1996/97 New Zealand Health Survey Statistical Methodology

Written by Associate Professor Stephen J Haslett,

Statistics Research and Consulting Centre

Massey University,

in conjunction with Statistics New Zealand

Contents

Introduction

The 1996/97 New Zealand Health Survey (1996/97 Health Survey) covered a one-year period starting 14 October 1996. During that year, there was a three-week period where no interviewing was conducted because of the Christmas holiday period, as well as a one-week break between the second and third quarter of data collection.

This document contains information on the objectives, design methodology and estimation / analysis procedures for the 1996/97 Health Survey.

Background

In April 1996, the Ministry of Health (MoH) and the four regional health authorities (RHAs) invited Statistics New Zealand (SNZ) to submit a tender to conduct a national health survey. The tender was submitted in late April 1996 and SNZ was advised that it was the successful tenderer in May 1996.

The tender document for the 1996/97 Health Survey specified that the questionnaire design and analysis of the data was the responsibility of the MoH and the RHAs. SNZ, as the successful tenderer was required to:

- design the sampling methodology within the parameters provided
- train interviewers
- contact and interview respondents
- assist with the final development of the questionnaire
- undertake the pretesting of the questionnaire
- assist with the ethics committee process
- take part in information/consultation meetings with Mäori and Pacific Islands groups
- give progress reports, including key results, at predetermined times during the fieldwork phase
- clean and process the data
- meet predetermined timeframes
- work collaboratively with the National Nutrition Survey provider.

The 1996/97 Health Survey was also used to generate a sample for a National Nutrition Survey (NNS), which was conducted by the University of Otago (LINZ) on behalf of the Ministry of Health. Targeted respondents in the 1996/97 Health Survey were asked to give their written consent to have their names, addresses, telephone numbers and coded information passed to the University of Otago, which then contacted them at a later date to participate in the NNS.

In addition to the requirements specified in the tender:

- additional Mäori and Pacific Island households were selected to ensure that the objectives of the NNS were met; and

- an increase in sample numbers within the Central RHA was made in order to obtain regional and sub-regional estimates.

The data collected from the 1996/97 Health Survey were made available in the form of a computerised survey data file to the Ministry of Health by Statistics New Zealand. The data file was used for further analysis carried out by the Ministry of Health in collaboration with the Statistics Research and Consulting Centre at Massey University.

Public access to the survey data file is possible, subject to formal application and approval. Application for access to the data file, for *bona fide* research purposes only, can be made to Statistics New Zealand through the Statistics New Zealand Data Laboratory.

Further detail on how to use the survey data file itself and the data dictionary for this data file, further detail on the scoring of the SF-36, additional definitions of terms, and a complete set of survey questionnaires are contained in "1996-97 New Zealand Health Survey: Unit Record Data File Documentation and User's Guide", prepared for the Ministry of Health and the Regional Health Authorities by Statistics New Zealand in December 1997.

Objectives

The Ministry of Health commissioned the 1996/97 Health Survey to obtain accurate national estimates of the health status of New Zealanders and their use of health services. The overall objectives of the survey were:

- to measure the health status of New Zealanders, including their self-perceived physical and mental health status (SF-36 questions), and prevalence of selected health conditions (asthma, diabetes, high blood pressure);

- to measure the utilisation of health services (GPs, specialists and other health professionals) and prescriptions;

- to measure individuals' perception of health services, including their satisfaction with health services and barriers to access of health services (satisfaction with GPs and other health services, unmet need for GPs and other health services, health card and health insurance);

- to provide information on selected health risks (smoking, exercise, alcohol consumption, injury, environment);

- to provide this information at both a national and RHA level and for key population groups, especially Māori and Pacific Islands people (age, sex, ethnic group);

- to compare results by key demographic and socioeconomic variables (income, income support, labour force status, occupation, education qualifications, age, sex, ethnic group, number of rooms in dwelling).

Population

The population for the 1996/97 Health Survey was defined as the total usually resident, non-institutionalised, civilian population of New Zealand of all ages, residing in private households.

The following people were excluded:
- those living in non-private dwellings
- long-term residents of rest homes, hospitals and psychiatric institutions
- inmates of penal institutions
- members of the permanent armed forces
- overseas diplomats
- overseas visitors who expect to be usually resident in NZ for less than 12 months
- people living on offshore islands (except for Waiheke Island).

Survey Design

Frame

The sampling frame used for the 1996/97 Health Survey is an area-based frame. New Zealand is divided into 18,800 small geographic areas, that are defined by meshblock boundaries. These geographic areas are called Primary Sampling Units (PSUs). Most PSUs contain between 50 and 100 dwellings, with an average size of about 65 dwellings per PSU.

The PSUs are divided into 122 groups (or strata) according to various characteristics derived from 1991 Census data. PSUs with similar characteristics are grouped together in the same strata. Characteristics used for this stratification include region (Regional Council area), urban/rural, high/low Maori population density, high/low Pacific Islands population density, plus other socio-economic variables such as level of employment, education, age and family structure. By stratifying PSUs in this way, SNZ was able to ensure that the sample selected from all 122 strata was a sample with characteristics representative of the total population.

Sample Design

Selecting Area Units (PSUs)

There were 1,752 PSUs selected for the 1996/97 Health Survey. These PSUs were selected by drawing a simple random sample from each of the 122 strata. In each of these PSUs, adult respondents were selected. In 970 of these PSUs, adult respondents were also asked to act as a proxy respondent for a selected child in the household.

Selecting Households within PSUs

A panel of households in each PSU was chosen to participate in the 1996/97 Health Survey. The number of households selected in each PSU varied across urban/rural breakdown.

Selection of Extra Maori and Pacific Island Households

Each of the 1,752 selected PSUs contained a panel of households which recently completed participation in the Household Labour Force Survey (HLFS). In all of these PSUs, households from this panel which contained an adult Pacific Islands person when last surveyed in the HLFS was selected for the 1996/97 Health Survey. In half of these 1,752 PSUs, households from this panel which contained an adult Maori when last surveyed in the HLFS was also selected for the 1996/97 Health Survey.

Only Maori and Pacific Islands people were eligible for selection within these extra households.

Selection of Oversample for the Central RHA

An additional sample of households was selected from each Central RHA PSU. This oversample ensured that reliable sub-regional estimates of the health status of the population living in the Central RHA could be produced at the completion of the survey.

The households in this oversample were not asked to consent to having their personal information passed to the University of Otago for the NNS.

Selection of a Respondent within a Household

A single eligible adult (aged 15 years and over) was uniquely selected in each household from all eligible adults within that household. The selected adult was then asked to participate in the 1996/97 Health Survey. If the household was also selected to provide child data for the 1996/97 Health Survey, the selected adult was asked if they were a parent or caregiver to any children under 15 years of age in the household. All children for whom the selected adult was a parent or caregiver were listed and one of these children was uniquely selected. The adult respondent was then asked to act as a proxy respondent for the selected child by completing the child health questionnaire.

The decision to select only one respondent per household was driven primarily by response burden issues. A high level of household response burden can be detrimental

to response rates and to the quality of data collected. As a single 1996/97 Health Survey adult questionnaire takes 30-45 minutes on average to administer, selection of multiple respondents within households was seen as inappropriate.

An additional advantage of single versus multiple respondent selection within households is the lower level of sample clustering. For a specified sample size, a less clustered sample generally provides more reliable survey estimates. There are however implications for variance (ie accuracy estimation) for sample survey estimates from the 1996/97 Health Survey and this issue will be discussed in the section on Estimation below.

Selection Probabilities and Sample Survey Weights

The selection mechanism for each household and respondent is a complex one, and not all individuals or household have a equal chance of selection. To adjust estimates of population quantities derived from the survey for this effect, it is necessary to weight the data by a weight that is approximately equal to the inverse of that person or household's selection probability. For example, a person who had a one in twenty chance of selection before sampling and is in the sample would count as approximately 20 people for survey purposes and have a sample survey weight of about 20. The reason this weighting mechanism is approximate is that the inverse selection weights are rescaled slightly, or poststratified, so that the sum of the sample survey weights over the whole sample exactly matches key demographic variables such as age or sex. Further detail is given below in the section on Estimation.

Sample Size

The 1996/97 Health Survey sample consisted of 11,921 households and a Household Form was completed for each. The Household Form listed all household members and collected the following basic demographic information: age, sex, ethnic group, marital status and relationship to the reference person. Of the households contacted, 9,382 were found to contain an eligible adult, who was asked to participate in the survey. Not all selected persons agreed to or were able to complete the interview; 7,869 interviews were obtained with enough information to be included in the files used for estimation and analysis.

In total 7,862 adult and 1,019 child records are included in the unit record file.

Sample Allocation

Each quarter's sample contained respondents spread over all the 122 strata. Where possible, selected PSUs from each stratum were allocated evenly across the four quarters. Even allocation of PSUs across quarters were also made at RHA and sub-regional level. As a result, each quarter's sample was reasonably representative of the population as a whole.

Response Rates

The adult response rate represents the proportion of eligible households visited during the survey period which provided an adult respondent. The response rate for the survey was 73.8%.

The response rate was determined by assigning each visited 1996/97 Health Survey household to one of five eligibility classes, based on the household's participation code (assigned by the survey interviewer) and the selected adult's response to core survey questions.

The five eligibility classes were:

1. **Ineligible pre-contact**: Households not found to contain any eligible adults, without the need to make contact with a person in the household.
    (Includes: vacant dwellings, dwellings under construction, derelict dwellings, dwellings demolished or removed, dwelling converted to non-private dwelling)

2. **Ineligible post-contact:** Households not found to contain any eligible adults, after contact with a person in the household is made. (Includes: All persons out of scope - all overseas visitors, diplomatic residence, no usual residents in survey population)

3. **Eligible non-responding:** Households found to contain at least one eligible adult but no adult selected OR selected adult failed to complete 90% of core survey questions. (Includes: Household non-contact or refusal by contact person where household known to be eligible (e.g. information from neighbour), adult selected but did not complete 90% of adult core questions due to refusal/non-contact)

4. **Eligible responding**: Households found to contain at least one eligible adult, one adult selected and at least 90% of survey core questions completed.

5. **Unknown eligibility**: Not known whether there are eligible adults usually resident. (Includes: full household non-contact, full refusal of contact person - eligibility unknown)

Eligible households refer to those which contained at least one eligible adult. The response rate was then calculated as:

$$\text{Response Rate} = 1 / (1 + (C/D) + (E/D)*(C+D)/(B+C+D))$$

where B = sum of weights of households ineligible post-contact
       C = sum of weights of households eligible non-responding
       D = sum of weights of households eligible responding
       E = sum of weights of households with unknown eligibility.

Questionnaire Content

The 1996/97 Health Survey used four different questionnaires - Household Form, Adult Questionnaire, Child Questionnaire, General Health Questionnaire. Copies of the questionnaire can be found in *Taking the Pulse: The 1996/97 New Zealand Health Survey* (Ministry of Health 1999).

Household Form

The purpose of the household form was to list the following characteristics of all people who usually live in the household; age, sex, ethnic group, marital status and relationship to other household members. From this list, one person, aged 15 and over, was randomly selected to participate in the 1996/97 Health Survey using a selection methodology described in the previous section. This list was also used to produce outputs on the respondent's family/household characteristics.

Adult Questionnaire

This questionnaire included questions on the following topics:

- asthma
- diabetes
- high blood pressure
- health care services used, satisfaction and unmet need
- prescription items
- injury and poisoning
- smoking
- environment
- exercise
- insurance
- health cards
- income support
- standard demographic socioeconomic variables (age, ethnic group, education qualifications, labour force status, income, occupation)
- number of rooms in the dwelling

The questionnaire was administered in a face-to-face interview with the selected respondent, i.e. a randomly selected person who must be 15 years of age or over. Proxy interviews for the adult questionnaire were not allowed but where the selected respondent had an illness, disability or language difficulty which prevented them from participating in the survey, assistance from another household member was accepted.

The questions on asthma, diabetes and high blood pressure were taken or adapted from previous surveys on these topics, after consultation with relevant New Zealand experts and the Ministry of Health.

The questions on use of GP and specialist services, use of other health services, prescriptions, health cards and health insurance were taken or adapted from those in the 1992/93 Household Health Survey. Many of the recall periods in the survey were lengthened after the experience of the 1992/93 Survey. Some of the questions on the use of other health services and health cards were also modified so that they were consistent with those asked in the 1996 Household Disability Survey.

The injury questions and the questions on physical activity were largely new questions, but were heavily influenced by the treatment of these topics in overseas health surveys. The Hillary Commission, and other New Zealand experts, were consulted when developing the physical activity questions.

The smoking questions were largely taken or adapted from the 1996 Census of Population and Dwellings, the 1992/93 Household Health Survey, and an Environmental Tobacco Smoke Survey which has twice been conducted for the Ministry of Health.

The Central RHA supplied some questions about people's attitudes to, and difficulty finding information about, health services, and about environmental health.

The demographic and socioeconomic questions were taken from the Census and/or the Household Labour Force Survey, in order to give comparability with the Census, other major social surveys, and the 1992/93 Household Health Survey.

Child Questionnaire

This questionnaire covered a subsection of the topics listed for the adult questionnaire:

- health care services used, satisfaction and unmet need
- prescription items
- injury or poisoning
- insurance
- health cards

The questions were asked of the parent/caregiver about their child. A question on disability was also asked about the child. The child was not asked to participate. The selected respondent was allowed to ask another primary caregiver in the household to help answer questions about the child, if some detail was not known by the selected respondent.

General Health Questionnaire

This questionnaire was made up of the following:

- the disability 'filter' questions from the Census;
- the SF-36 Health Survey on self-reported health status, an internationally used survey;
- the World Health Organization AUDIT questionnaire on alcohol use disorders.

All respondents aged 15 years and over were asked about any disabilities they have, their views about their health, how they feel, how well they are able to do their usual activities, and their alcohol consumption. This was a self-completion questionnaire. It was not asked of children less than 15 years of age. If for any reason the respondent could not read, then the interviewer could administer this questionnaire. Other household members were not asked to complete or interpret the General Health Questionnaire for the respondent.

The data collected in this questionnaire will enable the self-perceived health status of New Zealanders to be compared with that of people in other countries.

Collection

The questionnaires and procedures for this survey were field tested. In addition, the questionnaires were put through a series of cognitive tests. The questionnaires received approval from the Ministry of Health.

Interviewing for the survey which commenced on 14 October 1996 continued for one year, since seasonality is important in analysis of health issues. The interviews were conducted on an ongoing basis, excepting a three-week break around the Christmas holiday season, and a one-week break in April.

Consent Forms

Selected respondents were asked to sign two consent forms. Proxy signatures were not allowed, except if another member of the household had signing authority for a respondent who had a disability.

Health Consent
The New Zealand Health Survey Consent Form had two purposes:
- to ensure that respondents gave their informed consent to participate in the survey;
- to get the respondents' permission for SNZ to pass to the regional health authorities and the Ministry of Health coded answers to the survey, i.e. unit record data. No names, addresses, or any identifiable information are released.

The selected respondent was asked to sign this form before the interview started.

Nutrition Consent

At the end of the interview, the respondent was asked to sign a consent form which would give SNZ their permission to release the respondent's name and address from the household form to the University of Otago, who conducted the NNS. The university then undertook the process of asking people to participate in the NNS.

Collection Method

The interviewing was done face-to-face by trained SNZ interviewers, with an additional self-completed questionnaire component. No proxies were allowed, except in the cases specified above.

Follow-up Procedures

Interviewers were expected to make three contact attempts, at different times of the day and week, at each selected address, unless absence during the survey period was verified by a neighbour or it was clear the address did not exist. As well, two additional attempts were made to get the selected person.

Processing

The following is an overview of the processing steps for the 1996/97 Health Survey.

Data Capture

Data from the survey questionnaires were entered directly into computers in SNZ's Wellington office. The data capture program, which used Blaise software, allowed for a valid range of codes for each question and automatically followed the flow of the questionnaire.

Edit and Imputation

All survey records were subjected to computer edits to identify and correct invalid or inconsistent information. The variables required for weighting (age, sex and ethnic group) were imputed if missing. This was accomplished by going back to the original questionnaires. Records with other missing information were not imputed. For all other questions, in the event the respondent was unwilling or unable to answer, "refused" and "do not know" codes were used, as appropriate.

Classifications and Definitions

The survey used SNZ standard classifications related to income, education qualifications, occupation, labour force status, ethnic group, age, sex, marital status, family and household type, and number of rooms in dwelling. This allows the survey results to be integrated with data from other surveys.

Coding

Several questions allowed "write-in" responses. In the case of questions that had an "other, specify" category (for example, Question 6 in the Adult Questionnaire), each "write-in" response was examined and coded back to an existing code if applicable.

A number of questions (e.g. Question 15) allowed for multiple responses. For these "tick all that apply" categories, all responses were retained with each option shown as a separate variable on the data file.

Creation of Combined and Derived Variables

A number of variables on the computer dataset were derived by using items found in various parts of the questionnaires. In some cases, the derived variables are straightforward and involve collapsing of categories. In other cases, several variables have been combined to create a new variable. The data dictionary provides comments indicating the origin of these variables.

Amount of Detail on the Unit Record Data File

In order to guard against identifying an individual or any breach of confidentiality, the amount of detail included on this file is less than is available on the master file retained by SNZ. No identifiable information is included. All names, addresses, telephone numbers or any other identifiable information of respondents have been eliminated from this file.

Estimation

Calculation of Selection Probabilities and Sample Weights

General Principles: unstandardised weights

Each responding adult and child has been assigned a unique survey weight to be used in the calculation of survey estimates called an 'unstandardised weight'.

The two most important functions of these allocated weights are:
- to produce unbiased survey estimates by taking account of the varying probabilities of selection amongst members of the sample population.
- to 'rate up' the sample data to total population size, enabling simple calculation of estimates of population counts with given health characteristics of interest.

As a general guide to this process, three steps were involved in producing the final unstandardised weight assigned to each individual respondent at the conclusion of the 1996/97 Health Survey. These were:

1. determining the respondent's probability of selection;
2. adjustment for non-response;
3. post-stratification.

Probability of Selection: Each individual respondent was first assigned a design weight which represents the inverse of their probability of selection into the sample.

Non-response adjustment factor: This design weight was then adjusted to take account of the level of non-response among selected individuals in the same region.

The increase in respondents' design weights was carried out by calculating a factor RESPFACT, where

$$RESPFACT = \frac{\sum \text{estimated eligible individuals}}{\sum \text{eligible responding individuals}}$$

This response adjustment factor is calculated as:

$$RESPFACT = 1 + (C/D) + (E/D)*(C+D)/(B+C+D)$$

where B = sum of weights of households ineligible post-contact
C = sum of weights of households eligible non-responding
D = sum of weights of households eligible responding
E = sum of weights of households with unknown eligibility.

Response factors were calculated and applied to design weights within quarters at RHA level. The adjusted design weight is referred to as the response weight.

Post Stratification: The final stage of unstandardised weight adjustment scales response weights so that the weighted breakdown of the responding sample is consistent with SNZ usual resident population estimates for age, sex and RHA breakdowns as at March 31 1997.

Survey weights included in the final unit record files are consistent with March 1997 usual resident population estimates for:
- RHA (Northern, Midland, Central , Southern) *by*
- Age (0-4, 5-9, 10-14, 15-19, …, 75-79, 80+) *by*

- Sex (Male, Female)

General Principles: standardised weights

In addition to the unstandardised weight for a respondent, several sets of standardised weights were also developed. Each of these standardised weights is tailored for a particular set of analyses. Standardised weights are formed directly from unstandardised weights, and each such set of weights is scaled to allow better comparison of particular subpopulations. For example, for each sex, in the comparison of smoking preference between ethnic groups, the effect of different age structure within each ethnic group confounds the analysis if unstandardised weights are used, since differences between ethnic groups may simply be a consequence of differences in the age structure of the different ethnic subpopulations for each sex. Standardised weights avoid this problem by rescaling the standardised weights so that each ethnic group is treated as having the same age structure within each sex, ie, equivalent to the age structure for each sex for the New Zealand population as a whole. Any differences then found between ethnic groups, for either sex, using such a set of standardised weights, then cannot be due to differences in the age structure because this effect has been 'standardised' out.

Estimation for the Adult Population: unstandardised weights

In the 1996/97 Health Survey, only adult respondents who completed ninety percent of the core adult health questions are included on the final unit record data set and used in estimation. To identify these respondents, and to assist in adjustment for non-response, each selected household is first assigned an adult eligibility status.

Adult Household Eligibility

The 5 adult eligibility statuses are:

1. Ineligible Pre-contact
2. Ineligible Post-contact
3. Eligible Non-responding
4. Eligible Responding
5. Unknown Eligibility

The set of adult respondents used in estimation are those whose household is assigned the 'eligible responding' status.

Eligibility Rules

Rules differ for main sample and Maori and Pacific Island oversample households.

Main sample households

Adult Eligibility Status	HHPC	Adult Core Qns Completed
Ineligible Pre-contact	09	-
Ineligible Post-contact	08	-
Eligible Non-responding	20,22,40	-
	03,06	less than 90%
Eligible Responding	01	-
	03,06	90% or more
Unknown eligibility	02,21,04,07	-

Maori/PI oversample households

Adult Eligibility Status	HHPC	Adult Core Qns Completed
Ineligible Pre-contact	-	-
Ineligible Post-contact	-	-
Eligible Non-responding	20,22,40,08,09, 02,21,04,07	-
	03,06	less than 90%
Eligible Responding	01	-
	03,06	90% or more
Unknown Eligibility	-	-

(refer to section below for Household Participation Codes (HHPC))

Assigning Adult Weights

<u>Probability of Selection</u>

Each adult respondent is assigned a design weight which reflects their probability of selection into the survey. This selection probability depends upon the stratum and PSU of the respondent household and the respondent's ethnicity (Maori/ non-Maori).

<u>Maori</u>

Adult Maori have two chances of selection into the survey sample:

    through the main sample in any selected main sample PSU, OR
    through the Maori oversample selected in a set of Maori oversample PSUs.
For adult maori **k** ,in household **j** in psu **i** and stratum **h**, probability of

selection $\left(p_{hijk}\right)$ into the sample is:

$$p_{hijk} = \frac{n_{nzhs,h}}{N_h}\left(\frac{1}{K_{hi} * APF_{hi} * Adult_{hij}}\right) + \frac{n_{maor,h}}{N_h}\left(\frac{1}{K_{hi} * MPI_{hij} * HLFSFACT_q}\right)$$

where N_h = Total no. of PSUs in stratum h

$n_{nzhs,h}$ = No. of PSUs in stratum h selected for 1996/97 Health Survey main sample

$n_{maor,h}$ = No. of PSUs in stratum h selected for maori oversample

K_{hi} = No. of household panels in PSU hi

APF_{hi} = $\dfrac{\text{No. of households in panel in PSU hi}}{\text{No. of households selected for main sample in PSU hi}}$

$Adult_{hij}$ = No. of adults in household hij

MPI_{hij} = No. of Maori and Pacific island adults in household hij

$HLFSFACT_q$ = inverse of HLFS national response rate in qtr q

q = quarter from which PSU hi Maori oversample selected

The selection probability (p_{hijk}) for each adult maori is the sum of their selection probabilities into each of the main sample and maori oversamples.

<u>Pacific Islanders</u>

Pacific Islanders also have two chances of selection into the survey sample:

through the main sample in any selected main sample PSU, OR
through the Pacific Island oversample selected in a set of Pacific Island oversample PSUs.

For adult PI **k** ,in household **j** in psu **i** and stratum **h**, probability of selection (p_{hijk}) into the sample is:

$$p_{hijk} = \frac{n_{nzhs,h}}{N_h}\left(\frac{1}{K_{hi} * APF_{hi} * Adult_{hij}}\right) + \frac{n_{pi,h}}{N_h}\left(\frac{1}{K_{hi} * MPI_{hij} * HLFSFACT_q}\right)$$

where $n_{pi,h}$ = No. of PSUs in stratum h selected for PI oversample

(N.B. $n_{pi,h} = n_{nzhs,h}$ in all strata h)

q = quarter from which PSU hi PI oversample selected

The selection probability (p_{hijk}) for each adult Pacific Islander is the sum of their selection probabilities in to the main sample and Pacific Island oversample.

<u>Others (Non-Maori/ Non-PI)</u>

Non-Maori/Non-PI have only one chance of selection into the survey sample:

through the main sample in any selected 1996/97 Health Survey PSU

For non-maori/non-pi adult **k**, in household **j** in psu **i** and stratum **h**, the probability of selection (p_{hijk}) into the sample is:

$$p_{hijk} = \frac{n_{nzhs,h}}{N_h} \left(\frac{1}{K_{hi} * APF_{hi} * Adult_{hij}} \right)$$

Adult Design Weight

The design weight for each adult respondent is then assigned as the inverse of their selection probability:

For adult k, in household j, psu i, stratum h:

DSGNWGT$_{hijk}$= $\dfrac{1}{p_{hijk}}$.

Adult Household Weights

Household weights are calculated and used to adjust for survey non-response. For each type of sample household selected in the 1996/97 Health Survey (Main sample, Maori oversample, PI Oversample), the household weight reflects the inverse probability of the household being selected into this sample.

For a household in PSU **i** in stratum **h**:

- if selected into Main Sample

$$HHDWGT_{hij} = \frac{N_h}{n_{nzhs}} * K_{hi} * APF_{hi}$$

- if selected into Maori Oversample

$$HHDWGT_{hij} = \frac{N_h}{n_{maor,h}} * K_{hi} * HLFSFACT_q$$

- if selected into Pacific Island Oversample

$$HHDWGT_{hij} = \frac{N_h}{n_{pi,h}} * K_{hi} * HLFSFACT_q$$

Adult Non-Response Adjustment

<u>Adult Response Factor</u>

In order to adjust for adult non-response in the 1996/97 Health Survey a response factor is calculated and used to adjust adult design weights. Response factors are calculated for ethnic groups (Maori, PI, Other) by RHA by survey quarter (1-4).

Response factors for Maori and Pacific Islanders within RHA* Survey quarter cells are calculated differently from the corresponding Non-Maori/non-PI factor.

<u>Maori (PI) response factor</u>

The response factor for Maori(PI) takes account of non-response within both the main sample and the Maori (PI) oversample. Separate response factors are first calculated for the main sample and the Maori (PI) oversamples, with the final response factor being calculated as a weighted average of these two values.

For RHA * Survey quarter cell c, response factor for Maori (PI) is:

$$RESPFACT_c = \frac{(M_{c,main} * \mathrm{Re}\,spfact_{c,main}) + (M_{c.over} * \mathrm{Re}\,spfact_{c,over})}{M_{c,main} + M_{c,over}}$$

where:

$M_{c,main}$= sum of DSGNWGT for eligible responding Maori (PI) selected into the main sample in RHA * Survey qtr cell c

$M_{c,over}$= sum of DSGNWGT for eligible responding Maori (PI) selected into the Maori (PI) oversample in RHA * Survey qtr cell c

$Respfact_{c,main}$= response factor calculated to take account of Maori (PI) non-response in the main sample in RHA * Survey qtr cell c

$$= 1 + C_1/D_1 + (E_1/D_1) + (C_1+D_1)/(B_1+C_1+D_1)$$

where:

A_1=sum of HHDWGT for all households ineligible pre-contact in main sample in RHA*Qtr cell c (N.B. A_1 not used in respfact)

B_1=sum of HHDWGT for all households ineligible post-contact in main sample in RHA*Qtr cell c

C_1=sum of HHDWGT for all households eligible non-responding in main sample in RHA*Qtr cell c

D_1=sum of HHDWGT for all households eligible responding in main sample in RHA*Qtr cell c

E_1=sum of HHDWGT for all households with unknown eligibility in main sample in RHA*Qtr cell c

Respfact$_{c,over}$= response factor calculated to take account of Maori (PI) non-response in the Maori (PI) oversample in RHA * Survey quarter cell c

$$= 1 + C_2/D_2 + (E_2/D_2) + (C_2+D_2)/(B_2+C_2+D_2)$$
$$= 1 + C_2/D_2 \ (\text{as } E_2=0 \text{ for oversample households})$$

where :

A_2=sum of HHDWGT for all households ineligible pre-contact in Maori(PI) oversample in RHA*Qtr cell c

B_2=sum of HHDWGT for all households ineligible post-contact in Maori(PI) oversample in RHA*Qtr cell c

C_2=sum of HHDWGT for all households eligible non-responding in Maori(PI) oversample in RHA*Qtr cell c

D_2=sum of HHDWGT for all households eligible responding in Maori(PI) oversample in RHA*Qtr cell c

E_2=sum of HHDWGT for all households with unknown eligibility in Maori(PI) oversample in RHA*Qtr cell c

Non-Maori/Non-PI response factor

The response factor for Non-Maori/ Non-PI takes account of non-response within the main sample only.

For RHA * Survey quarter cell c, response factor for Non-Maori/Non-PI is:

$RESPFACT_c = \mathrm{Re}\,spfact_{c,main}$ (as calculated above)

$$= 1 + C_1/D_1+ (E_1/D_1) + (C_1+D_1)/(B_1+C_1+D_1)$$

where:

A_1=sum of HHDWGT for all households ineligible pre-contact in main sample in RHA*Qtr cell c (N.B. A_1 not used in respfact)

B_1=sum of HHDWGT for all households ineligible post-contact in main sample in RHA*Qtr cell c

C_1=sum of HHDWGT for all households eligible non-responding in main sample in RHA*Qtr cell c

D_1=sum of HHDWGT for all households eligible responding in main sample in RHA*Qtr cell c

E_1=sum of HHDWGT for all households with unknown eligibility in main sample in RHA*Qtr cell c

Adult Response Factors - Derivation

<u>Main/Oversample Response Factors</u>

$\text{Respfact}_{c,\text{main}}$ / $\text{Respfact}_{c,\text{over}}$

These response factors are given by

$$R = 1 + (C/D) + (E/D) \times (C+D)/(B+C+D)$$

where:
A=sum of HHDWGT for all households ineligible pre-contact in main/oversample
B=sum of HHDWGT for all households ineligible post-contact in main/oversample
C=sum of HHDWGT for all households eligible non-responding in main/oversample
D=sum of HHDWGT for all households eligible responding in main/oversample
E=sum of HHDWGT for all households with unknown eligibility in main/oversample

Rewrite the expression as

$$R = \frac{D + C + E \times \dfrac{C+D}{B+C+D}}{D}$$

$B + C + D$ is the total weight for hholds whose adult eligibility status we know.

$C + D$ is the total weight for those hholds we know have an eligible adult status.

So $E \times \dfrac{C+D}{B+C+D}$ estimates the eligible hhlds among those of unknown eligibility, assuming unknown eligibility hholds are like those we know about.

So the numerator estimates the total number of eligible hhlds. The denominator estimates the total eligibles responding. The adult response factor boosts adult design weights by this ratio.

<u>Maori (PI) Response Factor</u>

This response factor is given by:

$$RESPFACT_c = \frac{(M_{c,main} * \mathrm{Re}\, spfact_{c,main}) + (M_{c.over} * \mathrm{Re}\, spfact_{c,over})}{M_{c,main} + M_{c,over}}$$

where:

$M_{c,main}$= sum of DSGNWGT for eligible responding Maori (PI) selected into the main sample in RHA * Survey quarter cell c

$M_{c,over}$= sum of DSGNWGT for eligible responding Maori (PI) selected into the Maori (PI) oversample in RHA * Survey quarter cell c

Now, $M_{c,main}$ is the weighted number of eligible responding Maori (PI) in the main sample, $Respfact_{c,main}$ is an estimate of one over the proportion of eligible households containing Maori(PI) in the main sample which respond.

So $(M_{c,main}* Respfact_{c,main})$ is an estimate of the total number of eligible Maori (PI) in the main sample.

Similarly, $M_{c,over}$ is the weighted number of eligible responding Maori (PI) in the oversample, $Respfact_{c,over}$ is an estimate of one over the proportion of eligible households containing Maori(PI) in the over sample which respond.

So $(M_{c,over}* Respfact_{c,over})$ is an estimate of the total number of eligible Maori (PI) in the oversample.

The numerator is an estimator of the total number of eligible Maori (PI) in main and oversamples. The denominator is the weighted count of eligible responding Maori (PI) in both samples. The Maori (PI) response factor is the rate-up factor representing this ratio.

Adult Response Weight

For adult respondent hijk in RHA * survey quarter c, a response weight is calculated which reflects the sample selection weight adjusted for non-response:

$RESPWGT_{hijk}$= DSGNWGT$_{hijk}$* RESPFACT$_c$

Adult Post-Stratification

The final stage of the adult weighting process is a post-stratification adjustment which ensure that the final weighted totals of eligible adult respondents are consistent with independent population estimates for RHA by Age group * Sex breakdowns. This process improves the overall accuracy of survey estimates.

Post-stratification factors are calculated for RHA * Sex * 5-yr age groups.

For adult respondent hijk in RHA*Sex*Agegrp cell g:

$$\textbf{PSFACT}_{\textbf{g}} = P_g \; / \; 1996/97 \text{ Health Survey}_g$$

where P_g = population estimate within RHA*Sex*Agegrp **g**

        1996/97 Health Survey$_g$ = the sum of RESPWGT for all adult respondents
in RHA*Sex*Agegrp g

Population estimates apply to NZ usual residents from private dwellings and are
point estimates for March 31 1997 - the midpoint of the survey period.

Adult Final Unstandardised Weight

The final weight calculated for each selected adult k in 'eligible responding'
household hij and RHA* Sex*Age group g is:

$$\textbf{AFINWGT}_{\textbf{hijk}} = \text{RESPWGT}_{hijk} * \text{PSFACT}_g$$

Estimation for the Child Population: unstandardised weights

In the 1996/97 Health Survey, only selected children whose selected caregiver
completes ninety percent of the core child health questions are included in
estimation. Each household selected to provide child health data (we will call these
'child' households) is assigned to child eligibility status to assist in determining the
set of children used in estimation and to assist in weight adjustment for non-
response.

Child Household Eligibility

Each 'child' household is assigned one of 6 child eligibility statuses:

1. Ineligible Pre-contact
2. Ineligible Post-contact
3. Eligible - selected adult not a caregiver to child under15
4. Eligible Non-responding
5. Eligible Responding
6. Unknown Eligibility

The set of children used in estimation are those whose household is assigned the
child 'eligible responding' status.

Eligibility Rules

Only a subset of households selected to participate in the 1996/97 Health Survey are also asked to provide child data. Those households not selected to provide child data are ignored in the weighting procedure for children.

For those households selected to provide child data, the eligibility rules are outlined in the following table:

Child Eligibility Status	HHPC	Child under 15 in hhld	Selected adult a caregiver to child	Child Core Qns completed
Ineligible Pre-contact	09	-	-	-
Ineligible Post-contact	01,20,22,03, 40,06	No	-	-
	08	-	-	-
Eligible - adult not caregiver	01,20,22,03, 40,06	Yes	No	-
Eligible - Non-responding	20,22,40	Yes	Yes	-
	03,06	Yes	Yes	less than 90%
Eligible - Responding	01	Yes	Yes	-
	03,06	Yes	Yes	90% or more
Unknown Eligibility	02,,21,04,07	-	-	-

(refer to Section 9.2 for Household Participation Codes (HHPC))

Assigning Child Weights

Probability of Selection

For child k, in household j, PSU i, stratum h, probability of selection (p_{hijk}) into the child sample is:

$$p_{hijk} = \frac{n_{child,h}}{N_h} * \frac{1}{K_{hi} * CPF_{hi} * CISW_{hijk}}$$

where N_h = Total no. of PSUs in stratum h

$n_{child,h}$ = No. of PSUs in stratum h selected to provide child data in the 1996/97 Health Survey

K_{hi} = No. of household panels in PSU hi

CPF_{hi} = $\dfrac{\text{No. of households in panel in PSU hi}}{\text{No. of households selected for child sample in PSU hi}}$

$CISW_{hijk}$ = child individual selection weight
= inverse probability of selection of child k within household hij

$$= \frac{1}{\frac{1}{Adult_{hij}} \left(\sum_{a \in P} \frac{1}{Child_a} \right)}$$

and Adult$_{hij}$= number of adults eligible for selection in household hij

P = set of adults in household hij who are a parent/caregiver to child k

Child$_a$= number of dependent children under 15 of adult a in household hij

Family and age codes on the household form are used to determine set P and values of Child$_a$.

Child Design Weight

Design weight for each selected child is then assigned as the inverse of selection probability:

For child k, in household j, psu i, stratum h:

DSGNWGT$_{hijk}$= $\dfrac{1.}{p_{hijk}}$

Child Household Weight

Child household weights are calculated for each household selected to provide child data. These are used to adjust for survey non-response within the child population.

For child household **j** in PSU **i** in stratum **h**:

$$HHDWGT_{hij} = \frac{N_h}{n_{child,h}} * K_{hi} * CPF_{hi}$$

This value reflects the household's probability of selection into the 1996/97 Health Survey child sample.

Child Non-Response Adjustment:

<u>Child Response Factor</u>

Child response factors are calculated within RHA * Survey quarter and used to adjust child design weights to take account of non-response within selected child households.

For RHA * Survey quarter cell c, child response factor is calculated as:

$$RESPFACT_c = \frac{D + E + C + F * \left(\dfrac{D + E + C}{D + E + C + B} \right)}{(D + E + C) * \left(\dfrac{E}{D + E} \right)}$$

where   A=sum of HHDWGT for all households ineligible pre-contact in RHA*Qtr cell c

B=sum of HHDWGT for all households ineligible post-contact in RHA*Qtr cell c

C=sum of HHDWGT for all households eligible-adult not parent in RHA*Qtr cell c

D=sum of HHDWGT for all households eligible-non responding in RHA*Qtr cell c

E=sum of HHDWGT for all households eligible responding in RHA*Qtr cell c

F=sum of HHDWGT for all households with unknown eligibility in RHA*Qtr cell c

Child Response Factor - Derivation

Child response factor is:

$$RESPFACT_c = \frac{D + E + C + F * \left(\dfrac{D + E + C}{D + E + C + B} \right)}{(D + E + C) * \left(\dfrac{E}{D + E} \right)}$$

where:

 A=sum of HHDWGT for all households ineligible pre-contact
 B=sum of HHDWGT for all households ineligible post-contact
 C=sum of HHDWGT for all households eligible-adult not parent
 D=sum of HHDWGT for all households eligible-non responding
 E=sum of HHDWGT for all households eligible responding
 F=sum of HHDWGT for all households with unknown eligibility

Now (D+E+C)/(D+E+C+B) is the proportion of households we expect to contain children.
Therefore, F*(D+E+C)/(D+E+C+B) is an estimate of those households of unknown eligibility which contain children.
As (D+C+E) is the number of households we know to contain eligible children., then the numerator:

$$D + E + C + F * \left(\frac{D + E + C}{D + E + C + B} \right)$$

is an estimate of the total number of households containing eligible children.

D is the number of households which provide child data for the 1996/97 Health Survey.

(D+E+C)/(D+E) is the inverse of the proportion of households known to contain children where the selected adult is a parent/caregiver to an eligible child. This rate-up factor is implicitly included in the design weights applied to child data (through CISW).

Therefore, the denominator:

$$(D + E + C) * \left(\frac{E}{D + E} \right)$$

is an estimate of the number of child households represented by design weights of child respondents.

Response Factor is therefore an estimate of the ratio:

<u>   total number of households containing eligible children   </u> .
     number of child households represented by design wgts of responding children

Child Response Weight

For child respondent hijk in RHA * survey quarter c, response weight is calculated as:

RESPWGT$_{hijk}$= DSGNWGT$_{hijk}$* RESPFACT$_c$

Child post-stratification

The final stage of the child weighting process is a post-stratification adjustment which ensures that the final weighted totals of eligible child respondents are consistent with independent population estimates for RHA by Age group * Sex breakdowns. This process improves the overall accuracy of survey estimates.

Post-stratification factors are calculated for RHA * Sex * 5-yr age groups.

For child respondent hijk in RHA*Sex*Agegrp cell g:

PSFACT$_g$= P$_g$ / 1996/97 Health Survey$_g$

where   P$_g$ = SNZ total population estimate within RHA*Sex*Agegrp **g**

$$1996/97 \text{ Health Survey}_g = \text{the sum of RESPWGT for all adult respondents in}$$
$$\text{RHA*Sex*Agrgrp g}$$

Child Final Weight

The final weight calculated for each selected child k in 'eligible responding' household hij and RHA* Sex*Age group g is:

$$\mathbf{CFINWGT_{hijk}} = \text{RESPWGT}_{hijk} * \text{PSFACT}_g$$

Estimation for the Adult and Child Population: standardised weights

There are nineteen sets of standardised weights. These weights standardise each of 17 variables (1 child, 16 adult) to conform to the age structure and the sex structure of the NZ population. There is also two sets of weights, one for children the other for adults, standardised by age and sex only (ie, not by age*sex or by any other variable).

These 19 sets of standardised weights allow better comparison for each of the 17 variables (plus age and sex comparisons based on the age + sex standardised weights, for children and adults) because comparisons are then based on the subpopulations formed by these variables standardised to the NZ age and the NZ sex population structure for March 1997 (which is itself based on the projection from SNZ from the 1996 NZ Census of population and dwellings);

All standardised weights are standardised by age group and by sex, but not the interaction between agegroups within sex (ie, agegroup*sex interaction); adult variables are standardised by adult agegroup and sex, child variables by child agegroup and sex.

The 17 additional variables used for standardisation are child ethnic group (c_ethgrp), and adult ethnic group (a_ethgrp), marital (status), lfstatus, educ, smoking, asthma, A4, A7, hlthfam, urb, dur_act1, A64, A66_1, A83C, audit, g1. Definitions of these variables can be found in "1996-97 New Zealand Health Survey: Unit Record Data File Documentation and User's Guide". The 17 corresponding weights in the data file are casethwt, aasethwt, aasmarwt, aaslfswt, aaseduwt, aassmowt, aasastwt, aasa4wt, aasa7wt, aashfmwt, aasurbwt, aasda1wt, aasa64wt, aas661wt, aasa83wt, aasaudwt, aasg1wt. The two variables standardised by age group and sex only are caswt for children and aaswt for adults.

Each of these sets of standardisation weights were calculated via the PROC IML procedure of SAS, using a general iterative proportional fitting algorithm. The SAS program used a macro designed to fit a complete range of implicit hierarchical loglinear models (including those involving interactions of any order) based on the specified standardisation variables. (There were two or three such standardisation variables for each set of standardised weights.) All implicit loglinear models used for the 1996/97 Health Survey were complete independence models (ie, no interactions were fitted). For a discussion of Loglinear models see Agresti (1990). Where

standardisation is not possible because of some age or one sex has zero respondents in some subgroup, that particular set of standardised weights are all set to zero.

Response Criteria

The criteria for determining whether incomplete questionnaires were treated as responses or non-responses were as follows:

- *Response*: Questionnaire incomplete but 90% of the specified core health questions was completed, where a specified core question was defined as a question that was to be answered by all respondents, i.e. a skip out was not allowed.
- *Non-response*: No questions completed or the 90% criteria was not met.

Household Participation Codes

Code 01 *Full response*	This code is given when all the questionnaires (Adult, Child and General Health) have been fully completed for the selected respondent.
Code 02 *Full refusal* *(Unknown eligibility)*	This code is given when the <u>household</u> or initial contact person, refuses to help you complete the household form.
Code 20 *Full refusal* *(Eligibility established)*	This code is given when you have completed the household form, and the <u>selected person</u> refuses to take part in the 1996/97 Health Survey.
Code 21 *Full refusal* *(previous/house hold)*	This code is given when the <u>household </u>or an initial contact person refused to do the survey, because they had previously done another SNZ.
Code 22 *Full refusal* *(previous/respondent)*	This code is given when you have completed the household form, and the <u>selected person</u> refused to do the survey, because they had previously done another SNZ survey.
Code 03 *Part Refusal*	This code is given when the selected respondent has: completed some, but not all the required questionnaires e.g. they did the adult questionnaire, but refused the child. OR they have refused to fully complete a questionnaire, e.g. they refused to answer the smoking questions in the

adult questionnaire.

Code 04 *Full Non-* *Contact/* *(unknown* *eligibility)*	This code is given when despite making at least 4 visits at different times (morning, afternoon, evening), it has not been possible to establish contact with anyone at the selected dwelling.
Code 40 *Full Non-* *Contact* *(Eligibility* *established)*	This code is given when despite making 2 additional visits to the household, it has not been possible to establish contact with the selected respondent.
Code 06 *Part Non-* *Contact*	This code is given when the selected respondent has answered some of the required Health questionnaires, but could not be recontacted to complete the survey e.g. they did the adult questionnaire and arranged to do the child questionnaire at another time, but could not be recontacted.
Code 07 Death *or Illness*	This code is used when unforeseen circumstances (for example serious illness or death in the household) would make it inappropriate to conduct an interview during the survey period.
Code 08 *All persons out* *on scope*	This code is used when all the residents at the selected dwelling are out of scope for the survey e.g. they are all overseas visitors OR this is a Maori/ Pacific Island interview, but there are no Maori or Pacific Island residents in the household. See scope rules on page 43.

Code 09
Ineligible
Dwellings
	This code is a combination of the following HLFS codes:
	    09 Vacant dwelling
	    10 Dwelling under construction
	    11 Dwelling converted to non-dwelling
	    12 Derelict dwelling
	    13 Dwelling demolished or removed
	    15 Converted to a Non-private dwelling

Statistical Analysis

Estimation of standard errors for sample survey estimates

The statistical analysis of the 1996/97 Health Survey was carried out using both SAS (SAS Institute Inc, 1989; SAS Institute Inc, 1990) and Sudaan (Shah et al, 1997).

Preliminary analyses were carried out in SAS using the appropriate standardised or unstandardised sample survey weights. Sudaan was then used to calculate the accuracy of these estimates initially derived using SAS, and to provide accurate standard errors for the estimates.

In general, SAS procedures are unable to provide the correct accuracy measures (eg variance, standard errors, confidence intervals) for complex survey data such as the 1996/97 Health Survey. Sudaan makes the correct adjustments for the complex survey design once the survey design is correctly specified as part of the Sudaan program.

The extent to which the variance of an estimate is inflated or deflated due to a complex survey design is called a design effect. A parallel definition is used for the inflation or deflation of a test statistic (eg a chi-square test) due to a survey design. See, for example, Sarndal, Swennson, and Wretman (1992).

Sudaan was used to analyse categorical, Likert scale, binary, and continuous variables from the 1996/97 Health Survey. Variance calculations and test statistics were calculated for the survey design treated as a sampling 'without replacement' design. However at the final stage of sampling within households only one person was selected, so that joint selection probabilities of people within households was zero. Consequently no unbiased estimator of variance exists (see for example, Haslett, 1985), although an estimate that overestimates variances for positively valued variables, such as those collected in the 1996/97 Health Survey, does exist. (eg, rates, proportions). The consequent variance inflation is likely to be small for rates and proportions for the 1996/97 Health Survey, because the zero joint selection probabilities occur at the last stage of sampling. As a check on this assumption all tests were also run treating the design as a 'with replacement' design, ie, a design which would allow for re-selection of units at each stage of sampling. Because with replacement designs do not incorporate finite population corrections which generally reduce variance, using with replacement variance estimation generally inflates variance estimates. The effect for the 1996/97 Health Survey was small, usually in the region of 1-2% relative (ie the ratio of the with replacement variance estimate to the without replacement variance estimate was 1.01 to 1.02). Since with replacement variance estimation was, in this sense, generally conservative, and since it avoided problems with zero joint selection probabilities, variance estimates given in this report are calculated on a with replacement basis. Note that estimates themselves (eg, rates, proportions, SF-36 scores) are completely unaffected by this choice of variance estimation method.

Bootstrap Confidence Intervals

For rates and proportions, especially when these are small (or in the case of proportions near to one), symmetry of confidence intervals based on the Central Limit Theorem cannot be guaranteed even for samples as large as the 1996/97 Health Survey. The symmetry of confidence intervals was checked for all such child and adult variables detailed in this report using the bootstrap (Efron and Tibshirani, 1992). In essence the bootstrap uses the sample actually drawn, and by drawing a large number of samples from the original sample using with replacement sampling is able to produce a distribution which, after some adjustment, is able to provide estimates of confidence intervals. The simplest such bootstrap technique, sometimes called the naïve bootstrap, is to draw say 100,000 such samples, order the estimates from these 100,000 replications by the value of the estimate and choose the 2,500[th] and 97,500[th] as indicating the 95% confidence bounds. Although such a technique can provide asymmetric confidence intervals, unlike the Central Limit Theorem approach of estimate plus or minus 1.96 times the standard error, there may be non-negligible bias when this particular bootstrap is used. For this reason various smoothing and bias adjustment techniques for the empirical distribution of the 100,000 bootstrap estimates have been developed. The one used here was generally the ABC or BC$_a$ technique. See Efron and Tibshirani (1992) for details.

For sample survey data, the bootstrap had to be adjusted so that the bootstrap sampled without replacement at each stage of sampling in the calculation of each of the 100,000 bootstrap estimates used. A variation of the naïve bootstrap was used when ABC or BC$_a$ lead to bias adjustments which were not due to bootstrap sampling but rather to the complex sampling scheme (especially the unequal survey weighting for rare characteristics), where if ABC or BC$_a$ was used, the parametric bootstrap adjustment was found to considerably overcompensate for bias.

The conclusion reached for all relevant variables from the 1996/97 Health Survey was that asymmetry of confidence intervals was insufficient to require the bootstrap. For this reason ninety five percent confidence intervals for estimates were calculated in *Taking the Pulse: The 1996/97 New Zealand Health Survey* using the conventional formula: estimate plus or minus 1.96 (or more conservatively two) standard errors.

Multivariate Analyses and Principal Component Factor Analysis

Analyses using both these statistical techniques were undertaken in SAS using the relevant sample survey weights. Multivariate analyses were also undertaken using Sudaan to give the correct standard errors and statistical tests of significance for regression coefficients and least square means. Standard errors for correlation coefficients used in the psychometric analyses of the SF-36 were calculated using a technique based on ratio estimation for complex surveys due to Fuller *et al* (1986).

Approximate Sampling Error Guidelines

Sampling Variability Guidelines

The estimates derived from this survey were based on a sample of households and individuals within households. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used. The variability of a survey estimate due to the random nature of the sample selection process is measured by its sample error.

Errors which are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered into Blaise and errors may be introduced in the processing and tabulation of the data. These are all examples of non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the survey. Quality assurance measures were implemented at each step of the data collection and processing cycle to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training of interviewers with respect to the survey procedures and questionnaire, observation of interviewers to detect problems of questionnaire design or misunderstanding of instructions, procedures to ensure that data capture errors were minimised, and coding and edit quality checks to verify the processing logic.

A major source of non-sampling error in surveys is the effect of non response on the survey results. The extent of non response varies from partial non response (failure to answer just one or some questions) to total non response. Total non response may occur because the interviewer was either unable to contact the selected household, the household refused to participate in the survey, or the selected respondent refused to participate in the survey. Total non response was handled by adjusting the weight of households who responded to the survey to compensate for those who did not respond.

In most cases, partial non response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, or could not recall the requested information.

Since it is an unavoidable fact that estimates from a sample survey are subject to sampling error, sound statistical practice calls for researchers to provide users with some indication of the magnitude of this sampling error.

Sample Error Estimates

Estimates of Proportions

The following table provides a guide to the level of sampling error present when estimating population proportions. The approximate sample error is calculated using the value of the proportion estimate and the size of the sample used to derive it - refer to Table 1.

Table 1: Sample Errors

Sample Size	Proportion Estimate								
	3%	**5%**	**7%**	**10%**	**15%**	**20%**	**30%**	**40%**	**50%**
70	5.7	7.2	8.5	9.9	11.8	13.3	15.2	16.2	16.6
120	4.3	5.5	6.5	7.6	9.0	10.1	11.6	12.4	12.7
150	3.9	4.9	5.8	6.8	8.1	9.1	10.4	11.1	11.3
200	3.3	4.3	5.0	5.9	7.0	7.8	9.0	9.6	9.8
300	2.7	3.5	4.1	4.8	5.7	6.4	7.3	7.8	8.0
500	2.1	2.7	3.2	3.7	4.4	5.0	5.7	6.1	6.2
700	1.8	2.3	2.7	3.1	3.7	4.2	4.8	5.1	5.2
1000	1.5	1.9	2.2	2.6	3.1	3.5	4.0	4.3	4.4
1500	1.2	1.6	1.8	2.1	2.6	2.9	3.3	3.5	3.6
2000	1.1	1.4	1.6	1.9	2.2	2.5	2.8	3.0	3.1
3000	0.9	1.1	1.3	1.5	1.8	2.0	2.3	2.5	2.5
5000	0.7	0.9	1.0	1.2	1.4	1.6	1.8	1.9	2.0
7000	0.6	0.7	0.8	1.0	1.2	1.3	1.5	1.6	1.7
8800	0.5	0.6	0.8	0.9	1.1	1.2	1.4	1.4	1.5

Note that for a proportion estimate p which exceeds 50%, the sample error is equivalent to that of proportion 1-p calculated using the same sample size.

Example - Using sample error guidelines to determine confidence intervals

Suppose 500 1996/97 Health Survey respondents are used to estimate that 20% of a given subpopulation are regular smokers. Using the sample error table, we see that this estimate has an approximate sample error of 5.0. This means that we can be 95% sure that the proportion estimate lies within 5 percentage points of the true proportion of regular smokers n this sub-population. That is, a 95% confidence interval on the true percentage of regular smokers is (15, 25).

Estimates of Population Counts

The following table provides a general guide to the relative sampling error (RSE) of population count estimates derived using the unit record file.

Table 2: Sample Errors for Population Counts

Estimate	Adults		Children		Total	
	Sampling Error	Relative Sampling Error (%)	Sampling Error	Relative Sampling Error (%)	Sampling Error	Relative Sampling Error (%)
1,000	2,000	200	3,100	310	2,200	220
2,000	2,900	145	4,400	220	3,100	155
3,000	3,500	117	5,400	180	3,800	127
4,000	4,100	103	6,200	155	4,400	110
5,000	4,500	90	6,900	138	4,900	98
7,000	5,400	77	8,200	117	5,800	83
10,000	6,400	64	9,700	97	6,900	69
15,000	7,900	53	11,900	79	8,400	56
20,000	9,100	46	13,700	69	9,700	49
50,000	14,300	29	21,200	42	15,300	31
100,000	20,000	20	29,100	29	21,500	22
200,000	27,700	14	38,300	19	30,000	15
300,000	33,300	11	43,100	14	36,200	12
500,000	41,300	8	44,400	9	45,400	9
650,000	45,500	7	38,100	6	50,500	8
800,000	48,700	6	20,800	3	54,600	7
1,500,000	54,000	4	-	-	65,000	4
2,000,000	49,200	2	-	-	65,900	3
2,400,000	38,700	2	-	-	63,000	3
2,600,000	29,300	1	-	-	60,200	2
2,800,000	10,100	0	-	-	56,300	2
3,300,000	-	-	-	-	40,000	1
3,600,000	-	-	-	-	18,400	1

Estimate cut-off point	Adults	Children	Total
for RSE=30%	45,300	94,700	52,200
for RSE=50%	16,500	36,700	19,000

Guidelines for Future Analysis of 1996/97 Health Survey Data

The 1996/97 Health Survey used a stratified design, with significant differences in sampling fractions between strata. Thus, some areas are over-represented in the sample (relative to their populations) while other areas are relatively under-represented; this means that the unweighted sample is not representative of the target populations.

The following general guidelines should be used when analysing data or creating tables:

- When the relative sampling error is 30% to less than 50%, then the data should be used with caution. This cell should also be noted in the table. The user could use brackets or a single asterisk to denote the cell; i.e. ( ) or *

- When the relative sampling error is 50% or greater, then the data is generally too unreliable for most practical purposes. This data cell should be noted differently from the previous, such as using a double asterisks; i.e. **

References

Agresti A .1990. *Categorical Data Analysis* New York: John Wiley and Sons.

Efron B, Tibshirani RJ. 1993 *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Fuller WA, Kennedy W, Schnell D, et al, 1986. *PC-CARP* Ames, Iowa: Statistical Laboratory, Iowa State University.

Haslett SJ. 1985. The Linear Non-Homogeneous Estimator in Sample Survey. *Sankhya* B47:101-117.

Sarndal CE, Swennson B, Wretman J. 1992. *Model Assisted Survey Sampling* New York: Springer-Verlag.

SAS Institute Inc. 1990. *SAS Procedures Guide* Version 6 (3rd ed). Cary, N.C: SAS Institute Inc.

SAS Institute Inc. 1989. SAS/STAT Users Guide, Vol 1 and 2 (4th ed), Cary, N.C: SAS Institute Inc.

Shah BV, Barnwell BG, Bieler GS. 1997. *Sudaan User's Manual Release 7.5.* Research Triangle Park, N.C: Research Triangle Institute.

Skinner C J, Holt D, Smith TMF (eds). (1989). *Analysis of Complex Survey Data.* New York: John Wiley and Sons.

Statistics New Zealand. 1997. *1996-97 New Zealand Health Survey: Unit Record Data File Documentation and User's Guide*. Unpublished document, prepared for the Ministry of Health and the Regional Health Authorities. Statistics New Zealand.