# Multiple Imputation of Dental Caries Data Using a Zero Inflated Poisson Regression Model

**Bhavna T. Pahel, BDS, MPH, PhD**,
Cecil G. Sheps Center for Health Services Research, The University of North Carolina at Chapel Hill

**John S. Preisser, PhD**,
Department of Biostatistics, UNC Gillings School of Public Health, The University of North Carolina at Chapel Hill

**Sally C. Stearns, PhD**, and
Department of Health Policy and Management, UNC Gillings School of Public Health, The University of North Carolina at Chapel Hill

**R. Gary Rozier, DDS, MPH**
Department of Health Policy and Management, UNC Gillings School of Public Health, The University of North Carolina at Chapel Hill

## Abstract

Excess zeros exhibited by dental caries data require special attention when multiple imputation is applied to such data.

**Objective**—To demonstrate a simple technique using a zero-inflated Poisson (ZIP) regression model, to perform multiple imputation for missing caries data.

**Methods**—The technique is demonstrated using data (N=24,403) from a medical office-based preventive dental program in North Carolina, where 27.2% of children (N=6,637) were missing information on physician-identified count of carious teeth. We first estimate a ZIP regression model using the non-missing caries data (N=17,766). The coefficients from the ZIP model are then used to predict the missing caries data.

**Results**—This technique results in imputed caries counts that are similar to the non-missing caries data in their distribution, especially with respect to the excess zeros in the non-missing caries data.

**Conclusion**—This technique can be easily applied to impute missing dental caries data.

## Keywords

Dental caries; children; imputation; Zero Inflated Poisson regression; mixture model

## INTRODUCTION

Multiple imputation is increasingly the standard approach for addressing missing data in research studies (1). However, multiple imputation of missing dental caries data can require special attention. In most epidemiological studies, particularly with children, counts of carious

teeth and surfaces exhibit excess zeros, indicating a low prevalence of disease. Therefore, the imputation model should account for this distribution. To our knowledge, current studies of dental caries do not address implications of the distribution of dental caries data in imputing missing caries data. This paper describes a straightforward technique to impute missing caries data using a Zero-Inflated Poisson (ZIP) regression model. Although the ZIP model has previously been used to model epidemiological dental caries data (2,3), it has not, to our knowledge, been applied to imputing missing caries data.

We demonstrate this imputation technique using data on counts of teeth with caries, collected as part of the evaluation of a medical office-based preventive dental program in North Carolina (NC). The program trains pediatric primary care providers to conduct oral health assessments and provide other preventive services for children younger than 3 years of age (4). The physician-identified count of carious teeth in data (N=24,403) from this statewide program was missing for 27.2% (N=6637) of child observations. Children with missing caries information were more likely to live in dental underserved areas, less likely to be referred to a dentist by their physician, and were younger than those with non-missing caries data, indicating that the data are not missing completely at random (MCAR) and therefore should be imputed (1). Further, of those with non-missing caries information, about 94% had a value of 0, making a zero-inflated regression model appropriate for these data.

## METHODS

### The Zero Inflated Poisson (ZIP) regression model

Count data, including dental caries data, commonly exhibit zero inflation and overdispersion relative to the Poisson distribution. Zero-inflation refers to the presence of excess zeros, as observed with dental caries data. Overdispersion occurs when the variance exceeds its mean, which can have as its source excess zeros. The ZIP model allows for the modeling of zero-inflated count data (5) and provides superior fit for dental caries data compared to a Poisson model (2,3).

The ZIP model assumes that the observed counts are generated by a mixture of two possible processes. The first process determines the probability of an excess zero. If an excess zero is not generated in the first process, then the count is estimated using the second part of the model which models the dependent variable as a count using a Poisson distribution. These two processes are described in detail below.

**Process 1—**Process 1 is a Bernoulli process where, $z_{ij}$, a binary variable (0, 1), determines whether an excess zero is generated for the count variable $y_{ij}$ (e.g., number of teeth with dental caries for the $j_{th}$ child in the $i^{th}$ county).

Let $z_{ij} = 1$ indicate an excess zero, and $z_{ij} = 0$, otherwise. The model is:

$$\text{logit}(\varphi_{ij}) = 1/\{1 + \exp[-(x_{ij}\beta)]\} \qquad \text{Equation 1}$$

where $\varphi_{ij} = E(z_{ij}|x_{ij}) = P(z_{ij}=1|x_{ij})$ is the probability of an excess zero and $\beta$ are the coefficients of the covariates ($x_{ij}$) in the model.

**Process 2—**Given $z_{ij} = 0$, $y_{ij}$ is generated using a Poisson process:

$$\log(\theta_{ij}) = \exp(w_{ij}\alpha) \qquad \text{Equation 2}$$

where $\theta_{ij} = E(y_{ij}^*)$ and $y_{ij}^*$ has a Poisson distribution, denoted by $f(y_{ij}^*)$; and $\alpha$ are the coefficients for each of the covariates ($w_{ij}$) in the model.

The mixture of these two processes gives the ZIP model for the observed counts, $y_{ij}$. The probability function for the ZIP model is:

$$P[y_{ij}=0]=\varphi_{ij}+(1-\varphi_{ij})f(0)$$

Where $f(0)$ is the Process 2 Poisson distribution function evaluated at $y_{ij}^*=0$; and

$$P[y_{ij}=k]=(1-\varphi_{ij})f(k), k=1, 2, \ldots$$

## Data sources and file linkages

In NC, Medicaid reimburses medical practitioners to provide preventive dental care during well child visits for preschool age children through the Into the Mouths of Babes (IMB) program. During the demonstration phase (2000–2002), physicians voluntarily completed patient encounter forms (EFs) to provide dental information (including counts of teeth with caries) not available in Medicaid claims submitted for reimbursement. A total of 24,403 EFs were available for a study that merged EFs with Medicaid claims for IMB services to examine physicians' dental referral behaviors (6,7). Physicians recorded caries status on the EFs using 11 categories to indicate the number of teeth with decay (0=None, followed by 10 categories from 1= 1–2 to 10=19–20 primary teeth). The sensitivity and specificity of medical providers' caries assessment was not evaluated. However, in a previous study in NC, physicians achieved a sensitivity of .76 and specificity of .95, compared to a pediatric dentist (gold standard), in identifying children with cavitated carious lesions (8). In addition to caries information from the EFs, we used NC Medicaid claims to obtain the child's age, race and county of residence. These data were supplemented with county level information on dental and pediatric primary care providers (9) and water fluoridation (10).

## Steps in imputation

For comparison we first imputed the caries data using a Poisson regression model, which should predict fewer zeros than the ZIP model. We then estimated two ZIP models, one with and one without accounting for the clustering of child-observations at the county level to generate the caries predictions (see Appendix for SAS code and model results). Below we describe the process used to generate predictions using the ZIP regression model.

**Step 1: Estimate ZIP model with non-missing caries data—**None of the commercially available statistical software including SAS, Stata, SPSS and MLWiN offer built-in programs to impute count data with excess zeros. We therefore wrote a multiple imputation program for this purpose in SAS version 9.1). To impute the missing caries data, we first estimated a ZIP model using data from children with non-missing caries. The count of carious teeth was the dependent variable in this model. Processes 1 and 2 described above can have different covariates, we have $x_{ij} = w_{ij}$ in the dental caries example. These covariates included child's age in months, age squared and age cubed, whether the child is Hispanic, percent of the child's county population age 0–17 years living in poverty, and whether all or part of the county of residence is a dental health professional shortage area (HPSA). For the ZIP model with random county effects we also included random effects for the child's residence county in both parts of the model.

**Step 2: Generate predictions for level of caries based on estimated coefficients —**For each observation with missing caries status, the same covariates used in Step 1 are inserted into Equation 1 above. If $\hat{\varphi}_{ij}$, the predicted probability in the first part of the ZIP model

(Process 1), is less than a random number drawn from a uniform (0,1) random distribution, then the individual is assigned a value of zero (indicating no caries). Alternately, if the zero inflation predicted probability ($\hat{\varphi}_{ij}$) exceeds the random uniform draw, a random Poisson draw is used in Process 2 to generate the caries count. This process of generating a prediction for the caries count is repeated 20 times to obtain a dataset with 20 values for the imputed dental caries variable for each individual. It is important to note that our imputation technique requires that dental caries be the only variable with missing data.

## RESULTS

Children with missing caries were different from those with non-missing caries on a number of variables (see Table 1). A higher percent of those with non-missing caries were older and were referred to a dentist by their physician compared to those with missing caries (3% vs. 2%). Second, more children with missing caries were seen in medical practices located in counties (wholly or partially) designated a dental health professional shortage area. Table 2 provides a comparison of the distribution of the caries variable in the non-imputed and imputed datasets.

The distribution of imputed caries using the ZIP models (with and without random county effects) was similar to that in the non-imputed data. The Poisson model predicted far fewer children with no caries compared to the ZIP models. For the majority of the sample the value imputed for dental caries was zero. Across the 20 imputed datasets, the largest value imputed was a '10' indicating that the child had 19–20 teeth with decay.

## DISCUSSION

In the data used for this paper, children with missing caries data differed from those with non-missing caries on a number of important variables. Those with missing caries were likely to be younger and also less likely to have received a dental referral from their physician. Further, a higher proportion of children with missing caries information lived in an underserved area with respect to dental care. Therefore, it was important to impute the missing caries information as complete case analysis, by excluding children with missing caries information, would likely bias results of studies conducted with these data (1).

To our knowledge, this study is the first to impute dental caries data that exhibit a count distribution with excess zeros using a ZIP regression model. Because of the excess zeros common to population-level caries data it is important to account for them when imputing missing caries data. Similar to previous studies, we found that the ZIP model accounts for this over-inflation of the zero count and provides a better fit for dental caries data than the Poisson model (2). We have extended the application of the ZIP model to impute missing caries information collected as part of a population-based study, while also accounting for clustering of observations. Although this procedure is limited in not allowing imputation of missing observations on variables other than dental caries, techniques to impute such information for categorical and normally distributed continuous variables are widely available (1).

## Acknowledgments

## References

1. Allison, PD. Missing data. Thousand Oaks, Calif: Sage Publications; 2002.

2. Bohning D, Dietz E, Schlattmann P, Mendonca L, Kirchner U. The zero-inflated poisson model and the Decayed, Missing and Filled teeth index in dental epidemiology. Journal of the Royal Statistical Society Series A (Statistics in Society). 1999; 162(2):195–209.

3. Lewsey JD, Thomson WM. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socio-economic status. Community Dent Oral Epidemiol. 2004 Jun; 32(3):183–9. [PubMed: 15151688]

4. Rozier RG, Sutton BK, Bawden JW, Haupt K, Slade GD, King RS. Prevention of early childhood caries in North Carolina medical practices: implications for research and practice. J Dent Educ. 2003 Aug; 67(8):876–85. [PubMed: 12959161]

5. Lambert D. Zero-Inflated poisson regression models with an application to defects in manufacturing. Technometrics. 1992; 34:1–14.

6. Pahel, BT.; Rozier, RG.; Stearns, SC. Agreement between patient records and Medicaid claims in a medical office-based preventive dental program (# 413). Academy Health ARM; Orlando, FL: 2007.

7. Pahel, BT.; Rozier, RG.; Stearns, SC.; Preisser, JS.; Mayer, ML.; Clements, DA. J Dent Res. Vol. 87. 2008. Predictors and effectiveness of dental referrals by primary care physicians; p. 2434

8. Pierce KM, Rozier RG, Vann WF Jr. Accuracy of pediatric primary care providers' screening and referral for early childhood caries. Pediatrics. 2002 May; 109(5):E82–2. [PubMed: 11986488]

9. US Department of Health and Human Services. Area Resource File (ARF). Health Resources and Services Administration, Bureau of Health Professions; Rockville, MD: 2003.

10. North Carolina Oral Health Section. County-level water fluoridation data. Raleigh, NC: NC Department of Health and Human Services; 2007.

## APPENDIX: SAS CODE

## Multiple Imputation of Dental Caries Data Using a Zero Inflated Poisson Regression Model

### Code to estimate the Poisson model using data with non-missing dental caries information

```
proc genmod data = cariesnomissing;
model CARIES = AGE AGE_SQUARED AGE_CUBED HISPANIC PCT_POVERTY HPSA_WHOLE
HPSA_PARTIAL /dist=p;
output out=predpoi p=p;
run;
```

### Code to generate predictions for observations with the missing dental caries data using coefficients from the Poisson model

The number of imputations is specified by "numimp." All imputations are performed in one data step. "Seed =&seedval" sets the seed for the first random number call but has no effect thereafter.

```
%macro impute (output, seedval, numimp);
data &output;
set cariesmissing;
seed=&seedval;
do MInum = 1 to &numimp;
lambda= exp(-5.6120 + 0.3029*AGE - 0.0078*AGE_SQUARED + 0.0001*AGE_CUBED +
0.5562*HISPANIC -.0012*PCT_POVERTY + 0.6462*HPSA_WHOLE -0.1331*HPSA_PARTIAL);
```

```
CARIES = ranpoi(seed, lambda);
output;
end;
run;
%mend impute;
%impute (imputeall, 334, 20);
```

### Code to estimate the ZIP model using the data with non-missing dental caries information

Note: Although we use PROC NLMIXED (SAS version 9.1) to estimate the ZIP model, PROC GENMOD and PROC COUNTREG in SAS version 9.2 or higher are now available to estimate a ZIP model. However, PROC GENMOD and PROC COUNTREG do not allow inclusion of random effects in the ZIP model. Additional information about using SAS to estimate the ZIP model can be gained from referring to postings by Dale McLerran on the publicly accessible SAS-L listserv of the University of Georgia (http://www.listserv.uga.edu/archives/sas-l.html).

```
proc nlmixed data=cariesnomissing qpoints=15;
/* Enter starting values for grid search */
parms a0= 3.1999 a1= .05747 a2= −.00796 a3= .000126 a4= −.5660 a5= .00459
a6= −.00979 a7= .1397
b0= .1660 b1= −.03127 b2= .003687 b3= −.00007 b4= .05623 b5= −.02372
b6= −.07242 b7= −.05164;
linpinfl= a0 + a1*AGE + a2*AGE_SQUARED + a3*AGE_CUBED + a4*HISPANIC +
a5*PCT_POVERTY + a6*HPSA_WHOLE + a7*HPSA_PARTIAL;
infprob= 1/(1+exp(−linpinfl)); /* inflation probability for zeros */
lambda= exp(b0 + b1*AGE + b2*AGE_SQUARED + b3*AGE_CUBED + b4*HISPANIC +
b5*PCT_POVERTY + b6*HPSA_WHOLE + b7*HPSA_PARTIAL);
if CARIES = 0 then prob = infprob + (1−infprob)*exp(−lambda);
if CARIES = 0 then loglike = log(prob);
else loglike = log((1−infprob)) + CARIES*log(lambda) −
lambda − lgamma(CARIES+1);
model CARIES ~ general(loglike);
ODS output ParameterEstimates=p1;
run;
```

### Code to generate predictions for observations with the missing dental caries data

The number of imputations is specified by "numimp." All imputations are performed in one data step. "Seed =&seedval" sets the seed for the first random number call but has no effect thereafter.

```
%macro impute (output, seedval, numimp);
data &impute;
set cariesmissing;
seed=&seedval;
do MInum=1 to &numimp;
/* insert code for computing linear predictors for each observation */
linpinfl = 3.3269 + 0.03722*AGE −0.00674*AGE_SQUARED + .000106*AGE_CUBED
−0.617*HISPANIC + 0.007207*PCT_POVERTY −0.8311*HPSA_WHOLE
+ 0.1339*HPSA_PARTIAL;
infprob= 1/(1+exp(−linpinfl)); /* inflation probability for zeros */
```

```
if ranuni(seed) < infprob then CARIES =0;
else do;
lambda= exp(.05855 – 0.04261*AGE + 0.004643*AGE_SQUARED – 0.00008*AGE_CUBED
+ 0.07227*HISPANIC – 0.00621*PCT_POVERTY – 0.02893*HPSA_WHOLE
– 0.1033*HPSA_PARTIAL);
CARIES =ranpoi(seed, lambda);
end;
output;
end;
run;
%mend impute;
%impute (imputeall, 334, 20);
```

## Code to estimate the ZIP model with county random effects using the data with non-missing dental caries information

Note: The random effects for the two parts of the ZIP model are u1 and u2, and s2u1 and s2u2 are their respective variances. The model assumes that the random effects are normally distributed and have zero covariance.

```
proc sort data=cariesnomissing; by COUNTY_ID; run;
proc nlmixed data=cariesnomissing qpoints=15;
/* Enter starting values for grid search */
parms a0= 3.1999 a1= .05747 a2= –.00796 a3= .000126 a4= –.5660 a5= –.00459
a6= –.00979 a7= .1397 s2u1= .3820
b0= .1660 b1= –.03127 b2= .003687 b3= –.00007 b4= .05623 b5= –.02372
b6= –.07242 b7= .05164 s2u2= .06351;
linpinfl= a0 + a1*AGE + a2*AGE_SQUARED + a3*AGE_CUBED + a4*HISPANIC +
a5*PCT_POVERTY + a6*HPSA_WHOLE + a7*HPSA_PARTIAL + u1;
infprob= 1/(1+exp(-linpinfl)); /* inflation probability for zeros */
lambda= exp(b0 + b1*AGE + b2*AGE_SQUARED + a3*AGE_CUBED + a4*HISPANIC +
a5*PCT_POVERTY + a6*HPSA_WHOLE + a7*HPSA_PARTIAL + u2);
if CARIES = 0 then prob = infprob + (1-infprob)*exp(-lambda);
if CARIES = 0 then loglike = log(prob);
else loglike = log((1-infprob)) + CARIES1*log(lambda) –lambda –
lgamma(CARIES1+1);
model CARIES ~ general(loglike);
random u1 u2 ~ normal([0,0], [s2u1, 0, s2u2]) subject = COUNTY_ID
out=random_effects;
run;
```

## Code to generate predictions for observations with the missing dental caries data

The number of imputations is specified by "numimp." All imputations are performed in one data step. "Seed =&seedval" sets the seed for the first random number call but has no effect thereafter.

```
%macro impute (output, seedval, numimp);
data &output;
set cariesnomissing;
seed=&seedval;
do MInum =1 to &numimp;
```

```
/* Code for computing linear predictors for each observation */
linpinfl = 3.1979 + .05898*AGE −.00796*AGE_SQUARED + .000126*AGE_CUBED
−.5953*HISPANIC + .00547*PCT_POVERTY − .04483*HPSA_WHOLE
+ .1686*HPSA_PARTIAL + u1;
infprob= 1/(1+exp(−linpinfl)); /* inflation probability for zeros */
if ranuni(seed) < infprob then CARIES =0;
else do;
lambda= exp(.1729 − .02864*AGE + .003753*AGE_SQUARED −.00007*AGE_CUBED
+ .01028*HISPANIC −.02744*PCT_POVERTY −.03358*HPSA_WHOLE
−.08311*HPSA_PARTIAL + u2);
CARIES =ranpoi(seed, lambda);
end;
output;
run;
run;
%mend impute;
%
```

*impute*

```
(imputeall, 35, 20);
```

Note: Convergence of the estimation algorithm is aided by using starting values for the parameters as determined from simpler models that omit the random effects or the zero-inflation portion of the model. The imputation program generates a dataset with 20 imputed values for the CARIES variable for each individual in the dataset with missing Caries information. Little and Rubin state that ten imputations usually are sufficient for a broad range of applications (1). Further, once data have been imputed, special consideration needs to be given to within and between subject variance when interpreting regression estimates generated using the imputed data. Space limitations preclude us from describing in detail the proper analysis of multiply imputed data. However, such techniques are widely available (1,2). For example, the PROC MIANALYZE procedure in SAS can be used to adjust variance estimates for multiply imputed data.

Coefficient estimates from models estimated to impute the missing dental caries data

| Variables | Poisson model | | Zero-inflated Poisson model | | Zero-inflated Poisson model with county random effects | |
|---|---|---|---|---|---|---|
| | Coefficient estimate | Std. Error | Coefficient estimate | Std. Error | Coefficient estimate | Std. Error |
| | | | Estimates from Poisson part of ZIP model | | | |
| Intercept | −5.6[***] | 0.9 | 0.1[***] | 0.8 | 0.2[**] | 0.5 |
| Age | 0.3[*] | 0.1 | −0.04[***] | 0.11 | −0.03 | 0.08 |
| Age squared | −0.01 | 0.01 | 0.01[***] | 0.01 | 0.004 | 0.004 |
| Age cubed | 0.0001 | 0.0001 | −0.0001[***] | 0.0001 | −0.0001 | 0.0001 |
| Hispanic (vs. not Hispanic) | 0.6[***] | 0.1 | 0.1 | 0.1 | 0.01 | 0.07 |
| % population in child's county of residence living in poverty | −0.001 | 0.002 | −0.01 | 0.01 | −0.03[*] | 0.01 |

| Variables | Poisson model | | Zero-inflated Poisson model | | Zero-inflated Poisson model with county random effects | |
|---|---|---|---|---|---|---|
| | Coefficient estimate | Std. Error | Coefficient estimate | Std. Error | Coefficient estimate | Std. Error |
| | | | **Estimates from Poisson part of ZIP model** | | | |
| Health professional shortage area (HPSA) designation of child's residence county | | | | | | |
| Whole county is a dental HPSA | 0.7 | 0.4 | −0.03 | 0.08 | −0.03 | 0.12 |
| Part of county is a dental HPSA | −0.1 | 0.3 | −0.1 | 0.1 | 0.08 | 0.08 |
| Random county effect | | | | | 0.10[*] | 0.03 |
| | | | **Estimates from Zero-inflation part of ZIP** | | | |
| Intercept | | | 3.3 | 0.9 | 3.2[*] | 0.5 |
| Age | | | 0.04 | 0.57 | 0.1 | 0.1 |
| Age squared | | | −0.01[*] | 0.01 | −0.008[*] | 0.004 |
| Age cubed | | | 0.0001[*] | 0.0001 | 0.0001[*] | 0.0001 |
| Hispanic | | | −0.6[***] | 0.1 | −0.6[***] | 0.1 |
| % population in child's county of residence living in poverty | | | 0.01 | .01 | 0.01 | 0.02 |
| Health professional shortage area (HPSA) for dental care designation of child's residence county | | | | | | |
| Whole county is a dental HPSA | | | −0.8[***] | 0.2 | −0.1 | 0.2 |
| Parts of county is a dental HPSA | | | 0.13[**] | 0.09 | 0.2 | 0.1 |
| County random effect | | | | | 0.4[***] | 0.1 |

[*]
$P$  .05,

[**]
$P$  .001,

$P$  .0001

N = 17,766

## References

1. Little, R.; Rubin, D. Statistical analysis with missing data. 2. Hoboken, NJ: John Wiley and Sons, Inc; 2002.

2. Allison, PD. Missing data. Thousand Oaks, Calif: Sage Publications; 2002.

**Table 1**

Descriptive statistics for the samples with and without missing dental caries data

| Variable | Children with observed dental caries (N=17,766) | | Children with missing dental caries (N=6,637) | |
| --- | --- | --- | --- | --- |
| | **Mean or Percent** | **Std. Deviation** | **Mean or Percent** | **Std. Deviation** |
| Child's age in months[*] | 16.20 | 7.21 | 15.16 | 6.80 |
| Child is Hispanic (vs. not) | 15.41 | .36 | 14.56 | .35 |
| Child received a referral for dental care from a physician[*] | 3.09 | .17 | 1.97 | .14 |
| Percent of child's county population 0–17 yrs. of age living in poverty | 14.02 | 4.04 | 14.27 | 4.57 |
| Health Professional Shortage Area for dental care (HPSA) designation for child's county, 2000 | | | | |
| No part of county is a HPSA[*] | 63.85 | .48 | 56.67 | .50 |
| One or more parts of the county designated as HPSA[*] | 33.06 | .47 | 37.19 | .48 |
| Whole county designated as HPSA[*] | 3.09 | .17 | 6.15 | .24 |

[*] Chi-square test significant at P .01

**Table 2**

Comparison of the distribution of the dental caries variable in the non-imputed and imputed datasets

| Number of teeth with caries | Non-imputed data (N=17,766 ) | Imputed data (N=6,637) | | |
| --- | --- | --- | --- | --- |
| | | Poisson Model | Zero Inflated Poisson Model | Zero Inflated Poisson Model with County Random Effects |
| | % | % | % | % |
| None | 93.66 | 89.01 | 94.19 | 94.14 |
| 1 or 2 | 3.72 | 9.54 | 2.68 | 2.92 |
| 3 or 4 | 1.47 | 1.21 | 1.78 | 1.75 |
| 5 or 6 | 0.52 | 0.19 | 0.86 | 0.79 |
| 7 or 8 | 0.24 | 0.04 | 0.34 | 0.28 |
| 9 or 10 | 0.13 | 0.01 | 0.11 | 0.10 |
| 11 to 20 | 0.26 | 0.00 | 0.04 | 0.02 |
| Akaike's Information Criterion (AIC) for model fit[£] | | 0.6372 | 0.6062 | 0.5965 |

[£]Smaller value indicates better model fit