

Adding Value to NZSSDS

Underpinning Transparency in **Research**: establishing a template for a research repository with real-world examples

Modifying and Archiving Survey Data Sets for **Teaching** Purposes

Who are we?

Faculty of Science:

- Karl Parker- BSC PGDip/MSc in Medical Statistics
- Carmen Lim- BSC in Statistics

Faculty of Arts:

- Kaylyn Oon – BA/BCom in Statistics, Marketing and Operations Management

Outline

- Aim
- Other examples
- How NZSSDS could be improved
- What we did
- How to integrate our materials in NZSSDS
- What more could be done
- Challenges we faced
- Lessons we learnt
- Conclusion

Data Archiving - NZSSDS

- Aim
- Other examples
- How NZSSDS could be improved

Aim

- The NZ Social Science Data Service is an ongoing initiative being undertaken by COMPASS.
- The vision is that data sets, all well documented with metadata for users, will be made publicly available for examination, basic analysis online and authorised download (via Nesstar software) to the wider research community.
- Key published papers will be documented along with computer code for data manipulation and analysis.
- Allow greater transparency since results can be easily replicated
- Used for teaching purposes.

What ideas are out there on the web

➤ Economic and Social Data Service

- Includes both numeric data and qualitative material such as in-depth interviews, diaries, anthropological field notes and complete answers to survey questions.
- Data sets downloadable in various formats
- No code is given as to how to analyse data

The screenshot shows the homepage of the Economic and Social Data Service (ESDS). At the top, there is a navigation bar with links for Home, A-Z index, Site map, Contact, Login, and a search box. Below this is a main header with the ESDS logo and a navigation menu with tabs for About, Data, Support, Resources, Create and deposit, News, Events, and a dropdown for 'Which service?'. The main content area is divided into several sections: 'Search the Data Catalogue' with a search box and 'GO' button; 'Themes' listing 'Health and health behaviour' and 'Crime and social control'; 'Creating and depositing data' providing guidance on standards; 'Teaching and learning with ESDS' offering advice for educators; 'ESDS Services' listing 'ESDS Government', 'ESDS International', 'ESDS Longitudinal', and 'ESDS Qualidata'; 'In the spotlight...' featuring a 'Complex sample design guide and metadata'; 'NEW USERS' with a FAQ; 'ESDS GUIDES' listing various guides; and 'LATEST NEWS' with a photo of the UK Data Archive Director and links to 'ESDS Qualidata study', 'SDS moves to full service', 'OECD Forum', and 'Political science courses data'. There are also links for 'more news' and 'ESDS newsfeed'.

What ideas are out there on the web

➤ My Experiment

- Collaborative environment.
- Social networking elements, solely focused in sharing research based materials.
- User profiles, groups, tags

➤ Many others..

The screenshot shows the myExperiment website interface. At the top left is the logo "my experiment" with "my" in red and "experiment" in green. To the right is a yellow box with the text: "myExperiment makes it easy to **find, use and share scientific workflows** and other **Research Objects**, and to build **communities**." Below this is a search bar with a dropdown menu set to "All" and a "Search" button. The main content area is divided into three columns. The left column is titled "First time visitor? Try these videos:" and lists "Project Introduction" and "Bioinformatics Case Study". Below this is "Use myExperiment to..." with a list of actions: "Find Workflows", "Find Files", "Share Your Workflows and Files", "Create and Find Packs of Items", "Create and Join Groups", "Find People and Make Friends", "Send Messages", "Get Feedback", "Tag and Rate things", "Write Reviews and Comments", and "Build your Profile and Reputation". The middle column has an "Explore" button, a workflow diagram, and an "About myExperiment" section with links for "Join the Mailing List", "Give us Feedback For Developers", "The myGrid Project", "Taverna Workflow Workbench", "The BioCatalogue Project", and "myExperiment Publications". The right column has a "Register" button, a login section with "Username or Email:" and "Password:" fields, a "Remember me:" checkbox, an "Or use OpenID:" section with a field and "(eg: name.myopenid.com)" example, and a "Login" button. At the bottom, a blue bar states: "myExperiment has over 2700 users, 160 groups, 850 workflows, 200 files and 70 packs".

How could NZSSDS be improved

- Provide a single document – ‘user guide’: a pdf containing the questionnaire, variables, derived variables, interviewer instructions/notes, coding etc.
- New users page which tells them about essential/important features of the website (FAQs).
- Provide public with a list of “most searched items”.
- Dedicate an entire page to search results, where each result contains all the information concerning the project and related materials.
- Provide a space where users can give feedback.
- Make it easier to find the data sets section of the website.
- Separate yet integrated ‘research’ and ‘teaching’ spaces depending on need with maximum visibility and minimal drilling down.

Ideas implemented, what we did.

Articles/Surveys

- Karl's - **Do Hospital Bed Reduction and Multiple System Reform Affect Patient Mortality? A Trend and Multilevel Analysis in New Zealand Over the Period 1988 – 2001** – Medical care (2007)
- Carmen's –**Early Sex and Its Behavioral Consequences in New Zealand- (PRS)-** (1991)
- Kaylyn- **Quality of hospital care for Maori patients in New Zealand: retrospective cross-sectional assessment- (NZQHS) -** hospital records data (1998)
- Carmen and Kaylyn - **The Impact of Organizational and Professional Context on Social Variation in Clinical Activity: A multi-level analysis of primary care** NatMedCa 2001/2

Medical care - Karl

- Do Hospital Bed Reduction and Multiple System Reform Affect Patient Mortality? A Trend and Multilevel Analysis in New Zealand Over the Period 1988 – 2001.
- To assess the impact of hospital system restructuring on the quality of care.
- Used a sample of approx six million hospital admittance records over the time period 1988 to 2001.
- Used summary statistics and mixed models to analysis trends.
- The results show that despite a decline in hospital beds and multiple reforms did not affect 60 day post admission death.

The Project

- The Beginning
- The Process
- The Product (Master File)

The Beginning

➤ We started with the following files and info:

- SAS Files - between 80 to 160 files.
- Log/output – also between 80 to 160 files.
- Article
- Datasets – up to 6 million observations.
- Variables/original data dictionary.
- Additional files – excel spreadsheets for calculations.

The Process

- Read article
- Familiarised with variables
- Locating figures in tables from SAS output
- Organising code into one file for each table presented in the journal article
- Rerunning SAS programmes
- Read supplementary material to understand topic
- Check mistakes
- Create Master File

The Product

➤ Master File

- This is the documentation that contains everything that a user would need to know about the study
- The purpose of the creation of the master file is to ensure that the data and the analytical process are easy for users or students to understand and follow
- Adapted from the Scottish Health Survey, Joint Health Survey Unit from the University College London

Master File

- Instructions to user
- User Guide
- Code and Output

Quality of hospital care for Maori patients in New Zealand - Kaylyn

- Retrospective cross-sectional assessment
- NZ Quality of Healthcare Study (NZQHS) - hospital records data (1998)
- Objective: Assess possible disparities in quality of hospital care for Maori with data on preventable adverse events as an indicator of suboptimum treatment
- Cross sectional survey of general public hospitals with more than 100 beds
- 13 hospitals selected, 6579 patients sampled
- Findings:
 - Significant differences between groups: 14% of Maori admissions were associated with an adverse event vs. 11% for non-Maori/non-Pacific patients
 - No specific causes for this were evident, but various policy and system issues can be addressed

Instructions to User

- A contents page where NZSSDS users are able to quickly locate various types of information about any given journal article
- Provides hyperlinks under 3 clear sections – User guide, Supporting documents and Codes and output

ADVERSE EVENTS IN NEW ZEALAND PUBLIC HOSPITALS, 1998 INSTRUCTIONS TO USER

How to use this file:

This file provides links between the various documentation files associated with this data set. Click [OPEN](#) to go to the file you require. Each of the documentation files has a return link to this file in the bookmarks window.

OPEN	User Guide Includes article details, user guide, original and derived variables
OPEN	Supporting Documents Includes review forms (RF1 and RF2), codebook, the published journal article, related publications and corresponding dataset/s
OPEN	Codes and Outputs Includes SAS output of tables in the journal article, annotated code of analyses used in a word document and annotated code in a SAS program

User Guide

- **Purpose:**

Introduce the background & analysis of the study to the users who intend to plan on extending the analysis of the study

- **Divided into five sections:**

1. Article Details
 - Background, Objective, Methods, Conclusion
2. Purpose of Metadata
3. Using the Data
 - Variables on the files
4. Analyses
5. Data Dictionary
 - Original Variables
 - Derived Variables



Quality of hospital care for Maori patients in New Zealand: retrospective cross-sectional assessment

Adverse events in New Zealand public hospitals, 1998

User Guide

Principal findings from a national survey carried out by:

Peter Davis
Roy Lay-Yee
Lorna Dyal
Robin Bryant
Andrew Sporie
Deborah Brunt
Alastair Scott

[Acknowledgements](#)

We wish to acknowledge the National Centre for Social Research and the Department of Epidemiology and Public Health, University College of London for the structure in which we based our Master File and User Guide.

Article Details

- **Background** →
 - Contains a brief abstract of the study
- **Objective of the study** →
- **Methods** →
 - Survey that the study based on
 - Survey design
- **Conclusion** →
 - Implications
 - Reliability & Validity of the data

2.0 Article Details

2.1 Background

A steady trend towards earlier sexual intercourse is now well documented. However the relationship of this development to trends in other early sexual formative experiences has not been explored nor has there been an analysis of the longer term of the behavioural consequences. The data examined in this paper were drawn from a two-stage national survey of sexual lifestyles on a representative sample of 2361 adult New Zealanders between the ages of 18-54. Techniques of survival analysis and multiple logistic regression were used. Over time, there was a consistent decline and diminishing gap in age of onset for first experience, first intercourse, and first regular partnership. Multivariate analysis confirms that males, the young, less educated, ethnic minorities and respondents professing no religious affiliations were all more likely to report earlier onset of sexual experiences. However, while all three socio-sexual events showed the same trend, pre-intercourse experiences exerted the strongest influence on subsequent sexual practices and outcomes. This has implications for preventative strategies.

2.2 Objective

The objective of this study is to describe the sequence of initiation into early sexual activity and partnering and to analyse the behavioural consequences of this process.

2.3 Methods

This study was carried out as part of The New Zealand Partner Relations Survey (PRS) funded by the Health Research Council of New Zealand in 1991. PRS is modelled based on a protocol that is being developed by the World Health Organisation's global programme on AIDS (GPA). It is a two stage stratified national survey where households in New Zealand are selected at random using random digit-dialling (RDD) with a single eligible interviewee per selected household, followed by a subsample survey of non-contacts and refusals to determine eligibility. A 15 minute questionnaire based on (GPA) is administered using computer assisted telephone interviewing (CATI) to a nationally representative sample of 2361 respondents between the age of 18-54. Questionnaire items presented in the paper are grouped into three categories: contextual precursors (socio-demographic and cultural background factors), socio-sexual development (measures of early sexual activity and partnership), and behavioural consequences (reported sexual practices and outcomes).

2.4 Conclusion

The study confirms a steady trend towards earlier sexual experience and a rapid progress to a long term sexual partnership. Those that engage in early pre-intercourse sexual experience are twice as likely to report a greater number of partners, anal sex, same gender sex, and STD in their later life. If the trend carries on, this has clear implications towards future risk behaviours and related risk profiles. These implications also affect preventative strategies which are currently in place. For example, preventative strategies must be communicated at an earlier age with a different focus in order to educate the public who are to be sexually active. As a result, these strategies will potentially need to be altered to accommodate for the consequences of early pre-intercourse sexual experiences. In regards to the reliability and validity of the data, a number of methodological issues also have to be taken into account. Firstly, even though the findings conform to other literature regarding patterns of early sexual activity and partnership in Western Europe and North America, due to the relatively small population of New Zealand and thus smaller sample size, it is uncertain that such results can be generalised to other parts of the world. Secondly, it is important to note that as the survey was conducted via telephone, it follows that only people with telephones could be interviewed (frame errors). However, given that 95% of the population had a telephone at the time the study was conducted and little discrepancies of results were found between the telephone interviews and face to face interviews on health related topics implies that differences which do arise within the study are not detrimental to the findings. And finally, despite the sensitivity of the topic, the study provides and demonstrates acceptable quality of the data as the overall non response in the survey was less than one percent and not more than five percent for the most sensitive questions.

Purpose of Metadata

- Organization of the metadata
- Data and Analytical process is user friendly and easy to understand
- Instructions to the user on how to use the file
- Description of each files for the study
- (Data Dictionary, annotated SAS outputs, questionnaires of the survey etc.)

3.0 Purpose of Metadata

The documentation has been organised into the following sections:

- CATI questionnaires & Code book for the Survey
- File containing the dataset
- Data Dictionary
- Annotated code and associated output (SAS programme also included)
- Related publications

The metadata has been structured in such a way to ensure the data and analytical process is user-friendly and easy to understand. The CATI questionnaires and the codebook have been included to allow users to take note of how the interviews were administered. It is also made available for further reference of variables if needed. The data dictionary provides a list of all the variables, variable names, a description of variables and their values. It includes both original and derived variables created in the course of the analysis. The annotated SAS code and associated output shows and describes each section of the provided SAS code. More specifically, annotations are given to describe a SAS procedure and the related output. The code is organized by table, according to the journal article. The SAS program contains annotated code without the output. The code is sorted by table and can be downloaded and run in SAS. Related publications have also been included to further assist the user in understanding the study and also to contemplate the results relative to other studies of similar topics.

Using the Data

- Description of the datasets
- Variables on the files
 - Information on how to re-run or extend the current analysis
- Comments on the original and the derived variables

4.0 Using the Data

The 1991 data consists of one individual level file

prsall.ssd.sas7bdat	2361 records	Contains the original questionnaire variables plus new 'weighting' variable and 'socioeconomic-status' variables.
---------------------	--------------	---

4.1 Variables on the files

The dataset enclosed contains the codebook variables (original variables). In order to produce the derived variables which are also shown below, the corresponding SAS program is given in the Code and Output section of the Master File. The variables in the file are detailed in section six of this user guide. This section is the best place to start planning the analyses which you would like to conduct. It includes:

- Major categories of variables- e.g. Marital status, Ethnic groups, Region, Religion
- Sub categories of variables (values) - e.g. The types of religion within the religion category
- The variable name used in SAS

Once you have decided which of the derived variables to include in your analysis, you can look up further details of the variable in the original variable dictionary as well as the question wording used the CATI questionnaires. There are a number of analyses which can be performed on the given data set. Ones which have been used in the study include survival analysis and multiple logistic regression analysis. These analyses can be replicated using the code provided in the Code and Output section of the Master File. Alternatively, different analyses can also be conducted on the data set to answer questions of your own.

Analyses

- Analysis used for each table in the result section of the study
- SAS procedure
- Details of the analysis
- Comment on the result of the table, why analysis is used, weighting factor

5.0 Analyses

Table	Method	Details	Comment
1 Survival Analysis	PROC LIFETEST (product-limit) procedure	Censored variable is defined for the three socio-sexual activity Derive the number of censored observations	The results of the survey showed a large proportion of the respondents had not experienced any kind of sexual activity at the time of the interview. The exact survival time until the initiation of sexual activity of these respondents are unknown. These are called censored observation. Failure to take censored observation into account could produce serious bias in the study and hence survival analysis.
2 Survival Analysis	PROC LIFETEST (Life-table) procedure	Median age of onset for the three socio-sexual activity Raw median (R_{raw}): exclude censored observations Adjusted median (Adj): include censored observations Within each PROC LIFETEST, two separate analyses are performed using birth cohort and gender as the strata.	A raw, unadjusted median is likely to underestimate the true figure. Life-table method is used to produce non-integer medians for both raw and adjusted medians. The Wolcott and Log-Rank statistics are used to test the equality of distributions in the survival curves.
	PROC FREQ procedure	Sample Distribution	
3 Survival Analysis	PROC LIFETEST (Life-table) procedure	Median age of onset for the three socio-sexual activities. Strata: Birth Cohort By Gender	Adjusted median is used as censored value is taken into account.
4 Multiple Logistic Regression	PROC LOGISTIC procedure	Odds ratios for age of onset for the three socio-sexual activities.	The Socio-Demographic predictor variables are represented in a binary form.

Data Dictionary

- **Original variables**
 - Variables from the original questionnaire
- **Derived variables**
 - The original variable is processed, producing new variables in a format more appropriate for higher analysis

6.0 Data Dictionary

6.1 Original Variables

Variable Name	Description	Values
q53	Age of first sexual experience (years)	?? = hasn't happened yet
q57	Age of first sexual intercourse (years)	
q58	Age when first had regular partner (years)	
q54	Age group (years)	1= 18-19 2= 20-24 3= 25-29 4= 30-34 5= 35-39 6= 40-44 7= 45-49 8= 50-54
q56	Gender	1= male 2= female
q50	If never had regular partner, ever had sexual intercourse	1= yes 2= no
q48	Number of regular partners- ever	1-5= 1-5 6 = 6-9 7 = 10 8 = none
wt1	Weighting Factor (post-stratified and design-weighted)	
q57	Region	1= Northland 2= Auckland 3= Waikato/ Thames Valley 4= Bay of Plenty 5= Gisborne 6= Hawkes Bay 7= Tairāhaki 8= Central Districts 9= Wellington 10= Manukau 11= Westland 12= Canterbury 13= Otago 14= Southland
q52	Ethnic Group	1= European/ Pakeha 2= Maori 3= Pacific Island 4= Other
q53	Religion	1= Anglican 2= Presbyterian 3= Roman Catholic 4= Methodist 5= Baptist 6= Christian not specified 7= Rātana 8= Latter Day Saints 9= Brethren 10= Jewish 11= Jehovah's Witness 12= Pentecostal not specified 13= Assemblies of God 14= Seventh Day Adventist 15= Hindu 16= Buddhist 17= Other 18= No religion
ses	Socioeconomic status overall (respondent) after R. Johnstone's own occupation OR if married/de facto, maximum of own and partner's occupation where 1 is the highest class	1 to 6
q51	Highest level of education completed	1= No school 2= Primary 3= Secondary 4= Technical or vocational college 5= University 6= Other

Code and Output

- In order to make the analysis process easier for users of the website, we've located where the figures have come from, and have clearly shown which codes produce the corresponding output
- These are organized by table, in .rtf files

```

10:55 Monday, February 15, 2010 1
Table 2: weighted - means (sd)
The UNIVARIATE Procedure
Variable:
STAY
Weight:
wt
ethgrp=Maori

/*Table 2: means (sd)*/
*Univariate analysis to derive weighted hospital stay of adverse event cases by ethnicity (mean
and SD);
PROC univariate data=a;
TABLE 'Table 2: weighted - means (sd)';
WEIGHT wt;
WHERE sel=1 and (ethgrp='Maori' or ethgrp='nonMP');
BY ethgrp;
VAR stay;
RUN;

```

Weighted Moments			
N	135	Sum Weights	136.994806
Mean	7.58262353	Sum Observations	1038.78004
Std Deviation	13.8509026	Variance	191.847503
Skewness	6.97836579	Kurtosis	62.9050939
Uncorrected SS	33584.2434	Corrected SS	25707.5655
Coeff Variation	182.666563	Std Error Mean	1.18338491

ethgrp=nonMP

Weighted Moments			
N	683	Sum Weights	692.588157
Mean	10.0340332	Sum Observations	6949.45255
Std Deviation	15.0352331	Variance	226.058235
Skewness	4.08222916	Kurtosis	22.4957536
Uncorrected SS	223902.754	Corrected SS	154171.717
Coeff Variation	149.84237	Std Error Mean	0.57131107

Annotated Codes

- Made short notes about what the codes mean, what it does to the dataset and what analysis is used
- Annotated codes are available in SAS programs and in Word documents

Table 1

```
/*Table 1;
*unweighted frequencies of age group, sex, deprivation group, admission xxx
(acute) and diagnostic categories by ethnicity;
PROC FREQ DATA=a;
TITLE 'Table 1 - unweighted - freq';
TABLES agegp sex depgp acute mdcgp (agegp sex depgp acute mdcgp)*ethgp;
RUN;
*Sorting data file a by ethnic group for the univariate analysis of hospital
stay mean and SD;
PROC SORT DATA=a;
BY ethgp;
RUN;
*Univariate analysis to derive weighted total hospital stay mean and SD;
PROC UNIVARIATE DATA=a;
TABLES 'Table 1 - weighted - univariate';
WEIGHT wt;
VAR stay;
RUN;
*weighted percentages of socio-demographic characteristics and MDCs;
PROC FREQ DATA=a;
TITLE 'Table 1 - weighted - freq';
WEIGHT wt;
TABLES agegp sex depgp acute mdcgp (agegp sex depgp acute mdcgp)*ethgp;
RUN;
*Univariate analysis for weighted hospital stay mean and SD by ethnicity;
PROC UNIVARIATE DATA=a;
TITLE 'Table 1 - weighted - univariate';
BY ethgp;
WEIGHT wt;
VAR stay;
RUN;
```

Table 2

```
/*Table 2: means (sd)*/
*Univariate analysis to derive weighted hospital stay of adverse event cases
by ethnicity (mean and SD);
PROC UNIVARIATE DATA=a;
TITLE 'Table 2: weighted - means (sd)';
WEIGHT wt;
WHERE ael=1 and (ethgp='Maori' or ethgp='nonMP');
BY ethgp;
VAR stay;
RUN;
*age standardised incident characteristics of attributable days (mean and
SD);
PROC UNIVARIATE DATA=a;
TITLE 'Table 2: weighted, age-adjusted - means (sd)';
WEIGHT ethagewt;
WHERE ael=1 and (ethgp='Maori' or ethgp='nonMP');
BY ethgp;
VAR q15;
RUN;
```

Spreadsheet of Calculations

- This spreadsheet contains the calculations necessary to derive some of the figures that appear in the articles
- The resultant figures are **bolded**
- Formulas are displayed on the bottom of the page to help the user understand how figures were derived

European											
outcome= referral	fixed part (explained)	random part (unexplained)-residual variance		total variance	UN(1,1)	UN(2,2)	UN(2,1)	2*UN(2,1)	explained proportion of total variance(%)	proportion of total variance(%)	proportion of total variance(%)
	var(linear predictor)	var(doctor level 2)	var(patient:level 1)						(due to linear predictor)	(due to doctor)	(due to patient)
All	1.07116324	1.7145	3.28987	6.07553	1.8773	1.508	0.836	1.671	17.63	28.22	54.15
Independent	2.47967613	4.7471	3.28987	10.51665	3.1908	3.770	1.107	2.2134	23.58	45.14	31.28
Coordinated	1.43186229	1.4204	3.28987	6.14213	2.724	2.343	1.823	3.6466	23.31	23.13	53.56
Capitated	0.92934973	0.9497	3.28987	5.16892	1.0413	0.710	0.401	0.8014	17.98	18.37	63.65
Salaried	0.81725581	2.8571	3.28987	6.96423	1.3643	0.806	-0.344	-0.687	11.74	41.03	47.24

Deprived											
outcome= referral	fixed part (explained)	random part (unexplained)-residual variance		total variance	UN(1,1)	UN(2,2)	UN(2,1)	2*UN(2,1)	explained proportion of total variance(%)	proportion of total variance(%)	proportion of total variance(%)
	var(linear predictor)	var(doctor level 2)	var(patient:level 1)						(due to linear predictor)	(due to doctor)	(due to patient)
All	1.0226956	5.3544	3.28987	9.6697	1.9815	1.539	0.917	1.8338	10.58	55.39	34.03
Independent	1.8722994	8.6584	3.28987	13.82057	3.4243	2.849	1.193	2.3856	13.55	62.65	23.80
Coordinated	1.43154422	6.6846	3.28987	11.40601	2.3878	1.610	1.343	2.6866	12.55	58.61	28.84
Capitated	0.93405456	1.2430	3.28987	5.46692	1.4706	1.344	0.786	1.5718	17.09	22.74	60.11
Salaried	0.80548257	1.7247	3.28987	5.82007	1.1042	0.779	-0.079	-0.15828	13.84	29.63	56.53

Not deprived											
outcome= referral	fixed part (explained)	random part (unexplained)-residual variance		total variance	UN(1,1)	UN(2,2)	UN(2,1)	2*UN(2,1)	explained proportion of total variance(%)	proportion of total variance(%)	proportion of total variance(%)
	var(linear predictor)	var(doctor level 2)	var(patient:level 1)						(due to linear predictor)	(due to doctor)	(due to patient)
All	1.0226956	1.6868	3.28987	5.99937	1.9815	1.539	0.917	1.8338	17.05	28.12	54.8
Independent	1.8722994	3.8872	3.28987	9.04937	3.4243	2.849	1.193	2.3856	20.69	42.96	36.3
Coordinated	1.43154422	1.3114	3.28987	6.03281	2.3878	1.610	1.343	2.6866	23.73	21.74	54.5
Capitated	0.93405456	1.2430	3.28987	5.46692	1.4706	1.344	0.786	1.5718	17.09	22.74	60.11
Salaried	0.80548257	2.0413	3.28987	6.13653	1.1042	0.779	-0.079	-0.15828	13.13	33.26	53.6

Percentage of total variance:
 Ethnicity: Non Europeans
 Practice type: All, Independent, Coordinated, Capitated, Salaried
 Linear predictor (fixed) (variance for linear predictor/total variance)*100
 Practitioner (random) (variance for doctors [level 2]/total variance)*100
 Patient (random) (variance for patients [level 1]/total variance)*100

Codes and Outputs Compiled

- Provides direct hyperlinks to all the given code and output for each article by table

Code and Output

Table	Name	Size in Kilobyte	Description	Comments
1	CombinedTable1.docx	34	Annotated SAS codes and relevant outputs	
	Table1.sas	6	Table 1 SAS codes	
	Table1a.rtf	130	PROCFREQ output: Age group, Sex, Deprivation group, Acute, and MDC group by Ethnicity	<i>Figures in table 1 are bolded</i>
2	Table1b.rtf	187	PROCFREQ and PROCUNIVARIATE output: By Ethnicity	
	CombinedTable2.docx	51	Annotated SAS codes and relevant outputs	
	Table2.sas	4	Table 2 SAS codes	
	Table2a.rtf	135	PROC UNIVARIATE output: Weighted means by Ethnicity	<i>Figures in table 2 are bolded</i>
			PROCFREQ output: Total percentages by Ethnicity	
Table2b.rtf	195	PROCFREQ output: Weighted and unweighted figures by Ethnicity		
Table2c.rtf	202	PROCSURVEYREG output: Testing for significant differences of adverse events by Ethnicity		
3	CombinedTable3.docx	35	Annotated SAS codes and relevant outputs	
	Table3.sas	2	Table 3 SAS codes	
	Table3a.rtf	171	PROCFREQ output: Weighted and unweighted rates of In-hospital Adverse Events	<i>Figures in table 3 are bolded</i>
4	Table3b.rtf	135	PROCSURBEYREG output: Testing for significant differences of In-hospital Adverse Events by Ethnicity	
	CombinedTable4.docx	20	Annotated SAS codes and relevant outputs	
	Table4.sas	0.9	Table 4 SAS codes	
	Table4a.rtf	66	PROCSURVEYLOGISTIC output: Odds ratios of patient related factors	<i>Figures in table 4 are bolded</i>
5	Table4b.rtf	117	PROCSURVEYFREQ (Rao-Scott Chi square test) output: Testing for significant differences of Odds ratios within variables	
	CombinedTable5.docx	24	Annotated SAS codes and relevant outputs	
	Table5.sas	0.8	Table 5 SAS codes	
	Table5a.rtf	118	PROCFREQ output: Unweighted figures of preventable adverse events by Area of Effort for Prevention and Ethnicity	<i>Figures in table 5 are bolded</i>
	Table5b.rtf	118	PROCFREQ output: Weighted percentages of preventable adverse events by Area of Effort for Prevention and Ethnicity	

Summary

- How to integrate our materials in NZSSDS
- Challenges we faced
- Lessons we learnt
- What more could be done
- Conclusions

Early Sex and Its Behavioural Consequences in New Zealand - Carmen

- Objective: Investigate the sexual behaviour of adults during their adolescent years and analyse the behavioural consequences of this process
- Partner Relations Survey 1991 (PRS)
- 2-stage stratified national survey administered using CATI
- Findings: The study confirms a steady trend towards earlier sexual experience and a rapid progression to a long term sexual partnership.
 - Implication towards future risk behaviours
 - Preventative strategies

The Impact of Organizational and Professional Context on Social Variation in Clinical Activity

- Objective: To assess whether organizational and professional context of primary care is associated with practitioner clinical decisions for patients from lower socio-economic areas or of minority ethnic background
- NatMedCa 2001/02
- Multi-stage probability sample involving 244 physicians & 9,272 patients across 185 practices in New Zealand
- Findings: Practitioner variability was more defined for patients from ethnically and socio-economically disadvantaged backgrounds
- Influence of organizational setting

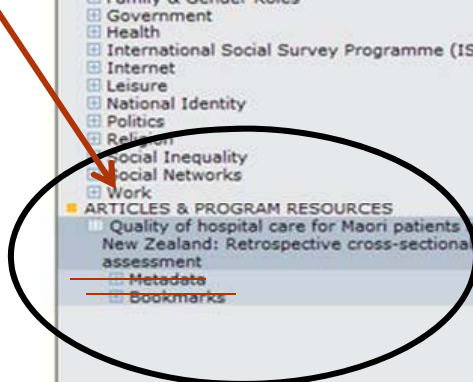
How to integrate our materials in NZSSDS

- Four journal articles to incorporate.
- These will be added to the “article and program resources” section of the website.
- Contains links to and from data sets
- **Suggestions:**
 - Mock up 1 (next slide)
 - Mock up 2

Mock up (I)

Contents page that contains links to other supporting document

1. Documented articles



2. Additional tab for "New Articles Documented" →

Instructions to User

- A contents page where NZSSDS users are able to quickly locate various types of information about any given journal article
- Provides hyperlinks under 3 clear sections – User guide, Supporting documents and Codes and output

ADVERSE EVENTS IN NEW ZEALAND PUBLIC HOSPITALS, 1998 INSTRUCTIONS TO USER

How to use this file:

This file provides links between the various documentation files associated with this data set. Click [OPEN](#) to go to the file you require. Each of the documentation files has a return link to this file in the bookmarks window.

OPEN	User Guide Includes article details, user guide, original and derived variables
OPEN	Supporting Documents Includes review forms (RF1 and RF2), codebook, the published journal article, related publications and corresponding dataset/s
OPEN	Codes and Outputs Includes SAS output of tables in the journal article, annotated code of analyses used in a word document and annotated code in a SAS program

Mock up (II)

Study Description

1. Documented studies & Content of the master file

2. Documented studies

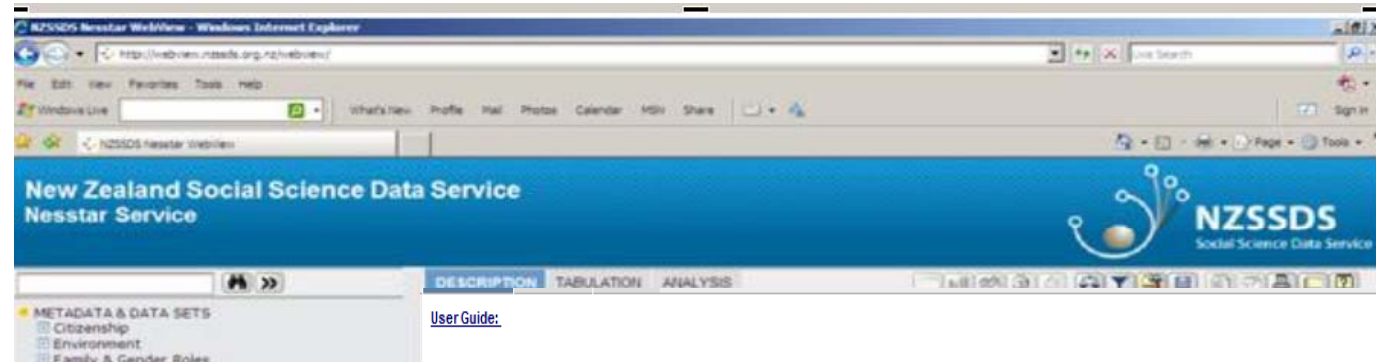


Table	Name	Size in Kilobyte	Description	Comments
1	AmotatedTables1.docx	34	Amotated SAS codes and relevant outputs	
	Table1.sas	6	Table 1SAS code	
	Table1.rtf	150	PROC FREQ output: Age group, Sex, Deprivation group, Acute, and VDC group by Ethnicity	Figures in table 1 are bolded
	Table2.rtf	187	PROC FREQ and PROC UNIVARIATE output: By Ethnicity	
2	AmotatedTables2.docx	51	Amotated SAS codes and relevant outputs	
	Table2.sas	4	Table 2SAS code	
	Table2.rtf	135	PROC UNIVARIATE output: Weighted means by Ethnicity	Figures in table 2 are bolded
	Table3.rtf	195	PROC FREQ output: Total percentages by Ethnicity	
	Table4.rtf	202	PROC FREQ output: Weighted and unweighted figures by Ethnicity	
3	AmotatedTables3.docx	35	Amotated SAS codes and relevant outputs	
	Table3.sas	2	Table 3SAS code	
	Table3.rtf	170	PROC SURVEYFREQ output: Testing for significant differences of adverse events by Ethnicity	Figures in table 3 are bolded
	Table4.rtf	135	PROC SURVEYFREQ output: Weighted and unweighted rates of In-hospital Adverse Events	
4	AmotatedTables4.docx	20	Amotated SAS codes and relevant outputs	
	Table4.sas	0.9	Table 4SAS code	
	Table4.rtf	66	PROC SURVEYFREQ output: Odds ratios of patient related factors	Figures in table 4 are bolded
	Table5.rtf	117	PROC SURVEYFREQ (Rao-Scott Chi square test) output: Testing for significant differences of Odds ratios within variables	
5	AmotatedTables5.docx	24	Amotated SAS codes and relevant outputs	
	Table5.sas	0.8	Table 5SAS code	
	Table5.rtf	118	PROC FREQ output: Unweighted figures of preventable adverse events by Area of Effort for Prevention and Ethnicity	Figures in table 5 are bolded
	Table6.rtf	118	PROC FREQ output: Weighted percentages of preventable adverse events by Area of Effort for Prevention and Ethnicity	

What more could be done (I)

➤ Project

- More integration between article and documentation/code used
 - Eventually place hyperlinks in the article itself
- Backlog of older journal articles to be packaged and put online
- For teaching purposes, a subset of the original datasets could be used
 - Additional weighting variable can be used to accommodate for the smaller number of observations
 - Variables may be recoded for the ease of use

What more could be done (II)

➤ On NZSSDS website

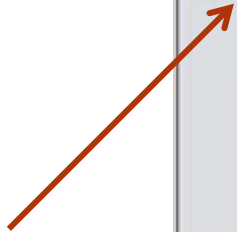
- Page containing MF of all major studies conducted
- Place a search engine on the home page
- Promote NZSSDS in the University of Auckland Statistics website
- Research space & teaching space in NZSSDS (next slide)

Proposed Research and Teaching Space

Research
Datasets



Teaching
Data sets



Challenges we faced

- Locating figures in the output that appear in the tables
- Deleting unnecessary output in the Code and Output i.e. Including output which contained figures in the article
- Moving output from SAS to Excel to conduct calculations
- Grouping all metadata components in a logical manner- data dictionary, SAS code and output, SAS program into a “Master File” with article details, purpose of metadata and instructions to the user

Lessons we learnt through this process

- Data not archived is rarely used or even lost
- Reduce the cost duplications of data collection
- Need to be meticulous when cleaning code
- Need to make instructions to users obviously clear- avoid miscommunications which could lead to costly errors
- Always need to understand theory behind analyses
 - Mixed modelling/multi-level analysis (easier to locate figures)
 - Survival analysis
 - Survey logistic regression, survey freq, survey regression

Conclusion

- 4 key published journal article has now been documented with metadata for users
- The 4 master files & final data sets with full mark up awaiting to be uploaded to NZSSDS archive
- All the articles has been documented along with SAS codes for data manipulation and analysis
- Hopefully, this would assist the users in understanding more about the study, hence fortify future investigations & scientific inquiries

We want to thank.....

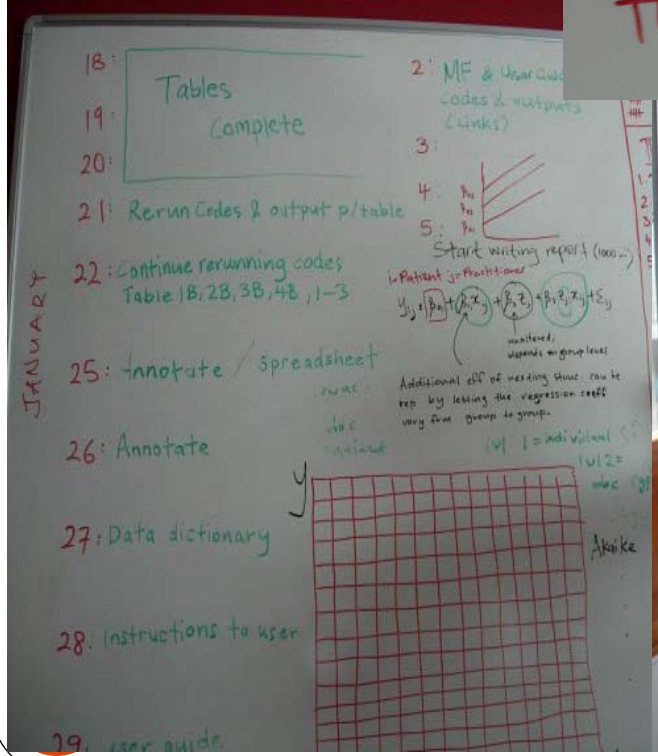
- Faculty of Science and Arts
- Roy Lay-Yee
- Peter Davis

- And of course.....
- **All** the lovely people at COMPASS

Studentship experience



The HEAT is on!!



QUESTIONS??

