Evaluation of a graphical method for detecting accelerated loss of lung function in occupational settings

Dr Robert Anthony Bell, 2010

ABSTRACT

Aim: To investigate whether a predictive graph generated by a spreadsheet is helpful to occupational physicians analysing serial spirometry for accelerated loss of lung function.

Method: Occupational physicians were recruited and randomised to receive a standard case set of serial spirometry printouts, with or without an associated predictive graph with each case, and asked to identify cases with accelerated loss of FEV₁ as well as rate their confidence in their classification. The standard case set comprised of eighteen cases with an average 6.61 (range 4-10) individual spirometry traces in each, covering an average of 6.9 years of testing (range 3.88-11.5 years). Predictive graphs included a lower limit of decline for FEV₁ calculated according to American College of Occupational and Environmental Medicine methodology, plus the plotted FEV₁ measures. Inter-observer agreement statistics were calculated, classification proportions and average rater confidence was compared between the two groups. Agreement with SPIROLA expert software classification was also calculated for each rater.

Results: Twelve participants completed the study, six in each group. Participant characteristics were similar. Raters in the graphical method group classified less abnormality than the usual method group, but the difference was not statistically significant ($X^2= 2.711 \ p=0.0996$). Kappa scores for inter-observer agreement were 0.607 (95%CI 0.488 to 0.727) for the raters in the usual method group, and 0.311 (95%CI 0.192 to 0.430) in the graphical method group. Mean rater confidence scores were 3.938 ($\sigma= 0.561$, SE=0.229) and 4.103 ($\sigma = 0.404$, SE=0.165) in the usual and graphical method groups respectively. This difference was not significant. Kappa scores for individual rater agreement with expert software classification were generally lower in the graphical method users.

Conclusion: The graphical tool did not improve rater agreement, either inter-observer or with reference to expert software, compared with usual methods in physicians attempting to identify rapidly declining lung functions. The tool cannot be recommended for general use without further evaluation.