/M ratios and SPECT defect scores using a low-dose MIBG cardiac imaging protocol in 74 patients with HF.²⁶ The intra-class coefficient of correlation, Lin's concordance correlation coefficient, and Bland-Altman analysis were used to evaluate the intra- and the inter-observer reproducibility.²⁷ With these approaches, the differences between measurements obtained twice by the same examiner and by two examiners were negligible for both early and late H/M ratios and for SPECT defect scores. We also used the κ statistic to evaluate the concordance rates for the identification of patients with a low H/M ratio (<1.60) on late planar imaging and the κ values were 0.90 for intra-observer concordance and 0.83 for inter-observer concordance. In patients with idiopathic dilated cardiomyopathy, κ was 0.91 for both intra- and inter-observer concordance, higher than in patients with ischemic HF ($\kappa = 0.81$ and 0.73 for intra-observer and inter-observer concordance, respectively).

We also evaluated the classification certainty of observer reproducibility of planar late H/M ratio in patients with HF using the standard formula utilized by Petraco et al²⁸ to assess the test-retest agreement of fractional flow reserve. We found that within the region of H/M values from 1.54 to 1.66 the agreement between paired intra-observer measurements falls below 80% (Figure 1A), reaching a nadir of approximately 50% around 1.60, the proposed clinical cutoff. Thus, when a single late H/M ratio measurement falls between 1.54 and 1.66, there is a substantial chance that the dichotomous classification will change if the measurement is repeated by the same observer. This measurement uncertainty creates a "measurement gray zone" that for inter-observer variability is slightly larger (1.52-1.68) (Figure 1B). These results might have several important implications both for interpretation of available MIBG studies and for application in patient's clinical decision making. Our analysis outlines the potential limitations of a dichotomous interpretation of late H/M ratio results. In fact, similarly to all other measurements in medicine, late H/M ratio does not carry strict dichotomous implications for which prognosis (or treatment) is best, and this is especially true close to the cutoff. Noteworthy, for observer reproducibility we can think of the observations for the same subject as a series of measurements of a quantity that does not vary over the period of

observation. On the other hand, when the quantity being measured is physiologically unstable or may vary under different conditions, as in a test-retest study, the classification certainty could be lower and the probability of misclassification higher than that we found in our observer reproducibility analysis of late H/M ratio.

Action items While a low H/M ratio on planar cardiac MIBG imaging is associated with increased risk for adverse outcomes in HF patients, the significance of MIBG SPECT findings is less clear. Moreover, it is not

clear whether predictors of sudden death are comparable to those predicting death of any cause or cardiac death. Sudden cardiac death might be related to the border zone between scar and normal myocardium and to regional denervation. In the current issue of the journal, Zhou et al²⁹ proposed a new quantitative tool to measure myocardial scar and border zone and to quantify MIBG uptakes in border zone. Their preliminary data suggest that MIBG uptake in the border zone might predict ventricular arrhythmia inducibility on cardiac electrophysiology testing with a promising accuracy. In particular, ROC curve analysis showed that the prediction accuracy for border zone extent (area under ROC = 0.75) was better than scar extent (area under ROC = 0.66). The prediction accuracy was further improved (area under ROC = 0.78), when assessing MIBG uptake in the border zone. A comprehensive review of Alba et al³⁰ sheds light on how far we are in providing useful predictive models for the management of HF patients. Validated risk prediction tools may be useful and should be used, but no tool (and even more a simple binary classification) can encompass all of the relevant information crucial for informed decision making. HF is a dynamic condition and prognosis should be reassessed frequently, particularly in patients for whom the results are critical for clinical and therapeutic decision making. The potential utility of MIBG SPECT in combination with perfusion SPECT imaging for prediction of cardiovascular events needs to be prospectively investigated in larger cohort of patients. In particular, the mismatch between perfusion (ischemia and scar in the border zone) and innervation could represent important areas for future research in HF patients. It would be clinically useful also to explore the impact of different therapies on cardiac innervation and if an improvement in MIBG variables will affect the clinical outcome.

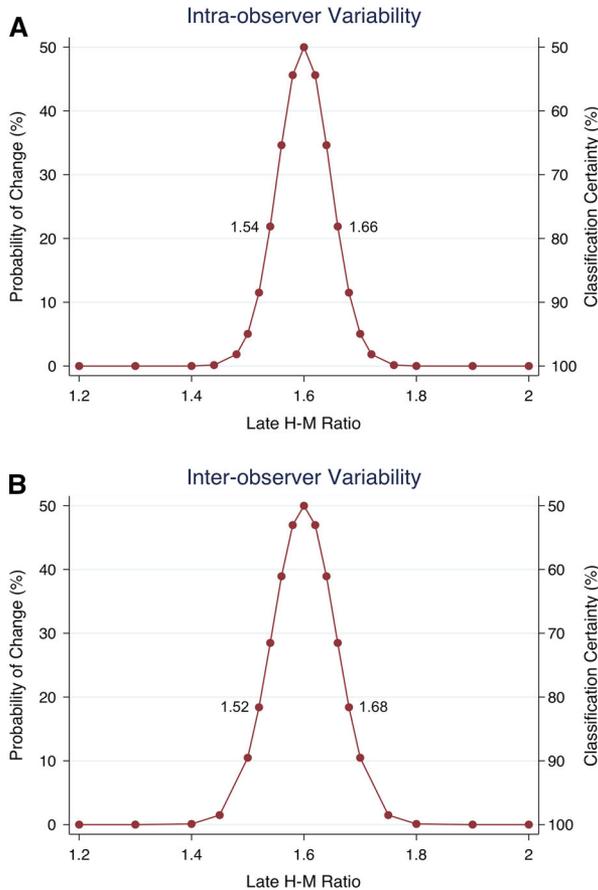


Fig. 1. Classification certainty of a single planar late H/M ratio measurement for values from 1.20 to 2.00 using the proposed cutoff of 1.60. For intra-observer reproducibility (A), the agreement between paired measurements falls below 80% within the region of values from 1.54 to 1.66. For inter-observer reproducibility (B), the agreement between paired measurements falls below 80% within the region of values from 1.52 to 1.68. For both intra- and inter-observer reproducibility, a nadir of approximately 50% is reached around a H/M ratio of 1.60. Classification certainty was calculated using the formula: $1 - \left(\frac{1}{2}e^{-((x-1.60)/SD)^2}\right)$, with x representing each late H/M ratio value; constant e is the base of the natural logarithm; 1.60 is the currently established cutoff for late H/M ratio and SD is the standard deviation of the difference between paired H/M ratio measurements. From our previous study,²⁶ this value was set to 0.066 for intra-observer reproducibility and to 0.080 for inter-observer reproducibility.

Conflict of interest

The authors have indicated that they have no financial conflict of interest.

References

1. Lymperopoulos A, Rengo G, Koch WJ. Adrenergic nervous system in heart failure: Pathophysiology and therapy. *Circ Res.* 2013;113:739-53.
2. Triposkiadis F, Karayannis G, Giamouzis G, Skoularigis J, Louridas G, Butler J. The sympathetic nervous system in heart failure physiology, pathophysiology, and clinical implications. *J Am Coll Cardiol.* 2009;54:1747-62.
3. Malpas SC. Sympathetic nervous system overactivity and its role in the development of cardiovascular disease. *Physiol Rev.* 2010;90:513-57.

4. Jacobson AF, Senior R, Cerqueira MD, Wong ND, Thomas GS, Lopez VA, ADMIRE-HF Investigators, et al. Myocardial iodine-123 meta-iodobenzylguanidine imaging and cardiac events in heart failure. Results of the prospective ADMIRE-HF (AdreView myocardial imaging for risk evaluation in heart failure) study. *J Am Coll Cardiol*. 2010;55:2212–21.
5. Jacobson AF, Lombard J, Banerjee G, Camici PG. 123I-mIBG scintigraphy to predict risk for adverse cardiac outcomes in heart failure patients: Design of two prospective multicenter international trials. *J Nucl Cardiol*. 2009;16:113–21.
6. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst*. 1994;86:829–35.
7. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: A bad idea. *Stat Med*. 2006;25:127–41.
8. Metz CE. Receiver operating characteristic analysis: A tool for the quantitative evaluation of observer performance and imaging systems. *J Am Coll Radiol*. 2006;3:413–22.
9. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J*. 2005;47:458–72.
10. Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW. Evaluation of markers and risk prediction models: Overview of relationships between NRI and decision-analytic measures. *Med Decis Making*. 2013;33:490–501.
11. Van Calster B, Steyerberg EW, D’Agostino RB Sr., Pencina MJ. Sensitivity and specificity can change in opposite directions when new predictive markers are added to risk models. *Med Decis Making*. 2014;34:513–22.
12. Nakata T, Nakajima K, Yamashina S, Yamada T, Momose M, Kasama S, et al. A pooled analysis of multicenter cohort studies of (123I)I-mIBG imaging of sympathetic innervation for assessment of long-term prognosis in heart failure. *JACC Cardiovasc Imaging*. 2013;6:772–84.
13. Merlet P, Valette H, Dubois-Randé JL, Moyse D, Duboc D, Dove P, et al. Prognostic value of cardiac metaiodobenzylguanidine imaging in patients with heart failure. *J Nucl Med*. 1992;33:471–7.
14. Germano G, Kavanagh PB, Slomka PJ, Berman DS. Tracking a therapeutic response: How reliable are serial measurements of LV perfusion and function? *J Nucl Cardiol*. 2012;19:360–3.
15. Iskandrian AE, Hage FG, Shaw LJ, Mahmarian JJ, Berman DS. Serial myocardial perfusion imaging: Defining a significant change and targeting management decisions. *JACC Cardiovasc Imaging*. 2014;7:79–96.
16. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30:1–15.
17. Bland JM, Altman DG. Applying the right statistics: Analyses of measurement studies. *Ultrasound Obstet Gynecol*. 2003;22:85–93.
18. Nakagawa S, Schielzeth H. Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biol Rev Camb Philos Soc*. 2010;85:935–56.
19. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96–106.
20. Watson PF, Petrie A. Method agreement analysis: A review of correct methodology. *Theriogenology*. 2010;73:1167–79.
21. ClinicalTrials.gov Identifier: NCT01936649. A Phase 4, Open-label test-retest study to assess the reproducibility of quantitative measurements of myocardial uptake of AdreView™ (Iobenguane I 123 Injection). <http://clinicaltrials.gov/ct2/show/NCT01936649>. Accessed February 26, 2014.
22. Matsui T, Tsutamoto T, Maeda K, Kusukawa J, Kinoshita M. Prognostic value of repeated 123I-metaiodobenzylguanidine imaging in patients with dilated cardiomyopathy with congestive heart failure before and after optimized treatments: Comparison with neurohumoral factors. *Circ J*. 2002;66:537–43.
23. Kasama S, Toyama T, Sumino H, Nakazawa M, Matsumoto N, Sato Y, et al. Prognostic value of serial cardiac 123I-MIBG imaging in patients with stabilized chronic heart failure and reduced left ventricular ejection fraction. *J Nucl Med*. 2008;49:907–14.
24. Drakos SG, Athanasoulis T, Malliaras KG, Terrovitis JV, Diakos N, Koudoumas D, et al. Myocardial sympathetic innervation and long-term left ventricular mechanical unloading. *JACC Cardiovasc Imaging*. 2010;3:64–70.
25. Veltman CE, Boogers MJ, Meinardi JE, Younis AI, Dibbets-Schneider P, Van der Wall EE, et al. Reproducibility of planar 123I-meta-iodobenzylguanidine (MIBG) myocardial scintigraphy in patients with heart failure. *Eur J Nucl Med Mol Imaging*. 2012;39:1599–608.
26. Pellegrino T, Petretta M, De Luca S, Paolillo S, Boemio A, Carotenuto R, et al. Observer reproducibility of results from a low-dose 123I-metaiodobenzylguanidine cardiac imaging protocol in patients with heart failure. *Eur J Nucl Med Mol Imaging*. 2013;40:1549–57.
27. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033–9.
28. Petraco R, Sen S, Nijjer S, Echavarría-Pinto M, Escaned J, Francis DP, et al. Fractional flow reserve-guided revascularization: Practical implications of a diagnostic gray zone and measurement variability on clinical decisions. *JACC Cardiovasc Intervent*. 2013;6:222–5.
29. Zhou Y, Zhou W, Folks RD, Manatunga DN, Jacobson AF, Bax JJ, et al. I-123 mIBG and Tc-99m myocardial SPECT imaging to predict inducibility of ventricular arrhythmia on electrophysiology testing: A retrospective analysis. *J Nucl Cardiol*. doi:10.1007/s12350-014-9911-7.
30. Alba AC, Agoritsas T, Jankowski M, Courvoisier D, Walter SD, Guyatt GH, et al. Risk prediction models for mortality in ambulatory patients with heart failure: A systematic review. *Circ Heart Fail*. 2013;6:881–9.