Generalised Linear Modelling of Childhood Caries

ongoing work presented as part of study for PhD by

John Stephenson

PhD student
School of Dentistry
University of Cardiff
Cardiff, GB-CF14 4XY

April 2008
Objectives

- To identify risk factors and clustering effects associated with the occurrence of childhood caries in primary teeth

- To assess the future risk to caries in primary teeth from a model of primary tooth lifetimes
Cohort study design

- Primary data source is cohort study of ~2650 children undertaken by Cardiff University School of Dentistry in 1999

- Children examined on 3 occasions:
  - School Year 1 – 2 (age 5 – 7)
  - School Year 3 (age 7 – 8)
  - School Year 5 (age 9 – 10)

- Arbitrarily interval censored data

- About 9% lost to follow-up

- Both primary and permanent teeth observed

- All surfaces of all teeth examined
  - Over 400,000 recorded observations in total
Study areas

**West Midlands** (fluoridated @ 1ppm)
- Industrial region: population ~2,500,000
- Dudley area (300,000)
- Walsall area (300,000)

**South Wales** (non-fluoridated)
- Industrial region: population ~1,500,000
- Mid Glamorgan area (500,000)
- South Glamorgan area (400,000)

~700 children sampled from each area
Cohort study: key recorded attributes

- Gender
- Date of birth
- Standardised measure of area deprivation
  - Estimated from recorded postcode
- Fluoridation status
  - Binary variable: correlated with geographical area
- Dentition type
- Tooth location/type (implicitly from tooth location)
  - Incisor, Canine, Pre-molar (permanent teeth only), Molar
- Surface type
  - Distal, Occlusal (Molars and pre-molars only), Mesial, Buccal, Lingual
Cohort study: response measures

- Responses recorded at surface level at each examination
- Outcomes dichotomised for modelling
- *Tooth*- and *child*- level responses created from recorded *surface*-level responses
- At least one positive *surface*-level outcome per tooth ➔ assumed positive response at *tooth* level
- At least one positive *tooth*-level outcome per child ➔ assumed positive response at *child* level
Analysis

- **Exploratory analysis**
  - Investigates data trends

- **Phase 1: multilevel logistic regression analysis**
  - Logistic generalised linear model
  - Identification of risk factors and hierarchical data structures

- **Phase 2: multilevel survival analysis**
  - Complementary log-log generalised linear model
  - Models survival curves of teeth
### Exploratory analysis: extent of primary dentition

<table>
<thead>
<tr>
<th>Primary tooth type</th>
<th>Proportion of teeth in primary state (non-appearance of corresponding permanent tooth)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1\textsuperscript{st} examination</td>
</tr>
<tr>
<td>Incisors</td>
<td>68.3%</td>
</tr>
<tr>
<td>Canines</td>
<td>~100%</td>
</tr>
<tr>
<td>Molars</td>
<td>~100%</td>
</tr>
</tbody>
</table>
Exploratory analysis: caries in primary dentition - some baseline results

• Significantly higher rates of occurrence in South Wales
  – West Midlands (1 ppm fluoridation): 29.3% ± 2.5%
  – South Wales (non fluoridated): 51.5% ± 2.6%

• Small bias towards occurrence in males
  – Males 42.9% ± 2.7%; Females 38.8% ± 2.7%

• Significantly higher rates of occurrence in molar teeth
  – Incisors 2.6%; Canines 1.4%; Molars 16.3%

• Significantly higher rates of occurrence on occlusal surfaces
  – Distal/Mesial 4.5%  Occlusal 12.9%
  – Buccal/Lingual 3.3%
Hierarchical Data Structure

• Existence of hierarchical structure ➔ clustering of data
  – lack of independence of units (e.g. teeth within children)

• Many possible hierarchical structures, for example:
  – surfaces nested within teeth nested within child nested within school

• Other interpretations of hierarchy are possible
  – Quadrant, tooth type, surface type, area etc.
  – May be more appropriately modelled as fixed effects

• Disregarding hierarchical structure may lead to:
  – spurious indications of parameter significance
  – Loss of information concerning variability at higher model levels
Possible 2 level structure

Tooth level 1; child level 2

Responses at tooth level

Child 1

Tooth 1

Tooth 2

Tooth 3

Child 2

Tooth 1

Tooth 2

Tooth 3

Child 3

Tooth 1

Tooth 2

Tooth 3
Possible 3-level structure

Surface level 1; tooth level 2; child level 3
Responses at surface level
Phase 1 Analysis

• A series of multilevel logistic regression analyses using surface, tooth and child level measures

• Analysis aims to determine:
  – appropriate model hierarchies
  – factors significantly associated with occurrence of caries
  – appropriate multilevel model type

• Current presentation considers response of occurrence of caries in primary dentition at 1st examination
Phase 1 analysis: key modelling approximations

• Assume no variation in ages of experimental units observed at any given examination

• Assume measured dentition has reached "steady state"

• Does not utilise most updated data
Phase 1 Analysis: assessment of possible model hierarchies

• Contribution of each level to model variance may be assessed by Variance Partition Coefficient (VPC)
  - VPCs calculated for variance components models
  - Simulation method developed for binary data

• Low VPC for a particular level suggests model structure should be reformulated excluding this level
## Phase 1 Analysis: Partitioning of model variance – *child* level variance components models

<table>
<thead>
<tr>
<th>Response</th>
<th>Model</th>
<th>Proportion of model variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level 1 (child)</td>
</tr>
<tr>
<td>Caries at 1\textsuperscript{st} exam</td>
<td>C1</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>92.5%</td>
</tr>
</tbody>
</table>
Phase 1 Analysis: Partitioning of model variance - *tooth* level variance components models

<table>
<thead>
<tr>
<th>Response</th>
<th>Model</th>
<th>Proportion of model variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level 1 (tooth)</td>
</tr>
<tr>
<td>Caries at 1<em>st</em> exam</td>
<td>T1</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>T2</td>
<td>75.3%</td>
</tr>
<tr>
<td></td>
<td>T3</td>
<td>78.4%</td>
</tr>
</tbody>
</table>
## Phase 1 Analysis: Partitioning of model variance - surface level variance components models

<table>
<thead>
<tr>
<th>Response</th>
<th>Model</th>
<th>Proportion of model variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level 1 (surface)</td>
</tr>
<tr>
<td>Caries at 1st exam</td>
<td>S1</td>
<td>100.0%</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>41.4%</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>42.7%</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>47.3%</td>
</tr>
</tbody>
</table>
Phase 1 analysis: covariate assessment

- Significance of risk factors assessed in logistic regression analysis by calculation of odds ratios and $p$-values

- Covariates initially tested using univariate analyses with increasing levels
  - tested for evidence of collinearity

- Significant covariates carried forward to multivariate analyses
Multivariate logistic regression – *tooth* level models

3-level hierarchy: tooth-child-school

<table>
<thead>
<tr>
<th>Covariate</th>
<th>p-value</th>
<th>Odds ratio</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.000</td>
<td>0.76 (0.65, 0.88)</td>
<td>Significant</td>
</tr>
<tr>
<td>Age at 1st examination</td>
<td>0.001</td>
<td>1.30 (1.06, 1.60)</td>
<td>Significant</td>
</tr>
<tr>
<td>SEC</td>
<td>0.000</td>
<td>1.10 (1.08, 1.13)</td>
<td>Significant</td>
</tr>
<tr>
<td>F - status</td>
<td>0.000</td>
<td>0.25 (0.21, 0.30)</td>
<td>Significant</td>
</tr>
<tr>
<td>Molar tooth</td>
<td>0.000</td>
<td>12.4 (11.3, 13.6)</td>
<td>Significant</td>
</tr>
</tbody>
</table>
# Multivariate logistic regression – surface level models

4-level hierarchy: surface-tooth-child-school

<table>
<thead>
<tr>
<th>Covariate</th>
<th>p-value</th>
<th>Odds ratio</th>
<th>Inference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.000</td>
<td>0.63 (0.52, 0.77)</td>
<td>Significant</td>
</tr>
<tr>
<td>Age at 1\textsuperscript{st} exam</td>
<td>0.065</td>
<td>1.30 (0.98, 1.71)</td>
<td>Not significant</td>
</tr>
<tr>
<td>SEC</td>
<td>0.000</td>
<td>1.13 (1.09, 1.17)</td>
<td>Significant</td>
</tr>
<tr>
<td>F - status</td>
<td>0.000</td>
<td>0.23 (0.18, 0.29)</td>
<td>Significant</td>
</tr>
<tr>
<td>Molar tooth</td>
<td>0.000</td>
<td>9.08 (8.33, 9.90)</td>
<td>Significant</td>
</tr>
<tr>
<td>Occlusal surface</td>
<td>0.000</td>
<td>2.00 (1.91, 2.10)</td>
<td>Significant</td>
</tr>
</tbody>
</table>
Phase 1 analysis: conclusions

- Most appropriate model hierarchies:
  - Surface – tooth – child
  - Surface – tooth – child – school
  - Tooth – child - school

- Tooth level largest contributor to model variance in most cases

- Significant risk factors associated with caries in primary dentition at baseline
  - all covariates generally significant

- Model type:
  - Random intercepts model adequate in most cases

- Inferences to be carried forward to Phase 2
Phase 2 analysis

- Survival analysis of primary dentition

- Makes use of inferences from Phase 1 analysis regarding model hierarchies and parameter significance
  - Requires assumption of survival function and proportional hazards model

- Modelling approximations
  - Caries and exfoliation failure modes assumed to be independent
  - Remineralisations disregarded
  - Interval censored data equivalent to left censored
Phase 2 analysis: method

- Survival data transformed using complementary log-log GLM
  - Leads to 2-parameter Weibull survival curve
    \[ S(t) = \exp(-\lambda t^\gamma) \]

- Baseline function extended to incorporate covariates
  \[ S(t) = \exp(-\lambda t^\gamma) \exp(B_1 x_1 + B_2 x_2 + \ldots) \]
Phase 2 analysis: resistance to caries and exfoliation - molar teeth
Phase 2 analysis: resistance to caries and exfoliation - non-molar teeth
Phase 2 analysis: effect of grouping data on resistance to caries - molar teeth
Phase 2 analysis: comparison of resistance to caries for molar teeth across regions
Phase 2 analysis: comparison of survival against caries for molar teeth between genders
Phase 2 analysis: conclusions

• Caries not life-limiting feature at any stage of lifetime of non-molar teeth

• Caries may limit lifetime of primary molar teeth surviving beyond ~10 years

• Grouping data by age of experimental unit does not significantly affect survival curve for primary molars

• Resistance to caries for primary molars distinct for children in fluoridated and non-fluoridated regions

• Resistance to caries for primary molars not significantly distinct between genders
Future work

• Future Phase 3 analysis: unit lifetime modelling
  – Models ultimate fate of sound, decayed and treated primary teeth

• Will incorporate additional treatment data from British Dental Practice Board (applies in ~51% of individuals)

• Will assume competing risks / multistate model with alternative “routes” to exfoliation possible
  – sound-exfoliation
  – sound-caries-exfoliation etc.

• Will assess effect of treatment on primary tooth survival and on subsequent state of permanent dentition
Future Work: Tooth lifetime model

- Competing risks multistate model
- Allows for extraction of carious teeth as “absorbing state”
Future Work: Tooth lifetime model (2)

- Competing risks multistate model
- Allows repeated transitions between carious and filled states