

Criteria for Acceptable Dietary Intake Biomarkers

Ross L. Prentice^{1,2}



ABSTRACT

Dietary intake biomarkers that can be written as actual intake, plus 'error' that is independent of actual intake and confounding factors can substitute for actual intake in disease association analyses. Also, such biomarkers can be used to develop calibration equations using self-reported diet and participant measures, and biomarker-calibrated intakes can be calculated in larger cohorts for use in disease association analyses. Criteria for biomarkers, and for biomarker-calibrated intakes, arise by working back from properties needed for valid disease association analyses. Accordingly, arguments for a potential biomarker are strengthened if error components are small relative to actual intakes, and important sources of reduced sensitivity or specificity

are not apparent. Feeding study biomarker development can then involve regression of actual intake on putative biomarkers, with regression R^2 values playing a role in biomarker evaluation. In comparison, 'predictive' biomarker status, as argued in this issue by Freedman and colleagues for 24-hour urinary sucrose plus fructose as biomarker for total sugars, involves regression of potential biomarker on actual intake and other variables, with parameter stability across populations and limited within-person variability as criteria. The choice of criteria for biomarkers and for biomarker-calibrated intakes, is discussed here, in the context of total sugars intake.

See related article by Freedman et al., p. 1227

Introduction

First, I would like to commend Dr. Tasevska, her mentor the late Dr. Bingham, and their colleagues for their leading research over many years toward dietary biomarker development, including for 24-hour urine sucrose plus fructose (24uSF) as biomarker for total sugars intake. Similarly, I would like to commend Drs. Freedman, Kipnis, and colleagues for their substantial research on dietary measurement error modeling and mitigation. The work by Freedman and colleagues (1) to be commented on here, derives from the combined efforts of these strong research teams.

Nutritional epidemiology research is greatly complicated by correlations among foods and nutrients of interest, and by correlations among their respective measurement errors. Established dietary intake biomarkers; for example, doubly-labeled water (DLW; ref. 2) for total energy, and urinary nitrogen (UN) for total protein (3), show that self-reported intake may be weakly associated with actual intake, and that there may be strong systematic biases related to participant characteristics [e.g., body mass index (BMI)], whether based on food frequency questionnaires, food records, or dietary recalls (4–6). Much nutritional epidemiology reporting relies fundamentally on self-reported diet for validity of findings. In comparison, dietary intervention trials can lead to randomization group outcome contrasts that do not rely on self-reported dietary data for their validity. However, cost and logistical challenges preclude the conduct of many full-scale dietary intervention trials with chronic disease outcomes, and self-reported dietary data may be needed for randomization group adherence assessment.

The use of dietary intake biomarkers has potential to strengthen the nutritional epidemiology research area, both for associations targeted in observational studies, and for adherence assessment in randomized trials. The 'biomarker' label is used ubiquitously in biomedical research, sometimes for a bio-specimen measure that simply shows some positive correlation with a dietary intake of interest. However, the principal utility of a dietary biomarker, in observational study contexts, is (arguably) to provide a replacement for actual intake in a manner that yields reliable disease association results. For example, with hazard ratio (Cox) modeling of clinical outcomes this reliability will typically follow if actual intake can be written as intake biomarker plus 'error' that is independent of actual intake and of confounding factors (7–9).

Feeding studies, as carried out in the United Kingdom and United States by Tasevska and colleagues (10, 11) for total sugars intake, are central to biomarker development. These studies include designs where an individualized diet is provided that approximates each participant's usual diet. Such designs have potential for pertinent bio-specimen measures to stabilize quickly and for retention of the dietary intake variation of the underlying study population.

Our Women's Health Initiative (WHI) research group conducted a 153-participant feeding study during 2010 to 2014 using this type of design (12). We reported related intake biomarkers derived from serum carotenoid and tocopherol concentrations and related these to cardiovascular disease, cancer, and diabetes in a WHI sub-cohort (13). In other recent reports, we used serum and urine metabolomics profiles, in conjunction with DLW and UN established dietary biomarkers, to define novel biomarkers for protein and carbohydrate intake and for their densities relative to total energy. Also for certain protein and carbohydrate components, and we used these in recent association studies with this same set of disease outcomes (14–16). In each of these contexts, biomarker development involved regression of feeding study intake on potential biomarker and other participant characteristics (with dietary and bio-specimen measures log-transformed). This framework allows participant characteristics that may improve the properties of an intake biomarker to be included in biomarker specification. We use a 36% or larger regression R^2 (or cross-validated R^2) as a key criterion in biomarker development, based on benchmark values of about 50% and 40% for DLW energy and UN protein in this feeding study context. Of course, complex metabolism

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington. ²Department of Biostatistics, University of Washington, Seattle, Washington.

Corresponding Author: Ross L. Prentice, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, P.O. Box 19024, Seattle, WA 98109-1024. Phone: 206-667-4264; E-mail: rprentic@whi.org

Cancer Epidemiol Biomarkers Prev 2022;31:1151–3

doi: 10.1158/1055-9965.EPI-22-0180

©2022 American Association for Cancer Research

and physiology may attend bio-specimen measures under consideration, and we also require that there be no known dietary sources or metabolic pathways that may reduce biomarker sensitivity or specificity in an important way. This latter criterion can typically only be assessed informally, though adherence to such a criterion may be improved by allowing biomarker specifications to include other available participant characteristics and measures.

These criteria can be compared with those used by Freedman and colleagues (1) to establish 24uSF as a ‘predictive biomarker’ for total sugars. They use linear regression the ‘other way around’ of log-transformed 24uSF on log-feeding study intake along with participant characteristics, including a person-specific random effect. Their argument for predictive biomarker establishment relies on stability of regression equations across UK and US study feeding studies and limited random effect variance compared with feeding study intake variance. Properties of biomarker transportability across populations, and limited within-person biomarker variability over time are certainly desirable properties, but do they ensure that the biomarker can substitute for actual intake in disease association analyses?

A putative biomarker, such as 24uSF may be on a different scale from actual intake. However, the regression model of Freedman and colleagues (1) can be ‘solved’ to obtain a total sugars intake value corresponding to a measured 24uSF. It may not matter much which regression is used to develop an equation connecting biomarker to feeding study intake, but we think there are advantages to regressing feeding study intake on potential biomarker as we do in WHI. For example, our modeling approach provides a convenient framework for incorporation of additional participant measures in biomarker specification toward enhancing biomarker correspondence with feeding study intake. It also provides an immediate allowance for an independent random error component for the feeding study assessed intake, because such errors shouldn’t bias the fitted linear equations, only reduce estimation precision.

In the total sugars context one could, for example, use the regression of feeding study intake on potential biomarker to examine whether additional days of 24uSF can augment the regression R^2 (e.g., see Figure 1 in reference 17 for scatter plot of 24uSF versus daily feeding study total sugars), or whether sugars other than sucrose and fructose (e.g., maltose) can be adequately acknowledged by the small fractions (10) of dietary sugars that are metabolized and excreted in the urine as sucrose or fructose. One can even consider the addition of the self-reported sugars intake that would be used in subsequent disease association analyses for possible R^2 increment. Interestingly, if the proposed biomarker is somewhat weak in the sense that these (typically cohort baseline) dietary self-report data explain feeding study intake variation beyond that explained by the biomarker, then bias in disease association analyses may arise unless the biomarker equation is augmented to include the self-reported intake, as has been confirmed in simulation studies (18).

It may not be practical to carry out the bio-specimen analyses needed for biomarker calculation on an entire study cohort. Then, a

two-stage approach may be considered in which biomarker values in a cohort sub-study, nonoverlapping with the feeding study, are used to generate calibration equations that estimate intake using self-reported intake and other measures in the sub-study. To do so, the feeding study biomarker equation is used to calculate sub-study biomarker values and these are regressed on concurrent self-reported dietary intakes and pertinent participant characteristics. Calibration equations also need to satisfy certain criteria: It turns out that, under biomarker modeling assumptions, the precision of related calibrated-intake and disease association analyses in larger cohorts depends primarily on properties of the calibration equations and little on the magnitude of the error component in the biomarker equations. Therefore calibration equation regression R^2 values can be considered in relative isolation, and we again use an $R^2 > 36\%$ for the regression of biomarker values on self-reported intakes and pertinent participant characteristics as a calibration equation criterion, along with the absence of known sources of reduced sensitivity or specificity. In addition, we examine the specific contribution of the self-reported dietary intake in calibration equation evaluation. For example, in WHI with DLW energy as biomarker the self-reported energy estimates mentioned above explain only a few percent of biomarker variation, while larger fractions relate to age, ethnicity and, especially, BMI (5, 6), adding complexity to the interpretation of related energy intake disease association analyses (19).

In summary, the use of dietary intake biomarkers has potential for a fresh examination of important nutritional epidemiologic associations. The criteria that biomarkers should satisfy is a worthy research topic. Two-stage approaches that use intake biomarkers to adjust self-reported intakes for measurement error also have an important role, though criteria then need to be satisfied for both biomarker and calibration equations. In relation to our biomarker equation R^2 criterion, the high correlation mentioned in Freedman and colleagues (1) between 24uSF and an estimate thereof from feeding study intake may suggest a strong correlation also between feeding study intake and estimated intake from the 24uSF measure, though the latter not given in (1). Transportability of biomarkers meeting criteria to other cohorts is also an important goal, for which plausibility could be reduced if biomarkers require substantial complexity to meet criteria. This topic too deserves further study in this most promising biomarker approach to future nutritional epidemiology research.

Authors’ Disclosures

No disclosures were reported.

Acknowledgments

This work was supported by the National Heart, Lung, and Blood Institute, NIH, U.S. Department of Health and Human Services (contracts HHSN268201600046C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, HHSN268201600004C, HHSN271201600004C, and 75N92021D00001); and NCI grants R01 CA119171.

Received March 10, 2022; revised March 17, 2022; accepted March 17, 2022; published online June 1, 2022.

References

1. Freedman LS, Kipnis V, Midthune D, Commins J, Barrett B, Sagi-Kiss V, et al. Establishing 24-hour urinary sucrose plus fructose as a predictive biomarker for total sugars intake. *Cancer Epidemiol Biomarkers Prev* 2022;31:1227–32.
2. Schoeller DA, Hnilicka JM. Reliability of the doubly labeled water method for the measurement of total daily energy expenditure in free-living subjects. *J Nutr* 1996;126:348S–54S.
3. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. *J Nutr* 2003;133:921S–4S.
4. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN Study. *Am J Epidemiol* 2003;158:1–13.

5. Neuhouser ML, Tinker L, Shaw PA, Schoeller D, Bingham SA, Horn LV, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am J Epidemiol* 2008;167:1247–59.
6. Prentice RL, Mossavar-Rahmani Y, Huang Y, Horn LV, Beresford SAA, Caan B, et al. Evaluation and comparison of food records, recalls and frequencies for energy and protein assessment using recovery biomarkers. *Am J Epidemiol* 2011;174:591–603.
7. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982;69:331–42.
8. Wang CY, Hsu L, Feng ZD, Prentice RL. Regression calibration in failure time regression. *Biometrics* 1997;53:131–45.
9. Carroll RJ, Ruppert D, Stefanski LA, Crainiceano CM. Measurement error in nonlinear models, a modern perspective, 2006. Boca Raton, FL: Chapman and Hall/CRC.
10. Tasevska N, Runswick SA, McTaggart A, Bingham SA. Urinary sucrose and fructose as biomarkers for sugar consumption. *Cancer Epidemiol Biomarkers Prev* 2005;14:1287–94.
11. Tasevska N, Pettinger M, Kipnis V, Midthune D, Tinker LF, Potischman N, et al. Associations of biomarker-calibrated intake of total sugars with the risk of type 2 diabetes and cardiovascular disease in the Women's Health Initiative Observational Study. *Am J Epidemiol* 2018;187:2126–35.
12. Lampe JW, Huang Y, Neuhouser ML, Tinker LF, Song X, Schoeller DA, et al. Dietary biomarker evaluation in a controlled feeding study in women from the Women's Health Initiative cohort. *Am J Clin Nutr* 2017;105:466–75.
13. Prentice RL, Pettinger M, Neuhouser ML, Tinker LF, Huang Y, Manson JE, et al. Application of blood concentration biomarkers in nutritional epidemiology: example of carotenoid and tocopherol intake in relation to chronic disease risk. *Am J Clin Nutr* 2019;109:1189–96.
14. Zheng C, Nagana Gowda GA, Raftery D, Neuhouser ML, Tinker LF, Prentice RL, et al. Development of potential metabolomics-based biomarkers of protein, carbohydrate, and fat intakes using a controlled feeding study. *Eur J Nutr* 2021;113:1083–92.
15. Prentice RL, Pettinger M, Neuhouser ML, Raftery D, Zheng C, Gowda N, et al. Biomarker-calibrated macronutrient intake and chronic disease risk among postmenopausal women. *J Nutr* 2021;151:2330–41.
16. Prentice RL, Pettinger M, Zheng C, Neuhouser ML, Raftery D, Nagana Gowda GA, et al. Biomarkers for components of dietary protein and carbohydrate with application to chronic disease risk among postmenopausal women. *J Nutr* 2022 Jan 7 [Epub ahead of print].
17. Tasevska N, Midthune D, Potischman N, Subar AF, Cross AJ, Bingham SA, et al. Use of the predictive sugars biomarker to evaluate self-reported total sugars intake in the Observing Protein and Energy Nutrition (OPEN) study. *Cancer Epidemiol Biomarkers Prev* 2011;20:490–500.
18. Huang Y, Zheng C, Tinker LF, Neuhouser ML, Prentice RL. Biomarker-based methods and study designs to calibrate dietary intake for assessing diet–disease associations. *J Nutr* 2022;152:899–906.
19. Zheng C, Beresford SA, Horn LV, Tinker LF, Thomson CA, Neuhouser ML, et al. Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. *Am J Epidemiol* 2014;180:526–35.