Examiner Calibration Exercise for the Assessment of Radiographic Caries

J.M. McClary*, P.A. Walters, S.A. Jacobs, R.W. Gerlach, R.D. Bartizek, J.L. Winston, A.R. Biesbrock
Procter & Gamble Co., Mason, OH, USA

662

ABSTRACT

The calibration of caries examiners for clinical studies often involves a clinical visual-tactile repeatability assessment and sometimes an assessment of radiographic repeatability. However, there are very few reports of structured radiographic calibration protocols in the literature. This novel examiner calibration study design was developed to assess the proficiency of eleven examiners to detect caries presentation and caries progression in the permanent teeth of children between the ages of 9 and 12. A test set of bitewing radiographs was selected from a one year clinical caries study involving 150 subjects who had four radiographs taken at baseline and four at one year. The test set consisted of the matched baseline and one year radiograph sets from 20 subjects that were selected based on film quality and the diverse presentation of radiographic caries. The radiograph film sets were blinded with respect to subject identifier and visit period and randomly ordered. This set was independently scored for interproximal caries using an illuminated light box by a panel of three experienced clinicians whose consensus served as the gold standard. Overall, a combined total of 88 lesions were detected at baseline and at one year. Eleven experienced caries examiners using their preferred methods of evaluation then evaluated the test set. Sensitivity of the eleven examiners ranged from 66% to 100%, while the number of false positive scores ranged from 4 to 119 representing positive predictive values ranging from 42% to 94%. Because of the matched nature of the radiograph sets, radiographic increment over one year could also be determined. Based on the increment observed by the standard panel, weighted kappa statistics were calculated for each of the eleven examiners. The weighted kappa values ranged from a low of 27% to a high of 88%. **This model has promise for assessing examiner sensitivity to radiographic caries detection.**

INTRODUCTION

The calibration of caries examiners for clinical studies often involves a clinical visual-tactile repeatability assessment and sometimes an assessment of radiographic repeatability. However, there are very few reports of structured radiographic calibration protocols in the literature.

This novel examiner calibration study design was developed to assess the proficiency of eleven examiners to detect caries presentation and caries progression in the permanent teeth of children between the ages of 9 and 12.

The detection of interproximal caries from standard radiographic films in caries clinical studies has not always yielded similar results across examiners. There have been instances where an individual examiner does not have the consistent sensitivity to follow radiographic progression from a sound surface status to caries as successfully as others reading the same set of films.

OBJECTIVE

To assess the ability of independent examiners to detect the presentation of caries and progression of caries in permanent teeth of children between the ages of 9 and 12.

MATERIALS AND METHODS

-Each set of 4 posterior bitewing radiographs (covering molars and premolars) was assessed for film quality and presentation of radiographic caries.

-Based on this assessment, radiographs from the Baseline and One-year clinical examinations for twenty (20) subjects were selected.

-Sets of films were blinded with respect to subject identifier and visit period and randomly ordered.

-The randomized set was independently scored for interproximal caries using an illuminated light box by a panel of three experienced clinicians whose consensus served as the standard for comparison.

-Eleven experienced caries examiners using their preferred methods of evaluation then independently evaluated the test set.

-The data were then decoded and matched into Baseline and One-Year examinations for 20 subjects. For the standard panel and for each of the eleven test examiners, net caries increment scores were calculated for each subject.

-Based on the increment scores, weighted kappa statistics were calculated for each test examiner vs. the panel of standard examiners.

-Standard diagnostic test statistics were calculated from the combined Baseline and One-Year data. A total of 1120 surfaces were included. Sensitivity (the percent of carious surfaces as determined by the standard panel that were also determined as carious by the test examiner) and Positive Predictive Value (the percent of carious surfaces as determined by the test examiner that were also determined as carious by the standard panel) were of most interest

SUBJECTS

The test set of bitewing radiographs was selected from a one year clinical caries study involving 150 subjects aged 9-12 years who had four radiographs taken at both the baseline and one year examinations.

RESULTS

-The 3-clinician standard panel detected a combined total of 88 lesions at baseline and at one year.

-The sensitivity of the eleven examiners ranged from 66% to 100%, while the number of false positive scores ranged from 4 to 119 representing positive predictive values ranging from 42% to 94%.

-Examiners with high sensitivity yet low positive predictive values might be considered overly liberal in their caries determinations. They have high sensitivity because they score any suspect area as caries, thus they correctly call nearly all truly carious lesions.

Research presented at the 30th Annual Meeting of the AADR, March 7-10, 2001

P&G

Examiner Calibration Exercise for the Assessment of Radiographic Caries

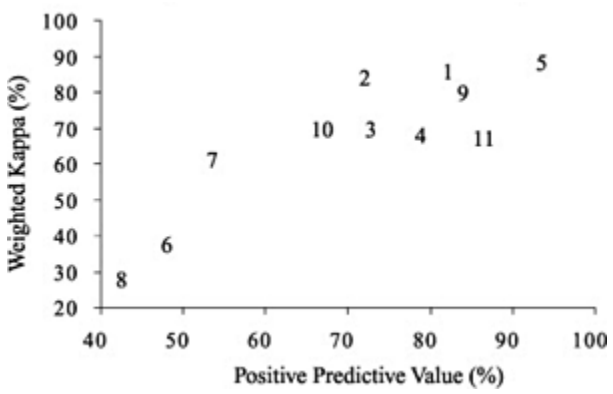
J.M. McClary*, P.A. Walters, S.A. Jacobs, R.W. Gerlach, R.D. Bartizek, J.L. Winston, A.R. Biesbrock
Procter & Gamble Co., Mason, OH, USA

-Examiners with high positive predictive values yet low sensitivity might be considered overly conservative in their caries determinations. They have high positive predictive values because they only score obvious caries, thus they rarely miscall a truly sound surface. The standard examiner panel was conservative in their caries determinations.

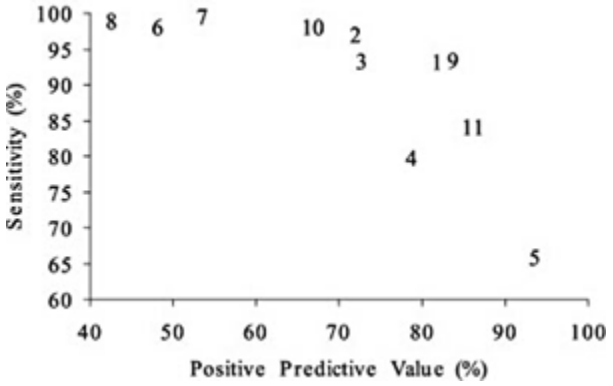
-Because of the matched nature of the radiograph sets, radiographic increment over one year could also be determined. Based on the increment observed by the standard panel, weighted kappa statistics were calculated for each of the eleven examiners. The weighted kappa values ranged from a low of 27% to a high of 88%.

-Examiners with a combination of relatively high sensitivity, positive predictive values, and weighted kappa values have demonstrated the best agreement to the standard examiner panel.

Weighted Kappa vs. Positive Predictive Value for 11 Examiners
(Referenced vs. 3 Standard Examiners)



Sensitivity vs. Positive Predictive Value for 11 Examiners
(Referenced vs. 3 Standard Examiners)



CONCLUSION

-Experienced caries examiners were found to differ widely in their level of agreement with a standard examiner panel.

-Examiners who were liberal in their determinations, who were conservative in their determinations, and who agreed well with the standard panel were identified.

-This model has promise for assessing examiner sensitivity to radiographic caries detection.

Research presented at the 30th Annual Meeting of the AADR, March 7-10, 2001

P&G