Evaluation of computer aided learning in developing clinical decision-making skills

E. J. Kay,[1] B. Silkstone,[2] and H. V. Worthington,[3]

Aim The aim of this study was to determine whether an educational intervention delivered by a computer aided learning package improved the sensitivity and specificity of dentists' restorative treatment decisions
Method The study was a randomised controlled study using a Solomon three-group design. Ninety-five dentists were randomly allocated to the three study groups. One group of dentists read the radiographs pre and post an educational intervention, a second group read the radiographs once, after the intervention, and a third group read the radiographs twice, but received no intervention. On each occasion the dentists read 24 surfaces on each of 15 radiographs and made 360 decisions on how certain they were about restoring the tooth surface. Comparisons of mean sensitivity, specificity and areas under ROC curves were made within and between the study groups. Kappa values were used to assess changes in the level of agreement between dentists.
Results There were no significant changes in sensitivity, specificity or area under ROC curves caused by the intervention. There was no evidence that the level of agreement between the dentists improved after the intervention.
Conclusion A computer aided learning package had no effect on dentists' treatment decision-making behaviour.

Research has shown large variations occur between dentists in their diagnoses, restorative decisions and treatment plans.[1–4] A previous study by Kay indicated that a key parameter in dentist's decision making process might be the way in which costs and benefits of false negative and false positive decisions are weighed against each other.[5] Other work has also suggested that if an individual understands the trade-off between sensitivity and specificity then they make fewer false positive and false negative choices.[6–7] A previous study, which developed the hypothesis tested in the research reported here, indicated that the accuracy of dentists' treatment decisions could be improved by educating practitioners about the use of probabilistic reasoning when making treatment decisions.[8]

The aim of this study was to determine whether an educational intervention delivered by a computer aided learning package improved the sensitivity and specificity of dentists' restorative treatment decisions. If the research showed that the new approach was effective in improving the validity of decisions, a teaching package was to be developed and widely disseminated.

[1]Professor of Dental Health Services Research, [2]Research Assistant, [3]Reader in Statistics, University Dental Hospital of Manchester, Higher Cambridge Street, Manchester, M15 6FH
*Correspondence to: E J Kay
email: E.Kay©man.ac.uk

The specific objectives of the study were to examine the effect of a previously piloted educative intervention on the sensitivity and specificity and variability of dentists' treatment decisions and to develop the educative intervention into an easily disseminated teaching package.

In order to achieve the aims and objectives of the study, two research hypotheses were tested. These were:

1. The intervention will increase the sensitivity and specificity of dentists' restorative treatment decisions.
2. The intervention will increase the agreement between dentists in the clinical treatment decisions made by them.

Method
Study material
The study used material constructed and previously reported by the lead author.[3,5,7] These were 15 sets of simulated bitewing radiographs. The radiographs were of models which simulated the posterior part of a natural dentition, using extracted teeth mounted in a medium which simulated human bone. One fifth of the teeth included in these simulated dentitions had some visible evidence of caries, ranging from incipient lesions to large cavities. After radiographs of the models had been taken, using a technique precisely equivalent to normal bitewing radiography, the teeth were serially sectioned into 300μm thick longitudinal sections. These sections were then examined for caries using a microscope and each approximal surface was graded on a 5-point scale:

0 – Sound
1 – Lesion limited to outer half of enamel
2 – Lesion more than half way through enamel but not penetrating dentine
3 – Lesion into dentine but not more than half way to pulp
4 – Lesion into dentine and more than half way to pulp

The radiographs were presented to the participating dentists along with an appropriate patient vignette, eg 'this radiograph is of a patient who you will see again within the next two years. His/her diet and oral hygiene are 'average'- that is, better than your worst patient but worse than your best patient.'

The participating dentists were asked to make treatment decisions about the approximal surfaces of the teeth on the radiographic films, grading their decisions on a 6-point scale:

1 – Definitely place a restoration
2 – Probably place a restoration
3 – Might consider placing a restoration
4 – Might consider leaving the tooth unrestored
5 – Probably leave the tooth unrestored
6 – Definitely leave the tooth unrestored

Since each radiograph contains information on 24 surfaces, each dentist made 15 x 24= 360 decisions. The analysis is therefore based

on the comparison of more than 30,000 decisions made before or after the educational intervention.

Study design

The study used a Solomon three-group design. After recruitment the participating dentists were randomly assigned to one of three groups as outlined below and described in Table 1.

Group 1 were asked to view the radiographs and made treatment decisions using the scale described above. On the subsequent day, they used the CAL package and then re-read the radiographs.

Group 2 used the CAL package prior to a single reading of the radiographs.

Group 3 read the radiographs and on the subsequent day re-read them. The procedure was the same as for group 1, except they did not utilise the CAL package until after the second viewing of the radiographs.

With O1 to O5 as defined in Table 1, the following effects were compared:

O2 – O1	Effect of intervention and dual reading
O2 – O3	Effect of dual reading and differences in variability between groups
O2 – O5	Effect of intervention and differences in variability between groups
O5 – O4	Effect of dual reading
O4 – O1	Effect of differences in variability between groups

Table 1 Summary of the research design, where 0 = observation and X = intervention

	Group 1	Group 2	Group 3
Read radiographs	O1		O4
Participate in seminar	X	X	
Read radiographs	O2	O3	O5

Educational intervention

The intervention used was a computer aided learning (CAL) package, which had the following teaching objectives:

After using the CAL programme, participants would understand:

1. The impossibility of perfection in the recognition of pathology.
2. The potential effects of uncertainty on the delivery of quality patient care.
3. That judgement involves accepting the possibility of making detrimental as well as beneficial decisions.

4. That the values and costs associated with any treatment decision relates to: the consumers views, the probability of success of that treatment, the acceptability of both the outcome and the process which occur subsequent to the decision.

Statistical methods

Sensitivity and specificity of the dentist's decisions to restore teeth were calculated at two thresholds of diagnosis, "definitely restore" versus all other ratings (T1), and 'definitely', 'probably', or 'might consider' restoring the tooth versus 'definitely', 'probably', or 'might consider' leaving the tooth unrestored (T2). These calculations used the microscopic criteria of caries penetrating dentine as the gold standard. Area under ROC curves were derived for each dentist in each group at all time points. Comparisons between the different groups were made by independent sample t-tests, and comparisons within the same group made, using paired t-tests. Kappa statistics were computed to assess the level of agreement between dentists within a group.

Results

Thirty-one dentists participated in group 1, 33 in group 2, and 31 in group 3. However, the majority of the data from one dentist in group 3 was unusable. The data from this dentist was, therefore, excluded.

The microscopic examination of the 360 tooth surfaces showed that 244 (68%) were sound and 52 (14%) had caries into enamel and 44 (12%) caries into dentine. Two surfaces (1%) had restorations and data for 18 (5%) were not available owing to the tooth shattering on sectioning.

The mean sensitivity and specificity values for threshold T1 for each study group are shown in Table 2. The initial mean sensitivity value for the dentists in group 1 was 0.318 and this increased to 0.336 after the intervention, the initial mean specificity value of 0.946 increased to 0.950, however these changes were not significant ($P = 0.50$; $P = 0.56$).

Table 3 P-values for the planned comparisons for sensitivity and specificity, at both thresholds (T1, T2) and area under the ROC curve

	Dentists decisions Threshold T1		Dentist's decisions Threshold T2		Area under ROC curve
	Sensitivity	Specificity	Sensitivity	Specificity	
O2 – O1*	0.50	0.56	0.98	0.11	0.24
O2 – O3	0.33	0.44	0.66	0.33	0.20
O2 – O5	0.81	0.20	0.59	0.83	0.53
O5 – O4	0.42	0.02	0.43	0.03	0.19
O4 – O1	0.89	0.53	0.94	0.66	0.51

*O1–O5 are defined in Table 1

Table 2 Mean (standard deviation) of the sensitivity and specificity values for the dentists in each study group, compared with the gold standard, where the dentist's decisions are at threshold T1

Group	Number	Pre Intervention or first time point (O1, O4)		Post intervention or second time point (O2, O3, O5)		P-value‡	
		Mean sensitivity (sd)†	Mean specificity (sd)	Mean sensitivity (sd)	Mean specificity (sd)	sensitivity	specificity
1	31	(O1) 0.318 (0.130)	(O1) 0.946 (0.045)	(O2) 0.336 (0.163)	(O2) 0.950 (0.037)	0.50	0.56
2	33			{O3} 0.299 (0.133)	(O3) 0.946 (0.040)		
3	30*	(O4) 0.313 (0.143)	(O4) 0.952 (0.029)	(O5) 0.326 (0.140)	(O5) 0.961 (0.026)	0.42	0.02

*There were insufficient readings for one dentist
†Standard deviation
‡P-value from paired t-test comparing pre versus post/second time point

Table 4 Mean (standard deviation) of sensitivity and specificity values for the dentists in each study group, compared with the gold standard, where the dentist's decisions are at threshold T2

Group	Number	Pre Intervention or first time point (O1, O4) (O2, O3, O5)		Post intervention or second time point		P-value‡	
		Mean sensitivity (sd)†	Mean specificity (sd)	Mean sensitivity (sd)	Mean specificity (sd)	Sensitivity	Specificity
1	31	(O1) 0.519 (0.130)	(O1) 0.852 (0.103)	(O2) 0.520 (0.140)	(O2) 0.876 (0.074)	0.98	0.11
2	33			(O3) 0.493 (0.134)	(O3) 0.857 (0.083)		
3	30*	(O4) 0.522 (0.130)	(O4) 0.862 (0.083)	(O5) 0.539 (0.130)	(O5) 0.880 (0.082)	0.43	0.03

*There were insufficient readings for one dentist
†Standard deviation
‡P-value from paired t-test comparing pre versus post/second time point

Table 5 Mean area under the ROC curve for the dentists in each study group, compared with the gold standard

Group	Number	Pre intervention or first time point (O1, O4)	Post intervention or second time point (O2, O3, O5)	P-value
1	31	(O1) 0.729 (0.063)	(O2) 0.744 (0.058)	0.24
2	33		(O3) 0.720 (0.085)	
3	30	(O4) 0.738 (0.052)	(O5) 0.752 (0.036)	0.19

The mean sensitivity values and specificity values for all the dentists in groups 1, 2 and 3, were similar at all time points, for this threshold T1, with the sensitivity ranging from 0.299 to 0.336, and the specificity ranging from 0.946 to 0.961 (Table 2). All the P-values obtained from the specified comparisons were non significant apart from a significant increase in the mean specificity of the two readings conducted by the dentists in group 3 (Table 2; $P = 0.02$). The P-values obtained from the specified comparisons are detailed in Table 3 and were all non-significant.

The results for the other threshold T2, were similar. For the dentists in group 1 there was an increase in the initial to post intervention mean sensitivity from 0.519 to 0.520 and in the specificity from 0.852 to 0.876, which were not significant ($P = 0.98$; $P = 0.11$). The mean sensitivity values and specificity values for all the dentists were similar at all time points, the sensitivity ranging from 0.493 to 0.539, and the specificity ranging from 0.852 to 0.880 (Table 4).

There was a small non-significant increase in the mean area under the ROC curves for the dentists in group 1, from 0.729 to 0.744 (Table 5). The range of values for the mean areas under the ROC curves for each group of dentist at both time points was small, ranging from 0.720 to 0.752.

The results of calculating kappa for each of the categories separately against a combination of the remaining categories are shown in Table 6. For example, the initial kappa value for the dentists in group 1, comparing 'definitely restore' with the other decision categories is 0.44. There is fair to moderate agreement at this threshold with kappa values ranging across groups and time points from 0.370 to 0.512. Comparing the 'definitely leave unrestored' category to the other categories, the range of kappa values was 0.276 to 0.326, which is considered fair agreement. Although the standard errors of kappa cannot be calculated when the number of readings per dentist varies, there was no evidence that the level of agreement between the dentists improved after the intervention.

Discussion

The project fulfilled all of its objectives. However, the educational intervention was found not to affect the sensitivity and specificity or the agreement between dentist's treatment decisions.

The range of mean sensitivity, specificity, and kappa values which assessed the agreement between dentists, were narrow showing that the results were consistent between the study groups, and for different time points within the study groups. The study had sufficient power to detect small differences in all the indices measured. For example a change of 0.1 in sensitivity, 0.03 in specificity, and 0.03 in

Table 6 Kappa statistics for the agreement between the dentists in each group, for each decision category

Outcome		Group 1		Group 2	Group 3	
		Pre intervention	Post intervention	Post intervention	First reading	Second reading
1	(definitely restore)	0.442	0.488	0.370	0.449	0.512
2		0.089	0.069	0.065	0.079	0.081
3		0.057	0.054	0.059	0.052	0.062
4		0.023	0.020	0.017	0.013	0.008
5		0.026	0.024	0.032	0.044	0.043
6	(definitely leave unrestored)	0.276	0.301	0.283	0.276	0.326
Combined		0.193	0.210	0.190	0.194	0.229

the area under the ROC curve would have been detectable with 95% power with the sample size of 30 dentists per group.

The results are surprising since previous evidence had strongly suggested that information about uncertainty enhanced the accuracy and consistency of dentists' restorative treatment decisions.[8] In that study, the intervention had been delivered as a seminar and a trend towards improved decision making was clearly shown. However, the power of the previous study was questionable. The results described in this report are more reliable and a definite lack of effect of the intervention was demonstrated. The discrepancy between the results of the two studies can be explained in two ways. The first and most plausible explanation is that the previous study's small sample size allowed any changes in the dentists' treatment decisions after the intervention to be overstated and that the current study has revealed the true lack of impact of education about uncertainty. A second, less plausible, but possibly more interesting explanation might be that delivery of information via a personally delivered interactive seminar is more effective for improving dentists' decision making skills than a computer aided learning package. The educational intervention in the current study was delivered via a CAL package for two reasons. Firstly, it was delivered by computer in order to ensure standardisation across the different experimental groups. Secondly, it was important to have a method of education which could be easily and widely distributed and used, should it prove to be effective.

The study has clearly shown that education about uncertainty delivered by a computer aided learning package had no effect on dentists' decision making. This finding, especially when viewed in the light of the previous study, has very important implications, given the burgeoning use of computer packages in education. The only difference between the two studies was the mode of delivery of the information. The information, study design, materials and methods were exactly the same. Further research is urgently needed to directly compare education delivered by CALs and by personnel. The findings of this study have enormous implications for education.

Conclusions

1. Further study of the psychology of treatment decision making is needed in order to determine the key factors influencing dentists' choices.
2. Direct comparisons of computer aided learning and traditional education are needed before the ease of distribution of such packages causes potentially ineffective educational methods to overtake traditional ones.

1 Elderton R J, Nuttall N M. Variation among dentists in planning treatment. *Br Dent J* 1983; **154**: 201-206.
2 Kay E J, Watts A, Paterson R C, Blinkhorn A S. An investigation into the validity of dentists' decisions to restore the occlusal surfaces of permanent teeth. *Comm Dent Oral Epidemiol* 1988; **16**: 91-94.
3 Kay E J, Nuttall N M. Knill-Jones R. Restorative treatment thresholds and agreement in treatment decision making. *Comm Dent Oral Epidemiol* 1992; **20**: 265-268.
4 Nuttall N M, Pitts N B. Restorative treatment thresholds reported to be used by dentists in Scotland. *Br Dent J* 1990; **169**: 119-126.
5 Kay E J, Nuttall N M. The relationship between the dentists' attitudes and their restorative treatment decisions planned on the basis of bitewing radiographs. *Comm Dent Oral Epidemiol* 1993; **21**: 43-46.
6 Hanley J A, McNeil B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29-36.
7 Kay E J, Knill-Jones R. Variations in restorative treatment decisions: application of ROC analysis. *Comm Dent Oral Epidemiol* 1992; **20**: 113-117.
8 Choi B C K, Jokovic A, Kay E J. Reducing variability in treatment decision making-effectiveness of educating clinicians about uncertainty. *Med Educ* 1998; **32**: 105-111.