Multilevel modelling of dental caries clinical trial data

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Girvan Burnside

October 2009

Abstract

Data from dental caries clinical trials has a naturally hierarchical structure, with surfaces clustered within teeth, clustered within individuals and individuals potentially also clustered within social or administrative units. Multilevel modelling allows analysis of clustered data using individual observations without aggregating data, but has been little used in the field of dental caries. The aim of the work in this thesis is to investigate the use of multilevel modelling in the analysis of data from clinical trials of agents designed to prevent dental caries.

The statistical methodology of recent caries clinical trials is assessed, with particular emphasis on appropriate analysis of clustered data. Multilevel models are fitted to a clinical trial data set, with various model specifications. The use of multilevel models of clinical trial data to predict tooth and surface specific caries incidence is explored. A simulation study investigates the power of multilevel modelling compared to traditional analysis methods.

Several cluster randomised caries trials have been published with analysis which incorrectly ignores the clustering. The hierarchical nature of caries data was rarely considered in trial analysis.

Multilevel modelling has the advantage over traditional analyses of allowing greater understanding of the patterns of caries development within the mouth.

Multilevel modelling of caries clinical trial data can also provide clinically useful methods of predicting tooth and surface specific caries incidence, based on baseline caries patterns. In the data set analysed, caries on the contralateral surface (the corresponding surface on the opposite side of the mouth), was a stronger predictor than caries in the corresponding surface on the opposing jaw, or caries on an adjacent tooth.

Multilevel modelling using the natural hierarchy of surfaces and teeth clustered within individuals may not allow significant reductions in the number of participants required in a caries clinical trial, compared to the use of traditional analyses, but investigators interested in exploring the effect of their intervention in more detail should consider the application of multilevel modelling to their clinical trial data.

Acknowledgements

I would like to thank my supervisors, Cynthia Pine and Paula Williamson for their guidance and support throughout the course of this PhD project.

I would also like to thank Rebecca Harris and my other colleagues in Dental Public Health, for their support and encouragement.

I am very grateful to Unilever Oral Care, and particularly Tony Roberts, for a generous research donation.

Thanks also to Valeria Marinho for supplying copies of many papers for the review in Chapter 4, Marta van der Hoek and Gleb Komarov for translating some of these papers, Mark Gilthorpe for his very useful comments on my work, and to many other colleagues in Liverpool Dental School, and the Centre for Medical Statistics and Health Evaluation for their help and advice.

Publications

Work included in this thesis has been published in the following peer-reviewed journal articles.

Chapter 4

Burnside G, Pine CM, Williamson PR. Statistical aspects of design and analysis of clinical trials for the prevention of caries. Caries Research. 2006;40(5):360-5.

Sections of Chapters 5 and 7

Burnside G, Pine CM, Williamson PR. The application of multilevel modelling to dental caries data. Statistics in Medicine. 2007 Sep 30;26(22):4139-49.

Chapter 6

Burnside G, Pine CM, Williamson PR. Modelling the bilateral symmetry of caries incidence. Caries Research. 2008;42(4):291-6

Table of contents

1. Introduction

The overall aim of the work undertaken for this thesis is to investigate the use of multilevel modelling in the analysis of data from clinical trials of agents designed to prevent dental caries. Specific research questions which will be addressed in this thesis are:

- Is multilevel modelling an appropriate method of analysis in dental caries clinical trials?
- Does the use of multilevel modelling have the potential to increase efficiency in clinical trials of caries preventive agents?
- Can multilevel analysis of data from caries clinical trials be used to develop models of tooth and surface specific caries incidence?

Dental caries, commonly known as tooth decay, has been identified by the World Health Organisation as a major oral health problem in industrialised countries, affecting 60-90% of schoolchildren, and the vast majority of adults (Petersen, 2003). Survey data from the USA has shown caries to be the most prevalent chronic childhood disease, despite being largely preventable (Dye et al., 2007).

If caries is left untreated, there can be an increased likelihood of sepsis, in the form of dental abscesses extending into the gingival tissue adjacent to the carious lesion. In

addition to pain and discomfort, these infections can, in very rare acute cases, result in serious problems such as orbital cellulitis and brain abscesses (Pine et al., 2006).

Dental caries data is collected during a clinical examination, where each surface on each tooth is assessed by the examiner as to whether it is sound, decayed or filled, or if the tooth is missing as a result of extraction due to caries. These data are traditionally aggregated into a single measure for each individual, known as the DMF (decayed, missing or filled) index. This approach means that the tooth and surface specific information is lost.

If these data are to be analysed by tooth or surface, appropriate statistical methods must be used to correctly adjust for the clustered nature of the data. Another level of clustering can be introduced where studies are randomised by allocating groups of participants, such as schools or general practices, to an intervention. Failure to use appropriate statistical methods to analyse clustered data can lead to incorrect conclusions.

Multilevel modelling is a statistical method which allows analysis of clustered data using the individual observations, without having to aggregate the data. This technique is common in many areas, but has been little used in the analysis of dental caries data.

The remainder of this thesis is structured as follows.

In Chapter 2 background on the disease process of dental caries is presented, and the main approaches to prevention of the disease are discussed. The use of caries as an outcome variable in statistical analysis, represented by the DMF index is described. The remainder of the chapter discusses the literature on statistical analysis of caries clinical trials, and considers suggestions which have been made to improve the efficiency of caries clinical trials.

In Chapter 3, the concept of clustered data is introduced, and the particular issue of clustering in caries data is described. Published work on methods suggested to analyse caries data at tooth and surface level are considered. The multilevel modelling method is discussed, with a review of how this has been used with periodontal data, which has a similar multilevel structure to caries data.

In Chapter 4, recent publications from clinical trials of caries preventive agents are considered, to assess whether the trials were cluster randomised, and if so, whether an appropriate method of analysis was used. Analysis is undertaken to estimate the degree of clustering which would be required to result in inaccurate conclusions being drawn from the incorrectly analysed trials. This chapter also assesses the quality of reporting of trials according to the CONSORT guidelines.

In Chapter 5, multilevel modelling is applied to a data set from a caries clinical trial of a caries preventive agent in 12-16 year olds. Different methods of estimation in multilevel modelling with binary outcome variables are described and compared. Methods of modelling the effect of tooth position within the mouth are investigated.

In Chapter 6, multilevel modelling is used on a caries clinical trial data set to investigate the prediction of caries incidence on specific teeth and surfaces based on the baseline caries experience of other related surfaces, such as the contralateral surface, the corresponding surface in the opposing jaw, and surfaces on adjacent teeth.

In Chapter 7, the model fit of the multilevel models is investigated. Methods of assessing model fit are considered, and the model fit of various model specifications is compared.

In Chapter 8, a simulation study is conducted to investigate the performance of multilevel modelling methods and standard analysis using caries increment. Data sets

are simulated from a three level binomial distribution with various sample sizes, treatment effects and random tooth level effects, to investigate robustness of two-level analysis and comparative power of multilevel models and traditional analysis.

In Chapter 9, the main findings of the thesis are discussed, and recommendations are made to investigators on the use of multilevel modelling for the statistical analysis of caries clinical trial data. Finally, some opportunities for further work are discussed.

2. Statistical analysis of dental caries data

2.1. Introduction

One of the research questions which will be addressed in this thesis is whether the use of multilevel modelling techniques has the potential to increase efficiency in clinical trials of caries preventive agents. In addition, the thesis considers how multilevel analysis of data from caries clinical trials can be used to model patterns of caries incidence.

This chapter first reviews the literature to provide a brief description of the disease process of dental caries, and considers the main approaches to prevention of the disease. This is followed by a review of literature on how caries is measured as an outcome variable using the DMF (decayed, missing or filled) index, and on the statistical analysis of this index. The literature on methods of analysis of clinical trials with caries outcome variables is also reviewed. The effect of changes over time in the prevalence of dental caries on the design and analysis of clinical trials is considered, with suggestions from the literature on improving efficiency of caries clinical trials. This leads into Chapter 3, which will discuss issues relating to clustered data.

2.2.Dental Caries

Dental caries is a process in which the structure of a tooth is attacked by acid generated by micro-organisms as a by-product of metabolism (Kidd, 2005). Over time, micro-organisms form a deposit, known as plaque, which adheres to the surfaces of the teeth. The presence of plaque alone is not sufficient to cause the disease. The caries process also requires a substrate, in the form of a suitable dietary carbohydrate such as sucrose or glucose. Some of the micro-organisms in the plaque can ferment this carbohydrate to produce acid, resulting in a rapid drop in the pH of the plaque. Repeated exposure to this acidic environment can cause demineralisation of the enamel on the tooth surface, producing a carious lesion. Caries is a continuous process, beginning with the demineralisation of a small area in the outer enamel layer of the tooth, which can progress into the dentine below, potentially resulting finally in the destruction of the tooth. In the early stages of the caries process, neutralisation of the acid by saliva can result in mineral being regained, and the progress of the lesion can reverse, a process known as remineralisation.

The process of development of a caries lesion has been described as a balance between protective and pathological factors (Featherstone, 2004). The pathological factors are cariogenic bacteria, salivary dysfunction, and frequency of ingestion of fermentable carbohydrates. The protective factors include salivary flow rate, many of the components of saliva, substances that stimulate salivary function, and fluoride. For caries to progress, the pathological factors need to outweigh the protective factors. Interventions seeking to reduce caries attempt to tip the balance in the other direction either by reducing the effect of the pathological factors, or by increasing the effect of the protective factors.

2.3. Prevention

Caries is a largely preventable disease, as a combination of reducing the frequency of ingestion of fermentable carbohydrates, particularly sugar, and increasing fluoride intake will in most cases be enough to weigh the balance in favour of the protective factors.

There is much evidence that fluoride has a preventive effect on dental caries. The association between fluoride and decreased caries prevalence was first observed in 1931 (Churchill, 1931), in a study comparing regions with different levels of fluoride content in the water supply. Today, by far the most common method of delivery of fluoride is in toothpastes. Since the early 1970s, caries prevalence has fallen markedly in most developed countries (Renson, 1986), and many experts believe this to be largely due to the widespread introduction of fluoride toothpaste in this period (Bratthall et al., 1996). A Cochrane review of the effectiveness of fluoride toothpastes for preventing dental caries in children and adolescents showed a pooled prevented fraction (PF) of 24% in number of decayed, missing of filled tooth surfaces when fluoride toothpastes are compared with placebos (Marinho et al., 2003b). Other approaches to the topical delivery of fluoride include mouthrinses (PF of 26%, (Marinho et al., 2003c)), gels (PF of 28%, (Marinho et al., 2002a)), and varnishes (PF of 33%, (Marinho et al., 2002b)). A Cochrane review combining these four delivery methods showed an overall prevented fraction of 26% for topical fluoride therapy compared to placebo or no treatment. Studies of the use of fluoride gels, mouthrinses, or varnishes in addition to toothpaste showed a small advantage over the use of toothpaste alone, with an additional prevented fraction of 10% (Marinho et al., 2004b). A review of trials comparing the different methods of delivery to each other proved inconclusive on whether any one method was superior to the others (Marinho et al., 2004a).

Another common approach to caries prevention is the pit and fissure sealant. These are coatings applied to the pit and fissure surfaces of the molar teeth of children, with the intention of preventing the growth of bacteria on these surfaces, which are the most susceptible to caries. A Cochrane review of trials of fissure sealants found a significant difference in favour of specific types of resin-based sealants compared to control, with a pooled relative risk of 0.13 for increased caries in the occlusal surfaces of first permanent molars at 12 months follow up (Ahovuo-Saloranta et al., 2008).

Other approaches to caries prevention include the use of antibacterials such as chlorhexidine, which has been shown to kill caries-associated bacteria, although there is little evidence for its efficacy in preventing caries (Forgie et al., 2000).

Many approaches to prevention focus on changing behaviour, particularly using oral health education to promote increased toothbrushing frequency, and the reduction of the frequency of sugar intake in the diet (Adair and Ashcroft, 2007).

Although levels of dental caries have declined since the widespread introduction of fluoride toothpaste in the early 1970s (Renson, 1986), the disease is still a significant public health problem in most of the developed world, with 60-90% of schoolchildren in industrialised countries affected (Petersen, 2003). Several studies have shown that caries risk is higher in more deprived areas (Ellwood and O'Mullane, 1996; Jain et al., 2007; Provart and Carmichael, 1995)

2.4. Measurement of caries

The traditional way to record caries is from a visual clinical examination by a trained dental examiner. Many guidelines and criteria systems have been published, and a review article compares 29 sets of criteria for caries detection (Ismail, 2004). Common

guidelines for epidemiological surveys recommend the use of a mirror and blunt-ended probe (Pitts et al., 1997).

On each tooth in the mouth, there are either four or five surfaces which can be assessed for caries (four on the anterior teeth, which have sharp incisal edges, and five on the posterior teeth). These surfaces can be classified into smooth surfaces, and surfaces which have pit and fissure systems. A clinical examination for dental caries entails the examiner inspecting each surface on each tooth individually and classifying it according to whether it is sound, has caries, is filled, has been extracted, etc. Epidemiological surveys usually score a surface as having caries if the caries is judged to have extended into the dentine. For example the World Health Organisation criteria for recording caries state that "Caries is recorded as present when a lesion in a pit or fissure, or on a smooth tooth surface, has an unmistakable cavity, undermined enamel, or a detectably softened floor or wall" (WHO, 1997).

In clinical trials, caries is often also recorded for non-cavitated lesions, which are confined to the enamel, and do not extend into the dentine. These early carious lesions can give a more sensitive measure of the onset of the disease (Pitts, 2004).

The clinical visual examination forms the basis of most caries assessments, but has been supplemented by several other techniques.

The traditional caries clinical trial has often used bitewing radiographs to supplement the visual assessment (Wenzel, 2004). These radiographs can detect demineralisation that may not be visible to the naked eye.

Another method which has been used in several clinical trials is fibre-optic transillumination (FOTI) (Mitropoulos, 1985). FOTI is a non-invasive procedure where a narrow beam of bright light is directed to the surfaces of teeth which contact each

other (the approximal surfaces). Areas of demineralisation will deflect the light beam and cause shadows, allowing diagnosis of lesions which would be difficult to detect visually without this aid.

Other diagnostic methods have been used to detect early caries, including DIAGNOdent, a laser examination tool (Lussi et al., 1999), and Quantitative Light-Induced Fluorescence (QLF), a tool which assesses the change in auto-fluorescence of teeth to detect caries (Stookey, 2004).

2.5. Analysis of caries data

2.5.1. The DMF Index

The degree to which an individual is affected by caries is commonly measured using the DMF index, which has been in use since the 1930s (Klein and Palmer, 1937). DMF stands for decayed, missing or filled, and the index counts either the number of teeth which are decayed missing, or filled (DMFT), or the number of surfaces (DMFS). The missing and filled components of the index should only include those teeth or surfaces which are missing or filled due to the effects of caries. The decayed, missing and filled components of the index can be considered separately, but together are referred to as caries experience, as the missing and filled teeth or surfaces, though not affected by caries at the present time, have experienced the disease in the past.

Epidemiological surveys around the world use estimates of the mean DMFT in a population as a measure of dental health. The World Health Organisation publishes mean DMFT estimates for 12-year-olds as a comparison of dental health between countries (Petersen, 2003).

2.5.2. Subjectivity of measurement

The measure is dependent on the examiner, who must make judgments on the status of each tooth or surface. Caries is a continuous process, with the structure of the tooth gradually decaying, although lesions in the early stages, sometimes presenting as white spots in the enamel, can reverse, a process known as remineralisation, or remain static without further progression. For an examiner to decide whether to classify a tooth or surface as decayed, a threshold must be set beyond which a classification of decayed is given. A common threshold is where caries extends into the dentine (often known as the D_3 threshold). This can encompass caries where a cavity is present in the tooth, and also where caries in dentine can be seen as a shadow below the enamel. If the earlier stages of the disease are to be included, the threshold can be set to include caries only in enamel, which has not yet extended into the dentine. This is often known as the D_1 threshold (WHO, 1997).

As these classifications are based on examiner judgement, there can be variation in how teeth and surfaces are classified both between and within examiners. Clinical studies involving several examiners will usually incorporate examiner training in an attempt to minimise this variation (Pine et al., 1997). The strength of agreement between examiners is usually measured using kappa statistics, which measure the observed proportion of agreement, corrected for the proportion of agreement which is expected by chance alone (Fleiss et al., 1979). This method is also used to assess intra-examiner agreement, where an examiner has re-examined a proportion of participants.

2.5.3. Distribution of DMF data

The analysis of dental caries data has mostly used parametric methods, assuming that the distribution of the DMF index approximates a normal distribution (although as

DMF data is count data, and not continuous, it cannot truly follow a normal distribution). However, the distribution of DMF within many populations has high degrees of skewness, as shown in a typical data set in Figure 2.1. As caries prevalence has decreased over time, many more people remain caries free, resulting in the large peaks at dmft=0. The skewed nature of the data has been recognised as an issue for many years. In the 1970s a study used simulation to investigate whether Analysis of Variance was appropriate for the analysis of dental clinical trials and found that ANOVA gave slightly conservative tests of significance on the simulated data, and that conclusions from ANOVA could be accepted with no increase in probabilities of type I or type II errors (Glass et al., 1972).

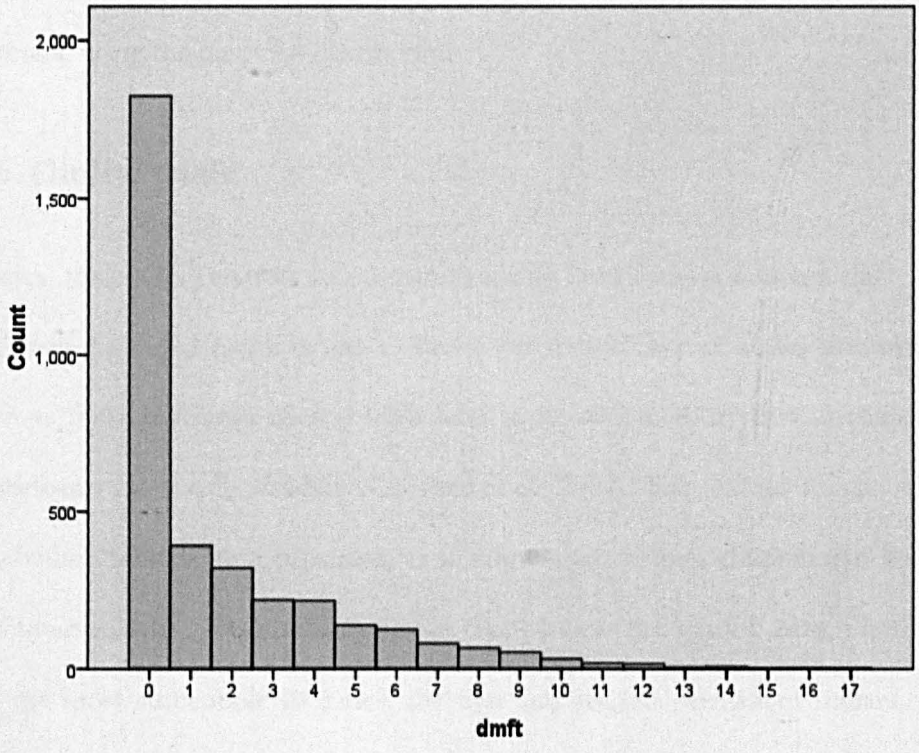

Figure 2.1 : Typical distribution of dmft in a 5-year-old population

Since this time, as caries prevalence has reduced further, there have been many suggestions in the literature that analysis of caries data should not use normal

distribution based methods. In particular, researchers have investigated the suitability of the Poisson and negative binomial distributions (Fabien et al., 1999; Worthington, 1984). These distributions are not always accurate fits to caries data sets, and some modifications have been suggested. One study considered DMFT data from a Brazilian caries prevention study (Böhning et al., 1999). The standard Poisson model did not adequately fit the data due to an excess of zeroes in the data set, so the authors used the zero-inflated Poisson distribution. This distribution essentially models the population as two sub-populations, one following a Poisson distribution, and one with all zero values. Another similar approach was taken in a study analysing a data set of UK children (Lewsey et al., 2000) which recommended employing a Poisson or negative binomial model for those children with caries, and modelling the caries free/caries present outcome using the binomial distribution.

2.6. Clinical trials

Clinical trials with caries as an outcome variable have been conducted since the 1950s, and have provided much of the evidence for the efficacy of caries preventive agents such as fluoride. Caries clinical trials tend to be two to four years in length, and the participants are usually children (Chesters et al., 2004). The reasons for the recruitment of children include ease of access, as if schools are willing, children can be recruited, and interventions and examinations can take place at the school. Also, the teeth which are the most susceptible to caries, the first and second permanent molars, will erupt while the child is school age. These teeth are most at risk of developing caries within the first few years after eruption. The length of study will usually be at least two years, to give carious lesions sufficient time to develop.

2.6.1. Analysis of caries data in clinical trials

In clinical trials with caries as an outcome, the interest is in the comparative number of carious lesions which have developed during the course of the trial. Therefore, rather than DMFT or DMFS, the outcome variable will be the number of teeth or surfaces which were sound at the baseline examination, and affected by caries at follow-up. This is equivalent to the change in DMF index, and is known as caries increment. Caries increment was first suggested as a measure of caries incidence in a study of schoolchildren in 1946 (Boyd and Cheyne, 1946).

As caries clinical trials tend to be two to four years in length, this represents the number of new lesions which have developed over a two to four year period. Therefore the skewed distributions found in DMF prevalence data will also apply to caries increment, often to a greater degree, as the number of zeroes is likely to be even higher than for prevalence data, as surfaces already decayed, which would contribute to the DMF score, will have a zero value for caries increment.

As with DMF prevalence data, the Poisson distribution has been suggested for modelling increment data. An article on current knowledge on statistical methodology for caries clinical trials published in 1984 observed that if caries increment levels were low, the Poisson distribution may provide a good fit to incremental data (Worthington, 1984). A reanalysis of a clinical trial data set using Poisson regression models (Hujoel et al., 1994a) confirmed the original result of the trial, and concluded that these models may have advantages over standard ANOVA based analyses as they include the concept of time at risk, using data from intermediate examinations.

2.6.2. Efficiency of clinical trials

There has been much discussion in the literature around ways to improve efficiency in caries clinical trials. Efficiency has many specific definitions in the literature, but in general refers to the principle of the effect achieved in relation to the resources expended (Hausen, 2004). As the prevalence of caries has declined, studies have shown that the number of participants required in a clinical trial to show the same relative treatment difference have greatly increased (Whelton, 2004). Reduction in the number of required participants is therefore a priority for efficiency of caries trials. Some studies have attempted to reduce the number of participants required by pre-selecting participants based on certain criteria. One study has looked at improving efficiency by identifying certain tooth surfaces which were most likely to show treatment effects, and suggesting including only participants with a high proportion of these surfaces unaffected by caries at baseline (Downer et al., 1977). A later study suggested only selecting participants with high levels of caries at baseline, as these participants had already demonstrated susceptibility to caries (Burchell et al., 1991).

Other approaches have included making adjustments to the DMFS index to make separate measurements of lesion initiation and progression (Kingman and Selwitz, 1997). Several other suggestions for improving efficiency relate to the issue of clustered data, and will be considered in the next chapter.

3. Analysis of clustered data

3.1. Introduction

This chapter gives an introduction to cluster randomisation, and describes how the effects of clustering can be quantified using the intra-cluster correlation coefficient and design effect. The natural clustering in caries data is described, and methods which have been suggested to analyse caries data at tooth and surface level are reviewed.

An introduction to multilevel modelling is given, and as this method has rarely been used for dental caries data, the literature on multilevel modelling with periodontal outcomes is reviewed, as the structure of periodontal data has many similarities to that of caries data.

Finally, this chapter will review the literature on the modelling of within-mouth patterns of caries, such as symmetry with respect to the midline, and caries aggregation.

3.2. Clustered data

Robert Fisher's classical principles of experimental design (Fisher, 1935) make the assumption that the unit of analysis should always be the same as the unit of randomisation. Statistical methods such as analysis of variance, assume independence between individuals, and randomisation at the same level as the unit of analysis ensures that this assumption is held. However, in many cases it is desirable to design a trial

which is randomised not at the individual level, but at a higher level, such as school, or medical practice. This is known as cluster randomisation and the unit of randomisation is known as the cluster (Donner and Klar, 2000). The effect of cluster randomisation is to introduce an additional source of variation into the data, as well as varying between individuals, outcomes can also vary between clusters, as individuals within a cluster may be more alike than individuals in different clusters. If the analysis is performed at the cluster level, the effective sample size will be decreased, the standard error will be increased, and therefore confidence intervals will be wider and p-values higher. If the clustering is ignored in the analysis of a cluster randomised trial, for example by randomly assigning schools to experimental groups, and then analysing using the pupil as the unit of analysis, the lower standard errors, narrower confidence intervals, and lower p-values obtained from the individual analysis will be incorrect, and may result in incorrect conclusions being drawn from the analysis, as there will be an increased probability of type I error (a false positive result, or finding a statistically significant difference when no difference exists in the underlying populations).

The effect of clustering on a data set can be quantified using the intracluster correlation coefficient (ICC), also known as the intraclass correlation coefficient, or the variance partition coefficient (VPC). This quantity represents the proportion of the variation in observations which can be attributed to variation between clusters, rather than to variation between individuals. The ICC can take any value from 0 to 1. The ICC can be characterised as the Pearson correlation coefficient between any two observations in the same cluster. An ICC of 0 would occur where there is no correlation between observations within a cluster, and an ICC of 1 would occur where every observation in a cluster is identical.

The definition of clustered data can be extended to more than one level of clustering. The example of pupils clustered within schools could be extended to consider the class within the school. If pupils in the same class are more similar than pupils in different classes in the same school, then we have an additional cluster effect. In addition we could have schools clustered in different towns. This is an example of a multilevel or hierarchical structure, illustrated in Figure 3.1.

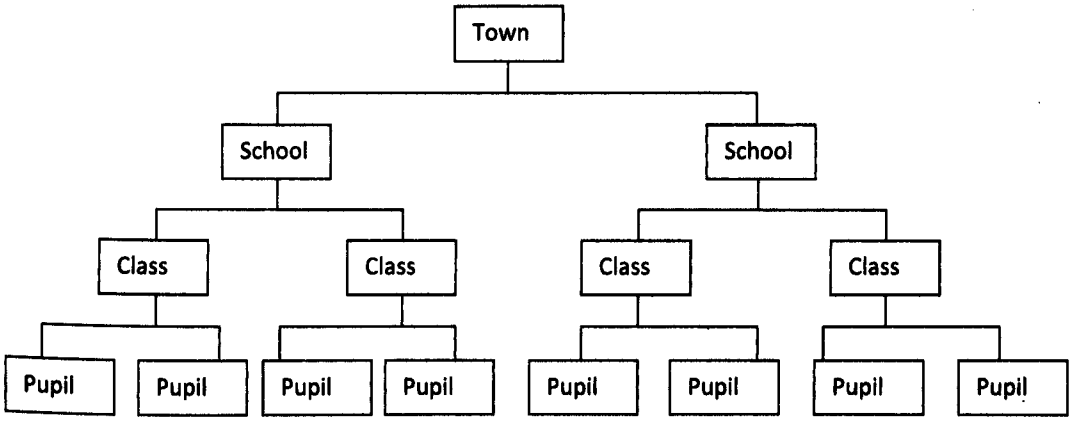

Figure 3.1 : Example of multilevel data structure

3.3. Clustered data in caries clinical trials

An important distinction to make in clinical trials is between explanatory and pragmatic trials. Explanatory trials aim to measure the efficacy of an agent, by recruiting as homogeneous a sample as possible, and tightly controlling the conditions of delivery. Pragmatic trials aim to measure the effectiveness of a preventive programme using the agent under real-life conditions, in a sample representative of the population where the agent would be used in practice (Schwartz and Lellouch, 1967). A study published in 1976 discussed the distinctions between these types in trials of caries-preventive agents (O'Mullane, 1976). As the efficacy of many caries preventive agents containing fluoride has been well established, it is likely to be more useful today to establish effectiveness in pragmatic trials (Hausen, 2004). Explanatory studies will normally be randomised by

individual participant. Pragmatic studies, however, to replicate real life conditions, are often conducted in multiple sites, e.g., schools or general practices. Interventions in these trials are sometimes allocated by site rather than by individual participant. This may be done for logistical reasons such as ease of delivery, or to avoid contamination between the intervention groups (Donner and Klar, 2000). These cluster randomised trials must be appropriately analysed to avoid the potential for incorrect conclusions.

Caries data is often collected from individuals who come from a multilevel structure as shown in Figure 3.1. However, if caries is assessed on each surface on each tooth, then the data collected will have a naturally multilevel structure as shown in Figure 3.2.

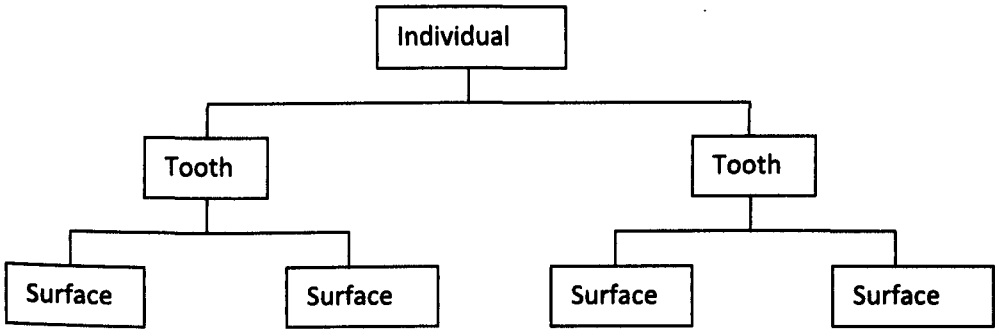

Figure 3.2 : Multilevel structure of caries data

Most clinical trials with caries as an outcome variable will collect data in this form, but as the randomisation will generally be performed at the individual level, it is inappropriate to analyse the data at tooth or surface level, assuming independence of observations. The traditional method of analysis has been to aggregate the data into the DMFT or DMFS index. This technique is appropriate, and does not suffer from the problem of dependent observations. However, by aggregating the data, it means that the detail of the data at tooth and surface level is lost.

This issue has been considered in the dental literature. A 1999 article observes that many dental studies collect hierarchical, or multilevel data, and highlights the need for correct analysis of this data, pointing out the availability of software to perform these analyses (Macfarlane and Worthington, 1999). The multilevel structure of caries data was also pointed out in another article (Gilthorpe et al., 2000). The issue was highlighted again in a paper on efficiency issues in statistical methods used in caries trials (Mancl et al., 2004).

3.3.1. Analysis of caries data at tooth and surface level

Methods have been suggested for the analysis of caries data using tooth or surface as the unit of analysis. An adjustment to the chi-square test (Ahn et al., 2002) can allow for the clustering within participant with tooth as the unit of analysis. This method can be used where a binary outcome variable is observed for each tooth, and subjects belong to two or more experimental or observational groups. The standard Pearson chi-squared test is inappropriate for the analysis of this design, as the observations are not independent. The method proposed by Ahn uses the intracluster correlation coefficient, as defined in section 3.1, to calculate a chi-squared statistic which is adjusted for clustering.

A limitation of this approach is that it only allows straightforward comparisons of proportions, and not more complex models, with adjustment for covariates. Work has been published on the analysis of caries data with more complex models.

Another approach which has been investigated is the use of survival analysis at surface level (Hannigan et al., 2001) This study uses a clinical trial data set where participants were examined at yearly intervals for three years. The authors discuss the over-simplified nature of analysis using DMF indices, and argue that the standard analyses

make little use of the vast quantities of data which are collected in a caries clinical trial. The authors point out that time has always been identified as one of the most important variables in the caries process. Therefore, as caries trials usually include clinical examinations at intermediate points between baseline and final examinations, it is proposed that these data are used to calculate the survival time of a tooth surface, defined by the time of the first examination where the surface is observed to have caries experience. As examinations in this data set are only at yearly intervals, the data is interval-censored, as the survival time can only be determined to somewhere within the 12-month period. The parameter estimates are calculated using standard survival analysis techniques, with the estimates of variance recalculated using jackknife estimators, adjusting for the clustering in the data. This is an example of a marginal model. These models allow for the dependence within subjects in calculating the variability of the regression coefficients. Marginal approaches treat the dependence structure as nuisance parameters, rather than explicitly modelling the variance at each level of the hierarchy. An alternative approach to modelling clustered data is multilevel modelling, which is discussed in the next section.

3.4. Multilevel modelling

Another method of analysis for clustered data is multilevel modelling. This method allows the analysis of data within a hierarchical or multilevel structure. Caries data naturally falls into a three-level structure, with individual participant as the top, or level 3 unit, tooth as the level 2 unit, and surface as the level 1 unit, as shown in Figure 3.2. In contrast to the marginal modelling methods, such as those used in the time-to-event analyses of caries data, this approach allows the random structure to be explicitly modelled, in addition to the fixed effects. In a single-level model the variance of the error term can be measured, and represents the amount of variation in the data which is

unexplained by the model. Multilevel models work by splitting the variance in outcome into components for each level of the model, so random effects at tooth and participant level are estimated in the modelling process. These random effects at the higher levels are assumed to follow a normal distribution with mean 0, and variance which is estimated in the modelling process (Goldstein, 2003). Simulation studies have shown that parameter estimates are fairly robust to violations of this assumption (Maas and Hox, 2004). In multilevel models, the intracluster correlation coefficient (ICC) can then be calculated to measure the proportion of variance which is attributable to each level of the model. This quantity is also known as the variance partition coefficient (VPC). In a three-level model, separate ICCs can be calculated measuring the proportion of total variance attributable to level 2 and level 3.

Multilevel modelling of the hierarchy shown in Figure 3.2 would treat the tooth level as a random effect, where teeth within an individual are treated as having random variation with respect to the outcome variable. However, in caries data, it has been shown that different teeth and surfaces have different susceptibilities to caries (Batchelor and Sheiham, 2004; Hannigan et al., 2000). Therefore, it may be of interest to use fixed effects to model the susceptibilities of particular teeth and surfaces.

3.5. Multilevel modelling of periodontal disease outcomes

The majority of analyses using multilevel modelling techniques that have been published in the dental literature are on periodontal disease outcomes. Periodontal data has a similar multilevel structure to caries data, with measurements taken at several sites around each tooth, giving a three level structure similar to that shown in Figure 3.2 (Gilthorpe et al., 2000).

Periodontal disease, as with dental caries, is associated with bacteria in the plaque which collects on teeth. The gingiva, or gums, become infected and inflamed, and the tissue which connects the gingiva to the tooth recedes, leaving pockets between the tooth and gingiva, where bacteria and food debris can collect. This can lead to loss of the alveolar bone which supports the tooth, and in extreme cases, the tooth can be lost (Armitage, 2004).

The severity of periodontal disease at a given point on a tooth can be measured using several outcome variables (Page and Eke, 2007). The distance from the cementoenamel junction (the point where the enamel of the tooth meets the root) to the gingival margin (the edge of the gum) gives a measure of *gingival recession*. The distance from the gingival margin to the bottom of the periodontal pocket is known as *pocket probing depth*. The sum of these two measurements is known as *lifetime cumulative attachment loss*, or sometimes just *attachment loss*.

Measurements of the height of the alveolar bone can also be assessed using radiographs. The *alveolar bone height* at the interproximal areas of teeth (the areas where two teeth meet) can be measured, and in longitudinal studies can be used to assess the level of bone loss. *Bone defects* can also be measured on radiographs, these are areas where bone has been lost which are surrounded by unaffected bone, thus allowing measurement of the extent of the defect (Needleman et al., 2006).

Data on these outcomes tend to follow a similar hierarchy to caries data, with sites nested within teeth, nested within individuals (individual – tooth – site).

This section reviews the published multilevel models of periodontal outcomes, concentrating on how tooth and site are modelled within individuals, and whether the

authors approach the modelling of tooth and surface level with random effects, fixed effects, or a combination of both.

The studies considered here all use as their outcome variable either one of the above measurements, or longitudinal change in one of the measurements between two examinations.

The effects of explanatory variables on these dependent variables are modelled within the hierarchical structure of the data set. It should be noted that many of the analyses discussed here were performed at a time when software availability is likely to have limited the complexity of the models fitted.

3.5.1. Two-level models

The simplest multilevel model encountered in the periodontal literature is the random intercept 2-level (individual – site) model. This model is fitted to a data set of 29 individuals with chronic periodontal disease, with pocket depth as the outcome variable (Sterne et al., 1988). The data came from a placebo-controlled clinical trial of the use of the antibiotic metronidazole in patients with chronic periodontal disease. The measurements have been taken on 4 sites per tooth (buccal, lingual, mesial and distal), and only 6 teeth are included (the first permanent molars in the upper right and lower left quadrants, and the central incisors and first premolars in the upper left and lower right quadrants). This gives a maximum of 24 site measurements per individual. These measurements were taken before and after the intervention. The actual number of sites measured per individual varies between 12 and 24, due to missing teeth. In total 936 observations were made.

The outcome variable is pocket depth measure after the intervention. Tooth and site type are included as covariates at site level in the fixed part of the model, using dummy

variables. The upper right first permanent molar is used as the reference category for the tooth variables, with each of the other 5 teeth examined assigned to a dummy variable. The buccal site is used as the reference category for the site types, with 3 dummy variables for mesial, lingual and distal.

The authors discuss some of the assumptions made in their model. They point out that the "tooth level" factors are assumed to have an effect only at their site. This is due to the lack of a tooth level in the multilevel structure. They do not discuss the possibility of fitting a 3-level model with tooth level included. Also, the covariance between sites is assumed to be explained completely by the individual level effect. This means that this model does not take account of potential relationships between particular sites within a tooth, other than to model the variation between individual. The authors discuss the possibility of overcoming this by allowing a different site level variance for each type of site, although do not present results on these models.

This model is also used in a study on 142 factory employees, with an outcome variable of alveolar bone height measured at the interproximal areas of teeth for a total of 5579 sites (Albandar and Goldstein, 1992). It is not clearly stated in the paper which sites or teeth are measured, but an average of 39 sites per individual was measured. The study looked at predictors of periodontal disease progression. Tooth was not included as a site level variable in the fixed part of the model. The paper describes recent advances in software for multilevel modelling techniques, and suggests how they could be used in periodontal research. The issue of random effects is discussed, reproduced below.

"We also note that in this model the sites are assumed to be sampled at random within individuals, independently. That is, the identification, i, of a site simply indexes a sample position. In reality, of course, we can identify the same sites for individuals, and this information can be used to improve the analysis. The appropriate multilevel model in

this case, however, is complicated and we do not consider it further. An alternative formulation is to fit a separate dummy variable for each site. This would then yield a standard unilevel analysis of covariance model. But since the number of sites is usually large this alternative approach would be cumbersome."

Another study fits a model with this specification, to data from three samples, 89 individuals under age 30, with juvenile periodontitis, 139 individuals under age 35 with severe periodontitis, and 309 individuals aged over 35 with chronic periodontitis (Gunsolley et al., 1994). The dependent variable in the analyses is attachment loss, and is measured at four sites for each tooth (mesial, distal, buccal and lingual). All teeth were included. Tooth positions are included in the fixed part of the model as explanatory variables at site level, categorised into incisors, canines, premolars, first molars and second molars, and separated into upper and lower, giving a total of ten tooth position categories. Site is also included as a covariate in the fixed part of the model, separated into three categories, with mesial and distal grouped together as one category.

3.5.2. Three level models

The above model can be extended to 3 levels, using the hierarchy individual – tooth – site.

One of the studies listed above (Gunsolley et al., 1994) extend their two-level model described above, to include the tooth level in the hierarchy. They state that "the purpose of this study is to investigate factors which influence the within individual correlation structure of attachment level measurements. Specifically, should the correlation structure take into account the tooth from which site specific measurements are made?" They conclude that including the tooth level is important in their data set, as the standard errors they find for the fixed effects of tooth position are smaller in the 2-

level model than the 3-level model, since in their analysis using the two-level model, the correlation structure of sites within teeth is not being modelled. This suggests that the standard errors in the two-level model are too narrow, potentially leading to falsely significant results.

This model is also used in a study on 22 patients aged 35-55 with periodontal disease (Axtelius et al., 1999). The dependent variables here are pocket probing depth, and change in pocket probing depth between two examinations. Four sites per tooth are measured on all teeth. The authors include dummy covariates for tooth type (incisors, canines, premolars and molars) in the fixed part of the model.

Another study models lifetime continuous attachment loss, and pocket probing depth using model 2, with measurements taken on 4 sites per tooth, on all teeth excluding the third molars (Tu et al., 2004b). The population is 100 white males aged 16-20, examined at 3 timepoints. This paper uses fixed effects for each tooth position within a quadrant as covariates. The central incisor is taken as the reference category, and a dummy variable is created for each position from lateral incisor to second permanent molar, a total of 6 dummy variables.

The authors conclude from their results that tooth level is an important part of the hierarchy, stating that "Variances at the tooth level were generally greater than those at the subject level, indicating that variation in LCAL (lifetime cumulative attachment loss) and PD (pocket depth) across different tooth positions in the mouth was more prominent than individual variability across subjects. Therefore, both tooth and subject levels need to be taken into account when specifying the periodontal data structure. Either due to the incorrect assumption that the tooth level has no pre-eminent role in the full hierarchy, or because the accepted methodology has limitations, studies that take

account of the clustering of sites in the mouth while ignoring the tooth-level structure might lead to erroneous results."

The same data set is also modelled using a longitudinal outcome variable (Tu et al., 2004a). The multilevel techniques used are the same. The outcome variable here is the change in measurement between baseline and final examination.

3.5.3. Summary

These periodontal studies have shown a variety of approaches to modelling the tooth level, both by including and excluding random effects at tooth level, and by including various tooth position indicator variables. In Chapter 5, the application of these methods to a caries data set will be investigated.

3.6. Within mouth patterns of caries incidence

One of the aims of this thesis is to investigate how multilevel analysis of data from caries clinical trials can be used to predict tooth and surface specific caries incidence using baseline caries status.

It has long been a commonly held view by clinical observation that caries develops symmetrically in similar teeth on the right and left sides of the mouth. This apparent symmetry with respect to the midline was interpreted in the early part of the last century, as evidence that caries was not an infective disease (Eckermann, 1919). A study of 300 bitewing radiographs (Scott, 1944) found that 73% of posterior decayed, missing or filled teeth (DMFT) were involved bilaterally (i.e. the corresponding tooth on the opposite side of the mouth was also decayed, missing or filled). A longitudinal study of dental caries in 666 English schoolchildren (Berman and Slack, 1972) found 'bilateral symmetry of caries attack at all ages'. Some more recent reports have challenged this

view. A study of 510 children aged 12 years (Wood, 1985) demonstrated, from a retrospective dental record analysis, that 44% of maxillary and 33% of mandibular pairs of occlusal surfaces of first permanent molars showed caries experience unilaterally, i.e. only on 1 side of the mouth. A trial of 15,132 adults, using data from the 1985–1986 National Survey of Oral Health in the USA (Hujoel et al., 1994b) found that 94% of adults with 2 or more decayed or filled surfaces had 2 or more 'discordant pairs', where a decayed surface on 1 side of the mouth had a sound contralateral counterpart (on the corresponding laterally opposite side of the mouth). The study showed that the distribution of these pairs was not random with respect to the midline and that the caries tended to be aggregated on 1 side of the mouth. An investigation using data from 20,000 UK children aged 5–16 years (Batchelor and Sheiham, 2004) used probit analysis to rank surfaces by their susceptibility to caries. The study did not show precise symmetry between equivalent surfaces on the left and right sides of the mouth but found that symmetry existed within groups of sites with similar susceptibility to caries. A recent evaluation of deciduous teeth in 7,074 children aged 3–7 years (Vanobbergen et al., 2007) concluded that associations of caries experience at the population level appeared to follow a symmetrical pattern, but at an individual level, when using the same method as the study from the 1985-86 US National Survey (Hujoel et al., 1994b), the study found similar results, i.e. that caries lesions tend to cluster on one side of the mouth.

3.7. Conclusions

This chapter, and the previous one, have demonstrated that the changing patterns of caries experience have impacted on the design of caries clinical trials, resulting in higher required sample sizes and concern over the statistical validity of the standard analyses using the DMF index. Also, clustered data is an important issue in caries trials. Studies

conducted in the community are often cluster randomised, and in addition the issue of aggregating caries data to an individual level variable has been raised, with many researchers suggesting the use of analysis at tooth and surface level. This thesis will first consider recent clinical trials of topical fluoride interventions in children, with particular attention to issues related to clustered data. Multilevel modelling is a common statistical technique for clustered data in many clinical areas, but has been little used in the caries literature. Therefore, the primary aim of the thesis is to investigate the potential use of multilevel modelling in caries clinical trials. There has also been considerable interest in modelling within-mouth patterns of caries, such as symmetry and caries aggregation. However, the literature is based on analysis of cross-sectional data sets, and has not used longitudinal data sets such as those from clinical trials. This thesis will also consider whether these within-mouth patterns can be used to predict caries incidence over the period covered by a clinical trial, using multilevel modelling techniques.

4. Statistical aspects of the design and analysis of clinical trials for the prevention of caries

4.1. Introduction

In Chapter 3, the issue of correctly analysing data from cluster randomised trials was raised. This chapter will consider in detail recent publications from caries clinical trials, and examine the design and analysis, and standard of reporting, of these publications, with particular reference to issues related to clustering.

An important distinction to make in clinical trials is between explanatory trials of the efficacy of an agent, and pragmatic trials of the effectiveness of a preventive programme using the agent in real-life conditions (O'Mullane, 1976). Explanatory studies will normally be randomised by individual participant. Pragmatic studies, however, are often conducted in multiple sites, e.g. schools. For logistical reasons, interventions in these trials are sometimes allocated by site rather than by individual participant.

Clinical studies where groups of participants rather than individuals are randomised to treatment groups are known as cluster randomised studies. These studies can be analysed taking the cluster as the unit of analysis, although this approach loses the information from the individual participants. If analysis is to be performed at the

individual level, the clustering must be accounted for, as failure to do so will result in a confidence interval for the relative treatment effect which is too narrow, and therefore possibly an incorrect conclusion. The effect of clustering can be quantified by calculating the intracluster correlation coefficient (ICC), defined as the proportion of the total variation which can be attributed to the variation between clusters. The value of the ICC can range from 0 to 1. An ICC of 0 would mean that all observations within a cluster were independent, i.e. there is no cluster effect. An ICC of 1 would arise when all observations within a cluster are identical, i.e. there is no variation within clusters. This can be used to calculate the effective sample size, which is defined as the number of participants in an individually randomised trial which would give the same power as the cluster randomised trial. The clustering can be taken into account without losing the information at the individual level using various methods, such as multilevel modelling (Goldstein, 2003).

Clinical studies examining dental caries experience generate multiple outcome data for each participant. Within each participant there are data for multiple teeth, and for multiple surfaces on each tooth. This type of clustering is usually handled by taking caries increment, the number of teeth or surfaces which have become affected by caries during the course of the trial, as the outcome variable. However, as the data have been summarised, tooth and surface specific information is lost. These data could provide important clinical information, as interventions may be more effective on particular teeth or surfaces within the mouth, depending on the method of application. Therefore, if analysis is to be performed at tooth or surface level, the clustering within participants should be accounted for to ensure accurate conclusions.

Studies have been published reviewing the issues of clustering in clinical trials in other fields (Eldridge et al., 2004; Varnell et al., 2004), however, a study of this type has not been previously published for trials in dental caries.

4.1.1. Reporting of randomised controlled trials

The CONSORT statement (CONsolidated Standards Of Reporting Trials), originally published in 1996, was revised in 2003 (Moher et al., 2003). This set of guidelines for the reporting of randomised controlled trials has been adopted by many medical journals. An extension to the CONSORT statement covering the reporting of cluster randomised trials has recently been proposed (Campbell et al., 2004). This extension requires reporting of the rationale for adopting a cluster design, how the effects of clustering were incorporated into the sample size calculations, and how the effects of clustering were incorporated into the analysis. The guidelines state that ICCs should be reported.

The aim of this study is to assess the design and analysis of recent randomised controlled trials in which dental caries is the outcome of interest, with particular emphasis on the potential impact of clustering on the assessment of relative treatment effects in cluster randomised trials.

4.2. Methods

Trials of topical fluoride interventions in children have been examined in this methodological study, as these interventions are likely to have been evaluated in both explanatory and pragmatic trials. Seven Cochrane reviews evaluating the efficacy of topical fluoride vehicles as caries preventive agents have been published (Marinho et al., 2003a; Marinho et al., 2003b; Marinho et al., 2002a; b; 2003c; Marinho et al., 2004a; b),

and have been used to identify trials for this methodological study. These reviews have been chosen as they examine the four main methods of topical fluoride intervention, gels, toothpastes, varnishes and mouthrinses.

The lists of both the included and excluded studies in the seven Cochrane reviews were scrutinised. Trials excluded from the reviews were considered for inclusion here, as the exclusion criteria for the Cochrane reviews were stricter than is necessary for this study. For example, trials with unblinded outcome assessment and additional non-fluoride interventions were excluded from the Cochrane reviews, but included here since design and analysis issues apply equally to these trials. Non-randomised trials excluded from the Cochrane reviews were also excluded from this study. In the case of a trial where randomisation was not reported in the paper, attempts were made to contact the authors to establish whether the trial was randomised. If the authors did not reply with this information, the trial was excluded. In addition, trials which were reported in theses and not published in peer-reviewed journals were excluded, as were trials only published in abstract form. Trials published before 1990 were excluded, since the aim of this review was to examine reasonably current practice. Since the Cochrane reviews identified studies up to 2000, this work covers the ten-year period 1990-2000.

The seven Cochrane reviews identified 303 trials, 29 of which had published results since 1990. Eleven were non-randomised and were thus excluded, one was excluded as it was only available in abstract form, and two were excluded as they were only published in theses. This left 15 trials for consideration in this study.

All published papers relating to eligible trials were obtained and studied. Papers not published in English were translated. Information was extracted from the papers on the type of trial, and the characteristics of the participants. The articles were studied to

establish the units of randomisation and analysis, and to assess whether cluster randomised trials were appropriately analysed.

Where clustering was not accounted for in the analysis, and a statistically significant difference between groups was reported, an analysis was undertaken as follows to assess how robust this result would be to various degrees of clustering. This analysis tests how robust the calculation performed by the authors of the papers would be to clustering in the data, and can be calculated using the summary data reported in the papers. For example, consider a hypothetical trial conducted in 10 schools, with 30 children in each school, and each school randomly assigned to either the intervention group or control. Ignoring the clustering, the control group have a caries increment of 1.5 (s.d. 0.9), and the intervention group 1.2 (s.d. 0.9). Analysing at the child level, ignoring the clustering, using a t-test gives a significant difference (t statistic= 2.89, S.E. = 0.104, d.f. = 298, $p<0.01$). To examine the effect of clustering, we use an adjusted t-test method (Donner and Klar, 2000), first estimating the *variance inflation factor,* or *design effect* for each group, given by $C_i = 1+(\overline{m}_{Ai}-1)\hat{\rho}$, for the *i*th group, where $\overline{m}_{Ai}$ is defined as $\sum_j \frac{m_{ij}^2}{M_i}$, m_{ij} is the size of the *j*th cluster in the *i*th group, M_i is the total number of individuals in the *i*th group, and $\hat{\rho}$ is the intracluster correlation coefficient (ICC). The design effect can be interpreted as the ratio of the number of individuals in the trial to the *effective sample size*, which is the number of participants in an individually randomised trial which would give the same power as the cluster randomised trial. If the individual cluster sizes are not published, the cluster sizes are assumed to be equal, and $\overline{m}_{Ai}$ is replaced by the mean cluster size for the group. In our example, as the cluster sizes are equal, $\overline{m}_{Ai}$ is equivalent to the mean cluster size m, and is equal to 30. Substituting various hypothetical values for the ICC into the formula $1+(\overline{m}_{Ai}-1)\hat{\rho}$ gives a design effect for

each ICC which can be used to adjust the standard error of the difference in means (S.E. = $S_P\left[\frac{C_1}{M_1}+\frac{C_2}{M_2}\right]^{1/2}$), where S_P is the pooled standard deviation over the two groups. The test statistic is then recalculated for each value of the ICC, by replacing the original standard error with the adjusted standard error calculated above. The degrees of freedom under the null hypothesis for this statistic are K-2, where K is the total number of clusters in the trial.

The smallest value of the ICC which results in p>0.05, or equivalently the 95% confidence interval just including zero, can be found. In this case, an ICC of 0.02 gives a design effect of 1.58 assumed to be equal in each group, and an adjusted standard error of 0.131. Substituting this standard error gives a new t-statistic of 2.297, which is lower than the critical value of the t-distribution with 8 degrees of freedom (2.306), resulting in a p-value of greater than 0.05, and the null hypothesis no longer being rejected.

In the case of binary outcomes, where the data have been analysed using the chi-square statistic, a similar method is used, where the design effect for each group is calculated exactly as above, and the adjusted chi-square statistic, with one degree of freedom is $\sum_{i=1}^{2}\frac{M_i(\hat{P}_i-\hat{P})}{C_i\hat{P}(1-\hat{P})}$, where $\hat{P}$ is the overall event rate observed in the study, and $\hat{P}_i$ is the event rate in the ith group (Donner and Klar, 2000).

The trials identified to be cluster randomised were also studied to identify whether they could be affected by consent bias. This is a potential source of bias in cluster randomised trials, and can occur when the participants are consented into the trial after the randomisation has been performed. The participants may know into which group they have been randomised before they decide whether to take part, and their decision

may be influenced by this knowledge (Puffer et al., 2003). The CONSORT guidelines recommend including a flow diagram of both clusters and participants in a trial, which if included, would allow the reader to judge the potential for consent bias.

The CONSORT guidelines state that papers should report how the sample size was determined, since if no sample size calculation is reported, the reader cannot judge whether studies with non-statistically significant results were sufficiently powered to detect a difference. The reports of the trials included in this study were examined to establish if a sample size calculation was reported.

4.3. Results

A description of the principal study design features is presented for each of the fifteen trials (Table 4.1). One of the trials (Brodeur et al., 1988) although first published prior to 1990, has been included, as an additional paper was published from the trial after the 1990 cut-off (Brodeur et al., 1990). From the character of the study design, and purpose and conduct of the trial, the trials were classified as either explanatory or pragmatic. Five trials were randomised at a higher level than the participant. Four of these were judged to be pragmatic trials, and were randomised by school, or by school class. The remaining cluster randomised trial (Ran et al., 1991) was classified as an explanatory trial, as it aimed to measure the efficacy of an agent, rather than the effectiveness of a programme, and was placebo controlled. This trial was randomised by school class.

Three of the trials included in this study were excluded from the original Cochrane reviews. One was excluded due to open outcome assessment, i.e. the examiner was not blinded to group allocation (Brodeur et al., 1988). Another was excluded due to it being unclear whether the study was randomised (Chikte et al., 1996). The author has been

Study	Location	Year of commencement	Intervention[1]	Mode of administration[2]	Age of children at baseline	Number of groups	Number of participants at randomisation	Number of randomisation units (if different)	Number of participants at end of study	Sample size calculation reported	Unit of randomisation	Unit of analysis	Explanatory / Pragmatic
Brodeur et al., 1988	Canada	1983	M	S	9-11	4	955	22	699	No	School	Child	P
Borutta et al., 1991	Germany	1988	V	P	12-14	4	400		360	No	Child	Child	E
Frostell et al., 1991	Sweden	1977	V	P	4	2	NR		206	Yes	Child	Child	P
Ran et al., 1991	Israel	1989	G	S	13	4	140	4	112	No	School class	Child	E
Spets-Happonen et al., 1991	Finland	1985	M	S	11	4	243		201	No	Child	Child	E
Tewari et al., 1991	India	1982	V	P	6-12	4	1335		1251	No	Child	Child	E
Heidmann et al., 1992	Denmark	1983	M	S	6-12	2	1306		1083	No	Child	Child	P
Olivier et al., 1992	Canada	1985	G	P	6-7	2	488		431	No	Child	Child	P
Karjalainen et al., 1994	Finland	NR	MT	S	7-8	2	313	16	206	No	School class	Child	P
Seppa et al., 1995	Finland	1991	VG	P	10-12	2	289		254	No	Child	Child	E
Bravo et al., 1996	Spain	1990	V	P	6-8	3	362	15	314	No	School class	Child, tooth	P
Chikte et al., 1996	S. Africa	1990	M	S	6-12	2	2041	4	1245	No	School	Child	P
Kleber et al., 1996	USA	1994	MT	S	10-11	3	260		156	No	Child	Child	E
Petersson et al., 1998	Sweden	1994	M	S	13	2	NR		139	No	Child	Child	P
Gisselsson et al., 1999	Sweden	1993	G	P	13	3	317		280	No	Child	Child	E

NR – Not reported

[1] G – gel, M – mouthrinse, T – toothpaste, V – varnish [2] P – professionally applied, S – self applied

Table 4.1 : Description of the clinical trials of caries-preventive agents included in this study

contacted, and confirmed that the trial was randomised. A third was excluded because the intervention included a non-fluoride component (Karjalainen et al., 1994).

4.3.1. Analysis of cluster randomised trials

Only one of these five trials (Bravo et al., 1996) reported that clustering was accounted for in the analysis. This trial had a randomisation unit of school class, and the authors included school, although not school class, as a variable in their logistic regression analysis. No statistically significant school effect was detected, and so the clustering was ignored. In addition, the authors performed an analysis with tooth as the unit of analysis. Here, they adjust the analysis to account for the lack of independence of the teeth in the mouth using an appropriate method (Donner and Banting, 1989).

The four remaining cluster randomised trials were not analysed using methods appropriate to cluster randomised data. Table 4.2 summarises the statistically significant results reported in these four trials, and gives details of the adjusted statistical tests which have been used to identify the critical values of the ICC which, if present in the data, would result in the loss of the statistically significant results. For one study (Ran et al., 1991) the adjusted t-test cannot be calculated as there is only one cluster per group, which results in 0 degrees of freedom. Any cluster related variation in this data would result in the loss of the statistically significant result.

4.3.2. Consent bias

One cluster randomised trial (Ran et al., 1991) did not mention consent, but the four groups all received supervised toothbrushing in class, with different gels, so consent bias was unlikely. Another trial (Brodeur et al., 1988) included a participant flow chart which clearly stated that parental consent for children to take part took place before randomisation, so no bias could be present.

Trial	Outcome variable	Statistical Test	Groups	Original data	Original p-value	Clustering details	Critical value of ICC	Design Effects	Adjusted standard error	Adjusted test statistic	Degrees of freedom
Brodeur et al., 1988*	DMFS	t-test on fluoridated communities only	Test (n=157)	Mean 0.44 s.d. 1.89	0.026	10 clusters (estimated)*	0.02	Test 1.03	0.264	t=2.28	18
			Control (n=131)	Mean 1.04 s.d. 2.52		10 clusters (estimated)*		Control 1.02			
Ran et al., 1991	DMFS	t-test, one pairwise significant result	Group I (n=29)	Mean 2.3 no s.d. reported	<0.02	1 clusters		See text			
			Group IV (n=27)	Mean 0.0 no s.d. reported		1 clusters					
Karjalainen et al., 1994	Number of caries free individuals	Chi-squared test	Test (n=112)	42 of 112 caries free	0.001	8 clusters	0.08	Test 2.04	NA	χ^2=3.341	
			Control (n=94)	53 of 94 caries free		8 clusters		Control 1.86			
Chikte et al., 1996	DMFT adjusted for baseline values	t-test	Test (n=603)	Mean 0.720 s.d. 1.137	<0.001	2 clusters	0.002	Test 1.6	0.082	3.958	3
			Control (n=642)	Mean 1.045 s.d. 1.138		2 clusters		Control 1.6			

* The number of clusters in fluoridated communities is not specifically stated in the paper. The maximum possible number of clusters for this comparison is 20 (there are 22 clusters total, but the unfluoridated communities must have at least one cluster per group). For this analysis, it is assumed 10 schools were in the mouthrinse group and 10 in the control group. This scenario minimises the effect of clustering for this result

Table 4.2 : Details of adjusted statistical tests for cluster-randomised trials originally analysed without adjustment for clustering

In the other three cluster randomised trials, it was not clear whether consent bias could be present. The first (Bravo et al., 1996) did not report when consent took place, but did state that there were different consent rates in the 3 groups, ranging from 75% in the varnish group, to 87% in the sealant group. The other two trials, (Chikte et al., 1996; Karjalainen et al., 1994) do not report whether the random allocation of schools took place prior to the consent process.

4.3.3. Sample size calculation

None of the cluster randomised trials presented a sample size calculation. Only one of the individually randomised trials (Frostell et al., 1991) included this information, but did not do so in enough detail to allow the calculation to be replicated.

4.4. Discussion

Since ICCs have not routinely been published for caries increment in schools, it is difficult to estimate how likely it is that the critical values found in this study could be present in the data. Data held by the authors on a clinical trial conducted in schools showed an ICC of 0.01 for caries increment between schools. This value would result in two of the trials considered here losing their statistically significant results. Since the actual ICCs from these trials have not been reported, there is no evidence that the conclusions drawn from the analyses without adjustment for clustering are accurate. One trial had only one cluster per group, and so could not be analysed using the cluster-adjusted method.

There is also an argument that school or class effect may be an issue in clinical trials which are randomised at the participant level. This is usually tackled by stratifying by school at randomisation. However, unequal dropout rates between schools can remove

this balance, resulting in some schools having a bias towards a particular treatment group. This could introduce bias in the overall results since different schools may have different caries levels. Applying multilevel modelling techniques to these trials would be one way of addressing this potential source of bias by accounting for the school effect in the analysis. Other appropriate methods of analysis for clustered data include generalised estimating equations, bootstrap or jackknife methods to estimate the standard error, and Poisson regression models.

Three of the five cluster randomised trials did not report the consent procedure in sufficient detail to establish whether there was potential for consent bias. Inclusion of a participant flow diagram as recommended by CONSORT would allow a consideration of this design feature to be made. This review suggests that sample size calculations have not been routinely detailed in reports of clinical trials of caries preventive agents.

When the work in this chapter was originally accepted for publication, the use of the CONSORT statement was not yet widespread in the dental literature, although two of the ten highest impact dental journals (Journal of Dental Research and Caries Research) required reports of randomised controlled trials to conform to these guidelines. The adoption of the CONSORT guidelines by other dental journals has increased, with several other high impact journals requiring the guidelines. This should help reduce the problems identified in this methodological study. This review may be considered a baseline study and it will be of interest to assess whether design, analysis and reporting practices improve in the future.

In summary, the majority of trials identified in this study found to be cluster randomised were analysed without accounting for the clustering in the analysis. This could result in the significance of the group differences being exaggerated. These issues are unlikely to be specific to the trials of topical fluorides considered here, and could

apply equally to all trials in dentistry which may be cluster randomised. In addition, the CONSORT guidelines were not followed in the reporting of the majority of these trials. This results in the reader being unable to assess whether non-significant results may have been due to underpowering. These methodological considerations mean that inaccurate conclusions may have been drawn about the clinical benefit of some interventions.

5. The application of multilevel modelling to dental caries data

5.1. Introduction

As discussed in Chapter 3, clinical studies examining dental caries experience generate multiple outcome data for each participant. The caries status of each tooth surface is diagnosed separately. Caries is a continuous and gradual process, and certain thresholds exist beyond which a surface is defined as having caries. The two common thresholds used are known as D_1, where the surface has a carious lesion affecting the outer layer of enamel, and D_3, where the lesion has affected the both the enamel and the dentine layer below the enamel (Forgie et al., 2000). Within each participant there are data for multiple teeth, and for multiple surfaces on each tooth. This clustering of data within an individual is usually handled by taking the caries increment, that is the number of surfaces which have become affected by caries during the course of the trial, as the outcome variable (Worthington, 1984). However, as the data have been summarised, tooth and surface specific information is lost. These data could provide important clinical information, as interventions may be more effective on particular teeth or surfaces within the mouth, depending on the method of application. However, if analysis is to be performed at the tooth or surface level, the clustering within participants should be accounted for to ensure accurate conclusions. Failure to account

for clustering will result in a confidence interval for a relative effect which is too narrow, and therefore possibly an incorrect conclusion.

Methods have been suggested previously for the analysis of caries data using tooth or surface as the unit of analysis. An adjustment to the chi-square test (Ahn et al., 2002) can allow for the clustering within participant with tooth as the unit of analysis. However, this approach is limited to straightforward comparisons of proportions, and does not allow for more complex models. Clustered survival analysis has been used (Hannigan, 2004), defining the survival time of a surface as the time from the start of the trial until the tooth becomes affected by caries. These data are then analysed using a marginal model, which allows for the dependence within subjects in calculating the variability of the regression coefficients. Marginal approaches treat the dependence structure as nuisance parameters, rather than explicitly modelling the variance at each level of the hierarchy.

Another method of analysis for clustered data is multilevel modelling. This method allows the analysis of data within a hierarchical structure. Caries data would naturally fall into a three-level structure, with individual participant as the top, or level 3 unit, tooth as the level 2 unit, and surface as the level 1 unit. In contrast to the marginal modelling methods, this approach allows the random structure to be explicitly modelled, in addition to the fixed effects. In a single level model the variance of the error term can be measured, and represents the amount of variation in the data which is unexplained by the model. Multilevel models work by splitting the variance in outcome into components for each level of the model, so random effects at tooth and participant level are estimated in the modelling process. These random effects at the higher levels are assumed to follow a normal distribution with mean 0, and variance which is estimated in the modelling process. Simulation studies have shown that parameter

estimates are fairly robust to violations of this assumption (Maas and Hox, 2004). In multilevel models, the intracluster correlation coefficient (ICC) can then be calculated to measure the proportion of variance which is attributable to each level of the model. This quantity is also known as the variance partition coefficient (VPC). In a three level model, separate ICCs can be calculated measuring the proportion of total variance attributable to level 2 and level 3.

Most of the work on the analysis of clustered dental data has come from the field of periodontology. Periodontology data sets tend to have a similar hierarchical structure to caries data sets, with measurements on various sites around each tooth, for each tooth in the mouth. In particular, multilevel modelling techniques have been applied to several periodontal data sets (Axtelius et al., 1999; Gilthorpe et al., 2001; Gunsolley et al., 1994; Nieri et al., 2002; Sterne et al., 1988; Tu et al., 2004a; b). These have been considered in detail in Chapter 3.

The earliest use of multilevel techniques for periodontal data (Sterne et al., 1988) fitted a two-level model, with observations on sites nested within individual participants. This model was subsequently extended to a three-level model, with measurement sites nested within individual teeth, which are nested within individual participants (Gunsolley et al., 1994).

Several of the periodontal models have used a covariate to investigate the effect of tooth position within the mouth. This issue is of interest in caries, as the disease tends to affect certain teeth more than others. One paper (Axtelius et al., 1999) uses four tooth position categories, namely incisors, canines, premolars and molars (referred to as Classification A here), whilst another (Tu et al., 2004a; b) uses seven categories, central incisor, lateral incisor, canine, first premolar, second premolar, first molar and second molar (Classification B). These two classifications assume that the effect of tooth

position in each of the four quadrants is equal. A third paper (Gunsolley et al., 1994) categorises teeth into five types, incisors, canines, premolars, first molars, and second molars, and treats upper and lower teeth separately, resulting in 10 categories for tooth position (Classification C).

None of these multilevel models or tooth position classifications has been previously used for the analysis of data from caries clinical trials and thus the potential of multilevel modelling to enhance efficiency and understanding of the therapeutic benefit of caries prophylactic agents is unknown. Therefore, the aim of this chapter is to explore the utility of multilevel modelling of data from a clinical trial with caries as the outcome variable, by investigating the effect of different level structures on the interpretation of the results. In addition, the implications of tooth groupings to model the fixed effects of tooth position are explored.

5.2. Methods

The dataset analysed in this chapter comes from a randomised controlled trial examining the caries preventive efficacy of a chlorhexidine varnish on the teeth of adolescents aged 12 at baseline, and followed up over 3 years (Forgie et al., 2000). Chlorhexidine is an anti-bacterial agent, and the varnish is intended to reduce the development of caries by reducing the level of caries-associated microflora (Emilson, 1994). The intervention groups received applications of the varnish a minimum of once per year, with additional applications if they were found to have high levels of bacteria.

The 1240 participants in the trial were randomised to one of four groups. Group A received chlorhexidine varnish and dental health advice, group B received a placebo varnish and dental health advice, group C received only dental health advice, and group D received no intervention. The original paper used caries increment as the outcome

variable, and compared the four groups via an analysis of variance. The results showed no significant difference in caries increment between the four groups.

The analysis here will use data from participants randomised to group A, the chlorhexidine varnish, and group C, the control group who received dental health advice only. The group sizes were 268 in the varnish group, and 324 in the control group at randomisation. Only participants who were examined both at baseline and at the 3-year follow up examination are included in this analysis, with group sizes of 222 in the varnish group, and 243 in the control group.

The outcome variable is caries increment by individual surface (rather than the aggregated measure commonly used). This is calculated for each surface, taking the value 1 if it was unaffected by caries at baseline (including surfaces on unerupted teeth), and had become affected by caries by the end of the study. Otherwise, the increment takes the value 0. If the surface is affected by caries at baseline, and is therefore unavailable for increment, it has been excluded from the analysis. This excluded 3158 of 28416 surfaces (11%) in the varnish group, and 3216 of 31104 surfaces (10%) in the control group. Children in the control group had a mean of 13.2 surfaces excluded, and those in the varnish group had a mean of 14.2 surfaces excluded. The threshold for determining whether a surface was affected by caries was the D_1 level, which includes any visible caries, whether only in the enamel of the tooth, or extending into the dentine. The data were also analysed using the D_3 threshold, but as the conclusions were similar, these results have not been presented here.

The first model fitted, model 0, is a simple logistic regression on the surface outcomes, with no multilevel structure. This model is fitted only as a baseline for comparison with later models.

$$\text{logit}(\pi_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + e_{0i}$$
$$y_i \sim Bin(1, \pi_i)$$
$$\text{var}(y_i \mid \pi_i) = \pi_i(1-\pi_i)$$

(Model 0)

The outcome variable for a specific surface i, y_i, which takes the value 0 if the surface does not develop caries and 1 if it does develop caries, is assumed to follow a binomial distribution with probability of success π_i. The model uses π_i as the dependent variable, which represents the probability that the ith surface becomes affected by caries during the course of the study. The logistic model uses the logit transformation on the left hand side of the equation, where $\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i})$. The right hand side of the equation includes covariates x_1 to x_n with associated coefficients β_1 to β_n and the error term e_{0i}.

This model makes the assumption that the caries status of a tooth surface is independent of the status of the other surfaces on that tooth, and that caries status of a tooth is independent of the status of other teeth of that individual. These assumptions are unlikely to hold in reality.

The next model fitted, model 1, is the two-level model, allowing for correlation between the results from multiple surfaces within the same individual, similar to that proposed by Sterne et al (Sterne et al., 1988) for a periodontal outcome.

$$\text{logit}(\pi_{ik}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \upsilon_{0k} + e_{0ik}$$
$$y_{ik} \sim Bin(1, \pi_{ik})$$
$$\text{var}(\upsilon_{0k}) = \sigma^2_{\nu 0}$$
$$\text{var}(y_{ik} \mid \pi_{ik}) = \pi_{ik}(1-\pi_{ik})$$

(Model 1)

The quantity σ^2_{v0} is the variance of the error term at the participant level, and is estimated in the modelling process.

Model 2 includes the tooth level in the random part of the model and has a three-level structure, allowing for surfaces within a tooth within an individual, as in Gunsolley et al (Gunsolley et al., 1994).

$$\begin{aligned}
&\text{logit}(\pi_{ijk}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \upsilon_{0k} + u_{0jk} + e_{0ijk} \\
&y_{ijk} \sim Bin(1, \pi_{ijk}) \\
&\text{var}(\upsilon_{0k}) = \sigma^2_{v0} \\
&\text{var}(u_{0jk}) = \sigma^2_{u0} \\
&\text{var}(y_{ik} \mid \pi_{ijk}) = \pi_{ijk}(1 - \pi_{ijk})
\end{aligned} \qquad \textbf{(Model 2)}$$

Here, σ^2_{v0} is the variance of the error term at the participant level, and σ^2_{u0} is the variance of the error term at the tooth level. Both quantities are estimated in the modelling process.

Each model was fitted with one explanatory variable denoting the treatment group. In addition, Model 2 was fitted using in turn each of the different classifications for tooth position as sets of indicator variables. Model 2 was also fitted with a set of indicator variables for each individual tooth, for comparison with the previously suggested classifications.

The ICCs for logistic models are not as straightforward to calculate as for models with continuous outcomes. Various approaches are possible, but the one chosen here is the threshold model (Snijders and Bosker, 1999). This model assumes that there exists an underlying continuous outcome variable where a value above some threshold corresponds to the value 1 in the dichotomous outcome variable. Since the

development of caries is a continuous process with a certain threshold at which an examiner will judge a lesion to be present, this model seems appropriate for these data.

Under the assumptions of a logistic regression model, the underlying continuous variable must follow a logistic distribution, which has variance $\frac{\pi^2}{3}$, which is approximately equal to 3.29 (Snijders and Bosker, 1999). This can then be used as the level 1 variance to calculate the ICC in the same way as for continuous outcomes.

5.3. Estimation in multilevel logistic models in MLwiN.

The multilevel analyses were performed using version 2 of the MLwiN software (Rasbash et al., 2005). This section gives details on how this software fits multilevel logistic models.

The standard two-level logistic model with binary outcome, as used in model 1, is given by the following equations.

$$\text{logit}(\pi_{ik}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \upsilon_{0k} + e_{0ik}$$
$$y_{ik} \sim Bin(1, \pi_{ik})$$
$$\text{var}(\upsilon_{0k}) = \sigma^2_{\nu 0}$$
$$\text{var}(y_{ik} \mid \pi_{ik}) = \pi_{ik}(1-\pi_{ik})$$

The outcome variable for a specific surface i, y_{ik}, which takes the value 0 if the surface does not develop caries and 1 if it does develop caries, is assumed to follow a binomial distribution with probability of success π_{ik}. This represents the probability that the ith surface becomes affected by caries during the course of the study. The logistic model uses the logit transformation on the left hand side of the equation, where $\text{logit}(\pi_{ik}) = \log(\frac{\pi_{ik}}{1-\pi_{ik}})$. The right hand side of the equation includes the intercept, β_0,

covariates x_1 to x_n, with associated coefficients β_1 to β_n. These form the fixed part of the model. The remaining terms in the equation, the random effect at level 2, v_{0k}, and the level 1 residual term e_{0ik} form the random part of the model. The quantity σ^2_{v0} is the variance of the random term at the participant level, and is estimated in the modelling process.

The unknown parameters to be estimated in this model are the β coefficients, and the variance of the random effect, σ^2_{v0}

5.3.1. Quasi-likelihood estimation

The default estimation methods in MLwiN are based on iterative generalised least squares (IGLS) (Goldstein, 1986). For models assuming a normal distribution, this procedure begins by generating initial values for the parameters in the fixed part of the model using ordinary least squares, then alternates between updating the estimates in the random and fixed parts of the model, using generalised least squares, until convergence, when the change in each parameter from one iteration to the next, is less than a given tolerance. The estimates from this method are biased, but a correction term can be applied to give unbiased estimates, which is particularly important in small samples. This method is known as restricted iterative generalised least squares (RIGLS) (Goldstein, 1989).

For models with other distributional assumptions, such as the logistic model considered here, one approach to estimation is to use a Taylor series expansion to linearise the model, and allow the IGLS or RIGLS procedures to be used.

The two main types of quasi-likelihood method differ based on which value is used for the Taylor series expansion. In marginal quasi-likelihood (MQL) the expansion is done

around zero, whereas in predictive quasi-likelihood (PQL) the expansion is done around the current estimates of the residuals. PQL is also sometimes known as penalised quasi-likelihood. Both methods also have 1st and 2nd order versions, depending on whether the Taylor series expansion includes 1st order terms only, or also incorporates 2nd order terms.

Both 1st order MQL, and 1st order PQL have been shown to be seriously biased, particularly when the underlying random parameter values are large, with the 2nd order MQL method only giving a modest improvement over its 1st order counterpart (Rodriguez and Goldman, 1995). However, in response to these results the 2nd order PQL method was developed, and was shown to be considerably less biased (Goldstein and Rasbash, 1996). Subsequent investigation has shown that 2nd order PQL can also show significant bias on certain datasets (Browne and Draper, 2006).

Another drawback of PQL is that it often fails to converge in cases where there are many level 2 units where the responses are all zero. This can often happen in models where there are few level 1 units in each level 2 unit (Goldstein, 2003). The three level model for caries data falls into this category, as each level 2 unit (tooth) will only have a maximum of 5 level 1 units (surfaces).

5.3.2. Markov Chain Monte Carlo Estimation

An alternative form of estimation available in MLwiN is the Markov Chain Monte Carlo (MCMC) method (Browne, 2005).

The modelling involves estimation of several parameters, the coefficients of the fixed part of the model, $\beta_0, \ldots, \beta_n$, and the variance from the random part, σ^2_{v0}, for two level models. These unknown parameters are denoted as a group by θ.

The software generates samples from the joint posterior distribution of these unknown parameters, using the likelihood form of Bayes' theorem, $p(\theta|y) \propto p(\theta)L(y|\theta)$. Here $L(y|\theta)$ is the likelihood function formed from the observed data y, and the distributional assumption used in the model. The default prior distributions used in MLwiN are uniform for fixed parameters, and gamma distributions for variances.

The estimation process requires starting values, so before the MCMC algorithm can be used, one of the quasi-likelihood methods described above must be used to generate these. The MCMC process then uses a combination of Gibbs sampling (Geman and Geman, 1984), and Metropolis-Hastings sampling (Hastings, 1970) to sample from the joint posterior distribution of the estimates.

Beginning with the starting values generated by the quasi-likelihood method, the Gibbs sampling method updates each parameter estimate in turn by sampling from the conditional posterior distribution of the parameter, given the observed data y, and the current values of all the other parameters. This new estimate then replaces the current value of that parameter, to be used in the conditional distribution for sampling the next parameter.

This method can be used for normally distributed models, as it is straightforward to write the conditional posterior distributions for both the fixed and random parameters. However, for more complex models, such as the logistic model considered here, the conditional distributions are more complicated, and an alternative approach is required, the Metropolis-Hastings method.

At each iteration, Metropolis-Hastings sampling forms a proposal distribution for the parameters, and samples from this distribution. The proposal distribution is usually a multivariate normal distribution. A rule is then applied to either accept or reject this

new set of parameters. This rule is based on calculating the ratio of the conditional probability of the new proposed estimates to the conditional probability of the previous set of estimates, known as r. If $r > 1$, then the new estimates are accepted. If $r < 1$, then the new estimates are accepted with probability r, using a random mechanism.

The aim of these MCMC iterations is not to generate individual point estimates for the values of the coefficients, but to generate a sample from the complete posterior distribution. Therefore, the procedure stores all the values generated, and this simulated distribution can be used to calculate both point estimates and coverage estimates for each parameter. As the chain may take some time to converge to the correct posterior distribution, a number of iterations from the beginning of the run are discarded, and not used in calculating the estimates. This is known as the burn-in period, and the default number of iterations in MLwiN is 500.

It is necessary to determine for how many iterations the MCMC routine should be run to give an acceptable accuracy of the estimates. This can be done using the Raftery-Lewis diagnostic (Raftery and Lewis, 1995). Once the routine has been run for around 5000 iterations, this statistic can be calculated. This gives the number of iterations that are required so that the 95% credible interval calculated from the simulated posterior distribution is accurate to within a given tolerance (default 1%). The procedure can then be continued for the required number of iterations, and the diagnostic checked again to ensure that convergence has been reached.

Simulation studies have shown that MCMC methods can give considerably more accurate estimates than quasi-likelihood methods (Browne and Draper, 2006), although they are considerably more computationally intensive.

The models in this chapter were fitted using the Markov Chain Monte-Carlo (MCMC) method of estimation. Convergence was checked using the Raftery-Lewis diagnostic to ensure that the chain was run for a sufficient number of iterations to estimate the 95% credible interval to an accuracy of 1%, with p=0.05.

5.4. Results

A total of 53146 surfaces from 465 participants were included in the analysis. There were a mean of 4.22 surfaces included per tooth (range 1 to 5), a mean of 114.29 surfaces per participant (range 49 to 128), and a mean of 27.09 teeth per participant (range 20 to 28).

Using the traditional method of analysis gives a mean caries increment of 10.63 (s.d. 8.07) in the varnish group, and 10.81 (s.d. 9.38) in the control group. These data would usually be analysed using a t-test given the relatively large sample size, giving no evidence of a difference between the groups.

Table 5.1 shows the estimated regression coefficients for models 0 to 2.

Model 0 is a logistic model treating all surfaces as independent observations. In this model the coefficient of group is given by –0.010. The associated odds ratio is 0.99. The interpretation of this is that the odds of a surface in the intervention group developing caries is almost identical to the odds of a surface in the control group developing caries, by a factor of 0.99. The 95% credible interval for this odds ratio is 0.94 to 1.05.

	Model 0 – single level logistic regression, surface only	**Model 1** – two-level model, individual - surface	**Model 2** – three-level model, individual – tooth - surface
Fixed part of model			
Constant β_0 (95% CrI)	-2.263 (-2.303,-2.223)	-2.547 (-2.674,-2.419))	-3.195 (-3.359,-3.034)
Coefficient of group covariate (95% CrI)	-0.010 (-0.066,0.049)	0.030 (-0.153,0.205)	0.018 (-0.206,0.242)
Odds ratio for group (95% CrI)	0.99 (0.94 ,1.05)	1.03 (0.86,1.23)	1.02 (0.81,1.27)
Random part of model			
Variance at individual level (95% CrI)		0.840 (0.713,0.985)	1.138 (0.955,1.347)
Variance at tooth level (95% CrI)			1.930 (1.732,2.134)
ICC estimate at individual level		0.20	0.18
ICC estimate at tooth level			0.30

Table 5.1 : Coefficients from models with different level structures

This interval contains 1, and therefore this model shows no evidence for a difference between the groups.

Model 1 adds individual as a source of variation, and the estimate of the variance of the random effect due to individual variation is 0.840. The ICC estimate shows that 20% of

the variance can be attributed to variations between individuals. The calculation of this value uses the value of 3.29 for the variance at surface level, from the threshold model described above. The ICC is the proportion of the variance attributable to the individual level, so is equal to 0.840 / (3.29 + 0.840) = 0.20. The estimate of the coefficient of group is higher than that derived from model 0 above, but the credible interval is much wider. This is to be expected, and is due to the added uncertainty of the effect of the individual. The odds ratio here is 1.03, and the 95% credible interval of 0.86 to 1.23 is much wider than in the previous model.

Model 2 adds tooth level to the analysis. Here, the ICC estimates show 18% of the variance attributable to the individual level, and 30% of the variance attributable to variation between teeth within individuals. In clinical practice, one may expect a higher percentage of the variance to be found between individuals. However, in this and most Phase III clinical trials, investigators seek to reduce variation attributable to the individual by pre-selecting people of similar age, socio-economic background and previous caries levels.

Table 5.2 shows the results for Model 2, with each of the covariates of tooth position, and the model including fixed effects for all teeth. The effects of group in all of these models are similar. The ICCs at tooth level are lower in these models than the 30% in the model without the tooth position covariates. This is due to the covariates explaining much of the variation at tooth level.

Table 5.3 shows the fixed effects of the three types of tooth position covariates, compared to the model with fixed effects for all teeth. The entries in the table are predicted probabilities (expressed as percentages) of a surface in the control group developing caries. An entry for a particular tooth can be interpreted as the predicted probability from the model, that a surface on that tooth, for a child in the control group

	Fixed effects for all teeth	Classification A	Classification B	Classification C
Fixed part of model				
Constant β_0 (95% CrI)	-3.081 (-3.394,-2.721)	-3.826 (-4.016,-3.634)	-3.834 (-4.054,-3.612)	-2.996 (-3.200,-2.771)
Coefficient of group covariate (95% CrI)	0.031 (-0.201,0.258)	0.028 (-0.202,0.259)	0.024 (-0.209,0.257)	0.036 (-0.193,0.267)
Odds ratio of group (95% CrI)	1.03 (0.82,1.29)	1.03 (0.83-1.30)	1.02 (0.81,1.29)	1.04 (0.82,1.31)
Random part of model				
Variance at individual level (95% CrI)	1.365 (1.101,1.627)	1.364 (1.157,1.602)	1.368 (1.162,1.606)	1.395 (1.174,1.639)
Variance at tooth level (95% CrI)	0.520 (0.001,0.738)	0.811 (0.693,0.941)	0.784 (0.659,0.909)	0.640 (0.125,0.779)
ICC estimate at individual level	0.26	0.25	0.25	0.26
ICC estimate at tooth level	0.10	0.15	0.14	0.12

Table 5.2 : Coefficients of models with tooth position covariates

will develop caries during the course of the trial. The predicted probabilities for the varnish group are very similar, and are not presented here. The top part of the table shows the predicted probabilities from the model with fixed effects for all teeth, and the bottom part shows the predicted probabilities for the three sets of tooth position covariates.

	Tooth number (position) within the quadrant						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Quadrant 1 (UR)	4	5	2	3	3	16	15
Quadrant 2 (UL)	5	5	2	2	4	15	14
Quadrant 3 (LL)	0	0	1	2	2	21	14
Quadrant 4 (LR)	1	1	1	1	4	25	18
All quadrants Teeth grouped by type (Classification A)	2		1	2		16	
All quadrants (classification B)	2	2	1	2	3	18	15
Quadrants 1 and 2 Upper Teeth (Classification C)	5		2	3		15	14
Quadrants 3 and 4 Lower Teeth (Classification C)	0		1	2		23	16

Table 5.3 Predicted probabilities (x100) of a surface in the control group developing caries from models with different tooth position covariates

The predicted probabilities from Classifications A and B are similar to each other, but both fail to account for some of the variation between individual teeth in the overall model, most noticeably the differences observed between the upper (quadrants 1 and 2) and lower (quadrants 3 and 4) incisors (tooth numbers 1 and 2), with the lower incisors being less likely to develop caries. Classification C treats the upper and lower teeth separately, and accounts for this difference. This is reflected in the ICCs, where

Classifications A and B have higher ICCs at tooth level (0.15 and 0.14) than Classification C (0.12), which has a closer ICC to the model with fixed effects for each tooth (0.10). Further investigation of model fit for these analyses will be presented in Chapter 7.

5.5. Discussion

Traditional analysis of caries data using the summary measure of caries increment at the individual level does not allow for any exploration of the data at tooth or surface level, although data is usually available at these levels. There may be a number of reasons why an investigator or healthcare provider would be interested in the efficacy of different methods, preventive agents or treatments at different levels. For example, understanding treatment effects or health benefit at level 3, i.e. for individuals allows a consideration for how a particular procedure could have different effect sizes for people of different ages, men or women, or whether people live in fluoridated water communities. Being able to quantify relative treatment effect sizes at the tooth level, may be important in looking at degenerative conditions that can present more commonly on different tooth types, for example the relative benefit of an erosion preventing agent on molars and incisors rather than on an overall whole mouth measure for a condition that may occur with a lower frequency on canines and premolars. Investigators commonly require to measure comparative benefit at the surface level, or level 1, for example being able to determine whether a preferential benefit is found on fissures as compared to smooth surfaces

The use of the threshold model for dental caries data allows straightforward modelling of relative variance estimates at the different levels of a multilevel logistic model. The model assumption that a dichotomous outcome is based on an underlying continuous

variable with a threshold point where the outcome changes from zero to one, is in reality how dental caries is classified in clinical trials, as caries is a continuous process, but is only identified as present in a clinical trial when it has reached a certain level of severity.

This chapter illustrated modelling issues using a data set from a group of adolescents, and the results here model the probabilities of a sound surface on a given tooth developing caries over the three year period between the approximate ages of 12 and 15, when the canines (position 3 within quadrants), premolars (positions 4 and 5), and second permanent molars (position 7) have recently erupted into the mouth. The other permanent teeth (incisors, positions 1 and 2, and first permanent molars, position 6) will have been in the mouth since around the age of 6 years. This age group is often included in caries clinical trials due to the large number of newly erupting teeth, which have the potential to become affected by caries. Due to the pattern of caries development, the tooth position results cannot be assumed to be generalisable to data from other age groups. Younger age groups will certainly show different patterns, due to the different teeth present, deciduous teeth in very young children, and a mixture of deciduous and permanent teeth in older children. In addition, the exclusion from the analysis of surfaces which already have caries at baseline means that models may differ in populations with higher or lower levels of existing disease. The finding of differing probabilities of caries according to differing tooth types, with molars most susceptible is well established and has been shown most commonly in analyses of cross-sectional survey data (Batchelor and Sheiham, 2004).

However, this work considers development and prevention of new carious lesions over time and the results also show the importance of differentiating between the upper and lower arches, when modelling the probabilities of caries developing on teeth,

particularly on the incisors. The lesser probability of developing caries in the lower arch is likely to be due to the protective effect of saliva, which contains calcium and phosphate and can be considered as a remineralising fluid as it collects in pools around the lower incisors (Featherstone, 2004). The model with fixed effects for all teeth also shows a potential difference between the left (quadrants 2 and 3) and right (quadrants 1 and 4) sides of the mouth in the probabilities of the molars developing caries, with slightly higher probabilities of developing caries in the right side of the mouth. The remineralising effect of saliva combined with topical fluorides e.g. from brushing with a fluoridated toothpaste can result in the repair of D_1 level lesions, so surfaces with caries at baseline can appear sound at follow-up (Kidd, 2005). However, the models here only predict caries at follow-up on surfaces considered sound at baseline, as surfaces with caries at baseline are excluded, so these reversals do not impact on the models.

The models here were fitted using MCMC estimates. The default method in the MLwiN software, is an iterative generalised least squares estimation, using either MQL (marginalised quasi-likelihood), or PQL (penalised-quasi-likelihood) approximations for transformation to a linear model. The first order MQL approximation is known to give estimates which are biased downwards, which could result in underestimates of the ICCs and treatment effects (Goldstein, 2003). In this data set, the MQL approximation of the variance at tooth level in model 2 was less than half that of the MCMC estimate. Second order PQL (penalised quasi-likelihood) methods are considered to give better estimates. However these are known to suffer from convergence problems (Goldstein, 2003), and in fact failed to converge for this data set. Therefore, although MCMC methods are far more computationally expensive, they give much more robust and reliable solutions, and should be the method of choice for this type of data.

The models used here are random intercept models, i.e. the probability of developing caries for a given surface varies depending on the tooth and the individual child. The treatment effect, however, is constrained to be equal for all teeth and individuals. These models can be extended to random slopes models, where the treatment effect is allowed to vary by tooth or by individual. If the data are to be analysed at the surface level, the clustering must be accounted for to avoid potential incorrect conclusions, as illustrated by the much wider confidence intervals for treatment effect observed in the multilevel models, compared to the single level model.

In summary, this chapter provides the first application of multilevel modelling to caries data from a clinical trial and has found that this statistical approach can increase understanding of the patterns of caries development within the mouth, and allows for full use of the data collected at surface level. Use of the threshold model allows estimation of the relative variances at individual, tooth and surface level, within a multilevel logistic regression model.

6. Prediction of patterns of caries incidence using multilevel modelling

6.1. Introduction

Studies of caries prediction have often come to the conclusion that past caries experience is the best predictor of the development of caries in the future. A review of the dental literature for caries prediction models (Powell, 1998) found that clinical variables such as the caries status of the most recently erupted teeth and surfaces were stronger predictors than non-clinical variables, and that past caries experience was the most significant predictor, with other important independent variables including socio-economic status, fluoride exposure, tooth morphology, and microbial agents.

6.1.1. Symmetry of caries development

It has long been a commonly held view by clinical observation that caries develops symmetrically in similar teeth on the right and left sides of the mouth. This apparent symmetry with respect to the midline was interpreted in the early part of last century, as evidence that caries was not an infective disease (Eckermann, 1919). A study of 300 bitewing radiographs (Scott, 1944), found that 73% of posterior decayed, missing or filled teeth were involved bilaterally (i.e. the corresponding tooth on the opposite side

of the mouth was also decayed, missing or filled). A longitudinal study of dental caries in 666 English schoolchildren (Berman and Slack, 1972) found "bilateral symmetry of caries attack at all ages."

Some more recent studies have challenged this view. A study of 510 children aged 12 years (Wood, 1985) demonstrates that, from a retrospective dental record study, 44% of maxillary and 33% of mandibular pairs of occlusal surfaces of first permanent molars showed caries experience unilaterally, i.e. only on one side of the mouth. A study of 15,132 adults, using data from the 1985-86 National Survey of Oral Health in the US (Hujoel et al., 1994b), found that 94% of adults with two or more decayed or filled surfaces had two or more "discordant pairs", where a decayed surface on one side of the mouth, had a sound counterpart on the laterally opposite side of the mouth. The study showed that the distribution of these pairs was not random with respect to the midline, and that the caries tended to be aggregated on one side of the mouth. A study using data from 20000 UK children aged 5 to 16 years (Batchelor and Sheiham, 2004) used probit analysis to rank surfaces by their susceptibility to caries. The study did not show precise symmetry between equivalent surfaces on the left and right sides of the mouth, but found that symmetry existed within groups of sites with similar susceptibility to caries. A recent study of deciduous teeth in 7074 children aged 3 to 7 years (Vanobbergen et al., 2007) concluded that associations of caries experience at the population level appeared to follow a symmetrical pattern, but at an individual level, when using the same method as in the US study above (Hujoel et al., 1994b) the study found similar results, i.e. that caries lesions tend to cluster on one side of the mouth.

6.1.2. Caries prediction

Although some of the studies above used longitudinal data sets, none have studied the issue of the symmetry of the disease over time, instead choosing a single timepoint and

examining the issue cross-sectionally. At a given time point, while there may be caries present on a surface, while the corresponding surface on the other side of the mouth is sound, which would be assessed as asymmetrical caries in a cross-sectional study at this particular timepoint, there is a possibility that the sound surface could develop caries at a later date.

Models of caries prediction generally assess the likelihood of a child developing caries anywhere in the mouth. However, it can also be of interest to predict individual teeth and surfaces which are likely to develop caries in order to inform the nature of the most effective intervention during a given time. For example, if fissure surfaces are most at risk, a fissure sealant programme may be optimal; but if approximal surfaces are most vulnerable and a fissure sealant policy is followed, the costs of repairing the approximal carious lesions present, which would usually involve access via the fissure, could outweigh the short-term benefit of the initial intervention with sealants. This may be confounded by the fact that the probabilities of caries development can vary significantly for different teeth, as seen in Chapter 5.

6.1.3. Multilevel modelling

Data collected at the surface level is clustered within teeth, which are clustered within individual participants. If statistical models use surface or tooth as the unit of analysis, then this clustering must be accounted for to avoid potential incorrect conclusions. A method which can be used to allow for this clustering is multilevel modelling, which in this case uses surface as the basic unit of analysis, but rather than assuming all surfaces are independent, assumes that the variability between surfaces affected can be explained as a combination of the variability at surface level, tooth level, and individual child level.

The purpose of this study is to investigate if caries on a given surface can be used as a predictor of future caries on the corresponding surface on the other side of the mouth. The importance of quantifying any such likelihood would be in the nature of the preventive intervention the dental care provider could give, and perhaps, more crucially in the nature of the very specific advice and quantification of risk that could be given to patients by their practitioner, so providing a significant and evidence-based motivator to self care.

6.2. Methods

The data come from a randomised controlled trial examining the caries preventive efficacy of a chlorhexidine varnish on the teeth of adolescents aged 12-14 years at baseline (mean age 13.4 years), and followed up over 3 years (Forgie et al., 2000). The study was conducted in Tayside, Scotland, an area with relatively high levels of caries, with mean DMFT in 14 year olds of 2.88, compared to a mean DMFT for the UK of 1.67 (Pitts et al., 2000). All schools in the region took part, and the participants selected for inclusion had high levels of mutans streptococci, and were thus regarded to be at high risk of developing caries. Chlorhexidine is an anti-bacterial agent, and the varnish is intended to reduce the development of caries by reducing the level of caries-associated microflora (Emilson, 1994). The intervention groups received applications of the varnish a minimum of once per year, with additional applications if they were found to have high levels of bacteria. The varnish was applied to all teeth. Dental examinations were performed by a single examiner, and intra-examiner reliability levels were good (kappa>0.80).

The 1200 participants in the trial were randomised to one of four groups. Group A received chlorhexidine varnish and dental health advice, group B received a placebo

varnish and dental health advice, group C received only dental health advice, and group D received no intervention. The study showed no significant difference in caries increment between the four groups, and for this analysis, all participants are treated as a single group.

The analysis here uses data from all participants who received both a baseline and a 3 year follow up examination.

The outcome variable is calculated for each surface, taking the value 1 if it was unaffected by caries at baseline, and had become affected by caries by the end of the study. If it was unaffected by caries at baseline, and remained unaffected by the end of the study, the increment takes the value 0. If the surface was affected by caries at baseline, and therefore cannot contribute to the outcome, it has been excluded from the analysis. This excluded 12930 of the 126336 possible surfaces (10%). The threshold for determining whether a surface was affected by caries was the D_1 level, which includes any visible caries, whether only in the enamel of the tooth, or extending into the dentine.

A logistic multilevel model was fitted using MLwiN software. Three levels were specified: individual participant, tooth, and surface. Only surfaces which were sound at baseline were included. The outcome variable was the development of caries into enamel or dentine, after three years. Covariates in the model were tooth position, the caries status of the contralateral surface at baseline, the caries status of the corresponding surface in the opposing jaw at baseline, whether caries was present on an adjacent tooth, and the individual's total number of decayed surfaces at baseline. In addition, interaction terms between tooth position and caries status of the opposite surfaces were investigated. The equations describing the multilevel model are shown here.

$$\text{logit}(\pi_{ijk}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \upsilon_k + u_{jk}$$

$$y_{ijk} \sim Bin(1, \pi_{ijk})$$

$$\text{var}(\upsilon_k) = \sigma_\upsilon^2$$

$$\text{var}(u_{jk}) = \sigma_u^2$$

$$\text{var}(y_{ijk} \mid \pi_{ijk}) = \pi_{ijk}(1 - \pi_{ijk})$$

The outcome variable for the ith surface on the jth tooth of the kth individual is denoted y_{ijk}, and takes the value 0 if the surface does not develop caries and 1 if it does develop caries. This variable is assumed to follow a binomial distribution with probability of success π_{ijk}. In this model, π_{ijk} represents the probability that the surface becomes affected by caries during the course of the study. The logistic model uses the logit transformation on the left hand side of the equation, where $\text{logit}(\pi_{ijk}) = \log(\frac{\pi_{ijk}}{1-\pi_{ijk}})$, to create a continuous outcome variable for use in the model.

The right hand side of the equation includes covariates x_1 to x_n with associated coefficients β_1 to β_n.

The multilevel model differs from a standard logistic regression model by including random effects at participant level (υ_k), and at tooth level (u_{jk}). The variances of the error terms at participant level, σ_υ^2, and tooth level, σ_u^2, are estimated in the model-fitting process.

The coefficients of logistic models can be interpreted using predicted probabilities. These represent the probability under the model assumptions, of a surface developing caries over the course of the study, given particular values of the covariates.

All analyses were performed using version 2 of the MLwiN software (Rasbash et al., 2005). The models were fitted using the Markov Chain Monte-Carlo (MCMC) method of estimation (Browne, 2005). This method constructs a chain of estimates which, if the chain is run for a sufficient number of iterations, will converge to the final estimate. Convergence was checked using the Raftery-Lewis diagnostic, which is an estimate of the required number of iterations to produce an estimate of a given accuracy. In this model, the chain was run for a sufficient number of iterations to estimate the 95% credible interval of all estimates to an accuracy of 1%, with p=0.05 (i.e. the probability that the estimates of the 95% CrI bounds are within 1% of the true value is less than 0.05).

6.3. Results

A total of 987 of the 1200 randomised children (82%) were included in the analysis, with 113406 surfaces which were sound at baseline. Of these surfaces, 4388 (4%) were decayed by 3 years. The dropouts had slightly higher caries levels at baseline (mean D_1MFS 13.9) than those who completed the study (mean D_1MFS 10.7).

The odds ratios from the final multilevel model are shown in Table 6.1. The model was initially run with all possible interaction terms between tooth position and the baseline caries status of the corresponding laterally opposite surface, corresponding surface in the opposing jaw, and adjacent teeth. The interaction terms which were significant were those between the caries status of the corresponding laterally opposite surface and the first molars, the caries status of the corresponding surface in the opposite jaw and the canines, and second molars, and caries status of the adjacent teeth and the first molars and upper second molars. All other interaction terms were removed to obtain the final model.

The effect of caries at baseline on the corresponding laterally opposite surface is shown to be highly significant (odds ratio 4.80, 95% credible interval 4.28 to 5.38). The effect of caries at baseline on the corresponding surface in the opposing jaw was also significant, but the effect was smaller in magnitude (odds ratio 1.66, 95% credible interval 1.49 to 1.83). The effect of caries at baseline on an adjacent tooth was also significant (odds ratio 1.89, 95% credible interval 1.74 to 2.05), but also smaller in magnitude than that of caries in the laterally opposite surface.

The effects of all tooth positions, compared to the reference category, the upper incisors, were significant. Odds ratios less than one indicate teeth where a given surface is less likely to develop caries than one on the upper incisors, and those greater than one indicate teeth where surfaces are more likely to develop caries.

The odds ratios for the interaction terms between the first molars and the baseline caries status of the corresponding laterally opposite surfaces are both less than one. This indicates that the effect on outcome of having baseline caries on the corresponding laterally opposite surface is smaller in the first molars than in the other teeth.

In contrast, the odds ratios for the interaction terms between the baseline caries of the corresponding surface in the opposing jaw and the canines and second molars were all greater than one. This indicates that the effect on outcome of having baseline caries in the corresponding surface in the opposing jaw is greater in these teeth than in the others.

The results are illustrated in Table 6.2, which shows the predicted probabilities of developing caries for a surface on each of the tooth classifications. These are shown for all four possible combinations of caries status of the corresponding laterally opposite surface, and caries status of the corresponding surface in the opposing jaw, where the

Variable	Odds ratio	95% CrI
Caries at baseline variables:		
On contralateral surface	4.80	(4.28, 5,38)
On corresponding surface in opposing jaw	1.66	(1.49, 1.83)
On adjacent tooth	1.89	(1.74, 2.05)
Number of decayed surfaces in mouth	1.06	(1.05, 1.06)
Tooth position variables		
Upper incisor (reference tooth)		
Upper canine	0.35	(0.30, 0.41)
Upper premolar	0.55	(0.49, 0.61)
Upper first molar	3.72	(3.27, 4.25)
Upper second molar	3.46	(2.97, 4.03)
Lower incisor	0.09	(0.07, 0.11)
Lower canine	0.16	(0.13, 0.20)
Lower premolar	0.41	(0.37, 0.46)
Lower first molar	6.42	(5.64, 7.32)
Lower second molar	2.54	(2.27, 2.83)
Interactions		
Upper first molar x Caries at baseline on contralateral surface	0.65	(0.53, 0.81)
Lower first molar x Caries at baseline on contralateral opposite surface	0.44	(0.36, 0.54)
Upper canine x Caries at baseline on corresponding surface in opposing jaw	2.89	(1.49, 5.55)
Lower canine x Caries at baseline on corresponding surface in opposing jaw	2.54	(1.41, 4.46)
Upper second molar x Caries at baseline on corresponding surface in opposing jaw	1.95	(1.57, 2.43)
Lower second molar x Caries at baseline on corresponding surface in opposing jaw	1.63	(1.24, 2.13)
Upper first molar x Caries at baseline on adjacent tooth	0.43	(0.36, 0.52)
Lower first molar x Caries at baseline on adjacent tooth	0.52	(0.43, 0.62)
Upper second molar x Caries at baseline on adjacent tooth	0.54	(0.45, 0.64)

Table 6.1: Odds ratios and 95% credible intervals from final model predicting the development of caries on a given surface during the study.

Caries status of laterally opposite surface at baseline	**Sound**	**Sound**	**Decayed**	**Decayed**
Caries status of corresponding surface in opposing jaw at baseline	**Sound**	**Decayed**	**Sound**	**Decayed**
Upper incisor	0.04	0.06	0.15	0.23
Upper canine	0.01	0.06	0.06	0.23
Upper premolar	0.02	0.03	0.09	0.14
Upper first molar	0.12	0.19	0.31	0.42
Upper second molar	0.12	0.30	0.39	0.67
Lower incisor	0.00	0.01	0.02	0.03
Lower canine	0.01	0.03	0.03	0.11
Lower premolar	0.02	0.03	0.07	0.11
Lower first molar	0.20	0.29	0.34	0.46
Lower second molar	0.09	0.21	0.32	0.55

Table 6.2 Predicted probabilities of a surface (sound at baseline) on a given tooth developing caries over the 3-year study, by baseline caries status of opposite surfaces, for a child with the mean number of decayed surfaces overall ($D_1MFS=10.7$), and no caries on the adjacent teeth.

other covariates are set at baseline DMFS=10.7 (the mean), and no caries present on adjacent teeth.

The predicted probabilities of surfaces on different teeth developing caries vary greatly, as expected. The relative effects of the status of the corresponding opposite surfaces

also vary as suggested by the significant interaction terms in the model. A surface on the lower incisors where the corresponding surface in the opposing jaw is also sound has a predicted probability of less than 1% of developing caries if the corresponding laterally opposite surface is sound at baseline, and 1% if decayed, whereas a surface on a upper first molar has a predicted probability of 12% of developing caries if both the corresponding laterally opposite surface and corresponding surface in the opposing jaw were sound at baseline, increasing to 67% if both were decayed.

The effect of the interaction terms for the first molars can be seen when comparing the predicted probabilities in Table 2 for the first and second molars. Clearly the differences between the predicted probabilities for those with and without a decayed corresponding laterally opposite surface are smaller for the first molars than for the second molars.

6.4.Discussion

Caries data is traditionally analysed at individual participant level, to avoid the issue of clustered data. The multilevel modelling technique used here allows the analysis to take place at tooth surface level, while directly modelling the clustering effects of the hierarchical structure. This allows both efficacy analyses and predictive models to explore the effects of covariates which operate at tooth or surface level, as well as those operating at participant level.

The differing probabilities of caries development in the various tooth types is well documented (Batchelor and Sheiham, 2004). The work here shows that there is also a marked difference in the probabilities of caries development on surfaces which have caries on the corresponding laterally opposite surface, and to a lesser degree, on the corresponding surface in the opposing jaw. This is particularly large in the upper second molars, where a surface has a 67% chance of developing caries in the following 3 years

when both the corresponding laterally opposite surface, and the corresponding surface in the opposing jaw are decayed at first examination between ages 12 and 14, compared to only 12% if both are sound. It is important to note that these results should not be generalised beyond the age group studied, due to the times of eruption of the different teeth. In addition, the exclusion of surfaces which have caries at baseline means that models may differ in populations with different caries levels. The participants in this study had a mean D_3MFT of 2.6 (standard deviation 2.9). A national UK survey of 14 year olds (Pitts et al., 2000) showed a mean D_3MFT for the UK of 1.8, so the group studied here have above average caries levels. However, the survey showed a mean D_3MFT for Scotland of 2.8, although as the surveyed children had a mean age of one year older than those in this study, it is likely that the study children have a slightly higher caries level than the overall population of Scotland. This is to be expected as the children were pre-selected based on levels of *Streptococcus mutans* found in the mouth.

Several other variables may impact on the development of caries, such as socio-economic status, fluoride exposure, tooth morphology, and microbial agents. These variables have not been included in the models here. Although some of these may be confounders, they are not readily available for this data set, and these models have been fitted to investigate the predictive power of existing caries patterns alone.

The significant interaction terms show that the effect of having caries on the corresponding laterally opposite surface is reduced on the first molars. This is likely to be due to the fact that the first molars are the first permanent teeth to erupt into the mouth, and do so several years before the canines, premolars and second molars. Since the children in this dataset are aged 12-14 at baseline, their first molars will have been erupted for 6-8 years by this point, whereas the other posterior permanent teeth will be fairly newly erupted, or indeed still be unerupted at the initial examination. Therefore,

there is a greater chance of caries on the first molars having been present for several years, having been filled or the tooth extracted. This historical caries is less likely to be a predictor of new caries on the same type of tooth than the more recent caries in the newer teeth, as teeth are particularly vulnerable to caries in the first few years after eruption (Carlos and Gittelsohn, 1965). In fact 91% of the children included here had some caries experience on the first molars at baseline. If this study was carried out at a younger age, it is likely that the effect on the first molars would be greater. This does not apply to the other teeth which have been present for several years, the incisors, as they are far less likely to have developed caries at a young age (Hannigan et al., 2000).

Conversely, the increased effect of caries on the corresponding surface in the opposing jaw for canines and second molars is likely to be due to the fact that these teeth have only recently erupted into the mouth at baseline. A child who has already developed caries in these teeth may be at greater risk of developing further caries in the next 2 years (Powell, 1998).

The remineralising effect of fluoride can result in reversals of carious lesions, with lesions diagnosed with D_1 level caries becoming apparently sound at a later date. This can cause problems with models of caries progression, but as all surfaces with caries at baseline were excluded from this analysis, these do not impact on the model.

In conclusion, this multilevel modelling technique provides a clinically useful method of estimating the probability of a surface developing caries over a period of time, based on the caries status of the laterally opposite surface, and the corresponding surface in the opposing jaw, while controlling for the natural clustering in tooth surface data.

This information can help inform preventive treatment planning decisions, by indicating particular surfaces which may be at risk in the short term for targeted interventions such

as fissure sealants. The use of whole-mouth treatments such as fluoride varnishes may also be indicated where there is caries on one side of the mouth which is not mirrored on the opposite side.

7. Investigation of model fit

7.1. Introduction

In Chapter 5, various specifications of multilevel model were applied to a data set from a clinical trial of a caries preventive agent. This chapter discusses aspects of model fit for the multilevel models applied to these data. The concept of assessing comparative model fit using the Akaike Information Criterion (AIC) and the Deviance Information Criterion (DIC) is introduced, and the various models assessed using this method. The residuals from the models are examined, leading to investigation of the possibility of fitting the zero-inflated binomial model to the data.

7.2. Comparative model fit using AIC and DIC

Goodness of fit in statistical models is often assessed using the likelihood function. The likelihood statistic $-2\log(\text{likelihood})$ can be calculated for a fitted model. This statistic can also be calculated for the full model, where a parameter is included for every observation, so the data fits the model exactly. The difference between the likelihood statistic for the fitted model and the full model is known as the deviance.

If two fitted models are nested, i.e. all parameters in the first model are also included in the second, then the difference in the deviances can be compared to a chi-squared distribution with degrees of freedom equal to the additional number of parameters in

the second model, to test whether the additional parameters significantly improve the model. However, if the models are not nested, this technique cannot be used. In this case, models can be compared using methods such as the Akaike Information Criterion.

7.2.1. Akaike Information Criterion

The Akaike Information Criterion, or AIC (Akaike, 1974) is a general method of comparing statistical models to assess which has the best fit to a given data set. The AIC is calculated using the formula $AIC = -2\log(\text{likelihood}) + 2k$, where k is the number of parameters in the model. A model with a lower AIC is considered to fit better. The statistic is designed to balance between accuracy (measured using the likelihood statistic) and the complexity of the model (measured by the number of parameters).

7.2.2. Deviance Information Criterion

The Deviance Information Criterion, or DIC (Spiegelhalter et al., 2002) is a generalisation of the AIC which is particularly useful in the multilevel models fitted here, as it is straightforward to calculate from the MCMC estimation used in MLwiN. First, a quantity known as $\bar{D}$ is calculated by calculating the deviance at every iteration of the MCMC run, and taking the mean. This is used as the measure of accuracy of the model, as with the likelihood statistic in the AIC. Calculating the deviance at the expected value of the unknown parameters, $D(\overline{\theta})$, and subtracting this from $\bar{D}$, gives the quantity p_d, which is the effective number of parameters, giving a measure of complexity. The DIC is then calculated as $\bar{D} + p_d$, the sum of terms measuring the accuracy and complexity of the model. As with the AIC, lower values of the DIC represent improved model fit.

Comparison of multilevel models using DIC

Table 7.1 shows the values of the DIC calculated for all models fitted to the data set in Chapter 5. The first 3 models did not include the tooth position covariates, and were comparing the different level structures. It can be seen from the table that the three level model has a better fit than either the two level model ignoring tooth level, or the simple logistic regression with no multilevel structure. The last four models compare the fitted model with different classifications of tooth position. Both classification A, which categorised teeth into incisors, canines, premolars and molars, and classification B, which categorised teeth into seven categories (central incisor, lateral incisor, canine, first premolar, second premolar, first molar and second molar) had a higher DIC than the model which included fixed effects for all teeth. Neither of these classifications separated the upper and lower jaw, whereas classification C categorised teeth into incisors, canines, premolars, first molars and second molars, separately for each jaw, giving a total of ten categories. This model had the lowest DIC of all fitted models, confirming that this model the best fit of all models tested.

7.3. Examination of residuals

The three level model with hierarchy participant-tooth-surface can be written as the following equation.

$$\text{logit}(\pi_{ijk}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \upsilon_{0k} + u_{0jk} + e_{0ijk}$$
$$y_{ijk} \sim Bin(1, \pi_{ijk})$$
$$\text{var}(\upsilon_{0k}) = \sigma^2_{\nu 0}$$
$$\text{var}(u_{0jk}) = \sigma^2_{u0}$$
$$\text{var}(y_{ijk} \mid \pi_{ijk}) = \pi_{ijk}(1 - \pi_{ijk})$$

Model 1

Model	DIC
Models without tooth position covariates	
Single level logistic regression, surface only,	33090.50
Two level model, individual – surface	30198.30
Three level model – individual – tooth surface	27656.18
Three level models with tooth position covariates	
Fixed effects for all teeth	25381.29
Classification A (four categories)	25717.33
Classification B (seven categories)	25677.81
Classification C (ten categories)	25283.02

Table 7.1: DIC values for all models fitted in Chapter 5

The outcome variable for the ith surface on the jth tooth in the kth individual, y_{ijk} takes the value 0 if the surface does not develop caries and 1 if it does develop caries, and is assumed to follow a single trial binomial distribution with probability of success π_{ijk}. The model predicts π_{ijk}, which represents the probability that the ith surface on the jth tooth in the kth individual becomes affected by caries during the course of the study. The logistic model uses the logit transformation on the left hand side of the equation, where $\text{logit}(\pi_i) = \log(\frac{\pi_i}{1-\pi_i})$. The right hand side of the equation can be separated into the fixed part of the model, and the random part of the model. The fixed part of the model includes covariates x_1 to x_n with associated coefficients β_1 to β_n. The random part includes the random effect of participant, υ_{0k}, and the random effect

of tooth, u_{0jk}. Also σ^2_{v0} is the variance of the random effect at the participant level, and σ^2_{u0} is the variance of the random effect at the tooth level. Both variances are estimated in the modelling process.

The model includes intervention group and tooth position as covariates. As the effect of intervention group was very small and non-significant, it has been excluded from the analyses in this section, and only the covariate describing tooth position included. This categorical variable has ten classifications for tooth position, incisors, canines, premolars, first molars, and second molars, each separately for the upper and lower jaw.

The coefficients from the fitted model are shown in Table 7.2, with those from the fixed part expressed as odds ratios, with 95% credible intervals produced from the MCMC estimation. The predicted probabilities in the table represent the probability that a sound surface at baseline on the given tooth will develop caries over the 3 years of the study. These probabilities are predicted from the fixed effects in the model only, which corresponds to the means of the distributions of the random effects i.e. at $v_{0k} = u_{0jk}$ =0.

In order to investigate model fit, the crude residuals have been calculated from the model, as the observed value minus the predicted value. The predicted values are calculated using both the fixed and random parts of the model. Figure 7.1 shows the distribution of residuals.

Fixed part			
	exp (ß)	95% CrI	Predicted probability
Upper incisors (reference category)			0.05
Upper canines	0.39	(0.32 ,0.48)	0.02
Upper premolars	0.57	(0.49 ,0.66)	0.03
Upper first molars	3.58	(3.06 ,4.19)	0.15
Upper second molars	3.36	(2.90 ,3.90)	0.15
Lower incisors	0.09	(0.07 ,0.12)	0.00
Lower canines	0.16	(0.12 ,0.21)	0.01
Lower premolars	0.42	(0.36 ,0.49)	0.02
Lower first molars	5.97	(5.10 ,6.98)	0.23
Lower second molars	3.81	(3.29 ,4.42)	0.16
Random part			
	σ	95% CrI	
Tooth level	1.393	(1.186, 1.632)	
Participant level	0.646	(0.530, 0.764)	

Table 7.2: Coefficients from Model 1

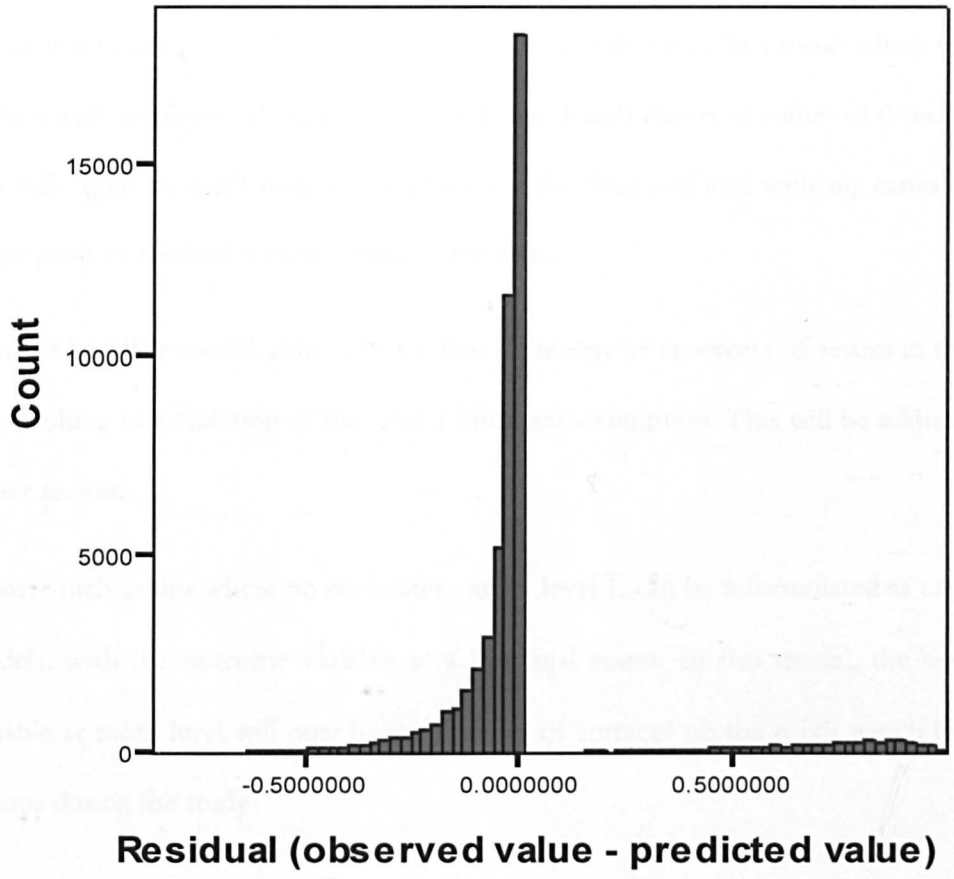

Figure 7.1 : Distribution of crude residuals from Model 1.

The left hand side of this graph corresponds to those cases where the observed value is zero, and seems to be reasonably distributed with most cases having a residual close to zero, with a tail of more negative values. However, the graph is very asymmetrical, with the right hand side of the graph, corresponding to those cases with an observed value of 1, being grouped together at the upper end of the scale. This suggests that the model may be underestimating the probabilities of surfaces developing caries.

One possible contributory factor to this is that the model does not differentiate between surfaces. None of the covariates in this model vary at level 1, as the tooth

position variable only varies at level 2, the tooth level. Therefore, where certain surfaces on a particular tooth may have differing probabilities of developing caries, the predicted values from the model will be an average for the whole tooth. In a tooth which has one surface with an observed value of 1 (caries), and 4 with observed values of 0 (no caries) this will result in small negative residuals for the four surfaces with no caries, and a larger positive residual for the surface with caries.

Another possible contributory factor is that there may be an excess of zeroes in the data set, resulting in a violation of the level 1 binomial assumption. This will be addressed in a later section.

Models such as this where no covariates vary at level 1, can be reformulated as two level models, with the outcome variable as a binomial count. In this model, the outcome variable at tooth level will now be the number of surfaces on the tooth which become carious during the study.

$$\text{logit}(\pi_{jk}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \upsilon_{0k} + u_{0jk} + e_{0jk}$$
$$y_{jk} \sim Bin(n_{jk}, \pi_{jk})$$
$$\text{var}(\upsilon_{0k}) = \sigma^2_{\nu 0}$$
$$\text{var}(u_{0jk}) = \sigma^2_{u0}$$
$$\text{var}(y_{jk} \mid \pi_{jk}) = n_{jk}\pi_{jk}(1-\pi_{jk})$$

Model 2

Here, n_{jk} is the number of surfaces on tooth j in individual k which are sound at baseline, and thus included in the analysis. The model still allows us to include a random effect at tooth level, which may seem counterintuitive, as we already have a residual error term at tooth level, e_{0jk}. However, since each tooth has an associated count value, of the number of sound surfaces at baseline, and since the level 1 variance depends on the values of the covariates, it is possible to model this random effect. If all n_{jk} values

were equal, and there were no covariates, it would be impossible to model the random effect at tooth level.

This model is in fact, equivalent to model 1, and produces identical parameter estimates to those shown in Table 7.2. However, as the outcome variable is measured at tooth level rather than surface level, the residuals are different. Residuals from model 2 are shown in Figure 7.2.

It can be seen from this graph that rewriting the outcome variable as the number of surfaces which become carious results in more symmetric residuals. It appears that the model is still underestimating the number of surfaces which develop caries, as the right hand side of the distribution has a peak above zero.

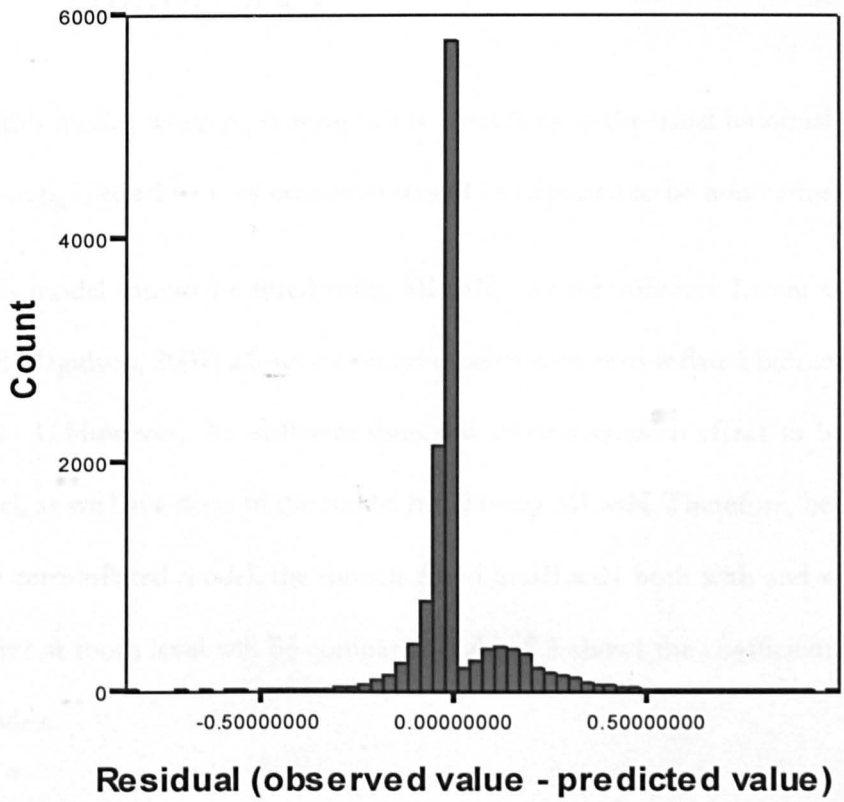

Figure 7.2: Distribution of crude residuals from Model 2.

7.4. Zero inflation

Another factor which could be contributing to the underestimation of the probabilities of caries incidence is an excess of zeroes in the data. The current model assumes that the data follows a binomial distribution at level 1. If the observed data has more teeth which do not develop caries during the study (i.e. have an outcome of zero) than would be expected from the binomial distribution, the data is zero-inflated. This can be modelled by assuming a zero-inflated binomial distribution at level 1, rather than the previous binomial assumption (Hall, 2000). The zero inflated distribution differs from the standard binomial distribution by adding a zero-inflation parameter, given by p_{jk} for the jth tooth in the kth individual. The distribution is then given by:

$$y_{jk} \sim \begin{cases} 0 & \text{with probability } p_{jk} \\ Bin(n_{jk}, \pi_{jk}) & \text{with probability } 1\text{-}p_{jk} \end{cases}$$

In this model, when p_{jk} is zero, this is equivalent to the usual binomial distribution, and when p_{jk} is equal to 1, all outcomes would be expected to be non-carious.

This model cannot be fitted using MLwiN, but the software Latent GOLD (Vermunt and Magidson, 2005) allows two-level models with zero-inflated binomial distribution at level 1. However, this software does not allow a random effect to be added at tooth level, as we have done in the model fitted using MLwiN. Therefore, before investigating the zero-inflated model, the models fitted in MLwiN both with and without a random effect at tooth level will be compared. Table 7.3 shows the coefficients from these two models.

Fixed part	**With random effect at tooth level**	**Without random effect at tooth level**
	exp (ß) (95% CrI)	exp (ß) (95% CrI)
Upper incisors (reference category)		
Upper canines	0.39 (0.32, 0.49)	0.43 (0.36, 0.52)
Upper premolars	0.57 (0.49, 0.66)	0.59 (0.52, 0.67)
Upper first molars	3.57 (3.06, 4.19)	3.10 (2.72, 3.53)
Upper second molars	3.36 (2.90, 3.90)	2.91 (2.59, 3.30)
Lower incisors	0.09 (0.07, 0.12)	0.11 (0.09, 0.14)
Lower canines	0.16 (0.12, 0.21)	0.18 (0.14, 0.23)
Lower premolars	0.42 (0.36, 0.49)	0.45 (0.39, 0.51)
Lower first molars	5.97 (5.10, 6.98)	5.01 (4.41, 5.71)
Lower second molars	3.81 (3.29, 4.42)	3.35 (2.97, 3.79)
Random part		
	σ	σ
Tooth level	0.65 (0.53, 0.76)	
Participant level	1.39 (1.19, 1.63)	1.19 (1.02, 1.39)

Table 7.3 : Comparison of parameter estimates for binomial count models fitted in MLwiN with and without random effects at tooth level

The residuals from the two models are shown in Figure 7.3.

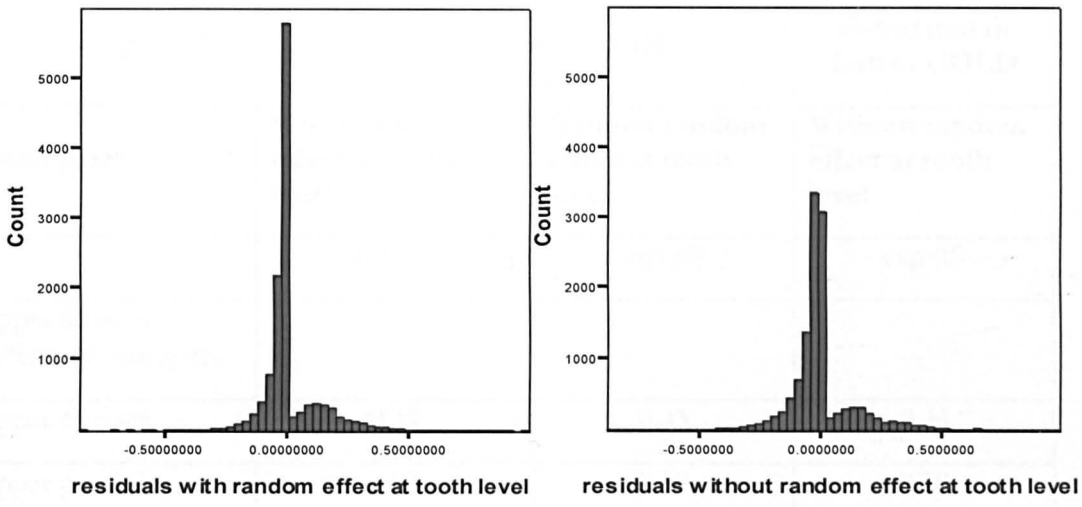

Figure 7.3: Comparison of residuals from models with and without random effects at tooth level

This figure shows that although the shapes of the graphs are similar, the model with random effects at tooth level has a much higher peak at the point immediately below zero, corresponding to accurate predictions of outcome in teeth which do not develop caries during the study. The model with random effects at tooth level appears to have a better fit to the data than the model without. The model with random effects at tooth level has a DIC of 15427, and the model without random effects at tooth level has a DIC of 15954, confirming that the model with random effects at tooth level is a better fit to the data than the model without.

7.4.1. Estimation of model using Latent GOLD

The estimation of the model using the Latent GOLD software differs from the MCMC routine in MLwiN, as Latent GOLD uses a different method of estimation, a combination of EM and Newton-Raphson methods. Table 7.4 extends Table 7.3 by

showing the parameter estimates from the model without random effect at tooth level fitted using Latent GOLD.

	Estimated in MLwiN		Estimated in Latent GOLD
Fixed part	**With random effect at tooth level**	**Without random effect at tooth level**	**Without random effect at tooth level**
	exp (ß)	exp (ß)	exp (ß)
Upper incisors (reference category)			
Upper canines	0.39	0.43	0.44
Upper premolars	0.57	0.59	0.59
Upper first molars	3.58	3.10	3.08
Upper second molars	3.36	2.92	2.90
Lower incisors	0.09	0.11	0.11
Lower canines	0.16	0.18	0.18
Lower premolars	0.42	0.45	0.45
Lower first molars	5.97	5.01	4.98
Lower second molars	3.81	3.35	3.33
Random part			
	σ	σ	σ
Tooth level	0.646		
Participant level	1.393	1.193	1.056

Table 7.4 : Comparison of parameter estimates for binomial count models fitted in MLwiN with and without random effects at tooth level, and fitted in Latent GOLD without random effects at tooth level

It can be seen from this table that the parameter estimates from Latent GOLD are fairly close to those of the same model fitted in MLwiN. Latent GOLD has the advantage that estimation is much quicker, with this model being fitted in under a minute, compared to several hours for the MCMC routine in MLwiN.

7.4.2. The zero-inflated binomial model

The model in Latent GOLD can now be extended to assume a zero-inflated binomial distribution at level 1. This model has additional parameters which estimate the probability that a given tooth will have a guaranteed zero outcome, and not be affected by the rest of the model. This will be referred to as the probability of being in the zero part of the model. The other part of the model, which has a binomial distribution will be referred to as the non-zero part of the model, although it should be noted that this part of the model can also predict zero outcomes, within the binomial distribution.

Rather than assume a constant probability of being in the zero-part of the model for all teeth, we can also allow this to vary with the covariates in the model. Therefore, tooth position will also predict the probability of being in the zero part of the model.

Table 7.5 shows the parameter estimates from this model. These are not directly comparable to the parameter estimates from the previous model, as the predictions come from a combination of the probability of being in the zero part of the model, and the prediction from the fixed part of the model. The table also shows the probabilities of falling into the zero part of the model for each value of the tooth position variable.

Fixed part for non-zero part of model	**With random effect at tooth level**
	exp (ß)
Upper incisors (reference category)	
Upper canines	0.57
Upper premolars	0.60
Upper first molars	1.09
Upper second molars	1.10
Lower incisors	0.70
Lower canines	0.35
Lower premolars	0.47
Lower first molars	1.98
Lower second molars	1.30
Random part for non-zero part of model	
	σ
Participant level	0.70
Probabilities of being in zero part of model	
Upper incisors (reference category)	0.64
Upper canines	0.77
Upper premolars	0.67
Upper first molars	0.01
Upper second molars	0.06
Lower incisors	0.95
Lower canines	0.85
Lower premolars	0.70
Lower first molars	0.16
Lower second molars	0.11

Table 7.5: Parameter estimates from zero-inflated binomial model

Figure 7.4 shows the comparison of the residuals from the two models fitted using Latent GOLD.

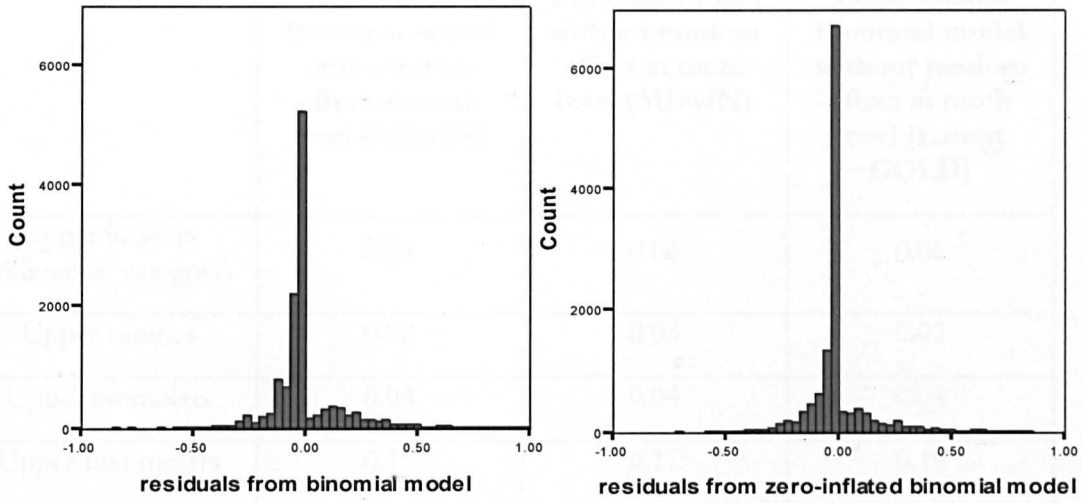

Figure 7.4 : Comparison of crude residuals (observed value – predicted value) from binomial model, and zero inflated binomial model, fitted in Latent GOLD.

These graphs show that the zero-inflated binomial model appears to be a better fit to the data than the non-zero inflated model with a higher peak around zero, and a flattening of the second peak on the right hand side of the graph. Comparing the models using the AIC gives 16026 for the zero-inflated model, and 16604 for the non-zero inflated model, confirming that the zero-inflated binomial model is a better fit to the data. The AIC has been used rather than the DIC as Latent GOLD does not use the MCMC estimation method.

In order to compare the results of the zero-inflated binomial model with those from the binomial models, we can calculate predicted probabilities of developing caries from

each model. For the zero-inflated binomial model, these are derived from a combination of the probability of being in the non-zero part of the model, and the predicted probability of a tooth in the non-zero part of the model developing caries. The probabilities are shown in Table 7.6.

	Binomial model with random effect at tooth level (MLwiN)	**Binomial model without random effect at tooth level (MLwiN)**	**Zero-inflated binomial model without random effect at tooth level (Latent GOLD)**
Upper incisors (reference category)	0.05	0.06	0.06
Upper canines	0.02	0.03	0.03
Upper premolars	0.03	0.04	0.04
Upper first molars	0.15	0.17	0.19
Upper second molars	0.15	0.16	0.18
Lower incisors	0.00	0.01	0.01
Lower canines	0.01	0.01	0.01
Lower premolars	0.02	0.03	0.03
Lower first molars	0.23	0.25	0.26
Lower second molars	0.16	0.18	0.20

Table 7.6 : Predicted probabilities from binomial models with and without random effects, and from zero-inflated binomial model.

This table shows differences between the predicted probabilities for the model with random effects at tooth level, and the zero-inflated binomial model, particularly on the molars. However, the differences between the two models without random effects at tooth level, with and without zero-inflation, are closer.

In summary, software limitations have affected the model fit investigations. Use of MLwiN shows that the binomial model with a random effect at tooth level is a better fit than the model without. Use of Latent GOLD shows that a zero-inflated binomial model is a better fit to the data than a standard binomial. A zero-inflated binomial model with random effect at tooth level may be the best fit to the data, but neither package is capable of fitting this model.

8. Simulation study

8.1. Introduction

One of the criticisms of analysing caries data using the DMF index, is that the detail of the observations at tooth and surface level are lost. By increasing the number of observations analysed, there is the potential for increased efficiency in caries clinical trials.

The analysis of model fit in Chapter 7 has suggested that a 3-level model, with random effects at both tooth level and individual participant level, is a better fit to the data than a 2-level model, with only random effect at individual participant level. However, the two level model is computationally easier to fit, and the procedure is available in a wider range of software than three level modelling. The two level model also avoids problems which can be caused by fitting models with few observations in each level 2 unit (Goldstein, 2003).

The purpose of this simulation study is to assess the power of analysis of caries data using multilevel modelling, compared to the traditional caries increment analysis. In addition, the effect of ignoring tooth level in the multilevel structure is investigated, to assess robustness of the data to this method of analysis, which is computationally simpler, and may be easier to interpret.

8.2. Simulation methods

The basis of the simulation study is the following model:

$$\text{logit}(\pi_{ijk}) = \beta_0 + \beta_1 x_1 + \sum_{m=2}^{10} \beta_m x_m + \upsilon_{0k} + u_{0jk} + e_{0ijk}$$

$$y_{ijk} \sim Bin(1, \pi_{ijk})$$

$$\text{var}(\upsilon_{0k}) = \sigma^2_{\nu 0}$$

$$\text{var}(u_{0jk}) = \sigma^2_{u0}$$

$$\text{var}(y_{ik} \mid \pi_{ijk}) = \pi_{ijk}(1 - \pi_{ijk})$$

In this model x_1 is the indicator variable for group allocation, and x_2, ...,x_{10} are indicator variables for the tooth position classification from previous chapters (Classification C in Chapter 5), with upper incisor as the reference category.

The data set that has been used in this thesis comes from a clinical trial which showed no significant group effect. In order to investigate whether a significant group effect would have an effect on the other parameters in the model, the data set analysed selected a different subgroup of participants.

The data set here includes all participants in the two groups who received chlorhexidine or placebo varnish from the clinical trial data set used in previous chapters. As no significant treatment effect was found in this trial, the effect of compliance to the varnish protocol has been used in the model instead of group allocation. There was a significant difference found in D_1FS caries increment between those who complied with the protocol (mean=9 .89, s.d.=7.59), and those who did not (mean=12.19, s.d.=9.23).

Clearly this is not an effect due to a caries preventive agent, but has been used here to generate two subgroups with significantly different caries levels.

Published work on the estimation methods in MLwiN (Browne and Draper, 2006) show that the default estimation method for multilevel logistic regression, 1st order MQL, is significantly biased, with the 2nd order PQL method, giving more accurate estimates, although still sometimes biased, particularly in estimating random effect . Another disadvantage of PQL is that it can fail to converge in some models. The MCMC routine has been shown to give the least biased estimates, but this method takes a prohibitively long time to converge, which makes it unrealistic for large scale simulations. The simulations here will be performed using 2nd order PQL estimation.

In order to check the likely accuracy of the PQL method, models have been fitted using the three estimation methods on the data set described above. Table 8.1 shows the fitted coefficients from the three level logistic multilevel model, fitted using the three estimation methods.

The PQL estimates here are mostly very similar to the MCMC estimates, with the exception of the random effect at tooth level where the PQL estimate is higher than the MCMC estimate. The 1st order MQL results are clearly very different from the MCMC results.

8.2.1. Choice of simulation parameters

In order to assess power in different situations, the simulations were run with various numbers of simulated individuals per group, and with various coefficients of intervention effect. As the effect of ignoring tooth level is also of interest, various random effects at tooth level were also simulated.

The parameters which vary in the simulations are number of individuals per group, coefficient of intervention effect, and variance of the random effect at tooth level. The values of the first two parameters were chosen based on observed values in trials

	1st order MQL	2nd order PQL	MCMC
Constant	-2.390 (0.072)	-3.247 (0.090)	-3.143 (0.087)
Group (compliant or not)	0.351 (0.115)	0.417 (0.130)	0.407 (0.127)
Upper incisor (reference category)			
Upper canine	-0.854 (0.093)	-1.040 (0.133)	-1.038 (0.113)
Upper premolar	-0.561 (0.064)	-0.655 (0.093)	-0.650 (0.080)
Upper 1st molar	1.107 (0.061)	1.439 (0.088)	1.420 (0.080)
Upper 2nd molar	0.891 (0.059)	1.200 (0.085)	1.183 (0.077)
Lower incisor	-2.109 (0.112)	-2.510 (0.160)	-2.503 (0.131)
Lower canine	-1.653 (0.126)	-1.928 (0.177)	-1.941 (0.145)
Lower premolar	-0.730 (0.067)	-0.883 (0.096)	-0.877 (0.082)
Lower 1st molar	1.467 (0.061)	1.930 (0.088)	1.886 (0.080)
Lower 2nd molar	1.046 (0.059)	1.361 (0.085)	1.338 (0.077)
Random effect – participant	1.175 (0.083)	1.332 (0.105)	1.276 (0.104)
Random effect - tooth	0.301 (0.035)	1.099 (0.063)	0.806 (0.066)

Table 8.1 : Coefficients from three different estimation methods in MLwiN.

reported in the Cochrane review on topical fluorides for preventing dental caries in children and adolescents (Marinho et al., 2003a). The review included 133 trials, with group size varying from 10 to 708. The median group size was 158, with quartiles at 93 and 254. The simulations will use values of 50, 150 and 300 individuals per group. The intervention effect in the 133 trials ranged from a prevented fraction (PF) of 0 to 80%. The prevented fraction is defined as the difference in caries increment between

intervention and control group, expressed as a percentage of the increment in the control group. The median PF was 25%, with quartiles of 17% and 32%. The three values chosen for the simulation are 15%, 25% and 35%. The observed coefficient of group effect in Table 8.1 (0.417), corresponds to a PF of 25%. As the trials in the Cochrane review generally do not use multilevel analysis, the values for the third parameter, variance of random effect at tooth level, cannot be chosen in the same way. The observed variance of the random effect at tooth level in the model fitted above was 1.099, from the 2nd order PQL analysis, and 0.806 from the MCMC analysis. The values for the simulation were chosen as 0 (no random effect at tooth level), 1 and 2. This gives a total of 27 combinations of parameter values.

It is also of interest to consider the effect of zero-inflation on the analysis, but due to serious convergence issues on this data set in Latent GOLD, the software which fits the zero-inflated binomial model, this has not been considered in this work, and will be returned to in the future (see the section in Chapter 9 on further work).

The coefficients of tooth position, and the random effect at individual participant level are kept constant for each simulation at the values from the 2nd order PQL model shown in Table 8.1.

1500 data sets were simulated for each of the 27 combinations of parameters detailed above using the MLwiN syntax language. For each simulated participant, predicted values of the outcome variable for each tooth surface were calculated based on the model parameters. Random effects at individual and tooth level were then applied to these predicted values, resulting in a probability of decay for each surface. The data were then sampled from the binomial distributions with these probabilities to obtain the simulated data sets. The two-level model ignoring the random effect at tooth level was fitted to each of the data sets. In addition, the full three-level model was also fitted to

each data set. All models were fitted using the second order PQL method. As PQL can have convergence problems (Goldstein, 2003) the maximum number of iterations was set at 100 for each data set, and if this maximum was reached, the model for that data set was considered to have failed to converge. All models which did converge, did so in fewer than 50 iterations. Finally, the traditional analysis of comparing DFS caries increment using t-tests was performed on each simulated data set. As a check, data sets were also simulated with no treatment effect, to check the value of α.

8.3. Results

8.3.1. Convergence

All the 2-level models successfully converged to a solution within the limit of 100 iterations. Some convergence problems were encountered in the 3-level models, and Table 8.2 summarises these.

50 per group	Random effect variance 0	Random effect variance 1	Random effect variance 2
PF 15%	100%	72%	0%
PF 25%	100%	84%	1%
PF 35%	100%	91%	3%
150 per group	Random effect variance 0	Random effect variance 1	Random effect variance 2
PF 15%	100%	72%	0%
PF 25%	100%	94%	0%
PF 35%	100%	100%	0%
300 per group	Random effect variance 0	Random effect variance 1	Random effect variance 2
PF 15%	100%	86%	0%
PF 25%	100%	99%	0%
PF 35%	100%	100%	0%

Table 8.2 : Percentage of 3-level models which successfully converged within 100 iterations

All models converged for the simulations where there was no random effect at tooth level. Where the variance of the random effect was 2, there were major convergence problems, and very few of the models successfully converged within 100 iterations. Of the models which did converge, none took more than 50 iterations. The models with random effect of variance 1, which is the closest value to that observed from the caries data set, had more success in converging in larger data sets, and where the prevented fraction was larger. In order to investigate if the failure to converge is systematic, Table 8.3 compares the coefficients of intervention group estimated in the 2-level model, between those data sets where the 3-level model converged, and those where it did not, for the simulations with the variance of the random effect at tooth level equal to one. The comparisons which are excluded are those where the number of failures was less than 20.

	Mean (s.d.) coefficient of group from 2-level model	
50 per group	Failed to converge	Converged
PF 15%	0.234 (0.212), n=427	0.267 (0.233), n=1073
PF 25%	0.332 (0.225), n=244	0.387 (0.222), n=1256
PF 35%	0.497 (0.195), n=135	0.530 (0.230), n=1365
150 per group		
PF 15%	0.226 (0.136), n=413	0.247 (0.132), n=1087
PF 25%	0.344 (0.141), n=84	0.372 (0.127), n=1416
300 per group		
PF 15%	0.224 (0.096), n=211	0.240 (0.094), n=1289

Table 8.3 : Group coefficients from 2-level model by convergence status of 3-level models, for simulated data sets with variance of random tooth effect equal to 1.

Table 8.2 and Table 8.3 both suggest that convergence difficulties increase as the group effect decreases. In Table 8.2, data sets simulated using lower preventive fractions have higher proportions of convergence failure. Table 8.3 shows that on average, the group coefficient estimated from the 2-level model appears to be slightly lower in those data sets where the three level model failed to converge. However, as this difference is quite small in most cases, the subsequent investigation of the observed power for the various analysis methods will use only those data sets for which the three-level model converged. As very few of the data sets with tooth level random effect of variance 2 converged, these have not been included in the remainder of this chapter.

8.3.2. Results of analyses of simulated data sets

Table 8.4 shows the mean group coefficient estimates for the simulated data sets using the 3-level and 2-level models. The analysis of data sets simulated with no treatment effect showed that α was within a sampling tolerance of 0.05 for all combinations of parameters and tests.

The mean estimated coefficients from the 3-level models of the simulated data sets with no random effect are mostly similar to the true values. As this model specification is the one used in the simulation, this is to be expected. For the data sets simulated with the random effect included with variance 1, the estimates from the 3-level model are biased upwards. This is likely to be partly due to the exclusion of the datasets where the 3-level model did not converge, as they tended to have smaller group effects.

The 2-level models gave very similar estimates to the 3-level models where there was no random effect in the simulated data sets, as would be expected. However, where a random effect was present, the estimates from the 2-level model are biased downwards.

		Random effect variance 0		Random effect variance 1	
	50 per group	3-level	2-level	3-level	2-level
Prevented fraction and true coefficient value	PF 15% 0.268	0.269 (0.246)	0.266 (0.244)	0.306 (0.266)	0.267 (0.233)
	PF 25% 0.417	0.419 (0.249)	0.416 (0.247)	0.443 (0.253)	0.387 (0.222)
	PF 35% 0.591	0.599 (0.246)	0.595 (0.244)	0.607 (0.258)	0.530 (0.226)
		Random effect variance 0		Random effect variance 1	
	150 per group	3-level	2-level	3-level	2-level
Prevented fraction and true coefficient value	PF 15% 0.268	0.260 (0.147)	0.258 (0.146)	0.283 (0.151)	0.247 (0.132)
	PF 25% 0.417	0.412 (0.144)	0.409 (0.142)	0.427 (0.145)	0.372 (0.127)
	PF 35% 0.591	0.593 (0.148)	0.590 (0.147)	0.606 (0.147)	0.528 (0.128)
		Random effect variance 0		Random effect variance 1	
	300 per group	3-level	2-level	3-level	2-level
Prevented fraction and true coefficient value	PF 15% 0.268	0.271 (0.101)	0.270 (0.100)	0.276 (0.108)	0.240 (0.094)
	PF 25% 0.417	0.415 (0.100)	0.413 (0.100)	0.425 (0.105)	0.371 (0.091)
	PF 35% 0.591	0.589 (0.103)	0.586 (0.102)	0.603 (0.104)	0.526 (0.091)

Table 8.4: Mean (s.d.) of group coefficient estimates for simulated data sets for which the 3-level model converged.

Table 8.5 shows the observed power for each combination of parameters, using the 3-level model, the 2-level model, and the traditional DFS increment analysis.

	Random effect variance 0			Random effect variance 1		
50 per group	3-level with tooth position	2-level with tooth position	DFS Increment	3-level with tooth position	2-level with tooth position	DFS Increment
PF 15%	18%	18%	21%	20%	21%	23%
PF 25%	39%	39%	40%	39%	41%	43%
PF 35%	68%	68%	67%	64%	66%	67%
	Random effect variance 0			Random effect variance 1		
150 per group	3-level with tooth position	2-level with tooth position	DFS Increment	3-level with tooth position	2-level with tooth position	DFS Increment
PF 15%	43%	43%	44%	44%	48%	48%
PF 25%	81%	81%	81%	80%	82%	81%
PF 35%	98%	98%	98%	98%	98%	98%
	Random effect variance 0			Random effect variance 1		
300 per group	3-level with tooth position	2-level with tooth position	DFS Increment	3-level with tooth position	2-level with tooth position	DFS Increment
PF 15%	76%	76%	74%	71%	73%	73%
PF 25%	99%	99%	97%	98%	98%	98%
PF 35%	100%	100%	100%	100%	100%	100%

Table 8.5 : Observed power for 3 analysis methods.

This table shows that ignoring the tooth level random effect generally only changed the observed power by a small amount, although where it did change it was always greater in the 2-level model than in the 3-level model. The observed power for the traditional DFS increment method was also very similar to that of the 3-level multilevel analysis, with more variation in the smaller data sets (50 per group).

8.4. Summary

This simulation study has shown that there is little difference between the observed power of multilevel models and traditional DFS increment analysis. It should be noted that although a range of parameters were simulated, all data sets were based on a population of Scottish adolescents with a particular disease level, and that these results cannot necessarily be assumed to be generalisable beyond this population. This work does give some indication that the advantages of multilevel modelling in dental caries clinical trials may lie in greater understanding of the data structure and within mouth patterns of caries development, rather than reduction of required sample size.

If multilevel modelling is to be used, then the results in this, and other chapters, indicate that the 3-level model should be used in preference to the 2-level model with no random effect at tooth level, due to underestimates of the treatment effect. However, this does not appear to have a large impact on the comparative power of the methods.

9. Discussion

The main aim of this thesis has been to investigate the use of multilevel modelling in the analysis of data from clinical trials of caries preventive agents. There has been discussion in the literature on the hierarchical structure of caries data, and work has been published on methods which allow for analysis at tooth or surface level. However, multilevel modelling has been little used in the caries literature, which has prompted the work presented here.

In this chapter, the main findings will be discussed, and recommendations will be made for investigators on the use of multilevel modelling in the analysis of caries clinical trial data. Some opportunities for further work in this area will be considered.

9.1. Summary of main findings

In Chapter 4, the statistical aspects of recent publications from clinical trials of topical fluorides in children were assessed, specifically considering issues related to clustering. This work found that several cluster randomised caries trials, with participants randomised in groups such as school class, have been published with analysis which incorrectly ignores the clustering. Some of these trials would only require the presence of a small intracluster correlation coefficient to result in an apparently statistically significant finding no longer showing significance with an appropriate analysis. In addition, the hierarchical structure of caries data, with surfaces and teeth clustered

within individual participants, was rarely considered in the analysis of these clinical trials.

In Chapter 5 multilevel modelling was applied to a data set from a caries clinical trial, and showed how modelling the random structure and calculating the intracluster correlation coefficient using the threshold model allows estimation of relative variances at individual, tooth and surface level, with computations which are much more straightforward than alternative methods. Use of tooth position fixed effects in the model can increase understanding of the patterns of caries development within the mouth. Estimation using MCMC methods was found to give more accurate estimates than the default quasi-likelihood methods in MLwiN.

In Chapter 6 it was demonstrated how multilevel modelling of caries clinical trial data could provide clinically useful methods of predicting tooth and surface specific caries incidence, within the age groups and populations examined, based on baseline caries patterns. The method applied to a data set of 12-16 year olds from Scotland showed that caries on the contralateral surface (the corresponding surface on the opposite side of the mouth), was a stronger predictor than caries in the corresponding surface on the opposing jaw, or caries on an adjacent tooth.

The investigation of model fit in Chapter 7 suggested that the three level model specification with the hierarchy individual – tooth – surface gave a better fit to the data than the two level model not including the tooth level. Also, there is evidence that a zero-inflated binomial model gives a better fit to the data than the standard binomial, although this could only be shown for the two-level model, due to software limitations.

The simulation study in Chapter 8 showed that where data are simulated with a random effect at tooth level, ignoring this in the multilevel analysis can result in biased estimates

of the treatment effect, which may explain why the observed power for this analysis was higher in some cases than for the model including random effect at tooth level. The investigation of observed power suggests that using multilevel modelling may not result in significant reductions in required sample size in planning caries clinical trials.

9.2. Limitations

The analyses presented here are on a population of 12 to 16 year olds in Tayside, Scotland. The results relating to tooth position are influenced by the eruption times of specific teeth. The canines, premolars and second permanent molars will have recently erupted, whereas the other permanent teeth will have been present for several years. This means that the tooth position results cannot be generalised beyond the age group examined. In addition to this, the exclusion of surfaces which had already developed caries from the analysis means that models may differ in populations with higher or lower caries levels.

The software which was used for analysis limited the full investigation of the best model fit. Model fit assessment suggested that a three-level zero-inflated binomial model may be a good fit to the data, but MLwiN does not support zero-inflated binomial modelling, and Latent GOLD does not support three-level models.

The simulation study did not use the most accurate estimation method (MCMC) as the time taken for each model to converge was prohibitive. However, the 2nd order PQL method used has been shown to give reasonably accurate estimates, although its failure to converge on some data sets has limited the data available for certain parameter combinations in the simulation study.

9.3. Comparison with other work

The results in Chapter 4 on the appropriateness of analysis, and quality of reporting, of cluster randomised trials of caries preventive agents are similar to those found in a general review of trials in primary care (Eldridge et al., 2004), and those found in a review of articles published in the American Journal of Public Health and Preventive Medicine (Varnell et al., 2004), which both found inappropriate analyses of cluster randomised trials.

The findings in Chapter 5 on the differing probabilities of caries according to different tooth types, with molars more susceptible, are similar to results from cross-sectional survey data e.g. (Batchelor and Sheiham, 2004).

Several previous studies on patterns of caries within the mouth have shown evidence that caries tends to aggregate on one side of the mouth, rather than show symmetry (Hujoel et al., 1994b; Vanobbergen et al., 2007). This is in contrast to the findings in Chapter 6 that the caries status of the contralateral surface was the best predictor of caries on a given surface. These previous studies both used different age groups (adults, and 3-7 year old children, respectively). It is possible that the result may not hold in these age groups, but as these previous studies used cross-sectional rather than longitudinal data, it may be that the lesions aggregated on one side of the mouth may be predictors of future caries on the contralateral surface, as shown in the results here. This thesis presents the first published work on this phenomenon using longitudinal data.

The lack of evidence for a reduction in required sample size in analysing at tooth and surface level was similar to conclusions drawn by Mancl et al., in relation to time-to-event analysis with surface as the unit of analysis, where the authors stated that "the gain in efficiency due to the use of surface-specific information will most likely be small

under most circumstances" (Mancl et al., 2004). The work in this thesis is the first to consider this issue in caries data using a multilevel modelling framework.

The findings here show that MCMC estimation gives more accurate estimates than the default quasi-likelihood methods in MLwiN for the specific data structures present in caries data. This is supported by previous work on other types of data (Browne and Draper, 2006).

9.4. Implications and recommendations to investigators

Investigators planning clinical trials of caries preventive programmes which are to be cluster randomised must ensure that the planned statistical analysis, and associated sample size calculation, is appropriate. Multilevel modelling is one appropriate analysis, which allows for covariates at both cluster and individual level to be included in the analysis. Multilevel modelling of binary data should use MCMC estimation in preference to quasi-likelihood based methods, to ensure accuracy of estimates.

There is no indication from this work that the use of multilevel modelling using the natural hierarchy of surfaces and teeth clustered within individuals will allow investigators to make significant reductions in the number of participants required in a clinical trial of a caries preventive agent, compared to the use of the traditional comparison of caries increments.

However, multilevel modelling does have the advantage of allowing greater understanding of the patterns of caries development within the mouth. Modelling the random structure allow estimates to be made of the relative variance at individual, tooth and surface level. Investigators who are interested in exploring the effect of their intervention in more detail should consider the application of multilevel modelling to their clinical trial data.

The use of multilevel modelling on clinical trial data sets to predict tooth and surface specific caries incidence using patterns of previous caries experience can be recommended.

9.5. Future work

The results in this thesis are based on data from a particular population in a specific age range, 12-16 year olds in Tayside, Scotland. Further investigations will involve the use of this multilevel modelling technique on data sets from different populations, and different age ranges. The results on groupings of tooth positions from Chapter 5 will be population dependent, and it is of interest to investigate how this varies in other populations. Different age ranges will have different teeth which are susceptible to caries, and the techniques used in Chapter 6 to predict tooth and surface specific caries incidence should be applied to other data sets to examine if the symmetry based predictions can be used in different populations.

The work presented here has concentrated on the differential probabilities of developing caries on different tooth types. It may also be of interest to investigate the effect of different surface types on these probabilities, specifically pit and fissure surfaces compared to smooth surfaces.

A possible application of multilevel modelling is to dental caries data collected on children with a mixed dentition, where some teeth are from the deciduous dentition, and some from the permanent. These data are usually analysed with separate outcome variables for each dentition. The application of multilevel modelling to these data sets may allow the modelling of both dentitions together, by including a covariate indicating to which dentition a particular tooth belongs.

The work presented in Chapter 7 on model fit suggested that for the two level model with hierarchy individual – surface, that a zero-inflated binomial distribution may give a better fit to the data than a standard binomial distribution. Also, for the standard binomial models, the three level model with hierarchy individual – tooth – surface fitted the data better than the two level model. These results suggest that the three level model with the zero-inflated binomial distribution may be worth investigating. At the present time, the software is not available to fit this model. One possible course of action is to program this model in the software package R, by modifying the currently available multilevel modelling routines. Another approach is to test the robustness of modelling using the standard binomial distribution in MLwiN to data sets simulated from zero-inflated multilevel binomial distributions. The introduction of a syntax module to the software Latent GOLD will facilitate this work by allowing models to be fitted to multiple data sets automatically, a feature which was unavailable until recently.

Another method of analysing caries clinical trial data at tooth and surface level which has been suggested is the use of survival analysis at surface level, with jackknife estimators for the variance, which adjust for the clustering in the data (Hannigan et al., 2001). A possible future study would involve comparing this method with the multilevel modelling method used in this thesis, both on real and simulated data sets, to investigate robustness to model mis-specification, and observed power to detect treatment differences.

10. List of References

Adair P, Ashcroft A (2007). Theory-based Approaches to the Planning and Evaluation of Oral Health Education Programmes. In: Community Oral Health. CM Pine and RV Harris editors. London: Quintessence.

Ahn C, Jung SH, Donner A (2002). Application of an adjusted chi2 statistic to site-specific data in observational dental studies. *Journal of Clinical Periodontology* 29(1):79-82.

Ahovuo-Saloranta A, Hiiri A, Nordblad A, Mäkelä M, Worthington HV (2008). Pit and fissure sealants for preventing dental decay in the permanent teeth of children and adolescents. *Cochrane Database of Systematic Reviews* Issue 4 Art. No. CD001830

Akaike H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6):716-723.

Albandar JM, Goldstein H (1992). Multi-level statistical models in studies of periodontal diseases. *Journal of Periodontology* 63(8):690-5.

Armitage GC (2004). Periodontal diagnoses and classification of periodontal diseases. *Periodontol 2000* 34:9-21.

Axtelius B, Soderfeldt B, Attstrom R (1999). A multilevel analysis of factors affecting pocket probing depth in patients responding differently to periodontal treatment. *Journal of Clinical Periodontology* 26(2):67-76.

Batchelor P, Sheiham A (2004). Grouping of tooth surfaces by susceptibility to caries: a study in 5-16 year-old children. *BMC Oral Health* 4(1):2.

Berman DS, Slack GL (1972). Dental caries in English school children: a longitudinal study. *British Dental Journal* 133(12):529.

Böhning D, Dietz E, Schlattmann P, Mendonça L, Kirchner U (1999). The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A* 162(Part 2):195-209.

Borutta A, Kunzel W, Rubsam F (1991). [The caries-protective efficacy of 2 fluoride varnishes in a 2-year controlled clinical trial]. *Dtsch Zahn Mund Kieferheilkd Zentralbl* 79(7):543-9.

Boyd JD, Cheyne VD (1946). Epidemiologic studies in dental caries: I. The pattern of dental caries in Iowa City children. *Journal of Pediatrics* 29:157-176.

Bratthall D, Hansel Petersson G, Sundberg H (1996). Reasons for the caries decline: What do the experts believe? *European Journal of Oral Science* 104:416-422.

Bravo M, Llodra JC, Baca P, Osorio E (1996). Effectiveness of visible light fissure sealant (Delton) versus fluoride varnish (Duraphat): 24-month clinical trial. *Community Dentistry and Oral Epidemiology* 24(1):42-6.

Brodeur JM, Simard PL, Demers M, Contandriopoulos AP, Tessier G, Lepage Y, et al. (1988). Comparative effects of FMR programs in fluoridated and unfluoridated communities. *Journal of the Canadian Dental Association* 54(10):761-5.

Brodeur JM, Simard PL, Demers M, Contandriopoulos AP, Tessier G, Lepage Y, et al. (1990). Recruitment and compliance in school-based FMR programs. *Journal of the Canadian Dental Association* 56(1):53-6.

Browne W (2005). MCMC Estimation in MLwiN version 2.0 Bristol: Centre for Multilevel Modelling.

Browne WJ, Draper D (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis* 1(3):473-514.

Burchell CK, Stephen KW, Schafer F, Huntington E (1991). Improved sensitivity in caries clinical trials by selection on the basis of baseline caries experience. *Community Dentistry and Oral Epidemiology* 19(1):20-2.

Campbell MK, Elbourne DR, Altman DG (2004). CONSORT statement: extension to cluster randomised trials. *British Medical Journal* 328(7441):702-8.

Carlos JP, Gittelsohn AM (1965). Longitudinal studies of the natural history of caries. II. A life-table study of caries incidence in the permanent teeth. *Archives of Oral Biology* 10(5):739-51.

Chesters RK, Ellwood RP, Biesbrock AR, Smith SR (2004). Potential modern alternative designs for Caries Clinical Trials (CCTs) and how these can be validated against the conventional model. *Journal of Dental Research* 83(Spec Iss C):C122-C124.

Chikte UM, Lewis HA, Rudolph MJ (1996). The effectiveness of a school-based fluoride mouth rinse programme. *Journal of the Dental Association of South Africa* 51(11):697-700.

Churchill HV (1931). The occurence of fluoride in some waters of the United States. *Journal of Dental Research* 12:141-148.

Donner A, Banting D (1989). Adjustment of frequently used chi-square procedures for the effect of site-to-site dependencies in the analysis of dental data. *Journal of Dental Research* 68(9):1350-4.

Donner A, Klar N (2000). Design and Analysis of Cluster Randomization Trials in Health Research London: Arnold.

Downer MC, Holloway PJ, Davies TG (1977). Efficiency of clinical trials: selective inclusion of study subjects. *Journal of Dental Research* 56 Spec No:C123-9.

Dye BA, Tan S, Smith V, Lewis BG, Barker LK, Thornton-Evans G, et al. (2007). Trends in oral health status: United States, 1988-1994 and 1999-2004. *Vital and Health Statistics Series 11* (248):1-92.

Eckermann R (1919). Dental caries in relation to oral osmosis Lund: Carl Bloms Boktryckeri.

Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC (2004). Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clinical Trials* 1:80-90.

Ellwood RP, O'Mullane DM (1996). Identification of areas with high levels of untreated dental caries. *Community Dentistry and Oral Epidemiology* 24(1):1-6.

Emilson CG (1994). Potential efficacy of chlorhexidine against mutans streptococci and human dental caries
Journal of Dental Research 73(3):682-691.

Fabien V, Obry-Musset AM, Hedelin G, Cahen PM (1999). Use of the generalised linear model with Poisson distribution to compare caries indices. *Community Dental Health* 16(2):93-6.

Featherstone JD (2004). The continuum of dental caries--evidence for a dynamic disease process. *Journal of Dental Research* 83 Spec No C:C39-42.

Fisher RA (1935). The Design of Experiments Edinburgh: Oliver and Boyd.

Fleiss JL, Slakter MJ, Fischman SL, Park MH, Chilton NW (1979). Inter-examiner reliability in caries trials. *Journal of Dental Research* 58(2):604-9.

Forgie AH, Paterson M, Pine CM, Pitts NB, Nugent ZJ (2000). A randomised controlled trial of the caries-preventive efficacy of a chlorhexidine-containing varnish in high-caries-risk adolescents. *Caries Research* 34(5):432-9.

Frostell G, Birkhed D, Edwardsson S, Goldberg P, Petersson LG, Priwe C, et al. (1991). Effect of partial substitution of invert sugar for sucrose in combination with Duraphat treatment on caries development in preschool children: the Malmo Study. *Caries Research* 25(4):304-10.

Geman S, Geman D (1984). Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEE Transactions on Pattern Analysis and Machine Intelligence* 6:721-741.

Gilthorpe MS, Maddick IH, Petrie A (2000). Introduction to multilevel modelling in dental research. *Community Dental Health* 17(4):222-6.

Gilthorpe MS, Griffiths GS, Maddick IH, Zamzuri AT (2001). An application of multilevel modelling to longitudinal periodontal research data. *Community Dental Health* 18(2):79-86.

Gisselsson H, Birkhed D, Emilson CG (1999). Effect of professional flossing with NaF or SnF2 gel on approximal caries in 13-16-year-old schoolchildren. *Acta Odontologica Scandinavica* 57(2):121-5.

Glass RL, Alman JE, Fleisch S, D'Agostino RB (1972). The appropriateness of analysis of variance to the statistical analyses of dental clinical trials. *Archives of Oral Biology* 17(4):633-43.

Goldstein H (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73(1):43-56.

Goldstein H (1989). Restricted unbiased iterative generalized least-squares estimation. *Biometrika* 76(3):622-623.

Goldstein H, Rasbash J (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society: Series A* 159:505-513.

Goldstein H (2003). Multilevel statistical models. 3rd ed. London: Arnold.

Gunsolley JC, Williams DA, Schenkein HA (1994). Variance component modeling of attachment level measurements. *Journal of Clinical Periodontology* 21(4):289-95.

Hall DB (2000). Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56(4):1030-1039.

Hannigan A, O'Mullane DM, Barry D, Schafer F, Roberts AJ (2000). A caries susceptibility classification of tooth surfaces by survival time. *Caries Research* 34(2):103-8.

Hannigan A, O'Mullane DM, Barry D, Schafer F, Roberts AJ (2001). A re-analysis of a caries clinical trial by survival analysis. *Journal of Dental Research* 80(2):427-31.

Hannigan A (2004). Using survival methodologies in demonstrating caries efficacy. *Journal of Dental Research* 83 Spec No C:C99-102.

Hastings WK (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* 57(1):97-109.

Hausen H (2004). How to improve the effectiveness of caries-preventive programs based on fluoride. *Caries Research* 38:263-267.

Heidmann J, Poulsen S, Arnbjerg D, Kirkegaard E, Laurberg L (1992). Caries development after termination of a fluoride rinsing program. *Community Dentistry and Oral Epidemiology* 20(3):118-21.

Hujoel PP, Isokangas PJ, Tiekso J, Davis S, Lamont RJ, DeRouen TA, et al. (1994a). A re-analysis of caries rates in a preventive trial using Poisson regression models. *Journal of Dental Research* 73(2):573-9.

Hujoel PP, Lamont RJ, DeRouen TA, Davis S, Leroux BG (1994b). Within-subject coronal caries distribution patterns: an evaluation of randomness with respect to the midline. *Journal of Dental Research* 73(9):1575.

Ismail AI (2004). Visual and Visuo-tactile Detection of Dental Caries. *Journal of Dental Research* 83(Spec Iss C):C56-C66.

Jain P, Shankar A, Ramaiah S (2007). Dental caries and social deprivation. *Lancet* 369(9562):639.

Karjalainen S, Eriksson AL, Ruokola M, Toivonen A (1994). Caries development after substitution of supervised fluoride rinses and toothbrushings by unsupervised use of fluoride toothpaste. *Community Dentistry and Oral Epidemiology* 22(6):421-4.

Kidd EAM (2005). Essentials of Dental Caries. Third ed. Oxford: Oxford University Press.

Kingman A, Selwitz RH (1997). Proposed methods for improving the efficiency of the DMFS index in assessing initiation and progression of dental caries. *Community Dentistry and Oral Epidemiology* 25(1):60-8.

Kleber CJ, Putt MS, Smith CE, Gish CW (1996). Effect of supervised use of an alum mouthrinse on dental caries incidence in caries-susceptible children: a pilot study. *ASDC Journal of Dentistry for Children* 63(6):393-402.

Klein H, Palmer CE (1937). Dental caries in American Indian children. Public Health Bulletin no. 239: US Government Printing Office.

Lewsey JD, Gilthorpe MS, Bulman JS, Bedi R (2000). Is modelling dental caries a 'normal' thing to do? *Community Dental Health* 17(4):212-7.

Lussi A, Imwinkelried S, Pitts N, Longbottom C, Reich E (1999). Performance and reproducibility of a laser fluorescence system for detection of occlusal caries in vitro. *Caries Research* 33(4):261-6.

Maas CJM, Hox JH (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica* 58(2):127-137.

Macfarlane TV, Worthington HV (1999). Some aspects of data analysis in dentistry. *Community Dental Health* 16(4):216-9.

Mancl LA, Hujoel PP, DeRouen TA (2004). Efficiency issues among statistical methods for demonstrating efficacy of caries prevention. *Journal of Dental Research* 83 Spec No C:C95-8.

Marinho VC, Higgins JP, Logan S, Sheiham A (2003a). Topical fluoride (toothpastes, mouthrinses, gels or varnishes) for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* Issue 4 Art. No. CD002782

Marinho VC, Higgins JP, Sheiham A, Logan S (2003b). Fluoride toothpastes for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* Issue 1 Art. No. CD002278

Marinho VCC, Higgins JPT, Logan S, Sheiham A (2002a). Fluoride gels for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* Issue 2 Art. No. CD002280

Marinho VCC, Higgins JPT, Logan S, Sheiham A (2002b). Fluoride varnishes for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* Issue 3 Art. No. CD002279

Marinho VCC, Higgins JPT, Logan S, Sheiham A (2003c). Fluoride mouthrinses for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* Issue 3 Art. No. CD002284

Marinho VCC, Higgins JPT, Sheiham A, Logan S (2004a). One topical fluoride (toothpastes, or mouthrinses, or gels, or varnishes) versus another for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* Issue 1 Art. No. CD002780

Marinho VCC, Higgins JPT, Sheiham A, Logan S (2004b). Combinations of topical fluoride (toothpastes, mouthrinses, gels, varnishes) versus single topical fluoride for preventing dental caries in children and adolescents. *Cochrane Database of Systematic Reviews* Issue 1 Art. No. CD002781

Mitropoulos CM (1985). The use of fibre-optic transillumination in the diagnosis of posterior approximal caries in clinical trials. *Caries Research* 19(4):379-84.

Moher D, Schulz KF, Altman DG (2003). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Clinical Oral Investigations* 7(1):2-7.

Needleman IG, Worthington HV, Giedrys-Leeper E, Tucker RJ (2006). Guided tissue regeneration for periodontal infra-bony defects. *Cochrane Database of Systematic Reviews* (2):CD001724.

Nieri M, Muzzi L, Cattabriga M, Rotundo R, Cairo F, Pini Prato GP (2002). The prognostic value of several periodontal factors measured as radiographic bone level variation: a 10-year retrospective multilevel analysis of treated and maintained periodontal patients. *Journal of Periodontology* 73(12):1485-93.

O'Mullane DM (1976). Efficiency in clinical trials of caries preventive agents and methods. *Community Dentistry and Oral Epidemiology* 4:190-194.

Olivier M, Brodeur JM, Simard PL (1992). Efficacy of APF treatments without prior toothcleaning targeted to high-risk children. *Community Dentistry and Oral Epidemiology* 20(1):38-42.

Page RC, Eke PI (2007). Case Definitions for Use in Population-Based Surveillance of Periodontitis. *Journal of Periodontology* 78(7s):1387-1399.

Petersen PE (2003). Continuous improvement of oral health in the 21st century - the approach of the WHO Global Oral Health Programme. The World Oral Health Report 2003. Geneva: World Health Organization.

Petersson LG, Svanholm I, Andersson H, Magnusson K (1998). Approximal caries development following intensive fluoride mouthrinsing in teenagers. A 3-year radiographic study. *European Journal of Oral Sciences* 106(6):1048-51.

Pine CM, Pitts NB, Nugent ZJ (1997). British Association for the Study of Community Dentistry (BASCD) guidance on the statistical aspects of training and calibration of examiners for surveys of child dental health. A BASCD coordinated dental epidemiology programme quality standard. *Community Dental Health* 14 Suppl 1:18-29.

Pine CM, Harris RV, Burnside G, Merrett MCW (2006). An investigation of the relationship between untreated decayed teeth and dental sepsis in 5-year-old children. *British Dental Journal* 200(1):45.

Pitts NB, Evans DJ, Pine CM (1997). British Association for the Study of Community Dentistry (BASCD) diagnostic criteria for caries prevalence surveys - 1996/97. *Community Dental Health* 14(Suppl 1):6-9.

Pitts NB, Evans DJ, Nugent ZJ (2000). The dental caries experience of 14-year-old children in the United Kingdom. Surveys coordinated by the British Association for the Study of Community Dentistry in 1998/99. *Community Dental Health* 17(1):48-53.

Pitts NB (2004). Modern concepts of caries measurement. *Journal of Dental Research* 83(Spec Issue C):C43-C47.

Powell LV (1998). Caries prediction: a review of the literature. *Community Dentistry and Oral Epidemiology* 26(6):361.

Provart SJ, Carmichael CL (1995). The relationship between caries, fluoridation and material deprivation in five-year-old children in Country Durham. *Community Dental Health* 12(4):200-3.

Puffer S, Torgerson D, Watson J (2003). Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *British Medical Journal* 327(7418):785-9.

Raftery AE, Lewis SM (1995). The number of iterations, convergence diagnostics and generic Metropolis algorithms. In: Practical Markov Chain Monte Carlo. WR Gilks, DJ Spiegelhalter and S Richardson editors. London: Chapman & Hall, pp. 1-15.

Ran F, Gedalia I, Fried M, Hadani P, Tved A (1991). Effectiveness of fortnightly tooth brushing with amine fluorides in caries-prone subjects. *Journal of Oral Rehabilitation* 18(4):311-6.

Rasbash J, Steele F, Browne W, Prosser B (2005). A User's Guide to MLwiN version 2.0 Bristol: Centre for Multilevel Modelling.

Renson CE (1986). Changing patterns of dental caries: a survey on 20 countries. *Annals, Academy of Medicine Singapore* 15:284-298.

Rodriguez G, Goldman N (1995). An Assessment of Estimation Procedures for Multilevel Models with Binary Responses. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 158(1):73-89.

Schwartz D, Lellouch J (1967). Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases* 20(637-648).

Scott D (1944). A study of the bilateral incidence of carious lesions. *Journal of Dental Research* 23:105-110.

Seppa L, Leppanen T, Hausen H (1995). Fluoride varnish versus acidulated phosphate fluoride gel: a 3-year clinical trial. *Caries Research* 29(5):327-30.

Snijders T, Bosker R (1999). Multilevel analysis London: Sage.

Spets-Happonen S, Luoma H, Forss H, Kentala J, Alaluusua S, Luoma AR, et al. (1991). Effects of a chlorhexidine-fluoride-strontium rinsing program on caries, gingivitis and some salivary bacteria among Finnish schoolchildren. *Scandinavian Journal of Dental Research* 99(2):130-8.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4):583-639.

Sterne JA, Johnson NW, Wilton JM, Joyston-Bechal S, Smales FC (1988). Variance components analysis of data from periodontal research. *Journal of Periodontal Research* 23(2):148-53.

Stookey GK (2004). Optical methods--quantitative light fluorescence. *Journal of Dental Research* 83 Spec No C:C84-8.

Tewari A, Chawla HS, Utreja A (1991). Comparative evaluation of the role of NaF, APF & Duraphat topical fluoride applications in the prevention of dental caries--a 2 1/2 years study. *Journal of the Indian Society of Pedodontics and Preventive Dentistry* 8(1):28-35.

Tu YK, Gilthorpe MS, Griffiths GS, Maddick IH, Eaton KA, Johnson NW (2004a). The application of multilevel modeling in the analysis of longitudinal periodontal data--part II: changes in disease levels over time. *Journal of Periodontology* 75(1):137-45.

Tu YK, Gilthorpe MS, Griffiths GS, Maddick IH, Eaton KA, Johnson NW (2004b). The application of multilevel modeling in the analysis of longitudinal periodontal data--part I: absolute levels of disease. *Journal of Periodontology* 75(1):127-36.

Vanobbergen J, Lesaffre E, Garcia-Zattera MJ, Jara A, Martens L, Declerck D (2007). Caries patterns in primary dentition in 3-, 5- and 7-year-old children: spatial correlation and preventive consequences. *Caries Research* 41(1):16-25.

Varnell SP, Murray DM, Janega JB, Blitstein JL (2004). Design and analysis of group-randomized trials: a review of recent practices. *American Journal of Public Health* 94(3):393-399.

Vermunt JK, Magidson J (2005). Technical Guide for Latent GOLD 4.0: Basic and Advanced. Belmont, Massachussetts: Statistical Innovations Inc.

Wenzel A (2004). Bitewing and digital bitewing radiography for detection of caries lesions. *Journal of Dental Research* 83 Spec No C:C72-5.

Whelton H (2004). Overview of the impact of changing global patterns of dental caries experience on caries clinical trials. *Journal of Dental Research* 83 Spec No C:C29-34.

WHO (1997). Oral health surveys. Basic methods Geneva: World Health Organization.

Wood PF (1985). Asymmetry of caries attack on the occlusal surfaces of first permanent molar teeth. *Australian Dental Journal* 30(2):123.

Worthington H (1984). Statistical methodology for clinical trials of caries prophylactic agents--current knowledge. *International Dental Journal* 34(4):278-84.