Adjusting for nonresponse bias corrects overestimates of food allergy prevalence

Lianne Soller, BSc, MSc, PhD candidate[a],
Moshe Ben-Shoshan, MD, MSc[b],
Daniel W. Harrington, MA, PhD[c], Megan Knoll, MSc[a],
Joseph Fragapane, BEng, MD[a], Lawrence Joseph, PhD[a,d],
Yvan St. Pierre, MSc[a], Sebastien La Vieille, MD[e],
Kathi Wilson, PhD[f], Susan J. Elliott, PhD[g], and
Ann E. Clarke, MD, MSc[a,h]

Clinical Implications

- We are the first to demonstrate that adjustment for nonresponse can lead to important changes in food allergy prevalence. Clinicians must be cautious when interpreting the literature because most authors do not account for nonresponse.

TO THE EDITOR:

Nationwide estimates of food allergy prevalence are frequently based on telephone surveys, as this allows population-based sampling from geographically diverse regions. The most recent telephone surveys from the United States and Canada estimate that the prevalence of self-reported food allergy ranges between 8.1% and 9.1%.[1,2] However, such studies are often limited as they provide prevalence estimates for a limited number of allergies[3,4] and do not consider nonresponse bias,[1-6] which may result in an overrepresentation of certain demographic groups who may tend to report more allergies.

Given these limitations, we used data collected in the Canadian population-based SPAACE (Surveying Prevalence of food Allergy in All Canadian Environments) study, which inquired about allergies to several foods and obtained information from households who refused or could not be reached to complete the study. This allowed us to: (1) provide population-weighted prevalence estimates of allergy to any food and (2) explore the influence of nonresponse bias on prevalence by presenting a range of estimates using different assumptions about food allergy prevalence among nonresponders.

METHODS

Survey methodology

The SPAACE study was a random cross-Canada telephone survey conducted between September 2010 and 2011, which targeted vulnerable Canadians (ie, those of low income, New Canadians, and of self-reported Aboriginal identity) using 2006 Canadian Census data (refer to Supplement E1 in this article's Online Repository at www.jaci-inpractice.org).[7,8] Households were telephoned and the initial adult respondent was queried using the Food Allergy Prevalence Questionnaire (FAPQ) on whether any household member had an allergy to peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and/or soy, or other foods.[7] Food allergy was defined as follows:

TABLE I. Weighted perceived and probable prevalence estimates of food allergy by age group

	Children under 18, % (95% CrI) (n = 4026)	Adults 18 and over, % (95% CrI) (n = 10,996)	All ages, % (95% CrI) (n = 15,022)
Perceived			
Peanut	2.4 (1.6, 3.2)	0.7 (0.5, 0.9)	1.1 (0.9, 1.3)
Tree nut	1.6 (1.0, 2.3)	1.2 (0.9, 1.5)	1.3 (1.0, 1.6)
Fish	1.0 (0.3, 1.8)	0.6 (0.4, 0.8)	0.7 (0.5, 0.9)
Shellfish	1.4 (0.6, 2.1)	1.9 (1.5, 2.2)	1.7 (1.4, 2.0)
Sesame	0.1 (0.0, 0.3)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)
Milk	0.7 (0.3, 1.1)	0.7 (0.5, 0.9)	0.7 (0.5, 0.9)
Egg	1.0 (0.6, 1.5)	0.5 (0.3, 0.7)	0.6 (0.4, 0.8)
Wheat	0.3 (0.0, 0.6)	0.4 (0.2, 0.6)	0.4 (0.2, 0.5)
Soy	0.1 (0.0, 0.3)	0.1 (0.0, 0.2)	0.1 (0.1, 0.2)
Other	2.2 (1.5, 3.0)	3.5 (3.0, 4.0)	3.2 (2.8, 3.6)
Any	6.9 (5.5, 8.2)	7.7 (6.9, 8.4)	7.5 (6.9, 8.1)
Probable*			
Peanut	2.2 (1.4, 2.9)	0.6 (0.4, 0.8)	1.0 (0.7, 1.2)
Tree nut	1.5 (0.9, 2.1)	1.0 (0.8, 1.3)	1.2 (0.9, 1.4)
Fish	0.9 (0.3, 1.6)	0.5 (0.3, 0.7)	0.6 (0.4, 0.8)
Shellfish	0.8 (0.4, 1.2)	1.6 (1.3, 2.0)	1.4 (1.2, 1.7)
Sesame	0.1 (0.0, 0.3)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)
Milk	0.2 (0.0, 0.3)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)
Egg	1.0 (0.5, 1.5)	0.5 (0.3, 0.6)	0.6 (0.4, 0.8)
Wheat	0.2 (0.0, 0.5)	0.2 (0.1, 0.4)	0.2 (0.1, 0.4)
Soy	0.1 (0.0, 0.3)	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)

*We collected only detailed information about food allergy to the 9 common foods; therefore, probable estimates for other foods and any food could not be calculated.

(1) *Perceived*: individuals self-reporting any food allergy, and
(2) *Probable*: individuals self-reporting a convincing history[9,10] and/or a physician diagnosis of allergy to peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and/or soy.

If the respondent refused to complete the FAPQ, the interviewer administered a much briefer Refusal Questionnaire (RQ) that queried if any household member had an allergy and if present, data on the household size, the respondent's education, the food(s) to which the individual was allergic, and whether the allergy was diagnosed by a doctor were collected.

Developing weighted estimates of prevalence

Point estimates and 95% credible intervals (CrIs) for the prevalence of perceived and probable allergy were weighted to account for the oversampling of vulnerable populations (refer to Supplement E2 in this article's Online Repository at www.jaci-inpractice.org).[7] Credible intervals are the Bayesian analogue to standard confidence intervals.

Developing nonresponse bias estimates

To develop nonresponse bias-adjusted estimates of prevalence of perceived allergy to any food, 4 groups were identified:

(1) *Full Participants*: households who completed the FAPQ,
(2) *Refusal Questionnaire Participants*: households who completed the RQ only,

TABLE II. Nonadjusted and bias-adjusted prevalence estimates of perceived allergy to any food

	Nonadjusted		Bias-adjusted			
Estimate number	Full participants (FP), % (95% CrI) (n = 15,022)	Refusal questionnaire participants (RQP), % (95% CrI) (n = 1393*)	Nonparticipants (NP), % (95% CrI) (n = 17,059*)		Never reached participants (NRP), % (95% CrI) (n = 8419*)	All participants, % (95% CrI) (n = 41,893)
					NRP same as NP	
1	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP half RQP	1.0 (0.7, 1.4)	1.1 (0.7, 1.5)	3.0 (2.8, 3.3)
2	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP same as RQP	2.1 (1.4, 2.8)	2.1 (1.5, 2.9)	3.7 (3.2, 4.2)
3	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP twice RQP	4.2 (2.8, 5.7)	4.3 (2.9, 5.9)	4.9 (4.1, 5.9)
					NRP mixture of FP, RQP, and NP	
4	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP half RQP	1.0 (0.7, 1.4)	3.5 (3.2, 3.8)	3.5 (3.2, 3.8)
5	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP same as RQP	2.1 (1.4, 2.8)	4.0 (3.6, 4.5)	4.0 (3.6, 4.5)
6	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP twice RQP	4.2 (2.9, 5.7)	5.1 (4.4, 6.0)	5.1 (4.4, 5.9)
					NRP same as FP	
7	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP half RQP	1.0 (0.7, 1.4)	6.4 (6.0, 6.9)	4.1 (3.8, 4.4)
8	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP same as RQP	2.1 (1.4, 2.8)	6.4 (6.0, 6.9)	4.5 (4.2, 4.9)
9	6.4 (6.0, 6.8)	2.1 (1.4, 2.9)	NP twice RQP	4.2 (2.8, 5.7)	6.4 (6.0, 6.9)	5.4 (4.8, 6.1)

*The number of people in all nonallergic households in the RQP group, and in all households in the NP and NRP groups, was imputed using the distribution of the number of people in each household in the FP group.

(3) *Nonparticipants*: households who were reached by telephone but refused to complete either questionnaire, and
(4) *Never Reached Participants*: households who could not be reached by telephone.

Food allergy data were available only from *Full* and *RQ Participants*. Multiple imputation (MI), the gold standard for adjusting for missing data,[11] was used to adjust the estimates for nonresponse bias that resulted from missing food allergy data within the *Nonparticipants* and the *Never Reached Participants* by using a model that included observed data (census tract [CT] and province of residence) to predict the missing data on the probability of food allergy.[12]

A range of assumptions regarding the prevalence of food allergy in the *Nonparticipants* and *Never Reached Participants* were investigated (refer to Supplement E3 in this article's Online Repository at www.jaci-inpractice.org). Compared with the prevalence in the *RQ Participants* living in the same CT, the prevalence in the *Nonparticipants* was assumed to be: (1) half, (2) equal to, and (3) twice as large as the *RQ Participants*.

Compared with the prevalence of those in the same CT, the prevalence among the *Never Reached Participants* was assumed to be: (1) equal to the *Nonparticipants*; (2) a weighted average of the *Full*, *RQ*, and *Nonparticipants*; and (3) equal to the *Full Participants*.

MI was implemented via a hierarchical logistic regression model with 4 levels: individual, household, CT, and province of residence. Weighting to account for the overrepresentation of vulnerable populations could not be done in this analysis because demographic information was only available for *Full Participants*. The analyses were performed using WinBUGS (version 1.4.3, MRC Biostatistics Unit, Cambridge, United Kingdom) (refer to Supplement E3 in this article's Online Repository at www.jaci-inpractice.org).

RESULTS

Participation rate

We telephoned 17,337 households, 14,113 of whom were actually reached. Of these 14,113 households, 1351 were ineligible due to a language barrier or unavailability of an adult. Of the 12,762 eligible households, 5734 households, representing 15,022 individuals, completed the FAPQ (45% response rate, or 5734 of 12,762) and were thus *Full Participants*, 524 households completed the RQ (an additional 4%, or 524 of 12,762) and were thus *RQ Participants*, and the remaining 6504 households answered the telephone but refused to provide any information (51%) and were thus *Nonparticipants*. An additional 3224 households were never reached, and were thus *Never Reached Participants*.

Prevalence estimates

Among *Full Participants*, the unweighted self-reported (perceived) prevalence of allergy to any food was 6.4% (6.0%, 6.8%). After weighting, this estimate increased to 7.5% (6.9%, 8.1%) (Table I).

Compared with the *Full Participants*, the unweighted perceived prevalence of allergy to any food was lower among the *RQ Participants* (6.4% [6.0%, 6.8%] vs 2.1% [1.4%, 2.9%]) (Table II). Applying the different assumptions regarding the prevalence of food allergy among the *Nonparticipants* and *Never Reached Participants*, 9 selection bias-adjusted estimates were obtained for the perceived prevalence of allergy to any food ranging from 3.0% (2.8%, 3.3%) to 5.4% (4.8%, 6.1%) (refer to Table II and Supplement E3 in this article's Online Repository at www.jaci-inpractice.org).

DISCUSSION

Comparison with previous studies

The unweighted perceived prevalence of food allergy in this study (6.4% [6.0%, 6.8%]) was less than that in our general population study conducted 2 years earlier (8.1% [7.5%, 8.7%]),[2] but these estimates are not directly comparable as our current study targeted vulnerable populations. The weighted perceived prevalence in the current study (7.5% [6.9%, 8.1%]) is also lower than that estimated in the NHANES study, a US population-based door-to-door survey conducted between 2007 and 2010 (9.0% [8.3%, 9.6%]).[13] The NHANES survey is weighted for nonresponse in general, but this weighting may not be sufficient to account for all possible nonresponse bias.[13] However, our weighted perceived prevalence in children (6.9% [5.5%, 8.2%]) is similar to that estimated by Gupta in a US population-based

internet survey conducted between 2009 and 2010 (8.0% [7.7%, 8.3%]).[6] Gupta's study also used weights to adjust for potential biases from sampling design and survey response.

Limitations

Although our response rate was only 45% (49% including the *RQ Participants*), other recent studies on food allergy prevalence have reported similar response rates.[1,3] In fact, research has shown that the majority of telephone surveys report response rates below 50%.[14] In addition, the information letter sent to participants before our telephone survey indicated (as required by our ethics board) that those with food allergy might need to complete a slightly longer questionnaire. It is possible, therefore, that those who participated were more likely to be allergic than those who did not. We have considered this by creating various imputation models, which assume different biases between responders and nonresponders. Finally, we had to impute the number of individuals in nonallergic households who completed the RQ because this information was not requested as we wanted to optimize the response rate by asking only a single question.

Conclusions and future directions

We are the first to consider the effect of nonresponse bias in the estimation of food allergy prevalence and have clearly demonstrated that doing so is crucial in developing accurate estimates. Despite survey response rates dropping in recent years, surveys remain an important methodology for population-based research. With low response rates, representativeness of survey participants is an important issue that must be addressed. We explored a range of assumptions for the prevalence of food allergy among *Nonparticipants* and *Never Reached Participants* and prevalence estimates ranged from 3.0% (2.8%, 3.3%) to 5.4% (4.8%, 6.1%). Given that the prevalence (unweighted) among *Full Participants* was 6.4% (6.0%, 6.8%), it is evident that nonresponse bias can substantially influence prevalence, and ignoring bias could result in an overestimation. Our research highlights the importance of minimizing nonresponse bias in designing a study, while acknowledging that bias is likely present and should be considered when performing the analysis.

[a]Division of Clinical Epidemiology, Department of Medicine, McGill University Health Center, Montreal, Quebec, Canada
[b]Division of Pediatric Allergy and Clinical Immunology, Department of Pediatrics, McGill University Health Center, Montreal, Quebec, Canada
[c]Department of Geography, Queen's University, Kingston, Ontario, Canada
[d]Department of Epidemiology and Biostatistics, McGill University, Montreal, Quebec, Canada
[e]Food Directorate, Health Canada, Ottawa, Ontario, Canada
[f]Department of Geography, University of Toronto, Toronto, Ontario, Canada
[g]Faculty of Applied Health Sciences, University of Waterloo, Waterloo, Ontario, Canada
[h]Division of Rheumatology, University of Calgary, Calgary, Alberta, Canada

The study was funded by Allergy, Genes, and Environment (AllerGen) Network of Centres of Excellence, Health Canada. The first author was supported by the Fonds de Recherche Santé-Quebec and the Research Institute of the McGill University Health Centre. Dr. Clarke holds the Arthritis Society Chair in Rheumatic Diseases at the Cumming School of Medicine, University of Calgary.
Conflicts of interest: L. K. Soller and K. Wilson have received research and travel support from AllerGen NCE. M. Ben-Shoshan has received lecture fees from Sanofi and Novartis. M. Knoll, S. J. Elliott, and A. E. Clarke have received research support from AllerGen NCE. The rest of the authors declare that they have no relevant conflicts.
Received for publication July 30, 2014; revised November 3, 2014; accepted for publication November 4, 2014.
Available online January 14, 2015.
Corresponding author: Lianne Soller, BSc, MSc, PhD candidate, Royal Victoria Hospital, 687 Pine Avenue West, V Building, Room V1.09, Montreal, Quebec, H3A 1A1, Canada. E-mail: liannesoller@gmail.com.
2213-2198
© 2015 American Academy of Allergy, Asthma & Immunology
http://dx.doi.org/10.1016/j.jaip.2014.11.006

REFERENCES

1. Vierk KA, Koehler KM, Fein SB, Street DA. Prevalence of self-reported food allergy in American adults and use of food labels. J Allergy Clin Immunol 2007;119:1504-10.
2. Soller L, Ben-Shoshan M, Harrington DW, Fragapane J, Joseph L, St Pierre Y, et al. Overall prevalence of self-reported food allergy in Canada. J Allergy Clin Immunol 2012;130:986-8.
3. Sicherer SH, Munoz-Furlong A, Godbold JH, Sampson HA. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. J Allergy Clin Immunol 2010;125:1322-6.
4. Ben-Shoshan M, Harrington DW, Soller L, Fragapane J, Joseph L, St Pierre Y, et al. A population-based study on peanut, tree nut, fish, shellfish, and sesame allergy prevalence in Canada. J Allergy Clin Immunol 2010;125:1327-35.
5. Sicherer SH, Munoz-Furlong A, Sampson HA. Prevalence of seafood allergy in the United States determined by a random telephone survey. J Allergy Clin Immunol 2004;114:159-65.
6. Gupta RS, Springston EE, Warrier MR, Smith B, Kumar R, Pongracic J, et al. The prevalence, severity, and distribution of childhood food allergy in the United States. Pediatrics 2011;128:e9-17.
7. Soller L, Ben-Shoshan M, Harrington DW, Knoll M, Fragapane J, Joseph L, et al. Prevalence and predictors of food allergy in Canada: a focus on vulnerable populations. J Allergy Clin Immunol 2014 (in press). http://dx.doi.org/10.1016/j.jaip.2014.06.009.
8. Knoll M, Soller L, Ben-Shoshan M, Harrington D, Fragapane J, Joseph L, et al. The use of incentives in vulnerable populations for a telephone survey: a randomized controlled trial. BMC Res Notes 2012;5:572.
9. Brown SG. Clinical features and severity grading of anaphylaxis. J Allergy Clin Immunol 2004;114:371-6.
10. Sicherer SH, Burks AW, Sampson HA. Clinical features of acute allergic reactions to peanut and tree nuts in children. Pediatrics 1998;102:e6.
11. Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. 2nd edition. Boca Raton, FL: Chapman and Hall; 2003.
12. Kmetic A, Joseph L, Berger C, Tenenhouse A. Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. Epidemiology 2002;13:437-44.
13. McGowan EC, Keet CA. Prevalence of self-reported food allergy in the National Health and Nutrition Examination Survey (NHANES) 2007-2010. J Allergy Clin Immunol 2013; 132:1216-9 e5.
14. Groves RM. Nonresponse rates and nonresponse bias in household surveys. Public Opin Q 2006;70:646-75.

SUPPLEMENT E1: SELECTION OF STUDY POPULATION

Canadians from low-income households, new Canadians (ie, less than 10 years living in Canada), and individuals who identify as Aboriginal were specifically targeted. Canadians with low levels of education were not targeted because it was anticipated that there would be substantial overlap between low income and low education, and by targeting low-income areas, those with low education would also be included.[1]

Adults having completed less than a postsecondary degree, trade certificate, or diploma were defined as being of low education. This group was restricted to individuals who are 18 years or older. Individuals were considered to be low income if their household income was below the low-income cutoff (LICO). The LICO is defined as an income level at which families or unattached individuals spend at least 70% of before-tax income on food, shelter, and clothing, and is determined according to family size and geographic location.[2] New Canadians were those who immigrated to Canada within 10 years of completion of the telephone survey. Individual are considered to be of Aboriginal identity if they report "Aboriginal" as their cultural background and identify with First Nations, Métis, or Inuit.

Using the 2006 Canadian census, the 100 census tracts (CTs) from within the census metropolitan areas (CMAs) that contained either the highest proportion of households living under the LICO or the highest proportion of new Canadians were selected.[3] Individuals of Aboriginal identity were selected in the same way using a lower threshold of 15%, which resulted in a total of 66 CTs included.

These CTs were then converted to postal codes using the 2006 Statistics Canada postal code conversion file. Then, InfoDirect ("White Pages" in Canada) selected a random sample of household telephone numbers with accompanying mailing addresses from these postal codes.

Due to this targeting strategy, CTs from the province of New Brunswick were not proportionately represented (only 2 CTs were included in the initial selection), and those from Nova Scotia and Newfoundland and Labrador were excluded from the initial selection because they were not among the top 100 in terms of proportion of low-income households or new Canadians, nor in the top 66 in terms of proportion of individuals of Aboriginal identity. Further, Prince Edward Island (PEI) and the 3 Canadian territories (Northwest, Yukon, and Nunavut) were excluded because they do not contain any CMAs, and hence there are no CTs.

Although our primary objective was to ensure adequate representation of the vulnerable populations, we also wanted to provide prevalence estimates involving populations from all Canadian provinces and territories. Hence, for New Brunswick, Nova Scotia, and Newfoundland and Labrador, CTs with the highest proportion of households under the LICO (range: 25.8% to 38.9% from 8 CTs in Saint John, New Brunswick; range: 24.1% to 40.9% from 10 CTs in Halifax, Nova Scotia; range: 27.4% to 41.4% from 5 CTs in St. John's, Newfoundland) were selected from the main CMAs. These areas contained too few new Canadians or individuals of Aboriginal identity to be included in the sampling for these populations. In PEI, we targeted the largest Census Subdivision in the province, Charlottetown. According to the 2006 Census, 13.2% of households in Charlottetown were below the LICO and 1.4% were new Canadians. In the Northwest and Yukon Territories, a random sample of households was selected from all areas. In Nunavut, all available records were purchased because of the large number of those of Aboriginal identity residing in this territory.

SUPPLEMENT E2: CREATING WEIGHTED ESTIMATES TO ACCOUNT FOR OVERSAMPLING OF VULNERABLE POPULATIONS

To create the weighted estimates, nonoverlapping subgroups of interest, each characterized by education, income, Canadian-born, and Aboriginal status, were created for both the study population and the 2006 Canadian Census database. The weight for each vulnerable group of interest was calculated by dividing the proportion of individuals in the Census who fall into this subgroup by the proportion of individuals in the SPAACE who fall into this same subgroup.

SUPPLEMENT E3: CREATING THE STATISTICAL MODELS FOR THE ANALYSIS OF NONRESPONSE

To account for missing data in our study, we used multiple imputation for both ignorable and nonignorable missing data, as proposed by Kmetic et al.[4] We created posterior distributions for the prevalence of food allergy for *Full Participants*, *Refusal Questionnaire Participants*, *Nonparticipants*, and *Never Reached Participants*, and mixtures of these posterior densities formed our final prevalence estimates. The prevalence of food allergy for the *Full* and *Refusal Questionnaire Participants* was estimated using data from the telephone survey, but to estimate the prevalence in the *Nonparticipants* and *Never Reached Participants*, for whom data on food allergy were missing, we created estimates across a range of clinically and statistically plausible assumptions.

Multiple imputation was used to adjust the estimates for nonresponse bias from missing food allergy data within the *Nonparticipants* and the *Never Reached Participants* by using a model that included observed data (CT and province of residence) to predict the missing data on the probability of food allergy.[4] Multiple imputation is the gold standard for adjusting for missing data.[5] It involves filling in missing values for the presence or absence of food allergy with a "best guess" that is based on the assumptions of bias described above. Ten thousand versions of the complete dataset were formed and data analysis was carried out on each dataset. To derive final inferences from the data, an average of the results from each of the ten thousand datasets was used as a point estimate for prevalence, with overall variance equal to the sum of within and between imputation variances.[5] Point estimates and 95% credible intervals (CrIs) were estimated. A 95% CrI implies that there is a 95% probability that the parameter of interest falls within the upper and lower limit of the interval, given the data and prior information used. If low information priors are used, the 95% CrIs essentially reflect the information in the data.

Before running the multiple imputation programs in WinBUGS, the following preliminary steps were completed:

1. In households who completed the Refusal Questionnaire and indicated that 1 or more members had a food allergy, the number of allergic individuals was imputed because it was unknown how many individuals had a food allergy. The

number of allergic individuals in each household was imputed based on the distribution of the number of allergic individuals from the *Full Participants*.
2. The total number of individuals in the household was imputed for the *Nonparticipants*, the *Never Reached Participants*, and the nonallergic *Refusal Questionnaire Participants*, based on the distribution of the total number of individuals from the *Full Participants*.
3. The prevalence of food allergy in the *Full Participants* was estimated by taking the observed number of allergic people divided by the observed total number of people in this group, assuming a binomial distribution.
4. The prevalence in the *Refusal Questionnaire Participants* was estimated by taking the imputed number of allergic people (described in step 1) divided by the observed total number of people in those households who reported allergy plus the imputed total number of people in households who did not report allergy (described in step 2).

As detailed in the article, 3 assumptions regarding the prevalence of food allergy in the *Nonparticipants* and 3 assumptions regarding the prevalence in the *Never Reached Participants* were investigated, which yielded 9 different models, as follows:

1. The prevalence in the *Nonparticipants* is half that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Nonparticipants*.
2. The prevalence in the *Nonparticipants* is the same as in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Nonparticipants*.
3. The prevalence in the *Nonparticipants* is twice that of the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Nonparticipants*.
4. The prevalence in the *Nonparticipants* is half that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is a mixture of the prevalence in the *Full Participants*, *Refusal Questionnaire Participants*, and *Nonparticipants*.
5. The prevalence in the *Nonparticipants* is the same as in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is a mixture of the prevalence in the *Full Participants*, *Refusal Questionnaire Participants*, and *Nonparticipants*.
6. The prevalence in the *Nonparticipants* is twice that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is a mixture of the prevalence in the *Full Participants*, *Refusal Questionnaire Participants*, and *Nonparticipants*.
7. The prevalence in the *Nonparticipants* is half that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Full Participants*.
8. The prevalence in the *Nonparticipants* is the same as in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Full Participants*.
9. The prevalence in the *Nonparticipants* is twice that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Full Participants*.

Multiple imputation was implemented via a hierarchical logistic regression model, with 4 levels: individual, household, CT, and province of residence. Each model had the same basic structure, as follows:

$$\begin{aligned}\text{logit}(\text{prevalence}) = {} & \text{intercept}_i + \text{household effect}(\text{number of individuals in household}) \\ & + \text{assumption about prevalence}(1-4,\ \text{unique for each of the 4 groups of participants})\end{aligned}$$

where the intercept depended on the CT and province (represented by "i" in the above equation). There were 13 provinces and 265 CTs. Province of residence and CT information was available for all households, regardless of participation level, and so was included for all subjects in the model. The analyses were carried out using WinBUGS (version 1.4.3, MRC Biostatistics Unit, Cambridge, United Kingdom).

REFERENCES

1. Pickett K, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health 2001;55:12.
2. Low Income Cutoffs. Available from, http://www.statcan.gc.ca/pub/75f0002m/2012002/lico-sfr-eng.htm; 2012.
3. 2006 Census Dictionary. Ottawa; 2007. Available from https://www12.statcan.gc.ca/census-recensement/2006/ref/dict/azindex-eng.cfm#C.
4. Kmetic A, Joseph L, Berger C, Tenenhouse A. Multiple imputation to account for missing data in a survey: estimating the prevalence of osteoporosis. Epidemiology 2002;13:437-44.
5. Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. 2nd edition. Chapman and Hall; 2003.