Meta Prevalence Estimates

Generating combined prevalence estimates from separate population surveys

Baohui Yang PhD
Biostatistician
Centre for Epidemiology and Research
NSW Department of Health

Background

Since 1998, Australian states and territories have made substantial investments in the population surveillance of chronic diseases and associated risk factors through the development and implementation of state based computer assisted telephone interview (CATI) systems.

Although these systems were not designed to generate national prevalence estimates, there is great interest in exploring the potential of these systems to be able to do so through data pooling. The reasons why this potential has not been realised are concerns about: technical issues, access and confidentiality, differences in jurisdictional data collection practice, and increased data management.

Aim

As an alternative to pooling unit record data, this report describes a method of combining separate population survey prevalence estimates to generate an overall combined estimate (or meta prevalence estimate).

Method to calculate a meta prevalence estimate

Calculating a meta prevalence estimate from separate survey estimates can be undertaken, taking into account the prevalence estimates, the standard errors, and the population size for each separate population survey, using the following formula.

If the estimate for the ith stratum is p_i, standard error s_i, and the population size in the ith stratum is N_i, $i = 1, 2, 3, ..., S$. Then the meta prevalence estimate would be

$$p = \frac{\sum_{i=1}^{S} N_i p_i}{\sum_{i=1}^{S} N_i},$$

and the standard error estimate would be,

$$Stde(p) = \sqrt{\frac{\sum_{i=1}^{S} N_i^2 s_i^2}{\left(\sum_{i=1}^{S} N_i\right)^2}}.$$

The above formula would also apply to any subgroup, such as sex, age group, and geographical area, as long as the required information on prevalence estimates, standard errors, and corresponding subpopulation sizes from each population survey, are available.

A meta prevalence estimate and a pooled estimate: an example

As a demonstration, we used 8 samples of the area health services in New South Wales, as separate population health surveys, to calculate a meta prevalence estimate for males and females for the state in 2003.

In each of the 8 surveys, list-assisted random digit dialling was used to select a household, then an eligible member of the selected household is randomly selected. In 2003, the adult sample size was approximately 1000 for each of the 8 area health services. The health indicators we included in the example were: current daily or occasional smokers, recommended vegetable and fruit consumption, any alcohol drinking, adequate physical activity, and high psychological distress.

Calculation of a meta prevalence estimate

First, each of the indicator prevalence estimates for males and females in the health areas were calculated as follows:

- The sample for a health area was weighted to adjust for differences in the probabilities of selection among subjects and then adjusted for differences between the age and sex structure of the survey sample and the Australian Bureau of Statistics 2001 mid-year health area population estimates.

- Point estimates and standard errors for the prevalence of a health indicator were calculated using SURVEYMEANS procedure in SAS 8.2. The SURVEYMEANS procedure calculates standard errors adjusted for the complex survey design and weighting. It uses the Taylor expansion method to estimate sampling errors of estimates.

We then calculated the meta prevalence estimate and the standard error estimate for each of the indicators for males and females in NSW using the above formula (Table 1).

Calculation of prevalence estimates from pooled unit record data

We pooled unit record data from the 8 surveys in 2003. Each of the indicator pooled prevalence estimates for males and females in NSW were calculated in a similar way as above:

- The pooled survey sample was weighted to adjust for differences in the probabilities of selection among subjects and then adjusted for differences between the age, sex and health area structure of the pooled survey sample and the Australian Bureau of Statistics 2001 mid-year population estimates.

- Point estimates and standard errors for pooled indicator estimates were calculated using SURVEYMEANS procedure in SAS 8.2, adjusted for the complex survey design and weighting, with the 8 area health services as different strata.

Table 1. Comparison of a meta prevalence estimates with a pooled prevalence estimate for males and females in New South Wales in 2003

Indicator	meta prevalence estimate %	pooled prevalence estimate %
Males		
Current smoking	24.74 (SE 0.8732)	24.74 (SE 0.8733)
Vegetables	11.91 (SE 0.5840)	11.91 (SE 0.5845)
Fruit	38.37 (SE 0.9424)	38.37 (SE 0.9478)
Alcohol	40.71 (SE 0.9600)	40.71 (SE 0.8733)
Physical activity	49.87 (SE 0.9854)	49.87 (SE 0.9838)
High psychological stress	9.22 (SE 0.5382)	9.22 (SE 0.5392)
Females		
Current smoking	19.78 (SE 0.6487)	19.78 (SE 0.6491)
Vegetables	26.23 (SE 0.6987)	26.23 (SE 0.6992)
Fruit	51.80 (SE 0.8122)	51.80 (SE 0.8139)
Alcohol	30.26 (SE 0.7443)	30.26 (SE 0.7457)
Physical activity	40.63 (SE 0.8039)	40.63 (SE 0.8080)
High psychological stress	12.82 (SE 0.5500)	12.82 (SE 0.5511)

Comparison of the meta prevalence estimates and the pooled prevalence estimates

The meta prevalence estimates and the pooled prevalence estimates are the same and the standard error estimates are the same to the second decimal place (Table 1). The reason for the standard error difference past the second decimal place is due to rounding.

In general, if each individual survey is weighted to adjust its sample to its population structure, and the weight sums to its population size, then the meta estimate calculated using the formula would be the same as the pooled estimate.

Discussion

Calculation of meta prevalence estimates using the formula provided is simple, when the prevalence estimates, standard error estimates, and subpopulation size from each separate population survey are available. The quality of the meta prevalence estimates relies on the quality of the individual population survey prevalence estimates, with the bias of any meta prevalence estimate being the weighted mean of the biases for the individual population survey prevalence estimates. If the individual population survey prevalence estimates were unbiased then the meta prevalence estimate would also be unbiased.

In contrast, pooling the individual population survey unit record data requires substantial data management both at the individual population survey level as well as at the level where the pooled data are managed. When the data are pooled technical issues such as differences in stratification and post stratification weighting schemes need to be addressed. Confidentiality of the pooled unit record data also needs to be addressed including compliance to legislations and policies from different jurisdictions. Processes for third party access will also need to be decided upon taking into consideration the legislations and policies from different jurisdictions.

Most importantly, pooling individual population survey unit record data will not provide a better quality prevalence estimate than the meta prevalence estimate.