Author's response to reviews

Title: Food composition database development for between country comparisons.

Authors:

Anwar T Merchant (merchant@ccc.mcmaster.ca)
Mahshid Dehghan (mahshid@ccc.mcmaster.ca)

Version: 2 Date: 24 December 2005

Author's response to reviews: see over
“Food composition database development for between country comparisons,” by Anwar T. Merchant ††, Mahshid Dehghan†

This paper clearly addresses an important project that needs to be documented in the literature. However, much of this draft is too rough and incomplete for publication. In the following review, I note several key concerns and suggestions. In addition, I am returning a file of the draft text, in which I have inserted many other comments and editing suggestions for the authors’ consideration.

Page 5, line 18. “The [USDA] database is kept current by a comprehensive and ongoing program of research and food testing.” This is far too strong a statement regarding at least some foods. Although revisions of the database are issued frequently, it should be made clear that the revision dates seldom reflect the data dates. Each revision generally updates only a small faction of the data.

Here are 4 examples handy to me (the first 4 of 5 that I checked): USDA’s 1999 data (release 13) for at least 13 nutrients are identical to its 1982 hard copy data for raw snap beans (NDB 11052), raw beet greens (11086) and raw Brussels sprouts (11098). In the case of raw asparagus (11011), the 1999 data are identical to a 1990 revision.

These 17- and 9-year lapses in USDA’s data may be less common among most foods of interest for the author’s PURE study and for INFOODS, but there seems no way of knowing, short of checking prior releases (many are available at USDA’s Website, and I have others) to learn when each food’s data of interest were last updated. Though this would be tedious, it seems to me worth documenting and correctly characterizing in this paper. A random sampling method might be considered in lieu of a full evaluation.

We agree that the USDA nutrient database, though it is updated regularly, may not be completely current. We have clarified that the ongoing updating is based on selected foods and is not necessarily comprehensive. (pages 5, 6).

Page 6, line 9 to page 7, line 5. “Our main goal in referring to local data sources was to select foods from the USDA nutrient database that were closest in nutrient content to those foods available in the country. This step was necessary because the USDA nutrient database [sometimes] has several entries for the same food, and the nutrient content of foods can differ because of variation in species, soil in which a plant is grown, or fortification policies.” These sentences, especially the last one, seem poorly considered. The point about fortification seems valid, but the available multiple entries in the USDA database rarely document variations in “species” (meaning varieties or cultivars; please reconsider the terminology) and probably never document variations due to the soil in which the plant was grown. Truly different species I would not usually consider “the same food.” Also, what is the definition of “match” here? Would “best match regarding xyz” be a more accurate term?
There is only one purported example here of matching nutrient content of “the same food”—Zimbabwe corn—and it is unclear. Whether it refers to dry or sweet corn (please specify), the only USDA entries appear to be for white and yellow varieties, which are dubious “same foods,” or for dried Navajo corn. The mentioned peeled and unpeeled apples are certainly not “the same food.” Further, the reference to “choosing” a “generic apple” seems inappropriate, as I find only one raw apple in the current USDA data. If the reference is instead to multiple entries for different types of cooking or drying, this is not stated, and again, I would not call dried foods “the same food.” If the reference is to rare multiple entries for seasonal variability, geographical variability within the US (soil, climate, latitude and probably cultivar), variability due to storage time, or the many different cuts of red meat with differing amounts of fat, all this should be made clear. I urge tabulation (not mere examples) of foods for which non-obvious choices were made among several USDA entries. If there were many, they could be enumerated in categories rather than individually. Otherwise, readers can hardly know or evaluate what was done.

Thank you for pointing this out. We have described the algorithm we used to select foods and have added examples of how it can be used (Figure 1, and Table 2, pages 6, 7).

Page 7, line 10. “We call this the PURE-USDA food composition table for each participating country.” Please give basic details: How many foods and which nutrients are currently included? Are these properties the same for each country included? If not, I urge a table with this information. Such a table could also include other worthy information, such as the numbers of foods and nutrients with missing data, the numbers of foods using unaltered USDA data, and the number of non-obvious choices among USDA foods (see prior heading).

We used the approach of using a set of key foods to select foods from the USDA nutrient database. Our PURE databases therefore had all the nutrients that were present in the USDA SR18 version of the nutrient database. For the purpose of this manuscript we have only selected those foods that we needed to calculate nutrient intake from beef stew. The country specific nutrient databases thus are not yet complete. We therefore cannot provide these data yet.

Page 7, line 12 to end of paragraph. “Estimation of nutrient content of mixed dishes” I cannot easily evaluate these steps. If they are published, a reference is needed. If they are unpublished, then these 12 lines seem skimpy, and this procedure seems to me to merit more careful description and scrutiny. Also, references are needed for Handbook 102 and for “retention factors.”

We have rewritten these sections and provided an example (pages 9, 10).
Page 12, lines 6-23 “Our approach also has some limitations.” What about usually (I presume usually) unavoidable mismatches in macronutrients between USDA data and local foods?

What about the apparent assumption that “matching” macronutrients also “matches” micronutrients that you did not or could not “match”? (E.g., implied on page 11, lines 13-18).

What about random uncertainties in the USDA means themselves? At least in vegetables, many USDA “means” are based on only zero, one or two samples of the food (N), and one has little idea about reliability, as USDA does not report a standard error (SE) for those means. Even when N = 3 or more, USDA’s SEs are sometimes substantial, because of inherent variability among samples of the same food (using that phrase in a narrow sense). We recently noted that coefficients of variation (CV = SD/mean) in USDA’s 1999 (SR13) data among samples of the same, raw vegetable ranged among 43 garden crops from 17% to 30% for protein, ash, P and niacin; 30% to 40% for Ca, vitamin A, thiamin, riboflavin and ascorbic acid; and 53% and 59% for Fe and fat (Davis, et al., J Am Coll Nutr 2004; 23(6): 669). It would be valuable for the PURE-USDA databases to characterize the underlying USDA data regarding Ns and relative SEs (RSE = SE/mean).

Our paper cited above also raises a question about the key assumption that foods sampled by USDA are usually good stand-ins for their international counterparts. We report apparent declines in some nutrient concentrations approximately during the transition in the US to intensive, industrial farming methods (USDA data published in 1950 and 1999). Such methods do not likely apply uniformly or to the same degree in the other countries considered here. This would create errors that are not systemic in the sense used on page 10, line 16.

We have included the limitations that have been suggested. We agree that the nutrient content of some of the foods we have selected from the USDA nutrient database will be different from the actual foods in the local countries. Actually testing the foods in all the countries would be the best way to estimate nutrient content, but that approach would be prohibitively expensive. The approach we are suggesting would not eliminate the possibility of these errors. However, it would reduce other errors, such as those introduced by differing assays or methods of selecting test foods. Moreover, the approach we have taken is consistent between countries and therefore the estimates would likely be more comparable.

Page 13 lines 1-8
We have prepared a food composition database primarily based on the USDA food composition database, from which to estimate nutrients using FFQ data from different countries. This is an entirely worthy goal, but I find it very difficult to know clearly what steps were taken, how extensively they were taken, and, most importantly, how well the goal was achieved. The broad claims are supported mainly by
unquantified generalities and questionable examples. The assumptions made are not clearly or fully identified, and are subjected to little or no evaluation. This paper describes a large and important project that deserves better documentation and a more careful assessment of remaining uncertainties.

A limitation of our approach is that it is not validated, primarily because there is no gold standard available to measure it against. However, we have identified potential sources from which errors could arise and addressed means of minimizing those, but some misclassification is inevitable. We have, however, proposed a system of database development that is replicable, and consistent between countries.

We have made the corrections in the tables. We prefer to use the intraclass correlation (ICC) rather than the Pearson correlation coefficient because the former takes into account not only the correlation but also differences in mean levels.