ABSTRACT

Research during the last several decades indicates the failure of existing nutritional labels to substantially improve the healthfulness of consumers' food/beverage choices. The present study aims to fill this void by developing a nutrition metric that is more comprehensible to the average shopper. The healthfulness ratings of 205 sample foods/beverages by leading nutrition experts formed the basis for a linear regression that places weights on 12 nutritional components (ie, total fat, saturated fat, cholesterol, sodium, total carbohydrate, dietary fiber, sugars, protein, vitamin A, vitamin C, calcium, and iron) to predict the average healthfulness rating that experts would give to any food/beverage. Major benefits of the model include its basis in expert judgment, its straightforward application, the flexibility of transforming its output ratings to any linear scale, and its ease of interpretation. This metric serves the purpose of distilling expert knowledge into a form usable by consumers so that they are empowered to make more healthful decisions.


Despite an increased standard of living in many developed countries, health problems attributable to poor nutrition persist, in part, because of consumers’ inability to translate the dietary advice of nutrition experts into behavioral change. Citing the improvement of public health as a primary objective, numerous studies have highlighted the need for a nutritional scoring system that reflects the current views of leading experts and can be used to score any food/beverage for its straightforward application, the flexibility of transforming its output ratings to any linear scale, and its ease of interpretation. This metric serves the purpose of distilling expert knowledge into a form usable by consumers so that they are empowered to make more healthful decisions.

J. M. Martin is a postdoctoral researcher, Harvard Law School Program on Negotiation, and J. Beshears is a doctoral candidate, Department of Economics, both at Harvard University, Cambridge, MA. K. L. Milkman is a doctoral candidate, Information Technology and Management, and Max H. Bazerman is Jesse Isidor Straus Professor of Business Administration, both at Harvard Business School, Harvard University, Boston, MA. L. A. Sutherland is an assistant professor, Department of Pediatrics and Adolescent Medicine and a senior nutrition scientist, Community Health Research Program, Hood Center for Children and Families, Dartmouth Medical School, One Medical Center Drive, Lebanon, NH.

Address correspondence to: Jolie M. Martin, MBA, PhD, Harvard Law School Program on Negotiation, Harvard University, 408 Dunster Mail Center, Cambridge, MA 02138. E-mail: joliemm@gmail.com

Manuscript accepted: October 24, 2008.

Copyright © 2009 by the American Dietetic Association.

0002-8223/09/10906-0015$36.00/0
doi: 10.1016/j.jada.2009.03.009
by Scarborough and colleagues (5), who advocate “a sys-
tematic, transparent and logical process” to categorize
foods by nutritional composition. Scarborough and col-
leagues (23) evaluated each of eight existing nutrient
profile models by the correlation of their ratings with
healthfulness categorizations by food and nutrition pro-
fessionals. The present research builds upon the premise
that expert assessments are, in some sense, the most
comprehensive embodiment of current scientific knowl-
edge, and goes one step further by actually employing
expert ratings to generate a nutrition metric.

METHODS
This study involved surveying US nutrition experts about
the healthfulness of sample food/beverage products, esti-
mating the regression equation that best predicts expert
ratings from the nutrient information on a Nutrition Facts
label, and finally analyzing the applicability of this model
to rating the healthfulness of products outside the initial
sample. All procedures were reviewed and exempted by the

Food/Beverage Sample
A large online grocer provided a database containing
nutritional information for >15,000 unique food/beverage
items. Also listed in the database were the 205 categories
used by the grocer to classify items and the unit sales of
each item in 2005. In order to create a sample representa-
tive of the items that consumers consume most regu-
larly, but also covering a range of food/beverage types, the
most frequently purchased item in each category was
selected to comprise a sample of 205 foods/beverages for
experts to rate. For each of these items, any nutritional
information missing from the database was collected by
researching similar items online (24,25).

Expert Sample
Participation in the survey was requested of 57 leading
nutrition experts belonging to three groups widely recog-
nized for their expertise in nutrition: chairs of the top
tree schools of public health nutrition departments (ie,
Harvard University, John Hopkins University, and the
University of North Carolina); directors of the eight US
Clinical Nutrition Research and Human Nutrition Cen-
ters; and directors of the 46 Coordinated Programs in
Dietetics with accredited status from the American Die-
etic Association. These experts—all of whom hold doc-
toral degrees in related fields—were each offered $250 for
their consultancy, which required them to complete a
1-hour Web survey. The overall response rate was 23%
(13 experts), and no repeated solicitation was undertaken
to include others because of strong correlation among
these respondents.

Data Collection
The survey asked experts to rate the healthfulness of
each of the 205 sample foods/beverages when consumed
in the recommended serving size. The item name was
displayed along with a picture of the item and a typical
Nutrition Facts label. The label listed serving size, serv-
ings per container, calories per serving, calories from fat
per serving, and the amount per serving of the following
12 components: total fat, saturated fat, cholesterol, so-
dium, total carbohydrate, dietary fiber, sugars, protein,
vitamin A, vitamin C, calcium, and iron.

Experts rated each of the 205 items on an 11-point
scale from −5 (“very unhealthy”) to 5 (“very healthy”). For
each of the 13 experts surveyed, an ordinary least squares
regression was run to predict their ratings for the 205
sample foods/beverages using the 12 nutritional compo-
nents listed on a Nutrition Facts label. For components
typically shown in both absolute amount and percentage
daily value on a Nutrition Facts label, only the absolute
amounts were included to avoid redundancy. Similarly,
“calories per serving” and “calories from fat” were ex-
cluded because they can be calculated directly from fat,
carbohydrates, protein, and alcohol (which was absent
from the foods/beverages in our sample).

Data Analysis
Each expert’s linear model was used to predict ratings for
the remaining 9,393 database items for which the 12
predictor variables were available. To measure similarity
of the 13 experts’ models for healthfulness, Cronbach’s α
was computed across both their original sample ratings
and across their model predictions for other items in the
database. Cronbach’s α is a measure of inter-rater reli-
ability, and values approaching 1 suggest that raters
have similar “representations” of the underlying con-
struct (here, healthfulness).

Next, a single linear model was generated to predict
the average expert opinion about the healthfulness of a
given food/beverage, with the 12 nutritional components
on each product’s Nutrition Facts label as predictor vari-
ables. As for the individual expert models, this model was
used to predict average expert ratings for the other 9,393
foods/beverages in the database.

RESULTS AND DISCUSSION
The 13 regression models resulting from individual ex-
erts’ survey responses indicate the implicit weighting
placed on each nutritional component, accounting for a
considerable amount of variance in sample ratings (aver-
age $r^2$ of 0.48; average adjusted $r^2$ of 0.45). Indicating a
high level of similarity between experts, Cronbach’s α
was .95 across their original sample ratings and .98
across their models’ predictions for the other 9,393 items
in the database. One can infer that the variation left
unexplained by each rater’s model was caused not by a
large rating error, but rather by the exclusion of predictor
variables that affect the healthfulness of foods/beverages
similarly for all experts. This suggests that the Nutrition
Facts label might be missing important components that
experts agree affect healthfulness. Despite limiting pre-
dictor variables to those available on a Nutrition Facts
label, high levels of correlation across experts’ judgments
justify generation of a single linear model to predict the
average expert opinion about the healthfulness of a given
food/beverage.

To generate such a model, the 13 expert ratings for
each food/beverage were first averaged. Across the 205
sample items, the average rating had a mean of 0.30 and a standard deviation of 2.2 on the −5 to 5 scale. Table 1 shows results of a regression model (using robust standard errors to allow for heteroscedasticity) to predict these average expert ratings for a food/beverage based on the 12 nutritional predictors. To summarize, the predicted average rating that experts would give to a food/beverage based on its nutritional components (to three significant digits) is:

\[
0.710 - 0.0538 \times \text{total fat} - 0.423 \times \text{saturated fat} - 0.00398 \times \text{cholesterol} - 0.00254 \times \text{sodium} - 0.0300 \times \text{total carbohydrate} + 0.561 \times \text{fiber} - 0.0245 \times \text{sugar} + 0.123 \times \text{protein} + 0.00562 \times \text{vitamin A} + 0.0137 \times \text{vitamin C} + 0.0685 \times \text{calcium} - 0.0186 \times \text{iron}
\]

Measurement units are specified in Table 1.

The model’s \( r^2 \) of 0.626 suggests that it captures almost two thirds of the variance in experts’ average ratings of foods/beverages, justifying use of the model to predict the average rating that would be given by experts to the other 9,393 foods/beverages in our database. The 10 highest and 10 lowest average predictions across items within a product category are shown in Table 2. To illustrate the usefulness of comparison within a single category, the 10 highest and 10 lowest predictions for items classified under “All Other Salty Snacks” are shown in Table 3.

**Table 1.** Model for predicting food/beverage healthiness based on 12 nutritional variables, derived by regressing average expert healthiness ratings (on an 11-point scale from −5 = “very unhealthy” to 5 = “very healthy”) on these 12 predictor variables for 205 sample foods/beverages

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.710*** (0.207)</td>
</tr>
<tr>
<td>Total fat (g)</td>
<td>−0.0538 (0.0414)</td>
</tr>
<tr>
<td>Saturated fat (g)</td>
<td>−0.423*** (0.0944)</td>
</tr>
<tr>
<td>Cholesterol (mg)</td>
<td>−0.00398 (0.00330)</td>
</tr>
<tr>
<td>Sodium (mg)</td>
<td>−0.00254*** (0.000445)</td>
</tr>
<tr>
<td>Total carbohydrate (g)</td>
<td>−0.0390 (0.0110)</td>
</tr>
<tr>
<td>Fiber (g)</td>
<td>0.561*** (0.109)</td>
</tr>
<tr>
<td>Sugar (g)</td>
<td>0.0245 (0.1090)</td>
</tr>
<tr>
<td>Protein (g)</td>
<td>0.123*** (0.0222)</td>
</tr>
<tr>
<td>Vitamin A (%DV)*</td>
<td>0.00562* (0.00234)</td>
</tr>
<tr>
<td>Vitamin C (%DV)</td>
<td>0.0137*** (0.00399)</td>
</tr>
<tr>
<td>Calcium (%DV)</td>
<td>0.0685*** (0.0137)</td>
</tr>
<tr>
<td>Iron (%DV)</td>
<td>−0.0186 (0.0186)</td>
</tr>
</tbody>
</table>

*Percentage daily value.

**Table 2.** Average model predictions (more negative numbers indicating greater unhealthiness and more positive numbers indicating greater healthiness) across items within product categories having the 10 highest average predictions and the 10 lowest average predictions

<table>
<thead>
<tr>
<th>Category name</th>
<th>Average predicted rating within category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dried beans (generic)</td>
<td>7.87</td>
</tr>
<tr>
<td>Natural supplements</td>
<td>7.86</td>
</tr>
<tr>
<td>Citrus (fresh)</td>
<td>3.68</td>
</tr>
<tr>
<td>Instant breakfast</td>
<td>3.67</td>
</tr>
<tr>
<td>Nutritional foods/beverage</td>
<td>3.37</td>
</tr>
<tr>
<td>Skim milk</td>
<td>3.35</td>
</tr>
<tr>
<td>Diet aids</td>
<td>3.34</td>
</tr>
<tr>
<td>Spinach (fresh)</td>
<td>3.26</td>
</tr>
<tr>
<td>Organic fruits (fresh)</td>
<td>3.26</td>
</tr>
<tr>
<td>Berries (fresh)</td>
<td>3.17</td>
</tr>
<tr>
<td>Beef (frozen)</td>
<td>−2.26</td>
</tr>
<tr>
<td>Cakes (fresh)</td>
<td>−2.29</td>
</tr>
<tr>
<td>Candy chocolate</td>
<td>−2.40</td>
</tr>
<tr>
<td>All other frozen breakfast</td>
<td>−2.46</td>
</tr>
<tr>
<td>Butter</td>
<td>−2.56</td>
</tr>
<tr>
<td>Sausage (fresh)</td>
<td>−2.70</td>
</tr>
<tr>
<td>Hot dogs/sausage/brats</td>
<td>−2.84</td>
</tr>
<tr>
<td>Pies (fresh)</td>
<td>−2.86</td>
</tr>
<tr>
<td>All other fresh bakery</td>
<td>−3.53</td>
</tr>
<tr>
<td>Premade lunch packs</td>
<td>−3.65</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

Although the valence of impact that most nutrients have on the healthfulness of a food/beverage might be common knowledge even to lay consumers, the contribution of the model outlined in Table 1 is the assignment of a magnitude weighting to each nutritional component. This allows for isolation of each component’s separate impact without compromising the ability to summarize their combined impact in a single metric. Indeed, the model demonstrates that some nutritional components have substantial positive effects on a food’s healthfulness, while others have substantial negative effects, implying that previous models focusing solely on one or the other have omitted critical information that experts take into account. While there was a necessary tradeoff between the explanatory power of the present model and its parsimony in predictor variables, it arguably encompasses the most important inputs to the professional judgments of nutrition experts.

This metric for rating healthfulness meets many of the criteria described in the literature, yet widely lacking from prior research. First, the present approach did not require experts to explicitly assign numerical valuations to different nutrients because their ratings for a broad sample of foods/beverages captured implicit judgments about the impact of different nutrients. Second, the decision to generate a model of healthfulness based on average expert ratings was validated by a high level of agreement across individual expert models, and the resulting predictions for average expert rating can be used to compare nutritional values of foods/beverages either across or within product categories. Lastly, the model’s predictions fall along a single linear spectrum, and can be easily transformed to any continuous distribution or discrete categorization that is deemed optimal for effectively conveying information to consumers in a particular context.

There are several possible applications for the model. Similar to the work of Scarborough and colleagues (23),
its predicted ratings could be correlated with those of lay consumers or those produced by other metrics to determine whether the latter reflect the knowledge of experts in their nutrient weightings. More importantly, the model can be used to generate healthfulness ratings that are displayed alongside food/beverage labels, allowing consumers to make more informed choices about which products to purchase and consume. One limitation, though, is the model’s inability to determine combinations of foods/beverages that comprise a well-balanced overall diet because it rates each item in isolation. It can prove most useful to consumers choosing between similar items within the same category. To this end, future studies will test the extent to which outputs of the model help consumers make decisions more closely aligned with recommendations of nutrition experts.

STATEMENT OF POTENTIAL CONFLICT OF INTEREST: No potential conflict of interest was reported by the authors.

FUNDING/SUPPORT: All expenses of this project (expert payment) were funded through the Harvard Business School.

ACKNOWLEDGEMENTS: All authors contributed substantially to the development of the study objective and design, and provided feedback throughout the writing and revision process.

References