PCORI Methodology Standards: Academic Curriculum

© 2016 Patient-Centered Outcomes Research Institute. All Rights Reserved.
Module 4: Assessing the Effect of Factors Known to Affect Diagnostic Test Evaluations

Category 10: Studies of Diagnostic Tests

Prepared by Najlla Nassery, MD, MPH
Jodi Segal, MD, MPH

Presented by Najila Nassery, MD, MPH
The accuracy of diagnostic tests and the impact on process of care and outcomes can be influenced by a variety of factors.

The STARD (Standards for the Reporting of Diagnostic Accuracy Studies) initiative, Food and Drug Association, and CONSORT (Consolidated Standards of Reporting Trials) group have issued guidelines for reporting on key factors that affect diagnostic performance in studies.

Guidelines specify that investigators need to evaluate factors, including:

- The condition of interest (i.e., how the condition is defined, how subjects with the condition are distinguished from those without the condition, and how a “positive” test is defined)
- The technical characteristics of the test
- Patient characteristics
- Characteristics of test interpreters
Test Characteristics

- Measures such as sensitivity and specificity vary with the threshold chosen for a “positive” diagnostic test

- This fundamental aspect of diagnostic testing has led to the development of Receiver Operating Characteristic (ROC) analysis (Zhou, et al., 2011) and the development of Summary ROC curves in meta-analysis of diagnostic test accuracy (Gatsonis and Paliwal, 2006)

- The technical specifications of a test include the criteria for a positive test result, machine types/settings, and assays

Sources:
Test Characteristics

Source:
When designing comparative-effectiveness research (CER) studies of diagnostic tests, it is important to ...

- Identify participant subgroups of interest
- When feasible, design the study with adequate precision to reach conclusions specific to these subgroups

Example:

- ACRIN Digital vs. Screen-Film Mammography Trial (Digital Mammographic Imaging Screening Trial (DMIST; Pisano, et al., 2005)
  - There was no significant difference between digital and film mammography for all women enrolled into the trial
  - Assessment of subgroups was able to identify that digital mammography had a higher area under curve (AUC) for pre-and perimenopausal women younger than 50 years with dense breasts

By emphasizing the need to consider patient subgroups at the design stage of studies, this standard helps clarify study goals and also minimize the likelihood of “fishing expeditions” after the data are collected.
“Clinical trial reports need a clearly defined policy on uses of baseline data, especially with respect to covariate adjustment and subgroup analysis. There is substantial risk of exaggerated claims of treatment effects arising from post-hoc emphasis across multiple analyses. Subgroup analyses are particularly prone to over-interpretation.”

—Assmann, et al., 2000
The approach to test interpretation includes a description of the population of test interpreters, when applicable, the amount and type of clinical information available to them, and any special training needed by interpreters.

Recruiting test interpreters with appropriate training can substantially improve diagnostic performance (Elmore, et al., 2009).

When comparing two or more tests, each patient should receive each test under the same protocol (e.g., using the same operator, giving tests in close temporal proximity) to minimize potential bias (to minimize variation not resulting from the test itself).

Standardization of testing technology and testing results across settings (i.e., sites, platforms, laboratories, institutions, geographic regions) is essential to draw reliable and generalizable scientific conclusions.

Adherence to this standard reduces the number of potential factors that could introduce heterogeneity and thereby the number of assumptions that need to be made to ensure validity.
Extensive methodology is available for study variations in diagnostic accuracy and test outcomes

- The literature includes hierarchical and mixed models and resampling-based approaches (Zhou, et al., 2011; Zou, et al., 2011)
- By addressing the sources of variability in test performance and outcomes, the results of the study can be better understood and applied to a particular clinical setting

When feasible, power and sample size calculations should take into account any additional variability created by factors known to affect diagnostic performance

- Any potential sources of bias that may be encountered should also be accounted for in the statistical planning

Sources:
Addressing these may lead to an increase in the number of participants who need to be enrolled into a trial.

Moreover, if subgroups are being tested, the statistical analysis plan should take into consideration the simultaneous testing of multiple groups.
References


References


