A search for truth in dietary epidemiology\textsuperscript{1–4}

Gary E Fraser

ABSTRACT Although results from epidemiologic studies of diet have taught us a great deal, much of the evidence remains mired in controversy because of the inconsistency of results among apparently good studies. I conclude that this can be largely explained by the combination of 2 problems: confounding and measurement error. This recognition allows some judgment as to which studies may be less prone to these difficulties and a search for new analytic methods that can produce less biased and more consistent results. The potential correlations between many nutrients, and to a lesser extent foods, make it difficult to know whether the nominated variable is actually the active principle or whether there is some other dietary risk factor that is closely associated. It is not generally recognized that all traditional analyses of this sort are based on a powerful but incorrect assumption: that there are no errors in dietary assessment. If the incorrect assumption is not satisfied, relative risk estimates become distorted—reduced by one-half or more in some cases. Regression calibration is a newer technique that uses a calibration substudy to provide information about errors and to correct results from the main study. There are a number of variants of this technique, all requiring assumptions about the data. Regression calibration methods that use carefully selected biological surrogates (correlates) of the dietary factor of interest in the calibration study seem to use more realistic assumptions.

KEY WORDS Measurement error, bias, confounding, dietary patterns, regression calibration

INTRODUCTION

Unlike many other behavioral risk factors, such as cigarette smoking or riding a high-powered motorcycle, eating is not optional. The question is not whether to eat but how to eat for optimal health. Implicit in this question is the assumption that diet matters. Although the accepted wisdom is that diet is an important risk factor for cancer, heart disease, and some other disease endpoints, the evidence is confusing enough to raise doubts in the minds of some. As evidence of this, entering the words diet and controversy in any Internet search engine yields many pages of “expert” commentary.

Over the past 40 y, dietary epidemiology has spawned thousands of investigations and many more peer-reviewed publications. Why have clear answers so often been difficult to find? By contrast, nearly all the large studies of blood cholesterol or blood pressure and heart disease find that these are important risk factors. Moreover, the epidemiologic results here are in quite good agreement with results from subsequent clinical trials. That a later age at birth of a first child increases a woman’s risk of breast cancer is undisputed. The studies of these factors, even if they differ in details, speak with a clear and unified voice.

It is generally accepted that a high intake of saturated fat is harmful and that low doses of alcohol protect some people against heart disease (ignoring other issues). Likewise, broadly speaking, higher consumption of fruit and vegetables almost certainly protects against many cancers. We know a good deal more than this about diet and disease, but much of the evidence for diet is complicated, and as recently described by Byers (1), for many it is unconvincing. The studies often do not provide uniform results, and there is the problem, for instance, that some apparently good studies do not support the idea that fiber or fish consumption protects against ischemic heart disease or that red meat is hazardous, even though there is a strong suspicion from basic science and other epidemiologic work that this should be so.

Below I discuss 2 main problems that probably explain much of this situation: confounding and measurement error. In combination, these 2 factors are powerful sources of confusion, not only about the magnitude of effects but also about statistical significance and confidence intervals. Finally, there is a nonmathematical discussion of regression calibration, which is one promising analytic technique to minimize the measurement error problem.

CONFOUNDING

Definitions of confounding often include the idea of confusion, because of the mixing of effects from 2 or more variables. Although the investigator may not be able to precisely identify the variables that are being mixed, the problem can be very real when different risk factors tend to clump together in the population. A multivariate analysis will in theory unravel this confusion, but only if all the important factors are measured with reasonable accuracy.

Dietary analyses are peculiarly predisposed to confounding simply because of the complexity of this variable. Most investigators have concluded that it takes at least 60 and perhaps as many as 130 questions about foods to even approximately

\textsuperscript{1} From the Department of Epidemiology and Biostatistics, Loma Linda University, Loma Linda, CA.
\textsuperscript{2} Presented at the Fourth International Congress on Vegetarian Nutrition, held in Loma Linda, CA, April 8–11, 2002. Published proceedings edited by Joan Sabaté and Sujatha Rajaram, Loma Linda University, Loma Linda, CA.
\textsuperscript{3} Supported in part by an NIH Senior Fellowship Grant (1F33CA66287).
\textsuperscript{4} Address reprint requests to GE Fraser, Department of Epidemiology and Biostatistics, Evans Hall, Room 203, Loma Linda University, Loma Linda, CA 92350. E-mail: gfraser@sph.llu.edu.
characterize a diet. In addition, a particular food contains thousands of discrete phytochemicals, along with recognized nutrients, vitamins, and minerals. Many of these are poorly characterized, and it is quite unclear which have the potential to affect disease risk.

Similarly, many chemicals are found together in particular foods. They are not randomly assorted. Thus, an index of one such factor (e.g., β-carotene) may also be a good index of another (e.g., vitamin C, or perhaps some other poorly defined phytochemical). If the first has no effect on disease but the second is powerfully protective, the index will probably indicate protection in the analysis. However, as the label for the index is β-carotene, quite the wrong conclusion will be reached. Perhaps it is issues like this that have led to the apparent disagreements between observational and clinical trial work on the possible effects of several antioxidant vitamins on risk of chronic disease.

There is no easy solution to this problem. However, if there turns out to be only a moderate-sized subset of phytochemicals that strongly affect the disease of interest, the problem may be manageable. If a true causal factor is closely linked to a more easily identified or currently favored second, inactive factor, then this confounding may prove difficult to find. Including all closely linked causal factors in the statistical model will break the confounding, but there will be a considerable cost in statistical power caused by the multicollinearity. Hence, very large studies may be necessary to prove the effect of the active variable. Of course, if the factors are that closely linked, a comfort is that foods containing the one will usually include the other as well.

It may be that a greater research focus on foods, rather than nutrients and phytochemicals, will be helpful. Cultural and other factors result in the consumption patterns of many foods being somewhat linked, an opportunity for confounding if several foods affect disease risk. However, the correlations between the use of different foods are usually relatively low, given the number of different eating patterns throughout a population and the large number of food choices available. This tends to limit the impact of confounding. The analyses may then be able to select foods that have particularly beneficial or hazardous combinations of nutrients and phytochemicals, although these components are never explicitly identified. Nevertheless, once such foods are found, this may open the door to productive basic science and other work.

A case in point is work (by myself and others) strongly suggesting that moderate nut consumption protects against heart disease (2). The reasons for this effect are not entirely clear. The phytochemical content of nuts is poorly characterized, as is any biological activity that these factors may possess. Clearly, there is room for additional productive work in this area.

Some have advocated a focus on dietary patterns. This markedly reduces the mathematical complexity by collapsing a complicated mix of foods, nutrients, and so on to a much smaller number of dietary patterns. Thus, the confounding problem described above disappears, as the patterns are mutually exclusive. However, the difficulty is how to define the dietary patterns to give the best understanding of the forces at work.

Defining patterns a priori, using knowledge from basic science or other sources, may produce patterns that would have markedly contrasting associated disease risks but to which no one subscribes. On the other hand, factor or cluster analyses may identify the patterns that actually occur in the population, but as they probably result from cultural, socioeconomic, or religious influences, there is no assurance that they will contrast much in their effects on health.

THE EFFECT OF MEASUREMENT ERROR ON RISK ESTIMATORS, STATISTICAL SIGNIFICANCE, AND CONFIDENCE INTERVALS

Many people have found themselves filling out a food-frequency questionnaire and wondering how accurate their information is. Diet is so complex, with seasonal differences and changing availability of proprietary products, that memory fails. These problems will potentially be exacerbated in case-control studies where the emphasis on memory is greater (to describe the premorbid diet) and there is no opportunity for repeat dietary assessment. There is a tendency for subjects to view their diets too favorably. It is well known that many obese subjects underreport intake of food (3, 4), probably without intending to. When calculating nutrient intakes from foods one is reliant on tables of values (e.g., US Department of Agriculture tables) that, while enormously useful, contain important errors.

Population scientists have always understood that there are serious errors from many sources in their dietary data. However, there has often been an unspoken assumption that random errors in dietary assessment will balance out and produce only random errors in relative risk estimates during statistical analysis. This is not the case.

First, there is no assurance that the dietary errors are random, compared with systematic errors that distort dietary assessment in one direction. Second, even random errors in dietary (or other exposure) assessment do systematically bias relative risk estimates. This is because the estimates of β coefficients incorporate the sums of squared errors. Even balanced positive and negative errors do not cancel in such a sum of squares but always increase the value of this sum.

If there is only one variable in the model, the distortions bias estimates of effect toward the null, or zero effect. If several variables are measured with error in the model, then although all effects are usually biased toward the null, there is no assurance of this, and circumstances exist where nonconservative biases are quite possible. Our recent work shows in a simulation study that the biases are not trivial (5) and that estimates of real effects will often be reduced to half their true values or even less.

Expressed a little differently, in multivariate work, the measurement error problem becomes one of confounding. However, this time it is not corrected by including all variables in the statistical model. This is because there is residual confounding, so named because it is confounding between the errors of the various dietary factors. For instance, if those who overreport their intake of fruits and vegetables are also those who underreport intake of red meat, their heart disease experience will be less favorable than expected from their dietary reports. This will usually diminish both the estimated beneficial effects of the fruits and vegetables, and the estimated hazardous effects of the meat (though there are other less likely possibilities).

Equally troubling is that the measurement error problem in multivariate analyses will usually result in the wrong P values and may produce confidence intervals that erroneously exclude the null value of zero. Again, our simulation work provided clear examples (5). Because of the improper characterization of the dietary variables, and the residual confounding, the estimated relative risks end up being a mix of the true relative risks of the
several intercorrelated dietary variables. Then it is easy to see that if a variable actually has no effect, its estimated effect may be null, because of the mixing of its zero true effect with the true effects of correlated variables.

This problem is not corrected by larger studies. Actually, the problem gets worse. A Wald test of statistical significance is the estimate of the effect divided by its standard error. We have seen that the estimated effect will often be quite wrong and nonnull even for a variable with no effect. Yet a large study size results in a smaller standard error; hence, greater calculated statistical significance (and confidence intervals that exclude the null) can easily be achieved by a null variable.

WHICH RESULTS ARE MORE CREDIBLE?

Dietary epidemiology has already taught us much that is important. However, it is also true that much less is known with confidence than most people think. How can conclusions, even tentative ones, be drawn in this complicated situation?

If a variable is only moderately correlated with a powerful risk factor, then confounding may still cause some biases, but they are of relatively small magnitude. This was shown by Flanders and Khoury (6). Second, if the effect of the variable is quite strong, then the usual result of measurement error to diminish estimates will still leave some apparent effect intact. A validation study where both the questionnaire method and a more accurate reference method are gathered on a representative subsample of subjects should be required. One minus the square of the correlation coefficient between the questionnaire estimate and an accurate reference method equals the proportion of the total variance of the questionnaire estimator that can be explained by error. Statistical significance will be a poor guide to a more accurate result, particularly in large studies and in highly multivariate models.

Thus, we should prefer results that suggest stronger effects, as weaker effects (eg, relative risks 0.7–1.3) are readily produced by confounding and hence measurement error, when in fact there is no effect. Second, where there is much error in estimates of effects, as is commonly the case (even if the calculated confidence interval is narrow), it is still true that averaging results from different studies tends to diminish the effects of some of the “random” errors, although other error effects will remain. Hence, it is wise to give greater weight to results that are consistent with those from other studies of different populations using different measurement techniques that will not closely duplicate the same errors.

These considerations may clarify why there is broad consistency for the finding that fruit and vegetable consumption protects against many cancers. As these foods are ubiquitous, their use overall will not be strongly correlated to other factors. Thus, confounding is somewhat limited. It may also be that their overall effect is quite strong; then some effect will be detectable despite the attenuation by measurement error. The consistent finding that health-conscious individuals, who often prefer these foods, are strongly protected from many causes of mortality (7–9) is compatible with this conclusion.

THE SEARCH FOR NEW ANALYTIC METHODS: REGRESSION CALIBRATION

Are more detailed conclusions beyond our present methods? For instance, it seems quite likely that meat consumption and low calcium intake are hazardous for colon cancer, that tomato products may protect against prostate cancer, that dietary fiber may protect against ischemic heart disease, that cruciferous vegetables may protect against several cancers, and that isoflavones may protect against some cancers and heart disease. Yet the epidemiologic evidence is inconsistent, probably in large part for the reasons discussed above.

There has been progress in developing methods that reduce bias when evaluating these more detailed dietary hypotheses. A number of methods are being explored at present, but they are usually not conceptually easy. A complicated problem will not have a simple solution.

An identifiable model has the same number of equations as unknowns. Then all the unknowns can be estimated. To work with identifiable mathematical models, some assumptions are necessary in all of these newer correction methods. However, the idea of assumptions is not new. It only seems new. All traditional dietary analyses in epidemiology share one strong but incorrect assumption: that we measure exposures such as foods, nutrients, and phytochemicals with great accuracy. Any analytic method that requires significantly weaker assumptions than this represents real progress even if it does not necessarily produce perfect validity. Such progress seems possible, although at some cost. The studies will need to be more complex and probably larger.

A brief nonmathematical review of some variants of the method that has been most thoroughly investigated, namely regression calibration (10, 11), follows. As indicated in the name, the method involves regression, which is familiar. It also requires a calibration substudy. In all the models described in Table 1, this substudy is used to establish the questionnaire’s misspecification of the true intake, on average. Then this information is used to correct the disease regression.

There is already a necessary assumption, the nondifferential error assumption. This amounts to assuming that those subjects whose errors under- or overestimate the dietary variable more than most do not systematically differ from others having the same true dietary intake in their experience of the outcome disease. The errors referred to are those remaining after taking account of other dietary and nondietary differences that are predicted by the remaining variables in the model. It may be necessary to add variables to the model to make this a tenable assumption. Then it is probably true that these residual errors in the questionnaire data are due to lapses of memory, misperceptions about diet, and random factors. Probably these will not relate strongly to the risk of most diseases, and the assumption is justified.

A difficulty with model 1 is the need to obtain both the true diet and the questionnaire diet on members of the calibration substudy to measure the questionnaire errors. Yet there is generally no way to measure the true diet with presently available technology in free-living individuals. This difficulty led to the identification of gold standard or reference dietary methods (12) that could

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison of the information to be gathered in 4 regression calibration models</strong></td>
</tr>
<tr>
<td><strong>Main study</strong></td>
</tr>
<tr>
<td>Model 1: questionnaire</td>
</tr>
<tr>
<td>Model 2: questionnaire</td>
</tr>
<tr>
<td>Model 3: questionnaire</td>
</tr>
<tr>
<td>Model 4: questionnaire</td>
</tr>
</tbody>
</table>

Downloaded from https://academic.oup.com/ajcn/article-abstract/78/3/521S/4689991 by guest on 09 October 2018
be used as surrogates of the truth. They are considered to be more valid than the questionnaire but are usually much more time-intensive and expensive. Examples are repeated 24-h recalls or multiday diet diaries.

When a specific requirement is placed on the reference instrument, namely that any errors that it incorporates are only random and on average unbiased, it may properly serve in place of the true diet for the purpose of regression calibration (model 2 of Table 1). This means that every subject’s data would approach his or her true value if many estimates from that subject were averaged. Unfortunately, we do not have reference methods that are likely to fulfill this rather stringent requirement. People do not remember their diets well for even 24 h, have trouble estimating portion size, and may not faithfully fill out diaries at the time that they eat. It is suspected that people tend to systematically minimize information at the extremes.

One reason for this requirement that there be only random errors in the reference is the necessity (to simplify the mathematics) that these errors not be related to the same subject’s errors in the questionnaire. Unfortunately, there are indications that these 2 sources of error are often related (13, 14). Subjects who erroneously report high or low on the questionnaire tend to do the same on the reference instrument.

Interestingly, if there is another variable, sometimes called an instrumental variable, that is correlated with the true dietary variable of interest, then it will (12), with certain assumptions, allow the troublesome correlation between errors in the reference and questionnaire to be estimated (model 3 of Table 1). If this error correlation can be estimated, a zero assumption is no longer necessary and this problem is solved. Instrumental variables are often biological in nature; when they are used as estimators of the true dietary variable the errors are unlikely to be correlated with the similar errors in the questionnaire. Thus, a zero for this new correlation is a new but easier assumption. An example of an instrumental variable would be erythrocyte folate when folate is the dietary variable of interest. It is unlikely that an individual who erroneously reports a high folate intake on the questionnaire will also have systematically high blood folate, for example.

However, still another assumption is necessary before it is possible to properly estimate the correlation between errors in the reference and questionnaire. With the instrumental variable in the model, the fact that many individuals will systematically report high or low on the reference method can now be accommodated, so long as averages across a whole group with the same true values are accurate, perhaps because there are equal numbers of subjects who systematically report high and low. Unfortunately, even this does not seem realistic, as it is quite possible that whole groups with the same true values may systematically report high or low, particularly if their diets tend toward the extremes. Thus, a scaling factor is necessary in the model to reduce the high reported intakes and/or increase the low reported intakes. However, this additional factor makes the model nonidentifiable. There are more variables than equations.

What if the reference method is dispensed with and substituted with a second instrumental variable? Then there would be the questionnaire estimate and 2 biological (instrumental) estimators, all of which may incorporate error (model 4 of Table 1). The key feature of both these instrumental variables is that their errors about the true intake should be uncorrelated with similar errors from the questionnaire. This removes the need to estimate both these error correlations. An example of 2 biological estimators of dietary folate may be erythrocyte folate and blood β-carotene. It is not necessary that the label on these variables be “folate” so long as they are correlated with dietary folate and satisfy the assumptions about error correlations.

Using standardized rather than traditional regression calibration allows all the model parameters to be solved with no further strong assumptions (see below). Then the requirement that groups on average report their true intake is unnecessary and a scaling factor can be incorporated into a fully identifiable model.

What is standardized regression calibration? This is when the parameter that is estimated is the product of the disease regression β coefficient and the standard deviation of the true dietary variable. A relative risk can be estimated that corresponds to the effect of moving one or more standard deviations of the true dietary variable through the population. This can generally be interpreted as a comparison of particular population quartiles, so long as there is a monotonic transformation for the true dietary variable from its actual distribution to normality.

CONCLUSIONS

Epidemiologic work over past decades has detected dietary patterns, specific foods, and nutrients that consistently appear to reduce or increase the risk of disease. However, much work in dietary epidemiology has been frustratingly difficult to interpret, because good studies do not always agree. The list of dietary items that have appeared to change risk in some but not other good studies undoubtedly includes extremely important risk factors whose effect sizes are seriously diminished by measurement error when traditional analytic methods are used. There is also the strong possibility that many moderate-sized, statistically significant effects from these same studies could be entirely spurious, again because of the measurement errors. The most controversial issue in the several versions of regression calibration is the validity of the assumptions required to simplify the mathematics. Methods that require 1, or preferably 2, appropriately chosen biological correlates hold promise, because they need only relatively weak assumptions to produce valid results. Given the strong and untenable assumption that underlies traditional analyses—namely, that dietary assessment is accurate—the regression calibration concepts described above are a worthwhile advance.

The author had no conflicts of interest.

REFERENCES