



Title	A Methodological Review of How Heterogeneity Has Been Examined in Systematic Reviews of Diagnostic Test Accuracy
Agency	NCCHTA, National Coordinating Centre for Health Technology Assessment Mailpoint 728, Boldrewood, University of Southampton, Southampton SO16 7PX, United Kingdom; Tel: +44 2380 595586, Fax: +44 2380 595639
Reference	Health Technol Assess 2005;9(12). March 2005. www.hta.ac.uk/execsumm/summ912.htm

Aim

To review how heterogeneity has been examined in systematic reviews of diagnostic test accuracy studies.

Conclusions and results

The 189 systematic reviews that met the inclusion criteria included a median of 18 studies. Meta-analyses had a higher median number (22 studies) compared to narrative reviews (11 studies). Graphic plots showing the spread of study results were provided in 56% of meta-analyses; in 79% these were plots of sensitivity and specificity in the receiver operating characteristic (ROC) space. Statistical tests to identify heterogeneity were used in 32% of reviews: 41% of meta-analyses and 9% of reviews using narrative syntheses. The χ^2 test and Fisher's exact test to assess heterogeneity in individual aspects of test performance were most common. In contrast, only 16% of meta-analyses used correlation coefficients to test for a threshold effect. A narrative synthesis was used in 30% of reviews. Of the meta-analyses, 52% carried out statistical pooling alone, 18% conducted only summary ROC (SROC) analyses, and 30% used both methods. In SROC analyses, the main differences between the models used were the weights chosen for the regression models, although in 42% of cases the use, or choice, of weight was not provided. The proportion of reviews using statistical pooling alone declined from 67% in 1995 to 42% in 2001, with a corresponding increase in the use of SROC methods, from 33% to 58%. Two-thirds of those using SROC methods carried out statistical pooling rather than presenting only SROC models. Reviews using SROC analyses often presented results as a combination of sensitivity and specificity. Three-quarters of meta-analyses attempted to investigate statistically possible sources of variation, using subgroup or regression analysis. The impact of clinical or sociodemographic variables was investigated in 74% of these reviews and test- or threshold-related variables in 79%. At least one quality-related variable was investigated in 63% of reviews. Within this subset, the most common variables were the use of blinding, sample size, the reference test used, and the avoidance of verification bias.

Recommendations

The emphasis on pooling individual aspects of diagnostic test performance and the under-use of statistical tests and graphic approaches to identify heterogeneity might reflect uncertainty about the most appropriate methods to use and greater familiarity with more traditional indices of test accuracy. This indicates the complexity of performing such reviews. In these cases it is strongly suggested that a statistician familiar with the field should be involved in the meta-analyses.

Methods

Systematic reviews that evaluated a diagnostic or screening test by including studies that compared a test with a reference test were identified from DARE. Reviews with structured abstracts (up to December 2001) were screened for inclusion. Data extraction was undertaken using standardized data extraction forms.

Further research/reviews required

Further methodological work on the statistical methods for combining diagnostic test accuracy studies is needed, as are sufficiently large, prospectively designed, primary studies of diagnostic test accuracy comparing two or more tests for the same target disorder. Use of individual patient data meta-analysis in diagnostic test accuracy reviews should be explored to consider heterogeneity in more detail.