



4324.0

Information Paper

National Health Survey

Confidentialised Unit Record Files

2001

Information Paper

National Health Survey

Confidentialised Unit Record Files

2001

**Dennis Trewin
Australian Statistician**

AUSTRALIAN BUREAU OF STATISTICS

EMBARGO: 11.30AM (CANBERRA TIME) FRI 19 SEP 2003

ABS Catalogue No. 4324.0
ISSN 0817-9344

© Commonwealth of Australia 2003

This work is copyright. Apart from any use as permitted under the *Copyright Act 1968*, no part may be reproduced by any process without prior written permission from the Commonwealth. Requests and inquiries concerning reproduction and rights in this publication should be addressed to The Manager, Intermediary Management, Australian Bureau of Statistics, Locked Bag 10, Belconnen ACT 2616, by telephone (02) 6252 6998, fax (02) 6252 7102, or email <intermediary.management@abs.gov.au>.

In all cases the ABS must be acknowledged as the source when reproducing or quoting any part of an ABS publication or other product.

Produced by the Australian Bureau of Statistics

INQUIRIES

For information about the 2001 National Health Survey or about the data contained on the confidentialised unit record files contact <health@abs.gov.au> or telephone Canberra (02) 6252 6415.

For Information about obtaining access to the files contact <curf.management@abs.gov.au> or telephone Canberra (02) 6252 5853.

CONTENTS

page

OVERVIEW

Introduction	5
About the survey	5
About the microdata	6
Estimation	7
Sampling error	7
About the data	8
About the files	9
Conditions of release	9
Conditions of sale	10
Price	10
How to order	10

APPENDIXES AVAILABLE SEPARATELY ON ABS WEB SITE

- 1** Differences between the BASIC and EXPANDED CURFs
- 2** List of data items contained on the BASIC and EXPANDED confidentialised unit record files
- 3** Selected population frequencies
- 4** File content

OVERVIEW

INTRODUCTION

This paper provides information about the release of microdata from the 2001 National Health Survey (NHS). The data are available through Confidentialised Unit Record Files (CURFs) released with the approval of the Australian Statistician. The files are available either on CD-ROM or through the Remote Area Data Laboratory (RADL).

The RADL is an on-line database query system under which microdata are held on a server at the ABS, to which users can submit programs to interrogate, analyse, model, etc. the data, and can access the results. Further information about this facility, can be found on the ABS web site, Access to ABS CURFs.

Two CURFs from the 2001 National Health Survey are available:

- The *BASIC file* — this file contains a basic set of data items at a reduced level of detail. It is available on CD-ROM and is also accessible through the RADL. The file was first released in April 2003. A revised edition of this file was released in September 2003.
- The *EXPANDED file* — this file contains an expanded set of data items at a more detailed level than the Basic file, and is accessible only through the RADL from September 2003. The earlier edition of the Expanded file which was released in April 2003 is superseded by the September edition.

Differences in the content between the Basic and Expanded files are summarised in Appendix 1. A full list of items contained on each of the files is provided as Appendix 2.

Unless otherwise indicated in this paper, the term CURF has been used to apply to both the Basic and Expanded versions of the file. The CURFs enable purchasers to tabulate, manipulate and analyse data to their own specifications. More detailed information to assist in using the CURFs and in interpreting the data is provided in the documentation on, or accompanying, the file, or otherwise through the ABS web site.

ABOUT THE SURVEY

The 2001 NHS is the fifth health survey of its type conducted by the ABS; previous surveys were conducted in 1977–78, 1983, 1989–90 and 1995. Published results and a CURF on CD-ROM are available from each of those surveys.

The 2001 NHS survey was conducted by the ABS in a sample of 17,918 private dwellings across Australia from February to November 2001. Both urban and rural areas in all states and territories were covered, but sparsely settled areas of Australia were excluded. Non-private dwellings, such as hotels, motels, hospitals, nursing and convalescent homes and short-stay caravan parks were not included in the survey.

Within each selected household, a random sub-sample of usual residents was selected for inclusion in the survey as follows:

- one adult (18 years of age and over)
- all children aged 0–6 years
- one child aged 7–17 years.

Trained ABS interviewers conducted personal interviews with selected residents of the sampled private dwellings. One person aged 18 years or more in each dwelling was selected and interviewed about their own health. An adult resident, nominated by the household, was interviewed about the health of each child selected within the dwelling. Adult female respondents were invited to complete a small additional questionnaire covering supplementary women's health topics.

OVERVIEW *continued*

ABOUT THE SURVEY

continued

The survey collected information about:

- the health status of the population, including long-term medical conditions experienced and recent injuries
- use of health services such as consultations with doctors and dentists, visits to hospital and other actions people have recently taken for their health
- health-related aspects of people's lifestyle, such as smoking, exercise and alcohol consumption
- demographic and socio-economic characteristics.

A supplementary survey of Aboriginal and Torres Strait Islander people was conducted in association with the 2001 NHS. Information about that survey, together with summary results were published in *National Health Survey: Aboriginal and Torres Strait Islander Results, Australia* (cat. no. 4715.0). Information from that survey is not included on any of the NHS CURFs outlined in this paper; confidentialised unit record information from that survey may be separately available through the RADL in late 2003.

A cross-section of results from the 2001 NHS was published in *2001 National Health Survey; Summary of Results, Australia* (cat. no. 4364.0) released in October 2001. Further results relating to specific topics covered by the survey will be published progressively over 2003–04, details of which will be posted on <www.abs.gov.au>.

Users will find on the ABS web site copies of the questionnaires used in the 2001 NHS, a listing of all data items available from the survey (a selection of which are included on the CURF), and a NHS Users Guide which contains detailed information about the survey design and operation, survey content, and data quality and interpretation. Further information about the NHS can be obtained by emailing <health@abs.gov.au> or telephoning (02) 6252 6415.

ABOUT THE MICRODATA

The NHS CURFs contains 26,862 confidentialised respondent records. Subject to the limitations of sample size and the data classifications used, it is possible to manipulate the data, produce tabulations and undertake statistical analyses to individual specifications.

The NHS microdata data are released under the *Census and Statistics Act 1905* which has provision for the release of data in the form of unit records where the information is not likely to enable the identification of a particular person or organisation. Accordingly, there are no names or addresses of survey respondents on the CURFs, and other steps have been taken to protect the confidentiality of respondents. These include removing some items from the CURFs, reducing the level of detail shown on the CURF for some other items, and swapping some characteristics between records.

Steps to confidentialise the datasets made available on the CURFs are taken in such a way as to ensure the integrity of the dataset and optimise its content, while maintaining the confidentiality of respondents. Intending purchasers should ensure that the data they require, at the level of detail they require are available on the CURFs; data obtained in the survey but not contained on the CURFs may be available as statistics in tabulated form on request. A full list of all the data items on each of the CURFs is provided as Appendix 2 to this paper.

OVERVIEW *continued*

ESTIMATION

As the survey was conducted on a sample of households in Australia, it is important to take account of the method of sample selection when deriving estimates from the CURF. This is particularly important as a person's chance of selection in the survey varied depending on the state or territory in which they lived. One of the fields on the CURF contains a 'main weight' for each person in the sample. This 'weight' is a value which indicates how many population units are represented by the sample unit. The weight provided is a 9-character field, which is the actual weight to 4 decimal places.

Where estimates are derived from the CURFs it is essential that they are calculated by adding the weights of persons in each category, and not just by counting the number of records falling into each category. If each person's 'weight' were to be ignored, then no account would be taken of a person's chance of selection or of different response rates across population groups, with the result that estimates produced could be seriously biased. The application of weights ensures that estimates will conform to an independently estimated distribution of the population by age, sex, state/territory and section of state, rather than to the distributions within the sample itself.

For weighting purposes, the 2001 NHS was benchmarked to the estimated population living in private dwellings at 30 June 2001, based on results from the 2001 Census of Population and Housing. Although persons in sparsely settled areas were not included in the NHS, these areas were not removed from benchmarks for practical reasons, except in Northern Territory. The effect on survey estimates is negligible.

SAMPLING ERROR

Sampling error is the difference between the published estimate and the value that would have been produced had all dwellings in scope of the survey been included.

In addition to the 'main weight' each record on the CURFs contains 30 'replicate weights'. The purpose of these replicate weights is to enable calculation of the sampling error on each estimate produced.

A basic idea behind the replication approach is to select subsamples repeatedly (30 times) from the whole sample. For each of these subsamples the statistic of interest is calculated. The variance of the full sample statistics is then estimated using the variability among the replicate statistics calculated from these subsamples. The formula which has been used by ABS in calculating the standard errors of estimates using the replicate weights technique is shown below:

$$\begin{aligned} & (\text{sqrt}((29/30)*(((V2-V1)*(V2-V1))+((V3-V1)*(V3-V1))+((V4-V1)*(V4-V1))+((V5-V1) \\ & *(V5-V1))+((V6-V1)*(V6-V1))+((V7-V1)*(V7-V1))+((V8-V1)*(V8-V1))+((V9-V1)* \\ & (V9-V1))+((V10-V1)*(V10-V1))+((V11-V1)*(V11-V1))+((V12-V1)*(V12-V1))+ \\ & ((V13-V1)*(V13-V1))+((V14-V1)*(V14-V1))+((V15-V1)*(V15-V1))+((V16-V1)* \\ & (V16-V1))+((V17-V1)*(V17-V1))+((V18-V1)*(V18-V1))+((V19-V1)*(V19-V1))+ \\ & ((V20-V1)*(V20-V1))+((V21-V1)*(V21-V1))+((V22-V1)*(V22-V1))+((V23-V1)* \\ & (V23-V1))+((V24-V1)*(V24-V1))+((V25-V1)*(V25-V1))+((V26-V1)*(V26-V1))+ \\ & ((V27-V1)*(V27-V1))+((V28-V1)*(V28-V1))+((V29-V1)*(V29-V1))+((V30-V1)* \\ & (V30-V1))+((V31-V1)*(V31-V1))))*V1), \text{ where} \\ & V1 \text{ is the standard weight for persons (WEIGHTP), and} \\ & V2-V31 \text{ being the 30 replicate weights (REPWT1-30)} \end{aligned}$$

OVERVIEW *continued*

SAMPLING ERROR *continued*

As well as enabling variances of estimates to be calculated relatively simply, replicate weights also enable unit record analyses such as chi-square and logistic regression to be conducted which take into account the sample design.

Further information about the use of replicate weights is contained in the *2001 National Health Survey: Users' Guide* (cat. no 4363.0). It should be noted that not all statistical computer packages may allow direct calculation of SEs using replicate weights. However, those packages that allow the direct use of Balanced Repeated Replication (BRR) methodology generally include the option of an adjustment factor. This factor can be incorporated to overcome the difference between the variance formulae.

ABOUT THE DATA

Detailed information about the data collected, comments regarding data quality and other points to assist in interpreting the data are contained in the *2001 National Health Survey: Users' Guide*. It is recommended that relevant parts of the Guide be read in conjunction with use of this CURF.

In addition, there are some general points regarding the data appearing on this CURF which should be noted.

- Adjustments have been made to the data in a small number of records in the course of confidentialising records included on the CURFs. Some adjustments were made to the BASIC file only, while others were made to EXPANDED file or to both files. These changes include:
 - changing income groupings
 - swapping weights between some records
 - changing height and/or weight values
 - removing a highly visible record from the file.
- In addition to these changes, some amendments have been made to the main NHS data file since the release of the summary publication, and some further changes between the main data file and the CURF have arisen through the systems in use to produce different types of output e.g. in the way some data are rounded.
- While each of these changes has been minor, collectively they mean that:
 - estimates produced from the CURFs may differ by a small amount from those published in *2001 National Health Survey; Summary of Results, Australia* (cat. no. 4364.0) or released in tables available electronically through the ABS web site
 - there will be small differences in estimates for some variables and populations produced from the BASIC and EXPANDED files.

The nature of the changes made, and the relatively small number of records involved have ensured that the effects on the data for analysis purposes is considered negligible.

- Some data items on the CURFs directly reflect responses to individual questions contained in the survey questionnaire, while others have been derived from responses to two or more questions. Due to the volume and complexity of the derivations, these derivations are not generally released. However, details may be available on request.

OVERVIEW *continued*

ABOUT THE DATA *continued*

- In previous NHSs, all persons in selected dwellings were enumerated, enabling users of the CURF to group people to form complete households, families and income units, and to analyse data in respect of those units. In the 2001 NHS, only single person households and some single adult households with children were fully enumerated. Linkages are provided on the CURFs enabling child records from a household to be analysed together and with the adult record from that household. However, as these units will be incomplete in most cases, it is recommended that analysis of the 2001 file be done at the person level only. Some household, family and income unit characteristics are held as person level characteristics on person records enabling analysis using these characteristics.

ABOUT THE FILES

The CURFs are available in SAS and SPSS. If you obtain a copy of the BASIC file on CD-ROM and your analysis software is other than SAS or SPSS you may require the services of a computer programmer to use the ASCII file version of the data. The RADL files can only be accessed in SAS and SPSS at this time. Additional software may be introduced in the future.

The CD-ROM contains data files, SAS and SPSS user files, and a set of information files. The RADL contains the SAS or SPSS dataset; text files, SAS and SPSS sample programs and a copy of this information paper will be available online through RADL. Details of the files are shown in Appendix 4.

CONDITIONS OF RELEASE

The NHS CURF is released in accordance with a Ministerial Determination (Clause 7, *Statutory Rules 1983, No. 19*) in pursuance of section 13 of the *Census and Statistics Act 1905*. As required by the Determination, the CURF has been designed so that the information on the file is not likely to enable the identification of the particular person or organisation to which it relates.

All organisations and individuals within organisations who request access to the CURF will, prior to being granted access to the CURF, be required to sign an Undertaking to abide by the legislative restrictions on use. The Australian Statistician's approval is required for each release of the CURF. Persons who purchase or access the NHS CURF are required to give an undertaking which includes, among other conditions, that in using the data they will:

- use the information only for statistical purposes;
- not attempt to identify particular persons or organisations
- not disclose, either directly or indirectly, the information to any other person or organisations
- not attempt to match the information with any other unit level list of persons.

Use of the data for statistical purposes means use, by persons who have signed the Undertaking, to produce information of a statistical nature. Examples of statistical purposes are:

- manipulation of the data to produce means, correlations or other descriptive or summary measures
- estimation of population characteristics
- use of data as input to mathematical models or for other types of analysis (e.g. factor analysis)

OVERVIEW *continued*

CONDITIONS OF RELEASE *continued*

- providing graphical or pictorial representations of the characteristics of the population or subsets of the population.

All CURF users are required to read and abide by the 'Responsible Access to ABS Confidentialised Unit Record Files (CURFs) Training Manual' available on the ABS web site, Access to ABS CURFs. Use of the data for unauthorised purposes may render the purchaser liable to severe penalties. Advice about the propriety of any particular intended use of the data is available from the contact officer.

CONDITIONS OF SALE

All ABS products and services are provided under conditions of sale. Any queries relating to these Conditions of Sale should be referred to <intermediary.management@abs.gov.au>.

PRICE

The price of the BASIC and EXPANDED file from the 2001 NHS is \$8,000 each, including freight and handling charges where appropriate. Access to both files can be provided at a cost of \$8,000 plus an administrative fee of \$420.

While the utmost care is taken in handling each CURF on CD-ROM, deterioration may occur between the time of copying and receipt of the file. Accordingly, if the CD-ROM is unreadable on receipt and this is reported to ABS within 30 days of receipt, it will be replaced free of charge.

HOW TO ORDER

All clients wishing to access the NHS CURFs should refer to the ABS web site, Access to ABS CURFs and read the Responsible Access to ABS Confidentialised Unit Record Files (CURFs) Training Manual, and other relevant information, before downloading the Application and Undertaking to apply for access.

Australian universities

University customers should refer to the ABS web site <www.abs.gov.au> (under 'Services to Universities'). The NHS CURF can be accessed by universities participating in the ABS/AVCC CURF agreement for research and teaching purposes.

Overseas customers

The CURF is not available to overseas customers.

FOR MORE INFORMATION . . .

- INTERNET* **www.abs.gov.au** the ABS web site is the best place to start for access to summary data from our latest publications, information about the ABS, advice about upcoming releases, our catalogue, and Australia Now—a statistical profile.
- LIBRARY* A range of ABS publications is available from public and tertiary libraries Australia-wide. Contact your nearest library to determine whether it has the ABS statistics you require, or visit our web site for a list of libraries.
- CPI INFOLINE* For current and historical Consumer Price Index data, call 1902 981 074 (call cost 77c per minute).
- DIAL-A-STATISTIC* For the latest figures for National Accounts, Balance of Payments, Labour Force, Average Weekly Earnings, Estimated Resident Population and the Consumer Price Index call 1900 986 400 (call cost 77c per minute).

INFORMATION SERVICE

Data already published that can be provided within five minutes will be free of charge. Our information consultants can also help you to access the full range of ABS information—ABS user pays services can be tailored to your needs, time frame and budget. Publications may be purchased. Specialists are on hand to help you with analytical or methodological advice.

- PHONE* 1300 135 070
- EMAIL* client.services@abs.gov.au
- FAX* 1300 135 211
- POST* Client Services, ABS, GPO Box 796, Sydney NSW 2001

WHY NOT SUBSCRIBE?

ABS subscription services provide regular, convenient and prompt deliveries of ABS publications and products as they are released. Email delivery of monthly and quarterly publications is available.

- PHONE* 1300 366 323
- EMAIL* subscriptions@abs.gov.au
- FAX* 03 9615 7848
- POST* Subscription Services, ABS, GPO Box 2796Y, Melbourne Vic 3001



2432400001017
ISSN 0817 9344

RRP \$10.00

