SoFIE-Health & NZHIS Data Integration Project

SoFIE-Health Report 4

Working Paper

Version 1.0

Authors: **Kristie Carter**
Michael Hayward
Ken Richardson

8 October 2009

Acknowledgements
SoFIE-Health is primarily funded by the Health Research Council of New Zealand as part of the Health Inequalities Research Programme. Establishment funding was also received from the University of Otago, Accident Compensation Corporation of New Zealand (ACC), and the Alcohol Liquor Advisory Council (ALAC). Comments on this report were received from Tony Blakely, Caroline Shaw and Statistics NZ.

Statistics New Zealand Security Statement
Access to the data used in this study was provided by Statistics New Zealand in a secure environment designed to give effect to the confidentiality provisions of the Statistics Act, 1975. The results in this study and any errors contained therein are those of the author, not Statistics New Zealand.

Copyright

Reference
Carter, K., Hayward, M., & Richardson, K. (2009). SoFIE-Health & NZHIS Data Integration project. SoFIE-Health Report 4. Working Paper. Wellington: University of Otago, Wellington.

Published by the Department of Public Health
University of Otago, Wellington
PO Box 7343
Wellington South 6242
New Zealand

Table of Contents

List of Figures

List of Tables

1 Introduction

This report contains the following information about the SoFIE-Health and NZHIS data integration project:

- Survey of Families, Income and Employment (SoFIE)
- SoFIE-Health module
- SoFIE-Health & NZHIS data linkage
- NZHIS data organisation and variables
- Initial results
- Appendices
    - SoFIE-Health consent and information sheet
    - Data dictionary
    - ICD10-AM chapter and subchapter headings

1.1 *Definitions*

UOW – University of Otago, Wellington.
NZHIS – New Zealand Health Information Service. Information Directorate, Ministry of Health.
OSM - Original Sample Member. At the first wave of the SoFIE, every eligible resident of the selected households became an OSM. These are the people who are interviewed over successive waves of the survey, regardless of whether or not they still reside in their original dwelling.
PQ - Personal Questionnaire. The individual questionnaire filled out by survey respondents.
HQ - Household Questionnaire. The questionnaire filled out by a nominated individual in the household.
HED - Household Enumeration Date. The date when all of the eligible members of the household are identified and recorded in the household questionnaire.
SAPD - Start of Annual Period Date is the reference date for the start of the annual period e.g. if a respondent's Wave 1 HED interview is on 01 December 2002, the SAPD is 01 December 2001 (12 months before the HED).
EAPD - End of Annual Period Date is the reference date for the end of the annual period for each respondent e.g. if a respondent's Wave 1 HED interview is on 01 December 2002, the EAPD is 30 November 2002.
PSU - Primary Sampling Unit. The geographical units or groups of units considered for selection in the first stage of a sampling process.
CAI - Computer-Assisted Interviewing. An interviewing technique that uses a computer to assist with data collection (usually face-to-face or self-interview).
CSR - Cross-sectional Respondent is someone who has moved into a household with an existing OSM, who is asked the PQ.
Point-in-time data – respondent data collected at a specific date (interview date).
Spell data – respondent data relating to a period with a defined start and end date.

1.2 *Getting more information*

Further information about the SoFIE study and links to other longitudinal studies can be found at http://www.stats.govt.nz/methods_and_services/completing-a-survey/individual-and-household-surveys/survey-of-family-income-and-imployment.aspx .

Information about the SoFIE-Health project is available at http://www.uow.otago.ac.nz/academic/dph/research/HIRP/SoFIE/SofieIndex.html

or by emailing the Principal Investigator (Kristie Carter) at kristie.carter@otago.ac.nz.

Information about the New Zealand Health Information Service is available at http://www.nzhis.govt.nz/.

2 Overview of SoFIE

2.1 *General Information*

Statistics New Zealand was granted funding from the Foundation for Research, Science and Technology in 1997 to conduct a feasibility study for a longitudinal survey of income, employment and family dynamics. Following on from the feasibility study, the SoFIE study was developed and first went into the field in October 2002.

SoFIE is a single fixed panel longitudinal survey with an expected maximum duration of eight years.[1] It collects information once a year from the same individuals on income levels, sources and changes; and on the major influences on income such as employment and education experiences, household and family status and changes, demographic factors and health status. Every two years (waves 2, 4, and 6) it collects information on assets and liabilities to monitor net worth and savings. Every other year (waves 3, 5, and 7) it collects detailed information on health.

The overall objective of SoFIE is to provide information about changes over time in the economic well-being of individuals and their families, and about factors influencing those changes.

2.1.1 Target population

The target longitudinal population for SoFIE is*:*

> *the usually resident population of New Zealand living in permanent, private dwellings on the main islands in the North and South Islands, including Waiheke Island as at the first wave of the panel.*

The survey excludes overseas visitors resident in New Zealand for less than 12 months and who intend to stay in New Zealand for less than 12 months; non-New Zealand diplomats and diplomatic staff and their dependants; members of non-New Zealand armed forces stationed in New Zealand and their dependants; and people living in institutions or in other non-private dwelling establishments such as boarding houses,

[1] Carter KN, Cronin M, Blakely T, Hayward M, Richardson K. Cohort profile: Survey of Families, Income and Employment (SoFIE) and Health Extension (SoFIE-health). *Int. J. Epidemiol.* 2009.

hotels, motels and hostels, as well as people living on offshore islands (excluding Waiheke Island). The cross-sectional population for a particular wave, wave *t*, can therefore be defined as: the usually resident New Zealand population living in private dwellings, as at wave *t* of the panel.

2.1.2 Sampling frame

SoFIE used the standard Statistics New Zealand sampling frame used for other Statistics New Zealand household surveys. This frame partitions the North Island, South Island and Waiheke Island into approximately 19,000 small geographical areas known as Primary Sampling Units (PSUs). On average, a PSU contains around 70 dwellings and about 150-200 people, but can range in size from 30 to 260 dwellings. The only areas not included in this sampling frame are offshore islands (except Waiheke Island) and very remote areas of New Zealand.

2.1.3 Sample selection

The selection of the random sample for SoFIE was a three-stage process. Firstly, PSUs were assigned to groups (strata) according to region, urban/rural, high/low Mäori population density and other socio-economic variables derived from the most recent census. Then systematic sampling was used to select a sample of PSUs independently from each stratum.

The next stage of sampling involved taking a systematic random sample of permanent private dwellings within the PSUs selected. All of the eligible residents of each selected household who agreed to participate were then included in the sample. These people were designated as OSMs and will be followed and interviewed over time. The sample comprised 1,500 PSUs, with an average 6.67 responding households obtained per PSU.

2.1.4 Sample Size

At wave 1 of SoFIE a total of 15,000 randomly-selected households were approached, of which approximately 11,500 agreed to be interviewed (response rate of 77%), with data collected from over 22,000 individuals aged 15 and over. All individuals who were interviewed in Wave 1 were asked to become OSMs. Children under the age of 15 at Wave 1 are interviewed as OSMs from the first wave after they turn 15. All OSMs are re-interviewed in subsequent years, regardless of changes in their place of residence. Other members of an OSM's household who are not OSMs ('non-OSMs' or cohabitants) are interviewed from Wave 2 onwards while they remain living with an OSM. Non-OSMs are asked a reduced set of questions and are not followed up if they leave the OSM's household.

2.1.5 Response rates and attrition

There were 29,685 OSMs (22,165 adults and 7,520 children) interviewed in Wave 1. In Wave 2 there were over 26,000 OSMs responding (89% of Wave 1) and in Wave 3 there were over 23,390 OSMs (88% of Wave 2 OSMs, 79% of Wave 1 OSMs). Attrition was greater in children and a child OSM who turns age 15 years before an interview is subsequently counted and interviewed as an adult OSM. The balanced panel across all three waves was 23,290 OSMs. The overall response rate of 79% of

Wave 1 responders re-interviewed in Wave 3 (based on unweighted estimates), combined with the household response rate at Wave 1 of 77%, gives an estimated effective response rate of 61%.[2] It was estimated that 55% of the original sample would be interviewed in the final Wave 8.

2.2 Survey instruments and Questionnaires

SoFIE is conducted using computer-assisted interviewing (CAI). In the field, interviewers use laptop computers to administer an electronic questionnaire (EQ), face-to-face, in respondents' homes. The EQ is organised into modules (a module is a series of questions on a particular topic).

There are two sets of questionnaires to SoFIE. For every household selected, a Household Questionnaire (HQ) is first administered to a single adult OSM, which includes questions on the characteristics of the entire household, and then a Personal Questionnaire is administered to every person aged 15 and over. In general, each adult answers their own Personal Questionnaire. However, in some cases a proxy is allowed to fill in the Personal Questionnaire for someone else, for reasons such as disability or language difficulties. Children (aged 14 or younger at the household enumeration date) do not fill out a Personal Questionnaire. Instead, the parent/caregiver nominated to answer for the child(ren) in the Household Questionnaire is asked a child module for each child as part of their Personal Questionnaire. Once a child turns 15, they answer the full PQ.

The HQ is answered by one OSM in each household. It consists of two modules, the first of which is used to determine family type and household composition at the time of the interview. The modules are:
1. Household
2. Standard of living

The PQ is administered to all OSMs and new respondents aged 15 and over, and consists of eight core modules:
1. Demographics
2. Child (if any)
3. Labour force involvement history
4. Education
5. Family
6. Current labour force involvement
7. Income
8. Contact

Depending on the module, the SoFIE questionnaire collects both point-in-time data and spell data. Point-in-time data relates to a single date, usually the interview date (e.g. the respondent's educational qualifications as at the interview date). Spell data relates to a period of time with a defined start and end date reported by the respondent (e.g. the

[2] Carter KN, Cronin M, Blakely T, Hayward M, Richardson K. Cohort profile: Survey of Families, Income and Employment (SoFIE) and Health Extension (SoFIE-health). *Int. J. Epidemiol.* 2009.

period of time a respondent lived with a family member, or the length of time a person worked for a particular employer). For example, the start of a new household spell occurs when a person enters or leaves a household. The analysis in each release includes incomplete spells, i.e. spells that were ongoing at the time of the interview.

The first wave (or interview cycle) of SoFIE was conducted from 1 October 2002 to 30 September 2003 and sought responses to these modules: Household questionnaire; Non-contact short form; Demographics; Child; Labour market history; Education; Family; Labour market; Income; Contact; Standard series modules.

The second wave (or interview cycle) of SoFIE was conducted from 1 October 2003 to 30 September 2004 and sought responses to these modules: Household questionnaire; Non-contact short form; Demographics; Child; Labour market history; Education; Family; Labour market; Income; Contact; Standard series, Asset and liability modules. These modules are repeated in Waves 4 (1 October 2005 to 30 September 2006), 6 (1 October 2007 to 30 September 2008), and 8 (1 October 2009 to 30 September 2010).

The third wave (or interview cycle) of SoFIE was conducted from 1 October 2004 to 30 September 2005 and sought responses to these modules: Household questionnaire; Non-contact short form; Demographics; Child; Labour market history; Education; Family; Labour market; Income; Contact; Standard series; Health modules. These modules are repeated in Waves 5 (1 October 2006 to 30 September 2007) and 7 (1 October 2008 to 30 September 2009).

3 SoFIE-Health

3.1 *General Information*

In 2003 researchers from the Department of Public Health, University of Otago, Wellington (UOW) applied for funding from the Health Research Council of New Zealand (HRC) to add on a health module to the core SoFIE questionnaires to be asked every alternative year. The SoFIE-Health add-on is comprised of 20 minutes of questionnaire time in waves 3 (2004-05), 5 (2006-07) and 7 (2008-09), in the following health-related domains: health status (SF36 & Kessler scale), perceived stress, chronic conditions (heart disease, diabetes, and injury-related disability), tobacco smoking, alcohol consumption, health care utilisation, and access and continuity of primary health care, and an individual deprivation score. In the Wave 3 interview OSMs were provided information and asked if they would consent to having their records linked to NZ hospitalisation, cancer registration and mortality data within the NZHIS (see Appendix 1).

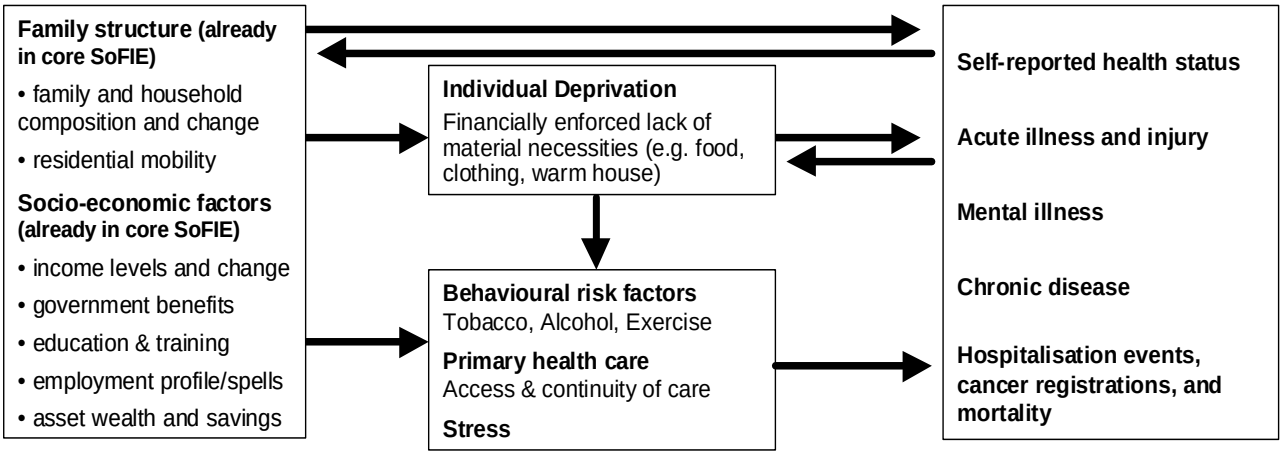

Figure 1: Domains of SoFIE-Health

In Figure 1 the information on family structure and socio-economic factors is placed on the left of the diagram, and constitutes the 'upstream determinants' of health already collected by the core of SoFIE. In the mid-stream determinants (data collected by SoFIE-Health) are individual deprivation, behavioural risk factors, primary care and stress. The right-hand box includes measures of self-reported health status and disease outcomes. The arrows identify likely causal relationships. Note that these domains cover in some way all four possible intervention points to reduce social inequalities in health identified by the Ministry of Health strategy for reducing inequalities in health. Further discussion of these issues can be found in Carter *et al*., 2008.[3]

4 SoFIE-Health & NZHIS data integration

4.1 *General Information*

The linkage of SoFIE data with hospital admissions, cancer register and mortality information allows us to investigate how factors such as labour market activity, family type and income are related to health events and health outcomes. As part of the health module OSMs were provided information and asked if they would consent to having their records linked to NZ hospitalisation, cancer registration and mortality data within the NZHIS (see Appendix 1). If the health module is being asked for the first time (i.e. an OSM has turned age 15 years) or if the respondent was previously unsure about consent they are provided with information on the linkage and asked if they would consent.

UOW access to the (microdata) unit record data is via the Data Laboratory at Statistics New Zealand, commencing in 2007. The dataset contains information for each adult OSM, with agreed variables from the core SoFIE, even-wave asset and SoFIE-health questionnaire modules, and summarised publicly funded hospital, cancer register and mortality data.

[3] Carter *et al*., 2008. SoFIE-Health Baseline Report. University of Otago, Wellington.

4.2 *Legal Context*

Ethical approval for this project was provided by the Central Region Ethics Committee.

The table below indicate the pieces of legislation that govern the different components and data integration in this project.

Dataset	Applicable Acts/Privacy codes
SoFIE dataset with details of consenting respondents	Statistics Act, including confidentiality obligations on Statistics NZ work, Privacy act
NZHIS dataset	Health Information (HI) Privacy Code
Integrated datasets	Statistics Act, including confidentiality obligations on Statistics NZ work, Privacy Act, Data Integration Policy

Security, privacy and confidentiality rules are operationalised through a series of Statistics NZ policies including:

- Confidentiality protocol and confidentiality principles of the *Statistics Act 1975*
- Policy for Publishing Official Statistics
- Microdata Access Protocols
- Data Integration Protocols
- Confidentiality Rules for release of aggregated SoFIE data

The *Statistics Act 1975* and *Privacy Act 1993* form a framework to protect information about individuals when used for statistical or research purposes. The *Privacy Act 1993* provides protection for living natural persons, while the *Statistics Act 1975* has a wider security coverage including living natural or deceased persons.

Compliance with the *Statistics Act 1975* and the *Privacy Act 1993* ensures that the privacy concerns of individuals are not set aside for the potential value and perceived benefits of record linking. Further, the *Health Information Privacy Code 1994* modifies the Information Privacy Principles in the *Privacy Act 1993*, and regulates the collection of health information by health agencies.

The purpose of the *Cancer Registry Act 1993* is to make better provision for the compilation of a statistical record of the incidence of cancer in its various forms, to provide a basis for the better direction of programmes for research and for cancer prevention. According to the *Cancer Registry Act 1993*, the purposes of the Cancer Registry are to provide information on the incidence of cancer; and to provide a basis for cancer survival studies and research programmes.

Data obtained by Statistics New Zealand for integration, and all integrated datasets, are considered to be furnished under the *Statistics Act 1975*, and therefore subject to the provisions of the Act.

Permission from supplier (SoFIE respondents & Statistics New Zealand)
Under Section 37A information may be disclosed if consent is obtained in writing from the person supplying it. See Appendix 1, which shows consent form for obtaining hospital admission and cancer register data and integrating SoFIE data obtained from respondents with hospital registry.

References on Legislation
The *Statistics Act 1975*, *Privacy Act 1993* and *Cancer Registry Act 1993* are available on http://www.legislation.govt.nz .

The Health Information Privacy Code 1994 is available on http://www.privacy.org.nz .

4.3 *Data Linkage Process*

To create linkage of hospitalisation, cancer registration and mortality data to the SoFIE (consenting) population, Statistics NZ compiled a file of names, date of birth, generated identifier and addresses of consenting SoFIE participants which staff at NZHIS matched to the National Health Index (NHI) file. NZHIS also requested sex data to be sent for matching purposes. The NHI number was then used to obtain the required health data which, along with the Statistics NZ generated id, was transferred to a Statistics New Zealand secured server (these data are called NZHIS registration data). NZHIS retained a file on a secure server containing the NHI numbers and Statistics NZ generated id to assist with future matches. The health data component of the NZHIS registration data, and a newly generated id were then placed in the Data Laboratory where approved UOW researchers, in consultation with Statistics NZ confidentiality experts, organised the data into a suitable form for analysis (and also ensured that no rare conditions could be identified). Figure 2 shows the flow diagram of information from Statistics NZ to NZHIS for linkage and matching.

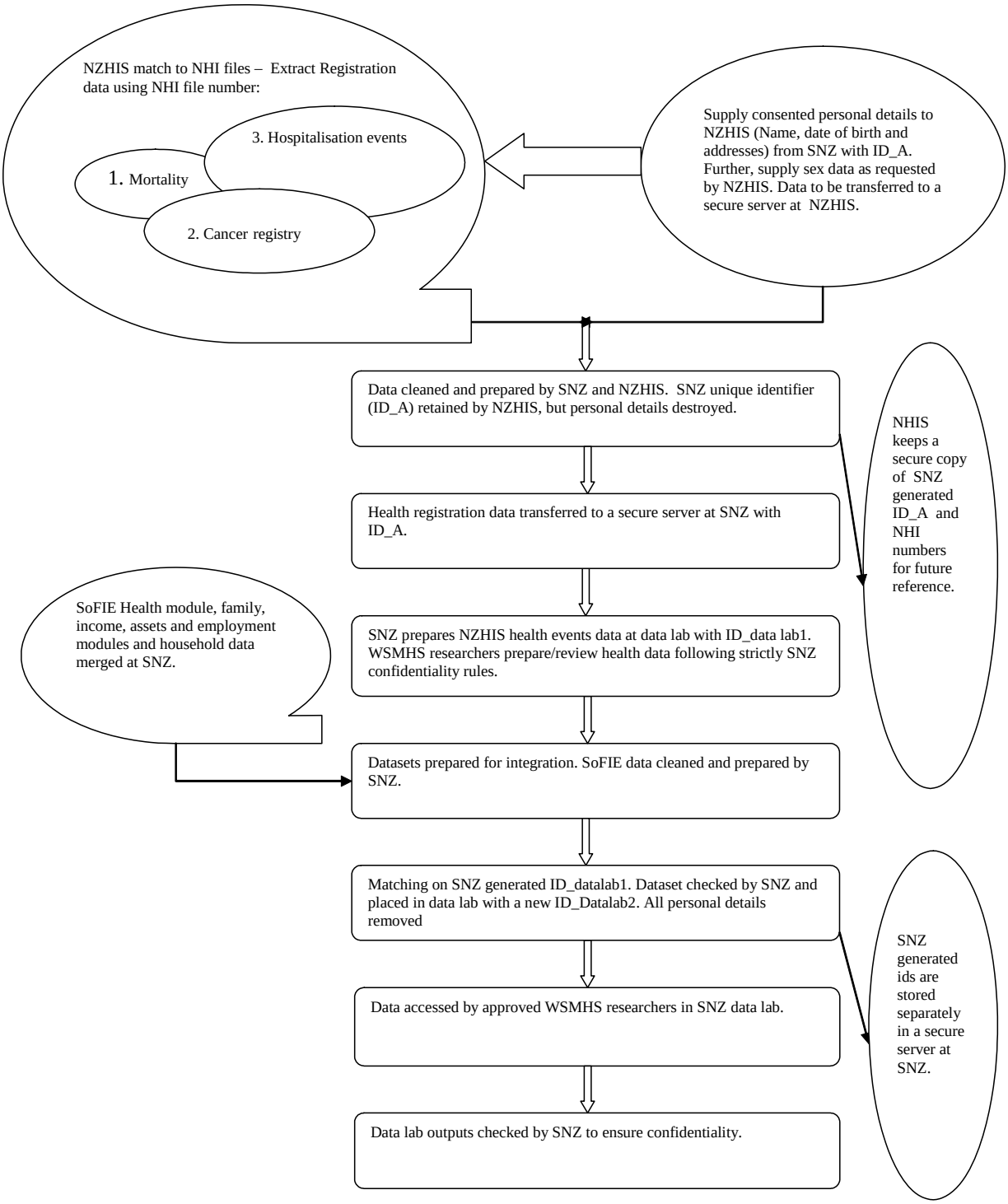

Figure 2: Flow of information from Statistics NZ to NZHIS then back to Statistics NZ

4.4 *Data Access and Confidentiality*

Unit record SoFIE data is only available for processing and analysis in the Statistics NZ data laboratory under protocols designed to protect the confidentiality of information. As noted in the SoFIE-Health Privacy Impact Assessment (PIA) Jan-Feb 2007, the protocols ensure:

- Accessed data are anonymised and confidentialised.
- Access is on-site and supervised by Statistics NZ employees.
- All output is subject to independent scrutiny by Statistics NZ staff to ensure the material is in the form of statistical aggregates that comply with the confidentiality requirements.

The current agreement with Statistics NZ for exporting summary SoFIE data from the Data Lab includes a requirement that all stratum counts are rounded to 5 units and counts less than 10 are replaced by 10.

Further details are available in the Output Guide for Data Laboratory Researchers, Statistics NZ, May 2007.

4.4.1 Data integration/Record matching

Statistics NZ assigned a unique identifier to cancer registration, mortality and hospital events records. The process of integration is explained in detail in the following sub-sections.

Statistics NZ will generate a new ID for consenting SoFIE respondents that is different to their SoFIE ID. NZHIS will keep a record of the relationship between the NHI number and the Statistics NZ generated ID number of consenting persons separately from the NZHIS registration dataset. Before the release of each dataset to UOW researchers, new IDs will be generated to prevent privacy risks.

Statistics New Zealand will keep a record of the link between the various generated IDs and SoFIE ID and securely store these ids separately from the NZHIS dataset.

Integration of data files: Treatment of source data

Record matching: Consent form personal details matched with NZHIS data at **NZHIS** using NHI file number.

To mitigate the privacy risk from transferring personal information to NZHIS, this data set was personally delivered by Statistics NZ and transferred immediately to a secure NZHIS server.

Name, address and date of birth information enable linking of individual records to the NZHIS health data and creation of health histories. In addition NZHIS requested sex data. As stated above, an individuals' personal data was supplied to NZHIS in the first

instance, but destroyed (after initial data cleaning and editing) once the NHI number was identified.

SoFIE health data matched with NZHIS data at **Statistics NZ**

The registration data (CDs of Mortality, cancer registry and publicly funded hospital events) was transferred from NZHIS to Statistics NZ by secure courier, transferred immediately to a secure server, and the CDs destroyed. Statistics NZ generated a Personal Identity Key (PIK). The various identifiers are stored securely and separately within Statistics NZ. The integrated data set will be stored securely on a Statistics NZ secured server until this project is completed in 2011.

5 Data Organisation and Variables

5.1 *General Information*

Unit record SoFIE data is only available for processing and analysis in the Statistics NZ Data Laboratory. Statistics NZ additionally require that: (i) linkage to each new wave of SoFIE data requires a separate contract to be negotiated between e.g., Statistics NZ and UOW; (ii) SoFIE data summaries must go through a confidentiality checking procedure before they can be taken outside the Data Laboratory. NZHIS data for merging/matching to SoFIE health dataset, cancer registry and publicly funded hospital events data from was obtained from 1990 onwards.

5.1.1 Mortality

Mortality data are only relevant for deceased persons, and will not be sought for the 1990 to October 2004 period. Death is likely to be a rare event among SoFIE OSMs. It will become a useful health outcome for those OSMs completing Wave 3 (or subsequent waves if missed in Wave 3 only), but is not relevant during Waves 1 and 2 prior to the health questionnaire administered in Wave 3. Variables from the mortality dataset in future matches/merges with NZHIS data will be from the Mortality Detail Table and Mortality Diagnosis Table.
The data will be summarised by UOW researchers once made available in the Statistics NZ Data Laboratory as follows:

- There will be a maximum of one record per OSM.
- Cause of Death data will be at the three digit level and then further summarised to major groupings – we expect these to be Cardiovascular, Cancer, Unintentional Injury, Suicide and other. However if possible, allowing for confidentiality requirements, we would prefer to split the CVD grouping to distinguish heart disease and cerebrovascular disease. If possible lung and other major cancers will be distinguished and road traffic accidents distinguished from other causes of unintentional injury.
- A number of flags will also be created including flags for cancer as a contributing cause and a flag identifying workplace deaths.

5.1.2 Cancer Registry

The data sought from the cancer register identified time of first notification, staging at first diagnosis, type and site of cancer. It was be sought for the 1990 to October 2006 period. The data was be summarised by UOW researchers once made available in the Statistics New Zealand Data Laboratory so that there was one record per type/site. The summary record had:

- One record per cancer type/site
- Variables identifying the type and site of the cancer.
- Variables identifying the date of first diagnosis
- Variables identifying the extent of the cancer at diagnosis

5.1.3 Publicly Funded Hospital Events (National Minimum Dataset)

Variables sought from the National Minimum Data-set enable the construction of summary records around the underlying illness or injury that is the reason for the hospital admissions. For each underlying event the data was summarised by UOW researchers once made available in the Statistics NZ Data Laboratory into unit records as follows:

- Variables identifying the date of first and latest admission for each hospitalisation event since 1990, and the principal ICD code (aggregated into groupings such as ‘coronary heart disease’, ‘stroke’, ‘other CVD’)
- Variables identifying surgical procedures.

5.2 *Final integrated datasets*

In the final integrated NZHIS datasets, date of birth was replaced with age, to enable longitudinal analyses, while still preserving confidentiality. UOW researcher access is through the Statistics NZ Data Laboratory located within Statistics NZ premises, and access is governed by Statistics NZ data integration and microdata access protocols.

These protocols ensure:

- NZHIS data accessed by UOW are anonymised and confidentialised.
- Access to the NZHIS data by UOW is on-site and supervised by Statistics NZ staff.
- All output is subject to independent scrutiny by Statistics NZ staff to ensure that it complies with the confidentiality requirements.

All the confidentiality rules for the release of aggregated census data will be applied to SoFIE health project. The summarised data will only be used in publication, which look at the interaction between health states and socio-economic and family position and change in position.

6 System design

6.1 *Introduction*

This chapter gives a brief overview of the integrated NZHIS data structure and design, describing the design ideas, directory structure, and file structure of the data in the data laboratory in Statistics NZ.

6.2 *Background*

The NZHIS datasets were first created after the wave three release of SoFIE-Health datasets, and were set up to be used with any current and future releases of SoFIE-Health information. All information provided in the raw datasets has been retained, with multiple cancer registrations and hospitalisations per person, in a restricted area on a Statistics NZ data server, only accessible by SoFIE-Health administrative staff. Each dataset is saved as a "long" dataset, where "long" refers to multiple records of cancer registration or hospitalisations per person.

The SoFIE-Health datasets have been set up as both "long" and "wide" datasets. The "long" SoFIE-Health dataset has multiple waves of information per person (i.e., 1 file record per person per wave), whereas the "wide" dataset just has one record of observations per person incorporating all waves of information (through the addition of a wave variable suffix). This means that neither dataset can readily accommodate the cancer registration or hospitalisation datasets without manipulation of the datasets and possible loss of information.

As a result all NZHIS datasets have been retained as separate datasets within the SoFIE-Health collection of datasets. The user can then select any relevant subset of NZHIS information from these datasets for their specific application and merge this with the SoFIE datasets as necessary.

6.3 *Structure*

The code that constructs the NZHIS datasets is designed in a top-down fashion (see Figure 3). There is a single main SAS code file (nh_create_ds) that controls construction of the NZHIS datasets, and this calls other SAS code files necessary for the construction of the datasets.

Each SAS file operates largely independently of other parts of the program, and variables derived by them are stored in an output NZHIS dataset. Other variables imported from the SoFIE-Health datasets, such as date of interview and respondent eligibility, are added after each dataset is constructed, by the main SAS code file.

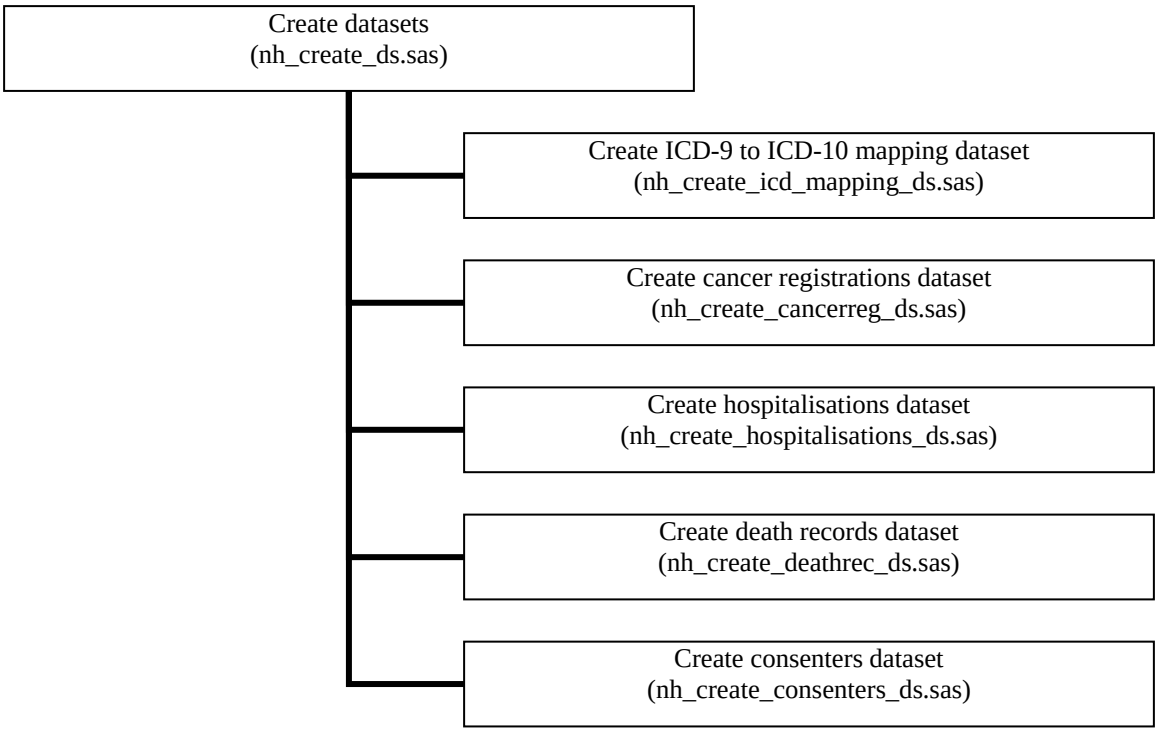

Figure 3: NZHIS Programme structure to create core datasets.

6.4 *Directories*

All code for the construction of the NZHIS datasets is located in the SASProgs directory under the NZHIS subdirectory (see Figure 4). There are four subdirectories under the NZHIS subdirectory: Admin; Datasets; Formats; and ICD.

The main SAS code file is located in the Admin subdirectory, and this references all of the other files. The Datasets directory contains the SAS code used in the construction of each individual dataset, and the Formats directory contains all formats relevant to the NZHIS and ICD datasets. Finally the ICD directory contains code and data files for the construction of the ICD mapping datasets, which is used in the construction of the hospitalisation dataset.

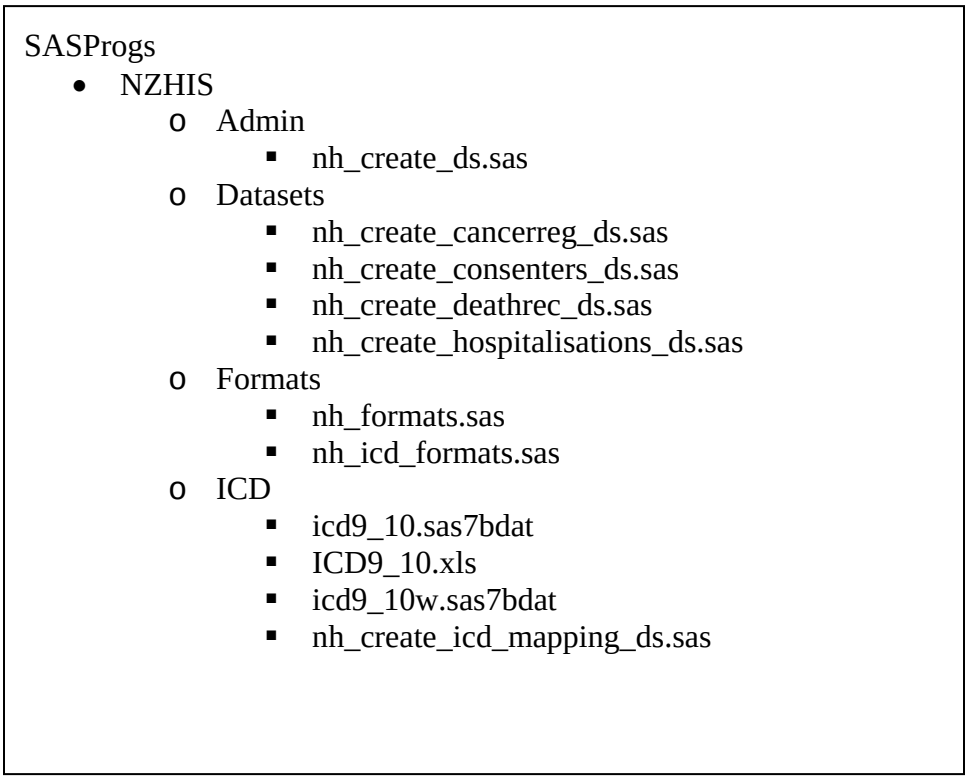

Figure 4: NZHIS directory structure in the data laboratory.

6.5 *Files*

The main SAS code file (nh_create_ds) includes the following steps:

- Set up file references
- Set up ICD datasets
- Set up temporary SoFIE information dataset
- Create cancer registrations dataset
- Create hospitalisations dataset
- Create death records dataset
- Create consenters dataset
- Clean up temporary information

The main SAS code file sets up two ICD mapping datasets, a temporary SoFIE information dataset, and the four NZHIS datasets. The two ICD datasets are used in the mapping of primary diagnosis information in the hospitalisation datasets from ICD-9 to ICD-10. Both datasets are set up and retained in a permanent location under the ICD folder.

The temporary SoFIE dataset contains information from the wide SoFIE-Health dataset and is used to derive additional variables (currently just event-wave indicators) which are added to each dataset created within the main SAS script. The temporary SoFIE dataset is then deleted at the end of the nh_create_xxx_ds scripts as part of the clean up process.

7 NZHIS Datasets

7.1 *Introduction*

This chapter provides information on the four NZHIS datasets created by the process described above. It contains general background information and sections on the ICD mapping process, formatting the variables, and the inclusion of derived variables from SoFIE-Health information. It also contains sections for each dataset, with brief information on some of the more important aspects of each dataset.

7.2 *Background*

As already noted, the Wave 3 SoFIE-Health module requested consent from SoFIE participants for the linkage of SoFIE-Health data with mortality, cancer registry, and hospitalisation data. The NZHIS then provided information for those that consented in the form of three datasets, along with a fourth dataset of consent and match indicators.

These four datasets are:

- Cancer registrations – information from the NZHIS New Zealand Cancer Registry (NZCR)
- Hospitalisations – information from the NZHIS National Minimum Dataset (Hospital Events) (NMDS)
- Death records – information from the NZHIS Mortality Collection
- Consenters – an indicator of match status from the matching process of SoFIE-Health information to NZHIS National Health Index information undertaken by NZHIS.

The individual SAS scripts that set up each dataset follow a standard structure involving three core steps:

- Preliminary work – the only significant section of this is in the hospitalisations, where primary diagnosis codes must be mapped from ICD-9 to ICD-10.
- Create derived variables – this section adds any variables derived from information within the respective datasets.
- Apply formats and labels – applies formats and labels to variables in the respective datasets.

The final step in the creation of each dataset is to merge information from SoFIE dataset, applying any additional formats and labels. This occurred after each dataset was set up, in order to preserve a separation of concerns between datasets (as per the Structure section in the Systems Design chapter).

All of the site and morphology codes in the cancer registry data were provided in ICD-10; however the hospitalisations datasets generally lacked ICD-10 primary diagnosis codes prior to August 1998. All hospitalisations had ICD-9 primary diagnosis codes, and these could be mapped to ICD-10 codes. This involved setting up a mapping dataset, the details of which are included in the next section.

7.3 *ICD mapping*

Mapping files were created in order to map the primary diagnosis variables in the hospitalisations dataset from ICD-9 to ICD-10. Setting up these datasets requires three pieces of information: the clinical code type, the ICD-9 code, and the corresponding ICD-10 code.

The clinical code type is a code that indicates which part of the clinical code table the respective ICD codes fall within. Each ICD-9 code can map to several ICD-10 codes, depending on the clinical code type. However each ICD-9 code only maps to a single ICD-10 code within a given clinical code table. Hence determining the clinical code type is critical to obtaining the correct mapping of ICD codes from ICD-9 to ICD-10

The excel spreadsheet ICD9_10.xls, originally supplied by NZHIS, contains a mapping of ICD-9 codes to ICD-10 codes, grouped by clinical code type. There is also a similar excel spreadsheet masterf4.xls on the NZHIS website, under the technical documentation, mapping files section.

The ICD9_10.xls spreadsheet contains two columns of ICD-10 code: ICD10L and ICD10H. The first is a logical mapping and the second a historical mapping. We have been advised to use the logical mapping, unless there is a good reason to do otherwise. The masterf4.xls spreadsheet only includes a logical mapping.

The first ICD mapping dataset (icd9_10.sas7bdat) is constructed through simply reading the excel spreadsheet into SAS. The data is sorted by clinical code and ICD-9 values, and restricted to just the logical mapping of ICD-9 to ICD-10 codes.

The second “wide” dataset (icd9_10w.sas7bdat) transposes the first dataset by ICD-9 codes, so that each row contains all of the possible mappings from ICD-9 to ICD-10. This is done by using column variables that correspond to the clinical code types “A”, “B”, “E”, “M”, “O”, or “V”. For example, if an ICD-9 code maps to a different ICD-10 code in the “A” and “V” tables, then these different ICD-10 codes now appear under the respective “A” and “V” variables.

A further primary diagnosis variable is created in the “wide” dataset using the “A”, “B” and “V” clinical code types. This process prioritises these codes by first copying the “A” clinical code table ICD-10 codes, then copying “B” codes if there is no “A” codes, and finally using the “V” codes if there are no “A” or “B” codes. The result of this is a

mapping of ICD-9 codes to a prioritised ICD-10 primary diagnosis code, which is then used in mapping the primary diagnosis variables in the hospitalisation dataset.

A final list of ICD-10 chapter and subchapter codes is included in Appendix 3.

7.4 *Formats*

NZHIS variable formats are supplied in the nh_formats.sas file. The file contains a section for general formats that are used in several of the NZHIS datasets, and then an individual section of formats specific to each dataset.

Formats are mainly derived from the code tables on the NZHIS website relating to variables in the New Zealand Cancer Registry and National Minimum Dataset (Hospitalisation Events). Additional formats are included from the CancerTrends project in order to group the cancer registry data, as well as a format for the consent indicator.

Formats for ICD-10 variables, excluding the CancerTrends formats, were supplied directly by NZHIS and are contained in the nh_icd_formats.sas file.

7.5 *Cancer Registration Dataset*

The cancer registrations dataset contains eleven variables from the original dataset, and five derived variables. The derived variables include site chapter codes, subchapter codes, and subchapter descriptions. A summary of the variables in the cancer registration dataset appears in the appendices.

The site chapter variable is just the first letter of each ICD-10 site code, dividing the site codes into malignant “C” codes and in situ “D” codes. The subchapter variables apply the ICD-10 subchapter groups supplied by NZHIS, providing subchapter codes and descriptions for each site code.

The site subchapter (CT) variables use slightly different groupings and descriptions from those in the Cancer Trends projects, hence the CT suffix. These are provided to help facilitate any comparison between the two projects.

7.6 *Hospitalisation Dataset*

The hospitalisation dataset contains a large number of variables, including two derived variables for ICD-10 primary diagnosis chapter codes. The ICD-10 primary diagnosis codes and the relevant subchapter codes and descriptions are also updated as part of the mapping of ICD-9 primary diagnosis codes to ICD-10 codes.

The hospitalisations SAS file first maps the ICD-9 primary diagnosis variable in order to fill in the missing ICD-10 primary diagnosis codes. This uses the “wide” mapping

dataset (icd9_10w.sas7bdat, set up previously) and fills in the missing ICD-10 primary diagnosis codes with the ICD-9 mapped code. Missing ICD-10 codes for operation/procedure variables could also be mapped as part of this process, but are not as there is no immediate intention to use these variables.

All ICD-10 subchapter codes and descriptions are overwritten after the mapping process is completed. The original subchapter descriptions seemed to have some special characters present and would result in some procedures listing identical new and original descriptions separately. There were also approximately half a dozen original descriptions that did not match the respective ICD-10 codes, some even referring to a different chapter. Therefore, for completeness, all ICD-10 subchapter codes and descriptions were updated using a common set of formats.

The ICD-10 chapter codes and descriptions were supplied by NZHIS, along with the subchapter codes and groups.

7.7 *Death Records Dataset*

Both the death records and consenter datasets are very brief, with few variables. There is a maximum of one observation per person and both are expected to be merged directly into the SoFIE-Health dataset in the future.

The death records dataset contains two variables and one derived variable. The two variables are just a person identifier and the death date. The derived variable uses this date to give a death year.

One death occurs between waves two and three, several months before wave three interview date. This is obviously not possible as each individual needs to agree for the data to be linked by signing the consent form at the wave three interview date. This has been briefly mentioned to Statistics New Zealand, and errors such as this may provide some indication of the reliability of the linkage process.

7.8 *Consenter Dataset*

The consent dataset has only two variables and no derived variables (at this stage). The first variable is the person identifier and the second is a match indicator. Both indirectly indicate consenters, as the dataset only contains person codes and match indicators for those that consented to their hospitalisation, cancer registration and mortality data being linked to the SoFIE-Health datasets.

The match indicator has three codes: 0 for no match, 1 for direct match and 2 for manual match. An automatic procedure was used by NZHIS to match National Health Information files with the SoFIE-Health information, and if this resulted in an exact match then the corresponding match indicator is set to a direct match. Of those records that do not automatically match, NZHIS checked close records and would manually match records with sufficient agreement. An example of a manual match would be where first and last names of an individual are in a different order on the NZHIS file.

7.9 *SoFIE variables*

A temporary SoFIE information dataset is constructed prior to NZHIS datasets, and includes cross sectional and longitudinal person identifiers and annual interview period dates. The variables are added as a final step in the main SAS code file (nh_create_ds.sas) to all NZHIS datasets, except the consenter dataset. The temporary SoFIE information dataset is then deleted after all the NZHIS datasets are created.

The addition of annual period dates as part of the SoFIE information allows the construction of wave indicator variables for the hospitalisation, cancer registration, and death records dataset. SoFIE interviews are conducted in monthly groups throughout the year, so there is no generic wave start and end dates. The individual period dates provided in the SoFIE information facilitate the construction of wave specific variables to indicate if a hospitalisations, cancer registration, or death record falls within the one-year period covered by a given wave.

8 Descriptive Data

8.1 *Introduction*

This chapter provides summary tables of a few key variables in the NZHIS datasets. Much of this work in the creation of these tables was coincidental to initial work on the NZHIS datasets so only provides a brief introduction to the type of information available.

All values are random rounded to base 5 and have a minimum value of 10, consistent with Statistics New Zealand confidentiality requirements.

8.2 *Consenters*

The following tables show the NZHIS consent and match indications by the SoFIE-Health consent indicator status. Overall 79% of the adult population at Wave 3 consented to having their data linked to NZHIS hospitalisation, cancer registration and mortality data. Of those, 96% were matched to an NHI number by NZHIS.

Table 1: Respondent consent to NZHIS linkage by the NZHIS matching process

NZHIS Match Indicator	SoFIE Consent Indicator		
	Yes	No	Total
Direct Match	11855		11855
Manual Match	2120		2120
No Match	520		520
Missing	130	3875	4005
Total	14625	3875	18500

There is a slight discrepancy in both tables, whereby 130 individuals consented according to the SoFIE consent indicator but have not had NZHIS information sent through. It is understood that there was an additional limitation on the data relating to "original sample member" status whereby these individuals were not included in the list of individuals sent through to NZHIS for linkage of cancer registration, hospitalisation, and mortality information.

8.3 *Hospitalisations*

The following tables show the total number of hospitalisations by year, number of hospitalisation events per year, and the number of hospitalisations per person, by length of stay in hospital. Note that a person can have multiple hospitalisations.

Table 2: Number of hospitalisations per year, 1990 to 2006.

Year of hospitalisation	Number of events
1990	1490
1991	1515
1992	1445
1993	1585
1994	1595
1995	1720
1996	1705
1997	1850
1998	1810
1999	2025
2000	2190
2001	2460
2002	2490
2003	2545
2004	2955
2005	3260
2006	10
Male	12065
Female	20585
Total	32655

Table 3: Number of hospitalisations within individuals, 1990 to 2006.

Year of hospitalisation	Events per year		
	1	2+	Total
1990	840	260	1100
1991	855	250	1105
1992	805	235	1040
1993	870	275	1145
1994	905	280	1185
1995	905	295	1200
1996	885	315	1200
1997	985	320	1305
1998	915	330	1245
1999	995	385	1380
2000	1040	425	1465
2001	1110	450	1560
2002	1120	480	1600
2003	1135	500	1635
2004	1225	540	1765
2005	1225	585	1810
2006	10	10	10
Total	15835	5920	21755

Table 4: Number of hospitalisations by length of stay in hospital

Admissions per person	Length of stay						
	0 days	1-5 days	6-10 days	11-20 days	21-50 days	51+ days	Total
1	925	1640	185	50	15	10	2825
2-3	2210	3960	690	195	60	10	7120
4-5	1820	3295	620	225	70	15	6045
6-10	2285	4455	820	335	120	25	8045
11-20	1230	2495	555	265	105	55	4705
21-30	445	890	200	120	45	25	1725
31-40	410	285	95	45	25	10	875
41-50	115	100	40	10	10	10	270
51-100	115	185	40	25	10	10	380
101+	615	45	10	10	.	.	660
Total	10170	17350	3250	1270	460	150	32650

The observations per person by length of stay (Table 4) have a concentration of individuals with few observations and shorter lengths of stay (as expected). However there are a number of individuals with a large number of short hospitalisation events, possibly the likes of dialyses patients, and a few individuals with very long stays.

8.4 *Cancer registrations*

Tables 5 and 6 show the number of cancer registrations for each individual in the cancer registration dataset and the total number of cancer registrations by year.

The cancer registration dataset again shows a steady increase in reported cancers over the period 1990 to 2005, possibly due to a mortality bias. There are also very few individuals with more than one cancer registration.

Table 5: Number of cancer registrations

Observations per person	Number of people
1	800
2	50
3	10
Total	860

Table 6: Number of cancer registrations by year of diagnosis

Year of Diagnosis	Site Chapter		
	C	D	Total
1990	15	10	20
1991	15	10	15
1992	15	10	25
1993	30	10	35
1994	20	10	30
1995	35	15	45
1996	30	10	40
1997	45	10	55
1998	30	10	45
1999	50	15	65
2000	50	30	80
2001	45	15	60
2002	60	25	85
2003	70	20	95
2004	70	25	100
2005	85	30	115
Total	670	250	920

* C = Malignant, D = In-situ

8.5 *Death records*

The following table show the NZHIS death records by year (after Wave 3):

Table 7: Number of deaths registered since 2004,

Year	Death Records
2004	10
2005	65
2006	110
2007	115
Total	280

The death records table shows the mortality information that was available at the time of the linkage of NZHIS and SoFIE datasets. There are very few records prior to 2006 as wave three of SoFIE was still in the field up until September 2005 (and individuals obviously had to be alive to sign the consent for linkage form). This information will also be incomplete as some sources of death record information can take several years to be entered into the NZHIS mortality tables (e.g. coroner's records).

9 APPENDICES

9.1 *SoFIE-Health Consent and Information Sheet*

The third SoFIE interview includes questions about your health and use of health services. The health questions in the Survey of Family, Income and Employment (SoFIE) will assist the Wellington School of Medicine and Health Sciences, University of Otago, and the Ministry of Health in researching the links between income, work and health. The research results will provide new information for planning policies and services for healthcare in New Zealand.

In addition to these questions, Statistics New Zealand (Statistics New Zealand) sought your consent to link data collected in SoFIE with hospital admissions hospitals, mortality and cancer register information held by the New Zealand Health Information Service (NZHIS). Linking this data will show how factors such as labour market activity, family type and income are related to health events and health outcomes. For example, it will show how hospitalisation impacts on employment and income.

Wellington Ethics Committee has given ethical approval for this data linking.

Even if you have never been admitted to a public hospital in New Zealand, we request your consent to pass your name, address and birth date to the NZHIS. The fact that you have never been in hospital is useful information for the research.

To ensure the data is correctly linked, Statistics New Zealand will provide the NZHIS with your name, address and date of birth. Using this information, the NZHIS will extract the following information (where it exists):

- Admission and discharge dates to New Zealand public hospitals together with the codes for diagnosis or type of accident, and type of treatment; and
- Year of cancer registration together with codes for site, type and extent of cancer.

Once NZHIS has extracted the information, all names, and addresses will be removed. The linked data will be summarised so that rare health conditions cannot be identified. The data will then be returned to Statistics New Zealand. Medical records or doctors' notes cannot be seen by anyone at the NZHIS, Statistics New Zealand or the Wellington School of Medicine.

Under the Statistics Act 1975, the information you provide, including this linked data, will be treated confidentially. Only people authorised by the Statistics Act 1975 are allowed to see your individual information, and they must use it only for statistical purposes. Your information will be combined with similar information from others to prepare summary statistics. The linked data will be held solely by Statistics New Zealand at our offices. Researchers who wish to use the statistical data for approved projects must apply for access to do so. Access is provided at the discretion of the Government Statistician.

If you choose not to sign the consent form, your details will not be passed to the NZHIS. Your participation in SoFIE will not be affected and your future health care will not be affected.

If you agree to sign the consent form, your details will be passed to the NZHIS to obtain information about past Cancer Register data, as well as any future data that may be obtained. There are no time restrictions placed on when the data may be obtained; data may be obtained up to and after death.

Further information
If you want more information about SoFIE, please telephone Statistics New Zealand toll free on 0800 104 744, or email us at sofie@stats.govt.NZ, or visit our website www.stats.govt.NZ/sofie. If you require further information about linking the health data, please contact the SoFIE Project Leader.

Thank you again for your help so far. Your continued participation in SoFIE is greatly appreciated and is important to the success of the survey. I hope you continue to enjoy being part of this major nation-wide survey.

Yours sincerely

Brian Pink
Government Statistician

Person number: _____

Consent form for obtaining Hospital Admission and Cancer Register Data

Statistics New Zealand (Statistics NZ) wishes to obtain your consent to link data collected in SoFIE with past and future hospital admissions, mortality and cancer register information held by the New Zealand Health Information Service (NZHIS). Linking this data will show how factors such as labour market activity, family status and income are related to health events and health outcomes. For example, it will show how hospitalisation impacts on employment and income.

Under the Statistics Act 1975, the information you provide, including this linked data, will be treated confidentially. Only people authorised by the Statistics Act 1975 are allowed to see your individual information, and they must use it only for statistical purposes. Your information will be combined with similar information from others to prepare summary statistics.

To ensure the data is correctly linked, Statistics New Zealand will provide the NZHIS with your name, address and date of birth. Using this information, the NZHIS will extract the following information (where it exists):

- Admission and discharge dates to New Zealand public hospitals together with the codes for diagnosis or type of accident, and type of treatment; and
- Year of cancer registration together with codes for site, type and extent of cancer; and
- Date and cause of death.

Once NZHIS has extracted the information, all names and addresses will be removed. The linked data will be summarised so that rare conditions cannot be identified. The data will then be returned to Statistics New Zealand. Medical records or doctors' notes cannot be seen by anyone at the NZHIS, Statistics New Zealand or the Wellington School of Medicine.

If you choose not to sign this form, no data from the NZHIS will be linked with your SoFIE data. Your participation in SoFIE will not be affected.

If you agree to sign this form, your details will be passed to the NZHIS to obtain information about past hospital admissions and Cancer Register data, as well as any future data that may be obtained. There are no time restrictions placed on when the data may be obtained; data may be obtained up to and after death.

I consent to my name, address and birth date being passed to the New Zealand Health Information Service for the purpose of obtaining information about hospital admissions, Cancer Register data, and date and cause of death. I understand this information will be linked with data collected in SoFIE.

Full Name:

Signed: Date: / /

9.2 *SoFIE-Health NZHIS data dictionary*

The following pages contains a brief listing (contents) of each dataset, including variable names, labels, formats, and any additional notes.

9.2.1 Cancer Registrations

Variable Name	Label	Format	Notes
age	Age at diagnosis		
basis	Basis of cancer diagnosis	$FBAS10C	
canc_diag	Cancer diagnosis	FYN124N	Derived using SoFIE dataset
canc_diagw1	Cancer diagnosis (W1)	FYN124N	Derived using SoFIE dataset
canc_diagw2	Cancer diagnosis (W2)	FYN124N	Derived using SoFIE dataset
canc_diagw3	Cancer diagnosis (W3)	FYN124N	Derived using SoFIE dataset
diag_date	Date of diagnosis	DDMMYY	SNZ confidentiality applied to variable
EAPDW1	End date of annual reference period (W1)	DDMMYY	Obtained from SoFIE dataset
EAPDW2	End date of annual reference period (W2)	DDMMYY	Obtained from SoFIE dataset
EAPDW3	End date of annual reference period (W3)	DDMMYY	Obtained from SoFIE dataset
HEDW1	Household enumeration date (W1)	DDMMYY	Obtained from SoFIE dataset
HEDW2	Household enumeration date (W2)	DDMMYY	Obtained from SoFIE dataset
HEDW3	Household enumeration date (W3)	DDMMYY	Obtained from SoFIE dataset
HOUSEHOLDW3	Randomised Household identifier (W3)		Obtained from SoFIE dataset
morph	Morphology code ICD-9		May be unreliable (particularly prior to 1997)
morph_descr	Morphology description ICD-9		May be unreliable (particularly prior to 1997)
PersonCS	Cross Sectional Randomised Person identifier		
PERSONCSW3	Cross Sectional Randomised Person identifier (W3)		Obtained from SoFIE dataset
PersonLT	Longitudinal Randomised Person identifier		Obtained from SoFIE dataset
regyr	Year of diagnosis		
SAPDW1	Start date of the annual reference period (W1)	DDMMYY	Obtained from SoFIE dataset

SAPDW2	Start date of the annual reference period (W2)	DDMMYY	Obtained from SoFIE dataset
SAPDW3	Start date of the annual reference period (W3)	DDMMYY	Obtained from SoFIE dataset
sex	Sex	$FSEXG4C	
site	Site code ICD-10		
site_chapter	Site chapter ICD-10		Derived from site variable
site_descr	Site description ICD-10		
site_subchapter	Site subchapter ICD-10		Derived from site variable (original overwritten)
site_subchapter_CT	Site subchapter ICD-10 (CT)		Derived variable (use CancerTrends groups)
site_subchapter_descr	Site subchapter description ICD-10		Derived from site variable (original overwritten)
site_subchapter_descr_CT	Site subchapter description ICD-10 (CT)		Derived variable (use CancerTrends groups)
stage	Extent of disease	$FSTG13C	May be unreliable (particularly prior to 1997)

9.2.2 Consenters

Variable Name	Label	Format	Description
consent	Consent	FYN124N	
HOUSEHOLDW3	Randomised Household identifier (W3)		Obtained from SoFIE dataset
MatchInd	Indicator of match status to NZHIS data	$FMSTA3C	
PersonCS	Cross Sectional Randomised Person identifier		Obtained from SoFIE dataset
PERSONCSW3	Cross Sectional Randomised Person identifier (W3)		Obtained from SoFIE dataset
PersonLT	Longitudinal Randomised Person identifier		Obtained from SoFIE dataset

9.2.3 Death records

Variable Name	Label	Format	Description
death	Death	FYN124N	Derived using SoFIE dataset
DeathDate	Date of recorded death	DDMMYY	
deathw1	Death (W1)	FYN124N	Derived using SoFIE dataset
deathw2	Death (W2)	FYN124N	Derived using SoFIE dataset

deathw3	Death (W3)	FYN124N	Derived using SoFIE dataset
deathyear	Year of recorded death		Derived from DeathDate variable
EAPDW1	End date of annual reference period (W1)	DDMMYY	Obtained from SoFIE dataset
EAPDW2	End date of annual reference period (W2)	DDMMYY	Obtained from SoFIE dataset
EAPDW3	End date of annual reference period (W3)	DDMMYY	Obtained from SoFIE dataset
HEDW1	Household enumeration date (W1)	DDMMYY	Obtained from SoFIE dataset
HEDW2	Household enumeration date (W2)	DDMMYY	Obtained from SoFIE dataset
HEDW3	Household enumeration date (W3)	DDMMYY	Obtained from SoFIE dataset
HOUSEHOLDW3	Randomised Household identifier (W3)		Obtained from SoFIE dataset
PersonCS	Cross Sectional Randomised Person identifier		
PERSONCSW3	Cross Sectional Randomised Person identifier (W3)		Obtained from SoFIE dataset
PersonLT	Longitudinal Randomised Person identifier		Obtained from SoFIE dataset
SAPDW1	Start date of the annual reference period (W1)	DDMMYY	Obtained from SoFIE dataset
SAPDW2	Start date of the annual reference period (W2)	DDMMYY	Obtained from SoFIE dataset
SAPDW3	Start date of the annual reference period (W3)	DDMMYY	Obtained from SoFIE dataset

9.2.4 Hospitalisations

Variable Name	Label	Format	Description
ADM_SRC	Admission source	$FADMS2C	
ADM_TYPE	Admission type	$FADM10C	
AGE_DIS	Age at discharge		
AGENCY	Agency	$FAGNCYC	
chapter_descr_icd10	Diagnosis chapter description ICD-10		Derived from prim_diag_icd10 (original overwritten)
chapter_icd10	Diagnosis chapter ICD-10		Derived from prim_diag_icd10 (original overwritten)
DRG_31	DRG code version 3.1	$	
DRG_CURRENT	DRG code current	$	
DRG_GROUPER_TYPE	DRG grouper type code	$FDRGG5C	
EAPDW1	End date of annual reference period (W1)	DDMMYY	Obtained from SoFIE dataset

EAPDW2	End date of annual reference period (W2)	DDMMYY	Obtained from SoFIE dataset
EAPDW3	End date of annual reference period (W3)	DDMMYY	Obtained from SoFIE dataset
END_TYPE	Event end type	$FEND19C	
EVENDATE	Event end date	DDMMYY	
EVENT_TYPE	Event type	$FEVT14C	
EVNTLVD	Event leave days	$	
EVSTDATE	Event start date	DDMMYY	
FAC_TYPE	Facility type	$FFCT21C	
FACILITY	Facility	$FFCLTYC	
GENDER	Gender	$FSEXG4C	
HEDW1	Household enumeration date (W1)	DDMMYY	Obtained from SoFIE dataset
HEDW2	Household enumeration date (W2)	DDMMYY	Obtained from SoFIE dataset
HEDW3	Household enumeration date (W3)	DDMMYY	Obtained from SoFIE dataset
HLTHSPEC	Health specialty	$FHLTSPC	
hosp_evnt	Hospitalisation	FYN124N	Derived using SoFIE dataset
hosp_evntw1	Hospitalisation (W1)	FYN124N	Derived using SoFIE dataset
hosp_evntw2	Hospitalisation (W2)	FYN124N	Derived using SoFIE dataset
hosp_evntw3	Hospitalisation (W3)	FYN124N	Derived using SoFIE dataset
HOUSEHOLDW3	Randomised Household identifier (W3)		
id_event	Event ID		
LOCAL_ID	Event local identifier	$FLOID9C	
LOS	Length of stay	$	
op01	Operation/procedure 01 ICD-9		
op02	Operation/procedure 02 ICD-9		
op03	Operation/procedure 03 ICD-9		
op04	Operation/procedure 04 ICD-9		
op05	Operation/procedure 05 ICD-9		
op06	Operation/procedure 06 ICD-9		
op07	Operation/procedure 07 ICD-9		
op08	Operation/procedure 08 ICD-9		
op09	Operation/procedure 09 ICD-9		
op10	Operation/procedure 10 ICD-9		
op1001	Operation/procedure 01 ICD-10		Partial data (has not been mapped forward)

op1002	Operation/procedure 02 ICD-10		Partial data (has not been mapped forward)
op1003	Operation/procedure 03 ICD-10		Partial data (has not been mapped forward)
op1004	Operation/procedure 04 ICD-10		Partial data (has not been mapped forward)
op1005	Operation/procedure 05 ICD-10		Partial data (has not been mapped forward)
op1006	Operation/procedure 06 ICD-10		Partial data (has not been mapped forward)
op1007	Operation/procedure 07 ICD-10		Partial data (has not been mapped forward)
op1008	Operation/procedure 08 ICD-10		Partial data (has not been mapped forward)
op1009	Operation/procedure 09 ICD-10		Partial data (has not been mapped forward)
op1010	Operation/procedure 10 ICD-10		Partial data (has not been mapped forward)
op1011	Operation/procedure 11 ICD-10		Partial data (has not been mapped forward)
op1012	Operation/procedure 12 ICD-10		Partial data (has not been mapped forward)
op1013	Operation/procedure 13 ICD-10		Partial data (has not been mapped forward)
op1014	Operation/procedure 14 ICD-10		Partial data (has not been mapped forward)
op1015	Operation/procedure 15 ICD-10		Partial data (has not been mapped forward)
op1016	Operation/procedure 16 ICD-10		Partial data (has not been mapped forward)
op1017	Operation/procedure 17 ICD-10		Partial data (has not been mapped forward)
op1018	Operation/procedure 18 ICD-10		Partial data (has not been mapped forward)
op1019	Operation/procedure 19 ICD-10		Partial data (has not been mapped forward)
op1020	Operation/procedure 20 ICD-10		Partial data (has not been mapped forward)
op11	Operation/procedure 11 ICD-9		
op12	Operation/procedure 12 ICD-9		
op13	Operation/procedure 13 ICD-9		
op14	Operation/procedure 14 ICD-9		
op15	Operation/procedure 15 ICD-9		
op16	Operation/procedure 16 ICD-9		
op17	Operation/procedure 17 ICD-9		
op18	Operation/procedure 18 ICD-9		
op19	Operation/procedure 19 ICD-9		
op20	Operation/procedure 20 ICD-9		
PersonCS	Cross Sectional Randomised Person identifier		
PERSONCSW3	Cross Sectional Randomised Person identifier (W3)		Obtained from SoFIE dataset
PersonLT	Longitudinal Randomised Person identifier		Obtained from SoFIE dataset

prim_diag	Diagnosis ICD-9	$	
prim_diag_icd10	Diagnosis ICD-10		Derived from prim_diag (original overwritten)
PURCHASER	Principal health service purchaser	$FPURCHC	
SAPDW1	Start date of the annual reference period (W1)	DDMMYY	Obtained from SoFIE dataset
SAPDW2	Start date of the annual reference period (W2)	DDMMYY	Obtained from SoFIE dataset
SAPDW3	Start date of the annual reference period (W3)	DDMMYY	Obtained from SoFIE dataset
SUB_SYS	Sub-system	$	Submitted system ID
subchapter	Diagnosis subchapter ICD-9		
subchapter_descr	Diagnosis subchapter description ICD-9		
subchapter_descr_icd10	Diagnosis subchapter description ICD-10		Derived from prim_diag_icd10 (original overwritten)
subchapter_icd10	Diagnosis subchapter ICD-10		Derived from prim_diag_icd10 (original overwritten)

9.3 *ICD-10AM chapter and subchapter headings*

The following pages contains a listing of the ICD-10AM chaper and subchapter headings.

CHAPTER	CODES	CHAPTER NAME
I	A00-B99	Certain infectious and parasitic diseases
II	C00-D48	Neoplasms
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	E00-E90	Endocrine, nutritional and metabolic diseases
V	F00-F99	Mental and behavioural disorders
VI	G00-G99	Diseases of the nervous system
VII	H00-H59	Diseases of the eye and adnexa
VIII	H60-H95	Diseases of the ear and mastoid process
IX	I00-I99	Diseases of the circulatory system
X	J00-J99	Diseases of the respiratory system
XI	K00-K93	Diseases of the digestive system
XII	L00-L99	Diseases of the skin and subcutaneous tissue
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue
XIV	N00-N99	Diseases of the genitourinary system
XV	O00-O99	Pregnancy, childbirth and the puerperium
XVI	P00-P96	Certain conditions originating in the perinatal period
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
XVIII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX	S00-T98	Injury, poisoning and certain other consequences of external causes
XX	V01-Y98	External causes of morbidity and mortality
XXI	Z00-Z99	Factors influencing health status and contact with health services

CODES	SUBGROUP NAME
A00–A09	Intestinal infectious diseases
A15–A19	Tuberculosis
A20–A28	Certain zoonotic bacterial diseases
A30–A49	Other bacterial diseases
A50–A64	Infections with a predominantly sexual mode of transmission
A65–A69	Other spirochaetal diseases
A70–A74	Other diseases caused by chlamydiae
A75–A79	Rickettsioses
A80–A89	Viral infections of the central nervous system
A90–A99	Arthropod-borne viral fevers and viral haemorrhagic fevers
B00–B09	Viral infections characterised by skin and mucous membrane lesions
B15–B19	Viral hepatitis
B20–B24	Human immunodeficiency virus [HIV] disease
B25–B34	Other viral diseases
B35–B49	Mycoses
B50–B64	Protozoal diseases
B65–B83	Helminthiases
B85–B89	Pediculosis, acariasis and other infestations
B90–B94	Sequelae of infectious and parasitic diseases
B95–B97	Bacterial, viral and other infectious agents
B99	Other infectious diseases
C00–C96	Malignant neoplasms
C00–C75	Malignant neoplasms, stated or presumed to be primary, of specified sites, except of lymphoid, haematopoietic and related tissue
C00–C14	Malignant neoplasms of lip, oral cavity and pharynx
C15–C26	Malignant neoplasms of digestive organs
C30–C39	Malignant neoplasms of respiratory and intrathoracic organs
C40–C41	Malignant neoplasms of bone and articular cartilage
C43–C44	Melanoma and other malignant neoplasms of skin
C45–C49	Malignant neoplasms of mesothelial and soft tissue
C50	Malignant neoplasms of breast

C51–C58	Malignant neoplasms of female genital organs
C60–C63	Malignant neoplasms of male genital organs
C64–C68	Malignant neoplasms of urinary tract
C69–C72	Malignant neoplasms of eye, brain and other parts of central nervous system
C73–C75	Malignant neoplasms of thyroid and other endocrine glands
C76–C80	Malignant neoplasms of ill-defined, secondary and unspecified sites
C81–C96	Malignant neoplasms, stated or presumed to be primary, of lymphoid, haematopoietic and related tissue
D00–D09	In situ neoplasms
D10–D36	Benign neoplasms
D37–D48	Neoplasms of uncertain or unknown behaviour
D50–D53	Nutritional anaemias
D55–D59	Haemolytic anaemias
D60–D64	Aplastic and other anaemias
D65–D69	Coagulation defects, purpura and other haemorrhagic conditions
D70–D77	Other diseases of blood and blood-forming organs
D80–D89	Certain disorders involving the immune mechanism
E00–E07	Disorders of thyroid gland
E10–E14	Diabetes mellitus
E15–E16	Other disorders of glucose regulation and pancreatic internal secretion
E20–E35	Disorders of other endocrine glands
E40–E46	Malnutrition
E50–E64	Other nutritional deficiencies
E65–E68	Obesity and other hyperalimentation
E70–E90	Metabolic disorders
F00–F09	Organic, including symptomatic, mental disorders
F10–F19	Mental and behavioural disorders due to psychoactive substance use
F20–F29	Schizophrenia, schizotypal and delusional disorders
F30–F39	Mood [affective] disorders
F40–F48	Neurotic, stress-related and somatoform disorders
F50–F59	Behavioural syndromes associated with physiological disturbances and physical factors
F60–F69	Disorders of adult personality and behaviour
F70–F79	Mental retardation
F80–F89	Disorders of psychological development

F90–F98	Behavioural and emotional disorders with onset usually occurring in childhood and adolescence
F99	Unspecified mental disorder
G00–G09	Inflammatory diseases of the central nervous system
G10–G13	Systemic atrophies primarily affecting the central nervous system
G20–G26	Extrapyramidal and movement disorders
G30–G32	Other degenerative diseases of the nervous system
G35–G37	Demyelinating diseases of the central nervous system
G40–G47	Episodic and paroxysmal disorders
G50–G59	Nerve, nerve root and plexus disorders
G60–G64	Polyneuropathies and other disorders of the peripheral nervous system
G70–G73	Diseases of myoneural junction and muscle
G80–G83	Cerebral palsy and other paralytic syndromes
G90–G99	Other disorders of the nervous system
H00–H06	Disorders of eyelid, lacrimal system and orbit
H10–H13	Disorders of conjunctiva
H15–H22	Disorders of sclera, cornea, iris and ciliary body
H25–H28	Disorders of lens
H30–H36	Disorders of choroid and retina
H40–H42	Glaucoma
H43–H45	Disorders of vitreous body and globe
H46–H48	Disorders of optic nerve and visual pathways
H49–H52	Disorders of ocular muscles, binocular movement, accommodation and refraction
H53–H54	Visual disturbances and blindness
H55–H59	Other disorders of eye and adnexa
H60–H62	Diseases of external ear
H65–H75	Diseases of middle ear and mastoid
H80–H83	Diseases of inner ear
H90–H95	Other disorders of ear
I00–I02	Acute rheumatic fever
I05–I09	Chronic rheumatic heart diseases
I10–I15	Hypertensive diseases
I20–I25	Ischaemic heart diseases
I26–I28	Pulmonary heart disease and diseases of pulmonary circulation

I30–I52	Other forms of heart disease
I60–I69	Cerebrovascular diseases
I70–I79	Diseases of arteries, arterioles and capillaries
I80–I89	Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified
I95–I99	Other and unspecified disorders of the circulatory system
J00–J06	Acute upper respiratory infections
J10–J18	Influenza and pneumonia
J20–J22	Other acute lower respiratory infections
J30–J39	Other diseases of upper respiratory tract
J40–J47	Chronic lower respiratory diseases
J60–J70	Lung diseases due to external agents
J80–J84	Other respiratory diseases principally affecting the interstitium
J85–J86	Suppurative and necrotic conditions of lower respiratory tract
J90–J94	Other diseases of pleura
J95–J99	Other diseases of the respiratory system
K00–K14	Diseases of oral cavity, salivary glands and jaws
K20–K31	Diseases of oesophagus, stomach and duodenum
K35–K38	Diseases of appendix
K40–K46	Hernia
K50–K52	Noninfective enteritis and colitis
K55–K63	Other diseases of intestines
K65–K67	Diseases of peritoneum
K70–K77	Diseases of liver
K80–K87	Disorders of gallbladder, biliary tract and pancreas
K90–K93	Other diseases of the digestive system
L00–L08	Infections of the skin and subcutaneous tissue
L10–L14	Bullous disorders
L20–L30	Dermatitis and eczema
L40–L45	Papulosquamous disorders
L50–L54	Urticaria and erythema
L55–L59	Radiation-related disorders of the skin and subcutaneous tissue
L60–L75	Disorders of skin appendages
L80–L99	Other disorders of the skin and subcutaneous tissue

M00–M25	Arthropathies
M00–M03	Infectious arthropathies
M05–M14	Inflammatory polyarthropathies
M15–M19	Arthrosis
M20–M25	Other joint disorders
M30–M36	Systemic connective tissue disorders
M40–M54	Dorsopathies
M40–M43	Deforming dorsopathies
M45–M49	Spondylopathies
M50–M54	Other dorsopathies
M60–M79	Soft tissue disorders
M60–M63	Disorders of muscles
M65–M68	Disorders of synovium and tendon
M70–M79	Other soft tissue disorders
M80–M94	Osteopathies and chondropathies
M80–M85	Disorders of bone density and structure
M86–M90	Other osteopathies
M91–M94	Chondropathies
M95–M99	Other disorders of the musculoskeletal system and connective tissue
N00–N08	Glomerular diseases
N10–N16	Renal tubulo-interstitial diseases
N17–N19	Renal failure
N20–N23	Urolithiasis
N25–N29	Other disorders of kidney and ureter
N30–N39	Other diseases of urinary system
N40–N51	Diseases of male genital organs
N60–N64	Disorders of breast
N70–N77	Inflammatory diseases of female pelvic organs
N80–N98	Noninflammatory disorders of female genital tract
N99	Other disorders of genitourinary tract
O00–O08	Pregnancy with abortive outcome
O09	Duration of pregnancy
O10–O16	Oedema, proteinuria and hypertensive disorders in pregnancy, childbirth and the puerperium

O20–O29	Other maternal disorders predominantly related to pregnancy
O30–O48	Maternal care related to the fetus and amniotic cavity and possible delivery problems
O60–O75	Complications of labour and delivery
O80–O82	Delivery
O85–O92	Complications predominantly related to the puerperium
O95–O99	Other obstetric conditions, not elsewhere classified
P00–P04	Fetus and newborn affected by maternal factors and by complications of pregnancy, labour and delivery
P05–P08	Disorders related to length of gestation and fetal growth
P10–P15	Birth trauma
P20–P29	Respiratory and cardiovascular disorders specific to the perinatal period
P35–P39	Infections specific to the perinatal period
P50–P61	Haemorrhagic and haematological disorders of fetus and newborn
P70–P74	Transitory endocrine and metabolic disorders specific to fetus and newborn
P75–P78	Digestive system disorders of fetus and newborn
P80–P83	Conditions involving the integument and temperature regulation of fetus and newborn
P90–P96	Other disorders originating in the perinatal period
Q00–Q07	Congenital malformations of the nervous system
Q10–Q18	Congenital malformations of eye, ear, face and neck
Q20–Q28	Congenital malformations of the circulatory system
Q30–Q34	Congenital malformations of the respiratory system
Q35–Q37	Cleft lip and cleft palate
Q38–Q45	Other congenital malformations of the digestive system
Q50–Q56	Congenital malformations of genital organs
Q60–Q64	Congenital malformations of the urinary system
Q65–Q79	Congenital malformations and deformations of the musculoskeletal system
Q80–Q89	Other congenital malformations
Q90–Q99	Chromosomal abnormalities, not elsewhere classified
R00–R09	Symptoms and signs involving the circulatory and respiratory systems
R10–R19	Symptoms and signs involving the digestive system and abdomen
R20–R23	Symptoms and signs involving the skin and subcutaneous tissue
R25–R29	Symptoms and signs involving the nervous and musculoskeletal systems
R30–R39	Symptoms and signs involving the urinary system
R40–R46	Symptoms and signs involving cognition, perception, emotional state and behaviour

R47–R49	Symptoms and signs involving speech and voice
R50–R69	General symptoms and signs
R70–R79	Abnormal findings on examination of blood, without diagnosis
R80–R82	Abnormal findings on examination of urine, without diagnosis
R83–R89	Abnormal findings on examination of other body fluids, substances and tissues, without diagnosis
R90–R94	Abnormal findings on diagnostic imaging and in function studies, without diagnosis
R95–R99	Ill-defined and unknown causes of mortality
S00–S09	Injuries to the head
S10–S19	Injuries to the neck
S20–S29	Injuries to the thorax
S30–S39	Injuries to the abdomen, lower back, lumbar spine and pelvis
S40–S49	Injuries to the shoulder and upper arm
S50–S59	Injuries to the elbow and forearm
S60–S69	Injuries to the wrist and hand
S70–S79	Injuries to the hip and thigh
S80–S89	Injuries to the knee and lower leg
S90–S99	Injuries to the ankle and foot
T00–T07	Injuries involving multiple body regions
T08–T14	Injuries to unspecified part of trunk, limb or body region
T15–T19	Effects of foreign body entering through natural orifice
T20–T31	BURNS
T20–T25	Burns of external body surface, specified by site
T26–T28	Burns of eye and internal organs
T29–T31	Burns of multiple and unspecified body regions
T33–T35	Frostbite
T36–T50	Poisoning by drugs, medicaments and biological substances
T51–T65	Toxic effects of substances chiefly nonmedicinal as to source
T66–T78	Other and unspecified effects of external causes
T79	Certain early complications of trauma
T80–T88	Complications of surgical and medical care, not elsewhere classified
T89	Other complications of trauma not elsewhere classified
T90–T98	Sequelae of injuries, of poisoning and of other consequences of external causes
V01–X59	Accidents

V01–V99	Transport accidents
V01–V09	Pedestrian injured in transport accident
V10–V19	Pedal cyclist injured in transport accident
V20–V29	Motorcycle rider injured in transport accident
V30–V39	Occupant of three-wheeled motor vehicle injured in transport accident
V40–V49	Car occupant injured in transport accident
V50–V59	Occupant of pick-up truck or van injured in transport accident
V60–V69	Occupant of heavy transport vehicle injured in transport accident
V70–V79	Bus occupant injured in transport accident
V80–V89	Other land transport accidents
V90–V94	Water transport accidents
V95–V97	Air and space transport accidents
V98–V99	Other and unspecified transport accidents
W00–X59	Other external causes of accidental injury
W00–W19	Falls
W20–W49	Exposure to inanimate mechanical forces
W50–W64	Exposure to animate mechanical forces
W65–W74	Accidental drowning and submersion
W75–W84	Other accidental threats to breathing
W85–W99	Exposure to electric current, radiation and extreme ambient air temperature and pressure
X00–X09	Exposure to smoke, fire and flames
X10–X19	Contact with heat and hot substances
X20–X29	Contact with venomous animals and plants
X30–X39	Exposure to forces of nature
X40–X49	Accidental poisoning by and exposure to noxious substances
X50–X57	Overexertion, travel and privation
X58–X59	Accidental exposure to other and unspecified factors
X60–X84	Intentional self-harm
X85–Y09	Assault
Y10–Y34	Event of undetermined intent
Y35–Y36	Legal intervention and operations of war
Y40–Y84	Complications of medical and surgical care
Y40–Y59	Drugs, medicaments and biological substances causing adverse effects in therapeutic use

Y60–Y69	Misadventures to patients during surgical and medical care
Y70–Y82	Medical devices associated with misadventures in diagnostic and therapeutic use
Y83–Y84	Surgical and other medical procedures as the cause of abnormal reaction of the patient, or of later complication, without mention of misadventure at the time of the procedure
Y85–Y89	Sequelae of external causes of morbidity and mortality
Y90–Y98	Supplementary factors related to causes of morbidity and mortality classified elsewhere
Z00–Z13	Persons encountering health services for examination and investigation
Z20–Z29	Persons with potential health hazards related to communicable diseases
Z30–Z39	Persons encountering health services in circumstances related to reproduction
Z40–Z54	Persons encountering health services for specific procedures and health care
Z55–Z65	Persons with potential health hazards related to socioeconomic and psychosocial circumstances
Z70–Z76	Persons encountering health services in other circumstances
Z80–Z99	Persons with potential health hazards related to family and personal history and certain conditions influencing health status