Using the Linked Scottish Health Survey to Predict Hospitalisation & Death

An analysis of the link between behavioural, biological and social risk factors and subsequent hospital admission and death in Scotland.

- Technical Report -

Analysis & Authorship by:

Richard Lawder[1] (Information Services NHS NSS)

Andrew Elders (Information Services NHS NSS)

David Clark (Information Services NHS NSS)

Project Group:

Professor Phil Hanlon (University of Glasgow)

Professor Matt Sutton (University of Aberdeen)

David Walsh (NHS Health Scotland)

Richard Lawder, Andrew Elders, David Clark, Bruce Whyte (Information Services NHS NSS)

This project was resourced and supported financially by NHS Health Scotland & Information Services NHS NSS

[1] Contact for main correspondence – richard.lawder@isd.csa.scot.nhs.uk

Acknowledgements

The project team wish to thank: NHS Health Scotland for financial support and contribution of staff members; ISD Scotland for staff members and substantial logistic support; The Chief Scientist Office, and in particular Dr. Peter Craig, for enabling access to the source dataset and for continuing financial support which allows ongoing linkage of this important resource.

Table of Contents

APPENDICES

1 Introduction

In 2004 a record linkage exercise was undertaken by the Information Services Division (ISD) of NHSScotland to link both the 1995 and 1998 Scottish Health Survey data to the linked Scottish hospital admission and mortality database. This is the first time that such a linkage has been undertaken on a national basis: it therefore provides an ideal opportunity to add to our understanding of the relationship between the broad range of factors measured in the Health Survey and subsequent hospital utilisation and, moreover, it provides a practical application in allowing us to plan more effectively for future health service provision in the light of rapidly changing lifestyle factors.

A project group[2] was set up consisting of researchers and analysts from NHS Health Scotland, Information Services (NHS NSS), University of Glasgow and University of Aberdeen, to take advantage of the new ability to link these lifestyle and hospital utilisation data across Scotland on a prospective (and retrospective) basis, to understand this relationship and to examine it within the broader context of the many factors that influence hospital usage.

The following report provides a detailed description of the linked survey data and how it was constructed and discusses specific issues surrounding the use of the data (e.g. survey design elements, missing values and representativeness of the sample) and how these have been addressed. Other issues covered in this report are migration, the definition of a "serious hospital admission" (used in regression analysis) and the modelling approach adopted (Cox-proportional hazard) for follow-up analysis.

This technical report will be updated in due course to include the linkage results of the 2003 Scottish Health Survey and the annual updates to the hospitalisation and mortality files.

The main report, which accompanies this technical paper (also available on the ScotPHO website - www.scotpho.org.uk/linkedshesreport), describes in detail the results of single-predictor[3] and multivariate regression analyses of the association between behavioural, biological, social and health status risk factors and outcome (hospital admission/mortality). All analysis is based on the 1998 Survey Respondents (Age-group: 16-74), Follow-up period - April 1998 to March 2004.

---

[2] Professor Phil Hanlon (University of Glasgow)
Professor Matt Sutton (University of Aberdeen)
David Walsh (NHS Health Scotland)
Bruce Whyte, Richard Lawder, David Clark, Andrew Elders (Information Services NHS NSS)

[3] Age & Sex standardised association models

2 Overview of Linkage of 1995 & 1998 Scottish Health Survey Records to Scottish Morbidity Records

This chapter summarises the process of linking Scottish Health Survey records with a linked file of hospital discharges, cancer registrations and death records to create the SHeS/SMR linked dataset.

Consent was granted for 15,668 Scottish Health Survey (SHS) responses including person-identifiable information to be made available to ISD. 7,363 responses were from the 1995 survey and 8,305 from the 1998 survey. Internal linkage of this dataset identified 23 repeat respondents i.e. participants who were surveyed in 1995 and again in 1998. As a result there is a combined total of 15,645 respondents.

The linkage of the SHS data to the September 2004 version of ISD's linked SMR01 'catalogue'[4] successfully linked 73% of the survey records i.e. 11,396 respondents (or 11,417 responses as this included 21 repeat respondents). An extract was taken, for each respondent, of details of SMR01 hospital admissions, SMR04 psychiatric admissions, and GRO death records up to 31 March 2004 and cancer registrations up to 31 December 2001, amounting to a total of 58,913 records. Each record comprises a standard set of dates, clinical information (including all diagnoses) and deprivation scores[5], with a total of 30 variables per record.

From the group of 11,396 respondents with a link to the SMR01 catalog, Table 1 below summarises hospital admissions, psychiatric admissions, cancer registrations and deaths.

---

[4] The SMR01 catalogue is a linked file that as well as SMR01 hospital discharge records, includes SMR04 psychiatric discharge records, cancer registrations and death records, and covers the period 1981 to the present day. This file only contains records from Scottish sources e.g. hospital admissions to Scottish hospitals, deaths registered in Scotland

[5] As defined by the Carstairs & Morris measure. Carstairs deprivation scores were originally developed by Vera Carstairs and Russell Morris. See the MRC Social and Public Health Sciences Unit website for more details - http://www.msoc-mrc.gla.ac.uk/Publications/pub/Carstairs_MAIN.html

Table 1: Summary of hospital admissions, psychiatric admissions, cancer registrations and deaths

Respondents with at least one hospital episode (acute or psychiatric)[6]	72.4%	11,325
Respondents with at least one psychiatric hospital admission	3.2%	502
Respondents with at least one cancer registration[7]	6.4%	1,003
Deaths	4.7%	743

These figures do not take migration into account. The following chapter discusses the potential impact of migration.

2.1 Age/sex distribution of the 15,645 respondents

Table 2 below reflects the age/sex distribution of the 15,645 survey respondents.

Table 2: Combined 1995&1998 age/sex distributions

		Sex of respondent from household grid		Total
		Male	Female	
Age band at time of interview	16-34	2,329 (15%)	2,988 (19%)	5,317 (34%)
	35-54	2,839 (18%)	3,415 (22%)	6,254 (40%)
	55+	1,789 (11%)	2,285 (15%)	4,074 (26%)
Total		6,957 (44%)	8,688 (56%)	15,645 (100%)

2.2 Summary of SMR output

Table 3 below shows, from the 58,913 morbidity & mortality records, the percentage and total number of survey respondents represented in each of the categories:

Table 3: SMR Summary Breakdown

Acute hospital admissions (SMR01)	94.3%	55,541
Psychiatric admissions (SMR04)	2.6%	1,523
Cancer registrations	1.9%	1,106
Death records	1.3%	743
Hospital admissions or cancer registrations before respondent participated in the survey	54.8%	32,259
Hospital admissions, cancer registrations or deaths after respondent was interviewed[8]	45.4%	26,735

[6] Note that the number shown in the table is 71 less than the total figure quoted earlier. The other 71 are made up of 26 orphan deaths i.e. (deaths with no accompanying cancer registrations or hospital episodes) and 45 cases where the respondent had at least one cancer registration but no hospital admissions.

[7] 269 of these respondents had died by 31 March 2004.

[8] This includes 18 records that relate to repeat respondents, which occurred between their 1995 and 1998 interviews.

In terms of disease-specific morbidity, 767 respondents (4.9%) experienced at least one hospital admission with a principal diagnosis of coronary heart disease (CHD). 253 respondents (1.6%) were admitted with a diagnosis of cerebrovascular disease.

2.3 Summary of death records

Table 4 below, shows the age/sex breakdown of the 743 survey respondent death records.

Table 4: Age/sex distribution of respondents who have died

	Sex of respondent		Total
Age band at time of interview	Male	Female	
16-34	18 (2.4%)	16 (2.2%)	34 (4.6%)
35-54	98 (13.2%)	76 (10.2%)	174 (23.4%)
55+	306 (41.2%)	229 (30.8%)	535 (72.0%)
Total	422 (56.8%)	321 (43.2%)	743 (100.0%)

From these deaths the five most common causes are as follows[9]:

I21 - Acute myocardial infarction (60 deaths)
C34 - Malignant neoplasm of bronchus and lung (59)
I25 - Chronic ischaemic heart disease (50)
J44 - Other chronic obstructive pulmonary disease (34)
K70 - Alcoholic liver disease (25)

The most common cause of death for each sex and age group can be seen in **Appendix1 – Table 5**.

[9] Causes of death on GRO death records up to 31 December 1999 were recorded according to the World Health Organisation's 9th version of the International Statistical Classification of Diseases (ICD9) and since then causes of death have been recorded as ICD10. In many cases there is no direct mapping between ICD9 and ICD10, so the following table has been restricted to the **545** records with a cause of death recorded as ICD10 (i.e. from 2000 onwards).

2.4 Summary of hospital utilisation

Tables 6 and 6a summarise the number of hospital episodes by age and sex. Just under half of hospital admissions occurred after the survey interview and these numbers are shown in brackets. Age is at the time of the survey. Table's 6b and 6c show the average number of inpatient stays and bed days, per respondent respectively.

Table 6: Total number of hospital episodes[10]

	Males	Females	Total
16-34	5,365 (2,034)	8,202 (3,347)	13,567 (5,381)
35-54	9,164 (4,096)	12,219 (5,030)	21,383 (9,126)
55+	10,717 (5,452)	11,397 (5,431)	22,114 (10,883)
Total	25,246 (11,582)	31,818 (13,808)	57,064 (25,390)

Table 6a: Average number of hospital episodes per respondent[10]

	Males	Females	Total
16-34	2.3 (0.9)	2.7 (1.1)	2.6 (1.0)
35-54	3.2 (1.4)	3.6 (1.5)	3.4 (1.5)
55+	6.0 (3.0)	5.0 (2.4)	5.4 (2.7)
Total	3.6 (1.7)	3.7 (1.6)	3.6 (1.6)

Table 6b: Average number of continuous inpatient stays[10]

	Both sexes	Male	Female
All ages	3.3 (1.4)	3.2 (1.4)	3.3 (1.4)
16-34	2.4 (0.9)	2.1 (0.8)	2.6 (1.0)
35-54	3.1 (1.3)	2.8 (1.2)	3.3 (1.3)
55+	4.7 (2.2)	5.1 (2.5)	4.4 (2.0)

Table 6c: Average number of inpatient bed days per respondent[10]

	Both sexes	Male	Female
All ages	12.5 (5.3)	12.8 (5.7)	12.3 (5.0)
16-34	6.3 (1.9)	6.4 (1.7)	6.2 (2.0)
35-54	10.2 (4.0)	9.6 (3.9)	10.7 (4.1)
55+	24.1 (11.9)	26 (13.8)	22.6 (10.4)

Tables 7 and 7a, shown in **Appendix 2,** summarise the most common diagnoses recorded on hospitalisation records by age and sex. The figures shown relate only to principal diagnoses that have been coded using ICD10. Only admissions since the date of the survey interview have been included.

[10] Figures in brackets refer to events since the survey.

3 Emigration - Linkage of Scottish Health Survey data to Community Health Index (CHI)

To identify the extent of emigration in both the 1995 & 1998 survey data sets, respondents were linked to the monthly CHI[11] download in March 2005.

Of the 15,668 SHeS subjects, 15,446 (98.6%) linked to CHI[12]. Of the 222 people who didn't link to CHI, 74 of these (33.3%) had already linked to SMR01.

The CHI extract contains current and any historic records (generated when a subject transfers out of a consortium[13]). This was converted into a single record per subject and the "CHI Status" field was taken from the most current/recent CHI record in the extract. A summary of the number of respondents by CHI status and survey year shown in 4 distinct groups can be seen in the Table 8 below.

Table 8: Summarised CHI Status of SHeS Respondents by Year of Survey

	YEAR					
	1995		1998		Total	
CHI status	**n**	**%**	**n**	**%**	**n**	**%**
1 - No link to CHI	140	1.90%	82	1.00%	222	1.40%
2 - Emigrants or untraced	384	5.20%	249	3.00%	633	4.00%
3 - Death	372	5.10%	481	5.80%	853	5.40%
4 - Currently on CHI	6,467	87.80%	7,493	90.20%	13,960	89.10%
Total	**7,363**	**100.00%**	**8,305**	**100.00%**	**15,668**	**100.00%**

It should be noted that those "Currently on CHI" are likely to include people who have emigrated but have not de-registered from their old practice and/or have still to register with a practice elsewhere in the UK. They also may include people who have emigrated for a period of time but then returned to Scotland. Taking those respondents in CHI Status 1 ('No link to CHI') as potentially being emigrants, and adding them to CHI Status 2 (Emigrants or

[11] The Community Health Index (CHI) is a population register, which is used in Scotland for health care purposes. The CHI number uniquely identifies a person on the index.

[12] Of the 15,668 SHeS subjects, 15,072 (96.2%) automatically linked to CHI. After clerical checking of a file of pairs of records which did not achieve the automatic linkage threshold, the latter was reduced and this, in addition to some other manual adjustments, resulted in the total of 15,446 (98.6%) linking. However, it should be noted that the "grey-area" around the threshold is quite wide due to insufficient and poor quality data on some of the SHeS records. A number of false positive links will be an inevitable consequence.

[13] A Consortium is made up of the GP practices covering one of the 8 geographical areas that represent one of the CHI databases.

untraced) we can thus estimate the number of emigrant respondents present in both survey years: **524 (7.1%) in 1995,** and **331 (4.0%) in 1998**.

The issue in relation to these emigrants is whether or not they should be excluded from the analysis. The CHI extract was examined in order to try and determine if there were appropriate start and end dates which could be incorporated into any analyses based around censoring of people lost to follow up. Due to uncertainty in the reliability of these dates, it was felt that if it was decided to utilise CHI data, then **any follow-up analyses using the linked SHeS/SMR1 catalogue should be undertaken only on those respondents who are known to have died or to be currently on CHI** (1995 - 92.9%; 1998 - 96.0%).

Note that the impact of emigration on the modelling analyses is discussed further in Chapter 7 of this report and in Chapter 7, section 7.3.2, of the Main Report.

4 Serious Hospital Admissions based on Healthcare Resource Groups

To enable regression analysis to be undertaken on the specific outcome of a *complicated or serious hospital admission,* the project group had to agree on a definition of such an admission. A 'serious' hospital admission can be defined in terms of relative case complexity, using estimated cost as a proxy measure for complexity. Our approach is to classify an admission as serious when the cost is above average. The seriousness of hospital admissions can be measured by analysing Healthcare Resource Groups (HRGs)[14]. For the purpose of our analysis we class a serious admission if it is at least as serious as an acute myocardial infarction i.e. (HRG weight = 1.1). The reasoning behind this is explained in the following paragraphs.

The HRG severity index provides a weight for each HRG, calculated as the reference cost (discussed further in the subsequent section) divided by the average cost for any admission. Therefore a weight of more than one indicates an above average cost and a potentially serious admission.

The mean estimated cost of post-survey episodes experienced by 1998 SHeS respondents is £1059. An initial analysis (based on the selection of: **an HRG weight greater than or equal to 1** AND the estimated cost being higher than the sample average AND the case being an inpatient episode) defined 28% of the subsequent admissions to be "serious", and indicated that 1499 respondents (18%) had experienced a subsequent serious admission. This can be compared to 44% of respondents experiencing *any* subsequent admission.

Table 9: Proportion of 1998 respondents experiencing subsequent hospital admissions

	Any admission	Serious admission
16-34	34%	6%
35-54	39%	13%
55-64	54%	27%
65-74	66%	42%
All	44%	18%

[14] Healthcare Resource Groups (HRGs) are standard groupings of clinically similar treatments, which use common levels of healthcare resource. They can be considered as 'units of currency' within the health service, allowing for costings across services.

As previously stated, the mean estimated cost of post-survey episodes experienced by respondents in 1998 is £1059. Having looked at conditions classified as serious by the above definition, it was clear that in some cases this was not appropriate: some conditions were not of a serious nature and were simply costly procedures. To try and account for this and assign a more appropriate point of reference for a serious admission, **we increased the serious threshold on the HRG index to 1.1**. In doing so, the proportion of serious hospital admissions drops slightly to 26%, with 1,435 respondents (17%) experiencing such a hospital admission. As an illustrative example, an AMI without complication has a HRG score of 1.1 and this is deemed an appropriate benchmark for the identification of such 'serious' admissions.

4.1 Limitations to this approach

Reference Costs for the NHS in England are published annually by the Department of Health. Costings are based on acute activity at episode level and are dependent on Healthcare Resource Group, type of admission (elective or non-elective), type of patient (inpatient or daycase) and length of stay.

The version of the linked SMR01 catalog that was used for the SHS linkage contains HRG version 3.1. The latest reference costs to be based on HRG version 3.1 was 2002/03; however, there are still outstanding issues relating to the mapping between English costs and Scottish records which means that 2002/03 costs cannot be applied accurately. The most recent appropriate reference costs are therefore from 2001/02. It should be noted that for 2001/02, additional costs are not available for excess lengths of stay.

In summary, the following points should be considered:

- only the cost of acute services are measured
- psychiatric admissions are excluded – this includes SMR04 records and also SMR01 records with a psychiatric diagnosis (these generate an HRG code beginning with a 'T' and their associated costs are not published)
- HRG 3.1 is not the most up-to-date version
- costs are based on an estimated average within HRG - there are some HRGs with a large variability of costs (e.g. C41)

- 2001/02 costs do not necessarily reflect the cost at the time of admission, nor do they necessarily represent current costs

- relative costings between HRGs may be different between Scotland and England but are assumed to be the same

- the cost of delivering the same services is higher overall in Scotland compared with England (recent estimates vary between approximately 6% and 10%)

- additional costs incurred for excess bed days are excluded (i.e. where the length of stay exceeds the nationally set trimpoint)

- HRGs are currently only available for admissions since April 1997. 0.6% of SMR01 records subsequent to the 1998 survey had no HRG assigned, so cost is estimated to be the mean cost within the specialty (depending on admission and patient type)

Despite these limitations, calculating the cost of admissions using HRGs and nationally published reference costs should hopefully be an effective measure of case complexity, and a useful estimate of the level of resources used.

5 Cox's Proportional Hazard Model

'Normal' multiple regression analysis is based around the risk of an outcome/event (e.g. death) at a given time. Cox's proportional hazards regression instead looks at the cumulative risk over time - it 'adds up' the hazards (risks) up to the time of the outcome, and is thus more suitable for studies with a reasonably long follow-up period. The follow-up period in our data set is 6 years and it is for this reason that Cox's proportional hazards regression was used for all the modelling. The hazard is modelled as follows:

$$H(t) = H_0(t) \times \exp(b_1 X_1 + b_2 X_2 + b_3 X_3 + . . . + b_k X_k)$$

where $X_1 \dots X_k$ are the collection of predictor variables (risk factors) and $H_0(t)$ is the baseline hazard at time t, representing the hazard for a respondent with the value 0 for all the predictor variables (risk factors). By dividing both sides of the above equation by $H_0(t)$ and taking logarithms, we obtain:

$$\ln(H(t) / H_0(t)) = b_1 X_1 + b_2 X_2 + b_3 X_3 + . . . + b_k X_k$$

$H(t) / H_0(t)$ represents the hazard ratio. The coefficients $b_1 ... b_k$ are estimated by Cox regression, and are interpreted in a similar manner to that of multiple logistic regression.

The '<u>stphplot</u>' function in STATA was used to provide graphical assessment for assessing violations of the proportional hazards assumptions. stphplot plots – ln(-ln(survival)) curves for each category of a nominal or ordinal covariate versus ln(analysis time). If the plotted lines were reasonably parallel, then the proportional hazard had not been violated. It was therefore appropriate to base the estimate for that variable on a single baseline survivor function.

6 Survey Design

6.1 Weighting

Each survey respondent was assigned a sampling weight. These weights were used in the modelling and have the advantage of controlling for the effects of unequal inclusion probabilities. The weighting ensures unbiased population estimates (as well as reducing the precision of estimates). However it is worth noting a limitation of the weightings used in the models: the sampling weights applied are based on the full 1998 Scottish Health Survey sample (9,047) and not the slightly smaller sample of respondents who consented to follow up (8,305).

6.2 Clustering

The Scottish Health Survey sample was selected using a multi-stage clustered design - a sampling methodology often used in national surveys as it is more cost-effective than designs without clustering. Respondents are only selected from a subset of primary sampling units (PSUs) instead of selecting the same proportion of respondents from every PSU in the population. The PSUs used in most instances for the 1998 survey were postcode sectors (around 5000 households per sector). However, to ensure a balanced set of PSUs, some postcode sectors covering large areas (or straddling two health-board based regions) were split into two or more sub-sectors, whilst some sparsely populated postcode sectors were combined with adjacent sectors.

312 PSUs were selected from a total of 935 (following implicit stratification by Carstairs deprivation score within each region), which resulted in respondents being clustered within specific geographical areas. If this feature of the sampling methodology were ignored in subsequent analyses, the model estimates would appear to be too precise (i.e. the standard errors produced would be underestimated) and could lead to spurious statistical significance. Our models are adjusted for this clustering, which has the effect of reducing the precision of the population estimates. In other words, adjusting for clustering accurately increases the size of the confidence intervals.

6.3 Stratification

Stratification serves to ensure that the sample is distributed over the strata (deprivation score) in the same way as the wider population. The analyses presented in the main report *do not allow for this stratification in the calculation of the standard errors and design effects.* The reason for this is that the analysis technique used in this project (Cox's proportional hazard model) is not available with the survey commands in Stata, which incorporate the effect of clustering and stratification as well as the effect of sampling weights when computing the variance, standard error, and confidence intervals. (See 'Carolina Population Centre' website[15] for further details on this). In such a scenario the alternative way to analyse the survey data is using the estimation commands with pweights and robust cluster() options to handle the sampling weights and clustering properly, however there is no option to specify the stratification variable, and as a result the standard error may be larger than it would be using another analysis technique allowing for stratification. As mentioned above if clustering effects are ignored, estimates appear too precise. Stratification effects (where they exist) act in the opposite direction; however, they are generally weaker than clustering effects. Consulting the ‘Practical Exemplars on the Analysis of Surveys’ website[16], the effects of each of the survey design elements were tested, and it was found that it is the PSU’s that are the main reason for reduced precision of estimates.

6.4 Exposure Time

Respondents were interviewed at differing times throughout the 13-month period starting in April 1998. Since we know the exact interview date, the exposure time for each respondent was defined as either: the end of follow-up date minus the interview date (for those who had no admission); the admission date minus the interview date (for those experiencing an admission); or the death date minus the interview date (for those that died without any admission).

[15] http://www.cpc.unc.edu/services/computer/presentations/statatutorial/example31.html
[16] http://www.napier.ac.uk/depts/fhls/peas/index.htm#

7 Emigration – Impact on Modelling

In order to analyse the impact of emigration on our modelling strategies, 2 Cox proportional hazard models were run on the 1998 respondents (aged 16-74) with "First Hospital Admission" full follow-up as the outcome variable, and age, sex and General Health status as explanatory variables:

Model 1: All respondents aged 16-74, including emigrants

Model 2: Excluding emigrants

In total 43.8% (un-weighted)/40.5% (weighted) of respondents had a subsequent hospital admission for any condition.

Table 10: CHI Status by Any Admission

CHI Status by Any Admission				
	No	**Yes**	**Total**	**% Any Admission**
Current	4,347	3,146	7,493	42.0%
Dead	61	420	481	87.3%
Non-emigrants	4,408	3,566	7,974	44.7%
Known Emigrants (or lost to follow-up)	194	55	249	22.1%
Non-links to CHI	62	20	82	24.4%
Emigrants	256	75	331	22.7%
Total	**4,664**	**3,641**	**8,305**	**43.8%**

Respondents have been categorised by their CHI status and the percentage of these who have had any hospital admission within each category is shown in Table 10 above. Of the 331 emigrants, 22.7% of them admissions following the survey compared with 44.7% of the non-emigrants (those with a current CHI status or flagged as dead on the CHI).

The results of the Cox Proportional Hazard regression model are shown in **Appendix 3**.

Conclusion

Appendix 3 shows that both models have identical statistically significant risk factor categories and that there is very little difference between the 2 models in terms of Hazard Ratios of age, sex and General Health. This suggests that whether emigrants are included or excluded from the modelling will have minimal impact on the results. It was therefore decided that all further analysis would be run **excluding the known 331 (un-weighted)/376 (weighted) emigrants**, as this would result in a cleaner data set.

8 Missing values

In this chapter we examine the extent of missing values in the dataset and their impact on modelling.

Many missing values were present among the chosen risk factors from the 1998 SHeS. This was due to respondents refusing to answer specific questions and/or refusing to have a biological measurement taken. The missing values were most evident in the Biological risk factors, with the highest percentage of missing values present for Fibrinogen, with 2,994 (37.6%) respondents in the working dataset (excluding emigrants) refusing to have a sample taken. Due to the extent of the many missing values present, it was important to **investigate the impact of missing values on modelling and how best to deal with them.** In investigating this, it was also possible to create a profile of those respondents who had refused to consent to follow-up.

For this analysis the full un-linked 1998 SHeS data for respondents aged 16-74 was used. This contains 9,040 cases, including the 8% of respondents who refused permission to link to administrative data.

Three Logistic modelling scenarios with – "Permission to flag respondent" (0=Given, 1=Refused) as the Response Variable were considered as follows:

Model 1: All missing values per variable are included as a single dump category called "Missing".

Model 2: Exclude all cases having missing values in any of the variables included for selection.

Model 3: Exclude the variables with large numbers of missing values (>=1,000), and exclude missing values (<1,000) from the model

Appendix 4 presents the results of the 3 modelling scenarios used. These results are summarised and discussed in the following section.

8.1 Model 1

Model 1 has the advantage of including all the cases in the analyses. However, the "Missing" categories may have undue influence on the results.

This shows that *age* is a significant factor with older people more likely to refuse permission but there is no difference between the sexes.

The only significant lifestyle factors show that people eating *root vegetables* and those eating *raw vegetables* at least once a day were significantly more likely to refuse.

Significant "Missing" categories in the biological variables *blood pressure* and *body mass index* suggest there is a correlation between non-participation in the nurse visit and in refusing permission to linkage. People who are obese are less likely to refuse. Confusingly, people who have a *limiting longstanding illnes*s and those who have no limiting longstanding illness are more likely to refuse than those with a non-limiting longstanding illness.

Significant social factors affecting "refusal" include: -
People in the *least deprived* quintile more likely to refuse compared to those in the middle quintile.
People not claiming *income related benefits* more likely to refuse compared to those that do.
People claiming *unemployment benefit* more likely to refuse compared to those that do not.
People *renting privately* more likely to refuse compared to house owners.
People with *no qualifications* more likely to refuse compared to those with degrees, Highers, A-levels or equivalent.
Retired people more likely to refuse compared to those in employment.
Skilled non-manual more likely to refuse compared to professional workers.
Residents of areas outside the *inner city* more likely to refuse.
People *born* outside Scotland more likely to refuse compared to Scots-born.
Residents of *areas* outside the Highlands and Islands more likely to refuse.

8.2 Model 2

Model 2 excludes the "Missing" categories. However, this results in a sample size of only 4,218 – less than half the total number of respondents.

By comparing the percentage distributions of each risk factor in this model with those in Model 1, it is noticeable that the profile of the respondents left in Model 2 is quite different from the full sample. For example, smaller proportions of the very young and the very old remain in the sample. The sample also contains "healthier" subjects in terms of vegetable consumption, blood pressure, BMI measurements, general health and longstanding illnesses. They are less deprived, less likely to be on benefits, more likely to own their own home, better educated and of a higher social class.

This should be kept in mind if modelling without the cases with missing values – i.e. that the sample is unlikely to be representative of Scotland. This is especially true if trying to make inferences about the nation at large.

In terms of the differences in outcome of "Consent to follow-up", the percentage non-consenting drops from 8.0% to 5.6% when cases with missing values are excluded.

Age becomes a stronger explanatory variable with all groups now significant compared to the 20-24 reference group, and with larger odds-ratios. Blood pressure, BMI, GHQ score, longstanding illness and economic activity are no longer significant variables and drop-out of the model to be replaced with *incapacity benefit* and *overcrowding*.

8.3 Model 3

Model 3 retains over 98% of the full sample but loses important data, particularly around biological risk factors. The variables root vegetable consumption, blood pressure, BMI, GHQ score, social class, C-reactive protein, fibrinogen, Total & HDL cholesterol, Gamma-GT, waist hip ratio, were all expressly excluded from entering the model to maintain the near complete sample size whilst removing missing values in the remaining variables.

Strangely this model reduces the influence of age as a risk factor for non-consenting with smaller odds-ratios and only 3 of the age groups now significant. Most of the other (remaining) factors have similar effects to Model 1 except there is one additional variable included – Drinking. The respondents who "never drink or trivial" have an odds-ratio of 2, indicating they are 2 times more likely to refuse consent than respondents who drink "heavily" (reference group).

Conclusion

It was reassuring to find that results from model 1 (including all missing values in a separate category) were very similar to model 3 (excluding variables with large numbers of missing values). It was therefore decided that adopting the approach of **model 1** (including missing values in separate categories) was the best option as all variables (especially the important biological variables) are retained, while still maintaining a full sample size. Other techniques for handling missing values - notably imputation, as referred to on the 'Practical Exemplars on the Analysis of Surveys' website[17] - were reviewed. However, due to complexity and time constraints, these were not adopted in this project.

[17] http://www.napier.ac.uk/depts/fhls/peas/index.htm#

9 Representativeness of Survey Samples

It is of interest to investigate whether or not the 1995 and in particular the 1998 final linked Survey-working data sets (adjusted for emigrants and weighted by allocated survey weights) have a similar age and sex breakdown as the published population estimates from the GRO(S). A Chi-Square Test for Independence was used to test for this. Tables 11a and 11b below show, for both the survey years, age and sex structures and the result of the Chi-square test.

Table 11a – 1995 Survey Respondent age/sex distribution vs. 1995 Scotland Population

1995		Males		Females		Total	
Age Group		**Survey**	**Scotland**	**Survey**	**Scotland**	**Survey**	**Scotland**
16 to 34	Count	1,420	697,900	1,404	708,626	2,824	1,406,526
	% of Total	20.8	21.2	20.5	21.5	41.3	42.7
35 to 54	Count	1,435	667,450	1,447	683,999	2,882	1,351,449
	% of Total	21.0	20.3	21.2	20.8	42.1	41.0
55 to 64	Count	550	255,067	583	280,062	1,133	535,129
	% of Total	8.0	7.7	8.5	8.5	16.6	16.2
Total	Count	3,405	1,620,417	3,434	1,672,687	6,839	3,293,104
	% of Total	49.8	49.2	50.2	50.8	100.0	100.0

The following Chi-square test summary was obtained – *X^2=6.71, dof=2, p-value=0.035, sig level=0.01*, indicating that there is a **no** significant difference between the 1995 survey sample and 1995 Scotland population in terms of their age and sex structure.

Table 11b – 1998 Survey Respondent age/sex distribution vs. 1995 Scotland Population

1998		Males		Females		Total	
Age Group		**Survey**	**Scotland**	**Survey**	**Scotland**	**Survey**	**Scotland**
16 to 34	Count	1,445	656,712	1,368	674,547	2,813	1,331,259
	% of Total	18.1	17.6	17.2	18.1	35.3	35.7
35 to 54	Count	1,538	696,401	1,547	717,301	3,085	1,413,702
	% of Total	19.3	18.7	19.4	19.2	38.7	37.9
55 to 64	Count	551	257,313	591	279,628	1,142	536,941
	% of Total	6.9	6.9	7.4	7.5	14.3	14.4
65 to 74	Count	428	197,891	505	247,593	933	445,484
	% of Total	5.4	5.3	6.3	6.6	11.7	12.0
Total	Count	3,963	1,808,317	4,011	1,919,069	7,974	3,727,386
	% of Total	49.7	48.5	50.3	51.5	100.0	100.0

The following Chi-square test summary was obtained – *X^2=7.96, dof=3, p-value=0.047, sig level=0.01*, indicating that there is a **no** significant difference between the 1998 survey sample and 1998 Scotland population in terms of their age and sex structure.

In summary both working files are representative of the age and sex structure of the Scottish population (as published by the GRO(S)). It should be stressed, however, that work carried out by Alistair Leyland[18], showed that in other ways (i.e. not just age & sex) the survey sample is not truly representative**.**

[18] *http://www.ccsr.ac.uk/esds/events/2004-10-29/slides/leyland.ppt*

10 Conclusions

The main strength of the SHeS/SMR dataset is as a resource to carry out follow-up analysis, combining lifestyle, biological and social variables with hospital admission data, cancer registrations and death records. The main report (www.scotpho.org.uk/linkedshesreport) describes the results of some initial analyses to investigate the relationship between a range of risk factors and the risk of hospitalisation or death. It is also worth noting that the utility of this resource will improve as further years of follow-up data accrue and with the addition of 2003 Scottish Health Survey data.

Project Group Contact E-mail Addresses

Professor Phil Hanlon (University of Glasgow) – Phil.Hanlon@clinmed.gla.ac.uk

Professor Matt Sutton (University of Aberdeen) – m.sutton@abdn.ac.uk

David Walsh (NHS Health Scotland, **now Glasgow Centre for Population Health (GCPH)**) – david.walsh@drs.glasgow.gov.uk

Bruce Whyte (Information Services NHS NSS, **now GCPH**) – bruce.whyte@drs.glasgow.gov.uk

Richard Lawder (Information Services NHS NSS) – richard.lawder@isd.csa.scot.nhs.uk

David Clark (Information Services NHS NSS) – david.clark@isd.csa.scot.nhs.uk

Andrew Elders (Information Services NHS NSS, **now Fife HB**) – andrew.elders@nhs.net

Appendix 1 - Most common cause of death for each sex and age group

TABLE 5 - Most common cause of death for each sex and age group
(ICD10 deaths only, Figures in brackets indicate actual numbers of death from each cause)

	Both sexes	Male	Female
All Ages	I21 - Acute myocardial infarction (60) *11% of deaths*	C34 - Malignant neoplasm of bronchus and lung (38) *12% of deaths*	I21 - Acute myocardial infarction (24) *10% of deaths*
16-34	E95 - Nutritional and metabolic disorders in diseases (5) *16% of deaths*	E95 - Nutritional and metabolic disorders in diseases (4) *24% of deaths*	Q05 - Spina bifida (2) *13% of deaths*
35-54	I21 - Acute myocardial infarction (13) *10% of deaths*	K70 - Alcoholic liver disease (9) *12% of deaths*	C50 - Malignant neoplasm of breast (6) I21 - Acute myocardial infarction (6) *10% of deaths*
55+	C34 - Malignant neoplasm of bronchus and lung (49) *13% of deaths*	C34 - Malignant neoplasm of bronchus and lung (31) *14% of deaths*	C34 - Malignant neoplasm of bronchus and lung (18) I21 - Acute myocardial infarction (18) *10% of deaths*

Although lung cancer is the single most common cause of death in males, 36 males have died from AMI, i.e. 11.7% of all male deaths – a higher proportion than female deaths caused by AMI.

Appendix 2 - Most common ICD10 diagnosis

TABLE7 – Most common ICD10 diagnosis chapter heading (number and percentage of respondents in brackets)

	Both sexes	Males	Females
All ages	Diseases of the digestive system (1956, 12.5%) Symptoms signs & abnormal clinical/laboratory findings NEC (1752, 11.2%) Diseases of the circulatory system (1286, 8.2%)	Diseases of the digestive system (851, 12.2%) Symptoms signs & abnormal clinical/laboratory findings NEC (722, 10.4%) Diseases of the circulatory system (679, 9.8%)	Diseases of the digestive system (1105, 12.7%) Symptoms signs & abnormal clinical/laboratory findings NEC (1030, 11.9%) Diseases of the genitourinary system (853, 9.8%)
16-34	Diseases of the digestive system (441, 8.3%) Symptoms signs & abnormal clinical/laboratory findings NEC (394, 7.4%) Health status and contact with health services (385, 7.2%)	Injury poisoning and other consequences of external causes (208, 8.9%) Diseases of the digestive system (181, 7.8%) Health status and contact with health services (133, 5.7%)	Pregnancy childbirth and the puerperium (280, 9.4%) Diseases of the genitourinary system (276, 9.2%) Symptoms signs & abnormal clinical/laboratory findings NEC (276, 9.2%) Diseases of the digestive system (260, 8.7%)
35-54	Diseases of the digestive system (750, 12%) Symptoms signs & abnormal clinical/laboratory findings NEC (659, 10.5%) Diseases of the genitourinary system (522, 8.3%)	Diseases of the digestive system (319, 11.2%) Symptoms signs & abnormal clinical/laboratory findings NEC (268, 9.4%) Diseases of the circulatory system (216, 7.6%)	Diseases of the digestive system (431, 12.6%) Diseases of the genitourinary system (405, 11.9%) Symptoms signs & abnormal clinical/laboratory findings NEC (391, 11.4%)
55+	Diseases of the digestive system (765, 18.8%) Diseases of the circulatory system (713, 17.5%) Symptoms signs & abnormal clinical/laboratory findings NEC (699, 17.2%)	Diseases of the circulatory system (415, 23.2%) Diseases of the digestive system (351, 19.6%) Symptoms signs & abnormal clinical/laboratory findings NEC (336, 18.8%)	Diseases of the digestive system (414, 18.1%) Symptoms signs & abnormal clinical/laboratory findings NEC (363, 15.9%) Diseases of the circulatory system (298, 13%)

TABLE7a – Most common diagnoses (number and percentage of respondents in brackets)

	Both sexes	Males	Females
All ages	R10 - Abdominal and pelvic pain (512, 3.3%) R07 - Pain in throat and chest (411, 2.6%) Z30 - Contraceptive management (352, 2.2%)	R07 - Pain in throat and chest (218 ,3.1%) I25 - Chronic ischaemic heart disease (164 2.4%) Z30 - Contraceptive management (154, 2.2%)	R10 - Abdominal and pelvic pain (378 ,4.4%) O04 - Medical abortion (217 ,2.5%) Z30 - Contraceptive management (198, 2.3%)
16-34	Z30 - Contraceptive management (227, 4.3%) O04 - Medical abortion (200, 3.8%) R10 - Abdominal and pelvic pain (193, 3.6%)	Z30 - Contraceptive management (74, 3.2%) M23 - Internal derangement of knee (37, 1.6%) R07 - Pain in throat and chest (29, 1.2%) R10 - Abdominal and pelvic pain (29, 1.2%)	O04 - Medical abortion (200, 6.7%) R10 - Abdominal and pelvic pain (164, 5.5%) Z30 - Contraceptive management (153, 5.1%)
35-54	R10 - Abdominal and pelvic pain (184, 2.9%) R07 - Pain in throat and chest (181, 2.9%) Z30 - Contraceptive management (125, 2%)	R07 - Pain in throat and chest (92, 3.2%) Z30 - Contraceptive management (80, 2.8%) R10 - Abdominal and pelvic pain (54, 1.9%)	R10 - Abdominal and pelvic pain (130, 3.8%) R07 - Pain in throat and chest (89, 2.6%) N92 - Excessive frequent and irregular menstruation (78, 2.3%)
55+	R07 - Pain in throat and chest (181, 4.4%) I25 - Chronic ischaemic heart disease (160, 3.9%) H26 - Other cataract (155, 3.8%)	I25 - Chronic ischaemic heart disease (110, 6.1%) R07 - Pain in throat and chest (97, 5.4%) I20 - Angina pectoris (81, 4.5%) I21 - Acute myocardial infarction (81, 4.5%)	H26 - Other cataract (94, 4.1%) R07 - Pain in throat and chest (84, 3.7%) R10 - Abdominal and pelvic pain (84, 3.7%) K57 - Diverticular disease of intestine (60, 2.6%)

Appendix 3 – Effect of Emigration on Modelling

Cox Proportional Hazard model for "General Health Status" on "First Hospital Admission"

Factor	Model 1 - Including Emigrants - N=8,305					Model 2 - Excluding Emigrants - N=7,974				
	N	Hazard Ratio	p-value	95% Lower CI	95% Upper CI	N	Hazard Ratio	p-value	95% Lower CI	95% Upper CI
Age										
16-19	358	0.83	*0.183*	0.63	1.09	323	0.80	*0.124*	0.84	0.98
20-24	508	0.78	*0.044*	0.61	0.99	452	0.78	*0.035*	0.60	1.06
25-29	723	0.86	*0.094*	0.71	1.03	663	0.84	*0.074*	0.61	0.98
30-34	898	0.95	*0.595*	0.80	1.14	850	0.91	*0.294*	0.70	1.02
35-39[1]	914	1.00				872	1.00			
40-44	789	0.95	*0.590*	0.79	1.14	772	0.92	*0.359*	0.77	1.10
45-49	704	0.99	*0.929*	0.82	1.20	687	0.95	*0.609*	0.78	1.16
50-54	765	1.17	*0.064*	0.99	1.37	750	1.12	*0.177*	0.95	1.32
55-59	683	1.37	*0.001*	1.14	1.65	673	1.32	*0.003*	1.10	1.59
60-64	668	1.43	*0.000*	1.21	1.70	657	1.36	*0.000*	1.15	1.62
65-69	687	1.80	*0.000*	1.52	2.12	680	1.72	*0.000*	1.46	2.02
70-74	608	2.24	*0.000*	1.89	2.64	595	2.15	*0.000*	1.83	2.54
Sex										
Male	3,664	0.91	*0.016*	0.84	0.98	3,507	0.91	*0.014*	0.84	0.98
Female[1]	4,641	1.00				4,467	1.00			
General Health										
Very Good[1]	2,917	1.00				2,783	1.00			
Good	3,280	1.40	*0.000*	*1.28*	*1.52*	3,128	1.39	*0.000*	*1.28*	*1.52*
Fair	1,568	2.12	*0.000*	*1.91*	*2.36*	1,533	2.11	*0.000*	*1.89*	*2.35*
Bad	450	3.45	*0.000*	*2.97*	*4.01*	442	3.49	*0.000*	*3.00*	*4.06*
Very Bad	90	3.60	*0.000*	*2.59*	*4.99*	88	3.76	*0.000*	*2.71*	*5.21*

[1] Reference Category

APPENDIX 4 - 'Missing Values' Modelling

Multivariate Models (forward stepwise logistic regression) controlled for both age and sex

Dependent variable	Model 1:	Including Missing Values		Model 2:	Excluding Missing Values		Model 3:	Subset of Variables (excl. missing values)	
Consent to follow-up via record linkage									
Survey year		1998			1998			1998	
	Mean	**Exp(B)**	**Sig.**	**Mean**	**Exp(B)**	**Sig.**	**Mean**	**Exp(B)**	**Sig.**
General									
Sex									
Male	49.2	Reference		49.9	Reference		49.0	Reference	
Female	50.8	1.082	*0.363*	50.1	1.021	*0.887*	51.0	1.129	*0.168*
Age Group									
16-19	6.9	1.255	*0.361*	2.4	6.006	*0.003*	6.0	1.568	*0.056*
20-24	8.5	Reference		6.7	Reference		8.6	Reference	
25-29	10.0	1.272	*0.269*	9.7	2.838	*0.047*	10.2	1.242	*0.317*
30-34	11.0	1.735	*0.008*	12.1	5.018	*0.001*	11.2	1.519	*0.043*
35-39	10.7	1.566	*0.035*	12.1	3.799	*0.008*	10.8	1.383	*0.123*
40-44	9.5	1.597	*0.034*	12.2	3.557	*0.012*	9.5	1.334	*0.185*
45-49	8.8	1.465	*0.101*	10.6	2.865	*0.046*	8.9	1.145	*0.556*
50-54	8.7	1.853	*0.006*	10.0	3.574	*0.014*	8.8	1.404	*0.125*
55-59	7.3	1.598	*0.048*	7.5	5.271	*0.001*	7.4	1.361	*0.184*
60-64	6.8	2.405	*0.000*	6.6	3.508	*0.022*	6.9	1.741	*0.015*
65-69	6.3	1.591	*0.066*	5.3	3.528	*0.028*	6.3	1.203	*0.450*
70-74	5.4	2.516	*0.000*	4.6	5.839	*0.001*	5.5	1.883	*0.007*
Behavioural/Lifestyle									
Cooked Root Vegetable Consumption									
Missing	8.8	1.236	*0.211*	-			-		
>= Once a day	1.8	2.609	*0.000*	2.1	4.185	*0.000*	-		
1-6 times per week	66.7	1.205	*0.083*	73.6	1.292	*0.167*	-		
< Once per week	22.7	Reference		24.3	Reference		-	Reference	
Raw Vegetable or Salad Consumption									
Missing	0.0	0.000	*1.000*	-					
>= Once a day	11.0	1.667	*0.000*	11.4	1.916	*0.008*	11.1	1.639	*0.000*
1-6 times per week	61.3	1.118	*0.274*	65.1	1.074	*0.714*	61.4	1.087	*0.402*
< Once per week	27.7	Reference		23.5	Reference		27.6	Reference	
Drinking									
Missing	1.2			-			-		
Never drank & trivial	5.8			4.3			5.9	1.968	*0.001*
Ex-drinker	4.1			3.4			4.2	1.287	*0.321*
Light drinker	45.3			46.5			45.9	1.375	*0.058*
Moderate drinker	21.0			22.8			21.2	1.303	*0.142*
Heavy drinker	9.0	Reference		9.6	Reference		9.1	Reference	
Excessive drinker	13.7			13.5			13.9	1.237	*0.276*
Biological									
Blood Pressure									
Missing	18.8	3.362	*0.000*	-			-		
Hypertensive	20.1	0.866	*0.256*	24.0			-		
Normotensive	61.1	Reference		76.0	Reference		-	Reference	
BMI									
Missing	9.5	1.815	*0.000*	-			-		
Underweight (<20)	4.9	1.342	*0.183*	4.5			-		
Desirable (20-25)	33.0	1.463	*0.004*	36.6			-		
Overweight (25-30)	33.9	1.338	*0.024*	39.6			-		
Obese (>30)	18.7	Reference		19.3	Reference		-	Reference	
General Health Measures									
GHQ Score									
Missing	6.1	2.272	*0.000*	-			-		
0 score	55.0	1.100	*0.364*	60.7			-		
1-3 score	24.3	Reference		25.1	Reference		-	Reference	
4+ score	14.5	1.113	*0.450*	14.2			-		
Longstanding Illness									
Limiting LI	23.6	1.449	*0.011*	20.6			23.7	1.371	*0.026*
Non-limiting LI	16.6	Reference		15.6	Reference		16.5	Reference	
No LI	59.9	1.381	*0.010*	63.9			59.7	1.465	*0.002*

Multivariate Models (forward stepwise logistic regression) controlled for both age and sex

Dependent variable	Model 1: Including Missing Values			Model 2: Excluding Missing Values			Model 3: Subset of Variables (excl. missing values)		
Consent to follow-up via record linkage									
Survey year		1998			1998			1998	
	Mean	**Exp(B)**	**Sig.**	**Mean**	**Exp(B)**	**Sig.**	**Mean**	**Exp(B)**	**Sig.**
Social									
Deprivation Quintiles (Carstairs 1991)									
1 - Least deprived	21.9	1.579	*0.001*	23.8	3.239	*0.000*	21.9	1.531	*0.001*
2	17.5	1.191	*0.221*	18.3	2.402	*0.002*	17.5	1.197	*0.204*
3	18.7	Reference		20.1	Reference		18.7	Reference	
4	20.3	1.125	*0.394*	19.1	1.936	*0.020*	20.3	1.165	*0.263*
5 - Most deprived	21.5	0.747	*0.052*	18.6	1.491	*0.175*	21.5	0.796	*0.124*
Income Related Benefit									
Yes	23.8	Reference		19.0	Reference		23.7	Reference	
No	76.2	1.780	*0.000*	81.0	2.484	*0.000*	76.3	1.638	*0.000*
Unemployment Benefit									
Yes	2.1	2.147	*0.013*	1.7	3.820	*0.003*	2.0	2.391	*0.004*
No	97.9	Reference		98.3	Reference		98.0	Reference	
Incapacity Benefit									
Yes	6.4			5.5	1.725	*0.049*	6.5		
No	93.6	Reference		94.5	Reference		93.5	Reference	
Overcrowding									
Yes	1.4			1.2	3.482	*0.014*	1.4		
No	98.6	Reference		98.8	Reference		98.6	Reference	
Housing Tenure									
Missing	0.1	4.399	*0.047*	-			-		
House owned or with mortgage	66.5	Reference		73.1	Reference		66.7	Reference	
Publicly rented	22.6	1.012	*0.924*	18.5	0.934	*0.767*	22.6	1.090	*0.463*
Privately rented	10.7	1.330	*0.046*	8.4	2.173	*0.002*	10.7	1.598	*0.001*
Highest Educational Qualification									
Missing	0.3	1.328	*0.857*	-			-		
A-level(s), degree or equivalent	52.8	Reference		54.4	Reference		52.8	Reference	
GCSE at A-c or equivalent	14.4	1.236	*0.098*	15.9	1.140	*0.535*	14.5	1.271	*0.053*
Other formal qualification	6.5	0.958	*0.824*	6.7	0.806	*0.543*	6.5	1.106	*0.587*
No formal qualification	26.1	1.613	*0.000*	23.0	1.796	*0.003*	26.2	1.694	*0.000*
Economic Activity									
Missing	0.3	1.287	*0.836*	-			-		
In employment	59.0	Reference		67.4	Reference		59.5	Reference	
Unemployed	3.8	0.554	*0.055*	3.0			3.8	0.557	*0.071*
Retired	36.9	1.279	*0.028*	29.6			36.8	1.338	*0.008*
Social Class									
Missing	3.9	2.215	*0.000*	-			-		
I - Professional	33.4	Reference		37.5	Reference		-		
IIIN - Skilled non-manual	14.6	1.429	*0.006*	15.0	2.009	*0.001*	-		
IIIM - Skilled manual	27.6	1.221	*0.088*	28.3	1.570	*0.021*	-		
IV - Semi-skilled manual	14.8	1.261	*0.108*	14.4	1.101	*0.718*	-		
V - Unskilled manual	5.2	1.304	*0.186*	4.4	1.916	*0.088*	-		
Others	0.4	2.043	*0.163*	0.4	2.392	*0.271*	-		
Area Type									
Missing	0.0	0.000	*1.000*	-			-		
Inner City	5.6	Reference		4.5	Reference		5.6	Reference	
Other dense urban/town centre	10.3	2.204	*0.001*	9.3	9.818	*0.004*	10.2	2.093	*0.003*
Suburban residential (city/large town outskirts)	53.6	1.840	*0.008*	53.6	6.080	*0.019*	53.6	1.623	*0.034*
Rural residential / village centre	25.2	2.103	*0.002*	26.5	8.386	*0.006*	25.2	1.822	*0.014*
Rural agricultural with isolated dwelling	5.4	1.724	*0.067*	6.0	5.270	*0.046*	5.4	1.550	*0.135*
Country of Birth									
Missing	0.2	1.019	*0.993*	-			-		
Scotland	86.0	Reference		85.9	Reference		86.2	Reference	
Other UK	10.2	1.401	*0.010*	10.7	1.730	*0.008*	10.2	1.386	*0.011*
Outside UK	3.6	1.382	*0.102*	3.5	2.005	*0.027*	3.6	1.406	*0.075*
Health Board Region									
Highlands & Islands	5.3	Reference		5.5	Reference		5.3	Reference	
Grampian & Tayside	18.1	2.538	*0.001*	18.9	14.142	*0.021*	18.2	2.463	*0.002*
Lothian & Fife	22.2	1.981	*0.021*	22.9	9.502	*0.051*	22.3	1.924	*0.025*
Borders, D & G	4.9	2.198	*0.019*	5.1	3.666	*0.303*	4.9	2.044	*0.033*
Glasgow	17.4	4.152	*0.000*	15.9	31.167	*0.003*	17.4	4.000	*0.000*
Lanarkshire, Ayr & Arran	18.3	4.086	*0.000*	17.5	23.578	*0.006*	18.2	3.977	*0.000*
Forth Valley, Argyll & Clyde	13.7	4.301	*0.000*	14.2	40.275	*0.001*	13.7	4.138	*0.000*

Model Summary:-

	Model 1	Model 2	Model 3
Un-weighted number of cases	*9,040*	*4,218*	*8,899*
Weighted number of cases	*8,995*	*4,226*	*8,839*
Weighted number with "Permission given"	*8,272*	*3,991*	*8,143*
Weighted number with "Refused"	*723*	*235*	*696*
Un-weighted percent missing	*0.0%*	*53.3%*	*1.6%*
Number of Steps	*16*	*13*	*12*
Cox & Snell R^2	*5.9%*	*5.9%*	*2.9%*
Nagelkerke R^2	*13.8%*	*16.9%*	*6.9%*
Cut-off for predicted values	*0.080*	*0.056*	*0.079*
"Permission given" predicted correctly	*69.4%*	*69.2%*	*62.0%*
"Refused" predicted correctly	*62.9%*	*69.5%*	*62.9%*
Overall predicted correctly	*68.9%*	*69.2%*	*62.1%*