Catalogue no. 11-522-XIE

Statistics Canada International Symposium Series - Proceedings

Symposium 2006 : Methodological Issues in Measuring Population Health

2006

Statistics Canada Statistique Canada

Canada

Proceedings of Statistics Canada's Symposium 2006
Methodological Issues in Measuring Population Health

Studying Scenarios of Nutrition Intervention: The Example of Soft Drinks

Didier Garriguet[1]

Abstract

Using data from the Canadian Community Health Survey, it is possible to estimate the distributions of usual nutrient intake. It is more difficult to estimate the usual consumption of specific food items. Consumption has to be estimated by combining the food item's consumption frequency with the distribution of consumers' usual intake of the food item. It may be difficult to estimate that distribution for less common food items, and it is virtually impossible to obtain reliable estimates of the food item's consumption frequency with only two days of data per respondent. Using an outside source or a parametric assumption may help to overcome this problem. One solution is to use an indirect approach to estimate a food item's impact on the distribution of a nutrient's usual intake by eliminating that food item or partly or completely replacing it with another food item.

KEYWORDS: Usual intake; daily intake; nutrition; food; nutrients.

1. Introduction

One of the objectives of the Canadian Community Health Survey (CCHS) – Nutrition (Cycle 2.2) was to measure the usual intake distributions of nutrients, food items and food groups using one or two 24-hour recalls per respondent. In a 24-hour recall, the respondent lists everything he/she ate and drank the day before the interview. It is fairly easy to calculate usual intake distributions using a repeated measurement model for variables reported every day, such as nutrient intake. It is much more difficult to try to estimate distributions for variables reported on a more irregular basis, such as specific food items. The problem is compounded for specific food items when the number of 24-hour recalls is small, as is the case for CCHS 2.2. How, then, can we go about measuring a food item's impact?

In this article, we will attempt to provide a partial answer, first by focusing on the differences between a distribution for a food item and a distribution for a nutrient. Then we will propose two different types of solutions, one that looks directly at the estimate for the food item and one that indirectly examines the food item's impact on a nutrient's usual intake in the population. The latter method we will call an intervention scenario, because it can be used to measure the potential effects of a policy or a new recommendation.

2. Usual Intake

2.1 Data source

The nutrition component of CCHS 2.2 was constructed to estimate the usual nutrient intake distribution in the population. To gather data, the survey used a 24-hour dietary recall; in other words, respondents were asked to report everything they ate and drank from midnight to midnight on the day before the interview. To maximize recall accuracy, the survey used an automated multiple-pass method developed in the United States (Moshfegh *et al*., 1999; Moshfegh *et al*., 2001). The list of foods and beverages reported is then converted into nutrients using the Canadian Nutrient File (Health Canada, 2005).

[1]Didier Garriguet, Statistics Canada, 150 Tunney's Pasture Driveway, Ottawa, ON, Canada, K1A 0T6, (didier.garriguet@statcan.ca)

Of course, an individual's consumption for one day does not reflect his or her usual consumption. The within-individual variation for a nutrient is often greater than the between-individuals variation. The variance calculated from daily consumption figures represents the total variance, i.e., the sum of the two sources of variation. For example, to estimate an individual's average usual energy intake with a 10% margin of error, we would need 31 days' worth of data; for vitamin A, 433 days would be needed (Basiotis, Welsh, Cronin, Kelsay and Mertz, 1987). Obviously, in a national population survey, obtaining such a large number of recalls would be impossible for budgetary and response-burden reasons. Instead, we collect a second 24-hour recall for a subset of the original sample. This second recall, collected three to ten days after the first recall to minimize correlation between the days, is used primarily to estimate the individual's contribution to the total variance and to adjust the one-day nutrient intake distribution to produce the usual distribution. This will be the nutrient's usual intake distribution for the population, and not for a particular individual.

To obtain the usual intake distribution, it is essential to have an estimate of the within-individual variance. Ideally, that estimate is generated with information from the first and second recalls. One problem is that if a particular food item is not consumed frequently, there will be very few recalls or even none at all for the item. It will then be impossible to generate the usual intake distribution for that food item.

2.2 Consumers' usual intake of a food item

If the usual intake distribution cannot be determined for a seldom-consumed food item, what can we get from the survey data? First, we can obtain the average consumption of a food item by the entire population either by including those who do not consume it or by excluding them (consumers only). The techniques used to produce the usual intake distribution of a nutrient or a food item assume that the average daily intake for the first recall and the average usual intake are equal. Consequently, the average usual intake is the same as the average intake using only the first 24-hour recall.

What about frequency of consumption? When there are not enough second 24-hour recalls, the only estimate that can be made from the first recall is the probability of consumption of a food item the day before the interview. This represents only the situation on one day, and cannot be interpreted as reflecting the frequency of consumption in the population. In other words, if 50% of the population consumes a food item the day before the interview, it does not mean that 50% of the population consumes that item every day. Some people may consume it every day, others once a week, and in the extreme case, still others may consume it only that one time during the year.

Having many second recalls may allow us to estimate a food item's within-individual variation among consumers, even if there is a low proportion of consumers in the population. However, that information will not be enough to enable us to estimate the item's usual consumption. To obtain a food item's usual intake distribution in the entire population, we need not only the usual intake of those who consume it, but also the frequency with which they consume it, i.e., the number of days they consume it per week or month.

2.3 Consumption frequency of a food item

The question, then, is as follows: How do we obtain a food item's consumption frequency? The simplest way is to ask a question like "How often do you consume food item X?" or "How many days a week (or month) do you consume food item X?" In a population survey such as the CCHS, the number of different food items reported is very large, greater than 8,000. Even the number of highly detailed categories of food items is in the hundreds. Therefore, it is impossible to obtain an individual consumption frequency for each item. The CCHS contains questions about the consumption frequency of salt and selected fruits and vegetables. Consumption frequency is not covered for most food items.

Is it possible to estimate consumption frequency with dietary recalls? In the previous section, we discussed the case in which there are not enough second recalls, and we concluded that only the consumption frequency for the day before the interview can be determined.

Now let's suppose that we have enough second recalls to obtain a usual intake distribution for the food item among those who consume it. The usual intake distribution is often produced with the SIDE or C-SIDE software application (Novenario, 1996; Dodd, 1996). Unlike SIDE, C-SIDE can do more than generate a nutrient's usual intake for the entire population or just for those who consume it. It also has a module that produces that consumption frequency. That module attempts to estimate the probability of consumption on l out of k days, taking into account the fact that we have a mix of binomials (k, p_m) with mix parameter $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_m)$. In practice, we model the population by dividing it into M+1 groups that have a probability {0, p1, p2, …, p_M} of consuming the item on a given day and represent $\{\theta_0, \ldots, \theta_M\}$ x 100% of the population respectively.

Using Dodd's notation (1996), we can write

$$\hat{\psi}_l = \left(\sum_{i=1}^{n} W_i\right)^{-1} \sum_{i=1}^{n} W_i I\left(\hat{\pi}_i = \frac{l}{k}\right) \tag{2.3.1}$$

where W_i, is the sampling weight of individual i, $\hat{\pi}_i = k^{-1}\sum_{j=1}^{k}\delta_{ij}$ is the proportion of consumption days of individual i, and δ_{ij} is an indicator variable that takes a value of 1 if the i^{th} individual consumes the food item on day j.
The expected value of the probability of consumption is

$$\psi_l(\theta) = \sum_{m\in E_l} \theta_m \binom{k}{l} p_m^l (1-p_m)^{k-l} \tag{2.3.2}$$

where $E_l = \{0,1,\ldots, M-1\}$ if $l < k$ or $\{1, 2, \ldots, M\}$ if $l = k$.

We then look for the solution of $\boldsymbol{\theta}$ that will minimize the modified minimum chi-square estimator

$$n\sum_{l=0}^{k}\left[\hat{\psi}_l - \psi_l(\theta)\right]^2 \tilde{\psi}_l^{-1} + \sum_{m-1}^{M} \frac{\theta_m}{1-\theta_0} \ln\left(\frac{\theta_m}{1-\theta_0}\right) \tag{2.3.3}$$

subject to the constraints $\sum_{m=0}^{M}\theta_m = 1$, $\theta_m \in [0,1]$, where $\theta \ln\theta = 0$ for $\theta = 0$,

The value of $\tilde{\psi}_l$ will be equal to $\max\{\hat{\psi}_0, (1-\overline{\psi})^k\}$ if l = 0, $\max\{\hat{\psi}_k, \overline{\psi}^k\}$ if l = k where $\hat{\psi}_l(1-\tilde{\psi}_0-\tilde{\psi}_k)(1-\hat{\psi}_0-\hat{\psi}_k)^{-1}$ for the other values of l between 1 and k-1. Finally,

$$\overline{\psi}_l = \left(k\sum_{i=1}^{n} W_i\right)^{-1} \sum_{i=1}^{n} W_i \sum_{j=1}^{k} \delta_{ij}\,. \tag{2.3.4}$$

In concrete terms, we are looking for the $\boldsymbol{\theta}$ that minimizes the distance between a distribution based on three consumption probabilities – 0%, 50% (l=1 day over k=2) and 100% (l=1 day over k=1 or l=2 days over k=2) – and a distribution that yields an identical proportion of the population for each consumption probability. For example, to obtain the weekly consumption under this part of the solution, each case – not consuming, consuming on one day out of seven, consuming on two days out of seven, and so on, up to consuming every day – would have an equal chance (1/8) of occurring. By default, C-SIDE tries to obtain a distribution for M=50 days.

Empirical results show that this equation's optimum solution with only two 24-hour recalls is very unstable. Furthermore, C-SIDE was developed using four 24-hour recalls, so that the distribution is based on five points instead of just three. In short, there is reason to doubt the estimate's accuracy when only two dietary recalls are used.

What options do we have left? In an ideal world, an outside source would provide answers to questions about consumption frequency. In the worst case, we may be able to make parametric assumptions about the consumption frequency distribution. Ideally, that assumption will take into account the consumption frequency on the day before the interview and, if possible, information from outside sources. The parametric assumption will undoubtedly vary from one food item to another, or even for the same item for different age groups. An example will be discussed in the next section.

2.4 Example: Usual consumption of soft drinks

To illustrate the theory described in the preceding sections, we will look at the example of ordinary soft drink consumption by adolescents aged 14 to 18. That is the age group which consumes the most ordinary soft drinks; 52.6% of its members drank ordinary soft drinks the day before the interview. Consumers in this group drank an average of 718 grams of soft drinks a day, the equivalent of nearly two drink cans. For the age group as a whole, average daily consumption is 378 grams.

We are able to use SIDE to estimate the food item's usual intake distribution, since there are enough second recalls from consumers. Now we have to determine the consumption frequency. The CCHS has no questions on soft drink consumption frequency. Fortunately, we have an external data source to help us estimate teenagers' consumption frequency.

Table 1
Frequency of consumption of ordinary soft drinks, 15-year-olds, Canada, 2002

Consumption frequency	Percentage of 15-year-olds	Total percentage for the categories
Never	4	21
Less than once a week	8	
Once a week	9	
2 to 4 times a week	23	23
5 to 6 times a week	19	19
Once a day, every day	16	37
Every day, more than once	21	

Source: CANSIM Table 110-0064, Public Health Agency of Canada

Is there a parametric distribution that will yield a similar distribution? After several trials, we selected the exponential distribution in λ parameters, where $1/\lambda$ is equal to the probability of not consuming any soft drinks the day before the interview multiplied by the number of days for which we want to know the distribution. Using a probability of not consuming of 47.4%, we obtain $\lambda = 0.301$ for one week and $\lambda = 0.070$ for one month (30 days). Summing the monthly results to produce a table similar to Table 1, we obtain the results shown in Table 2.

Table 2
Actual and predicted frequency of consumption of ordinary soft drinks, 15-year-olds, Canada, 2002

Consumption frequency	Actual percentage	Predicted percentage
Once a week or less	21	18.5
2 to 4 times a week	23	24.5
5 to 6 times a week	19	18.1
Every day	37	38.9

Source: CANSIM Table 110-0064, Public Health Agency of Canada

2.5 Probability of consumption above a specified limit

With the usual distribution of a food item among consumers and the consumption frequency, we can determine the proportion of the population that consumes more or less than a specified quantity of the item. For nutrients, the limits are set on the basis of the health risks and the level of knowledge and research concerning the nutrient in question. In the case of food items, the limits are not necessarily linked to observations that support a conclusion that the proportion of the population above or below a particular threshold is at risk.

To determine the proportion of the population that consumes more than a particular quantity, we multiply the probability of consuming the food item by the proportion of the population that consumes above the threshold, as shown in the equation

$$\sum_{j=1}^{M} P\left[I = j\right] * P\left[c \geq \tfrac{S*M}{j}\right] \qquad (2.5.1)$$

where M is the number of days for the consumption frequency, c is the consumption, and S is the maximum level for one day. Note that the probability of consuming the food item is assumed to be independent of the probability of consuming more than a particular quantity. C-SIDE tests that assumption.

If we replace M with 30 and S with an arbitrary threshold of 1 kg (about 1 litre) a day, and use the results we obtained earlier for the usual distributions and the consumption frequencies of ordinary soft drinks, we find that the proportion of adolescents aged 14 to 18 that consumes more than a litre of ordinary soft drinks a day is 4.3%.

3. Intervention Scenarios

3.1 Eliminating a food item

A limitation of the above method is that we do not necessarily know the health risks associated with consuming too much or too little of a particular food item. How, then, can we assess the impact that a particular food item has on the population's dietary intake?

We can provide an initial answer by measuring the average intake for a nutrient from the food item concerned. For example, instead of measuring the average quantity consumed in grams, we measured the amount of energy associated with consumption of that food item or the amount of sugar associated with it. Dividing the average consumption of a nutrient for that food item by the average total intake of the nutrient for an individual would give us an estimate of the item's importance in the population's diet.

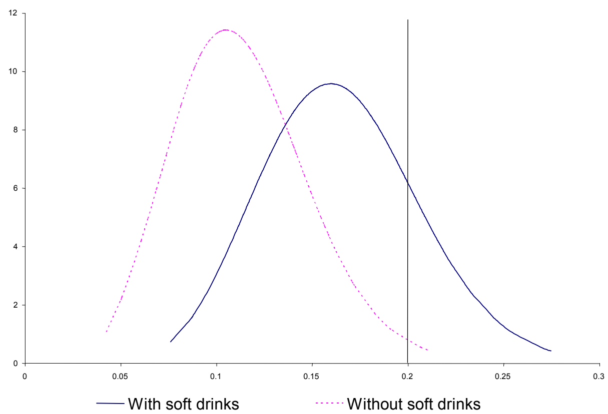
For example, in the population aged 14 to 18, ordinary soft drinks provide 30.2% of the daily sugar intake, 40.4% of the daily caffeine intake, and 5.4% of the daily energy intake. In the United States, the National Health and Nutrition Examination Survey (NHANES) makes it possible to compare estimates over time and identify trends. The CCHS is the first nutrition survey in more than 30 years, and there are no recent Canadian data to compare the CCHS's estimates with.

However, the estimate of the proportion of a nutrient's consumption that is due to a particular food item does not help us estimate the proportion of the population that consumes excessive amounts of a nutrient because of that food item. To estimate that proportion, we propose to estimate the distribution of the nutrient's usual intake with and without the food item in question. Of course, it can only be an approximation of the item's impact, since the selection of food items is necessarily correlated. In other words, one does not necessarily drink the same beverage with a salad as with a hamburger. It is also clear that eliminating a food item does not reflect its real impact, because in many cases, another item is consumed in its place. In the soft drink example, however, eliminating the food item may actually mean replacing it with a similar quantity of water, which does not contribute to the intake of most nutrients.

3.2 Example: Impact of eliminating soft drinks

Returning to the example of ordinary soft drink consumption in the population aged 14 to 18, we previously mentioned that soft drinks contributed 30.2% of the daily sugar intake. In Figure 1, we estimate the daily sugar intake distribution for the original population, and the distribution of the daily sugar intake of the same population after the quantity of sugar derived from ordinary soft drinks has been removed for each recall of each individual. The effect on the daily sugar intake distribution is noticeable not only in the estimate of the average distribution but also in the estimate of the between-individuals variance. Initially, 19.5% of the population aged 14 to 18 derived more than 20% of their daily energy from sugars. After ordinary soft drinks are eliminated from their diet, the proportion falls to 1.5%.

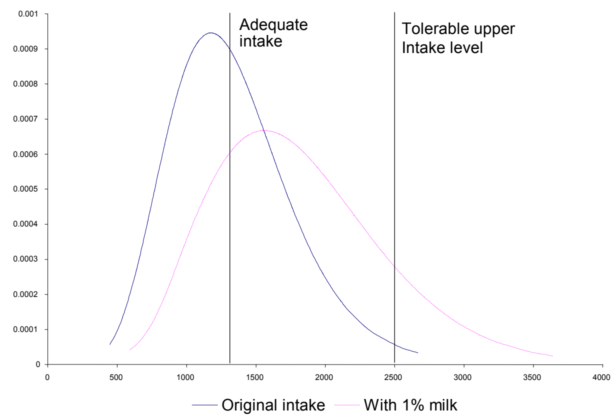
Figure 1
Usual caloric intake from sugars with and without ordinary soft drinks, household population, boys aged 14 to 18, Canada excluding territories, 2004



3.3 Replacing a food item

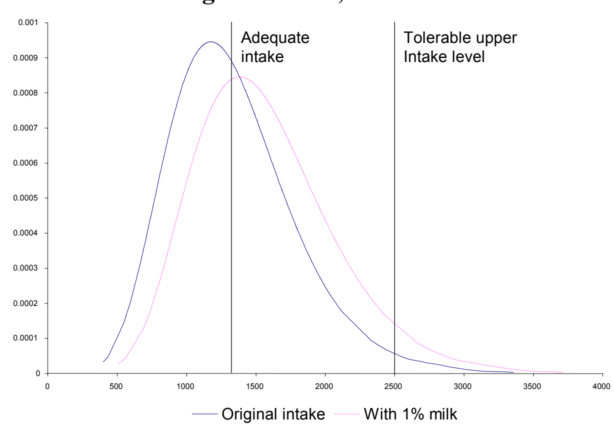
The usual intake distributions with and without a particular food item can also be used to determine the impact of a recommendation or a potential change in a population's eating habits. For example, since soft drink consumption has sometimes been associated with lower calcium intake due to reduced milk consumption, we might want to measure the effect that replacing ordinary soft drinks with an equivalent quantity of 1% milk would have on the usual calcium intake distribution. This scenario is illustrated in Figure 2. Initially, 52% of the population was below the adequate intake level of 1,300 mg per day and 2% was above the tolerable upper intake level of 2,500 mg per day. It should be noted that while the people above the tolerable upper intake level are running a higher risk of negative health effects, the adequate intake level is a recommendation based on observations of healthy people. The proportion of the population below that level is not the proportion of the population that has inadequate intake of the nutrient. When 1% milk is substituted for soft drinks for the entire population, the proportion of the population below the adequate intake level falls to 22%, while the proportion above the level rises to 14%.

Figure 2
Usual calcium intake when ordinary soft drinks are or are not replaced with 1% milk, household population, boys aged 14 to 18, Canada excluding territories, 2004



An intervention can have more realistic goals, however. We do not expect the entire population to substitute milk for ordinary soft drinks, but we are interested in determining the impact if 50% of the population changed its habits. To do so, we can conduct a Monte Carlo experiment with random samples representing 50% of the population for which consumption of ordinary soft drinks is replaced with consumption of 1% milk. If we repeat the experiment 100 times and take the average distribution, we find that 30% of the population is below the adequate calcium intake level and 5% are above the tolerable upper intake level (Figure 3).

Figure 3
Usual calcium intake when ordinary soft drinks are or are not replaced with 1% milk in 50% of the cases, household population, boys aged 14 to 18, Canada excluding territories, 2004



4. Conclusion

It is far from easy to measure the impact that a particular food item has on eating habits. When the impact of consuming specific quantities is known or specific quantities are being studied, the food item's usual consumption must be determined directly. That consumption has two parts: the food item's usual intake distribution among those who consume it, and the food item's consumption frequency. The first part can be determined from two 24-hour recalls if there are enough consumers over two days. The second part is difficult to estimate with only two 24-hour recalls. C-SIDE can estimate the frequency, but the estimates produced with two recalls are unstable, and their quality is suspect. The ideal solution is to obtain one week's or one month's consumption frequency from the survey or an outside source. Parametric assumptions can also be made about the consumption frequency distribution.

In many cases, it is more useful to indirectly examine the impact that the food item has on the usual intake distribution of one or more nutrients. Historical data may be sufficient for estimating averages, but the nutrient's usual intake distributions with and without the food item in question can also be estimated. The same approach can be used to test intervention scenarios or the effect of dietary recommendations: replace the food item being studied partially or completely with another item and examine the resulting differences in the nutrient's usual intake distributions.

References

Basiotis, P. P., Welsh, S. O., Cronin, F. J., Kelsay, J. L., and Mertz, W. (1987), "Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence", *Journal of Nutrition*, 117, pp. 1638-1641.

Dodd, K. W. (1996), *Technical Guide to C-SIDE (Software for Intake Distribution Estimation), Version 1.0, Dietary Assessment Research Series Report 9, A, September 1996*, disponible à l'adresse : http://www.card.iastate.edu/publications/DBS/PDFFiles/96tr32.pdf.

Moshfegh , A. J. , Borrud, L., Perloff, B. and coll. (1999), "Improved method for the 24-hour dietary recall for use in national surveys", *The FASEB Journal: official publication of the Federation of American Societies for Experimental Biology,* vol. 13, p. A603 (résumé).

Moshfegh, A. J., Raper, N., Ingwersen, L. and coll. (2001), "An improved approach to 24-hour dietary recall methodology", *Annals of Nutrition and Metabolism,* vol. 45 (supplement), p. 156 (summary).

Novenario, M. J. (1996), *User's Guide to SIDE, A, August 1996,* available at the website: http://www.card.iastate.edu/publications/DBS/PDFFiles/96tr32.pdf.

Santé Canada (2005), *Canadian Nutrient File, 2007b,* available at the site: http://www.hc-sc.gc.ca/fn-an/nutrition/fiche-nutri-data/index_e.html