

Exercise and Bayes' Theorem: Some things never go out of style

Hong Jun Yun, MD,^a Ravi V. Shah, MD,^b and Venkatesh L. Murthy, MD, PhD, FASNC^{a,c}

^a Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI

^b Division of Cardiovascular Medicine, Department of Internal Medicine, Beth Israel Deaconess Medical Center, Boston, MA

^c Division of Nuclear Medicine and Molecular Imaging, Department of Radiology, University of Michigan, Ann Arbor

Received Aug 26, 2015; accepted Aug 27, 2015
doi:10.1007/s12350-015-0281-6

See related article, pp. 366–378

Exercise stress testing with or without imaging is a mainstay in the diagnosis and management of known or suspected coronary artery disease (CAD). In an era where physicians face a dual mandate of reduction of risk of CAD morbidity while also minimizing economic costs, appropriate use of stress testing may provide valuable prognostic data to guide medical therapy and to select patients for invasive angiography. In order to limit excess testing which likely are to be low yield, stress imaging is usually reserved for those with an intermediate to high pretest risk of CAD and its complications, while low-risk patients who are able to exercise may be managed with exercise EKG alone.¹

In a seminal paper by Mark et al, the Duke treadmill score (DTS), which incorporates exercise capacity, electrocardiographic ST segment changes, and exercise-induced angina pectoris, was shown to predict survival in patients with suspected CAD.² Patients with low-risk scores (≥ 5) reflecting longer exercise times and little or no ST-segment deviation had an annual mortality rate of 0.25%. In contrast, patients with high-risk scores (< -10) had an average annual mortality rate of 5%.

Other models using more detailed clinical and exercise data have been developed and have even greater performance for diagnosis of CAD and for risk stratification.³ This approach has previously been endorsed in AHA/ACC guidelines for the management of chronic stable angina which recommended medical management for those with a risk of cardiac mortality $< 1\%$ per year and invasive angiography for those with a risk of cardiac mortality $\geq 3\%$ per year.⁴ For those with 1-3% risk of cardiac mortality, stress imaging was recommended.

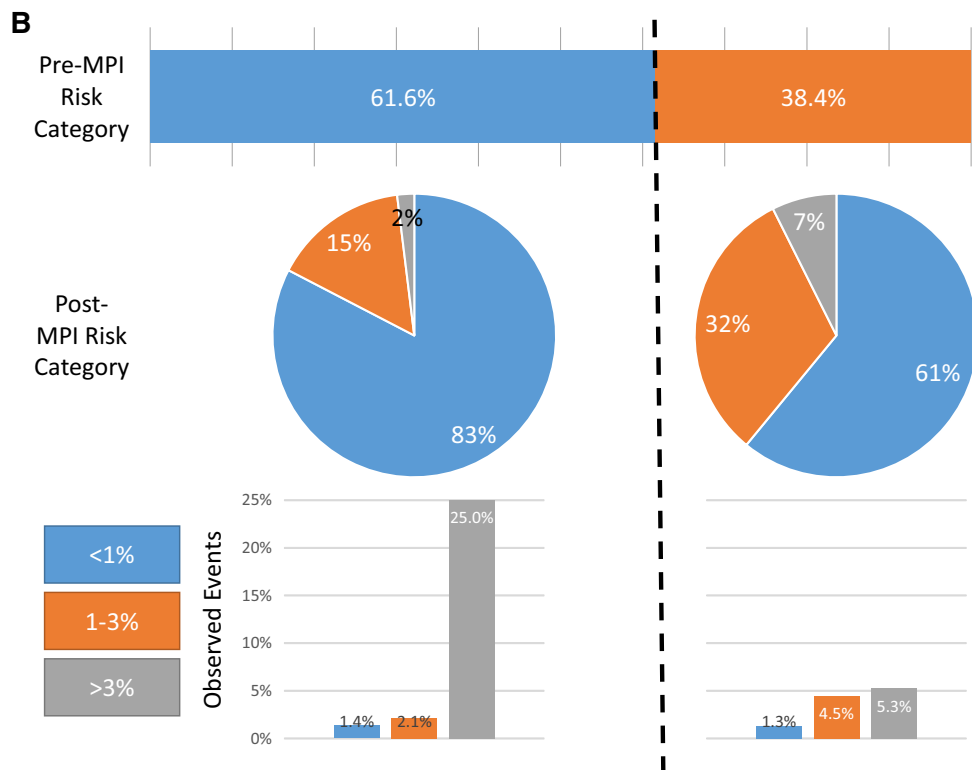
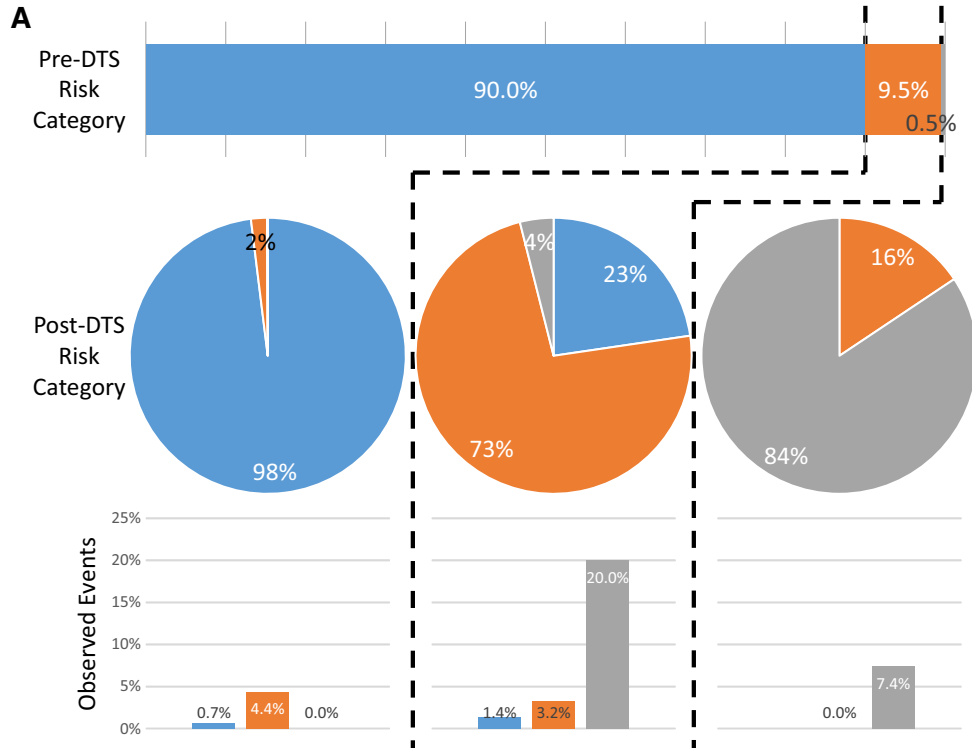
Supporting this approach, exercise testing-derived measures of functional capacity are powerful predictors of prognosis, irrespective of age, gender, or symptoms.⁵⁻⁸ High exercise capacity along with optimal medical therapy has been associated with excellent prognosis, even without revascularization.⁹ Furthermore, in the CASS registry, exercise parameters were useful in risk-stratifying patients who would derive the most benefit from CABG.¹⁰ Specifically, the benefits of CABG were largely confined to those with the poorest exercise capacity. Low exercise capacity patients who underwent CABG had an approximately 31% lower mortality than patients managed medically.

While maximum exercise capacity and ECG changes offer great insight into the presence of obstructive CAD and the risk of adverse cardiovascular events, non-invasive diagnostic imaging, particularly nuclear myocardial perfusion imaging (MPI), provides far more sensitive and detailed delineation of the extent of at-risk myocardium. To date, an overwhelming body of evidence has established that a normal or low-risk exercise nuclear MPI is associated with a $< 1\%$ per year risk of hard cardiac events.¹¹⁻¹³ In a large meta-analysis, the

Reprint requests: Venkatesh L. Murthy MD, PhD, FASNC, Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109-5873; vlmurthy@med.umich.edu

J Nucl Cardiol 2016;23:379–83.
1071-3581/\$34.00

Copyright © 2015 American Society of Nuclear Cardiology.



◀**Figure 1.** Risk reclassification by (A) addition of DTS to clinical variables among patients with suspected CAD and (B) addition of MPI variables to clinical variables and DTS among patients with established CAD in Koh et al. Low risk (<1%/year) are shown in *blue*, intermediate risk (1-3%/year) are shown in *orange*, and high risk (>3%) in *gray*. The *top* bar graph shows the pre-test risk distribution. Pie graphs in the *middle row* demonstrate the portions of patients in each pre-test risk stratum who are reclassified. The bar graphs at the *bottom* demonstrate observed mortality among reclassified groups. Nearly the entire suspected CAD group (A) is low risk and minimal reclassification was seen in that group. Among those with established CAD (B), no subjects were high risk based on clinical variables and DTS and only a small proportion of the cohort was reclassified as high risk based on MPI. Events in those reclassified to high risk were more common, however.

negative predictive value for MI and cardiac death of a normal exercise MPI was 98.8% and the annual event rate was 0.45%.¹² These rates remained low even in subgroup analyses of patients with an elevated pretest probability for CAD, such as a positive exercise treadmill test. Conversely, high-risk features such as reversible perfusion defects $\geq 10\%$ of the myocardium, LVEF <45% or post-stress LVEF reduction, transient ischemic LV dilation, increased lung or right ventricular uptake, and abnormal coronary flow reserve correlate with worse outcomes.¹⁴

Furthermore, there is a close relationship between exercise capacity and ischemia on MPI. Generally, the rate of ischemic perfusion scans is inversely proportional to exercise capacity. Bourque et al showed that patients with high exercise capacity (≥ 10 METs) had a very low prevalence of clinically significant ischemia (defined as $\geq 10\%$ of the left ventricle) and low rates of adverse cardiac outcomes.^{15,16} In another series of low-risk patients undergoing exercise testing with thallium SPECT imaging, functional capacity and stress defect score were both independent predictors of all-cause mortality, although functional capacity was the strongest predictor.¹⁷ Moreover, individuals with high-risk DTS on exercise testing but normal MPI also had low hard event rates.¹⁸

In this issue of the *Journal of Nuclear Cardiology*, Koh et al investigate the incremental prognostic value of exercise testing and MPI parameters when added to clinical variables for all-cause mortality in two distinct subsets of patients referred for clinically indicated stress nuclear perfusion imaging: patients with suspected CAD and those with established CAD. The primary conclusion is that the value of exercise parameters was largely confined to those with suspected but not established CAD, and the value of perfusion imaging was largely confined to those with established CAD.

The conclusions of this study need to be viewed in the context of the cohort examined. The study comprised

6702 patients with suspected and 2008 patients with known CAD evaluated at a single center in Singapore during a three-year period. In the suspected CAD group, 87% of patients had low-risk DTS and only 3.5% had moderate to severely abnormal MPI based on summed stress score. Patients with known CAD had similarly low-risk DTS, but 24% had abnormal MPI. All-cause mortality was low in both cohorts, occurring in 1% of patients without established CAD and 2.2% of patients with established CAD, respectively. Multivariable predictors of adverse outcome in both groups included age, male sex, >2 cardiovascular risk factors, DTS, summed difference score, and LVEF. In evaluating a series of models comprised clinical variables, clinical variables + DTS, and clinical variables + DTS + MPI, the addition of DTS to clinical variables in patients with suspected CAD significantly improved the prognostic model ($P = 0.001$), whereas MPI parameters did not ($P = 0.10$). In contrast, for patients with established CAD, the addition of DTS to clinical variables did not result in incremental prognostic information, whereas the addition of MPI strengthened the model.

This study also employed net reclassification improvement (NRI) analysis which is a relatively novel statistical methodology for quantification of the potential clinical impact of additional information from testing.¹⁹ NRI is computed as the sum of two fractions: (1) the percent of subjects who experienced the outcome in whom risk estimates were correctly increased with the additional test data (NRI for events) and (2) the percent of subjects who did not experience the outcome in whom risk estimates were correctly decreased (NRI for non-events). Koh et al report both of these fractions separately as well as their sum. A positive NRI indicates favorable risk reclassification, so it may be somewhat confusing to readers that the authors of this study have chosen to report a favorable NRI for non-events as a negative number, the inverse of usual practice. Another caveat of this study is that the authors have computed NRI in a manner which ignores the time interval from SPECT MPI to the outcome and the possibility that not all patients were followed for equal duration. Adaptations of NRI to address this have been developed in the statistical literature and may have resulted in somewhat smaller estimates for NRI and wider confidence intervals.²⁰ Importantly, NRI is most useful when clinically established risk categories are available. In this study, the authors adapted cutpoints defined in the 1999 ACC/AHA guidelines for management of chronic stable angina as has previously been done.⁴ The 1 and 3% cutpoints defined in these guidelines for low, intermediate, and high risk are based on cardiac mortality, and generalization to all-cause mortality as was done by Koh et al may introduce further biases. Although

approximately half of mortality in most societies is due to cardiovascular causes, this may vary greatly by age and gender. Nonetheless, it is likely that even more patients in this paper would have been assigned to low-risk categories if cardiac mortality data were used.

The results of the net reclassification analysis in both cohorts are summarized in Fig. 1. Pre-DTS, nearly all (90%) patients with suspected CAD were in the low-risk category with only a fraction in the high-risk group (Fig. 1A). As noted, if cardiac mortality had been used instead of all-cause mortality, an even smaller percentage would likely have been classified as high risk. Overall, there was minimal reclassification in the low-risk category with the addition of DTS (only 2% were reclassified as intermediate risk). Similarly, only a small minority (16%) of high-risk patients were reclassified based on DTS. As expected, the additional value of DTS was more substantial among intermediate-risk patients. Approximately, one in four intermediate-risk patients was reclassified, largely to low risk. Importantly, the 4% of intermediate-risk patients reclassified as high risk based on DTS had a 20% mortality during follow-up of approximately three years.

In patients with established CAD (Fig. 1B), the majority were low risk based on clinical variables and DTS (62%) and only 38% were considered intermediate risk. Importantly, no high-risk patients were identified based on clinical variables and DTS. Again the greatest risk reclassification was seen in patients with an intermediate pre-MPI risk. Approximately, 2 in 3 intermediate-risk patients were reclassified as low risk based on MPI findings, and these patients only had a 1.3% mortality during follow-up. Conversely, observed mortality rates among those who remained at intermediate risk and those reclassified as high risk based on MPI were similar (4.5% and 5.3%). Very few low-risk patients were reclassified, although there was a 25% mortality among the 2% of low-risk patients reclassified as high risk based on MPI findings.

As expected, the results of this study reconfirm that patients with suspected or established CAD who are low risk based on clinical factors and DTS are unlikely to have significant ischemia on MPI and have a favorable prognosis. A key finding of this analysis is that much of the added prognostic value of SPECT MPI is contingent on suitable pretest risk of CAD—a recapitulation of Reverend Thomas Bayes' Theorem. Although this study did find that even among patients with low risk based on clinical variables and DTS, those with abnormal MPI are more likely to have adverse outcomes, this finding was quite rare. Importantly, the findings also demonstrate that the greatest incremental value of MPI in risk assessment is through downgrading risk status in intermediate-risk patients, particularly in those with

established CAD. Indeed, the proportions of patients reclassified to a high risk based on either DTS or MPI in both cohorts were low, underscoring the importance of careful pretest assessment of risk.

There are several important limitations of this study. First, the low observed event rate raises concern for possible overfitting of the multivariate model. Moreover, the patient population was a geographically homogenous group from a single center in Singapore which may not be generalizable to other populations. The high proportion of low-risk DTS suggests that referral patterns and lifestyles may be different from that seen in many U.S. centers.

To further place the results of this study in context, the incremental prognostic value of exercise stress MPI has previously been analyzed in large cohorts. Hachamovitch et al found that while overall disease risk was low (hard event rate 1.8%) in patients with no previous evidence of CAD, myocardial perfusion SPECT did indeed have additional prognostic information beyond clinical and exercise variables.²¹ Potentially, the lack of benefit of MPI in the present study may be due to limited statistical power in the study of Koh et al due to the small number of clinical outcomes.

In the end, achieving an optimal balance between incremental prognostic value from diagnostic stress testing and avoidance of costly but low yield studies will always require careful selection of appropriate patients based on global pretest risk assessment. In the absence of unusual circumstances, low-risk patients defined by clinical variables and exercise capacity will experience low rates of adverse events and are unlikely to benefit from non-invasive imaging. The greatest net benefit and post-test reclassification will almost always be seen among intermediate-risk patients—a recapitulation of Reverend Bayes' theorem.

Disclosures

Dr. Murthy has received research support from INVIA Medical Imaging Solutions and has minor stockholdings in General Electric, Cardinal Health and Mallinckrodt Pharmaceuticals.

References

1. Wolk M, Bailey S. ACCF/AHA/ASE/ASNC/HFSA/HRS/SCAI/SCCT/SCMR/STS 2013 multimodality appropriate use criteria for the detection and risk assessment of stable ischemic heart. *J Card Fail* 2014;20:65-90. doi:10.1016/j.cardfail.2013.12.002.
2. Mark DB, Shaw L, Harrell FE, et al. Prognostic value of a treadmill exercise score in outpatients with suspected coronary artery disease. *N Engl J Med* 1991;325:849-53. doi:10.1056/NEJM199109193251204.
3. Lauer MS, Pothier CE, Magid DJ, et al. An externally validated model for predicting long-term survival after exercise treadmill

- testing in patients with suspected coronary artery disease and a normal electrocardiogram. *Ann Intern Med* 2007;147:821-8.
4. Gibbons RJ, Chatterjee K, Daley J, et al. ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable anginal: this document was approved by the American College of Cardiology Board of Trustees in March 1999, the American Heart Association Science Advisory and Coordinating Committee. *J Am Coll Cardiol* 1999;33:2092-197. doi:10.1016/S0735-1097(99)00150-3.
 5. Ekelund LG, Haskell WL, Johnson JL, et al. Physical fitness as a predictor of cardiovascular mortality in asymptomatic North American men. The Lipid Research Clinics Mortality Follow-up Study. *N Engl J Med* 1988;319:1379-84. doi:10.1056/NEJM198811243192104.
 6. Myers J, Prakash M, Froelicher V, et al. Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 2002;346:793-801. doi:10.1056/NEJMoa011858.
 7. Blair SN, Kohl HW, Barlow CE, et al. Changes in physical fitness and all-cause mortality. A prospective study of healthy and unhealthy men. *JAMA* 1995;273:1093-8.
 8. Spin JM, Prakash M, Froelicher VF, et al. The prognostic value of exercise testing in elderly men. *Am J Med* 2002;112:453-9.
 9. Hung RK, Al-Mallah MH, McEvoy JW, et al. Prognostic value of exercise capacity in patients with coronary artery disease: The FIT (Henry Ford Exercise Testing) project. *Mayo Clin Proc* 2014;89:1644-54. doi:10.1016/j.mayocp.2014.07.011.
 10. Weiner DA, Ryan TJ, Parsons L, et al. Long-term prognostic value of exercise testing in men and women from the Coronary Artery Surgery Study (CASS) registry. *Am J Cardiol* 1995;75:865-70.
 11. Hachamovitch R, Hayes S, Friedman JD, et al. Determinants of risk and its temporal variation in patients with normal stress myocardial perfusion scans: What is the warranty period of a normal scan? *J Am Coll Cardiol* 2003;41:1329-40. doi:10.1016/S0735-1097(03)00125-6.
 12. Metz LD, Beattie M, Hom R, et al. The prognostic value of normal exercise myocardial perfusion imaging and exercise echocardiography: A meta-analysis. *J Am Coll Cardiol* 2007;49:227-37. doi:10.1016/j.jacc.2006.08.048.
 13. Shaw LJ, Iskandrian AE. Prognostic value of gated myocardial perfusion SPECT. *J Nucl Cardiol* 2004;11:171-85. doi:10.1016/j.nuclcard.2003.12.004.
 14. Fihn SD, Gardin JM, Abrams J, et al. 2012 ACCF/AHA Guideline for the diagnosis and management of patients with stable ischemic heart disease: A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol* 2012;60:e44-164. doi:10.1016/j.jacc.2012.07.013.
 15. Bourque JM, Charlton GT, Holland BH, et al. Prognosis in patients achieving ≥ 10 METS on exercise stress testing: Was SPECT imaging useful? *J Nucl Cardiol* 2011;18:230-7. doi:10.1007/s12350-010-9323-2.
 16. Bourque JM, Holland BH, Watson DD, Beller GA. Achieving an exercise workload of ≥ 10 metabolic equivalents predicts a very low risk of inducible ischemia: Does myocardial perfusion imaging have a role? *J Am Coll Cardiol* 2009;54:538-45. doi:10.1016/j.jacc.2009.04.042.
 17. Snader CE, Marwick TH, Pashkow FJ, et al. Importance of estimated functional capacity as a predictor of all-cause mortality among patients referred for exercise thallium single-photon emission computed tomography: Report of 3,400 patients from a single center. *J Am Coll Cardiol* 1997;30:641-8.
 18. Vítola JV, Wanderley MRB, Cerci RJ, et al. Outcome of patients with high-risk Duke treadmill score and normal myocardial perfusion imaging on spect. *J Nucl Cardiol* 2015;. doi:10.1007/s12350-015-0156-x.
 19. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157-72. doi:10.1002/sim.2929 **discussion 207-12**.
 20. Steyerberg EW, Pencina MJ. Reclassification calculations for persons with incomplete follow-up. *Ann Intern Med* 2010;152:195-6. doi:10.7326/0003-4819-152-3-201002020-00019 **author reply 196-7**.
 21. Hachamovitch R, Berman DS, Kiat H, et al. Exercise myocardial perfusion SPECT in patients without known coronary artery disease : Incremental prognostic value and use in risk stratification. *Circulation* 1996;93:905-14. doi:10.1161/01.CIR.93.5.905.