NSW Population Health Survey: Description of Methods

Margo Barr, Deborah Baker, Matthew Gorringe, and Lindy Fritsche
Centre for Epidemiology and Research
NSW Department of Health

Introduction

The New South Wales Population Health Survey is a continuous survey of the health of people of New South Wales using computer assisted telephone interviewing (CATI). The main aims of the survey are to provide detailed information on the health of the people of New South Wales, and to support the planning, implementation, and evaluation of statewide health services and programs.

Prior to the introduction of the continuous survey in 2002, the Centre for Epidemiology and Research conducted adult health surveys in 1997 and 1998, an older people's health survey in 1999, and a child health survey in 2001.

Since 2002, the continuous survey has produced: an annual report on adult health for the whole state, a monthly report on adult health for the whole state, an annual report on adult health for each health area, biennial reports on child health for the whole state, triennial reports on adult health for the divisions of general practice, and occasional reports on the health of specific populations including.

This brief paper describes the methods used in the survey.

New South Wales Population Health Survey

Survey instrument

The survey instrument was developed by the NSW Health Survey Program in consultation with key stakeholders, the state's area health services, other government departments, and a range of experts. In any given year, the survey instrument includes: questions used in previous surveys, new questions developed specifically for that year, and questions developed specifically for some of the area health services. All questions not previously used are submitted to a lead ethics committee for approval prior to use. New questions are field tested prior to inclusion in the survey. The survey instrument is translated into 5 languages: Arabic, Chinese, Greek, Italian and Vietnamese.

Field testing

The survey conducts a rigorous two-stage process of field testing new questions prior to inclusion in the survey. The first phase involves a review of the proposed questions by epidemiologists and experienced interviewers for comprehensibility, cohesiveness and readability. Problems with interpretation and terminology are identified and category refinements recommended as required.

The second phase addresses reliability and convergent validity issues using a test–retest protocol. Reliability assesses the ability of a question to provide consistent measures. Convergent validity assesses whether prevalence differs when alternate forms of a question are used. When the questions are tested for test–retest reliability the same question is programmed for the initial questionnaire as well as the repeat questionnaire. When the questions are tested for

convergent validity then either of the questions are programmed to be randomly presented in the initial questionnaire and the other question is presented in the repeat questionnaire, thus allowing for the crossover design.

The target sample for field testing survey questions is 200 persons. This sample size ensures that a kappa of 0.6 (good or excellent) is able to be detected at a significance level of 5% and a power of 80% when compared to a kappa of 0.4 or less (fair or poor) for response frequencies greater than 20%.[7,8]

Test–retest reliability is estimated by Cohen's kappa statistic for binary and nominal variables, and weighted kappa with Cicchetti-Allison weights for ordinal variables.[1] Unbalanced tables are corrected using the method described by Crewson.[2] Since erroneously low values of kappa can arise from skewed data,[3–7] per cent agreement is also presented for categorical variables, calculated as the proportion of respondents in the same category at test and retest. Test–retest reliability for continuous data is assessed using spearman rank correlation coefficients (r_s) as a non-parametric analog of the intraclass correlation coefficient (icc).[9,10]

Survey sample

The target population for the survey is all state residents living in households with private telephones. The target sample is approximately 1,500 persons a year in each of the state's 8 area health services (a total sample of 12,000 persons a year). This sample ensures that a difference of + or – 1% can be detected for New South Wales and + or – 3% can be detected for the area health services.

The sampling frame is developed as follows. Records from the Australia on Disk electronic white pages (phone book) are geo-coded using MapInfo mapping software.[11,12] The geo-coded telephone numbers are assigned to statistical local areas and area health services. The proportion of numbers for each telephone prefix is calculated by area health service. All prefixes are expanded with suffixes ranging from 0000 to 9999. The resulting list is then matched back to the electronic white pages. All numbers that match numbers in the electronic white pages are flagged and the number is assigned to the relevant geo-coded area health service. Unlisted numbers are assigned to the area health service containing the greatest proportion of numbers with that prefix. Numbers are then filtered to eliminate contiguous unused blocks of greater than 10 numbers. The remaining numbers are then checked against the business numbers in the electronic white pages to eliminate business numbers. Finally, numbers are randomly sorted.

Households are contacted using list assisted random digit dialling. One person from the household is randomly selected for inclusion in the survey, using age order, having firstly identified the position of the household reporter. If the selected respondent is a child under the age of 16 years, a parent or carer is selected as a proxy respondent.

Interviews

Interviews are carried out continuously between February and December. Selected households with addresses in the electronic white pages are sent a letter describing the aims and methods of the survey 2 weeks prior to initial attempts at telephone contact. An 1800 freecall contact number is provided for potential respondents to verify the authenticity of the survey and to ask any questions regarding the survey.

Trained interviewers from the in-house New South Wales Health Survey Program CATI facility carry out interviews. Up to 7 calls are made to establish initial contact with a household, and 5 calls are made in order to contact a selected respondent.

Response rates

The overall response rate is calculated as completed interviews divided by completed interviews and refusals.

Data analysis

For analysis, the survey sample is weighted to adjust for differences in the probabilities of selection among subjects. These differences are due to the varying number of people living in each household, the number of residential telephone connections for the household, and the varying sampling fraction in each health area.

Post-stratification weights are used to reduce the effect of differing non-response rates among males and females and different age groups on the survey estimates. These weights are adjusted for differences between the age and sex structure of the survey sample and the Australian Bureau of Statistics mid-year population estimates for each area health service (excluding residents of institutions). This enables calculation of prevalence estimates for the state population rather than for the respondents selected.

Call and interview data are manipulated and analysed using SAS version 9.[13] The SURVEYMEANS procedure in SAS are used to calculate point estimates and 95 per cent confidence intervals for the estimates by dichotomizing the responses. The SURVEYMEANS procedure calculates standard errors adjusted for the design effect factor or DEFF (the variance for a non-random sample divided by the variance for a simple random sample). It uses the Taylor expansion method to estimate sampling errors of estimators based on the stratified random sample.[13]

The 95 per cent confidence interval provides a range of values that should contain the actual value 95 per cent of the time. A wider confidence interval reflects less certainty in the estimate. If confidence intervals do not overlap then the observed estimates are significantly different. If confidence intervals overlap slightly the observed estimates may be significantly different but further testing needs to be done to establish that significance.

For a pairwise comparison of subgroup estimates, the p-value for a two-tailed test is calculated using the normal distribution probability function PROBNORM in SAS, assuming approximate normal distribution of each individual subgroup estimates with the estimated standard errors, and approximate normal distribution for the estimated difference.

For predicted prevalence estimates, values are calculated using the FORECAST procedure in SAS version 9. The underlying model used in this procedure is the Holt exponential smoothing model. This model is designed to use all of the observed annual prevalence estimates and takes into account the increasing (or decreasing) trend in the prevalence estimates over time. In this way, the model uses past data as a basis for estimating annual prevalence estimates into the future. In this forecasting model later annual prevalence estimates are given more importance than earlier prevalence estimates.

References

1. Fleiss JL. *Statistical methods for rates and proportions*. 2nd edition. New York: John Wiley, 1981.
2. Crewson PE. *A correction for unbalanced kappa tables. SAS 6.12*. Paper 194-26. Available online at www2.sas.com/proceedings/sugi26/p194-26.pdf.
3. Feinstein AR and Cichchetti DV. High agreement but low kappa I. The problem of two paradoxes. *J Clin Epidemiol* 1990 43: 543–549.
4. Feinstein AR and Cichchetti DV. High agreement but low kappa II. Resolving the paradoxes. *J Clin Epidemiol* 1990; 43: 551–558.
5. Byrt T, Bishop J, and Carlin B. Bias, prevalence and kappa. *J Clin Epidemiol* 1993; 46(5): 423–429.
6. Kraemer HC and Korner AF. Statistical alternatives in assessing test–retest reliability, consistency and individual differences for quantitative measures: Application to Behavioral Measures of Neonates. *Psychol Bull* 1976; 83(5): 914–921.
7. Flack VF, Afifi AA, Lachenbruch PA, Schouten HJA. Sample size determinations for the two rater kappa statistic. *Psychometrika* 1988; 53: 321–325.
8. Sim J, Wright CC. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy* 2005; 85: 257–68.
9. Bland MJ and Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8: 135–160.
10. Kraemer HC and Korner AF. Statistical alrternatives in assessing test–retest reliability, consistency, and individual differences for quantitative measures: Application to Behavioural Measures of Neonates. *Psychol Bull* 1976; 83(5): 914–921.
11. Australia on Disk [software]. Sydney: Australia on Disk, 2000.
12. MapInfo [software]. Troy, NY: MapInfo Corporation, 1997.
13. SAS Institute. The SAS System for Windows version 9 [software]. Cary, NC: SAS Institute Inc., 2002.