International Journal of Cardiology 182 (2015) 534–540

Contents lists available at ScienceDirect

International Journal of Cardiology

journal homepage: www.elsevier.com/locate/ijcard

External validation of models for estimating pretest probability of coronary artery disease among individuals undergoing myocardial perfusion imaging ☆

Brent A. Williams [a,*], Joseph A. Ladapo [b], Michael E. Merhige [c,1]

[a] Geisinger Health System, Danville, PA, United States
[b] New York University School of Medicine, New York, NY, United States
[c] Coronary Disease Reversal and Prevention Center, Tonawanda, NY, United States

ARTICLE INFO

Article history:
Received 1 August 2014
Received in revised form 21 November 2014
Accepted 3 January 2015
Available online 6 January 2015

Keywords:
Cardiac imaging
Coronary artery disease
Positron emission tomography

ABSTRACT

Background: Clinical decisions regarding the appropriateness of noninvasive cardiac imaging among individuals with suspected coronary artery disease (CAD) rely heavily on the pretest probability of coronary artery disease (pCAD), often estimated from clinical prediction models. These models have not been validated among individuals undergoing noninvasive myocardial perfusion imaging (MPI) for suspected CAD. Thus, the objective of this study was to compare the extent of agreement and predictive performance between four published pCAD models among individuals undergoing positron emission tomography (PET MPI).

Methods: This cross-sectional study performed at a cardiac referral center included 2383 patients with stable symptoms undergoing PET MPI for the evaluation of suspected CAD. pCAD was estimated on a per-patient basis using four distinct pCAD estimation models. All pCAD estimates were calibrated to a common standard to allow fair comparisons of agreement and predictive performance. Pairwise pCAD model disagreement was defined as percent discordance in classifying patients as low versus intermediate pCAD ($<10\%$ vs. $\geq 10\%$). Predictive performance was quantified by c-statistics with abnormal myocardial perfusion as a binary outcome.

Results: Pairwise pCAD estimates demonstrated non-negligible disagreement with percent discordance between models ranging from 11% to 23%. Agreement worsened when higher thresholds for distinguishing low-intermediate pCAD were employed. All pCAD models demonstrated poor predictive performance for identifying abnormal stress perfusion with c-statistics ranging from 0.554 to 0.616.

Conclusions: pCAD estimation models showed suboptimal agreement and poor predictive performance in patients undergoing PET MPI. The transportability of pCAD models to MPI patients should be questioned and further evaluated in future studies.

1. Introduction

Among individuals with stable symptoms suggestive of occlusive coronary artery disease (CAD), proper patient selection for diagnostic noninvasive cardiac imaging should rely heavily on the estimated pretest probability of obstructive CAD (pCAD) using any one of multiple clinical prediction models designed for this purpose [1–5]. Appropriate use criteria for noninvasive cardiac imaging generally regard an *intermediate* pCAD (>10%) as an appropriate indication for imaging, while patients with a *low* pCAD below this threshold are typically considered inappropriate for imaging [6–8]. Though pCAD estimation models were initially developed in patients undergoing elective, invasive diagnostic coronary arteriography up to 35 years ago, they have been applied in more contemporary settings as well among patients undergoing noninvasive imaging [9–11]. The application of pCAD models to the noninvasive setting seems appropriate given the increasing role of noninvasive stress testing in the diagnostic evaluation of suspected CAD and the desire to minimize use of invasive coronary arteriography for solely diagnostic purposes [12–16]. Previous studies evaluating the external validity of pCAD models have been restricted to patients undergoing anatomic CAD evaluation with either invasive or noninvasive modalities. These studies generally found that pCAD models had fair discriminatory power for detecting obstructive CAD, but tended to overestimate CAD prevalence in lower-risk, noninvasive imaging cohorts [9,11,17].

To our knowledge, no study has externally validated the commonly employed pCAD estimation models in patients undergoing functional

☆ All authors take responsibility for all aspects of the reliability and freedom from bias of the data presented and their discussed interpretation.

* Corresponding author at: Geisinger Health System, 100 North Academy Avenue, Danville, PA 17822, United States.
E-mail address: bawilliams2@geisinger.edu (B.A. Williams).

[1] Dr. Michael Merhige is Medical Director of Positron Corporation.

http://dx.doi.org/10.1016/j.ijcard.2015.01.015
0167-5273/

CAD assessment with noninvasive myocardial perfusion imaging (MPI). Validating pCAD models in the MPI setting is particularly important given the growing recognition that functionally significant CAD may be a more valuable guide to selecting patients for coronary artery revascularization than anatomic CAD measures [18,19]. Accordingly, the current study was designed to compare and contrast four pCAD estimation models among a group of patients with suspected CAD undergoing positron emission tomography myocardial perfusion imaging (PET MPI) for diagnostic purposes. The primary goals of this study were to (1) evaluate the strength of correlation in pCAD estimates between models; (2) quantify the extent of disagreement between pCAD models in classifying patients as low versus intermediate pCAD; and (3) compare the predictive ability of pCAD models for identifying functionally significant CAD as determined by PET MPI.

2. Methods

2.1. Study patients

The current study is comprised of 2383 patients who underwent clinically-indicated, rest–stress Rubidium-82 (Rb82) PET MPI for the evaluation of stable symptoms suggestive of obstructive CAD [20]. Patients with a known history of myocardial infarction, percutaneous coronary intervention, or coronary bypass surgery were excluded from the current study, as were those undergoing PET MPI for the evaluation of a possible acute coronary syndrome. At the time of PET MPI, information pertaining to demographics, physical examination, cardiac risk factors, medical history, medications, and indication(s) for PET MPI were gathered and stored in patients' medical records. According to a pre-specified research protocol, the aforementioned data elements were retrospectively gathered by trained personnel from paper medical records and transcribed into a research database as previously described [20]. The parent study was approved by the Institutional Review Board at the University at Buffalo who granted a waiver of patient consent [20].

2.2. PET MPI perfusion measurements

The PET MPI protocol in place at the study institution has been previously described [21]. Briefly, attenuation-corrected rest–stress Rb82 PET MPI was performed for determining the fractional volume of the left ventricular (LV) myocardium affected by resting and stress-induced perfusion defects [22]. For stress PET, coronary hyperemia was achieved pharmacologically, typically with dipyridamole. Subjective visual interpretation of LV perfusion defect size and severity is complemented by automated computer software permitting precise, objective quantitation of multiple perfusion parameters [23]. Briefly, automated software identifies the 2% of the LV myocardium with maximal radiotracer uptake (perfusion) as a reference region, and size and severity of perfusion defects in the remaining myocardium are normalized to this maximally-perfused area. The software allows estimating the size of relative perfusion defects at multiple severity thresholds. In the current study, the PET parameter of interest is *perfusion defect size* defined as the percent of the LV myocardium with relative tracer uptake (perfusion) less than 60% of maximum [20,24]. Perfusion defect size was calculated at rest and during stress. Stress perfusion defects, including any combination of resting and stress-induced defects, were chosen as the primary perfusion parameter in this study as both types of perfusion defects are clinically important to identify among individuals without an established history of coronary disease. Stress-induced perfusion defects were analyzed individually as a secondary outcome. Perfusion defect size was analyzed on both continuous and categorical scales. A clinical threshold of 2.5% of the LV was used to differentiate abnormal from normal perfusion by PET MPI (i.e. perfusion defect size ≥ 2.5% was considered abnormal). This software-based threshold for defining abnormal was considered the clinical equivalent to a visual-based threshold often incorporated into clinical practice (≥5% of the LV). Baseline patient characteristics are reported across LV perfusion defect size groups using the following categories: 0%, >0–2.5%, >2.5–5%, >5–<10%, and <10%. Continuous variables are reported as means and standard deviations, with differences across groups tested for significance by Kruskal–Wallis tests; categorical variables are reported as relative frequencies, with differences across groups tested for significance by chi-square tests.

2.3. Models for estimating pretest probability of coronary artery disease

The estimated pretest probability of significant CAD was calculated individually for all study patients using four distinct models as described by (1) Diamond et al. ("CADENZA"), (2) Pryor et al. ("Duke"), (3) Morise et al. ("Morise"), and (4) Patel et al. ("NCDR" — National Cardiovascular Data Registry) [3–5,25]. Briefly, these models consist of 12 to 23 variables whose predictive information for obstructive CAD is aggregated into a single quantitative pCAD estimate. A set of covariates is common to all four models: age, gender, smoking, dyslipidemia, diabetes, and qualitative chest pain characteristics. All models were developed in patients undergoing invasive diagnostic coronary arteriography with significantly obstructive anatomic CAD as the endpoint. CADENZA pCAD estimates were calculated using computerized software available at the study institution [5]. The Duke, Morise, and NCDR pCAD estimates were obtained by applying the respective logistic regression models to study patients as reported in the original publications. All model variables were included in the calculations with the exception of myocardial infarction- and Q wave-related variables in the Duke model, and estimated glomerular filtration rate (not available), ejection fraction, and noninvasive testing results in the NCDR model. Importantly, many of these variables cannot appropriately be considered pretest information in patients with suspected CAD undergoing noninvasive imaging.

2.4. Assessing agreement between pCAD models

By design, this study did not intend to evaluate systematic differences in between-model calibration resulting largely from variation in disease prevalence across the four studies. Accordingly, all pCAD models were calibrated to a common standard to allow a more unbiased evaluation of agreement. The CADENZA pCAD model served as the common standard as this model was actively employed at the study institution. Recalibration of the remaining three models was achieved by updating the respective intercept terms in the logistic regression models such that the mean of predicted values from the respective models equaled the mean pCAD estimate from the CADENZA model (31%) [26]. This recalibration serves to adjust pCAD estimates for differences in study populations not captured by the predictors in the respective models which cause a systematic miscalibration of model estimates. After recalibration, the strength of association between pairwise pCAD estimates was evaluated by Spearman rank correlation coefficients. Scatterplots were created to graphically display the strength of association between pairwise estimates. The extent of disagreement between pCAD estimates was also quantified according to various probability thresholds for differentiating low from intermediate pCAD (10%, 15%, and 30%). In this analysis, percent disagreement reflects the percentage of patients for whom classification into low versus intermediate pCAD is discordant when applying two different pCAD models (i.e. one model labels as "low", and a second comparison model labels as "intermediate").

2.5. Assessing predictive capacity of pCAD models

The associations between pCAD estimates and PET MPI parameters were evaluated by logistic and linear regression models for binary and continuous outcomes respectively. Stress and stress-induced perfusion defects were analyzed separately. pCAD estimates were divided into four groups: <10%, 10–30%, 30–50%, and >50%. The 10% and 30% thresholds have been commonly applied for differentiating low from intermediate pCAD [6–8,17,19,27]. As a binary variable, the percentage with abnormal PET (perfusion defect size ≥ 2.5%) is reported across groups for each pCAD model. Odds ratios and 95% confidence intervals from a logistic regression model are reported with pCAD < 10% as the reference. Model discrimination was evaluated with c-statistics, and the percent of outcome variation explained by a model was evaluated with Nagelkerke's R^2 [26]. Furthermore, metrics for evaluating the performance of diagnostic tests, including sensitivity, specificity, positive predictive value, and negative predictive value were calculated comparing various pCAD thresholds (10%, 15%, 30%) against abnormal perfusion by PET MPI. As a continuous variable, perfusion defect size is reported as mean ± standard deviation across pCAD groups. The percent of outcome variation explained by a model is evaluated by R^2 for linear regression models after a normalizing transformation of perfusion defect size.

3. Results

The baseline characteristics of the 2383 study patients both overall and stratified by stress perfusion defect size are shown in Table 1. The mean age of study patients was 60 (±13) years, and 49% were male. The overall mean size of stress perfusion defects was 2.5% (±6.4) of the LV, and 473 (20%) patients had abnormal stress perfusion involving more than 2.5% of the LV. The overall mean size of stress-induced perfusion defects was 1.5% (±4.9) of the LV, and 275 (12%) patients had abnormal stress-induced perfusion involving more than 2.5% of the LV. As expected, the size of stress perfusion defects was positively associated with several known cardiac risk factors, including older age, male gender, systolic blood pressure, hypertension, diabetes, peripheral vascular disease, and cerebrovascular disease. Accordingly, patients with larger perfusion defects were also more likely to be taking several cardioprotective medications such as statins, nitrates, diuretics, beta blockers, calcium channel blockers, angiotensin converting enzyme inhibitors, and antiplatelet agents. By design, all pCAD estimation models were calibrated such that the mean of predicted values (31%) was identical across all four models (Table 1). All four pCAD estimates were significantly associated with perfusion defect size.

Scatterplots and Spearman correlation coefficients comparing pairwise pCAD estimates are presented in Fig. 1. Correlation coefficients ranged from a minimum of 0.36 when comparing pCAD estimates from CADENZA and Duke, to a maximum of 0.80 when comparing pCAD estimates from Duke and NCDR. All scatterplots revealed important dispersion around the line of identity (y = x). When incorporating a