CLS

Centre for Longitudinal Studies

CLS Cohort Studies

Working Paper 2008/5

Optimal modelling of hearing impairment in middle age in relation to hearing in childhood as measured by audiograms

Russell Ecob

July 2008

Optimal modelling of hearing impairment in middle age in relation to hearing in childhood as measured by audiograms

Russell Ecob

July 2008

First published in 2008 by the
Centre for Longitudinal Studies
Institute of Education, University of London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

ISBN 1 898453 69 1

The Centre for Longitudinal Studies (CLS) is an ESRC Resource Centre based at the Institution of Education. It provides support and facilities for those using the three internationally-renowned birth cohort studies: the National Child Development Study (1958), the 1970 British Cohort Study and the Millennium Cohort Study (2000). CLS conducts research using the birth cohort study data, with a special interest in family life and parenting, family economics, youth life course transitions and basic skills.

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875
email: info@cls.ioe.ac.uk

Contents

Contributors

Russell Ecob
Ecob Consulting
36 Prospecthill Road
Glasgow G42 9LE
+44(0)141-649-9354
www.ecob-consulting.com
russell@ecob-consulting.com

Acknowledgements

I am grateful to Alicia Rudnicka, Adrian Davis, Chris Power and David Strachan for their stimulation to produce this paper as well as to all participants in the 2002 -2004 biomedical follow up and to nurses, office and laboratory staff who contributed to the successful completion of the fieldwork. Also to Graham Sutton and Pauline Smith for their work identifying correlates of childhood hearing loss from the dataset for use in the Multiple Imputations.

Abstract

NCDS (National Child Development Survey) data include detailed audiograms at childhood ages 7,11,16 years and a ‘cut down’ audiogram at 45 years (frequencies 1 and 4 kHz only). A range of models relating data in childhood to adulthood is examined. Hearing losses at all childhood ages contribute independently and with comparable relationships to adult hearing loss: such models should include adjustments at all ages and at all frequencies in combination. This is possible through polynomial contrasts between frequencies at each age.

Models which use the logarithm of the adult hearing measure as dependent variable are compared with those which use the raw score. Residuals from raw score models are found to have unacceptable distributional properties which are avoided by models of the log hearing loss. However, care needs to be taken in the interpretation of the coefficients in these models which can be easily transformed into an additive contribution due to the effects of further explanatory variables at specified values of existing variables. For categorical variables such as gender this will be a particular category and for continuous variables, such as childhood hearing, usually the mean.

A range of multiple imputation models is compared with a complete case analysis. The complete case analysis excludes more than half of the data in models. In these cases I recommend the use of multiple imputation models. These models are can be estimated using recent macros in ‘off-the-peg’ statistical software (e.g. Mice in STATA).

Mixed-effects and latent trajectory growth models are alternatives to the conditional models used here. These allow each individual to have a unique (growth) trajectory over time. Their advantages are limited in the case of these data due to the concentration of hearing loss data in childhood resulting from the absence of measurements of hearing loss in early adulthood.

1. Introduction

This paper is concerned with comparing alternative models for longitudinal data on hearing loss It is divided into four sections.

Section 2 describes the study sample and the measurement of hearing, both in the childhood waves and, on the current biomedical wave, by audiogram.

Section 3 describes possible methods of adjustment for childhood hearing loss, as measured by audiogram, and recommends a model which efficiently summarises information over age and frequency in terms of polynomials in both a reference (or base) frequency together with contrasts with other frequencies. Separate models are produced for adult hearing threshold levels (HTL) as dependent variables at each of 1 and 4 kHz.

Section 4 compares log and raw outcome measures – namely adult hearing threshold levels (HTL) (with corresponding transformations to childhood HTLs) in terms of the distribution of residuals from the chosen models. I justify our decision to log all hearing data.

Section 5 examines alternative multiple imputation models. In common with all longitudinal studies, NCDS has missing data. For the analyses of hearing in NCDS this occurs particularly for the childhood hearing threshold levels amounting to around a third of the possible data. Multiple imputation methods give consistent estimates of all model parameters subject to the ‘missing at random’ assumption. However choices of adjustment variables, number of imputation cycles and imputation method used will to some extent determine the estimates obtained. Estimates of effects of noise at work are given for a few selected choices of multiple imputation methods and a complete case analysis.

Finally, Section 6 considers alternative modelling procedures to the conditional models used here and evaluates their advantages and disadvantages. I conclude that when measurements are unequally distributed over the life course, as in these data, with a substantial concentration of hearing threshold measurements at childhood ages and then no hearing measure until the current, biomedical, wave (age 45 years) the advantages of random trajectory growth and mixed-effects models are generally outweighed by the conditional models used.

2. Data characteristics and study sample

2.1 Study sample

Participants were originally enrolled in the Perinatal Mortality Survey (PMS) of all those born in England, Scotland and Wales during one week in March 1958 (Power & Elliott 2006) and followed up throughout childhood and adulthood, most recently at 44-45 years. A total of 17,415 individuals participated in the PMS from an eligible sample of 17,638. Immigrants with the same birth dates were recruited up to age 16 years (n = 920), giving 18,558 eligible study participants (Total Cohort Sample). At 44 - 45 years, 12,069 participants, still in contact with the study, and who at 42 years had not required a proxy interview, were invited to a clinical examination undertaken in their home by a trained nurse. Of these, 9377 participants were seen between September 2002 and March 2004, 8894 of these having a valid hearing measure at 1 and 4 kHz. Note that in these analyses, fewer participants had an unskilled manual class (IV or V) in childhood compared to the total cohort sample, although the difference was small (22.3% versus 24.7%).

2.2 Measurement of hearing

Hearing is measured using a pure tone audiogram which measures the hearing threshold level (HTL) in decibels for a pure tone at each of several frequencies (British Society of Audiology, 1981). At all childhood ages (7,11,16 years) the full audiogram was used using frequencies 0.25, 0.5, 1, 2, 4, 8 kHz. In adulthood (45 years), due to resource constraints, a ‘cut down’ audiogram was used with frequencies 1,4 kHz only. 4 kHz is usually thought of as the frequency most indicative of hearing loss that is specifically noise induced.

As reported by Fogelman (1983) these tests, in childhood, were carried out using the audiometer facilities and procedures locally available and were not closely standardised in accord with general audiological practice at this time. This has several implications for the quality of data collected as the quality of audiometer calibration was probably variable and the type of earphone can affect output. Moreover no information is available concerning ambient noise conditions and, though instructed specifically for this task, only around 60% of testers had audiometric testing qualifications. The values at these ages therefore may not reflect true audiometric thresholds at these ages. At age 7 years step size was 10dB rather than the standard 5dB: this would be expected to artificially raise apparent HTLs at age 7 years on average by 2.5dB and result in adjustments at this age with less than optimal precision. Conductive hearing loss was identified by proxy information obtained at ages 7,11 (see Appendix 1).

In adulthood, testing was carried out by the study research nurses who received training from experienced Audiologists. Only information from completed tests was used. Information on whether there was ambient noise at test (yes/no) was available and was included in all analyses as a ‘nuisance’ variable.

3. Methods of adjustment for childhood hearing loss

3.1 Overview of adjustment methods

The effect of a range of explanatory variables (lifetime exposure to noise, social class, body mass index etc) on adult hearing loss can only be ascertained once proper adjustments for hearing in childhood and other relevant factors (gender, family history of hearing loss, medical problems etc) are included in the models used.

We examine alternative methods of modelling hearing thresholds in childhood over the range of frequencies and ages at which measured. Analyses are carried out both separately at each of the ages 7,11 and16 years and in combination over all ages. Both outcome frequencies (1 and 4 kHz) are examined.

The combination of the three ages in childhood and the six frequencies (0.25, 0.5, 1, 2, 4, and 8 kHz) at which a hearing threshold is obtained can potentially give rise to substantial multicollinearity. This chapter examines ways of efficiently utilising this available information whilst minimising any increase in standard errors of the estimates of the effects of interest. We evaluate this in relation to retrospective recall of noise exposure at work, a variable of particular substantive interest (see Ecob et al, 2008). In this process it is hoped that further insights will be gained into the influence of hearing impairment in childhood on the natural development of hearing loss in adulthood.

The approach taken is, for each outcome frequency, to:

a) identify the frequency (the base or reference frequency) at a given age at which hearing impairment is maximally related to outcome (hearing impairment at 45 years)

b) identify the nature of the relationship to this base frequency at each age via polynomial terms in this frequency

c) include any other frequencies which are independently predictive of the outcome via polynomial contrasts with the base frequency

d) identify a 'best' model

e) compare this model with other simpler models (e.g. averaging frequencies, within two ranges of high and low frequency, single childhood age, base frequency only)

Note that slightly different models are used in the different sections of this paper. For the decision on choice of base frequency only gender and whether ambient noise at test is adjusted for. For comparisons of alternative adjustments we adjust in addition for family history of hearing loss and for proxies for conductive hearing loss in childhood as well as for noise at work (on which the estimates are compared across models). The random part of models includes nurse only and a combination of nurse and instrument in different sections (nurse only for models in section 2 and nurse and instrument in combination for models in sections 3 and 4)[1].

STATA was used for all modelling using, in this chapter, the procedure 'xtreg', giving random effects analyses.

Hearing impairment as dependent variable is transformed using a log transformation having added a constant term with a value chosen in order to minimise the skewness of the variable[2]. This transformation correlates highly (0.998) with the optimal Box-Cox transformation. Hearing impairment at all childhood ages is similarly transformed. Hearing impairment in the better ear is used throughout.

Details of variables used organised by area/domain together with NCDS code, comments, and recoding, where relevant, are found in Appendix 1.

3.2 Choice of base frequency.

There are a number of ways in which the best choice of base or reference frequency for a given frequency of dependent variable (adult hearing at each frequencies 1 and 4 kHz) can be assessed. We carry out an examination taking all the explanatory frequencies at a specified age (7,11, and 16 years) and determining the frequency giving the highest t value:

a) taking all frequencies in combination
b) taking each frequency separately
c) taking frequencies in pairs with a range of possible combinations of the best (base) frequency with another frequency.

The best choice of base frequency will be:

a) that which takes the highest t value
b) that for which the maximum contribution of any additional frequency to explaining the variation in adult HTL outcome (as measured by the t value) is a minimum.

If any conflicts between these two criteria occur we chose criterion (a). All analyses adjust for gender and for ambient noise at adult test.

Table 1 shows, for 1 kHz and 4 kHz outcomes, the relation to each explanatory frequency **separately** at each age separately. At all childhood ages 2 and 4 kHz show maximal t values in relation to 1 kHz outcome and 2,4 and 8 kHz maximal relation to 4 kHz outcome.

---

[1] An examination of the effect of the random effects choice on the decision on base frequency showed no effect.

[2] Note that an additional constant term is necessary for the log transformation to accommodate the legitimate range of the hearing variables down to -10.

Table 1: t values for relation to adult HTLs at 1 or 4 kHz (each explanatory frequency separately)

Frequency of childhood	**1 kHz in adulthood**			**4 kHz in adulthood**		
	Childhood HTL at age:			*Childhood HTL at age:*		
HTL (kHz)	7 yrs	11 yrs	16 yrs	7 yrs	11 yrs	16 yrs
8	6.03	8.80	9.55	**12.39**	9.19	13.00
4	6.43	**11.21**	10.39	12.34	12.06	**15.14**
2	**8.00**	6.98	**11.73**	8.29	**13.84**	9.01
1	6.98	5.79	11.28	6.03	12.22	6.50
0.50	6.01	4.88	9.74	6.25	9.51	5.08
0.25	5.19	4.94	9.10	5.75	7.87	4.28

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Table 2 shows, for 1 kHz and 4 kHz outcomes, the relation to each explanatory frequency **in combination** at each age separately. In most cases 2 kHz shows maximal t values in relation to 1 kHz outcome and 4 kHz maximal relation to 4 kHz outcome. As expected, the t values in Table 1 are larger than in Table 2.

Table 2: t values for relations to adult HTLs at 1 or 4 kHz (each explanatory frequency in combination)

Frequency of childhood	**1 kHz in adulthood**			**4 kHz in adulthood**		
	Childhood HTL at age:			*Childhood HTL at age:*		
HTL (kHz)	7 yrs	11 yrs	16 yrs	7 yrs	11 yrs	16 yrs
8	1.08	3.44	2.14	5.63	0.90	5.61
4	-0.27	**6.76**	-0.27	**6.76**	2.16	**10.14**
2	**4.20**	-0.26	**4.30**	0.81	**3.63**	-3.18
1	1.74	-1.52	2.32	-2.94	3.03	-2.06
0.50	-0.03	-0.23	-0.36	0.68	0.19	0.79
0.25	-1.72	-0.72	1.11	-0.45	-0.94	-2.00

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Tables 3 - 5 show, for 4 kHz outcome, the relation to each other frequency in turn in relation to each base frequency in turn (two frequencies included in all models) at ages 7,11,16 years respectively. Setting the base frequency at 4 kHz is seen to result, at all ages, in a minimum for the largest t value for any other frequency. Note that this other frequency is 8kHz at ages 7, 11 and 1 kHz at age 16. When a frequency other than 4 kHz is set as base frequency, 4 kHz always has the highest t value.

Table 3: 4 kHz adulthood HTLs - t values for relation to hearing impairment at Age 7 – specified (explanatory) frequencies in addition to base frequency, taken one at a time

frequency of additional childhood HTL(kHz)	1	2	4	8	Base frequency (kHz)
8	6.74	5.68	**2.68**	-	
4	**9.68**	**8.65**	-	**5.85**	
2	4.67	-	-0.95	1.64	
1	-	0.69	-2.09	0.04	
0.5	1.08	0.76	-1.40	-0.02	

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Table 4:. 4 kHz adulthood HTLs- t values for relation to hearing impairment at Age 11 - specified (explanatory) frequencies in addition to base frequency, taken one at a time

frequency of additional childhood HTL (kHz)	1	2	4	8	Base frequency (kHz)
8	10.86	9.35	**4.67**	-	
4	**11.91**	**8.99**	-	**6.13**	
2	6.09	-	-1.59	1.54	
1	-	-0.53	-3.67	-0.78	
0.5	1.91	0.71	-1.99	-0.24	

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Table 5: 4 kHz adulthood HTLs - t values for relation to hearing impairment at Age 16 - specified (explanatory) frequencies in addition to base frequency, taken one at a time

Frequency of additional childhood HTL (kHz)	1	2	4	8	Base frequency (kHz)
8	11.16	10.05	3.56	-	
4	**13.22**	**13.53**	-	**8.93**	
2	6.18	-	-4.76	0.76	
1	-	-0.39	**-5.20**	-0.98	
0.5	-0.40	-0.57	-4.58	-1.50	

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Tables 6 - 8 show similar results for the 1 kHz outcome, the relation to each other frequency in relation to the base frequencies (two frequencies included in all models) at ages 7,11,16 respectively. Setting the base frequency at 2 kHz results, at all ages, in a minimum for the maximum t value for any other frequency. When a frequency other than 2kHz is set as the base frequency in all cases except at age 16 (4 and 8 kHz) 2 kHz is always the frequency having the highest t value.

Table 6: 1 kHz adulthood HTLs - t values for relation to hearing impairment at Age 7 - specified (explanatory) frequencies in addition to base frequency, taken one at a time

frequency of additional childhood HTL(kHz)	1	2	4	8	Base frequency (kHz)
8	2.08	**0.95**	2.58	-	
4	2.34	0.01	-	3.55	
2	**4.74**	-	**5.66**	**5.73**	
1	**-**	1.14	4.04	4.15	
0.5	-0.15	-0.16	2.21	2.65	

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Table 7: 1 kHz adulthood HTLs - t values for relation to hearing impairment at Age 11 - specified (explanatory) frequencies in addition to base frequency, taken one at a time

frequency of additional childhood HTL (kHz)	1	2	4	8	Base frequency (kHz)
8	3.86	3.11	4.34	-	
4	3.18	1.93	-	5.56	
2	**5.18**	-	**7.62**	**8.29**	
1	-	**3.81**	6.19	7.35	
0.5	1.30	2.99	5.02	5.66	

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Table 8: 1 kHz adulthood HTLs - t values for relation to hearing impairment at Age 16 - specified frequencies in addition to base frequency, taken one at a time

frequency of additional childhood HTL (kHz)	1	2	4	8	Base frequency (kHz)
8	3.83	2.07	2.39	-	
4	5.53	3.58	-	6.78	
2	6.17	-	7.22	9.24	
1	-	4.46	7.48	9.95	
0.5	-0.36	2.38	4.88	6.44	

Notes: 1 maximum t value in each column is in bold
2 linear terms only

Thus, both for separate and combined analyses (frequencies within age) setting a base frequency of 4 kHz for an outcome of 4 kHz and 2 kHz for an outcome of 1 kHz generally results both in maximal t values for this frequency and minimal t values for any other frequency, when taken in pairs. Thus each of these analyses point to the same conclusion as regards the base frequency at all ages; **that 4 kHz be used as a base frequency for 4 kHz outcome and that that 2 kHz be used as a base frequency for 1 kHz outcome.**

3.3 Choice of polynomial terms in base frequency

A series of models is now run in which polynomial terms in relation to the (log) base frequency at each age separately are included, these being centered on the (log) base frequency.

The relationships of adult hearing loss the base frequency at each childhood age is tested using polynomials of up to degree 3 (cubic) (see Table 9). Resulting best models (after backward elimination) are, for 1 kHz (adulthood), at 7 years of age, cubic; 11 years quadratic and, at 16 years, linear. At 4 kHz (adulthood) all relationships are linear. Relationships at all childhood ages are comparable for each of 1 kHz and 4 kHz outcomes (for 1 kHz R squared is 0.018, 0.027, 0.029 at 7,11,16 years respectively; for 4 kHz R squared is 0.074, 0.079, 0.010 at 7,11,16 years respectively) . Standard deviation of childhood HTL is 0.24, 0.51, 0.40 at ages 7,11,16 respectively for 4 kHz and 0.23, 0.61, 0.48 at ages 7,11,16 respectively for 1kHz. The smaller standard deviations at 7 years correspond to different coding conventions at this age. (see section 2.2),

Table 9: 1 and 4 kHz adulthood HTLs – estimates (standard errors) of relation to base frequency of childhood HTLs

	Linear	**Quadratic**	**Cubic**
1 kHz			
7 yrs	0.051 (0.017)	-0.051 (0.029)	-0.041 (0.016)
11 yrs	0.048 (0.004)	0.014 (0.007)	-
16 yrs	0.110 (0.008)	-	-
4 kHz			
7 yrs	0.102 (0.009)	-	-
11 yrs	0.096 (0.008)	-	-
16 yrs	0.335 (0.022)	-	-

3.4 Choice of contrasts between base and other frequencies

Contrasts with base frequency are now formed based on a log scale. The lowest frequency was eliminated as this showed minor relationships to outcomes in combination with other frequencies (see Table 1). Thus, for 4 kHz, the terms, corresponding to 0.5, 1, 2, 4, and 8 kHz were -3, -2, -1, 0, 1 respectively and for 1 kHz, the terms, corresponding to 0.5, 1, 2, 4, and 8 kHz were -2, -1, 0, 1, 2 respectively. Linear, quadratic and cubic terms in these contrasts were created and tested using backward elimination. Table 10 shows the polynomial contrasts used in the following models. None of the cubic contrasts achieve statistical significance but both linear and quadratic contrasts do at all ages and at both frequencies. At 4 kHz all the terms have positive sign, indicating a further positive partial relationship at higher frequencies (8kHz) to adult hearing loss and/or a negative partial relationship to lower frequencies (0.5 to 2 kHz). In contrast, at 1 kHz all signs of linear terms are negative, indicating a negative partial relationship for higher childhood frequencies (4 and 8 kHz) to adult hearing loss (1 kHz) and vice versa for lower frequencies . Relationships at 16 years are comparable but stronger for the 4 kHz outcome than at ages 7 and 11 (see Figures 1a.b).

Table 10: 1 and 4 kHz adulthood HTLs – estimates (standard errors) for contrasts with base frequency of childhood HTLs

	Linear	Quadratic	Cubic
1 kHz			
7 yrs	-0.051 (0.029)	-0.041 (0.016)	-
11 yrs	-0.006 (0.003)	0.006 (0.002)	-
16 yrs	-0.021 (0.004)	0.016 (0.002)	-
4 kHz			
7 yrs	0.042 (0.013)	0.012 (0.005)	-
11 yrs	0.050 (0.007)	0.015 (0.003)	-
16 yrs	0.101 (0.017)	0.024 (0.007)	-

Figure 1a: Adult HTL (1 kHz) in relation to contrasts between childhood HTL frequencies (in relation to 2 kHz)

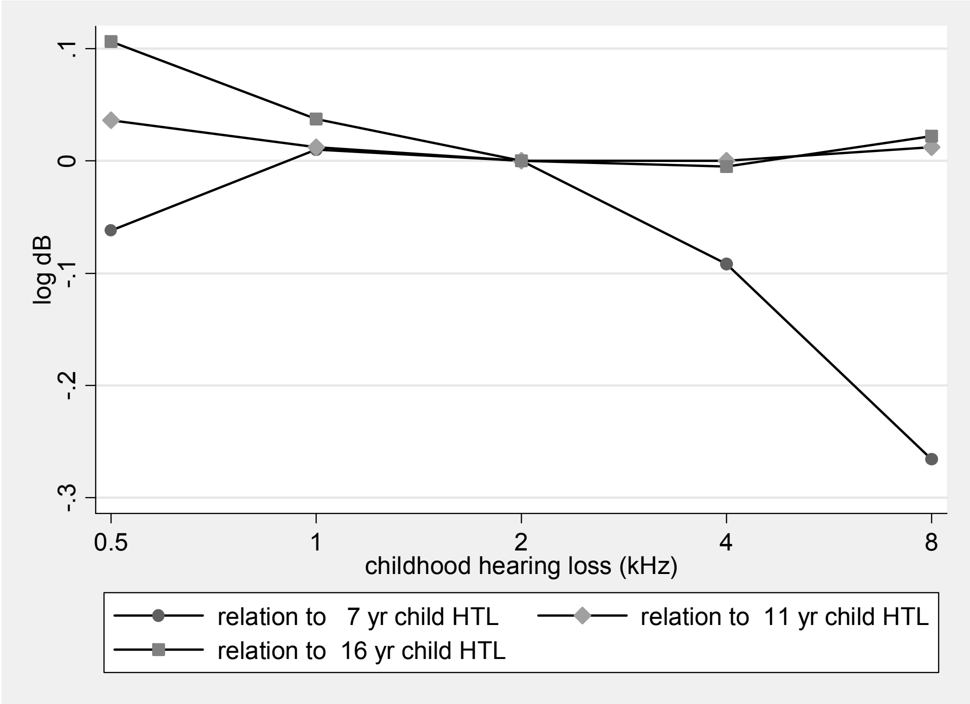
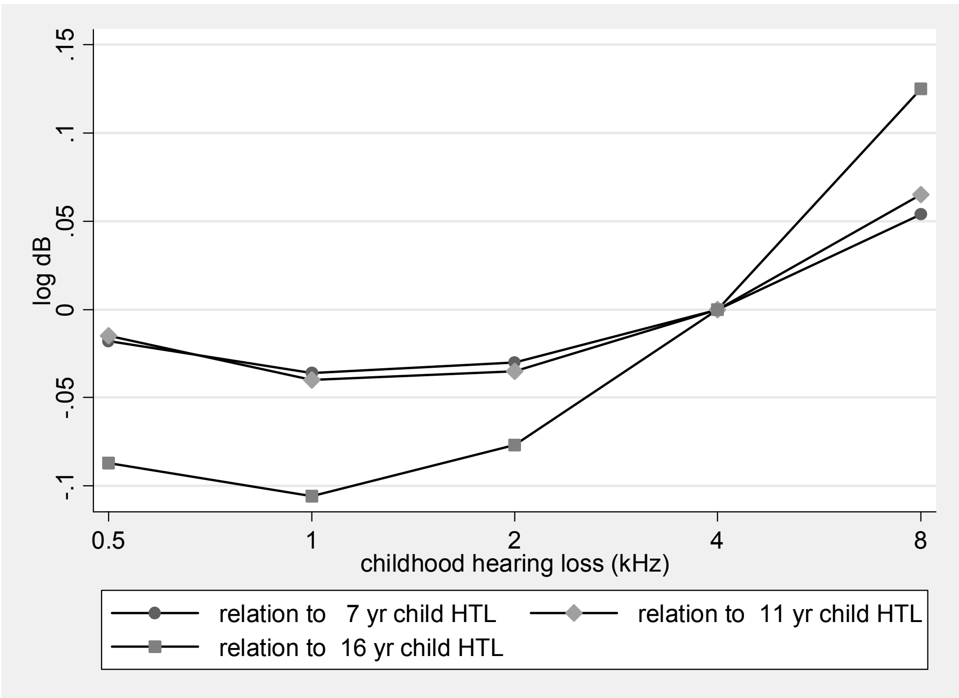


Figure 1b: Adult HTL (4 kHz) in relation to contrasts between child HTL frequencies (in relation to 4 kHz)



Note that for 1 kHz we have 6 terms in the base frequencies (all childhood ages) and 6 contrasts. For 4 kHz we have 3 terms in the base frequencies (all childhood ages) and 6 contrasts.

For all substantive papers (e.g. Ecob et al, 2008), which may adjust for a range of further factors, these terms are assessed independently for statistical significance (and included in the models only when so) in the context of the most basic model used in the paper.

3.5 Comparison of models with different adjustments for childhood hearing loss

We now compare our ‘full’ adjustment model (model 7 in Table 12), on both 1 kHz and 4 kHz adult hearing loss, to a range of other models which includes:

a) model (1) with no childhood hearing loss controls

b) models (2, 3, and 4) which include childhood hearing loss at each frequency separately (using appropriate terms extracted from the full model)

c) a model (5) which using the base frequencies (only) at each age

d) a model (6) which uses the high (4 kHz and 8 kHz) and low (0.25, 0.5, 1, and 2 kHz) frequencies averaged.

All models here include proxies for conductive hearing loss at ages 7 and 11 family hearing loss and noise (see Ecob et al, 2008) in addition to the variables in the previous models. The adjustment terms in the full model are pruned on the basis of statistical significance on the basis of relationships to hearing outcome in models (without noise or any of the additional variables above included).

This resulted in the following terms:

	1 kHz		4 kHz	
	Base	Frequency contrasts	Base	Frequency contrasts
7	l,q,c	-	l	-
11	l	-	l	l,q
16	l	l,q	l	l
age				

Key l=linear, q=quadratic c=cubic

Note that the number of statistically significant frequency contrasts in the models above are lower than in the models in Table 10. This is because in Tables 10 each childhood age was considered separately

This gives 7 terms for 1 kHz models and 6 terms for 4 kHz models

Model 6 first averages the low and high raw frequencies and then constructs Y=loge(α +x) where α is chosen to minimise skewness (using procedure lnskew0 in STATA). Values of a α re shown in Table 11. The higher values of α for 16 years reflects the lower values of skewness of hearing loss at this age (*see Figure 2).*

Table 11: Constant addition terms ( ) in order to obtain logged childhood hearing loss with zero skewness

Age	Low frequency	High frequency
7	12.61	7.30
11	13.50	4.64
16	26.27	20.38

Table 12 shows the coefficient of the highest value (>5 years) of noise exposure (in relation to no noise exposure) for a range of models including a 'full adjustment' model described above with base frequency and polynomial contrasts at each age. For 4 kHz with increasing adjustment the coefficient of noise decreases whereas for 1 kHz with increasing adjustment the coefficient of noise increases[3].

Table 12: 1 and 4 kHz adulthood HTLs – estimates of coefficient of highest value (> 5 years versus none) of noise exposure (standard error) for a range of models for adjustment for childhood hearing loss

Model No.	description	1 kHz	4 kHz
1	No adjustment	0.021 (0.006)	0.122 (0.012)
2	Adjustment age 7 years only	0.017 (0.007)	0.113 (0.015)
3	Adjustment age 11 years only	0.023 (0.006)	0.103 (0.016)
4	Adjustment age 16 years only	0.026 (0.008)	0.100 (0.019)
5	Adjustment at all ages using base frequency only	0.026 (0.008)	0.088 (0.019)
6	Adjustment at all ages using high/low frequencies	0.029 (0.008)	0.094 (0.020)
7	Full adjustment	0.028 (0.009)	0.085 (0.019)

Model 6 (which uses high and low averaged frequencies at each age) is seen to under adjust in relation to the full adjustment for 4 kHz (with little difference in standard error), giving larger estimates of the effects of noise, though the estimates for 1 kHz outcome are very similar.

[3] Note that we are assuming that the best model is that with the fullest adjustment for childhood hearing loss. After a certain point, further adjustment will, through the entailed multicollinearity, result in increases in the standard errors resulting in an increase in Mean Squared Error from a particular standard. However the adjustments made here do not show such increases in standard errors. Also the effect of further adjustment for childhood hearing loss on the coefficients of a related variable, such as noise at work, depends on the partial correlation of this variable with childhood hearing loss given the dependent variable. The raw correlations between noise at work and childhood hearing loss are relatively low (values 0.042, 0.023, 0.016 with childhood hearing loss in base frequency for 4 kHz at ages 16,11,7 respectively). This will limit the differential effect of alternative controls on the criterion of choice, the coefficient of the maximal noise at work category.

In all other cases model 5 (which adjusts at all ages for the base frequency only) comes closest to the full adjustment, followed by adjustment at age 16 only, then at age 11 only, then at age 7 only. Note that changes in coefficients between models are in opposite directions for 1kHz and 4 kHz

The manner of change in the coefficient of high noise levels with adjustment is partly explained by the higher correlation of this indicator variable with the 4 kHz outcome (Pearson's ρ=0.18) than with the 1 kHz outcome (Pearson's ρ =0.04)

The full adjustment model involves some increase, as expected, in standard error (in relation to no adjustment) though this is negligible in relation to other models which include adjustments at more than one childhood HTL and the increases in relation to models which adjust for HTL at one childhood age only are modest.

Note that alternative models of change over the life course are shown in de Stavola et al (2006). These allow, through reparametrisation of models such as those above, for a re-expression of the models to show directly the effect of differences in hearing loss at the different childhood ages on HTL in adulthood.

4. Comparisons of raw and log outcome measures

For models to be correctly specified the residuals must obey the following criteria:

1. be normally distributed
2. have constant variance over the range of values of key explanatory variables

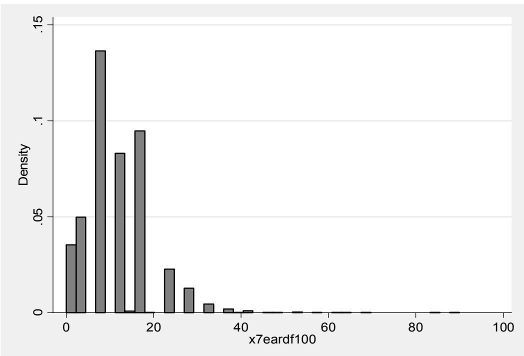
Models which do not satisfy (1) can be estimated using robust estimation methods. Heteroscedasticity (2) can be taken into account through generalised or weighted least squares estimation methods.

We aim, for simplicity, to run models without involving these estimation methods where possible and so we now investigate whether these assumptions are satisfied for our data through the modelling of raw hearing loss at adulthood in relation to raw childhood hearing loss and, if not, whether a suitable logarithmic transformation of either or both adult and childhood hearing measures will result in improvements in the criteria above. Note that this procedure does not automatically transform the residuals in the model into the appropriate form and the decision to transform the childhood hearing loss measures is based on knowledge that the relationships of log outcome to log explanatory variables tend to be more linear and with a higher degree of explanation than the relationship to the raw measures.

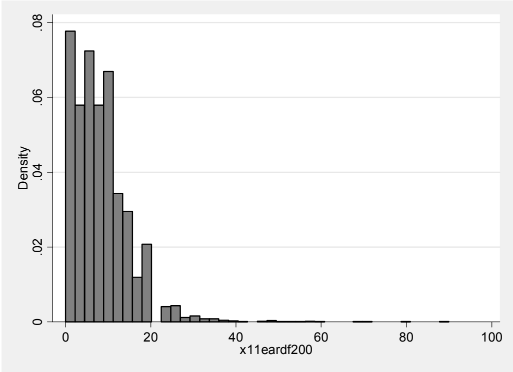
Figure 2 shows the distributions of HTL's at all childhood ages. Skewness is more marked at both frequencies at the younger ages. Figure 3 shows small differences in the skewness of the adult HTLs as a function of noise at work but not by gender.

Figure 2. Distributions of childhood hearing threshold levels; base frequencies at each age

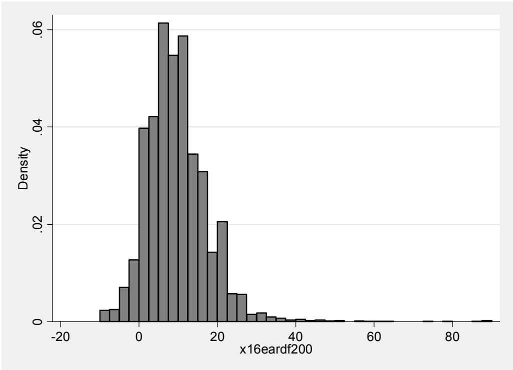
7 year 2kHz



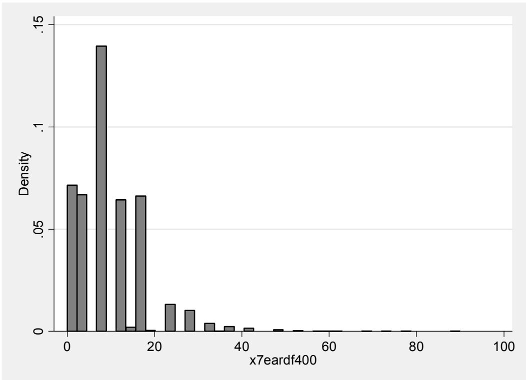
11 year 2kHz



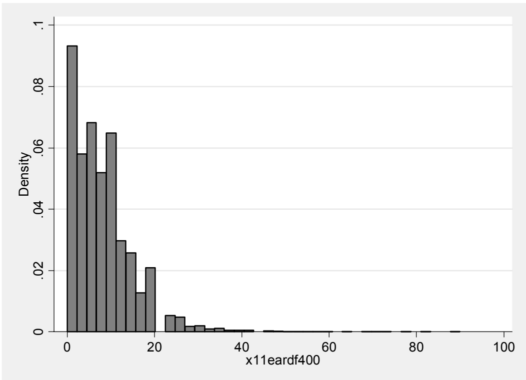
16 year 2kHz



7 year 4 kHz



11 year 4 kHz



16 year 4 kHz

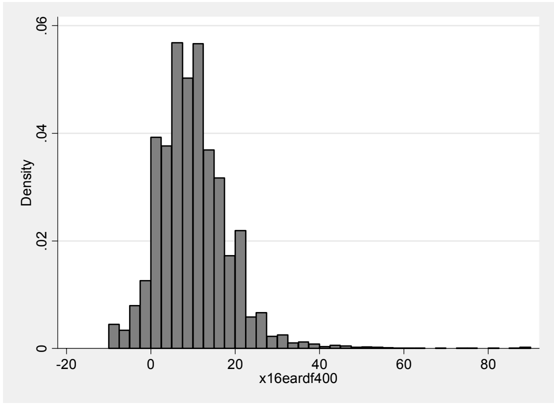
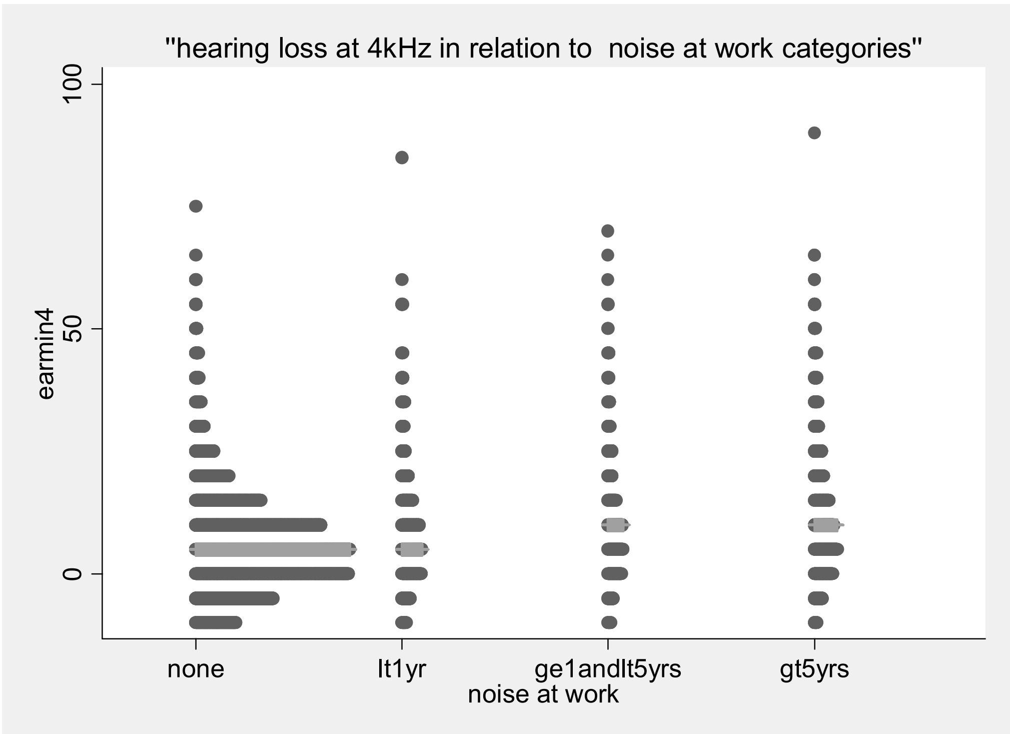


Figure 3. Distribution of outcome variables in relation to gender and noise at work

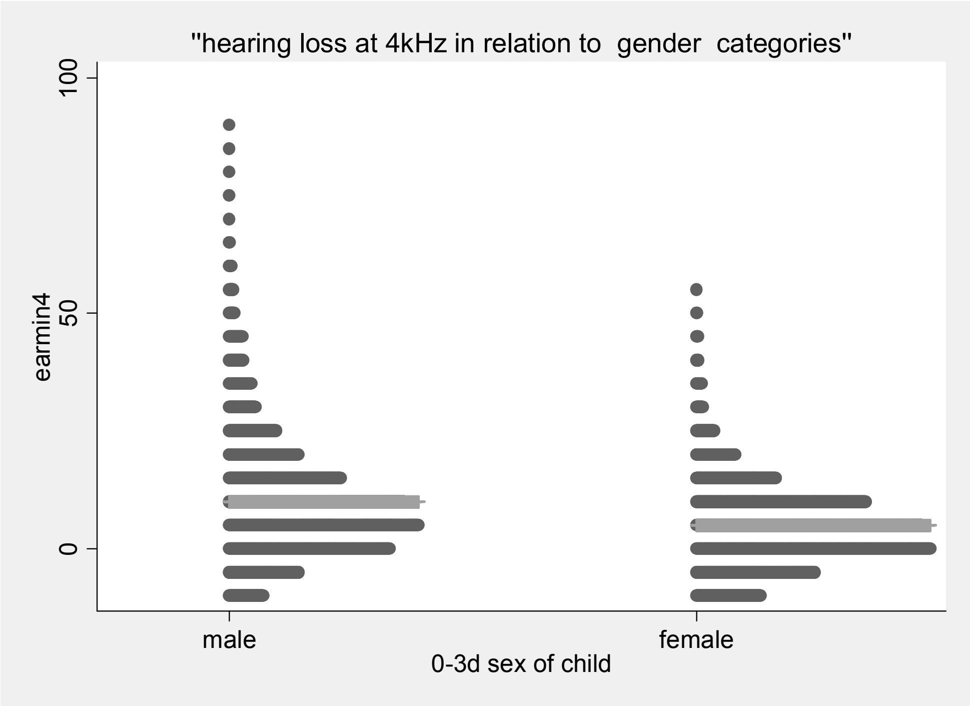
a) 4 kHz by noise at work categories



Skewness

None	<1 year	1-5 years	>5 years
0.51	**0.56**	**0.48**	**0.47**

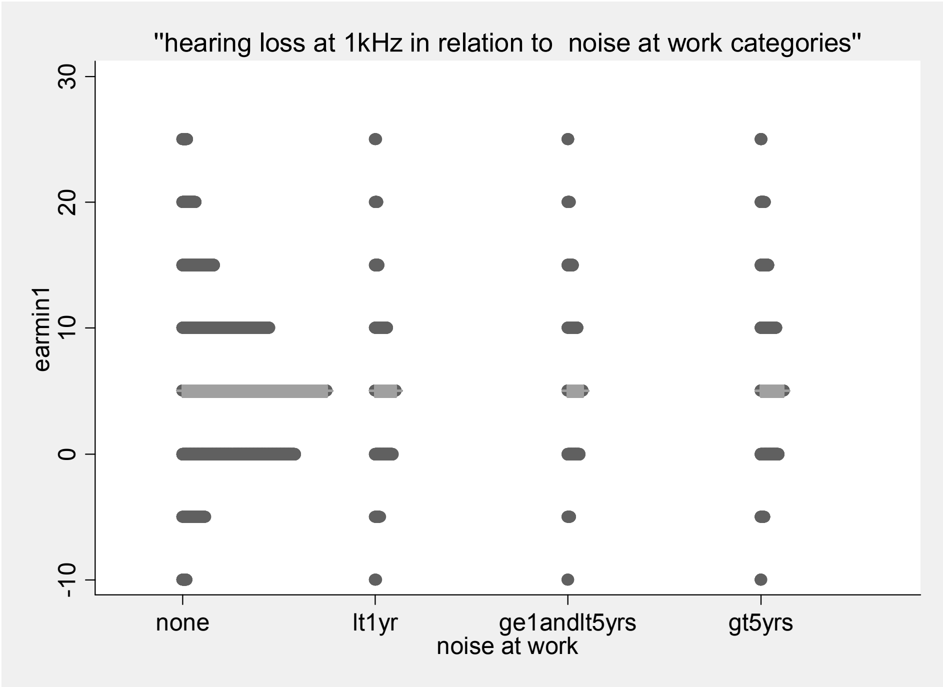
b) 4 kHz by gender categories



Skewness

Male	female
1.27	**0.93**

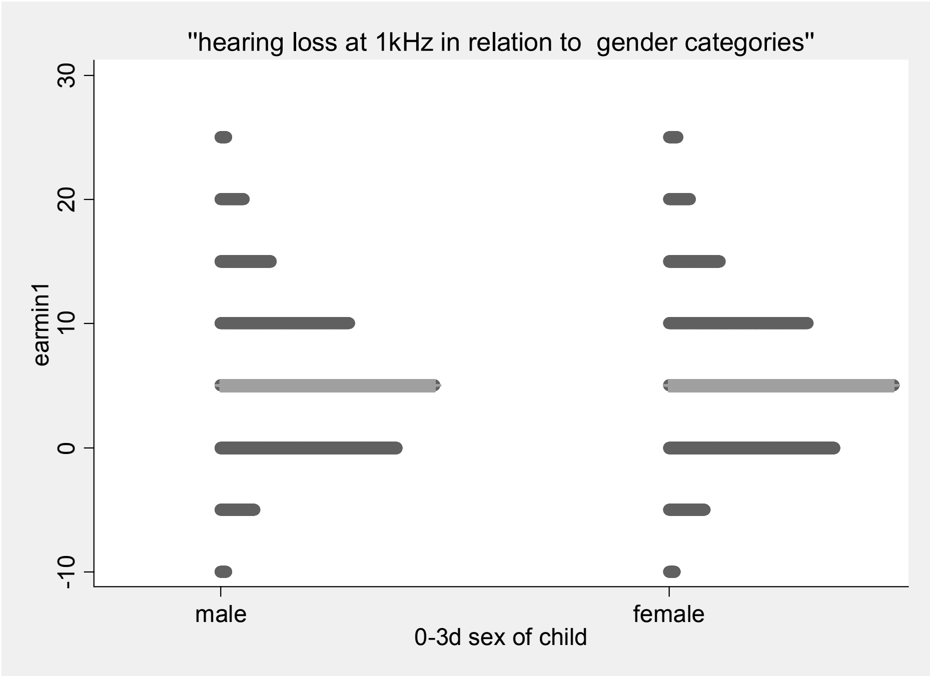
c) 1 kHz by noise at work categories



Skewness

None	<1 year	1-5 years	>5 years
1.32	**1.21**	**1.18**	**1.54**

d)1 kHz by gender categories



Skewness

Male	**female**
0.52	**0.52**

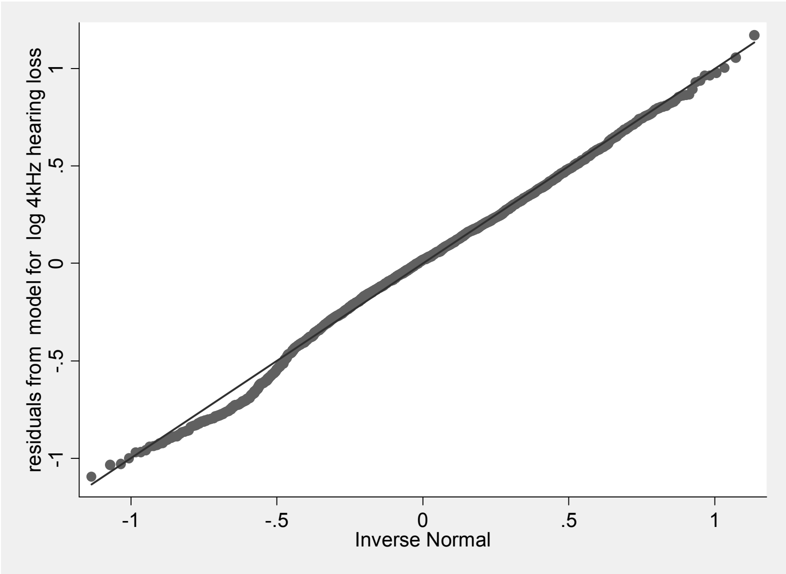
We address only criterion (1) here. Table 13 shows the skewness and kurtosis of the residuals for 4 kHz and 1 kHz for models both with and without adjustment for childhood hearing loss. Skewness (the crucial component) is substantially reduced as is kurtosis though this remains quite high. These reductions are larger at 4 kHz than at 1 kHz. This is also illustrated in Figure 4 which shows p-p plots for 4 kHz for both raw and logged outcome measures.

Table 13: 1 and 4 kHz adult HTLs; raw and logged model (without, then with adjustment for childhood hearing loss)

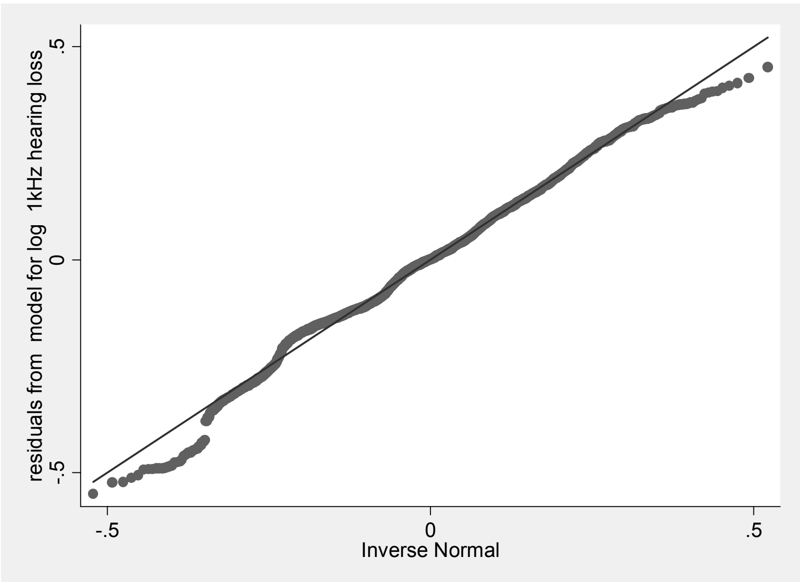
Dependent variable		Raw		Log	
frequency		Skewness	Kurtosis	Skewness	Kurtosis
1 kHz	Without childhood hearing loss adjustment	0.49	3.44	-0.02	3.30
	With childhood hearing loss adjustment	0.34	3.32	-0.13	3.31
4 kHz	Without childhood hearing loss adjustment	1.16	5.89	-0.12	3.18
	With childhood hearing loss adjustment	0.95	5.26	-0.22	3.24

Figure 4: Normal Probability plots for residuals for 4 kHz and 1 kHz , both with log transformed and raw data (change from 1 to 2 kHz) – with childhood hearing loss adjustments

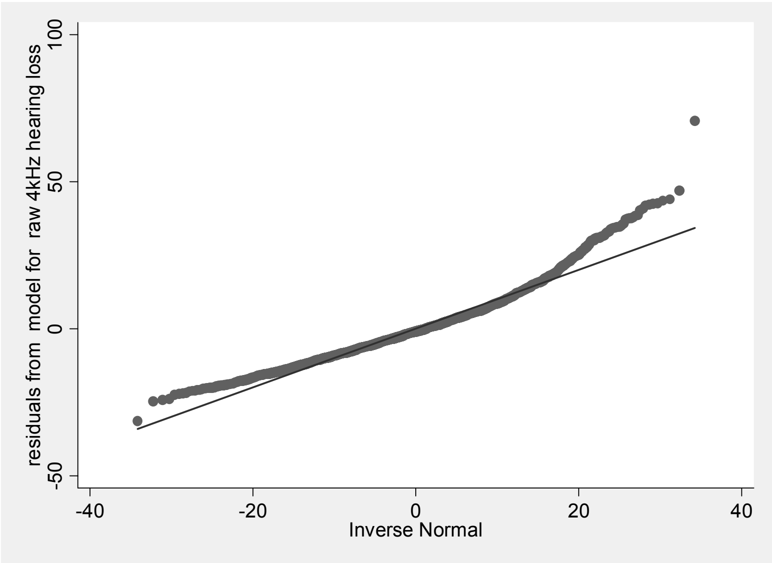
a) log hearing loss, 4 kHz



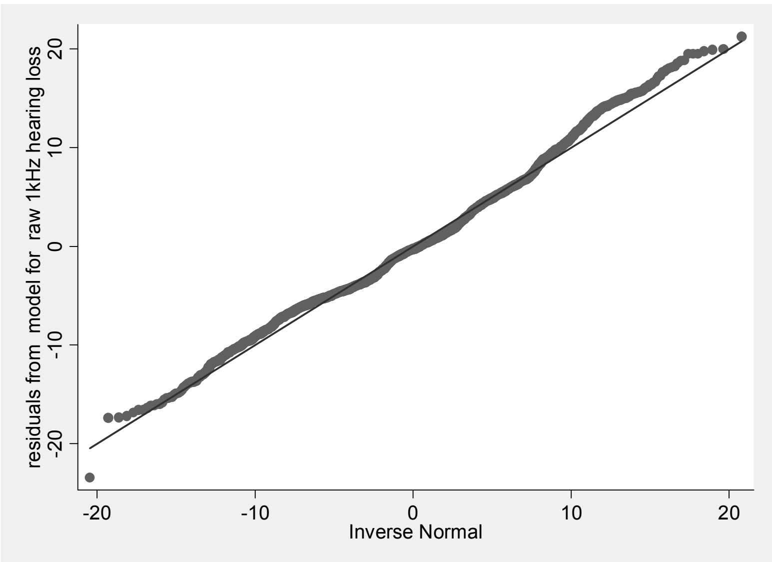
b) log hearing loss, 1 kHz



c) raw hearing loss, 4 kHz



d) raw hearing loss, 1 kHz



Note that an alternative modelling framework, the three parameter log normal distribution model is described by Bowater et al (1996) in a hearing loss context. A recent application to models of gene expression is Konishi (2004). These models simultaneously estimate values of α μ σ^2. representing respectively the intercept, mean and variance and allow these to vary in relation to other explanatory variables (e.g. age, sex, noise). These models therefore extend our log models to allow variability in these estimated parameters within the sample. These models are not currently implemented in the software (STATA 9) used. However two key components of the models shown here, the logging of the dependent variable and the estimation from the data of an intercept term (□) are within the spirit of this work. Adjusting for childhood hearing loss reduces but does not eliminate the need to log the dependent variable for 4 kHz outcomes.

The corollary to the use of the log model is that, though additive in the logs, when transformed back to raw scores the effects become multiplicative. Thus the effect of, say, higher noise, in an additive log model, is to multiply the estimated hearing threshold level by a constant factor. For simplicity and to conform with other research results these are transformed into the additive effect on the raw scale (with correspondingly estimated standard error). When childhood hearing loss measures are standardised this corresponds to a person with zero values on all explanatory variables, and so with mean hearing loss at each childhood age within all explanatory variable combinations.

5. Multiple imputation: a comparison of alternative methods

The extent of missingness varies for the different variables used as explanatory variables in our models as shown in Table 14.

Table 14: Number and Proportion missing by explanatory variable (given valid hearing loss at age 46 years at 4 kHz) (8899 cases)

Childhood HTL (all frequencies)	Valid	Missing	Missing (proportion)
7	6193	2706	0.304
11	6972	1927	0.217
16	6160	2739	0.307
Current social class (not other/unknown)	8508	391	0.044
Social class at birth (not other/unknown)	8570	329	0.037
Noise at work	8204	695	0.078
Current smoking	8610	289	0.032
Current drinking	8256	643	0.072

Hearing thresholds in childhood are seen to contain the highest proportions of missing data.

Following recommendations of Carpenter and Kenward (2007) a trawl of the variables in NCDS which were considered to be substantively related to hearing loss (for example as either a predictor, correlate or consequence) at each of the ages 7,11, and 16 years was undertaken in conjunction with audiology professionals (GS, PS). Those variables judged to be so related were selected at each age and examined in relation to the hearing loss measure (4 kHz) at the same age. These are listed in Appendix 2.

These were then analysed in relation to the childhood hearing thresholds and those which

- showed statistically significant relationship to hearing loss
- had the expected sign for the univariate relation to hearing loss

were selected. Finally a multiple regression with backward elimination (p=0.05) gave the variables at each age listed in Appendix 3. This gives details of recodings used, percentage in the positive pole and t values (both when included separately and simultaneously). Explanation at 7,11, and 16 years was 6.6, 7.7, and 2.7% respectively.

Three imputation models were then created (basic, Intermediate, full). This description refers to the 4 kHz HTL – similar imputations were carried out using the 2 kHz HTL. For these models social class (other/not known), both current and at birth, were defined as missing.

Basic
All variables (Childhood HTLs, Noise at work, gender Social class, drinking, smoking) were imputed using one other variable only. HTLs are all imputed using other information on HTLs only.

The following imputations are used:

HTL at each age was imputed using an equivalent HTL at lower age, the 7 year HTL being imputed using the 16 year HTL. All polynomial terms in the base frequency at any age were imputed using the polynomial term of next lower degree at that age. All polynomial contrasts at a given age were imputed using the polynomial contrasts at that age of next lowest degree.

Noise at work and gender were imputed using the linear term in the base frequency at age 16. Social class at birth and current were imputed using noise at work. Drinking, smoking were imputed using current social class.

Intermediate
All variables (Childhood HTLs, Noise at work, gender Social class, drinking, smoking) were imputed using other variables most closely associated with regard for patterns of missing values.

The following imputations are used:

16 year HTL was imputed using 7 and 11 year HTL together with explanatory variables at this age (see Appendix 3, Tables 1-3). 11 year HTL was imputed using 7 and 16 year HTL measures together with explanatory variables at this age (see Appendix 3, Tables 1-3). 7 year HTL was imputed using 11 and 16 year HTL measures together with explanatory variables at this age (see Appendix 3, Tables 1-3). All polynomial terms in the base frequency at any age were imputed using polynomial terms of lower degree at that age. All polynomial contrasts at a given age were imputed using lower degree polynomial contrasts at that age and all terms in base frequency in that age. Noise at work was imputed using the linear term in the base frequency at each age together with sex and social class at birth. Social class at birth was imputed using the linear term in the base frequency at each age together with sex, current social class and noise. Current social class was imputed using the linear term in the base frequency at each age, sex, social class at birth, and noise. Drinking and smoking were imputed using each other and sex, noise and current social class.

Full
Each variable was imputed using each other variable in the model.

An assessment of the effect of number of cycles (3,5,10) was made on the intermediate imputation (only). For all other imputations 5 cycles were used. When categorical variables (e.g. social class, noise, smoking , drinking) were used to impute other variables these are replaced by their constituent dummy variables using the commands ‘passive’ and ‘substitute’.

In addition a bootstrapped imputation with three cycles was run which does not require multivariate normality of the distribution of the regression coefficients. The full imputation model would not run with message “error running uvis” .

Results are shown in Table 15 and 16. For estimates of noise at work, of specific substantive interest (Ecob et al, 2008), little difference is found between the number of cycles used in the intermediate imputation methods or between the bootstrap and other methods (Table 15). The intermediate estimates can therefore be considered robust to non-normality of the distribution of regression coefficients in this case. In contrast all of these differed substantially (see Table 16) from the complete case analysis which, in relation to the imputed analysis, overestimates the coefficient of ‘noise less than 1 year’ in relation to ‘no noise’ and underestimates the coefficient of ‘noise greater than 5 years’ in relation to ‘no noise’.

However the different MI options do differ in the coefficients for the childhood HTL frequencies, especially at 16 years. The basic multiple imputation estimates for the effects of noise are similar to the intermediate though the coefficients of the relationship to childhood hearing loss differ.

Table 15: Comparison of alternative multiple imputation methods (with complete case) – for noise at work and for childhood HTL (base frequency)

	Multiple Imputation - intermediate				**Complete Case**
	3 cycles	5 cycles	10 cycles	3 cycles-bootstrap	
Noise at work (versus no noise at work)					
<1 year	0.032 (0.013)	0.032 (0.013)	0.031 (0.013)	0.031 (0.013)	0.057 (0.019)
=1 and <5 years	0.081 (0.014)	0.080 (0.014)	0.082 (0.014)	0.082 (0.014)	0.072 (0.021)
> 5 years	0.115 (0.012)	0.115 (0.012)	0.115 (0.012)	0.115 (0.012)	0.085 (0.018)
Childhood HTL (base frequency)					
16 years	0.257 (0.022)	0.249 (0.024)	0.253 (0.025)	0.261 (0.023)	0.307 (0.030)
11 years	0.069 (0.010)	0.067 (0.010)	0.069 (0.011)	0.065 (0.011)	0.058 (0.014)
7 years	0.039 (0.010)	0.046 (0.011)	0.046 (0.011)	0.045 (0.010)	0.051 (0.012)

Table 16: Comparison of numbers of cycles (and bootstrap) for intermediate multiple imputation methods) – for noise at work and for childhood HTL (base frequency)

	Multiple Imputation - Basic	Multiple Imputation- intermediate	Complete case
Noise at work (versus no noise at work)			
<1 year	0.032 (0.013)	0.032 (0.013)	0.057 (0.019)
1 and <5 years	0.080 (0.014)	0.080 (0.014)	0.072 (0.021)
> 5 years	0.115 (0.012)	0.115 (0.012)	0.085 (0.018)
Childhood HTL (base frequency)			
16 years	0.239 (0.024))	0.249(0.024)	0.307 (0.030)
11 years	0.056 (0.010)	0.067 (0.010)	0.058 (0.014)
7 years	0.048 (0.010)	0.046 (0.011)	0.051 (0.012)

6. Alternative models for hearing loss over time.

All models considered here are ‘conditional’ models (Plewis, 1985) in which the outcome measure is regressed on a number of explanatory variables including childhood hearing loss.

A further group of models are mixed-effects models, applied to longitudinal studies of hearing by Larry Brant and co-workers in Baltimore, USA (Pearson et al 1995; Morrell and Brandt, 1991). Here the path of each individual over time or with age is modelled in relation to explanatory variables which are either constant over time (sex) or time varying (hypertension). Mixed-effects models are described in Laird & Ware (1982) and can be estimated within a range of software (for a review see http://www.cmm.bristol.ac.uk/learning-training/multilevel-m-software/index.shtml). Alternative, growth mixture models (GMM) (Muthen & Muthen, 2005) allow for data reduction into a small number of latent trajectories, the number determined by the data. These are described in terms of their variation over time and of their composition in terms of relationships to other explanatory variables. GMM models (in contrast to latent trajectory models, or Latent Class Growth Analysis (LCGA)) allow Individual trajectories to vary in relation to that of the group to which allocated (Muthen, 2006),

Advantages of Mixed-effects Models are that the variability of individual change within groupings of all other explanatory variables is allowed for and accurately modelled. Such models are therefore well suited to studies in which the prime interest is in the form of changes over time, as in the examination of the patterns of hearing loss with ageing.

However for the NCDS data, with the concentration of data at childhood ages and one later (current wave) observation, the precise nature of change over time cannot be assessed. This limits the potential for mixed-effects models.

The incorporation of the wealth of data on a range of frequencies though audiograms is possible for mixed-effects models, as for other models; though these require repeat observations, they do not necessarily have to be all at the same frequency (we have shown that a 2 kHz childhood measure predicts the 1 kHz adult measure better than a 1 kHz childhood measure and so the 2 kHz measure in childhood may be the most appropriate prior measure). Thus, hearing loss at other frequencies, included in the models used here in terms of contrasts with the base frequency, could be similarly included in mixed-effects models as further explanatory variables.

7. References

Bowater, R.J., Copas, J.B., Machado, O.A. & Davis, A. 1996. Hearing impairment and the log-normal distribution. *Applied Statistics*, 45, 203-217.

British Society of Audiology. 1981. Recommended procedures for pure tone audiometry using a manually operated instrument. *British Journal of Audiology,* 15,213-6.

Carpenter, J & Kenward, M. (2007) Guidelines for handling missing data in social science research. http://www.lshtm.ac.uk/msu/missingdata/guidelines.pdf

de Stavola, B.L., Nitsch, D., dos Santos Silva, I., McCormack, V. et al (2006). Statistical issues in life course epidemiology. *American Journal of Epidemiology,* 163, 1-13.

Ecob, R., Sutton, G., Rudnicka, A., Smith. P., Power, C., Strachan, D., Davis A (2008). Is the relation of social class to change in Hearing Threshold Levels from childhood to middle age explained by noise, smoking and drinking behaviour**?** *(International Journal of Audiology – in press)*

Fogelman K. 1983. Growing up in Great Britain – papers from the National Child Development Study*.* Macmillan for National Children's Bureau, London.

Konishi, T (2004) Three parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment. BMC *Bioinformatics,5:5 (see also correction BMC Bioinformatics,5:82)*

Laird, N.M & Ware, J.K. (1982) Random effects models for longitudinal data. *Biometrics* 38, 963-974.

Morrell, C.H. & Brant, L.J (1991) Modeling hearing thresholds in the elderly. *Statistics in Medicine*, 10, 1453-1464.

Muthen, B. & Muthen, L.K. (2005) Mplus: Statistical analysis with latent variables; Users Guide. Version 3. 4th Printing.

Muthen, B. (2006). The potential of growth mixture modeling. Commentary*. Infant and Child Development*, 15, 623-625.

Pearson, J.D., Morrell, C.H., Gordon-Salant, S., Brant, L.J., Metter, E.J. et al (1995) Gender differences in a longitudinal study of age-associated hearing loss. *J. Acoustical Society Of America.* 98 (2) 1196-1205.

Plewis, I. (1985). Analysing Change; Measurement and Explanation using Longitudinal Data. London, Wiley.

Power, C. Elliott, J. 2006. Cohort Profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology,* 35, 34-41.

8. Appendices

Appendix 1: Plan of variables used in analyses to date (Note D= derived variable)

Area/domain	**Variable**	**NCDS code**	**Comments**	**Actions (analyses in which to use – (all=all analyses)/codings**
1. Outcome (hearing)	Tested ears at 1, 4 kHz (First then second)	AUDEARA (first ear tested, 1 kHz) AUDEARA3(second ear tested, 1 kHz) AUDEARC (first ear tested, 4 kHz) AUDEARC3 (second ear tested, 4 kHz)	Transformed for analysis as in Appendix 2 Construct hearing in better ear then log as shown in working paper	
2.Confounding/ nuisance variables: Testing environment	Noise at test (current wave)	AUDNOISE	Background noise Acceptable(1)/ Distracting (2)	
	Tester – nurse (current wave)	NURSENO	Range 101-222	Generate combined variable an=nurseno+120*audsn (all)
	instrument (current wave)	AUDSN	Range 1-89	" "
	Audiometer working? (current wave)	AUDCHK	Do not use at all	Audiometer working (1)/ not working (2)
	Measurements complete? (current wave)	AUDOC	Screen out all options apart from 'all measurements completed' (AUDOC =1) (all)	Completed (1) Not completed (2-4)
3.Confounding/ nuisance variables: other				

Area/domain	Variable	NCDS code	Comments	Actions (analyses in which to use – (all=all analyses)/codings
	Wax in ear ; 11 years (D)	WAX11	From auroscope; see \extravars.doc 20/09/05	(1- yes,0-no)
	Recurrent throat/ear infection (D)	EARILL11	; see \extravars.doc 20/0/905	(1- yes,0-no)
	Accidents/injuries to ear/head/neck	ACC	From ACCINJE1-36 (at age 42) ; see \extravars.doc 20/09/05	(1- yes,0-no)
	Ever had ear operation	EAROPX	From EAROP (at age 42) ; see \extravars.doc 20/09/05	(2- yes,1-no)
	Type of hearing or ear problem (age 42)	EARPROB1		Exclude cholesteatoma (4), repeated ear infections/discharge (1)
	Type of hearing or ear problem (age 42)	EARPROB2		Exclude cholesteatoma (5), repeated ear infections/discharge (2)
	Proxy for conductive hearing loss (7 years)	OTM	For constituent variables, see Appendix 2	recode n349 (2=1)(else=0),gen(otm)

Area/domain	Variable	NCDS code	Comments
4. Childhood; audiological control	By frequency within age (averaged over ears)	X7eard025 (D) X7eard050 (D) X7eard100 (D) X7eard200 (D) X7eard400(D) X7eard800 (D)	6 frequencies at three ages (7,11,16) At 7 years n558 –n563, n564-n569 for right then left ear (frequencies 250 Hz – 8kHz within ear)

Area/domain	Variable	NCDS code	Comments
		X11eard025 (D) X11eard050 (D) X11eard100 (D) X11eard200 (D) X11eard400(D) X11eard800 (D) X16eard025 (D) X16eard050 (D) X16eard100 (D) X16eard200 (D) X16eard400(D) X16eard800 (D)	At 11 years n1635 (3) 50, n1653 (3) 68 for right then left ear (frequencies 250 Hz – 8kHz within ear) At 16 years n2060 (2) 70, N2072 (2)82 for right then left ear (frequencies 250 Hz – 8kHz within ear) (note n1635 (3) 50 denotes n1635, n1638,n141, n1644, n1647, n1650)
5. Childhood; social			
	Social class – at birth	SC0_1	
	Region – at birth	REG0_58	
	Gender	N622	(female=2, male=1)
6. Childhood; biological/ Psychological (not audio)			
Birthweight/gestation	Birthweight (given gestation, percentiles)	N516	
	Gestation (days)	N497	
Possible illness, causes of deafness	Maternal rubella	N1837	
	Meningitis (see Peckham BMJ 1986)	N1294	

Area/domain	Variable	NCDS code	Comments
7. Adult; social			
	Noise exposure (adult)	NOISWORK	
	Social class – 42 years	Sc42_1	
	Job/industry		Data available potentially at all waves (ages 23,33,42)
8. Adult; hearing related			
	Do you have difficulty having a conversation with several people in a group?	HEARGRP	

Appendix 2: Candidate variables for inclusion in multiple imputations of childhood hearing loss (All variables are measured at the same age).

Age 7

Questions relating to health (health problems, service related, 'handicap') which may relate to audiometric hearing loss data

Immunisations/conditions

N212	Immunisation against diphtheria
N213	Immunisation against polio
N214	Immunisation against smallpox
N219	Mumps
N215	Measles
N217	Whooping cough
N216	German measles
N218	Chicken pox
N220	Scarlet fever
N221	Glandular fever/TB

Contact with services

N241	Seen at specialist hearing clinic
N242	Seen audiology/outpatient clinic
N250	Hosp admission – road accidents
N251	Hosp admission –home accidents
N252	Hosp admission – other accidents
N253	Hosp admission – illness/tests
N254	Hosp admission –other reasons
N249	Hosp admission –other operation

Other

N224	Speech therapy
N225	Stammer/stutter ever
N385	Stammer present
N226	Other speech difficulties
N349	Present/past otitis media
N352	Deformity external ear
N386	Assessment of speech intelligibility
N389	Doctor's assessment of hearing
N1827-9(2)	Handicaps – audio loss
N1827-9(4)	Handicaps - poor speech
N471(1)	Health factors – poor respiratory
N472-6(1)	Health factors – infected ears

Age 11

Questions relating to health (health problems, service related, 'handicap') which may relate to audiometric hearing loss data

Variables with 'hearing' mentioned – either in heading or response options

N1263	child always good hearing in both ears
N1266	age poor hearing started
N1267	hearing aid worn
N1391	seen specialist for hearing
N1392	treatment nose/palate/ears
N1477-9(4)	reason special education (partial hearing)
N1482	ever child impaired hearing
N1483	ever non-transitory ENT problems
N1552	ever hearing aid
N1553	would hearing loss affect schooling
N1089-92 (4)	BSAG (poor hearing)

Contact with services

N1282-3	child accident – unconscious
N1268	ever in hospital after accident
N1397	medical treatment – any operation
N1398	in hospital overnight
N1400	no. times admitted to hospital
N1399	hospital outpatient

Other

N1268-9	ever speech defect
N1270	speech therapy
N1284	swallowed poison
N1285	nearly drowned
N1267-93	illnesses (incl. mumps, measles, rubella)
N1301	time off school due to ill health last year
N1321-25	reason for this (infections, accident/injury)
N1332	prescribed medicines (liquid, tablets, inhaler, injections)
N1349	non-wax ear discharge
N1519-22	examine left (right) ear with auroscope (normal, inflamed, scarred, wax etc)
N1548-51	number of words misheard –right/left ear

Age 16

Questions relating to health (health problems, service related, ‘handicap’) which may relate to audiometric hearing loss data

Variables with ‘hearing’ mentioned – either in heading or response options

N2416	handicap for which require help (hearing)
N2602-8	hospital inpatient (hearing disorder)
N2663-67	child disability (hearing defect)
N1893-5	category child handicap (partial hearing)
N1910	reason hospital outpatient(hearing)
N1939	hearing aid prescribed
N1944	doctor’s assessment hearing
N2036	hearing defect
N2332	poor hearing – teacher’s view

Contact with services

N2554-9	reason absent from school (accident/injury)
N2566	accidents to child necessitating. hospital attendance
N2571-2580	injury 1st (2nd, 3rd) most recent accident
N2589	child hospitalised overnight (accident/other)
N2596-7	other operations (accidents/other)
N2598-2601	hospital admissions (upper respiratory)
N2601-8	hospital outpatient (hearing disorder)
N1904	reason hospital admission (injury to head)

Other

N2507	stammer/stutter
N2508	other speech difficulties
N2613	speech therapy in past 12 month
N2663-7	child disability (hearing defect)
N1893-5	category child’s handicap (deaf, partially hearing)
N1900	immunisation/vaccination (rubella)
N1907	reason accident/casualty (fracture –skull/laceration – head etc)
N1947	stammer/stutter
N1990	upper abnormal respiratory tract
N2037	speech defect

Appendix 3. Variables used in Multiple Imputation of childhood hearing loss

Table A1: Variables at 7 years related to hearing loss at 7 years (in order of t value –simultaneous)

Col No.	Variable	code	Variable name	Recoding (positive category)	% positive pole	T value (when included separately	T value when included simultaneously
1	Doctors assessment of hearing	N389	DAH	Some impairment or worse	3.67	22.84	18.82
2	Handicap	N1827	SDEAF	Severe deafness	0.20	16.82	13.64
3	Otitis media (past/pres)	N349	OTM	yes	5.42	9.73	5.50
4	Assess speech intelligibly	N386	ASI	Poor intelligible/ unintelligible	1.03	7.85	4.67
5	Deformity in external ear	N352	DEE	Yes	0.98	3.36	3.35
6	Hospital admission (any)	N249/54	HOSP_ADM	Yes	27.10	2.64	2.55

TableA2: Variables at 11 years related to hearing loss at 11 years (in order of t value –simultaneous)

Col no.	Variable	code	Variable name	Recoding (positive category)	% positive pole	T value when included separately	T value when included simultaneously
1	Hearing loss affects schooling	N1553	HLAS	Loss, no effect; loss, some effect; can't say	5.58	26.36	19.82
2	Age at which poor hearing first noted	N1266	APHN	Any age	6.59	17.17	5.48
3	Has child ever had impaired hearing?	N1482	CEIH	Yes congenital, acquired temp; acquired perm; cause uncertain	7.71	13.38	5.44
4	Auroscope exam	N1520/1	EEXA	Inflamed/scarred/obscured by wax/abnormal-other	14.44	11.23	5.39
5	Bad hearing (parental?)	N1263	BHOE11	Bad in one or both ears/ok now but not in past	3.25	17.22	3.28
6	Last year non-wax discharge	N1349	PYNW	yes	1.71	8.23	3.16
7	Has hearing ever been worn?	N1297	HAEW	yes	1.93	9.27	2.93

Table A3: Variables at 16 years related to hearing loss at 16 years (in order of t value –simultaneous)

Col no.	Variable	code	Variabl e name	Recoding (positive category)	% positive pole	T value when included separately	T value when included simultan eously
1	Child poor hearing (teacher)	N2332	PHTE	yes	1.51	15.32	6.38
2	Hearing defect	N2036	HEDE	yes	3.08	25.19	5.71
3	Doctor's assessment of hearing	N1944	DAHE	Hearing loss, no effect; interference	2.64	24.18	4.06
4	Has hearing aid been prescribed	N1939	HAID	yes	0.28	11.59	3.61
5	Nature of child's disability	N2663	NCHD	Hearing defect	0.31	17.99	2.93
6	Hospital outpatient – diag hearing disorder	N2603	HOCD	yes	0.18	10.54	2.92
7	Any other speech difficulties (teacher)	N2508	OTSD	yes	1.62	3.86	2.29
8	Does child stammer or stutter?	N2507	STST	Yes, mildly/severely	1.17	2.21	2.00

Table A4: Number of variables selected at each stage

Age	Selected by expert	Selected into multiple regression	Selection by multiple regression	% variance explained (adjusted)
7	13	12	6	6.60
11	29	25	7	7.73
16	21	18	8	2.72

Further notes

Selected proxies for Otitis Media in childhood

Age (yrs)	Variable	code	modes
7	N349	2 yes	questionnaire
11	N1519/N1522	2 Inflamed/ 5 abnormal other	Exam right/left ear with auroscope
16	None		

Summary of results of relevant regressions:

Glue ear relates to hearing loss at concurrent age (4 kHz)
Glue ear at 11 years but not at 7 years relates to loss in hearing from 11 to 16 (4 kHz)
Glue ear at 7 years does not relate to loss in hearing from 7 years to 11 years

Centre for Longitudinal Studies
Institute of Education
20 Bedford Way
London WC1H 0AL
Tel: 020 7612 6860
Fax: 020 7612 6880
Email cls@ioe.ac.uk
Web http://www.cls.ioe.ac.uk