

University of Wollongong

Research Online

Faculty of Health and Behavioural Sciences -
Papers (Archive)

Faculty of Science, Medicine and Health

1-1-2008

Assembling a nutrient database for a large cohort study: Blue Mountains Eye Study

Victoria M. Flood

University of Wollongong, vflood@uow.edu.au

W Smith

Centre for Clinical Epidemiology, Uni of Newcastle

E Rochtchina

Centre for Vision Research, Uni of Sydney

Jie J. Wang

Centre for Vision Research, Uni of Sydney

Paul Mitchell

Centre for Vision Research, Uni of Sydney

Follow this and additional works at: <https://ro.uow.edu.au/hbspapers>



Part of the [Arts and Humanities Commons](#), [Life Sciences Commons](#), [Medicine and Health Sciences Commons](#), and the [Social and Behavioral Sciences Commons](#)

Recommended Citation

Flood, Victoria M.; Smith, W; Rochtchina, E; Wang, Jie J.; and Mitchell, Paul: Assembling a nutrient database for a large cohort study: Blue Mountains Eye Study 2008, 37-40.
<https://ro.uow.edu.au/hbspapers/344>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Assembling a nutrient database for a large cohort study: Blue Mountains Eye Study

Abstract

To construct a nutrient database for use with a large population-based cohort study, the Blue Mountains Eye Study, dietary intakes were estimated using the Australian electronic food composition nutrient databases to which values for additional nutrients were added, including folate, vitamin B12, carotenoids and fatty acids. The addition of nutrient data from alternative, overseas, database sources may be useful in relation to the assessment of outcome measures, however the data obtained from these sources need to be interpreted with care, especially in relation to absolute quantities of intake.

Keywords

eye, assembling, nutrient, database, study, blue, cohort, large, mountains

Disciplines

Arts and Humanities | Life Sciences | Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details

Flood, V. M., Smith, W., Rochtchina, E., Wang, J. J. & Mitchell, P. 2008, 'Assembling a nutrient database for a large cohort study: Blue Mountains Eye Study', *Food Australia*, vol. 60, no. 1-2, pp. 37-40.

Assembling a nutrient database for a large cohort study: the Blue Mountains Eye Study

V. M. Flood, W. Smith, E. Rochtchina, J. J. Wang and P. Mitchell

To construct a nutrient database for use with a large population-based cohort study, the Blue Mountains Eye Study, dietary intakes were estimated using the Australian food composition electronic nutrient databases to which values for additional nutrients were added, including folate, vitamin B12, carotenoids, and fatty acids. The addition of nutrient data from alternative, overseas, database sources may be useful in relation to the assessment of outcome measures, however the data obtained from these sources needs to be interpreted with care, especially in relation to absolute quantities of intake.

Development of a nutrient database for a population cohort study is fraught with difficulties, given that the nutrient data is collected over a long period of time and both the food supply and laboratory analysis methods may change during that period. In addition, research questions of interest may change over time, and may therefore require additional information of particular nutrients that were not available or not anticipated in the original baseline data collection period. There are also practical considerations in relation to storage of data, construction of the nutrient database, and software used for the analyses of the dietary assessment tool. The database needs to be flexible enough to meet the current and the future needs of researchers involved with a longitudinal study.

This paper reflects personal experiences and practical considerations in relation to these issues, based on involvement with a population-based cohort study over ten years.

The study population

The Blue Mountains Eye Study (BMES) is a population-based cohort study of eye diseases and other chronic health outcomes in residents aged 49 years and over of a defined area, west of Sydney, Australia. Participants were first examined during the period January 1992 to January 1994. A door knock census of 38 ABS census collection districts (CCD) in two adjoining postcode areas (2780 and 2782) was conducted. Of 4433 eligible residents, 3654 aged 49 to 97 years, attended an eye examination (response rate 82.4%). Prior to attending the eye examination, participants were sent a 145 item semi-quantitative Food Frequency Questionnaire (FFQ). A total of 3267 participants attempted and returned the FFQ, of which 2897 were usable (79.3% of the participants examined; 88.6% of those attempted to complete the FFQ). Non-participants were of a similar age and gender distribution, but had a lower self-reported prevalence of spectacle wearing and hypertension (Smith 1997).

The second BMES study followed up participants from the original cohort, five years later, from 1997–1999. From the original 3654 people who participated at baseline, 2334 attended eye examinations for the second data collection period, representing 75% of survivors ($n = 3111$); 543 (14.9%) people had died, 383 (10.5%) had moved and 394 (10.8%) refused to participate. Of these, 2108 attempted the food frequency questionnaire, of which 2005 were usable (85.9% of those examined).

The third BMES study followed up participants from the baseline cohort, ten years after the original data collection, in 2002–2004. From the 3654 people who participated at baseline, 1952 attended interviews for the third data collection period, representing 75.6% of survivors. Of those who attended interviews, 1600 attempted the food frequency questionnaire, of which 1532 were usable (78.5% of those examined).

The Food Frequency Questionnaire

Dietary data were collected using a 145 item self-administered FFQ, modified for Australian diet and vernacular from an early FFQ from Willett & others (1988), and included portion size estimates, as well as frequency, strength, brand and type of supplements. Participants used a 9-category frequency scale to indicate the usual frequency of consuming individual food items during the past year. Participants were asked to include brand and type of certain foods, in particular breakfast cereals, margarine and oil. There were no specific questions about ripeness of fruit consumed, which can influence the nutrient content of most fruits. However, an allowance for seasonal variation of fruit and vegetables was made during analyses by weighting seasonal fruits and vegetables.

The FFQ was validated against three separate 4-day weighed food records, collected over the period of one year to account for seasonal variation, in a randomly selected sub-sample of the BMES cohort ($n = 79$). Validation of the FFQ has been described in peer-reviewed papers (Smith & others 1998, Flood & others 2004).

Data entry of the FFQ

The data from each FFQ was entered into a purpose built form in Microsoft Access. This form creates a spreadsheet of data with subjects ID in rows and FFQ items (coded with an item number) in the columns. A nine category frequency scale for each food is coded with numbers 1 (never) to 9 (4+/day) for each individual. Specific food types, such as margarine and breakfast cereals are given a precise code which is used in the corresponding nutrient database.

Data cleaning of FFQ

Participants with more than 25 missing values (17%) in their FFQ were excluded from the final data set. Participants with 13–25 missing values in the FFQ were checked for data entry errors, which were corrected where possible. If, after data checking, more than 12 FFQ questions remained blank or an entire page remained blank, these FFQs were also excluded from the final data set. Participants with FFQ daily energy intakes of less than 2500 kJ or greater than 18 000 kJ were excluded from the final data set. Finally, nutrient data were screened for extreme values, inspecting values in

the upper and lower 2% of the distribution to locate and correct any data entry errors and to check for plausibility. This process has been repeated for each data collection period.

Dietary analysis software

Dietary intakes in this study were estimated using a purpose-built software package, “the FFQ Masher” (Lazarus 1999), based on a Dbase software system. In the initial testing of this system, the nutrient values for a sub-sample of completed FFQ generated using the FFQ Masher were compared to values generated from a commercially available nutrient analysis package, Diet/1 (Xyris 1988), using an FFQ template created in that package. Any errors identified were corrected and the purpose built system has been used for the remaining analyses.

Compiling the nutrient database

The compilation of nutrient databases for epidemiological research poses many challenges, in particular, ensuring the use of reliable and valid data (Baghurst & Baghurst 1990).

The FFQ Masher incorporated the electronic form of the nutrient database for use in Australia (Nuttab90 in BMES1 and Nuttab95 in BMES2 and BMES3) (Department of Community Services and Health, 1990, 1995) to which additional nutrient values for the 145 food items were added. The Nuttab values represent analytical data of Australian food products. However, there were several nutrients not included in that database for which we sought further information and so we explored additional nutrient databases. Our preference was to use Australian analytical data, and where these were not available to use analytical data from other sources, with a focus to use national nutrient databases (ie borrowed data). However, some data used were based on a mixture of analysed, calculated, imputed or borrowed data.

For example, folate values for foods were added from the 1995 Australian National Nutrition Survey database, AUSNUT (ANZFA, 1999); data on folate values of fortified products available in Australia from 1997–1999 and then 2002–2004 were also included (Abraham & Webb, 2001, Lawrence & others 1999; FSANZ 1999, 2004). Since fortification of foods was progressively changing during the period of data collection, a series of seven different time periods was identified to categorise the folate values of fortified foods on the market. This information was a combination of data obtained from FSANZ and manufacturers. Each individual’s intake was then matched to the corresponding time period in which his or her FFQ was collected. Folate values for foods for BMES1 reflect pre-fortified values; most of the pre-fortified folate values used in Australia at that time were derived from UK (Holland & others 1991) data and did not reflect analytical data of Australian foods. The McCance and Widdowson tables from the UK use a single enzyme microbiological assay to determine folate, but the preferred method now receiving support in Australia is the tri-enzyme analytical technique which generally yields higher folate values than the single enzyme method (Arcot & Shreskha 2006). Additionally, it is probable that there are real differences between products analysed in the UK to those in Australia

(Leemhuis & others 2006).

Separating the data by time period allowed the analyses of pre-fortified foods, so calculation of dietary folate equivalents was possible. An allowance for the increased bioavailability of folate from supplements and fortified food products was calculated, expressed as Dietary Folate Equivalents (DFE) in accordance with US recommendations as follows: $1 \mu\text{g DFE} = \{\mu\text{g food folate} + (1.7 \times \mu\text{g synthetic folic acid})\}$ (IOM 1998). The difference between the pre-fortified folate estimate and the fortified folate estimate was taken, and supplement data was added to this, and then multiplied by 1.7. The pre-fortified value was then added back to this new figure, providing an estimate of DFE from diet and supplements.

Vitamin B12 nutrient estimates were obtained from the UK Composition of Foods (Holland & others 1991), and were then added to the nutrient database used in the FFQ Masher. There are likely to be differences between the composition of UK foods and the true content of Australian foods; the differences could result from variation of food varieties, climate, growing conditions, food manufacturing techniques and preparation (Scheelings 1996). The UK Composition of Foods includes foods stored in varying conditions (Holland & others 1991). Acid, alkali, light and oxidising agents can destroy vitamin B12, and approximately 30% of its activity is lost during cooking (Krause & Mahan 1984). However, in the absence of any other database, the UK tables provided an estimate of food composition in Australia.

The major health outcome of interest in this study has been eye disease and subsequently there has been a particular interest to estimate the intakes of various carotenoids (Manzi & others 2002, Flood & others 2002, Tan & others 2007). Where we had Australian carotenoid data in our national electronic database, we preferentially used this, because it was considered more comprehensive and accurate for Australian foods. However, after this we used the United States Department of Agriculture Carotenoid (USDA 1998) database to estimate other carotenoids, including α -carotene, β -cryptoxanthin, combined lutein and zeaxanthin, and lycopene. It should be noted that the 1998 USDA Carotenoid database included the combined lutein and zeaxanthin estimates. Later databases have separated these carotenoids using High Performance Liquid Chromatography (HPLC) laboratory analysis.

Food items were matched by food descriptor and, where available, information about macronutrient intake of a food. However, not all the items on our FFQ could be exactly matched with one of the items in the USDA carotenoid database. For example, our questionnaire uses cooked (boiled) or baked (with added oil) pumpkin, and boiled silverbeet, and the closest matches on the USDA database were canned pumpkin, and boiled spinach, described as ‘Spinach, cooked, boiled, drained without salt’. The database does not list English spinach, simply spinach. In these examples the β -carotene content of the food items from the two databases were very different (for example spinach (US carotenoid database 5242 $\mu\text{g}/100 \text{g}$), silverbeet, boiled (Australian Nuttab database 410 $\mu\text{g}/100 \text{g}$); ‘pumpkin, canned without salt’ (USDA Carotenoid database 6940 $\mu\text{g}/100 \text{g}$), ‘pumpkin, unspecified type, peeled, boiled (Australian

Nuttab database 2676 µg/100 g)). However, as previously indicated, we used the Australian β-carotene content of the food item. Many other items were closer in values to the Australian database; these are the most extreme examples where differences occurred. Differences such as these in food preparation, food varieties and possible food fortifications (from colouring, for example) are likely to limit the accuracy of our estimates for the Australian food supply and our population (Manzi & others 2002). A useful addition to the database would have been to search for other published analytical carotenoid data, but this was not done at that time, as it was considered valuable to use the relatively comprehensive USDA database. In addition, the food preparation methods of foods may influence the biological availability of carotenoids, in particular for fruit and vegetables. The FFQ described fruit as 'fresh' and in these cases fruit intake were assumed to be of raw fruit. Some particular fruits in the FFQ description, specified dried or canned fruits, which were then assessed appropriately from the nutrient database. The majority of vegetables were assumed to be cooked and generally the nutrient database option was 'boiled'. However, there were a few exceptions: for vegetables commonly consumed as raw items, such as lettuce, celery, mushrooms and shallots 'raw' nutrient data was used. An exception was carrot which was analysed as 50% boiled and 50% raw of unspecified carrot type.

The original fatty acid profile of nutrients in our database was limited to saturated fats, monounsaturated fats, and polyunsaturated fats as provided in the NUTTAB databases. Later, we added data from the NUTTAB fatty acid supplement database, but found that database limiting as it had too few decimal places for fatty acids found in very small amounts in foods, such as the long-chain omega-3 fatty acids. Subsequently, we used an alternative source of fatty acid data from the Royal Melbourne Institute of Technology (Mann & others 2002), which provided data in gram units to two decimal places, and included an estimate for trans fatty acids. While data described to two decimal places may have some limitations from the point of view of the biological variability of food sources, it has been useful when determining the estimated nutrient intake of some fatty acids which are available in only small but important amounts in foods, such as long-chain omega-3 fatty acids. This new data was added to appropriate foods in the database by matching other macronutrient content, in order to provide a close match for the food item, since it could not be matched by ID code. However, it is acknowledged that there are likely to be limitations to the accuracy of the data using this system.

Dietary supplement intake

Dietary supplement intake was assessed by questionnaire, which sought details about whether a supplement was used, and if so, the frequency, brand and strength. Supplement data were coded using an updated supplement database, previously developed for use in BMES 1 (Ashton & others 1997), which was then updated for each data collection period. Each nutrient's supplement data for each participant reflects the addition of all supplements used by that participant.

Interpretation of the data

Previous literature has documented the difficulties with using nutrient data from FFQ to provide estimates of absolute intake, but generally they can be used to rank order quintiles of intakes, and our own validity analyses have also confirmed this to be the case (Smith & others 1998, Flood & others 2004). In addition, using different nutrient database sources from other countries and different methods of laboratory analysis provides additional potential for measurement error. However, if data are analysed by ranking individuals into groups and then investigating these in relation to a health outcome, then the effect of this potential measurement error is likely to be reduced. An additional technique is not only to consider nutrients and their relationship to a health outcome but also to investigate related foods to a health outcome of interest. We have been able to successfully do this on a number of occasions and it has been a useful way of investigating associations of diet with disease. For example, in a paper investigating fatty acid intake and macular degeneration, we investigated not only fatty acid types but also major foods contributing to those fatty acid types (Chua & others 2006). In this case, both forms of analysis supported each other; we found those who had the highest quintile of omega-3 fatty acid intake versus the lowest quintile had a 60% reduced risk of developing early macular degeneration after five years (OR 0.41, 95% CI 0.22–0.75). In the same research we found those who consumed fish once a week had a 40% reduced risk of developing early AMD (OR 0.58, 95% CI 0.37–0.9).

Lessons learned

If given an opportunity to re-create our nutrient database for this cohort, there are several aspects of the database we would consider doing differently:

1. Create the database using a software platform which was more likely to allow easier modifications into the future. The initial data entry was conducted using Access forms. In hindsight, it would have been easier to create the entire nutrient analyses using that database, rather than switching to Dbase and using the Borland Database Engine, which has become difficult to support over time. While we are still able to use this software it is not easy to make additions to the system.
2. Explore other published sources of nutrient data, particularly from laboratory analysed data of Australian food sources. We tended towards national nutrient databases which is likely to have limited our use of some valuable data.
3. Document the details and exact source of all added nutrients for each individual food item. As we progressed with additions we learnt to specifically document all sources of information and any known details. Some earlier versions of additions were only described in general terms.

Conclusions

Over time we have created a comprehensive food and nutrient dataset to examine health outcomes over the long period of data collection. The use of a purpose-built software program to analyse our dietary data has provided some flexibility with including additional nutrients to the dataset. We have taken care about the

approach to consider additions to reflect changes to the food supply, such as fortification, and to consider how we can use nutrients from other nutrient datasets. It is important to remember the limitations of the dataset, generally using the data in a ranked form, interpreting both nutrient and food consumption. Anticipating the changing needs of researchers involved with the study also adds to the utility of a nutrient database, and ideally this should be considered in the planning stages of database construction.

References

- Abraham, B & Webb, K. 2001. Interim evaluation of the voluntary folate fortification policy. Australian Food and Nutrition Monitoring Unit. AusInfo, Canberra.
- Arcot, J & Shrestha, A. 2005. Folate: methods of analysis. Trends in Food Science and Technol. 16: 253-266.
- Ashton, BA & others. 1997. Development of a dietary supplement database. Aust. NZ J Pub. Health 21: 696-702.
- ANZFA. 1999. National Nutrition Survey Nutrient Database. Australia and New Zealand Food Authority. Canberra.
- Baghurst, KI & Baghurst, PA. 1990. The uses of food compositional data in nutritional epidemiology. In (Greenfield, H. ed). Uses and abuses of food composition data. Food Aust. 42(8): 511-513.
- Chua, B & others. 2006. Dietary fatty acids in 5-year incidence of age-related maculopathy. Arch Ophthalmol. 124: 981-986.
- Chug-Ahuja, JK & others. 1993. The development and application of a carotenoid database for fruits, vegetables and, selected multi-component foods. J Am. Diet. Assoc. 93 (3): 318-23.
- Department of Community Services and Health. 1990. Nuttab90. Nutrient Database for use in Australia. Australian Government Publishing Service, Canberra.
- Department of Community Service and Health. 1995. Nuttab95. Nutrient Database for use in Australia. Australian Government Publishing Service, Canberra.
- Flood, V & others. 2002. Dietary antioxidant intake and incidence of age-related maculopathy: the Blue Mountains Eye Study. Ophthalmology. 109: 2272-8.
- Flood, VM & others. 2004. Issues in assessing the validity of nutrient data obtained from a food frequency questionnaire: folate and vitamin B12 examples. Pub Health Nutr. 7: 751-756.
- Holland B & others. 1991. McCance and Widdowson's The Composition of Foods. 5th Ed. The Royal Society of Chemistry, Cambridge and Ministry of Agriculture, Fisheries and Food, London.
- IOM. 1998. Dietary references for thiamin, riboflavin, niacin, vitamin B6, folate, vitamin B12, pantothenic acid, biotin and choline. Institute of Medicine, National Research Council. National Academy Press, Washington, DC.
- Krause, MV & Mahan, LK. 1984. Food, Nutrition and Diet Therapy. 7th ed. W.B. Saunders Co, Philadelphia USA.
- Lawrence, MA, Rutishauser, IHE & Lewis, JL. 1999. An analysis of the introduction of folate-fortified products into stores in Australia. Aust. J Nutr. & Diet. 56: 15-21.
- Lazarus, R. 1999. FFQ Masher: Food frequency questionnaire software analysis package.
- Leemhuis, C, Cunningham, J & Arcot, J. 2006. Comparing Australian and United Kingdom folate values. Food Standards Australia New Zealand. Poster.
- Mann, NJ & others. 2003. Development of a database of fatty acids in Australian foods. Aust. J Nutr. & Diet. 60: 42-45.
- Manzi, F & others. 2002. The intake of carotenoids in an older Australian population: The Blue Mountains Eye Study. Public Health Nutr. 5: 347-352.
- Scheelings, P. 1996. Prospects for measuring folate in Australian foods. Aust J Nutr. & Diet. 53S: S23-S28.
- Smith WT. 1997. PhD thesis: Risk factors for age-related maculopathy. 1-287.
- Smith, W & others. 1998. Validity and reproducibility of a self-administered food frequency questionnaire in older people. Aust. NZ J. Publ. Health 22: 456-463.
- Tan, J & others. 2007. Dietary antioxidants and the long-term incidence of age-related macular degeneration: Blue Mountains Eye Study. Ophthalmol. In Press. On-line doi: 10.1016/j.opththa.2007.03.083
- USDA. 1998. USDA – NCC Carotenoid database.
- United States Department of Agriculture <http://www.nal.usda.gov/fnic/foodcomp/Data/Carot/>.

Willet, W & others. 1988. The use of a self-administered questionnaire to assess diet four years in the past. Am. J. Epidemiol. 127: 188-199.

Xyris. 1988. Software. Diet 1.3. Dietary analysis software. Brisbane.

Dr Victoria Flood is Nutritional Epidemiologist with the NSW Centre for Public Health, Nutrition, Human Nutrition Unit, University of Sydney, 2006 and the Centre for Vision Research, Westmead Millennium Institute, University of Sydney, 2145; email: v.flood@mmb.usyd.edu.au. Professor Wayne Smith is with the Centre for Clinical Epidemiology and Biostatistics, University of Newcastle, NSW 2300. Elena Rochchina, Senior Statistician, Dr Jie Jin Wang is Deputy Director and Professor Paul Mitchell is Director with the Centre for Vision Research.