

Recommended labels for approaches to evaluate diagnostic accuracy: the STARD ReLabel project

Maria Olsen¹, Bada Yang¹, Patrick Bossuyt¹ and Chris Hyde²

¹*Amsterdam University Medical Centers, dept. of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.*

²*Exeter Test Group, Institute of Health and Research, College of Medicine and Health, University of Exeter*

Background: There is no standardized terminology for describing diagnostic test accuracy (DTA) studies, which presents a barrier to clear and informative reporting of primary studies and hinders efforts towards making valid evidence synthesis. In a previous project, we observed a heterogeneous and sometimes confusing use of terminology for describing DTA study design features in reviews prepared for NICE guidelines^[1].

Aims: To develop a coherent set of terms for describing DTA study design features.

Methods: Based on data from our previous study, and newly collected data on features and terms, we are performing an iterative clarification, sorting, and categorization of all the terms and features we identified. These will be integrated in a coherent and complete set of terms, as a prototype. The strengths and limitations of this prototype are evaluated through an electronic survey. Participants are experienced DTA researchers and non-academic stakeholders and include health technology assessment groups, DTA guideline developers, and collaborators from industry. The survey responses are used to adapt and modify the set of terms. In the last phase, the set of terms will be piloted among end users with varying levels of DTA experience, to evaluate if it facilitates informative descriptions of DTA study designs.

Results: Our set of terms, developed with the input from a large group of experts and stakeholders, can be used to describe a DTA study in sufficient detail, without ambiguity.

Perspectives: We believe that having a standardized and agreed upon set of terms can reduce the use of misleading, subjective, ambiguous and heterogeneous wording when describing DTA research. This will eventually enable secondary researchers and health care decision-makers to better assess the validity and generalizability of DTA evidence.

Keywords

Diagnostic test accuracy, Study designs, Terminology, Labelling