The complicated business of testing our diagnostic tests!

Prem Soman, MD, PhD, FRCP (UK), FACC, MASNC,^{a,b} and Syed Bukhari, MD^a

^a Division of Cardiology, University of Pittsburgh Medical Center, Pittsburgh, PA

^b Division of Cardiology and The Heart and Vascular Institute, University of Pittsburgh Medical Center, Pittsburgh, PA

Received Dec 23, 2019; accepted Dec 26, 2019
doi:10.1007/s12350-019-02026-x

See related article, pp. 2126–2137

Care providers faced with a patient with chest pain have a number of options for noninvasive testing. They also have the option of “not testing” in patients with a low clinical likelihood of angina and a low-risk profile. A strong scientific basis to justify the use of one test over others is lacking and difficult to procure given the magnitude of the undertaking required. Diagnostic tests are usually compared using metrics of test efficacy (diagnostic accuracy). However, metrics of test effectiveness may be equally relevant, including patient outcome after a test, test effectiveness in the real world, downstream resource utilization, and societal cost.

In this issue of the journal, Qinli Ma and colleagues report on their analysis of a large dataset of Medicare patients, which was designed to compare patient outcomes (unstable coronary syndromes, all-cause and cardiac hospitalizations, and all-cause death), downstream resource utilization (CT coronary angiography, invasive angiography, PCI, or CABG), and cost (cardiac and all-cause medical costs reported per patient per month) between patients with suspected coronary artery disease (CAD) who underwent clinically indicated positron emission tomography (PET) myocardial perfusion imaging (MPI) without absolute flow quantification or conventional testing (stress

echocardiography or SPECT MPI; notably, patients who had exercise stress without imaging or CT coronary angiography were not included) as the first step in their evaluation.¹ The question is relevant because PET has established advantages over SPECT including a superior diagnostic accuracy for obstructive coronary artery disease (CAD), but is also more expensive.^{2,3} Furthermore, based on its advantages over SPECT, societal position papers have recommended PET as the preferred test for patients with suspected CAD who are unable to exercise (taking into consideration the generally accepted notion that functional parameters derived from exercise testing provide incrementally useful information and should be obtained when possible, thus making SPECT the preferred modality in such patients).⁴ The authors analyzed data from a nationally representative random sample of 5% of Medicare Fee for service subjects registered with a single large US health insurance provider (Anthem, Inc), who had undergone index imaging for chest pain during a defined 4-year period. Using a standard statistical approach to minimize differences between groups in retrospective data analysis, the authors applied a propensity score to match 4691 patients who underwent PET (3.2% of sampled patients) to an equal number of patients who underwent conventional testing (matched from a total of 139,884 patients). A propensity score is the probability that a subject with certain characteristics will be assigned to one of the two groups.⁵ The investigators used a multivariable logistic regression analysis to estimate the propensity score (likelihood) of having PET as the first imaging test (discriminant analysis and random forests are other potential approaches for deriving a propensity score) and used a “greedy” algorithm to create a one-to-one matched group (as opposed to matching each subject from one group to many subjects in the other group) sampled without replacement (each subject used in only one pairing). In the greedy matching algorithm used, the subject with the

Reprint requests: Prem Soman, MD, PhD, FRCP (UK), FACC, MASNC, Division of Cardiology and The Heart and Vascular Institute, University of Pittsburgh Medical Center, A-429 Scaife Hall, 200 Lothrop Street, Pittsburgh, PA 15213; premsoman@upmc.edu

J Nucl Cardiol 2021;28:2138–40.
1071-3581/\$34.00

Copyright © 2020 American Society of Nuclear Cardiology.

Figure 1. Intrinsic and extrinsic factors that influence the relationship between a diagnostic test and patient outcome (see text).

highest propensity score from one group is matched to the subject with the closest propensity score from the other group (choosing the “greedy nearest neighbor”), and the process is repeated for all subjects.⁵ An exhaustive list of covariates was used to create the propensity score, including a number of demographic and clinical variables (including the Elixhauser Comorbidity index score), baseline utilization, and cost characteristics. The average duration of follow up was 17 months.

The investigators found that the PET group had a 3.8-mSv less radiation dose per year and a lower incidence of unstable coronary syndromes (incidence rate ratio, IRR 0.77). However, the PET group also experienced more utilization of percutaneous coronary intervention (PCI, IRR 1.24) and hospitalizations, presumably related to PCI (IRR 1.10), while the mortality rate was similar (Hazard ratio, HR 0.95). In addition, PET was associated with a significantly higher Medicare spending of \$82.3 and \$539.9 per patient per month for cardiac and all-cause medical services, respectively. They conclude that PET imaging, compared to stress echo or SPECT MPI as the initial test in patients with chest pain, was not associated with reduced levels of utilization and spending despite “mostly” similar outcomes.

The derived propensity score is just an estimate, always raising the possibility of residual bias due to unrecognized confounders. A sensitivity analysis (a post hoc “what -if” analysis done to check the robustness of the primary results) using multivariate regression with variations of baseline characteristics and an inverse probability of treatment weighting was performed and

yielded similar results to the primary analysis. Furthermore, similar downstream risk for pneumonia and sepsis also suggests that the groups were matched in terms of co-morbidity risk.

While the authors’ conclusions pertaining to downstream resource utilization and cost are justified by the data, they curiously relegate the finding of a lower incidence of unstable coronary syndromes to the background, with just a single sentence referring to it in the discussion. This outcome measure is not explicitly defined in the manuscript, but can be reasonably assumed to include unstable angina and non-fatal myocardial infarctions. The PET group had a significant 23% reduction in this endpoint compared to the conventionally tested group. Given the earlier and 24% higher application of PCI to the PET group, is it not possible that this was causally related to the lower incidence of unstable coronary syndromes? At the very least, this finding is entirely consistent with our current understanding of the effects of coronary revascularization in chronic coronary syndromes, which indicates favorable effects on symptoms and quality of life, but not on mortality.

This paper reminds us of two important considerations relevant to the assessment of our diagnostic tests.

First, as illustrated in Figure 1, the relationship between a diagnostic test and patient outcome is influenced by many factors beyond test performance. The patient population studied is important, since even a highly accurate and effective test applied to subjects with a low probability of disease is likely to result in false-positive results and unnecessary downstream testing. Outcome after a diagnostic test is inherently dependent upon accurate test interpretation and the institution of appropriate therapeutic strategies. Finally, the choice of an appropriate outcome measure is critical, as illustrated by the discussion above. The use of a mortality endpoint may impose an unfairly high bar for the assessment of diagnostic tests for CAD.

Second, the emerging trend of papers using effectiveness endpoints (real world applicability, downstream resource utilization, and cost) in addition to efficacy endpoints (diagnostic accuracy) in the assessment of diagnostic tests for CAD is welcome. This paper by Ma et al is not the first to question the effectiveness of cardiac PET despite its superior accuracy. In a recent randomized comparison of the clinical effectiveness of PET and SPECT MPI in symptomatic patients with optimally medicated CAD, Patel and colleagues showed similar rates of 60-day diagnostic failure, defined creatively by the investigators as invasive angiography without finding obstructive CAD or the need for additional noninvasive diagnostic testing.⁶ In another randomized trial, Sibelink and colleagues showed

comparable outcomes when patients with CAD and LV systolic dysfunction were revascularized based on viability information derived from PET or SPECT imaging, despite the known superiority of PET over SPECT for the detection of myocardial viability.⁷ In a healthcare environment with an increasing emphasis on value, such effectiveness measures give us incremental guidance in the complicated business of selecting our diagnostic tests.

Disclosures

Prem Soman and Syed Bukhari have nothing to disclose.

References

1. Ma Q, Sridhar G, Power T, Agiro A. Assessing the downstream value of first-line cardiac positron emission tomography (PET) imaging using real world Medicare fee-for-service claims data. *J Nucl Cardiol* 2019. <https://doi.org/10.1007/s12350-019-01974-8>.
2. Mc Ardle BA, Dowsley TF, deKemp RA, Wells GA, Beanlands RS. Does rubidium-82 pet have superior accuracy to spect perfusion imaging for the diagnosis of obstructive coronary disease? A systematic review and meta-analysis. *J Am Coll Cardiol* 2012;60:1828-37.
3. Parker MW, Iskandar A, Limone B, Perugini A, Kim H, Jones C, Calamari B, Coleman CI, Heller GV. Diagnostic accuracy of cardiac positron emission tomography versus single photon emission computed tomography for coronary artery disease/clinical perspective. *Circul Cardiovasc Imaging* 2012;5:700-7.
4. Bateman TM, Dilsizian V, Beanlands RS, DePuey EG, Heller GV, Wolinsky DA. American society of nuclear cardiology and society of nuclear medicine and molecular imaging joint position statement on the clinical indications for myocardial perfusion pet. *J Nucl Cardiol* 2016;23:1227-31.
5. Statistics How To. Propensity score matching: Definition & overview. (February 3, 2017). <https://www.statisticshowto.datasciencecentral.com>. Accessed on 22 Dec 2019
6. Patel KK, Al Badarin F, Chan PS, Spertus JA, Courter S, Kennedy KF, Case JA, McGhie AI, Heller GV, Bateman TM. Randomized comparison of clinical effectiveness of pharmacologic spect and pet mpi in symptomatic cad patients. *JACC Cardiovasc Imaging* 2019;12:1821-31.
7. Siebelink HM, Blanksma PK, Crijns HJ, Bax JJ, Kingma T, Piers DA, Pruim J, Jager PL, Vaalburg W. No difference in cardiac event-free survival between positron emission tomography-guided and single-photon emission computed tomography-guided patient management: a prospective, randomized comparison of patients with suspicion of jeopardized myocardium. *J Am Coll Cardiol* 2001;37:81-8.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.