The Prevalence of food allergy in Canada:

A focus on vulnerable populations

Lianne Soller, PhD candidate

Department of Epidemiology, Biostatistics and Occupational Health

McGill University, Montreal

May, 2015

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of

Doctorate of Science

TABLE OF CONTENTS

ABSTRACT

To better characterize the burden of food allergy in Canada, this thesis: 1) provides prevalence estimates for common food allergies among those of low education, low income, new Canadians, and those of Aboriginal identity (vulnerable populations), 2) investigates prescription and availability of the epinephrine auto-injector (EAI) among these populations, and 3) explores the effect of non-response bias on the prevalence of allergy to any food.

Using 2006 Canadian Census data, postal codes with high proportions of vulnerable populations were identified and households within these postal code areas randomly selected to participate in a telephone survey. Information on food allergies, prescription and availability of the EAI, and demographics, were collected. Prevalence estimates were weighted using Census data to account for the targeted sampling. Multivariable logistic regression was used to identify predictors of food allergy, and prescription and availability of the EAI. Multiple imputation was used to generate non-response bias adjusted prevalence estimates.

Of the 12,762 eligible households, 5,734 households (15,022 individuals) completed the food allergy prevalence questionnaire (45% response rate), 524 households completed the Refusal Questionnaire (an additional 4%), and the remaining 6,504 households answered the telephone but refused to provide any information. An additional 3,224 households were never reached. Food allergy was less common among adults without post-secondary education and new Canadians. There were no differences according to income or Aboriginal identity. In the non-response bias analyses, nine estimates were obtained for the perceived prevalence of allergy to any food, all of which were lower than the prevalence of food allergy among full participants. In

the multivariate model for prescription of the EAI, higher education was associated with being more likely to be prescribed an EAI. There were no differences in terms of availability of the EAI.

Our data suggest that certain groups of vulnerable Canadians self-report fewer allergies and EAI prescriptions, which may be real and/or a result of lack of appropriate healthcare or awareness of allergies. This suggests important policy gaps that must be addressed to ensure equal opportunity for all Canadians to seek and receive healthcare. The non-response bias analysis highlights the importance of minimizing non-response bias and considering it when performing data analysis.

RÉSUMÉ

Pour mieux comprendre la charge des allergies alimentaires au Canada, cette thèse : 1) fourni des estimations de la prévalence des allergies aux aliments communs parmi ceux avec un faible niveau d'éducation, un revenu en bas de la moyenne, des immigrants qui sont arrivés au Canada dans les derniers dix ans, et ceux qui s'identifient comme ayant une identité autochtone (les populations vulnérables), 2) étudie la préscription et la disponibilité de l'autoinjecteur d'épinephrine (AIE) parmi ces populations, et 3) explore l'effet des bias causés par un taux de réponse bas sur la prévalence des allergies alimentaires.

Nous avons utilisés les donnés du recensement canadien de l'an 2006 pour pouvoir identifier les codes postaux avec les proportions de populations vulnérables les plus hautes. Des maisons parmi ces codes postaux ont été sélectionnées en façon aléatoire pour participer a un sondage téléphonique. Des informations sur les allergies alimentaires, la prescription et la disponibilité de l'AIE, et les démographiques, ont été receuillies. Pour ceux qui ont refusés de participer au sondage complet, un questionnaire de refus très court a été administrer, qui a évalué si les membres de la famille avaient des allergies alimentaires. Des estimations de la prévalence des allergies alimentaires ont été adjuster en utilisant les donnés du recensement canadien pour prendre en consideration l'échantillonnage ciblé. La regression logistique multivariable a été utilisé pour identifier les prédicteurs des allergies alimentaires, de la prescription et de la disponibilité de l'AIE. L'imputation multiple a été utilisée pour générer des estimations de prévalences ajustées pour le bias causés par un taux de réponse bas.

Des 12,762 maisons éligibles, 5,734 maisons (15,022 individus) ont complétés le questionnaire (taux de réponse de 45%), 524 maisons ont completes le questionnaire de refus (un 4% additionnelle), et les 6,504 maisons qui restent ont répondus au téléphone mais ont refusés de répondre aux questions. 3,224 maisons additionnelles n'ont jamais répondu au téléphone. Parmi ceux qui ont complétés le questionnaire, la prévalence d'allergies auto-évalué non-ajusté était de 6.4%, et avec l'ajustement, 7.5%. Les allergies alimentaires étaient moins communes chez les adultes sans éducation post-secondaire et les nouveaux immigrants. Il n'y avait aucune différence en ce qui concerne le revenue ni l'identité autochtone. Pour l'analyse du bias causés par un taux de réponse bas, neuf estimations ont été obtenus pour la prévalence auto-évaluée, allant de 3.0% à 5.4%. En ce qui concerne l'analyse multivariable pour la prescription de l'AIE, ceux ayant une éducation plus élevé ont une chance plus élevés d'avoir été prescrit l'AIE. Il n'y avait aucune différence en ce qui concerne la disponibilité de l'AIE.

Notre étude suggère que certains groupes vulnérables au Canada auto-évalue moins d'allergies alimentaires et de prescriptions pour l'AIE, qui pourrait être une observation réele ou qui pourrait être causé par un manque de services de santé ou un manque d'éducation pour les allergies alimentaires. Cela suggère des lacunes dans le système politique qui doivent être addréssées pour que tout les canadiens aient la même opportunité pour chercher et recevoir des soins de santé. En étant que la prévalence non-ajustée parmi les participants qui ont completes le questionnaire était plus élevé que tout les estimations ajustées, c'est evident que le taux de réponse peux causer un bias important dans l'estimation de la prévalence des allergies alimentaires, et que nous devrions prendre soin de ne pas ignorer ce bias. Notre projet souligne l'imprtance de maximiser le taux de réponse durant la phase conception de l'étude, et en même

temps, c’est important de reconnaître que le bias est probablement encore présent et devrait être considéré dans l’analyse des données.

ABBREVIATIONS

CI: confidence interval

CrI: credible interval

CT: census tract

DBPCFC: double-blind, placebo-controlled food challenge

EAI: epinephrine auto-injector

FAPQ: food allergy prevalence questionnaire

FP: Full Participants

LICO: low income cut-off

NP: Non-Participants

NRP: Never-Reached Participants

OR: odds ratio

RQ: Refusal Questionnaire

RQP: Refusal Questionnaire Participants

SCAAALAR: Surveying Canadians to Assess the Prevalence of Common Food Allergies and Attitudes towards Food Labelling and Risk

SPAACE: Surveying Prevalence of food Allergy in All Canadian Environments

SPT: skin prick test

STATEMENT OF ORIGINALITY

This thesis presents original work on the prevalence of food allergy, prescription and availability of the epinephrine auto-injector, and the impact of non-response bias on the prevalence of food allergy, in a telephone survey specifically targeting those of low education, low income, new Canadians, and those of Aboriginal identity (vulnerable populations). First, while the prevalence of food allergy in Canada was previously estimated by our group, this is the first study to specifically target vulnerable populations of Canadians to estimate food allergy prevalence in these groups and to compare vulnerable with non-vulnerable populations. We also present prevalence estimates for probable food allergy for milk, egg, wheat, and soy in the current study, whereas our previous study only estimated self-reported (perceived) prevalence. In addition, this is the first study on food allergy to explore the role of non-response bias on prevalence estimates, using Bayesian methodology and making various about the magnitude and direction of such bias to see the effect of different assumptions about prevalence. Finally, although we have previously estimated the proportion of food-allergic Canadians with the epinephrine auto-injector (EAI), the current study queried specifically on prescription and availability of the EAI and examined differences between vulnerable and non-vulnerable populations.

CONTRIBUTION OF AUTHORS

L Soller designed the questionnaire, wrote the protocol, supervised data collection, performed all data analysis, and wrote all manusripts.

M Ben-Shoshan helped design the study questionnaire, the protocol, and with manuscript preparation. He also sits on the thesis supervisory committee, and helped determine which sections would be included in the thesis.

S Cherkaoui helped with analysis of some of the data on epinephrine auto-injectors, presented the data at a Canadian conference, wrote the preliminary abstract, and helped with manuscript preparation.

M Knoll helped test the questionnaires prior to the beginning of data collection. She also collected some of the data for the telephone interviews, helped analyze the data, and helped with manuscript preparation.

D W Harrington accessed the Census 2006 data, determined the areas with the highest proportion of the groups of interest, and extracted postal codes from these areas. These postal codes were then sent to L Soller, who sent them to Info Direct, where random households from these postal codes were chosen. He also participated in data analysis and helped with manuscript preparation.

J Fragapane provided the technical support for the study and programmed the telephone questionnaire. He also assisted with manuscript preparation.

L Joseph helped design the study, determine the sample size, and develop the refusal questionnaire. He also provided direct guidance to L Soller during data analysis and manuscript preparation, including a lengthy discussion about the non-response bias analysis and help with preparing the WinBUGS programming. He also read through each section of this thesis, and helped determine what content would be included.

Y St Pierre provided statistical expertise including preparation of the dataset for analysis, and consultations with L Soller regarding the procedure for weighting the prevalence estimates. He also assisted in manuscript preparation.

S La Vieille provided insight regarding the needs of the federal government, and how the study was going to answer these needs. He also assisted in manuscript preparation.

K Wilson provided expertise on vulnerable populations. She helped develop the questionnaire so that it was culturally relevant, and consulted with the team regarding how to increase response rates. She also assisted in manuscript preparation.

S J Elliott provided expertise in social science and medicine, with regards to vulnerable populations and how to best approach them to participate in the study. She also provided invaluable resources to L Soller in terms of background material for this thesis, and she helped prepare the manuscripts.

A E Clarke provided direct supervision to L Soller and read all parts of this thesis. She helped in determining the content of the thesis and what analysis would be performed. She participated greatly in manuscript preparation.

I: INTRODUCTION AND OBJECTIVES

Introduction

The prevalence of food allergy

Food allergy has become a topic of increasing interest in today's society because of its unknown aetiology, unpredictable progression, difficult diagnosis, and potentially devastating consequences on the quality of life of affected individuals.[1] Estimates of the prevalence of food allergy vary from 3% to 35%, depending on the population and geographic area surveyed, and the study methodology used.[2]

Individuals with self-reported (perceived) food allergy are often found not to have true IgE-mediated food allergy after complete evaluation by an allergist; their signs and symptoms after ingesting the suspected allergen are not typical or do not occur in the appropriate time frame, or food-specific IgE cannot be demonstrated with appropriate diagnostic testing. Rona found that the overall prevalence of food allergy decreased when confirmatory tests were used to diagnose food allergy; estimates based on food challenges (1% to 10.8%) were generally lower than those based on Skin Prick Test (SPT) or IgE blood test (2% to 5%), but there was still inconsistency in the prevalence of allergy across studies.[2]

In assessing the prevalence of food allergy at a national level, it is clearly much more challenging to pose multiple questions which attempt to characterize the symptoms, time course, and treatment of possible allergic reactions and to seek results of confirmatory testing than to

pose a single question about the presence or absence of food allergy. Hence, there are only a few population-based studies on prevalence that have attempted to more fully describe adverse reactions to food. Sicherer in the United States[3-6] and our group in Canada[7, 8] are the only groups in North America who attempted to characterize food allergy by asking detailed questions about symptoms, diagnosis, and treatment of food-allergic reactions. Our previous nationwide Canadian study (SCAAALAR: Surveying Canadians to Assess the prevalence of common food Allergies and Attitudes towards food LAbelling and Risk) estimated the perceived prevalence of allergy to all nine common allergens (peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and soy) and allergy to any food.[7] This study did not collect probable allergy (see Appendix A) information on all of the nine common food allergens-it excluded milk, egg, wheat, and soy to reduce the length of the survey.

Prevalence of food allergy among vulnerable populations

It is hypothesized that those of low socioeconomic status (SES), new Canadians (immigrating to Canada in the last 10 years), and individuals of Aboriginal identity, hereafter termed *vulnerable populations*, experience fewer food allergies (see Appendix A). These populations are considered 'vulnerable' because they are more likely to experience issues accessing adequate healthcare services.[9] This is possibly explained by differences in environmental exposures[10, 11] and diet, [12, 13] as well as by issues with access to education and healthcare for food allergy.[14-16] However, there is very little data to support these hypotheses.

In SCAAALAR, participants were predominantly those of high SES and born in Canada.[7] Although these groups were found to have the highest prevalence of food allergy overall,[8] and for specific foods,[17] it is possible that this finding is a result of the non-response bias inherent in telephone surveys, whereby those with food allergy are more likely to participate in the survey, and those who participate in the survey are also more likely to be of high SES and Canadian-born.

It is therefore important to gather further information on the prevalence of food allergy among vulnerable populations in which these groups are specifically targeted. However, collecting information on vulnerable populations using a targeted sampling strategy via telephone survey is unlikely to solve the problem of non-response bias, and hence, statistical techniques to adjust for such bias must also be considered.

Prescription and availability of the Epinephrine Auto-injector

Since there is no widely accepted cure for food allergy, the only way to prevent an allergic reaction is to avoid the known allergen. Avoidance is often difficult because of unclear or absent precautionary statements on packaged foods,[18] and accidental exposures continue to occur even if the patient takes all the necessary precautions.[19] Patients must therefore rely on effective treatment in the case of an accidental exposure. This involves the prompt administration of epinephrine as soon as symptoms appear or exposure to the known allergen is suspected.[20, 21] Since the majority of allergic reactions occur outside the hospital, allergists recommend that all individuals with a history of anaphylaxis, a severe allergic reaction, carry an epinephrine auto-

injector (EAI).[22-29] However, even with specific guidelines from allergists, the EAI is still under-prescribed by physicians,[30] and, in cases where it is prescribed, the patient may not carry it[31, 32] or may not know how or when to use it.[33-35]

The SCAAALAR study found that only 55% of Canadians diagnosed with peanut, tree nut, fish, shellfish and/or sesame allergy self-reported having an epinephrine auto-injector (EAI), and adults, males, and those residing in households where the respondent was single were less likely to have one.[36] Vulnerable populations are thought to be even less likely to be prescribed and carry the EAI due to lower accessibility to a regular healthcare provider, greater use of alternative healthcare, and limited access to employee health benefits due to lower employment rates.[16, 37-40] However, as was previously mentioned, SCAAALAR did not adequately represent the vulnerable populations, so we could not make conclusions about differences in prescribing patterns. Further, that study did not query patients on actual availability of the EAI (see Appendix A).

Objectives

To better characterize the burden of food allergy in Canada, the specific objectives of this thesis are:

1) To estimate the prevalence of perceived and probable allergy to peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and soy, and allergy to any food, among vulnerable populations and compare with the prevalence in non-vulnerable populations.
2) To explore the effect of non-response bias on the prevalence of allergy to any food, and

3) To investigate prescription and availability of the epinephrine auto-injector among vulnerable populations.

This thesis will contain eight chapters, including the current one. Chapter 2 will review the literature on the prevalence of food allergy among vulnerable populations. Chapter 3 includes a manuscript responding to Objective 1. Chapter 4 will discuss non-response bias and various techniques to reduce and adjust for it. Chapter 5 presents a paper that addresses Objective 2. Chapter 6 reviews the literature on prescription and availability of the epinephrine auto-injector. Chapter 7 presents a paper responding to Objective 3. Finally, Chapter 8 will provide a summary and final conclusions.

II: LITERATURE REVIEW-PREVALENCE AND PREDICTORS OF FOOD ALLERGY AMONG VULNERABLE POPULATIONS IN CANADA

Introduction

Food allergy has become an increasingly important condition in western society due to its unpredictable nature and the need for extreme dietary vigilance, both of which can substantially compromise the quality of life of affected individuals and their families.[1, 41] Although immune modulatory therapies appear promising, these likely will not induce long term tolerance,[42] and food allergy will remain largely incurable. Those affected must rely on strict avoidance of the offending food and rescue therapy with epinephrine. Although allergic reactions to a large variety of foods have been reported,[43] the majority of reactions in North America are caused by nine main allergens: peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and soy.[2, 7, 8, 44-46]

Definition of food allergy

Food allergy is an adverse reaction arising from a specific immune response that occurs reproducibly upon exposure to a food.[43] A food allergen is defined as the specific component of a food or an ingredient in the food (usually a protein) that elicits the allergic reaction.[43] Allergic reactions to food can be either IgE-mediated or non-IgE-mediated, but only IgE-mediate will be considered in this thesis.[47] Symptoms of an IgE-mediated allergic reaction involve several organ systems including the skin (pruritis, urticaria, erythema, angioedema), eyes (pruritis, edema), respiratory tract (nasal congestion, rhinorrhea, sneezing, cough, chest tightness, wheezing), oral cavity (angioedema of the palate, lips and tongue), gastrointestinal tract (nausea, abdominal pain, reflux, diarrhea, vomiting) and cardiovascular system (tachycardia, hypotension, dizziness,

fainting, loss of consciousness).[43] Symptoms of an allergic reaction almost always occur within a few minutes to a few hours after ingestion of the allergenic food.[43] The exact mechanism of food allergy is still unknown, although several hypotheses exist including the nature of the food allergens and the interplay of genetic and environmental factors.[48, 49]

Overall prevalence of food allergy

Several studies have presented widely varying estimates of the overall prevalence of food allergy. The first systematic review to summarize the literature on this topic found estimates of self-reported food allergy ranging from 3% to 35%.[2] The authors note that there were a variety of potential reasons for the discrepancy in prevalence estimates across studies including the population and geographic area surveyed, and the study methodology. In an attempt to estimate the prevalence of food allergy in a more homogeneous population, researchers synthesized the literature from European prevalence studies only. Although prevalence estimates were even more discrepant than the first review (range: 1.6% to 38.1%), in general, children self-reported more food allergies than adults, and the prevalence was highest in Northwestern European countries and lowest in Southern Europe.[50] Although it is unclear why such differences were observed, a recent European study found country-specific differences in a wide range of factors that are hypothesized to play a role in the development of food allergy: allergic family history, obstetrical practices, and environmental exposures.[51] North American studies are more consistent in their results, with two groups from the United States and our group from Canada reporting estimates between 8% and 9%.[8, 52, 53]

Prevalence of food allergy among vulnerable populations

As we have seen in the previous section, there is a wide discrepancy in estimates of food allergy prevalence among the general population, particularly outside of North America. As anticipated, the situation is similar for estimates of food allergy prevalence in vulnerable populations - those of low education, low income, new Canadians (immigrating to Canada in the last 10 years), and individuals of Aboriginal identity. These populations are considered 'vulnerable' because they are more likely to experience issues accessing adequate healthcare services.[9] It is therefore of utmost importance to estimate the burden of food allergy among these populations. Researchers have attempted to do so, but existing studies are limited in that the majority focus only on children, do not collect data on specific food allergies, and/or do not employ an appropriate targeting strategy to ensure that an adequate sample of these vulnerable groups, who are particularly difficult to reach, are included.[15, 52-59] These limitations make it difficult to form any definitive conclusions about how the prevalence of food allergy in these vulnerable groups compares with that in the non-vulnerable populations.

Prevalence of food allergy according to socioeconomic status

Data exist comparing the prevalence of food allergy in different socioeconomic (SES) groups, but they are incomplete. Pawlinska-Chmara found a higher prevalence of food allergy in Polish children whose parents reported high SES compared to those reporting low SES [10.4% (95% CI, 7.0%, 14.7%) versus 2.2% (95% CI, 0.8%, 4.8%)].[54] In the United States, Gupta found that children from households with a higher income had a higher odds of having food allergy than those with a lower income [OR: 2.0 (95% CI, 1.4, 2.5)][53] and a recent report from the United States National Center for Health Statistics found that childhood food allergy prevalence

increased with increasing income level (4.0% among children with family income less than 100% of the poverty level, 5.0% among children with family income between 100% and 200% of the poverty level, and 5.4% among children with family income above 200% of the poverty level).[60] These studies are limited in that they only collected data on children, employed an ambiguous definition of socioeconomic status,[54] did not collect data on education, [53, 60] and used an unconventional definition of food allergy that does not differentiate between anaphylactic allergies and gastrointestinal intolerance and only considers food allergies in the past year.[60]

One American study estimated the prevalence of food allergy among adults, and found a higher prevalence in adults of higher education compared to those with lower education [7.4% (95% CI, 6.7%, 8.2%) versus 3.1% (95% CI, 2.6%, 3.6%)].[52] This study did not collect information on income. In addition, the study is more than a decade old, and warrants an update to determine whether the differential based on educational attainment is still present, more pronounced, or if it has decreased.

A study in the United Kingdom (UK) found that peanut allergy was more prevalent among adults and children of higher SES, but the study sample only included those registered with a family practitioner, and is hence not representative of the general population in the UK.[61] Another limitation is that this study only collected data on peanut allergy.

Our group in Canada has shown that those residing in a household where the primary respondent had a post-secondary education, compared with those who did not, had a higher odds of reporting food allergy overall [OR: 1.24 (95% CI, 1.03, 1.51)],[8] and of reporting tree nut allergy

[OR 1.90 (95% CI, 1.18, 3.04)].[17] No differences were observed according to household income. However, this study did not specifically target those of lower education or income, and hence, with such a small sample of individuals in these two groups, it is difficult to make conclusions about differences in prevalence according to education or income. Another limitation is that educational attainment was collected from the primary respondent and not from all adults in the household.

The Surveying Prevalence of food Allergy in All Canadian Environments (SPAACE) study, which forms the basis for this thesis, targeted vulnerable populations specifically, to estimate the prevalence of food allergy in these groups. In addition, SPAACE collected data on the level of education for all adults in the household, thereby allowing for the calculation of individual-level prevalence estimates according to education level. Data on household income was also collected to enable a comparison of food allergy prevalence according to income level.

There are several hypotheses as to why those of higher SES may have more food allergies. One reason is that highly educated and wealthier parents may have been more likely to follow the guidelines from the American Academy of Pediatrics which, until recently, recommended restriction of allergenic foods during pregnancy, lactation, and in infancy.[62] It is now thought that delayed introduction may actually increase, rather than decrease, the likelihood of developing food allergy.[12]

The higher prevalence of allergic disease in those of higher SES may also be explained by a phenomenon known as the hygiene hypothesis. The idea is that individuals of higher

socioeconomic status are exposed to fewer bacterial infections, which causes a skewing of the immune response away from Th1, the immune cells that fight infection, towards Th2, which leads to an increase in allergy.[63]

In addition, it has been shown that individuals of higher SES are more likely to seek medical attention for their ailments, and hence may be more likely to obtain a physician diagnosis of food allergy than those of lower SES, thereby causing an artificially inflated prevalence of food allergy in this group. They may also have better access to family doctors than those of lower education or income.[14]

Prevalence of food allergy according to country of birth

Peanut allergy has been reported to be less common in Asian children (0.43% and 0.64%) [64, 65] than in Canadian (1.8%), American (0.9% to 2%), and British (1.2% to 1.8%) children; the prevalence is 1.8% in Canada,[7] 0.9% to 2.0% in the United States,[4, 53, 66] and between 1.2% and 1.8% in the UK.[67, 68] The Asian study found a higher prevalence of food allergy among children of expatriates born in western countries compared with children born in Singapore and the Philippines.[64] Although it appears that children born in westernized countries have a higher prevalence of food allergy, the study populations were not selected in the same way and the ages of children differed between studies, precluding any definitive conclusions about differences in prevalence according to country. The Asian study collected data using a structured written questionnaire that was administered to local and expatriate school children in Singapore (4-6 and 14-16 years old) and the Philippines (14-16 years old),[64] while the Canadian and American studies randomly selected households and administered telephone surveys asking about children

of all ages.[4, 7, 53, 66] One study in the UK conducted face-to-face interviews with mothers of primary school-aged children,[67] and the other UK study collected data using a combination of medical records of children 4 years of age at one allergy centre on the Isle of Wight, and where data were missing from the medical records, physicians interviewed mothers.[68]

A recent population-based study from the United States reported that foreign-born children had a decreased odds of having food allergy than those born in the United States [OR: 0.48 (95% CI, 0.38, 0.61)].[55] A recent paper by our group at McGill suggests that food allergy may be less common among immigrants to Canada compared with individuals born in Canada, although the sufficient evidence for a difference was only available for shellfish allergy [OR: 0.49 (95% CI, 0.26, 0.95)].[17] Unfortunately, this study did not collect individual-level data on immigrant status, and the sample size of immigrants was quite small, making it difficult to draw definitive conclusions regarding differences in prevalence between immigrants and Canadian-born individuals.

These studies seem to support the hypothesis that individuals born in western countries have a higher prevalence of food allergy than those born in other countries, no matter what their country of residence. However, these studies are limited in that they collect data only among children,[53, 55, 64, 67, 68] only collected data on a few allergens[4] and/or do not collect immigration information for all family members.[17] In addition, there is no data that estimates prevalence according to number of years since immigration. The SPAACE study bridges the gaps identified in the previous paragraphs; specifically, data were collected, for individuals of all ages, on allergies to all foods and on the number of years since immigration.

Although some evidence supports that individuals born in westernized countries like Canada, the United States, and the UK have a higher prevalence of food allergy than those born in other countries, it is sparse and has limitations. There are a few hypotheses that seem to support such claims. The "healthy immigrant effect" states that immigrants to westernized countries tend to have a low prevalence of chronic conditions when they first arrive in their host country, and that their health status converges to that of the host population with increasing time since migration.[69, 70] Data supporting this hypothesis with regards to food allergy are lacking, however.

Differences in dietary habits may partially explain the apparently higher prevalence of food allergy among those born in westernized countries.[71] Individuals from western and non-western countries introduce certain foods into the diet of their children at different ages,[71] and age of introduction of allergenic foods may determine whether or not an individual develops tolerance or becomes sensitized to a food.[72, 73] Alternative preparation of the allergenic food, such as boiling peanuts, which is commonly done in many Asian countries, instead of roasting, as is the norm in North America, reduces allergenicity,[74, 75] which may explain the higher prevalence of peanut allergy in western-born individuals compared with those born in Asian countries.[64]

Barriers to adequate health care may also be responsible for the apparent difference in prevalence between those born in westernized countries versus other countries. For example, it has been reported that recent immigrants to Canada lack family doctors and consequently do not have access to appropriate diagnostic testing for possible food allergy.[14] An additional problem is language and cultural barriers, which prevent physicians from conveying information about food

allergies to new Canadians and hence, they may not appreciate that they have been diagnosed with a food allergy.[76]

Prevalence of food allergy according to Aboriginal status

Population-based data on the prevalence of food allergy among Aboriginal people in Canada are currently unavailable. A recent study showed that allergies and asthma are the second and third most frequently reported health concerns among on-reserve First Nations and Inuit children in Canada.[56] The 2006 Aboriginal Peoples Survey (APS) estimated the prevalence of allergies amongst Inuit children to be 10%,[57] and the Manitoba First Nations Regional Health Study of 2002/2003 found a prevalence of allergies of 8.8%.[58] Similarly, "A Shared Vision" reports that the rate of allergies in Aboriginal children (12.2%) is comparable to that of children in the general population of Canada (16.4%),[59] but that they are less likely to receive treatment for their allergy due to inadequate healthcare access.[57] A recent study estimated that 4.3% of off-reserve Canadian Aboriginal children were reported to have a food allergy, but this percentage decreased to 2.8% when patients were asked if a doctor had diagnosed the allergy.[15]

Although it is possible that food allergy is less common among Aboriginal children compared with the general population, the studies published are either not population-based, focus only on children, and/or do not differentiate between food and other allergies.[56-59] The SPAACE study is the first population-based study to estimate the prevalence of all food allergies among individuals reporting Aboriginal ancestry, regardless of age.

Although there is a lack of complete information regarding prevalence of food allergy amongst Aboriginal people, as we have seen in the previous paragraph, researchers have postulated that the prevalence of food allergy may be lower than in non-Aboriginal people. Potential explanations for a perception of lower prevalence include decreased access to healthcare on-reserve and the consequent decreased access to diagnosis and treatment for food allergy.[57, 77-79]

In addition, real differences in prevalence between Aboriginal and non-Aboriginal individuals may exist. In Canada, on-reserve Aboriginal communities experience substandard housing conditions and overcrowding due to a lack of municipal infrastructure.[58, 59] In fact, 15% of First Nations live in a crowded dwelling, compared to just 3% of the general population.[59] The hygiene hypothesis suggests that more siblings, early childhood infections, and poor sanitation more generally may affect the development of allergic diseases such as food allergy.[10, 11, 13, 59] Investigation into the relationship between intestinal bacteria and food allergy has found that the presence of intestinal microbiota, caused by infections during infancy, plays an important role in preventing the development of allergic disease.[80, 81]

Differences in diet may also explain differences in food allergy prevalence between individuals with and without Aboriginal ancestry. A recent study compared the traditional diet of Aboriginal Australians, which consists of lean red meat, plenty of seafood, and a large variety of fruits and vegetables, to the western diet of Australians, which consists of high levels of sugar and saturated fats, and reduced access to fresh fruits and vegetables.[13] This study found that asthma and allergies were more prevalent in those who followed a western diet compared with those who followed a more traditional diet, and suggested that a western diet may increase the

likelihood of developing food allergy by decreasing the diversity of microbiota in the gastrointestinal tract.[13]

Summary

In this review, we have shown that food allergy is an important health concern. Obtaining reliable and accurate data on the prevalence of food allergy is challenging, and particularly so among vulnerable populations, who are less likely to receive appropriate healthcare for their illnesses. Unfortunately, the current data are incomplete in that they are not population-based and/or do not collect information about all food allergens or across all ages. To better understand the burden of food allergy among vulnerable populations in Canada, the SPAACE study used Census 2006 data to specifically target and survey randomly selected households from areas across Canada with a high proportion of vulnerable populations. This sampling strategy allowed us to have sufficient sample sizes to present valid comparisons between prevalence estimates in vulnerable and non-vulnerable Canadians. In Chapter III, data from SPAACE on the prevalence of common food allergies and allergy to any food among vulnerable populations, and a comparison between vulnerable and non-vulnerable populations, will be presented in the form of a manuscript that was published in JACI: In Practice.

III: PREVALENCE AND PREDICTORS OF FOOD ALLERGY IN CANADA: A FOCUS ON VULNERABLE POPULATIONS

Lianne Soller, Moshe Ben-Shoshan, Daniel W. Harrington, Megan Knoll, Joseph Fragapane, Lawrence Joseph, Yvan St. Pierre, Sebastien La Vieille, Kathi Wilson, Susan J. Elliott, and Ann E. Clarke.

Abstract

Background

Studies suggest individuals of low education and/or income, new Canadians (immigrated <10 years ago), and individuals of Aboriginal identity may have fewer food allergies than the general population. However, given the difficulty in recruiting such populations (hereafter referred to as vulnerable populations), by using conventional survey methodologies, the prevalence of food allergy among these populations in Canada has not been estimated.

Objectives

To estimate the prevalence of food allergy among vulnerable populations in Canada, to compare with the non-vulnerable populations, and to identify demographic characteristics predictive of food allergy.

Methods

By using 2006 Canadian Census data, postal codes with high proportions of vulnerable populations were identified and households randomly selected to participate in a telephone survey. Information on food allergies and demographics was collected. Prevalence estimates

were weighted by using Census data to account for the targeted sampling. Multivariable logistic regression was used to identify predictors of food allergy.

Results

Of 12,762 households contacted, 5,734 households completed the questionnaire (45% response rate). Food allergy was less common among adults without post-secondary education versus those with post-secondary education [6.4% (95% CI, 5.5%, 7.3%) versus 8.9% (95% CI, 7.7%, 10%)] and new Canadians versus those born in Canada [3.2% (95% CI, 2.2%, 4.3%) versus 8.2% (7.4, 9.1)]. There was no difference in prevalence between those of low and high income or those with and without Aboriginal identity.

Conclusions

Analysis of our data suggests that individuals of low education and new Canadians self-report fewer allergies, which may be due to genetics, environment, lack of appropriate health care, or lack of awareness of allergies, which reduces self-report.

Introduction

Food allergy has become an increasingly important condition in western society due to its unpredictable nature and the need for extreme dietary vigilance, both of which can substantially compromise the quality of life of affected individuals and their families.[41] Although immune modulatory therapies appear promising, these likely will not induce long term tolerance,[42] and food allergy will remain largely incurable. Those affected must rely on strict avoidance of the offending food and rescue therapy with epinephrine. In the United States, estimates of the prevalence of self-reported food allergy range between 8.0% and 9.1%.[52, 53] However, until recently, the prevalence of food allergy in Canada was unknown.

From 2008 to 2009, our research team estimated that approximately 8% of Canadians self-report at least 1 food allergy and that the prevalence differs across socioeconomic groups and geographic regions (the SCAAALAR study: Surveying Canadians to Assess the prevalence of food Allergy and Attitudes towards food LAbelling and Risk).[8] However, given that the data were collected by using a large-scale telephone survey, it is not surprising that the resulting sample under-represented important parts of the Canadian population, specifically those of low education and low income, new Canadians, and individuals of Aboriginal. These 4 population groups are hereafter referred to as vulnerable populations. Although others have attempted to estimate the prevalence of food allergy in these vulnerable populations, existing studies are limited in that the majority focus only on children, do not collect data on specific food allergies, and/or do not use an appropriate targeting strategy to ensure an adequate sample of these vulnerable groups, who are particularly difficult to reach, are included. [52-59, 82] These limitations

make it difficult to form any definitive conclusions about how the prevalence of food allergy in these groups compares with that in the non-vulnerable populations.

The current study (SPAACE: Surveying Prevalence of food Allergy in All Canadian Environments) attempts to bridge these gaps, by specifically targeting and evaluating prevalence of specific food allergies in vulnerable populations of children and adults in all Canadian provinces and territories, comparing vulnerable with non-vulnerable populations, and examining potential socio-demographic determinants of food allergy.

Methods

Selection of study population

Canadians of low income, new Canadians, and individuals of Aboriginal identity were specifically targeted (see Appendix B for more details). Canadians of low education were not targeted since 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 targeted.[83]

An individual is considered to be of Aboriginal identity if they report "Aboriginal" as their cultural background, and identify with First Nations, Métis, or Inuit. New Canadians were those who immigrated to Canada within 10 years of completion of the telephone survey. Adults having completed less than a post-secondary degree, trade certificate, or diploma, were defined as being of low education. This group included individuals who are 18 years or older only. Individuals were considered to be low income if their household income was below the Low income cut-off (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.[84]

Using the 2006 Canadian census, the 100 census tracts (CTs) from within the census metropolitan areas (CMAs)[85] containing either the highest proportion of households living under the LICO (range: 41.5% to 91%) or the highest proportion of new Canadians (range: 31.9% to

66%) were selected. Individuals of Aboriginal identity were selected in the same way using a lower threshold of 15% (range: 15% to 94.6%), 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 (available via the Computing for Humanities and Social Sciences server at the University of Toronto) and Info-Direct (a company that maintains the 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, and those from Nova Scotia and Newfoundland and Labrador were excluded, from the initial selection since they were not among the top 100 in terms of proportion of low income households or new Canadians, or in the top 66 in terms of proportion of individuals of Aboriginal identity. Further, Prince Edward Island (PEI) and the three 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 55.0% from 10 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.

Participant recruitment

All households, with the exception of those in Nunavut, were mailed a letter informing them that the research team would contact them to complete a ten to fifteen minute telephone survey about dietary habits and the environment (see Appendix C). To help avoid selection bias, the letter did not mention that the study's purpose was to examine food allergy prevalence, but did advise (as required by our ethics board) that those with food allergies may have to complete a slightly longer survey. Included in the letter was a five-dollar coupon for a major restaurant chain or food product. Previous research has shown that incentives as small as five dollars provided before the survey, i.e., *a priori* incentives, increase response rates, especially among low income and minority populations.[86-88] A small pilot study, which provided a five-dollar *a priori* incentive to some households, chosen at random, and no incentive to others, was conducted prior to the beginning of data collection and confirmed previous findings.[89]

The recruitment strategy in Nunavut was different from the rest of Canada because the White Pages provides only the telephone numbers and does not provide addresses for these households.

Hence, we could not send the information letter and incentive to households in Nunavut prior to the interview. To advertise the study, a public service announcement was broadcast on a major northern Canadian news network during the period phone calls were being made to Nunavut residents. A five-dollar compensation was sent to those households after they completed the telephone survey and provided their address.

Telephone survey

Approximately two weeks after mailing the information letter, households were contacted to complete the telephone survey. The surveys were conducted by a team of similarly trained interviewers, based at McGill University in Montreal, Quebec, Canada, using Computer Assisted Telephone Interview (CATI) software (WinCati 4.2, Copyright 1986-2004 Sawtooth Technologies Inc, Northbrook, Illinois). Respondents were eligible to participate if they were eighteen years or older, were living in the household, appeared to have no cognitive or hearing barriers, could respond in either of Canada's official languages (English or French), and could answer questions about dietary habits and food allergies of all household members. Once eligibility was established, the respondent was invited to participate and asked whether any household member had an allergy to peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and/or soy, or any other foods. If the respondent reported that an individual had an allergy to one of the nine foods specified above, they were queried further using the Food Allergy Prevalence Questionnaire (FAPQ) (see Appendix D).

The FAPQ was initially developed by Sicherer et al to determine the general population prevalence of peanut, tree nut, fish and shellfish allergy in the United States,[3, 5, 6] and modified

by our team to include questions regarding sesame allergy for the SCAAALAR study.[7] In the current study, questions regarding a potential allergy to milk, egg, wheat and soy were added. As described previously by Ben-Shoshan et al,[7] individuals were queried on the history of the most severe allergic reaction, interval between exposure and symptom onset, and if the allergy was diagnosed by a physician.

Information on the age, sex, country of origin, number of years in Canada (for those not born in Canada), cultural/ethnic background (including Aboriginal identity status), education level (for those over eighteen years), and household income was obtained.

To optimize response rates and minimize selection bias, a maximum of fifteen attempts were made to contact households on different days and times between 9:00 AM and 9:00 PM (local time) Monday through Friday, and 10:00 AM and 5:00 PM on Saturdays and Sundays.

The questionnaires were translated into French and back-translated into English.

The study was approved by the Institutional Review Board of the McGill University Health Centre.

Definitions of food allergy

Two definitions of food allergy were used in this analysis:

1) Perceived food allergy; includes all individuals reporting any food allergy, and

2) Probable food allergy; a more conservative definition which includes all individuals reporting an allergy to peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and/or soy, who report a convincing history of food allergy and/or who self-report a physician-diagnosed food allergy. To be considered to have a convincing history,[90-92] an individual had to report experiencing at least two mild symptoms (pruritus, urticaria, flushing, or rhinoconjunctivitis), one moderate (angioedema, throat tightness, gastrointestinal complaints, or breathing difficulties (other than wheeze)) or one severe symptom (wheeze, cyanosis, or circulatory collapse) after ingestion or contact (or inhalation for fish, shellfish, egg or soy) within 2 hours after exposure to the food. To ensure that participants who were either lactose intolerant or had celiac disease were not mistakenly considered to have a milk or wheat allergy, those who reported either of these conditions or had symptoms which were limited to the gastrointestinal tract or who could tolerate either dairy or wheat products occasionally without experiencing a reaction were excluded from the estimates for probable milk or wheat allergy.[93, 94]

Statistical analysis

Estimating prevalence of food allergy among those completing the FAPQ and creating weighted estimates

Point estimates and 95% confidence intervals (CIs) for the prevalence of perceived and probable allergy for each of the vulnerable and non-vulnerable groups were calculated using the Clopper-Pearson exact method.[95] Given the targeted sampling strategy of this study, which purposely oversampled the vulnerable populations, the prevalence estimates were weighted. Even though

prevalence estimates were calculated for each vulnerable and non-vulnerable group separately, weighting was still necessary because the distribution of the other demographic characteristics may be distributed differently across vulnerable and non-vulnerable groups (see Appendix E).

For example, if we were interested in comparing prevalence between low income and high income individuals, it is possible that the other demographic characteristics of interest (education, immigrant status, Aboriginal status) may be distributed differently across levels of income. Hence, our groups of low and high income are neither representative of the general low income or high income population unless we account for the education, immigration, and Aboriginal population weights.

To create the weighted estimates, non-overlapping sub-groups 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. A weight variable was then created for each mutually exclusive group by dividing the number of individuals in the Census in that subgroup by the number of individuals from the SPAACE study in the corresponding subgroup. These weights were then used to calculate prevalence by specifying the weight variable that was generated, using the survey functions available in Stata 12.

As a non-trivial percentage of the sample did not report household income, a sensitivity analysis was performed where the prevalence of food allergy for those who did and did not provide their household income was compared.

Identifying predictors of food allergy

To identify predictors of food allergy, multivariate logistic regression analysis was performed for perceived allergy to any food. The following variables were included as covariates: education (<post-secondary degree versus ≥post-secondary degree; defined for adults only), household income (income < LICO versus income ≥LICO), a three-level variable for immigrant status (new Canadian, the reference group, immigrated ≥10 years ago, born in Canada), Aboriginal status (those of Aboriginal identity versus without Aboriginal identity), child (<18 years old), sex, and an interaction term between child and male, since food allergy prevalence has been shown to be higher in male children, although this trend is reversed in adulthood.[96]

A sensitivity analysis was conducted where the immigrant variable was either dichotomized as born in Canada versus immigrant, or continuous, expressing the number of years since immigrating to Canada. Additionally, a sensitivity analysis for missing income was performed where a multivariate model, which included individuals reporting their income, was compared to a model, which included those not reporting their income.

Results

Participation rate

Between September 2010 and September 2011, we attempted to reach 15,986 households by telephone of which 12,762 households were actually reached. Of the 12,762 households that were reached, 5,734 households, representing 15,022 individuals, completed the FAPQ (45% response rate). Given the targeting strategy employed, the sample consisted of a much higher percentage of vulnerable populations than are present in the general Canadian population. In the sample, 22.8% of participants were below the LICO, 11.8% were new Canadians, and 15.1% were of Aboriginal identity versus 15.7%, 7.2%, and 3.8%, respectively, of the general Canadian population.[97]

Prevalence of food allergy

Adults with low education had a lower prevalence of perceived allergy to any food than those with higher education [6.4% (95% CI, 5.5%-7.3%) versus 8.9% (95% CI, 7.7%-10%)] (Table I). This difference was most notable for tree nut. There was a trend for the perceived prevalence to be greater than the probable for most of the nine allergens. It should be noted that the prevalence of probable allergy to any food cannot be calculated as detailed history regarding allergy was collected for only nine food allergens and not for any other reported food allergen. To enable children to be included in this analysis, children were stratified based on highest educational attainment in the household and a trend towards lower prevalence in households with lower educational attainment was observed.

The perceived prevalence of tree nut and wheat allergy was lower in individuals living in households below the LICO (Table II). In a sensitivity analysis, perceived and probable prevalence estimates were similar in those reporting and not reporting household income.

New Canadians had a perceived prevalence of any food allergy of 3.2% (95% CI, 2.2%-4.3%), those who had immigrated at least ten years prior had a prevalence of 5.5% (95% CI, 4.5%-6.4%), and those born in Canada had a prevalence of 8.2% (95% CI, 7.4%-9.1%) (Table III). This difference was most notable for peanut and tree nut.

The prevalence of food allergy in individuals of Aboriginal identity was similar to the rest of the respondents (Table IV).

Socio-demographic predictors of perceived allergy

In the multivariate analysis, adults with low education [Odds Ratio (OR): 0.74 (95% CI, 0.58-0.95)] and men [OR: 0.61 (95% CI, 0.48-0.78)] were less likely to report an allergy; those born in Canada [OR: 2.80 (95% CI, 1.88-4.17)] or immigrating to Canada more than 10 years prior [OR: 1.74 (95% CI, 1.12-2.70)] were more likely to report an allergy (Table V). When the immigrant variable was dichotomized, immigrants were less likely than those born in Canada to report an allergy [OR: 0.54 (95% CI, 0.42-0.68)]; similarly, when the variable was continuous, the prevalence of perceived food allergy increased with increasing number of years since immigrating to Canada [OR: 1.02 (95% CI, 1.01-1.03)]. The predictors of perceived allergy to any food were the same in the multivariate model which was restricted to individuals who did

not report their income and in the model which was restricted to individuals who did report their income.

Discussion

SPAACE is the first Canadian study to specifically target and estimate the prevalence of food allergy in those of low education, low income, new Canadians, and individuals of Aboriginal identity. The sampling strategy used in this study was much more successful at targeting the vulnerable groups than our previous SCAAALAR study which used random sampling (households below the LICO: 22% in SPAACE versus 8.9% in SCAAALAR; new Canadians: 11.8% in SPAACE versus 1.9% in SCAAALAR), and similar strategies should be considered by others planning to conduct telephone surveys in the future.

Food allergy was less commonly reported among adults and children living in households with lower educational attainment, which may be both real and a reflection of under-diagnosis. It is possible that the more educated truly have a higher prevalence of food allergy as they may have been more likely than those with lower education to have followed recommendations suggesting that the restriction of allergenic foods early in life may prevent the development of food allergy.[12] Recent studies, however, have suggested that delayed introduction may, in fact, promote food allergy, potentially resulting in a higher prevalence in those who were more adherent to these guidelines. Consequently, this advice has since been retracted.[72] It is also possible that the lower prevalence of food allergy in those of lower education results partially from less awareness of food allergy because of lower levels of health literacy. They therefore may not recognize symptoms that may be suggestive of food allergy, and are less likely to consult a physician and be diagnosed. Although health care access is theoretically universal in Canada, differential access still exists and may contribute to under-diagnosis in the less educated.[14] Access may be limited by geographic remoteness from urban health care facilities

and by social and cultural factors.[98, 99] Others have also observed that low socioeconomic status is associated with fewer self-reported food allergies, but did not specifically target under-represented groups[52] or only included children.[54]

Immigrants were less likely to self-report food allergy and the odds of self-reporting food allergy increased by 2% for each additional year since immigrating to Canada. These findings support the "healthy immigrant effect," i.e., new Canadians tend to have a low prevalence of chronic conditions, but their health status worsens with time and eventually converges to that of the Canadian-born population.[100, 101] Additionally, many immigrants may become more aware of food allergy with increasing time in Canada, and potentially more likely to self-report. Our results are consistent with a recent American study, which reported that foreign-born children had a lower odds of having food allergy, but this study did not assess adult immigrants.[55]

Although the overall prevalence of food allergy may be hypothesized to be lower in individuals of Aboriginal identity because of larger household size, higher number of early childhood infections, and poorer sanitation, which may protect against allergic diseases,[10, 11, 13, 59] and less access to specialist health care,[57, 77-79] we observed that the prevalence was similar between those with and without Aboriginal identity. This may be because of an inadequate sample size or because our sample consisted of urban and off-reserve Aboriginal populations rather than on-reserve, where poor municipal infrastructure is more likely to be problematic. In contrast, a recent publication by our research team, using the 2006 Aboriginal Children's Survey, did demonstrate a lower prevalence among off-reserve Aboriginal children aged 0 to 5 years.[82]

Our study was limited by our inability to perform telephone interviews in languages other than English and French even though one of the targeted groups was recent immigrants. However, given the extensive ethnic diversity in Canada, it would have been logistically very difficult and expensive to translate the lengthy telephone questionnaire into multiple languages and complete the data collection within a realistic timeframe.

Our estimates of prevalence of allergy to specific foods are based on self-report of a convincing history or self-report of a physician diagnosis. In previous work, we had attempted to confirm self-report by requesting permission from participants to contact their physician and request results of diagnostic testing.[102] However, this was unsuccessful as many participants self-reporting food allergy either had not consulted a physician or refused to grant permission; in cases where participants consented, few physicians returned results. It is possible that the estimates in our study may have been lower if we required that self-report be confirmed with diagnostic testing. However, estimates for peanut allergy in Montreal school children where diagnosis was based on confirmatory testing[103] were very similar to estimates based on history alone in our previous population-based telephone survey (the SCAAALAR study).[102] Hence, estimates generated in this study by self-report of a convincing history or physician diagnosis likely should not represent a substantial overestimation.

This study suggests that those with lower education and immigrants have fewer food allergies. The difference may be real or apparent and the reasons are largely unknown. It is possible that the lower prevalence in these vulnerable populations is partially due to under-diagnosis due to their inadequate access to health care services because of geographic, bureaucratic, cultural, and

language barriers. These issues highlight important gaps in health care policy, and more research is needed to identify and address these impediments to ensure that all Canadians have an equal opportunity to seek and receive appropriate care. Indeed, our research team is undertaking in-depth studies with low-income families and new Canadians in order to explore the lived experiences of food allergies in these vulnerable populations.[82, 104]

Table I: Weighted Perceived and Probable Prevalence Estimates of Food Allergy according to education*

	A-Low education (n=5,332) % (95% CI)	B-High education (n=5,363) % (95% CI)	Difference A-B % (95% CI)
Perceived			
Peanut	0.6 (0.3, 0.9)	0.8 (0.4, 1.1)	-0.1 (-0.6,0.3)
Tree nut	0.7 (0.4, 1.0)	1.7 (1.2, 2.3)	-1.0 (-1.6,-0.4)
Fish	0.4 (0.2, 0.6)	0.8 (0.4, 1.1)	-0.4 (-0.8,0.0)
Shellfish	1.5 (1.1, 2.0)	2.2 (1.6, 2.8)	-0.6 (-1.3,0.1)
Sesame	0.1 (0.0, 0.3)	0.3 (0.1, 0.5)	-0.2 (-0.4,0.1)
Milk	0.7 (0.4, 1.1)	0.7 (0.3, 1.0)	0.1 (-0.4, 0.5)
Egg	0.6 (0.3, 0.8)	0.4 (0.2, 0.6)	0.1 (-0.2,0.5)
Wheat	0.4 (0.2, 0.7)	0.4 (0.2, 0.7)	0.0 (-0.3, 0.4)
Soy	0.1 (0.0, 0.2)	0.2 (0.0, 0.4)	-0.1 (-0.3, 0.1)
Any**	6.4 (5.5, 7.3)	8.9 (7.7, 10)	-2.4 (-3.8,-0.9)
Probable			
Peanut	0.4 (0.2, 0.6)	0.8 (0.4, 1.1)	-0.3 (-0.7, 0.1)
Tree nut	0.6 (0.3, 0.9)	1.5 (1.0, 2.0)	-0.8 (-1.4, -0.3)
Fish	0.3 (0.1, 0.5)	0.7 (0.4, 1.0)	-0.4 (-0.8, 0.0)
Shellfish	1.3 (0.9, 1.7)	2.0 (1.4, 2.5)	-0.7 (-1.3, 0.0)
Sesame	0.1 (0.0, 0.3)	0.2 (0.0, 0.4)	-0.1 (-0.3, 0.1)
Milk	0.1 (0.0, 0.3)	0.2 (0.0, 0.4)	-0.1 (-0.3, 0.1)
Egg	0.6 (0.3. 0.8)	0.4 (0.2, 0.6)	0.2 (-0.2, 0.5)
Wheat	0.3 (0.1, 0.5)	0.2 (0.0, 0.4)	0.0 (-0.3, 0.3)
Soy	0.1 (0.0, 0.2)	0.2 (0.0, 0.4)	-0.1 (-0.3, 0.1)

*A total of 15,022: 10,695 adults provided this information, 301 adults did not provide this information, and 4026 children were not asked about education.
**Any perceived allergy refers to self-report of allergy to 1 of the 9 common food allergies and other foods, such as fruit, vegetables, meat, chocolate, seeds, spices, legumes, and grains
Note: Cells where the 95% CI does not include the null are shaded in grey.

Table II: Weighted Perceived and Probable Prevalence Estimates of Food Allergy according to income*

	A-Low income (n=2,424) % (95% CI)	B-High income (n=8,205) % (95% CI)	Difference A-B % (95% CI)
Perceived			
Peanut	1.4 (0.8, 2.0)	1.2 (0.8, 1.6)	0.2 (-0.5, 0.9)
Tree nut	0.6 (0.2, 1.1)	1.6 (1.2, 2.0)	-1.0 (-1.6, -0.4)
Fish	0.4 (0.1, 0.7)	0.8 (0.4, 1.1)	-0.3 (-0.8, 0.1)
Shellfish	1.6 (0.9, 2.3)	1.9 (1.5, 2.4)	-0.3 (-1.1, 0.5)
Sesame	0.3 (0.0, 0.7)	0.2 (0.1, 0.3)	0.1 (-0.3, 0.6)
Milk	0.7 (0.3, 1.2)	0.7 (0.4, 1.0)	0.1 (-0.5, 0.6)
Egg	0.3 (0.1, 0.6)	0.7 (0.4, 0.9)	-0.3 (-0.7, 0.0)
Wheat	0.0 (0.0, 0.1)	0.3 (0.2, 0.5)	-0.3 (-0.5, -0.1)
Soy	0.1 (0.0, 0.3)	0.1 (0.0. 0.2)	-0.1 (-0.2, 0.2)
Any	7.2 (5.7, 8.6)	7.8 (6.9, 8.7)	-0.6 (-2.3, 1.1)
Probable			
Peanut	1.2 (0.6, 1.8)	1.1 (0.7, 1.4)	0.1 (-0.6, 0.8)
Tree nut	0.6 (0.1, 1.1)	1.4 (1.0, 1.8)	-0.8 (-1.4, -0.2)
Fish	0.4 (0.1, 0.7)	0.7 (0.4, 1.0)	-0.3 (-0.8, 0.1)
Shellfish	1.3 (0.7, 1.8)	1.6 (1.2, 2.0)	-0.4 (-1.1, 0.3)
Sesame	0.2 (0.0, 0.5)	0.2 (0.1, 0.3)	0.0 (-0.3, 0.3)
Milk	0.2 (0.0, 0.4)	0.2 (0.1, 0.3)	0.0 (-0.3, 0.3)
Egg	0.3 (0.1, 0.6)	0.7 (0.4, 0.9)	-0.3 (-0.7, 0.0)
Wheat	0.0 (0.0, 0.0)	0.2 (0.1, 0.4)	-0.2
Soy	0.0 (0.0, 0.1)	0.1 (0.0, 0.2)	-0.1 (-0.2, 0.0)

* Data on household income are missing for 4,393 individuals because participants refused to provide this information.
Note: Cells where the 95% CI does not include the null are shaded in grey.

1

Table III: Weighted Perceived and Probable Prevalence Estimates of Food Allergy according to immigrant status*

	A-New Canadian (n=1,754) % (95% CI)	B-Immigrant ≥10 years (n=2,851) % (95% CI)	C-Born in Canada (n=10,299) % (95% CI)	Difference A-B % (95% CI)	Difference B-C % (95% CI)	Difference (A-C) % (95% CI)
Perceived						
Peanut	0.4 (0.1, 0.7)	0.5 (0.2, 0.8)	1.3 (0.9, 1.6)	-0.1 (-0.6, 0.4)	-0.8 (-1.2, -0.3)	-0.8 (-1.3, -0.4)
Tree nut	0.2 (0.0, 0.5)	0.6 (0.2, 0.9)	1.5 (1.2, 1.9)	-0.3 (-0.8, 0.1)	-1.0 (-1.5, -0.5)	-1.3 (-1.7, -0.9)
Fish	0.4 (0.1, 0.8)	0.6 (0.2, 0.9)	0.7 (0.4, 1.0)	-0.1 (-0.6, 0.3)	-0.2 (-0.6, 0.3)	-0.3 (-0.7, 0.2)
Shellfish	1.3 (0.6, 1.9)	1.5 (1.0, 2.0)	1.8 (1.4, 2.2)	-0.2 (-1.0, 0.6)	-0.3 (-1.0, 0.3)	-0.6 (-1.3, 0.2)
Sesame	0.2 (0.0, 0.4)	0.1 (0.0, 0.1)	0.2 (0.1, 0.4)	0.2 (-0.1, 0.4)	-0.2 (-0.3, -0.1)	0.0 (-0.3, 0.2)
Milk	0.4 (0.0, 0.7)	0.5 (0.2, 0.8)	0.8 (0.5, 1.0)	-0.2 (-0.6, 0.3)	-0.3 (-0.7, 0.2)	-0.4 (-0.9, 0.0)
Egg	0.4 (0.1, 0.7)	0.6 (0.3, 1.0)	0.6 (0.4, 0.8)	-0.2 (-0.7, 0.3)	0.0 (-0.4, 0.4)	-0.2 (-0.6, 0.2)
Wheat	0.0 (0.0, 0.0)	0.5 (0.2, 0.8)	0.4 (0.2, 0.6)	-0.5	0.1 (-0.3, 0.5)	-0.4
Soy	0.1 (0.0, 0.3)	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)	0.0 (-0.2, 0.2)	0.0 (-0.2, 0.1)	0.0 (-0.2, 0.2)
Any	3.2 (2.2, 4.3)	5.5 (4.5, 6.4)	8.2 (7.4, 9.1)	-2.2 (-3.7, -0.8)	-2.8 (-4.1, -1.5)	-5.0 (-6.3, -3.7)
Probable						
Peanut	0.4 (0.1, 0.7)	0.4 (0.1, 0.7)	1.1 (0.8, 1.4)	-0.1 (-0.5, 0.4)	-0.7 (-1.2, -0.3)	-0.8 (-1.2, -0.4)
Tree nut	0.2 (0.0, 0.4)	0.4 (0.1, 0.7)	1.4 (1.1, 1.7)	-0.2 (-0.6, 0.2)	-1.0 (-1.5, -0.6)	-1.2 (-1.6, -0.8)
Fish	0.4 (0.1, 0.8)	0.5 (0.2, 0.9)	0.6 (0.4, 0.9)	-0.1 (-0.6, 0.4)	-0.1 (-0.5, 0.3)	-0.2 (-0.6, 0.2)
Shellfish	1.1 (0.5, 1.7)	1.2 (0.8, 1.7)	1.5 (1.2, 1.8)	-0.2 (-0.9, 0.6)	-0.3 (-0.8, 0.3)	-0.4 (-1.1, 0.3)
Sesame	0.2 (0.0, 0.4)	0.0 (0.0, 0.1)	0.2 (0.1, 0.3)	0.1 (-0.1, 0.3)	-0.2 (-0.3, 0.0)	0.0 (-0.3, 0.2)
Milk	0.2 (0.0, 0.4)	0.2 (0.0, 0.4)	0.2 (0.1, 0.3)	-0.1 (-0.3, 0.2)	0.1 (-0.2, 0.3)	0.0 (-0.2, 0.2)
Egg	0.4 (0.1, 0.7)	0.6 (0.2, 1.0)	0.6 (0.4, 0.8)	-0.2 (-0.7, 0.3)	0.0 (-0.4, 0.4)	-0.2 (-0.6, 0.2)
Wheat	0.0 (0.0, 0.0)	0.3 (0.1, 0.6)	0.3 (0.1, 0.4)	-0.3	0.1 (-0.2, 0.4)	-0.3
Soy	0.1 (0.0, 0.3)	0.1 (0.0, 0.2)	0.1 (0.0, 0.2)	0.0 (-0.2, 0.2)	0.0 (-0.2, 0.1)	0.0 (-0.2, 0.2)

*Immigrant status was only available for 14,904 participants.

Note: Cells where the 95% CI does not include the null are shaded in grey.

Table IV: Weighted Perceived and Probable Prevalence Estimates of Food Allergy according to Aboriginal identity*

	A-Aboriginal (n=2,265) % (95% CI)	B-Non-aboriginal (n=12,732) % (95% CI)	Difference A-B % (95% CI)
Perceived			
Peanut	1.2 (0.0, 2.4)	1.1 (0.8, 1.4)	0.1 (-1.2, 1.3)
Tree nut	0.7 (0.0, 1.7)	1.3 (1.0, 1.6)	-0.6 (-1.6, 0.4)
Fish	1.4 (0.1, 2.6)	0.7 (0.4, 0.9)	0.7 (-0.6, 2.0)
Shellfish	2.1 (0.5, 3.6)	1.7 (1.4, 2.1)	0.3 (-1.3, 2.0)
Sesame	0.4 (0.0, 1.1)	0.2 (0.1, 0.3)	0.2 (-0.5, 0.9)
Milk	0.6 (0.0, 1.2)	0.7 (0.5, 0.9)	-0.1 (-0.8, 0.5)
Egg	0.7 (0.0, 1.5)	0.6 (0.4, 0.8)	0.1 (-0.8, 0.9)
Wheat	0.2 (0.0, 0.5)	0.4 (0.2, 0.5)	-0.2 (-0.6, 0.2)
Soy	0.0 (0.0, 0.1)	0.1 (0.1, 0.2)	-0.1 (-0.2, 0.0)
Any	8.5 (5.3, 11.6)	7.4 (6.7, 8.1)	1.1 (-2.2, 4.3)
Probable			
Peanut	1.1 (0.0, 2.4)	1.0 (0.7, 1.2)	0.2 (-1.1, 1.4)
Tree nut	0.7 (0.0, 1.6)	1.2 (0.9, 1.5)	-0.5 (-1.4, 0.5)
Fish	1.0 (0.0, 2.2)	0.6 (0.4, 0.8)	0.4 (-0.7, 1.6)
Shellfish	2.1 (0.5, 3.6)	1.4 (1.1, 1.7)	0.6 (-1.0, 2.2)
Sesame	0.4 (0.0, 1.1)	0.2 (0.1, 0.3)	0.2 (-0.5, 0.9)
Milk	0.0 (0.0, 0.1)	0.2 (0.1, 0.3)	-0.2 (-0.3, 0.0)
Egg	0.7 (0.0, 1.5)	0.6 (0.4, 0.8)	0.1 (-0.7, 0.9)
Wheat	0.2 (0.0, 0.5)	0.2 (0.1, 0.4)	-0.1 (-0.4, 0.3)
Soy	0.0 (0.0, 0.1)	0.1 (0.1, 0.2)	-0.1 (-0.2, 0.0)

*Aboriginal identity was available for 14,997 individuals

Table V: Socio-demographic predictors of perceived allergy to any food

	Model 1 OR (95% CI)	Model 2 OR (95% CI)	Model 3 OR (95% CI)
Predictors			
Low education	0.74 (0.58-0.95)	0.75 (0.58-0.96)	0.73 (0.57-0.94)
Low income	1.00 (0.76-1.31)	0.98 (0.75-1.28)	1.00 (0.77-1.31)
Immigrated ≥10 years ago	1.74 (1.12-2.70)		
Born in Canada	2.80 (1.88-4.17)		
Immigrant to Canada		0.54 (0.42-0.68)	
Years since immigration			1.02 (1.01-1.03)
Aboriginal identity	0.96 (0.58-1.59)	0.95 (0.58-1.57)	0.95 (0.58-1.56)
Female Child	0.83 (0.55-1.24)	0.81 (0.54-1.21)	0.84 (0.56-1.25)
Male Adult	0.61 (0.48-0.78)	0.61 (0.48-0.78)	0.61 (0.48-0.78)
Male Child	1.78 (1.11-2.87)	1.79 (1.11-2.87)	1.78 (1.11-2.87)

All three models contained the following variables: education, household income, Aboriginal status, child, sex, and an interaction term between child and male (reference group: female adult). These models differed in terms of the definition of immigrant status, as follows:
Model 1 contained a 3-level variable for immigrant status (new Canadian, the reference group, immigrated ≥10 years ago, born in Canada),
Model 2 contained a dichotomous variable for immigrant status (born in Canada, the reference group, versus immigrant), and
Model 3 contained a continuous variable for immigrant status, expressing the number of years since immigrating to Canada.
Note: Cells where the 95% CI does not include the null are shaded in grey.

IV: LITERATURE REVIEW-NON-RESPONSE BIAS IN TELEPHONE SURVEYS

Introduction

Given the importance of population-based estimates of food allergy prevalence to characterize the burden of disease, many researchers have relied on telephone surveys for data collection. This methodology allows a large, geographically diverse population to be included; however, non-response is common in telephone surveys and can lead to possibly biased inferences. This literature review will summarize previous research on non-response and non-response bias. Specifically, study design techniques to minimize non-response bias and data analysis techniques to adjust for non-response bias will be described.

Non-response

Non-response means failure to obtain a measurement on one or more study variables selected for a survey.[105] There are two types of non-response: 1) Unit non-response, where an individual does not respond, and 2) Item non-response, where an individual responds to the questionnaire incompletely.[106] In recent years, telephone survey response rates have declined, with many studies reporting rates lower than 50%.[107] Several reasons for the decline in response rates for telephone surveys have been proposed, including the availability of caller identification and answering machines,[108] as well as a general disinterest in answering telephone surveys due to the overwhelming number of market research surveys.[109] With such low response rates, the presence of bias due to non-response cannot be ignored.[110]

Non-response bias

Non-response bias is a form of selection bias that occurs when participants in the study are systematically different from non-participants.[111] In other words, if the distribution of characteristics, whether known or unknown, of the individuals participating in a study differ from those of the individuals not participating, and if these characteristics are associated with the outcome of interest, then there exists the potential for bias to occur.[107] For example, research has shown that for certain types of studies non-responders tend to be less healthy, are less likely to use the healthcare system, are younger, male, living alone, unmarried, and have a lower educational attainment than those who respond.[112-114] Differences in participation rates across certain segments of the population reduce the possibility to generalize study observations to the total population, and this can lead to biased estimates of the association between the outcome and other variables of interest.[113]

Non-response bias in telephone surveys

People tend to participate more often in research that directly or indirectly affects them or their loved ones. Hence, telephone surveys to assess the prevalence of food allergy may be biased because those who are more aware of their food allergies may be more likely to participate.[49] There is also a general tendency for those of higher socioeconomic status (SES) and are Canadian-born to participate. Therefore, the prevalence of food allergy obtained from a telephone survey may be biased in the direction of more allergies in high socioeconomic strata. Our previous Canadian study found that those who participated in the study (35% of those contacted) were predominantly of high SES and born in Canada.[7] Although these groups reported the highest prevalence of food allergy both overall[8] and for specific foods,[17] this finding

may be (at least in part) explained by non-response bias. We acknowledged the possibility of non-response bias in this study, but we did not go any further to attempt to quantify this bias or quantitatively explore the impact of non-response bias on prevalence estimates.

Another recent study to estimate food allergy prevalence among American adults reported a 35.8% response rate for a nationally-representative, random digit dialing telephone survey.[52] Although the authors mention that weights were used to adjust for oversampling of certain groups in the study, it is unlikely that these completely eliminated non-response bias. Sicherer's study to estimate the prevalence of peanut, tree nut, and sesame allergy in the United States mentioned that their study likely had similar limitations to other telephone surveys and noted that individuals participating in the survey were not aware that the study was about food allergy.[4] They do not elaborate on why this is an important point, although presumably, they are alluding to non-response bias.

Although many prevalence studies, including those mentioned here, discuss non-response bias as a limitation of their study, most do not attempt to characterize the magnitude or direction of bias.[3-6, 52] Before presenting the various strategies that can be used to minimize or adjust for non-response bias, other limitations of telephone surveys that can lead to biased results will be described.

Other limitations of telephone surveys

In addition to non-response bias, there are other issues with using telephone surveys that will be briefly described here. Non-coverage bias is a term that has been used to describe the problem of

landline users not necessarily being representative of the general population, possibly leading to bias if only landline users are included in telephone surveys.[115] Researchers have concluded that the assumption that landline users are representative of the general population is no longer valid.[116-120] Although the inclusion of cellular phone users in telephone surveys is likely to be the way of the future,[121-123] the cost of performing telephone surveys with cellular phones are much higher than with landline phones,[120] and the response rate has been shown to be lower with cellular phones.[120, 121, 124] In addition, there are issues with privacy and participants being charged for incoming calls, which warrant further exploration.

Bias due to reliance on self-report of the disease of interest is another issue with telephone surveys. Most studies, including the current one, rely on self-report of the disease of interest to estimate prevalence because it is difficult to obtain confirmatory testing for diseases reported by telephone survey participants.[7] In particular, prevalence estimates based on self-report are often higher than those based on stricter criteria such as the requirement of a clinical history compatible with an IgE-mediated reaction combined with a positive diagnostic test result.[125, 126]

While non-coverage bias and self-report bias are important limitations of telephone surveys, the main purpose of this thesis is to obtain unbiased estimates of allergy prevalence. Self-report bias was addressed in Chapter II, in the way the questions on food allergy prevalence were designed, which allowed us to differentiate between self-report, convincing history of an allergic reaction, and doctor diagnosis of food allergy. Non-response bias was considered to be more serious than coverage bias here, since non-response rates were so high. The following sections will

summarize methods for handling non-response in the design, data collection, and analysis stages of a study.

Minimizing non-response bias in the design and data collection stages

Several researchers have set out a number of detailed strategies for minimizing the number of non-contacts and refusals in surveys. These include both questionnaire and interview techniques such as the use of short, personalized letters sent prior to the interview.[127] Inclusion of a study brochure with frequently asked questions like "Why is the study important?", "Who is being asked to participate?" and "How will the interviews be conducted?", and answers to these questions also increases the chance that an individual will participate in a telephone survey.[128] Incentives for cooperation have a positive effect on the response rate, as has the importance of the survey's topic. Previous research has shown that incentives as small as five dollars provided before the survey, i.e., *a priori* incentives, increase response rates, especially among low income and minority populations.[86-88] A pilot study performed prior to data collection for the SPAACE study confirmed this.[89]

The use of a short Refusal Questionnaire in longer surveys has proven to be useful for collecting data from non-responders on a small number of the most important study variables. For example, a study on osteoporosis collected the most important risk factors from people who did not want to fully participate in the study, and then used these variables to predict the probability of osteoporosis in non-complete respondents.[129] This method worked quite well, with an additional 30% participation rate for the shorter questionnaire. However, in other cases, participants do not even stay on the phone long enough to be asked the Refusal Questionnaire. In the SPAACE

study, only 4% of households who refused to complete the full questionnaire agreed to complete the Refusal Questionnaire.[130]

Another way to elicit information from non-responders is through the administration of a questionnaire following the initial survey. Rupp et al[112] administered a telephone survey among non-responders to a mailed questionnaire about rheumatoid arthritis six months following the initial questionnaire, to determine whether there were any differences in terms of sociodemographic characteristics (age, gender, type of household, marital status, and educational level) and health status (disease duration, pain, co-morbidity, general health, functional status, etc) between responders and non-responders, and to increase their response rates. Siemiatycki[131] and Fowler,[132] among others, have also used this strategy to increase response rates. However, these studies all used mailed questionnaires as the initial survey tool, and then followed up with a telephone survey. Therefore, it is unclear whether a follow up telephone survey would have the same benefit when a telephone survey was used as the initial survey tool. In addition, there can be ethical issues to repeatedly contacting households that do not respond or have explicitly refused to respond.

There are several ways that researchers can try to minimize non-response, as summarized in the paragraphs above. Unfortunately, no method is perfect, and it is almost guaranteed that any study will have some level of non-response. Therefore, when designing a study, it is important to not only include details on how researchers will attempt to increase the response rate during the data collection phase, but also to include details on how researchers will attempt to adjust for or

measure the bias due to non-response in the data analysis phase. Different statistical methods for adjusting for non-response bias will be discussed in the next section.

Adjusting for non-response bias in the analysis stage

Non-response creates a missing data problem whereby some or all study variables are missing.[106] There are different ways to analyze missing data, and deciding which statistical technique is appropriate requires one to first determine what type of missing data is present in a given dataset. The next section will discuss the types of missing data for the non-expert. Those interested in more rigorous definitions should consult a textbook on the subject, such as that by Rubin.[133]

Types of missing data

Missing data can be classified as: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).[134] As the name implies, MCAR data are equivalent to having complete data, and then randomly deleting certain entries, regardless of their values. This, in turn, implies that MCAR data will, on average, return unbiased inferences, similar to a complete dataset but with lower sample size.[106] When missing data are MCAR, then simply ignoring the missing data and performing analysis on complete data will yield an unbiased point estimate, albeit with wider confidence intervals compared to the complete dataset with no missing data.

Roughly speaking, MAR missing data implies that any biases created by the missing data can be adjusted for using information that is contained within the observed data.[106] Unfortunately, this

condition is generally unverifiable because, of course, we do not know the values of missing data. Hence, we cannot directly compare the inferences that are obtained with and without the missing data.

MCAR and MAR data are termed ignorable, since in either case valid inferences can be obtained from information in the data alone.[135] On the other hand, MNAR data are non-ignorable, and valid inferences become more difficult to obtain, often depending on outside information, if available. Roughly, MNAR missing data implies that, even after accounting for all of the information in the observed dataset, inferences may still be biased.[106]

Missing data can be analyzed in different ways, and each of the strategies to analyze missing data has its own strengths and weaknesses. Some of these methods will be discussed in the next section.

Conventional methods for analyzing missing data

The conventional methods for analyzing missing data include list-wise deletion, maximum likelihood methods, and various methods based on imputation, or filling in the missing data with estimated values. These methods, including their benefits and drawbacks, will be described in the following paragraphs.

List-wise deletion, also known as complete case analysis, deletes any cases where there is missing data on one or more variables of interest. Advantages of this technique include the

ability to use any standard statistical analysis without special methods, and the ability to compare univariate statistics since these are all based on the same sample size. This technique is only valid if subjects with complete data are representative of the study population in terms of inferences of interest. For example, if the missing data are MCAR, then complete case analysis will result in unbiased inferences, albeit with reduced precision owing to smaller sample sizes. The main drawback is that the method produces unbiased inferences only under very limited circumstances, i.e., that the missing data actually satisfies the MCAR assumption. It is most often the case that those with missing data are at least somewhat different from those with complete data, and the MCAR assumption is violated. As the number of cases with missing data increases, the method can become very inefficient because all cases with any missing data will be discarded, thereby increasing the standard errors of estimates.[136]

Single imputation involves replacing missing values in the dataset with some reasonable guess or more formal estimate from a model, and then performing statistical analysis as if there were no missing data. One obvious advantage of single imputation is that once the missing data points are filled in, standard complete-data analysis can be performed.[133] Another advantage is that imputation allows for the incorporation of outside information into the final inferences. For example, if the analyst has substantial knowledge about the reasons for non-response, this information can be considered in the imputation procedure.[133] The main disadvantage is that once the data have been filled in, analysis proceeds as if all data were known, which leads to variance estimates that are too low. In addition, even when the non-response mechanism is poorly understood, this uncertainty is not taken into account in the imputation model.

There are several ways to impute missing values. One of the simplest methods is called marginal imputation and involves filling in missing values with the mean of the observed values for each variable.[106] This method is known to yield biased estimates with too little variation, since not all subjects with missing data will have values close to the mean.[133]

Conditional mean imputation is used when analyses are based on means, variances, and correlations. It involves estimating the mean and variance of a variable using all cases that respond to that variable, and estimating the correlation between two variables using all cases that respond to both variables.[133] This method yields reasonable estimates for means if the normality assumption is plausible and missing data are MAR. However, the size of the variance and co-variance are generally underestimated, and negative variances can sometimes result.

'Hot deck' imputation involves finding a matching respondent for each non-respondent, where matching means that the two respondents are close with respect to the observed variables.[133] Matching criteria are determined by the analyst, and many trials might be run using different criteria to ensure that every respondent is matched with a non-respondent. Unfortunately, this method also underestimates variability because it treats the imputed values as if they were known with certainty.[133] In addition, it does not take into account the mechanisms by which data come to be missing.

All of the simple methods discussed above for salvaging information from cases with missing data are sub-optimal. List-wise deletion can introduce substantial bias if the data are not MCAR. Single imputation may adjust for bias but produces standard error estimates that are too low.

None of the methods covered so far are satisfactory, and two other techniques have been developed that work better: maximum likelihood and multiple imputation. Maximum likelihood and multiple imputation will be described in the next section.

Maximum likelihood

The goal of this literature review is not to present substantial statistical details of techniques to analyze missing data, especially for those not used in the rest of the thesis. Therefore, Maximum likelihood (ML) will only be briefly described here. ML estimation chooses as estimates those values that, if true, would maximize the probability of observing the data that has been observed. ML estimators have a few very appealing properties: they are consistent, meaning that they are approximately unbiased in large samples, and they are efficient, meaning that the standard errors are at least as small as the standard errors for any other estimator.[106] These properties hold under some general conditions, including that the missing data are MAR.[136] Unfortunately, since ML is a large sample inferential approach, when the sample size is small, the likelihood function may have a non-normal shape, maybe with local maxima, and asymptotic theory may not work very well.[106] Although some simpler models can be handled by standard software packages, implementation may be less straightforward or even impossible for more complex models. ML requires a model for the joint distribution of all variables with missing data, which can be quite difficult to obtain and which complicates the estimation process.[106] Further, it cannot be extended to handle missing data that are non-ignorable. Therefore, this method was not used in this thesis, where missing data may well be non-ignorable.

Multiple imputation

Under certain conditions described by Rubin, Multiple Imputation (MI) retains the optimal properties of ML-estimates that are consistent, efficient, and asymptotically normal when data are MAR, but eliminates some of the limitations. Unlike ML, MI can easily be used with any kind of data and any model.[106] In other words, if there is a way to analyze the complete dataset, then MI can be used, regardless of model complexity. Another advantage of MI over other methods is how the calculation of variance is performed. Whereas other methods yield variances that are usually too small because the data are essentially considered to be complete, MI takes into account this uncertainty and yields a final variance estimate that is apropriately adjusted.[137] Unlike other methods, MI also allows the sensitivity of inferences to various models of non-response to be investigated. This is an important feature of MI because, in general, the statistical analysis that is performed usually relies on assumptions that are unverifiable.[133] For all of these reasons, MI is considered to be the "gold standard" for analyzing missing data.

Multiple Imputation involves the following steps:[129]

1) The parameter of interest, say theta, is identified. This parameter, which could be a vector, includes all unknown quantities of interest (prevalences, odds ratios, regression coefficients, and missing data). Let x be the observed data, and y be the missing data.
2) A prior distribution for theta is specified. The prior distribution summarizes what is known about the parameter prior to collection of new data. If there is little prior information, a non-informative or diffuse prior is used, and the data themselves drive final results.[106] The user may choose to include information in the prior distribution if

there is information available on the missing data mechanism. In this case, the results obtained will be formed from a combination of the data and the prior information.

3) The distribution of the data, y, given the parameter value, is then specified as a likelihood function, f(y|theta).
4) The posterior distribution is determined using Bayes theorem, which states that the posterior distribution is proportional to the prior times the likelihood. The posterior distribution summarizes the knowledge about the unknown parameter theta given the information contained in the data (as represented by the likelihood function) and the prior information.
5) The missing data is imputed by drawing from the distribution for the missing data y, given the observed data x, and unknown parameters.
6) Now that the missing data has been "filled in", the desired analysis can be performed on each complete dataset separately. Each imputed dataset uses information from the previous datasets, thereby "updating" the estimates of the model parameters, ensuring that final estimates take into account uncertainty of the parameter estimates.
7) The average of the point estimate over all datasets is taken as the final result, and the variance is calculated as the sum of the within and between imputation variances.

As with all methods, there are drawbacks to imputation as well. Specifically, the process of replacing missing values with a suitable estimate can itself create bias in the resulting estimates when the user applies an incorrectly specified informative prior distribution to the missing data. Another problem with MI is that the analyst must model the distribution of every variable with

316 missing data, and this process can introduce bias if the models used are incorrect.[138] MI is useful
317 for situations where missing data are MAR and hence the missing data is ignorable.
318
319 Unfortunately, techniques to adjust for missing data that do not satisfy MAR and are thus non-
320 ignorable often yield biased results. Performing a sensitivity analysis examining the effect of
321 different assumptions about the missing data mechanism is one way to proceed when missing
322 data do not satisfy MAR or are non-ignorable.[133] A common technique is to consider a range of
323 plausible assumptions about the parameter of interest, and perform the MI analysis using each of
324 these assumptions to see the effect on final inferences.[133] Results from MI, like with all other
325 methods, must be interpreted with caution, because it is impossible to know exactly the missing
326 data mechanism, and hence, all one can do is make the best educated guess possible.
327
328 Summary of methods
329 We have discussed the various simple methods for handling missing data and concluded that
330 they are wrought with issues. Most of these issues have been addressed by more complex
331 analytical techniques such as maximum likelihood and multiple imputation, the latter being
332 preferred because it can handle any statistical model. As was pointed out in the previous
333 paragraph, when missing data are not MAR and are non-ignorable, MI with sensitivity analysis
334 seems to be the best approach, but results must be interpreted with caution. We have therefore
335 chosen to use this method for the thesis. Further details are provided in Chapter XX.
336

Conclusions and future directions

In this literature review, we have discussed the important problem of non-response bias in telephone surveys and the various strategies that can be used to minimize or adjust for this bias at the design, data collection, and analysis stage.

Many of the strategies for minimizing non-response were implemented during the design and data collection phase of SPAACE. We tried to increase the response rate by providing a five dollar incentive and an information letter to all households chosen to participate in the survey, and households were contacted multiple times, on different days and times.[38] A refusal questionnaire was used to collect information from households who refused to participate in the full telephone survey.[130]

At the analysis stage, multiple imputation was used to create a range of plausible prevalence estimates that adjusted for non-response bias using different assumptions about the prevalence in the non-responders. In Chapter V, data on non-response-adjusted prevalence estimates of food allergy will be presented in the form of a manuscript accepted in JACI: In Practice.

V: ADJUSTING FOR NON-RESPONSE BIAS CORRECTS OVERESTIMATES OF FOOD ALLERGY PREVALENCE

Lianne Soller, Moshe Ben-Shoshan, Daniel W. Harrington, Megan Knoll, Joseph Fragapane, Lawrence Joseph, Yvan St. Pierre, Sebastien La Vieille, Kathi Wilson, Susan J. Elliott, and Ann E. Clarke.

Introduction

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%.[8, 52] However, such studies-are often limited as they provide prevalence estimates for a limited number of allergies[4, 7] and do not consider non-response bias,[4, 6-8, 52, 53] which may result in an over-representation 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 non-response bias on prevalence by presenting a range of estimates using different assumptions about food allergy prevalence among non-responders.

Methods

Survey methodology

The SPAACE study was a random cross-Canada telephone survey conducted between September 2010 and 2011, which targeted vulnerable Canadians (i.e., those of low income, New Canadians, and of self-reported Aboriginal identity) using 2006 Canadian Census data (see Chapter III).[38, 89] 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.[38] Food allergy was defined as:

1) Perceived: individuals self-reporting any food allergy, and
2) Probable: individuals self-reporting a convincing history[91, 92] 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 (see Appendix F).

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 (see Chapter III for more details).[38] Credible intervals are the Bayesian analogue to standard confidence intervals.

Developing non-response bias estimates

To develop non-response bias-adjusted estimates of prevalence of perceived allergy to any food, four groups were identified:

1) *Full Participants*: households who completed the FAPQ,
2) *Refusal Questionnaire (RQ) Participants*: households who completed the RQ only,
3) *Non-Participants*: households that were reached by telephone but refused to complete either questionnaire, and
4) *Never Reached Participants*: households that could not be reached by telephone.

Food allergy data are available only from *Full* and *RQ Participants*. Multiple imputation (MI), the gold standard for adjusting for missing data,[137] was used to adjust the estimates for non-response bias resulting from missing food allergy data within the *Non-Participants* 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.[129]

A range of assumptions regarding the prevalence of food allergy in the *Non-Participants* and *Never Reached Participants* were investigated. Compared with the prevalence in the *RQ Participants* living in the same CT, the prevalence in the *Non-Participants* 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 *Non-Participants*, 2) a weighted average of the *Full*, *RQ*, and *Non-Participants*, and 3) equal to the *Full Participants* (see Appendix G for more details)

MI was implemented via a hierarchical logistic regression model with four 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) (see to Appendix H).

Results

Participation rate

We telephoned 17,337 households of which 14,113 were actually reached. Of these 14,113 households, 1,351 were ineligible due to a language barrier or unavailability of an adult. Of the 12,762 eligible households, 5,734 households, representing 15,022 individuals, completed the FAPQ (45% response rate, or 5,734 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 6,504 households answered the telephone but refused to provide any information (51%) and were thus *Non-Participants*. An additional 3,224 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 VI).

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%) versus 2.1% (1.4%, 2.9%)] (Table VII). Applying the different assumptions regarding the prevalence of food allergy among the *Non-Participants* and *Never Reached Participants*, nine 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 VII and Appendix G).

Discussion

Comparison with previous studies

The unweighted perceived prevalence of food allergy in this study [6.4% (6.0%, 6.8%)] was less than in our general population study conducted 2 years earlier [8.1%, 7.5%, 8.7%)][8] 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%)].[66] The NHANES survey is weighted for non-response in general, but this weighting may not be sufficient to account for all possible non-response bias.[66] 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%)].[53] 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.[4, 52] In fact, research has shown that the majority of telephone surveys report response rates below 50%.[107] In addition, the information letter sent to participants prior to 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 non-responders. Finally, we had to impute the number of individuals in non-allergic 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 non-response 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 which must be addressed. We explored a range of assumptions for the prevalence of food allergy among *Non-Participants* 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 non-response bias can substantially influence prevalence, and ignoring bias could result in an overestimation. Our research highlights the importance of minimizing non-response

bias in designing a study, while acknowledging that bias is likely present and should be considered when performing the analysis.

Table VI: Weighted Perceived and Probable Prevalence Estimates of Food Allergy by age group

	Children, Under 18 (n=4,026) **% (95% CrI)**	**Adults, 18 and over** (n=10,996) **% (95% CrI)**	**All ages** (n=15,022) **% (95% CrI)**
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 only collected detailed information about food allergy to the nine common foods; therefore, probable estimates for other foods and any food could not be calculated.

Table VII: Non-adjusted and bias-adjusted prevalence estimates of perceived allergy to any food

	NON-ADJUSTED		BIAS-ADJUSTED			
Estimate Number	**Full Participants, FP** (n=15,022) **% (95%CrI)**	**Refusal Questionnaire Participants, RQP** (n=1,393*) **% (95%CrI)**	**Non-Participants, NP** (n=17,059*) **% (95%CrI)**		**Never Reached Participants, NRP** (n=8,419*) **% (95%CrI)**	**All participants** (n=41,893) **% (95%CrI)**
					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 non-allergic 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.

500

VI: LITERATURE REVIEW-PRESCRIPTION AND AVAILABILITY OF EPINEPHRINE AUTO-INJECTORS

Introduction

Although immune modulatory therapies for food allergy appear promising, these likely will not induce long term tolerance,[42] and food allergy will remain largely incurable. Those affected must rely on strict avoidance of the offending food and rescue therapy with epinephrine. Failure to administer epinephrine promptly after suspected ingestion of a food allergen can have severe and even fatal consequences.[139]

Given that symptoms of anaphylaxis (a severe allergic reaction) can become life threatening quite quickly, guidelines regarding the importance of an appropriate diagnosis and prescription of the epinephrine auto-injector (EAI) have been published in many countries, including the United Kingdom,[139] Europe,[109] the United States[43] and Canada.[108] These guidelines address the need for a management and prevention plan for patients with food allergy. Specifically, avoidance of the food allergen and nutrition counseling are recommended. In addition, age and culturally-appropriate information on food allergen avoidance and emergency management of allergic reactions should be provided, and a prescription for the EAI,, instructions on its use, and the importance of having it readily available at all times, should be given at the time of diagnosis.

There is much evidence to suggest that these recommendations are often not practiced; a substantial proportion of individuals who report food allergy have not been diagnosed by a physician and are therefore not equipped with the knowledge or the tools necessary to prevent or treat an allergic reaction. In fact, a national survey from the United States found that 74% of children and only 44% of adults with peanut and/or tree nut allergy sought a diagnosis for food allergy, and that less than half of these were given a prescription for an EAI.[5]

Unfortunately, even those individuals who visit a healthcare professional for their suspected food allergy do not receive adequate information regarding the importance of avoiding the offending food or a prescription for the EAI. Our previous population-based telephone survey found that only 55% of Canadians diagnosed with peanut, tree nut, fish, shellfish and/or sesame allergy self-reported having an EAI, and adults, males, and those residing in households where the respondent was single were even less likely to have one.[36] However, this study was limited as it under-represented some of the most vulnerable populations-those of lower socioeconomic status, new Canadians, and individuals of Aboriginal identity. Further, we did not query patients on actual availability of the EAI nor on other common allergies-milk, egg, wheat, or soy.

This review will first consider the importance of epinephrine in the treatment of anaphylaxis, and current evidence regarding the prescription and availability of the EAI will be critically reviewed. Finally, this review will discuss how the current SPAACE study addresses the gaps in the literature on prescription and availability of the EAI, particularly with respect to vulnerable populations in Canada.

The importance of epinephrine in the treatment of anaphylaxis

Since the 1960s, epinephrine has been the treatment of choice for anaphylactic reactions in the community due to its ability to decrease the release of inflammatory mediators, which play a role in anaphylaxis.[43, 140, 141] However, studies have shown that many allergic reactions are left untreated, are treated with other therapies which are not as effective as epinephrine, such as antihistamines or steroids,[142, 143] or are treated too late, resulting in death in some cases.[144-146] A recent systematic review by Canadian researchers identified the infrequent treatment of allergic reactions with epinephrine as a major gap in anaphylaxis management.[147] One study included in this review surveyed daycares in the suburbs of Chicago and found that only 24% of centers would administer the EAI for a severe allergic reaction, even though each center had an average of seven children with food allergies.[148] Child care employees are not routinely provided with training regarding food allergy or administration of epinephrine. Rather, parents are left with the responsibility to provide the EAI and training on when and how to use it. It is alarming that child care centers are not permitted to keep emergency kits containing epinephrine on site and that very few child care workers know how and when to administer epinephrine, especially since children rely on their caregivers to administer the EAI in case of an allergic reaction.

A multicenter study involving twenty-one North American emergency departments found that only 19% of all patients admitted to hospital for a food-induced allergic reaction and only 24% of patients admitted for a severe reaction were treated with epinephrine.[33] Even more worrisome is that between 1993 and 2004, the use of epinephrine for allergic reactions in emergency departments in the United States decreased from 19% to 7%.[149] Recent data from an urban adult tertiary care emergency department in Montreal, Canada found that epinephrine was not

administered in almost half of moderate-to-severe cases of anaphylaxis.[150] Children experiencing severe anaphylactic reactions who presented to an emergency department in Montreal were all given epinephrine either outside or inside the hospital, but only three-quarters of moderate reactions and two-thirds of mild reactions received epinephrine.[151] The situation in Europe is not much better, with data from an anaphylaxis registry in Germany, Switzerland and Austria demonstrating that only 13.8% of anaphylaxis cases received epinephrine, while 50.1% received antihistamines and 51.3% corticosteroids.[152]

Timing of administration of epinephrine

Studies have shown that rapid administration of epinephrine after the onset of an anaphylactic reaction can be life-saving. Sampson documented deaths and near-deaths in children and adolescents caused by accidental exposure to a known food allergen, and found that only 2 of the 6 patients who died received epinephrine in the first hour following the onset of symptoms, but neither received it before the onset of severe symptoms.[145] Of the 7 who survived, all but 1 patient received epinephrine before the onset of severe symptoms. In Australia, none of the individuals who experienced a fatal food-induced anaphylactic reaction had been prescribed the EAI; epinephrine was only administered later when the patient was brought to the hospital.[153]

Yunginger assessed adults who experienced fatal anaphylactic reactions due to food, and concluded that the primary reason for these deaths was failure to administer epinephrine immediately after the onset of symptoms.[154] Bock, Pumphrey, and Greenberger also attributed fatal episodes of anaphylaxis to delayed administration of epinephrine.[32, 146, 155] These studies all

came to the same conclusion: individuals who receive epinephrine early are less likely to experience a fatal reaction than those who receive it late or not at all.

Prescription of the EAI

We have seen that epinephrine is the treatment of choice for food-allergic reactions, and prompt administration of the EAI can be life-saving. Unfortunately, many patients are not prescribed the EAI by their physician and hence do not have it available in case of an allergic reaction in the community. Studies have shown that emergency physicians often discharge patients following an anaphylactic reaction without a prescription for an EAI, education regarding avoidance of the suspected food allergen, or a referral to an allergist.[156, 157]

Even more surprising is that even after consultation with an allergist, patients with a history of anaphylaxis are still not always prescribed an EAI.[5] Our previous population-based telephone survey found that approximately half of Canadians with a diagnosed food allergy had an EAI.[36] However, data on prescription of the EAI was not collected in this study. Therefore, it is possible that even though patients reported not having the EAI at the time of the survey, they may have received a prescription from their physician in the past. A study from the United States found that only 46% of children and 23% of adults with a diagnosed peanut and/or tree nut allergy were prescribed an EAI.[5] However, this study did not collect data on prescription of the EAI for other allergies besides peanut and tree nut.

The situation of extremely low prescription rates of EAIs among food-allergic patients is not unique to the North American context. In Japan, physicians were asked to describe situations in

which they would prescribe the EAI to a patient. Of the physicians who had ever prescribed the EAI (47% of the participants), only 41.6% agreed that cases with a history of at least one anaphylactic episode should have an EAI, and 88% agreed that repeated cases of anaphylaxis warranted prescription of the EAI.[158] The authors did not report on the percentage of individuals with food allergy who were prescribed the EAI or the type of physician participating in the survey.

A Dutch study looked at the frequency of EAI ownership among adolescents aged 11 to 20 years old from high schools in 4 provinces of the Netherlands.[159] All participants were asked questions regarding symptoms and diagnosis of food allergy. Of the 2,284 participants surveyed, 396 indicated an issue with food and 168 agreed to be interviewed. Forty-eight adolescents were classified as probably food-allergic, of which eight were not aware of their food allergy. Twenty-three adolescents were considered candidates for an EAI, whereas only two of them had been prescribed this medication. The calculated questionnaire-based prevalence of EAI need was 3.0%, whereas only 0.09% of adolescents owned an EAI.. Although this statistic cannot be generalized to the entire Dutch population because it focuses on only a few schools and does not cover all age groups, the results are alarming, especially because adolescents are more likely to experience anaphylactic reactions due to risk-taking behaviour such as ignoring precautionary labels on packaged foods.[160]

In Germany, Mehl questioned pediatricians about prescription of the EAI for children below the age of 12 years who had experienced anaphylaxis to foods, insect stings, medication, and immunotherapy in the previous year.[150] Only 17% of children with an episode of anaphylaxis in

the past year were prescribed the device. Unfortunately, data on prescription of the EAI for children with food-induced anaphylaxis specifically were not reported. In addition, this study only looked at prescription of the EAI after an anaphylactic event had taken place, which would presumably be an important reason for prescribing the EAI. Although data are unavailable from this study on the prescription rate among children with allergies who had not experienced an anaphylactic episode in the last year, it is likely that the rate would be even less than 17%.

In addition to studies in children and adolescents, one Italian study explored the rate of EAI prescription among adults with food allergy in 19 allergy clinics.[161] The authors reported that only 25% of adults with probable food allergy were prescribed the EAI. This study collected data from patients who had visited allergy clinics, who are not expected to be representative of the general population of allergic individuals, many of whom likely do not seek a diagnosis from an allergist. Therefore, these results cannot be extrapolated to the general population.

In Canada, only one study has assessed the rate of EAI prescriptions. Using the Drug Programs Information Network in Manitoba, a database containing information on 279,638 children, the authors found that 1.2% of children were dispensed an EAI.[162] However, this study did not address whether prescribing practices were appropriate as it does not link prescription with any form of diagnosis. Hence, it provides no information on whether those prescribed an EAI actually require one, or on the percentage of children with food allergy who were prescribed an EAI. Further, this study provides no information on adults, does not provide information on what the EAI was prescribed for, and does not provide nationwide Canadian estimates of EAI dispensation rates.

Prescription of the EAI among vulnerable populations

It is hypothesized that those of low education, low income, new Canadians and those of Aboriginal identity would have fewer EAI prescriptions due to poorer access to regular healthcare,[99, 163] lower health literacy,[37] and/or use of alternative healthcare such as healers.[164] However, there is very little data regarding prescription of the EAI among these groups. Our previous study estimated the percentage of Canadians who had the EAI; specifically, we compared various education and income levels, and immigrants versus individuals born in Canada, but no significant differences were observed, potentially due to a small sample size.[36] Another limitation of this study is that it did not specifically address prescription of the EAI; hence, it is possible that those who did not have the EAI at the time of the survey had been prescribed the device at some point in the past but chose not to fill or renew the prescription. Finally, data on prescription of the EAI for allergies to milk, egg, wheat, and soy were not collected.

Availability of the EAI

Because anaphylaxis is largely unpredictable, occurs frequently in the community,[31, 165] progresses rapidly,[43] and is potentially life-threatening,[144-146] it is essential that all patients at risk of anaphylaxis be provided with a prescription for an EAI and be made aware of the importance of always carrying it with them. Unfortunately, many studies suggest that individuals at risk of anaphylaxis do not carry the device with them at all times.[31, 155, 166, 167] The European Academy of Allergy and Clinical Immunology recently released a public declaration on food allergy and anaphylaxis, and emphasized the need for individuals at risk to carry an EAI at all times.[168] They highlight that those patients at risk of anaphylaxis are not only those who have experienced a

severe allergic reaction in the past but also those who have suffered an allergic reaction after eating a very small amount of food and those who have concomitant asthma.

Many reasons for not carrying the EAI have been identified. A qualitative study performed in Scotland with teenagers found that many of them did not carry the EAI for various reasons, including the fact that the device was inconvenient to carry due to its size.[167] Like the Scottish study, researchers from Southampton found that most teenagers did not carry the EAI, and that the decision not to carry it was based on six factors: circumstances, the type of allergy, device design, the responsibility and attitude of others, and the teenager's feelings and concerns.[16]

An American study determined that only 25% of elementary school children at least 5 years of age, and 42% less than 5 years of age, have their EAI available with them during meals and snacks.[39] A study done by our team in Montreal found that 48% of children with peanut allergy did not carry their EAI with them at school.[31] Although these studies provide interesting data, unfortunately only elementary school children from small geographic locations were included.

In the United States, a study to assess risk-taking behaviours of adolescents with food allergy found that 61% of participants carried their EAI at all times, but this varied depending on the activity; 94% carried the EAI while travelling but only 43% while doing sports.[160] This is startling given that the likelihood of experiencing anaphylaxis outside the home is quite high.[19] As with the studies described in the previous paragraph, the study methodology is limited-this study only collected information on teenagers, and the participants answered an internet-based survey, which is likely not representative of the general population.

As with prescription of the EAI, the data on availability of the device among vulnerable populations is sparse. It is anticipated that vulnerable populations, potentially due to nonconventional healthcare beliefs, lower employment rates and therefore limited access to employee health benefits to cover the cost of purchasing the EAI, may be less likely to carry the EAI.[169] One qualitative study performed in Waterloo, Canada, queried physicians about issues surrounding food allergy among recent Asian immigrants, and found that educating immigrant patients regarding the importance of avoiding the offending food and carrying the EAI is difficult due to cultural and language barriers.[76] This study revealed a gap in the current healthcare system in its ability to adequately manage individuals of different cultural and linguistic backgrounds. Unfortunately, this study only collected data from a few Asian immigrant families from a small geographic area, and therefore, more data is needed from other immigrant populations from across Canada to see if the issues identified in this small population are generalizable.

Summary

Epinephrine is the first-line treatment for anaphylaxis, and the EAI should be prescribed to all those at risk of anaphylaxis, including those without a previous reaction. To prevent a fatality, the EAI must be available at all times. Unfortunately, not all allergic individuals are prescribed the EAI, and of those who are, many of them do not have it readily available. In order to ensure the safety of those at risk of severe anaphylaxis, it is important to identify those who are particularly unlikely to receive a prescription and/or to carry the EAI, and ensure that their healthcare needs are met. Although it is believed that vulnerable populations may be in danger due to lower accessibility of healthcare services and other reasons summarized above, data on prescription and availability of the EAI among vulnerable populations are currently unavailable.

The current study (SPAACE: Surveying Prevalence of food Allergy in All Canadian Environments) bridges the gaps in the literature by targeting vulnerable populations, and providing data on prescription and availability of the EAI for nine priority food allergens (peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and soy). In the next chapter, data from the SPAACE study on prescription and availability of the EAI will be presented in the form of a manuscript entitled “Likelihood of being prescribed an epinephrine autoinjector in allergic Canadians with lower educational levels,” which was published in Annals of Allergy, Asthma and Immunology.

VII: LIKELIHOOD OF BEING PRESCRIBED AN EPINEPHRINE AUTOINJECTOR IN ALLERGIC CANADIANS WITH LOWER EDUCATIONAL LEVELS

Lianne Soller, Sabrine Cherkaoui, Moshe Ben-Shoshan, Daniel W. Harrington, Megan Knoll, Joseph Fragapane, Lawrence Joseph, Yvan St. Pierre, Sebastien La Vieille, Kathi Wilson, Susan J. Elliott, and Ann E. Clarke.

Failure to administer epinephrine promptly after suspected ingestion of a food allergen can have severe and even fatal consequences.[170] Our previous population-based telephone survey found that only 55% of Canadians diagnosed with peanut, tree nut, fish, shellfish and/or sesame allergy self-reported having an epinephrine auto-injector (EAI), and adults, males, and those residing in households where the respondent was single were less likely to have one.[36] However, this study was limited as it under-represented some of the most vulnerable populations. Further, we did not query patients on actual availability of the EAI nor on other common allergies-milk, egg, wheat, or soy.

The current study (SPAACE: Surveying Prevalence of food Allergy in All Canadian Environments) bridges these gaps by targeting vulnerable populations-those of low education, low income (Low income refers to household income below the Low Income Cut-off (LICO), defined as the 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), new Canadians (immigrated to Canada <10 years ago), and individuals of Aboriginal identity (First Nations, Métis, or Inuit). It also provides data on prescription and

availability of the EAI for nine priority food allergens (peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and soy).

We performed a random cross-Canada telephone survey in 2010-2011 targeting regions with a high proportion of low-income households, new Canadians, and individuals of Aboriginal identity.[89] Using 2006 Census data, we first targeted postal codes from the 100 census tracts with the highest proportion of our groups of interest, then randomly selected households from within these postal codes. If the respondent reported that any household member had an allergy to any of the nine foods, they were queried on whether an EAI had ever been prescribed for this individual, and whether the device was always carried outside the home.

Multivariate logistic regression was performed to identify predictors of being prescribed or always carrying the EAI among individuals with a diagnosed food allergy (self-report of a physician diagnosis of allergy to at least one of the nine food allergens [36]).

The following variables were included as covariates: age (<18 years), sex, post-secondary education (college/university degree for adults or highest educational attainment in the household for children), low income, immigrant status (new Canadian, immigrated ≥10 years ago, born in Canada), Aboriginal status, marital status, urban location of household (Residing in a Canadian metropolitan area with a population ≥ 100,000), presence of peanut/tree nut allergy, allergy to > one priority allergen, age at most severe reaction, treatment with epinephrine during most severe reaction, multiple allergic reactions, and self-report of diagnostic allergy testing. As 24% of the sample did not report household income, a sensitivity analysis for missing income

was performed where a multivariate model, which included individuals reporting income, was compared to a model that included those not reporting income.

Of the 12,762 households reached, 5,734 households, representing 15,022 individuals, completed the survey (45% response rate). Of 348 participants with self-reported diagnosed food allergy, 44.0% (95% CI, 38.7%, 49.4%) were prescribed an EAI, and 56.9% (95% CI, 48.6%, 64.8%) reported always carrying it (Table I).

In the multivariate model, adults with higher education and children residing in a household with an adult having a post-secondary degree [OR: 3.30 (95% CI, 1.69, 6.44)], individuals reporting Aboriginal identity [OR: 2.50 (95% CI, 1.09, 5.75)], those with peanut/tree nut allergy [OR: 3.01 (95% CI, 1.75, 5.17)], those who experienced their most severe reaction at a younger age [OR: 0.98 (95% CI, 0.96, 0.99)], and those reporting confirmatory testing [OR: 3.44 (95% CI, 1.58, 7.48)] were more likely to be prescribed an EAI. Prescription did not differ according to income or immigrant status. The predictors of being prescribed the EAI were the same in the multivariate models that were restricted to individuals who did and did not report their income. There were no factors associated with carrying the EAI.

This is the first study to examine prescription and availability of the EAI using a population-based survey targeting vulnerable populations. We found that many Canadians with food allergy are not prescribed an EAI, particularly adults with lower education and children residing in households with low educational attainment; further, almost half of those prescribed the device do not carry it. As has been demonstrated previously, self-report of EAI availability is likely to

be overestimated, perhaps because of social desirability bias,[171] whereby patients tend to report carrying the device because they know they are supposed to and this response is thus "socially desirable."

Those of higher education were more likely to be prescribed the EAI, which could partly be attributed to greater access to long-term follow-up with a specialist or higher health literacy, making them more likely to request an EAI prescription from their physician.[37] We hypothesized that those of low income, new Canadians and those of Aboriginal identity would have fewer prescriptions due to poorer access to regular healthcare,[99, 163] and/or use of alternative healthcare such as healers.[164] Further, we anticipated that these populations, potentially due to nonconventional healthcare beliefs and lower employment rates [169] and therefore limited access to employee health benefits to cover the cost of purchasing the EAI, may be less likely to carry the EAI. However, we did not observe any differences between new Canadians, low income and the rest of the sample, perhaps because of inadequate sample size, or because those participating in our study were more likely to have access to healthcare and follow western medicine than non-participants. Interestingly, individuals of Aboriginal identity were more likely to be prescribed an EAI than non-Aboriginal people. It is possible that limited availability of emergency healthcare services in remote northern regions,[172] where most of Canada's Aboriginal population resides, prompted physicians to prescribe the EAI more readily than in urban centres, where the majority of the country's non-Aboriginal population resides.

It was interesting that no differences were observed in availability of the EAI across all of the characteristics we evaluated. It is possible that variables that predict who carries the device were not captured in our study and these deserve further attention.

Our findings may not be generalizable to the American context because of differences in the healthcare systems in Canada compared with the United States. In Canada, access to healthcare, and a large percentage of prescription drug costs, are covered by either the employer or by the government. In the United States, access to healthcare is not yet universal, and the vast majority of Americans must pay out of pocket for their prescription drugs.

As epinephrine is the only means of arresting progression of an allergic reaction into anaphylaxis, the importance of immediate accessibility of the EAI needs to be disseminated to the general population, particularly to those of lower socioeconomic status. In addition, clinicians must be reminded of the importance of prescribing the EAI as it will enhance the quality of life of patients and families affected with food allergy.

Table VIII: Prescription and availability of the EAI among individuals with diagnosed food allergy

	Total with diagnosed food allergy (n)	**Prescribed the EAI** % (95% CI)	**Total prescribed the EAI** (n)	**Carrying the EAI** % (95% CI)
Allergic to any of the nine priority food allergens	348	44.0 (38.7, 49.4)	153	56.9 (48.6, 64.8)
Children (<18 years)	96*****	68.8 (58.5, 77.8)	66	57.6 (44.8, 69.7)
Adults	244*****	34.8 (28.9, 41.2)	85	56.5 (45.3, 67.2)
Male	143	47.6 (39.1, 56.1)	68	51.5 (39.0, 63.8)
Female	205	41.5 (34.6, 48.5)	85	61.2 (50.0, 71.6)
High education	259**	49.0 (42.8, 55.3)	127**	59.8 (50.8,68.4)
Low education	8**	28.7 (19.5, 39.4)	25**	44.0 (24.4, 65.1)
High income	216$^{\neq}$	48.6 (41.8, 55.5)	138	56.2 (46.2, 65.9)
Low income	49$^{\neq}$	30.6 (18.3, 45.4)	15	53.3 (26.6, 78.7)
New Canadian	24	33.3 (15.6, 55.3)	8	75.0 (34.9, 96.8)
Immigrated ≥10 years ago	57	29.6 (18.0, 43.6)	16	56.3 (29.9, 80.2)
Born in Canada	267	48.3 (42.2, 54.5)	129	55.8 (46.8, 64.5)
Aboriginal identity	41	56.1 (39.7, 71.5)	23	43.5 (23.2, 65.5)
Non-Aboriginal identity	307	42.3 (36.8, 48.1)	130	59.2 (50.3, 67.8)
Married/living with partner	236$^{\pm}$	46.6 (40.1, 53.2)	110$^{\pm}$	55.5 (45.7, 64.9)
Single	104$^{\pm}$	38.5 (29.1, 48.5)	40$^{\pm}$	65.0 (48.3, 79.4)
Urban household	214	39.7 (33.1, 46.6)	85	62.4 (51.2, 72.6)
Rural household	134	50.7 (42.0, 59.5)	68	50.0 (37.6, 62.4)
Peanut and/or tree nut allergy	164	61.6 (53.7, 69.1)	101	59.4 (49.2, 69.1)
Did not have peanut and/or tree nut allergy	184	28.3 (21.9, 35.4)	52	51.9 (37.6, 66.0)
Multiple allergies	100	61.0 (50.7, 70.6)	61	65.6 (52.3, 77.3)
Single allergy	248	37.1 (31.1, 43.4)	92	51.1 (40.4, 61.7)
Treated with epinephrine during most severe reaction	67	43.3 (31.2, 56.0)	29	62.1 (42.3, 79.3)
Did not receive epinephrine during most severe reaction	281	44.1 (38.2, 50.1)	124	55.6 (46.5, 64.6)
Multiple allergic reactions	280	44.3 (38.4, 50.3)	124	58.1 (48.9, 66.9)
Single allergic reaction	68	42.6 (30.7, 55.2)	29	51.7 (32.5, 70.6)
Report diagnostic testing	293	48.1 (42.3, 54.0)	141	58.2 (49.6, 66.4)
Did not report diagnostic testing	55	21.8 (11.8, 35.0)	12	41.7 (15.2, 72.3)

*Data on age were missing for 8 individuals for prescription of the EAI.

**For adults, education refers to their personal educational attainment. For children, education refers to the highest educational attainment in the household. Data on education were missing for 2 individuals for prescription of the EAI and 1 for availability.

[≠]Data on income were missing for 83 individuals.
[±]Data on marital status missing for 8 individuals for prescription of EAI and 3 for availability.

VIII: FINAL SUMMARY AND CONCLUSIONS

Observed differences in prevalence between the vulnerable and non-vulnerable populations and potential explanations

Food allergy is an important condition affecting between 8 and 9% of North Americans, according to self-report.[8, 52, 53, 66] Prior to this study, there was very little information about how the burden of food allergy may differ across various sociodemographic groups. More specifically, it was unknown whether those of low education and income, immigrants, and individuals of Aboriginal identity experience food allergy differently than the general population. Given that vulnerable populations in Canada tend to report lower access to medical care,[14] it was hypothesized that they would be less likely to receive a doctor diagnosis for their food allergy and proper management for their food allergy, including prescription of the epinephrine auto-injector (EAI).

The current study indicates that those of lower education and new Canadians self-report fewer allergies, but there is no difference in terms of income or Aboriginal identity.[38] We also found that those of lower educational attainment were less likely to be prescribed the EAI, and those of Aboriginal identity were more likely to be prescribed the device, but we did not observe any differences in terms of availability of the EAI.[173]

One plausible reason for the differences in prevalence of food allergy we observed is that vulnerable populations are less able to recognize food allergy symptoms and are therefore less

likely to report food allergy than non-vulnerable populations. Another reason is that vulnerable populations experience barriers to accessing health care, which could result in fewer diagnoses of food allergy and a lower likelihood of being prescribed the EAI. It is also possible that non-vulnerable populations truly experience more food allergies due to delayed introduction of allergenic foods in children, less exposure to bacterial infections and other environmental exposures that may promote allergic disease, and other factors,[48, 49] as discussed in detail in Chapter II.

Educating key stakeholders on food allergy diagnosis and management

To ensure that all Canadians have equal access to information about and health care for food allergy, our results must be shared with important stakeholders at the government level, and with physician associations and community organizations where vulnerable populations live.

At the government level, our results could help inform the creation of targeted education campaigns around how to recognize food allergy symptoms, and the importance of receiving a diagnosis and follow up with an allergist. Specifically, public campaigns on the web, radio, and television, in addition to pamphlets or flyers that could be distributed to areas where a high number of vulnerable populations reside, may be very useful. Recently, a series of television commercials funded by EpiPen spread the message that individuals with allergies should carry the EpiPen with them in case of an accidental exposure. However, it is unclear whether such a campaign would be effective in targeting vulnerable populations because they may have limited access to a television and, if they do, they may not be fluent in English or French. To ensure that

campaigns to increase the knowledge and awareness of food allergy are effective in the vulnerable communities, more research is needed to determine the type of campaign that would reach the most people and the wording that needs to be used so that vulnerable populations grasp the main message.

Our research group is currently seeking funding for a knowledge mobilization project that will attempt to characterize the knowledge and awareness of food allergy among the general public as part of the SPAACE to SPAACE study, a follow up to the research presented in this thesis. The goal of this knowledge mobilization project is to first determine the gaps in knowledge and awareness around food allergy in terms of prevalence, diagnosis, and management, using a nationwide telephone survey (as outlined earlier in this thesis), then sharing the results of the questionnaire with key stakeholder groups (food allergy advocacy associations, government policy makers, physician groups, and more), and finally, creating dissemination tools such as information/fact sheets, videos, and other vehicles, that are adapted to the needs of the Canadian public. The hope is that these tools will lead to a more informed Canadian community who are able to better understand the burden of food allergy, know how to prevent an allergic reaction, recognize symptoms of an allergic reaction and when to treat with epinephrine if a reaction were to occur accidentally, and the importance of seeking a diagnosis where a food allergy is suspected.

As was previously mentioned, another potential reason for the lower prevalence of food allergy observed among vulnerable populations in this study is the lower access to health care services.

We know that Aboriginal communities in northern regions have difficulty accessing even basic medical care, new Canadians may sometimes not be immediately eligible for provincial health insurance, and those of low education and income may not have access to medication insurance to pay for the EAI. [99, 163] Differential access to health care in Canada is a problem that needs to be rectified by building more hospitals and clinics in under-served communities. In addition to infrastructure changes, we also need to train more healthcare professionals who can seamlessly navigate the cultural, language, and other barriers that are currently preventing vulnerable individuals from seeing a physician and having their health care needs met.

At the physician association level, there are a few things that can be done to improve the current situation. First, education around how to recognize food allergy and when to test for food allergy must be disseminated to all physicians, especially general practitioners and emergency room physicians, who are often the first point of contact for individuals with suspected food allergy. In the emergency rooms, many individuals with a suspected food allergy are not treated with epinephrine, not prescribed the EAI, and/or not given a referral to see an allergist. Among general practitioners, there is a huge problem of mis-diagnosis of food allergy because of lack of awareness of the diagnostic procedures, which includes a careful targeted clinical history and appropriate usage and interpretation of several tests, including a prick skin test, allergen-specific levels of IgE in the serum, and a food challenge. In summary, emergency room physicians and general practitioners must be taught how to 1) correctly assess patients who have had symptoms that may or may not be consistent with food allergy in a way that is appropriate and understandable to each patient, 2) recognize when and how to perform diagnostic tests to

confirm or refute the allergy, and 3) understand the importance of prescribing the EAI when a food allergy is diagnosed.

Most plausible estimate among bias-adjusted estimates

Since our study was a telephone survey, and it is well accepted that this type of survey is susceptible to low response rates, which can lead to bias,[107] we decided to explore the potential influence of non-response bias on our estimates of food allergy prevalence. We tested various clinically plausible assumptions about the magnitude and direction of such bias and created a total of nine prevalence estimates.[130] We found that even a conservative estimate of bias could alter the prevalence of food allergy significantly, and we concluded that non-response bias should not be ignored. Of course, our methods are not perfect, and caution must be applied when interpreting the results of our analysis. However, we believe that this exercise was an important one as it sheds light on the drastic differences in study results that can be obtained if non-response bias is present. Since one can never be certain if non-response bias is present in a specific study, it is important to implement strategies to maximize the response rate and collect information that would allow researchers to adjust for non-response after data collection.

If we were to present our research on bias-adjusted estimates of prevalence of food allergy to the government or other stakeholders, it would of course be necessary to summarize the data and provide the most plausible estimate. Although it is possible that any of these nine models are correct, or conversely, that none of these nine models is correct, We believe that the true prevalence of food allergy in Canada lies somewhere between model 4 and 5, where the *Non-*

Participants are assumed to have half of or the same prevalence of food allergy as the *Refusal Questionnaire Participants*, and the *Never Reached Participants* are assumed to have a mixture of the prevalence in the *Full Participants*, *Refusal Questionnaire Participants*, and *Never Reached Participants*, for a bias-adjusted overall prevalence of food allergy of between 3.5% to 4.0%. The reasoning behind the choice of these assumptions as the most plausible are two-fold. First, it is unlikely that the *Non-Participants* are two times *more* likely to report a food allergy than the *Refusal Questionnaire Participants*, due to data on the behavior of participants and non-participants in research which states that the likelihood of participating if the individual has an interest in the topic of the study is higher. Therefore, those not participating would have a lower prevalence than those who did participate. It is therefore very likely that the prevalence of food allergy among *Non-Participants* is the same as or half that of *Refusal Questionnaire Participants*, but there are no specific reasons to favor one over the other assumption, and therefore, both are deemed equally plausible. Second, we believe that the group of *Never Reached Participants* is likely to be a mixture of *Full*, *Refusal Questionnaire*, and *Non-Participants*, simply because there are multiple reasons why a particular household was never reached. Some of them would have probably completed the survey if they were available when the interviewer called, just like some would have completed the Refusal Questionnaire, and others would have refused to participate. On the other hand, it is unlikely that all *Never Reached Participants* would have fully participated, or that all would have refused, and therefore, these assumptions are likely to be incorrect. Therefore, we are left with an estimate between 3.5% and 4.0%, which is a much narrower range than that from the original nine estimates, and which is much more informative to stakeholders when trying to compare the prevalence of food allergy

adjusted and non-adjusted for non-response bias. However, our study is the first to provide such bias-adjusted estimates and hence, these estimates cannot be directly compared to those from other population-based food allergy prevalence studies.

Final Summary

This thesis presented data from the SPAACE nationwide telephone survey, a survey that specifically targeted vulnerable Canadians because of the lack of data regarding prevalence of food allergy among this segment of the Canadian population. We found that, in general, vulnerable populations were less likely to report food allergy than their non-vulnerable counterparts, and were also less likely to be prescribed the EAI. These data must be disseminated to key decision-makers so that measures can be taken to ensure equal opportunity for all Canadians, regardless of their socioeconomic or cultural background, to receive a proper diagnosis and long-term management for their food allergy, including prescription of the EAI.

This thesis also explored the important role of non-response bias in telephone surveys, specifically as it relates to the prevalence of food allergy. We found that even conservative estimates of non-response lead to important differences in the prevalence of food allergy obtained from a study. These results highlight the importance of taking all of the necessary steps to reduce non-response in a study, and taking into account the potential influence of non-response on the results obtained.

REFERENCES

1. Soller L, Hourihane J, DunnGalvin A. The impact of oral food challenge tests on food allergy health-related quality of life. Allergy 2014; 69:1255-7.
2. Rona RJ, Keil T, Summers C, Gislason D, Zuidmeer L, Sodergren E, et al. The prevalence of food allergy: a meta-analysis. The Journal of allergy and clinical immunology 2007; 120:638-46.
3. Sicherer SH, Munoz-Furlong A, Burks AW, Sampson HA. Prevalence of peanut and tree nut allergy in the US determined by a random digit dial telephone survey. The Journal of allergy and clinical immunology 1999; 103:559-62.
4. Sicherer SH, Munoz-Furlong A, Godbold JH, Sampson HA. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. The Journal of allergy and clinical immunology 2010; 125:1322-6.
5. Sicherer SH, Munoz-Furlong A, Sampson HA. Prevalence of peanut and tree nut allergy in the United States determined by means of a random digit dial telephone survey: a 5-year follow-up study. The Journal of allergy and clinical immunology 2003; 112:1203-7.
6. Sicherer SH, Munoz-Furlong A, Sampson HA. Prevalence of seafood allergy in the United States determined by a random telephone survey. The Journal of allergy and clinical immunology 2004; 114:159-65.
7. 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. The Journal of allergy and clinical immunology 2010; 125:1327-35.
8. 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. The Journal of allergy and clinical immunology 2012; 130:986-8.
9. Shi L, Stevens GD. Vulnerability and unmet health care needs. The influence of multiple risk factors. J Gen Intern Med 2005; 20:148-54.
10. Prescott SL. Allergy: the price we pay for cleaner living? Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology 2003; 90:64-70.
11. Fishbein AB, Fuleihan RL. The hygiene hypothesis revisited: does exposure to infectious agents protect us from allergy? Current opinion in pediatrics 2012; 24:98-102.
12. Fox AT, Sasieni P, du Toit G, Syed H, Lack G. Household peanut consumption as a risk factor for the development of peanut allergy. J Allergy Clin Immunol 2009; 123:417-23.
13. Walton SF, Weir C. The interplay between diet and emerging allergy: what can we learn from Indigenous Australians? International reviews of immunology 2012; 31:184-201.
14. Talbot Y, Fuller-Thomson E, Tudiver F, Habib Y, McIsaac WJ. Canadians without regular medical doctors. Who are they? Canadian family physician Medecin de famille canadien 2001; 47:58-64.
15. Harrington DW, Wilson K, Elliott SJ, Clarke AE. Diagnosis and treatment of food allergy in off-reserve Aboriginal children in Canada. The Canadian Geographer/Le Géographe canadien 2013; 57:10.
16. Macadam C, Barnett J, Roberts G, Stiefel G, King R, Erlewyn-Lajeunesse M, et al. What factors affect the carriage of epinephrine auto-injectors by teenagers? Clinical and translational allergy 2012; 2:3.
17. Ben-Shoshan M, Harrington DW, Soller L, Fragapane J, Joseph L, Pierre YS, et al. Demographic predictors of peanut, tree nut, fish, shellfish, and sesame allergy in Canada. Journal of allergy 2012; 2012:858306.

18. Sheth SS, Waserman S, Kagan R, Alizadehfar R, Primeau MN, Elliot S, et al. Role of food labels in accidental exposures in food-allergic individuals in Canada. Ann Allergy Asthma Immunol 2010; 104:60-5.
19. Nguyen-Luu NU, Ben-Shoshan M, Alizadehfar R, Joseph L, Harada L, Allen M, et al. Inadvertent exposures in children with peanut allergy. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology 2012; 23:133-9.
20. Sampson HA. Anaphylaxis and emergency treatment. Pediatrics 2003; 111:1601-8.
21. Ellis AK, Day JH. Diagnosis and management of anaphylaxis. CMAJ 2003; 169:307-11.
22. Ellis AK, Day JH. The role of epinephrine in the treatment of anaphylaxis. Curr Allergy Asthma Rep 2003; 3:11-4.
23. Fitzharris P, Empson M, Ameratunga R, Sinclair J, Crump V, Steele R, et al. Anaphylaxis management: the essential role of adrenaline (epinephrine) auto-injectors. Should PHARMAC fund them in New Zealand? N Z Med J 2006; 119:U1965.
24. Pongracic JA, Kim JS. Update on epinephrine for the treatment of anaphylaxis. Curr Opin Pediatr 2007; 19:94-8.
25. Ronborg SM, Olsen OT, Heinig JH, Malling HJ. [Adrenaline for self treatment of anaphylactic reactions. Indications, available preparations and prescription rules]. Ugeskr Laeger 1996; 158:4539-43.
26. Sicherer SH, Simons FE. Quandaries in prescribing an emergency action plan and self-injectable epinephrine for first-aid management of anaphylaxis in the community. J Allergy Clin Immunol 2005; 115:575-83.
27. Sicherer SH, Simons FE. Self-injectable epinephrine for first-aid management of anaphylaxis. Pediatrics 2007; 119:638-46.
28. Simons FE. First-aid treatment of anaphylaxis to food: focus on epinephrine. J Allergy Clin Immunol 2004; 113:837-44.
29. Soar J, Pumphrey R, Cant A, Clarke S, Corbett A, Dawson P, et al. Emergency treatment of anaphylactic reactions--guidelines for healthcare providers. Resuscitation 2008; 77:157-69.
30. Sicherer SH, Forman JA, Noone SA. Use assessment of self-administered epinephrine among food-allergic children and pediatricians. Pediatrics 2000; 105:359-62.
31. Ben-Shoshan M, Kagan R, Primeau MN, Alizadehfar R, Verreault N, Yu JW, et al. Availability of the epinephrine autoinjector at school in children with peanut allergy. Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology 2008; 100:570-5.
32. Greenberger PA, Rotskoff BD, Lifschultz B. Fatal anaphylaxis: postmortem findings and associated comorbid diseases. Annals of Allergy, Asthma & Immunology 2007; 98:252-7.
33. Clark S, Bock SA, Gaeta TJ, Brenner BE, Cydulka RK, Camargo CA. Multicenter study of emergency department visits for food allergies. J Allergy Clin Immunol 2004; 113:347-52.
34. Pulcini JM, Sease KK, Marshall GD. Disparity between the presence and absence of food allergy action plans in one school district. Allergy Asthma Proc 2010; 31:141-6.
35. Rhim GS, McMorris MS. School readiness for children with food allergies. Ann Allergy Asthma Immunol 2001; 86:172-6.
36. Soller L, Fragapane J, Ben-Shoshan M, Harrington DW, Alizadehfar R, Joseph L, et al. Possession of epinephrine auto-injectors by Canadians with food allergies. The Journal of allergy and clinical immunology 2011; 128:426-8.

37. 2006 Census Dictionary. Ottawa; 2007.] Available from https://www12.statcan.gc.ca/census-recensement/2006/ref/dict/azindex-eng.cfm#C.
38. 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. The Journal of Allergy and Clinical Immunology: In Practice 2014.
39. DeMuth KA, Fitzpatrick AM. Epinephrine autoinjector availability among children with food allergy. Allergy and asthma proceedings : the official journal of regional and state allergy societies 2011; 32:295-300.
40. Cantor D, O'Hare BC, O'Connor KS. The Use of Monetary Incentives to Reduce Nonresponse in Random Digit Dial Telephone Surveys. In: Advances in Telephone Survey Methodology: John Wiley & Sons, Inc.; 2007. p. 471-98.
41. Ravid NL, Annunziato RA, Ambrose MA, Chuang K, Mullarkey C, Sicherer SH, et al. Mental health and quality-of-life concerns related to the burden of food allergy. Immunology and allergy clinics of North America 2012; 32:83-95.
42. Sicherer SH, Leung DY. Advances in allergic skin disease, anaphylaxis, and hypersensitivity reactions to foods, drugs, and insects in 2012. The Journal of allergy and clinical immunology 2013; 131:55-66.
43. Boyce JA, Assa'ad A, Burks AW, Jones SM, Sampson HA, Wood RA, et al. Guidelines for the Diagnosis and Management of Food Allergy in the United States: Summary of the NIAID-Sponsored Expert Panel Report. The Journal of allergy and clinical immunology 2010; 126:1105-18.
44. Food allergy: a practice parameter. Ann Allergy Asthma Immunol 2006; 96:S1-68.
45. Sicherer SH, Sampson HA. 9. Food allergy. J Allergy Clin Immunol 2006; 117:S470-5.
46. Zuidmeer L, Goldhahn K, Rona RJ, Gislason D, Madsen C, Summers C, et al. The prevalence of plant food allergies: a systematic review. The Journal of allergy and clinical immunology 2008; 121:1210-8 e4.
47. Sampson HA. Food allergy. Part 1: immunopathogenesis and clinical disorders. J Allergy Clin Immunol 1999; 103:717-28.
48. Kremmyda LS, Vlachava M, Noakes PS, Diaper ND, Miles EA, Calder PC. Atopy risk in infants and children in relation to early exposure to fish, oily fish, or long-chain omega-3 fatty acids: a systematic review. Clinical reviews in allergy & immunology 2011; 41:36-66.
49. Shu SA, Chang C, Leung PS. Common methodologies in the evaluation of food allergy: pitfalls and prospects of food allergy prevalence studies. Clinical reviews in allergy & immunology 2014; 46:198-210.
50. Nwaru BI, Hickstein L, Panesar SS, Muraro A, Werfel T, Cardona V, et al. The epidemiology of food allergy in Europe: a systematic review and meta-analysis. Allergy 2014; 69:62-75.
51. McBride D, Keil T, Grabenhenrich L, Dubakiene R, Drasutiene G, Fiocchi A, et al. The EuroPrevall birth cohort study on food allergy: baseline characteristics of 12,000 newborns and their families from nine European countries. Pediatric Allergy and Immunology 2012; 23:230-9.
52. Vierk KA, Koehler KM, Fein SB, Street DA. Prevalence of self-reported food allergy in American adults and use of food labels. The Journal of allergy and clinical immunology 2007; 119:1504-10.
53. 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.

54. Pawlinska-Chmara R, Wronka I, Muc M. Prevalence and correlates of allergic DISEASES among children. Journal of Physiology and Pharmacology 2008; 59:549-56.
55. Silverberg J, Simpson EL, Durkin HG, Joks R. Prevalence of allergic disease in foreign-born american children. JAMA Pediatrics 2013:1-7.
56. MacMillan HL, Jamieson E, Walsh C, Boyle M, Crawford A, MacMillan A. The health of Canada's Aboriginal children: results from the First Nations and Inuit Regional Health Survey. Int J Circumpolar Health 2010; 69:158-67.
57. Aboriginal Peoples Survey-Inuit Health, Education and Country Food Harvesting. 2008.] Available from http://www.statcan.gc.ca/pub/89-637-x/89-637-x2008004-eng.pdf.
58. Masuda J, Stewart M, Blood R, Latycheva O, Hampson C. Asthma and Allergy Support and Education for Aboriginal Communities. AllerGen NCE Inc. Sixth Annual Research Conference. Vancouver, BC, 2011.
59. Canada ASo. A Shared Vision: Ensuring quality of life for adults and children with asthma and allergies in First Nations and Inuit communities in Canada. In: Brotherhood AoFNNI, ed, 2009.
60. Jackson KD, Howie LD, Akinbami LJ. Trends in allergic conditions among children: United States, 1997-2011. NCHS data brief 2013:1-8.
61. Kotz D, Simpson CR, Sheikh A. Incidence, prevalence, and trends of general practitioner-recorded diagnosis of peanut allergy in England, 2001 to 2005. The Journal of allergy and clinical immunology 2011; 127:623-30 e1.
62. American Academy of Pediatrics. Committee on Nutrition. Hypoallergenic infant formulas. Pediatrics 2000; 106:346-9.
63. Yazdanbakhsh M, Kremsner PG, van Ree R. Allergy, parasites, and the hygiene hypothesis. Science 2002; 296:490-4.
64. Shek LP, Cabrera-Morales EA, Soh SE, Gerez I, Ng PZ, Yi FC, et al. A population-based questionnaire survey on the prevalence of peanut, tree nut, and shellfish allergy in 2 Asian populations. The Journal of allergy and clinical immunology 2010; 126:324-31, 31 e1-7.
65. Shek LP, Lee BW. Food allergy in Asia. Curr Opin Allergy Clin Immunol 2006; 6:197-201.
66. McGowan EC, Keet CA. Prevalence of self-reported food allergy in the National Health and Nutrition Examination Survey (NHANES) 2007-2010. The Journal of allergy and clinical immunology 2013; 132:1216-9 e5.
67. Hourihane JO, Aiken R, Briggs R, Gudgeon LA, Grimshaw KE, DunnGalvin A, et al. The impact of government advice to pregnant mothers regarding peanut avoidance on the prevalence of peanut allergy in United Kingdom children at school entry. The Journal of allergy and clinical immunology 2007; 119:1197-202.
68. Venter C, Hasan Arshad S, Grundy J, Pereira B, Bernie Clayton C, Voigt K, et al. Time trends in the prevalence of peanut allergy: three cohorts of children from the same geographical location in the UK. Allergy 2010; 65:103-8.
69. Kennedy S, Kidd M, McDonald J, Biddle N. The Healthy Immigrant Effect: Patterns and Evidence from Four Countries. Journal of International Migration and Integration 2014:1-16.
70. Corlin L, Woodin M, Thanikachalam M, Lowe L, Brugge D. Evidence for the healthy immigrant effect in older Chinese immigrants: a cross-sectional study. BMC public health 2014; 14:603.
71. Du Toit G, Katz Y, Sasieni P, Mesher D, Maleki SJ, Fisher HR, et al. Early consumption of peanuts in infancy is associated with a low prevalence of peanut allergy. The Journal of allergy and clinical immunology 2008; 122:984-91.

72. Fleischer DM, Spergel JM, Assa'ad AH, Pongracic JA. Primary Prevention of Allergic Disease Through Nutritional Interventions. The Journal of Allergy and Clinical Immunology: In Practice 2013; 1:29-36.
73. Koplin JJ, Osborne NJ, Wake M, Martin PE, Gurrin LC, Robinson MN, et al. Can early introduction of egg prevent egg allergy in infants? A population-based study. The Journal of allergy and clinical immunology 2010; 126:807-13.
74. Blanc F, Vissers YM, Adel-Patient K, Rigby NM, Mackie AR, Gunning AP, et al. Boiling peanut Ara h 1 results in the formation of aggregates with reduced allergenicity. Molecular nutrition & food research 2011; 55:1887-94.
75. Beyer K, Morrow E, Li XM, Bardina L, Bannon GA, Burks AW, et al. Effects of cooking methods on peanut allergenicity. The Journal of allergy and clinical immunology 2001; 107:1077-81.
76. Lu SK, Elliott SJ, Clarke AE. Exploring Perceptions and Experiences of Food Allergy among New Canadians from Asia. Journal of allergy 2014; 2014:7.
77. Crighton EJ, Wilson K, Senecal S. The relationship between socio-economic and geographic factors and asthma among Canada's Aboriginal populations. International journal of circumpolar health 2010; 69:138-50.
78. Sin DD, Wells H, Svenson LW, Man SF. Asthma and COPD among aboriginals in Alberta, Canada. Chest 2002; 121:1841-6.
79. Derose KP, Bahney BW, Lurie N, Escarce JJ. Review: immigrants and health care access, quality, and cost. Medical care research and review : MCRR 2009; 66:355-408.
80. Russell SL, Finlay BB. The impact of gut microbes in allergic diseases. Current opinion in gastroenterology 2012.
81. Kalliomaki M, Isolauri E. Pandemic of atopic diseases--a lack of microbial exposure in early infancy? Current drug targets. Infectious disorders 2002; 2:193-9.
82. Harrington DW, Wilson K, Elliott SJ, Clarke AE. Diagnosis and treatment of food allergies in off-reserve Aboriginal children in Canada. The Canadian Geographer / Le Géographe canadien 2013; 57:431-40.
83. Pickett K, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health 2001 Feb; 55:12.
84. Canada Go. Low income cut-offs. In: Canada S, ed. Ottawa, 2013.
85. Canada S. 2006 Census Dictionary. Ottawa, Ontario, 2007.
86. Gibson PJ, Koepsell TD, Diehr P, Hale C. Increasing response rates for mailed surveys of Medicaid clients and other low-income populations. American journal of epidemiology 1999; 149:1057-62.
87. Schweitzer M, Asch DA. Timing payments to subjects of mail surveys: cost-effectiveness and bias. Journal of clinical epidemiology 1995; 48:1325-9.
88. Whiteman MK, Langenberg P, Kjerulff K, McCarter R, Flaws JA. A randomized trial of incentives to improve response rates to a mailed women's health questionnaire. Journal of women's health 2003; 12:821-8.
89. 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 research notes 2012; 5:572.
90. Hourihane JO, Kilburn SA, Dean P, Warner JO. Clinical characteristics of peanut allergy. Clin Exp Allergy 1997; 27:634-9.

91. Brown SG. Clinical features and severity grading of anaphylaxis. J Allergy Clin Immunol 2004; 114:371-6.
92. Sicherer SH, Burks AW, Sampson HA. Clinical features of acute allergic reactions to peanut and tree nuts in children. Pediatrics 1998; 102:e6.
93. Bahna SL. Cow's milk allergy versus cow milk intolerance. Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology 2002; 89:56-60.
94. de Boissieu D, Dupont C. [Differentiating celiac disease and wheat allergy]. Archives de pediatrie : organe officiel de la Societe francaise de pediatrie 2009; 16:873-5.
95. Clopper C, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934; 26:404–13.
96. Kelly C, Gangur V. Sex Disparity in Food Allergy: Evidence from the PubMed Database. Journal of allergy 2009; 2009:159845.
97. 2006 Census Highlight Tables. Ottawa; 2006.] Available from http://www12.statcan.gc.ca/census-recensement/2006/dp-pd/hlt/index-eng.cfm.
98. Sibley LM, Weiner JP. An evaluation of access to health care services along the rural-urban continuum in Canada. BMC Health Serv Res 2011; 11:20.
99. Asanin J, Wilson K. "I spent nine years looking for a doctor": exploring access to health care among immigrants in Mississauga, Ontario, Canada. Soc Sci Med 2008; 66:1271-83.
100. Chen J, Ng E, Wilkins R. The health of Canada's immigrants in 1994-95. Health reports / Statistics Canada, Canadian Centre for Health Information = Rapports sur la sante / Statistique Canada, Centre canadien d'information sur la sante 1996; 7:33-45, 37-50.
101. McDonald JT, Kennedy S. Insights into the 'healthy immigrant effect': health status and health service use of immigrants to Canada. Social science & medicine 2004; 59:1613-27.
102. 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.
103. Ben-Shoshan M, Kagan RS, Alizadehfar R, Joseph L, Turnbull E, St Pierre Y, et al. Is the prevalence of peanut allergy increasing? A 5-year follow-up study in children in Montreal. J Allergy Clin Immunol 2009; 123:783-8.
104. Lu S. They Thought I was Just Joking About It": Experiences and Perceptions of Food Allergy in New Canadians from Asia. School of Public Health & Health Systems Waterloo, Canada: University of Waterloo, 2013.
105. Sarndal CE, Swensson B, Wretman J. Model Assisted Survey Sampling. New York: Springer-Verlag; 1992.
106. Allison PD. Missing Data. Thousand Oaks, CA: Sage Publications; 2001.
107. Groves RM. Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly 2006; 70:646-75.
108. Waserman S, Chad Z, Francoeur MJ, Small P, Stark D, Vander Leek TK, et al. Management of anaphylaxis in primary care: Canadian expert consensus recommendations. Allergy 2010; 65:1082-92.
109. Muraro A, Werfel T, Hoffmann-Sommergruber K, Roberts G, Beyer K, Bindslev-Jensen C, et al. EAACI Food Allergy and Anaphylaxis Guidelines: diagnosis and management of food allergy. Allergy 2014; 69:1008-25.

110. Voigt LF, Koepsell TD, Daling JR. Characteristics of telephone survey respondents according to willingness to participate. American journal of epidemiology 2003; 157:66-73.
111. Berg N. Non-response bias. 2005.
112. Rupp I, Triemstra M, Boshuizen HC, Jacobi CE, Dinant HJ, van den Bos GA. Selection bias due to non-response in a health survey among patients with rheumatoid arthritis. Eur J Public Health 2002; 12:131-5.
113. Demarest S, Van der Heyden J, Charafeddine R, Tafforeau J, Van Oyen H, Van Hal G. Socio-economic differences in participation of households in a Belgian national health survey. The European Journal of Public Health 2013; 23:981-5.
114. Haapea M, Miettunen J, Veijola J, Lauronen E, Tanskanen P, Isohanni M. Non-participation may bias the results of a psychiatric survey. Social Psychiatry and Psychiatric Epidemiology 2007; 42:403-9.
115. Anderson JE, Nelson DE, Wilson RW. Telephone coverage and measurement of health risk indicators: data from the National Health Interview Survey. Am J Public Health 1998; 88:1392-5.
116. Lee S, Brick JM, Brown ER, Grant D. Growing cell-phone population and noncoverage bias in traditional random digit dial telephone health surveys. Health Serv Res 2010; 45:1121-39.
117. Blumberg SJ, Luke JV. Reevaluating the need for concern regarding noncoverage bias in landline surveys. Am J Public Health 2009; 99:1806-10.
118. Kempf AM, Remington PL. New challenges for telephone survey research in the twenty-first century. Annu Rev Public Health 2007; 28:113-26.
119. Shebl F, Poppell CE, Zhan M, Dwyer DM, Hopkins AB, Groves C, et al. Measuring health behaviors and landline telephones: potential coverage bias in a low-income, rural population. Public Health Rep 2009; 124:495-502.
120. Force ACPT. New Considerations for Survey Researchers When Planning and COnducting RDD Telephone Surveys in the U.S. With Respondents Reached Via Cell Phone Numbers. Lenexa, KS., 2010.
121. Hu SS, Balluz L, Battaglia MP, Frankel MR. Improving public health surveillance using a dual-frame survey of landline and cell phone numbers. American journal of epidemiology 2011; 173:703-11.
122. Barr ML, van Ritten JJ, Steel DG, Thackway SV. Inclusion of mobile phone numbers into an ongoing population health survey in New South Wales, Australia: design, methods, call outcomes, costs and sample representativeness. BMC Med Res Methodol 2012; 12:177.
123. Livingston M, Dietze P, Ferris J, Pennay D, Hayes L, Lenton S. Surveying alcohol and other drug use through telephone sampling: a comparison of landline and mobile phone samples. BMC Med Res Methodol 2013; 13:41.
124. Link MW, Battaglia MP, Frankel MR, Osborn L, Mokdad AH. Reaching the US cell phone generation: comparison of cell phone survey results with an ongoing landline telephone survey. Public Opinion Quarterly 2007; 71:26.
125. Altman DR, Chiaramonte LT. Public perception of food allergy. Environmental Toxicology and Pharmacology 1997; 4:95-9.
126. Jansen JJ, Kardinaal AF, Huijbers G, Vlieg-Boerstra BJ, Martens BP, Ockhuizen T. Prevalence of food allergy and intolerance in the adult Dutch population. J Allergy Clin Immunol 1994; 93:446-56.
127. Barriball KL, While AE. Non-response in survey research: a methodological discussion and development of an explanatory model. J Adv Nurs 1999; 30:677-86.

128. Kessler RC, Little RJ, Groves RM. Advances in strategies for minimizing and adjusting for survey nonresponse. Epidemiol Rev 1995; 17:192-204.
129. 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.
130. Soller L, Ben-Shoshan M, Harrington DW, Knoll M, Fragapane J, Joseph L, et al. Adjusting for non-response bias corrects overestimates of food allergy prevalence Journal of Allergy and Clinical Immunology: In Practice Accepted.
131. Siemiatycki J, Campbell S. Nonresponse bias and early versus all responders in mail and telephone surveys. Am J Epidemiol 1984; 120:291-301.
132. Fowler FJ, Jr., Gallagher PM, Stringfellow VL, Zaslavsky AM, Thompson JW, Cleary PD. Using telephone interviews to reduce nonresponse bias to mail surveys of health plan members. Med Care 2002; 40:190-200.
133. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons; 1987.
134. Rubin DB. Inference and missing data. Biometrika 1976; 63:11.
135. Schafer JL, Graham JW. Missing data: Our view of the state of the art. Psychological Methods 2002; 7:147-77.
136. Little RJ, Rubin DB. Statistical Analysis with Missing Data. 2 ed. New York: Wiley; 2002.
137. Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. 2nd Edition ed: Chapman and Hall; 2003.
138. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls; 2009.
139. Working Group of the Resuscitation Council UK (Soar J PR, Cant A, et al.). Emergency treatment of anaphylactic reactions: guidelines for healthcare providers. Resuscitation Council (UK) 2008:4.
140. Simons KJ, Simons FE. Epinephrine and its use in anaphylaxis: current issues. Curr Opin Allergy Clin Immunol 2010; 10:354-61.
141. Cianferoni A, Muraro A. Food-induced anaphylaxis. Immunology and allergy clinics of North America 2012; 32:165-95.
142. Sheikh A, Ten Broek V, Brown SG, Simons FE. H1-antihistamines for the treatment of anaphylaxis: Cochrane systematic review. Allergy 2007; 62:830-7.
143. Choo KJ, Simons E, Sheikh A. Glucocorticoids for the treatment of anaphylaxis: Cochrane systematic review. Allergy 2010; 65:1205-11.
144. Sampson HA. Fatal food-induced anaphylaxis. Allergy 1998; 53:125-30.
145. Sampson HA, Mendelson L, Rosen JP. Fatal and near-fatal anaphylactic reactions to food in children and adolescents. N Engl J Med 1992; 327:380-4.
146. Bock SA, Munoz-Furlong A, Sampson HA. Fatalities due to anaphylactic reactions to foods. The Journal of allergy and clinical immunology 2001; 107:191-3.
147. Kastner M, Harada L, Waserman S. Gaps in anaphylaxis management at the level of physicians, patients, and the community: a systematic review of the literature. Allergy 2010; 65:435-44.
148. Bansal PJ, Marsh R, Patel B, Tobin MC. Recognition, evaluation, and treatment of anaphylaxis in the child care setting. Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology 2005; 94:55-9.
149. Gaeta TJ, Clark S, Pelletier AJ, Camargo CA. National study of US emergency department visits for acute allergic reactions, 1993 to 2004. Ann Allergy Asthma Immunol 2007; 98:360-5.

150. Asai Y, Yanishevsky Y, Clarke A, La Vieille S, Delaney JS, Alizadehfar R, et al. Rate, Triggers, Severity and Management of Anaphylaxis in Adults Treated in a Canadian Emergency Department. International archives of allergy and immunology 2014; 164:246-52.
151. Ben-Shoshan M, La Vieille S, Eisman H, Alizadehfar R, Mill C, Perkins E, et al. Anaphylaxis treated in a Canadian pediatric hospital: Incidence, clinical characteristics, triggers, and management. The Journal of allergy and clinical immunology 2013; 132:739-41 e3.
152. Grabenhenrich L, Hompes S, Gough H, Rueff F, Scherer K, Pfohler C, et al. Implementation of anaphylaxis management guidelines: a register-based study. PloS one 2012; 7:e35778.
153. Liew WK, Williamson E, Tang ML. Anaphylaxis fatalities and admissions in Australia. The Journal of allergy and clinical immunology 2009; 123:434-42.
154. Yunginger JW, Sweeney KG, Sturner WQ, Giannandrea LA, Teigland JD, Bray M, et al. Fatal food-induced anaphylaxis. JAMA 1988; 260:1450-2.
155. Pumphrey RS. Lessons for management of anaphylaxis from a study of fatal reactions. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology 2000; 30:1144-50.
156. Clark S, Camargo CA, Jr. Emergency management of food allergy: systems perspective. Curr Opin Allergy Clin Immunol 2005; 5:293-8.
157. Schwartz HJ. Acute allergic disease in a hospital emergency room: a retrospective evaluation of one year's experience. Allergy Proc 1995; 16:247-50.
158. Imai T, Sugizaki C, Ebisawa M. [Investigation on the usage and prescription of adrenaline self-injector against anaphylaxis]. Arerugi 2008; 57:722-7.
159. Flokstra-de Blok BM, Doriene van Ginkel C, Roerdink EM, Kroeze MA, Stel AA, van der Meulen GN, et al. Extremely low prevalence of epinephrine autoinjectors in high-risk food-allergic adolescents in Dutch high schools. Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology 2011; 22:374-7.
160. Sampson MA, Muñoz-Furlong A, Sicherer SH. Risk-taking and coping strategies of adolescents and young adults with food allergy. Journal of Allergy and Clinical Immunology 2006; 117:1440-5.
161. Asero R, Antonicelli L, Arena A, Bommarito L, Caruso B, Colombo G, et al. Epinephrine autoinjector prescription in food-allergic adults: symptom-based only or allergen-based also? An Italian multi-centre study. Eur Ann Allergy Clin Immunol 2010; 42:25-31.
162. Simons FE, Peterson S, Black CD. Epinephrine dispensing for the out-of-hospital treatment of anaphylaxis in infants and children: a population-based study. Ann Allergy Asthma Immunol 2001; 86:622-6.
163. Grabovschi C, Loignon C, Fortin M. Mapping the concept of vulnerability related to health care disparities: a scoping review. BMC Health Serv Res 2013; 13:94.
164. Watt L, Gulati S, Shaw N, Sung L, Dix D, Poureslami I, et al. Perceptions about complementary and alternative medicine use among Chinese immigrant parents of children with cancer. Supportive Care in Cancer 2012; 20:253-60.
165. Mehl A, Wahn U, Niggemann B. Anaphylactic reactions in children – a questionnaire-based survey in Germany. Allergy 2005; 60:1440-5.
166. Song TT, Worm M, Lieberman P. Anaphylaxis treatment: current barriers to adrenaline auto-injector use. Allergy 2014; 69:983-91.
167. Smith W, Chey T, Jalaludin B, Salkeld G, Capon T. Increasing response rates in telephone surveys: a randomized trial. Journal of Public Health 1995; 17:33-8.

168. Food Allergy & Anaphylaxis Public Declaration. Zurich: EAACI; 2013.] Available from http://webr.emv2.com/eaaci/EAACI_Anaphylaxis_Flyer_English.pdf.
169. Grenier G, Xue L. Canadian Immigrants' Access to a First Job in Their Intended Occupation. Journal of International Migration and Integration 2011; 12:275-303.
170. Pumphrey R. When should self-injectible epinephrine be prescribed for food allergy and when should it be used? Curr Opin Allergy Clin Immunol 2008; 8:254-60.
171. Grimm P. Social Desirability Bias. In: Wiley International Encyclopedia of Marketing: John Wiley & Sons, Ltd; 2010.
172. Richardson KL, Driedger MS, Pizzi NJ, Wu J, Moghadas SM. Indigenous populations health protection: a Canadian perspective. BMC public health 2012; 12:1098.
173. Soller L, Cherkaoui S, Ben-Shoshan M, Harrington DW, Knoll M, Fragapane J, et al. Likelihood of being prescribed an epinephrine autoinjector in allergic Canadians with lower educational levels. Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology 2014.

APPENDIX A: DEFINITIONS

Definitions of vulnerable populations

Vulnerable populations

The populations we targeted in this study, i.e., those of low income and education, new Canadians, and individuals of Aboriginal identity, are considered to be 'vulnerable' because they are more likely to experience issues accessing adequate healthcare services and are more likely to suffer from ill health, compared with the general population.

Low income

Low income households are considered to be those with a household income below the Low Income Cut-Off (LICO), defined as the income level at which families or unattached individuals spend at least 70% of before tax income on food, shelter and clothing. It is determined according to family size and geographic location.

Low education

Adults at least 18 years old with less than a post-secondary degree, or highest educational attainment in the household less than a post-secondary education for those under 18.

New Canadian

A new Canadian is someone who has immigrated to Canada in the last 10 years.

Aboriginal identity

Under Section 35 of the Constitution Act (1982), "Aboriginal Peoples of Canada" includes Indian (First Nation), Inuit, and Métis peoples of Canada. (http://laws-lois.justice.gc.ca/eng/Const/page-16.html).

Definitions of food allergy

Perceived food allergy

Includes all individuals reporting any food allergy.

Probable food allergy

A more conservative definition which includes all individuals reporting an allergy to peanut, tree nut, fish, shellfish, sesame, milk, egg, wheat, and/or soy, who report a convincing history of food allergy and/or who self-report a physician-diagnosed food allergy. To ensure that participants who were either lactose intolerant or had celiac disease were not mistakenly considered to have a milk or wheat allergy, those who reported either of these conditions or had symptoms which were limited to the gastrointestinal tract or who could tolerate either dairy or wheat products occasionally without experiencing a reaction were excluded from the estimates for probable milk or wheat allergy.

Convincing history

To be considered to have a convincing historyan individual had to report experiencing at least two mild symptoms (pruritus, urticaria, flushing, or rhinoconjunctivitis), one moderate (angioedema, throat tightness, gastrointestinal complaints, or breathing difficulties (other than wheeze)) or one severe symptom (wheeze, cyanosis, or circulatory collapse) after ingestion or contact (or inhalation for fish, shellfish, egg or soy) within 2 hours after exposure to the food.

Definitions for the Epinephrine Auto-Injector (EAI)

Prescription of the EAI

The individual reports having ever been prescribed the epinephrine auto-injector for their food allergy.

Availability of the EAI

The individual has been prescribed the EAI, and reports carrying it with them at all times when they leave the house.

APPENDIX B: ADDITIONAL INFORMATION ON SAMPLING METHODS

In order to involve the vulnerable populations, we used an approach that would target low socioeconomic status (SES) Canadians, new Canadians, and residents of the territories. We first selected all census metropolitan areas (CMAs) from the 2006 Canadian census, then a master file of census tracts (CTs) from within the CMAs. Two files were constructed from this list; one which sorted all CTs based on the proportion of households living under the low income cut-off (LICO), a second with all CTs sorted based on the proportion of residents who reported migrating to Canada since 1996. The 100 CTs with the highest proportion of each were then selected and merged into one file. Duplicate CTs were removed. The resulting list were converted to postal codes using the 2006 Statistics Canada postal code conversion file (PCCF) available via the Computing for Humanities and Social Sciences server at the University of Toronto. The resulting postal code file contained 12,785 six-digit postal codes from which Info-Direct could select a random sample of telephone numbers. The aboriginal population was targeted in a similar way, focusing on the CTs that reported the highest proportion of self-identified Aboriginal people living in urban areas. Using a lower threshold of 15% Aboriginal, this added 8,344 six-digit postal codes to the Info-Direct file.

Due to the targeted strategy employed, certain regions were not represented in the postal code file. The provinces of New Brunswick, Nova Scotia, Newfoundland, and Prince Edward Island were not represented proportionally. Specifically, Nova Scotia, Newfoundland, and PEI were excluded completely. The three territories (Nunavut, Yukon, and the Northwest Territories) were also completely excluded based on the sampling strategy. A multi-pronged strategy was used in

order to include these regions. For the provinces of New Brunswick, Nova Scotia and Newfoundland, CTs were selected from the main CMAs (Saint John, Halifax and St. John's) based on the proportion of households under the LICO .

St. John's - There are 8 total CTs in this CMA. We included the top 5 by LICO in the sample (ranged from 27.4% to 41.4%)

Halifax - There are 13 total CTs in this CMA. We included the top 10 by LICO in the sample (ranged from 24.1% to 40.9%)

Saint John - There are 14 total CTs in this CMA. Two of these were captured by the original sampling strategy, so we included another 8 (range of LICO from 25.8% to 38.9%).

These areas were not targeted based on the proportion of new Canadians based on the low proportions of these populations in these regions. After conversion to six-digit postal codes, Info-Direct randomly selected a proportional number of phone numbers that were added to the initial file.

Since Prince Edward Island does not have any CMAs as defined by Statistics Canada, we targeted the largest Census Subdivision in the province, Charlottetown (population = 32174). A proportional number of households were drawn randomly from the postal codes in this area by

Info-Direct, and included in the file. 13.2% of households in Charlottetown were below the LICO in 2006, and the percent of new Canadians was 1.4%.

For the Yukon and Northwest territories, numbers and addresses were randomly selected from all areas via Info-Direct. A similar strategy was used for Nunavut, however, addresses were not available to accompany the phone numbers provided by Info-Direct. Alternatively, the study was advertised through relevant media outlets to create awareness, and increase response rates.

APPENDIX C: INFORMATION LETTER

Centre universitaire de santé

McMaster University

We are interested in your views on the effects of the environment on the health of Canadians!

Your household has been chosen at random to participate in a national telephone survey designed to gather the views of Canadians about the impacts of the environment on our health. Your participation in this survey is completely voluntary; your household was chosen randomly from public telephone listings in your area. Below, you will find answers to key questions about the study. You can also find out more information on the study's web site (http://smaart.mcgill.ca) or you can call or e-mail one of the researchers directly (see contact information below).

You will also find a 5$ **gift card** enclosed as a token of our appreciation for sharing your views with us. Your opinions as well as the views of others living in your community ***are very important*** to get a true picture of the effects of the environment on the health of Canadians, especially our children. ***We sincerely hope you will consider participating.***

What will happen next? About a week after receiving this letter, you will receive a phone call from one of our research assistants at McGill University. If you have call display, the display will read: *Univ. McGill*. That's how you'll know it's us calling!

How long will this survey take? Around 15 minutes if there are no food allergies in your house. If there is a food allergy, it may take a bit longer. If you're busy when the research assistant calls, we can re-schedule for a time that is convenient for ***you***.

Who's paying for this study? This study is being funded by Health Canada and the AllerGen research network, which is based at McMaster University in Hamilton, Ontario.

We can't do this without your help.
If you do participate in the study, all information will be kept anonymous and confidential. If you decide you do not want to participate, you can let us know by (1) Telling the interviewer when they call; (2) Call our toll free number (1-866-431-7344); (3) Send an e-mail to: SPAACE@epimgh.mcgill.ca.

Thank you in advance for considering our request!

Sincerely,

Principal investigators for this research study:
Ann Clarke, MD, McGill University Health Centre - Montreal, Quebec
Susan Elliott, PhD, McMaster University - Hamilton, Ontario

Santé Canada Health Canada

AllerGen

APPENDIX D: PEANUT ALLERGY PREVALENCE QUESTIONNAIRE

Now I am going to ask you a few questions about your experience with peanuts.

1.0) Have you ever had a reaction to TOUCHING peanuts?

No	Refused
Yes	Don't know

1.1) Have you ever had a reaction to SMELLING or INHALING peanuts?

No	Refused
Yes	Don't know

1.2) Have you ever had a reaction to EATING peanuts?

No	Refused
Yes	Don't know

1.4) Have you ever eaten peanuts?

No	Refused
Yes	Don't know

1.6a) How many allergic reactions have you had to peanut in your lifetime?

1.6b) It's important that we try to get an estimate for the number of reactions, can you tell me if it was..

Only 1 reaction	5-10 reactions	Refused
2-5 reactions	More than 10 reactions	Don't know

1.7) About how old where you when you had your FIRST allergic reaction to peanuts?

1.8) It's important that we try to get an estimate...if you're not sure, can you at least tell me if it was...

Before you started school	In High school
In Elementary school	After High school
In Middle school	

1.9a) About how old where you when you had your LAST allergic reaction to peanuts?

1.9b) It's important that we try to get an estimate...if you're not sure, can you at least tell me if it was...

Before you started school	In High school
In Elementary school	After High school
In Middle school	

2.0) How old were you when you had your MOST SEVERE reaction to peanut?

2.1) It's important that we try to get an estimate...if you're not sure, can you at least tell me if it was...

- Before you started school
- In Elementary school
- In Middle school
- In High school
- After High school

2.2) Was the most severe reaction caused by eating, touching, or inhaling peanuts?

- Eating
- Touching
- Inhaling
- Refused
- Don't Know

I am going to read a list of symptoms that may or may not have occurred during the MOST SEVERE reaction, please indicate which one(s) occurred.

3.0) Did you have hives (skin rash, welts, urticaria)?

- No
- Yes
- Refused
- Don't know

3.1) Did you have swelling (edema)?

- No
- Yes
- Refused
- Don't know

3.2) Where did you have the swelling? (Click on all that apply)

- Eyes(eyelids)
- Tongue
- Lips
- Face
- Refused
- Don't know
- Other(specify)

3.3) Did you have nausea or stomach pain?

- No
- Yes
- Refused
- Don't know

3.4) Did you vomit?

- No
- Yes
- Refused
- Don't know

3.5) Did you have diarrhea?

- No
- Yes
- Refused
- Don't know

3.6) Did you start coughing?

- No
- Yes
- Refused
- Don't know

3.7) Did you have trouble breathing?

No	Refused
Yes	Don't know

3.8) Did you start wheezing?

No	Refused
Yes	Don't know

3.9) Did you have an itchy mouth?

No	Refused
Yes	Don't know

3.10) Did you feel any closing or tightening of the throat?

No	Refused
Yes	Don't know

3.11) Did you feel lightheaded or as if you were going to faint?

No	Refused
Yes	Don't know

3.12) Did you have any other symptoms?

No	Refused
Yes(specify)	Don't know

3.13) Now I would like you to think back to your most severe reaction. We would like to know how long it was from when you were exposed to peanuts and when your symptoms started?
INTERVIEWER NOTES: Exposed = eat,inhaled,touched

HH--MM--SS
-- --

Record time
Immediately
Refused
Don't Know

3.14) It's really important that we get an estimate...if you're not sure, can you at least tell me if the symptoms started...

In less than an hour	Refused
More than an hour	Don't know

4.0) Was adrenaline used to treat your most severe reaction to peanuts?

No	Refused
Yes	Don't know

Now think about your allergic reaction(s) to peanut in general...

4.1) Have you ever used any alternative treatments or health care providers for your peanut allergy?

No	Refused
Yes	Don't know

4.2) Which treatments and/or health care providers have you used?

Now we would like to talk to you about how your peanut allergy was diagnosed.

5.0) Has your allergy to peanuts ever been confirmed by a doctor?

No	Refused
Yes	Don't know

5.1) Did the doctor do a skin test?

No	Refused
Yes	Don't know

5.2) Did the skin test show that you are allergic to peanut?

No	Refused
Yes	Don't know

5.3) Did the doctor do a blood test?

No	Refused
Yes	Don't know

5.4) Did the blood test show that you are allergic to peanut?

No	Refused
Yes	Don't know

5.5) Did the doctor do a food challenge?

No	Refused
Yes	Don't know

5.6) Did the test show that you are allergic to peanut?

No	Refused
Yes	Don't know

I am now going to ask you about what you have done SINCE your diagnosis with a peanut allergy

6.0) Since your diagnosis, have you stopped eating peanuts completely?

No	Refused
Yes	Don't know

6.1) Since your diagnosis, have you continued to eat peanuts occasionally with no reaction?

No	Refused

Yes Don't know

6.2) Since your diagnosis, have you continued to eat peanuts occasionally with a reaction?

No Refused
Yes Don't know

APPENDIX E: STATA DO-FILE FOR CALCULATION OF WEIGHTED AND UNWEIGHTED PREVALENCE ESTIMATES AND DEMOGRAPHIC PREDICTORS

```
clear
set more 1
***Lianne work computer

cd "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Prevalence analysis"

***Lianne home computer
***cd "C:\Users\Lianne\Desktop\Documents\My Dropbox\PhD Thesis\Prevalence analysis\"

*** local macros for allergen in dataset and variable names

local a1 "pnvars"
local a2 "tnvars"
local a3 "fhvars"
local a4 "sfvars"
local a5 "ssvars"
local a6 "mkvars"
local a7 "egvars"
local a8 "wtvars"
local a9 "syvars"
local b1 "pn"
local b2 "tn"
local b3 "fh"
local b4 "sf"
local b5 "ss"
local b6 "mk"
local b7 "eg"
local b8 "wt"
local b9 "sy"



*** in the following loop each allergen is dealt with in turn (`i' goes from 1 to 9)

local i=1
while `i'<10 {

*** I use each allergy-specific dta file
```

```
use `a`i", clear

keep if resp!=""
* g obs=_n
g all=`i'

destring `b`i"1_0-`b`i"6_2, replace

*** now I check the elements of the algorithms

*** way=1 if most severe rx's mode of exposure is eating or touching/drinking (or inhaling but
that's
*** only for allergies to fish, shellfish, egg or soy i.e. for `i'==3 or `i'==4 or `i'==7 or `i'==9)

g way=`b`i"2_2<3|((`i'>2&`i'<5|`i'==7|`i'==9)&`b`i"2_2==3)

*** mild=1 if 2 mild symptoms are reported (out of hives, diarrhea and itchy mouth)

g mild=(`b`i"3_0==2&`b`i"3_5==2)|(`b`i"3_0==2&`b`i"3_9==2)|(`b`i"3_5==2&`b`i"3_9==2)

*** sx are convincing, i.e. symp=1 if mild=1 or one moderate or severe sx reported
*** (that's either edema, stomach pain, vomiting, coughing, trouble breathing, wheezing,
*** throat tightening or fainting) - note: manual check are to be made afterwards for
*** non-probable cases reporting open-ended other sx

g
symp=mild==1|`b`i"3_1_==2|`b`i"3_3==2|`b`i"3_4==2|`b`i"3_6==2|`b`i"3_7==2|`b`i"3_8==2|`b`
i"3_10_==2|`b`i"3_11_==2

*** time=1 if sx occurred immediately, or no more than 2 hours afer exposure, or estimated
*** as within an hour if specific timing is not recorded

g time=`b`i"3_13_==2|`b`i"3_13h_+(`b`i"3_13m_/60)<=2|`b`i"3_14_==1

***nongisymp =1 if non-gi symptoms present (mk or wt only)

g
nongisymp=(`i'~=6&`i'~=8)|`b`i"3_0==2|`b`i"3_1_==2|`b`i"3_6==2|`b`i"3_7==2|`b`i"3_8==2|`b`
i"3_9==2|`b`i"3_10_==2|`b`i"3_11_==2

*** cvhx (convincing history) = 1 if way=1 and symp=1 and time=1 and nongisymp=1

g cvhx=(way+symp+time+nongisymp==4)
```

```
*** now, this is when pt says her allergy was confirmed by an md

g mdcf=`b`i"5_0==2

*** probable cases are then defined as subjects with cvhx=1 or mdcf=1

g prC=cvhx+mdcf>0

***remove those with dx of lactose intolerance or who can tolerate yogurt and eat dairy
sometimes with no rxn

if `i'==6 {
replace prC=0 if `b`i"5_9==1|`b`i"1_51==1|`b`i"6_1==2
}

***remove those with dx of celiac disease or who can eat wheat sometimes with no rxn

if `i'==8 {
replace prC=0 if `b`i"5_8==1|`b`i"6_1==2
}

*** the rest of the loop just builds the resulting datasets

keep resp subject all-prC

if `i'>1 {
append using allergy0
}
sort all resp subject
save allergy0, replace
local i=`i'+1
}

keep resp subject all prC
reshape wide prC, i(resp subject) j(all)
sort resp subject
save allergy1, replace

*** now define self-reported allergies using 2 nested loops
*** note that `k'=`i'+1 for allergies 6 to 9 because 6 is other allergy in the data

use pnvars
local i=1
while `i'<10 {
```

```
local k=`i'+(`i'>5)
g srA`i'=0
local j=1
while `j'<8 {
replace srA`i'=1 if real(which_allergy_`j'_)==`k'
local j=`j'+1
}
local i=`i'+1
}
keep resp subject srA*
sort resp subject

*** now merging the self-reports with the probable cases

merge resp subject using allergy1
drop _merge
sort resp subject
save allergy1, replace

*Creating the variable for other allergy

use pnvars
g srAoth=0
local j=1
while `j'<8 {
replace srAoth=1 if real(which_allergy_`j'_)==6
local j=`j'+1
}
keep resp subject srAoth
sort resp subject
merge resp subject using allergy1
drop _merge
egen srAny=rowtotal(srA*)
replace srAny=srAny>0
sort resp subject
save allergy1, replace

* For our first step, we start by handling a few implausible values for year of birth and
* create our merge key (mgID) in the demographic and allergy data.

use sdvars1a, clear
replace demo1_3=demo1_3+1900 if demo1_3>9&demo1_3<100
replace demo1_3=1979 if demo1_3==979
replace demo1_3=1985 if demo1_3==198
```

```
replace demo1_3=demo1_3+100 if demo1_3>9&demo1_3<1900
g mgID=string(demo1_2)+"/"+string(demo1_3)+"/"+string(demo1_4)
keep resp subdemo mgID
sort resp mgID subdemo
save M1
use syvars
replace resp_yob="19"+resp_yob if real(resp_yob)>9&real(resp_yob)<100
g mgID=resp_sex_+"/"+resp_yob_+"/"+resp_mob_
sort resp subject
merge resp subject using pnvars
keep if real(which_allergy_1_)~=.
keep resp subject mgID
sort resp mgID

* we now proceed with a first attempt at merging, and then identify problem cases,
* i.e. allergy subjects that don't have a counterpart in demographics (_merge==1),
* and those that have more than 1 (moreth1==1), and export them in spreadsheet format
* with the other observations from their households.

merge resp mgID using M1
sort resp mgID
save M1, replace
qui by resp mgID: g moreth1=(_N>1)&_merge==3
g check=_merge==1|moreth1==1
egen problem=max(check), by(resp)
keep if problem==1
drop _merge problem
sort resp subject subdemo
outsheet using problems

* next, decisions made wrt "problems.xls" are manually implemented (from problems.out/xls
* to merge_chge.csv) and then returned in Stata format for the second step - prvsetup2.do


* Here's the rest of the data setup and production of the statistics.

* First, we replace the missed merges with our manual changes.
* Note that we exclude some households if merging was impossible (exclA) or prepare to do if
* merging was uncertain wrt either education, income or gender (exclE,I,G).

insheet using merge_chge.csv, clear
append using M1
sort resp subdemo _merge
qui by resp subdemo: drop if _merge==1|(_N>1&_merge~=.)
```

```
sort resp subject
qui by resp subject: g x=_N
tab x
replace subject=. if _merge==3&x==2
egen exclA=max(exa), by(resp)
drop if exclA==1
egen exclE=max(exe), by(resp)
egen exclI=max(exi), by(resp)
egen exclG=max(exg), by(resp)
drop exa-exg _merge x exclA
sort resp subdemo

* we now retrieve the demographic data and define our stratification variables.

merge resp subdemo using sdvars1a
drop if _merge<3|(demo1_2>2& demo1_3<10& demo1_9>13)
replace demo1_3=demo1_3+1900 if demo1_3>9&demo1_3<100
replace demo1_3=1979 if demo1_3==979
replace demo1_3=1985 if demo1_3==198
replace demo1_3=demo1_3+100 if demo1_3>9&demo1_3<1900
replace exclE=0 if exclE==.
replace exclI=0 if exclI==.
replace exclG=0 if exclG==.
g male=demo1_2==1 if demo1_2<3
g dob=date("15/"+string(demo1_4)+"/"+string(demo1_3),"DMY")
replace dob=date("1/7/"+string(demo1_3),"DMY") if dob==.|demo1_4<0
g immig=index(upper(demo1_5),"CAN")~=1|demo1_6~=. if demo1_6~=.|demo1_5~=""
g immlt10y=immig==1&(demo1_6num<10|demo1_6a==1) if
immig~=.&demo1_6num>=0&demo1_6~=2&(demo1_6~=3|demo1_6a<4)
g aborg=demo1_8>1&demo1_8<5 if demo1_8<5|demo1_8==.
g postsecd=demo1_9==8|demo1_9>9 if demo1_9<14
drop mgID demo* _merge
sort resp
save M1, replace

* the next part is for computing age and setting territory and low income variables.
* note that 2 files (csize.dta and lico7910.csv) are derived from external sources -
* "csize" is based on mapping postal codes (fsa = 1st 3 digits) to community size,
* and "lico7910" is a file with low income cut-off values from Stats Can.

insheet using sampleinfospaace.csv, clear
g resp=string(v1)+"s"
g dateQ=date(v3,"MD20Y")
rename v5 prov
```

```
g fsa=substr(v6,1,3)
drop v*
save hhdata
insheet using sampleinfonunavut.csv, clear
g resp=string(v1)+"n"
g dateQ=date(v3,"MD20Y")
g prov="NU"
drop v*
append using hhdata
sort resp
save hhdata, replace
use hhdemo
sort resp
merge resp using hhdata
keep if _merge==3|demo2_2i~=""
replace fsa=upper(substr(demo2_2i,1,3)) if fsa==""&prov~="NU"
replace fsa="H7G" if fsa=="X7G"
replace prov="NS" if fsa=="B2L"
replace prov="NB" if fsa=="E2K"
replace dateQ=18696 if dateQ==.
g tno=prov=="NU"|prov=="NT"|prov=="YT"
g income=real(demo2_3) if real(demo2_3)>=0
replace income=5000+10000*(real(demo2_3a)-1) if income==.&real(demo2_3a)<11
replace income=160000 if income==.&real(demo2_3a)==11
keep resp dateQ tno fsa income
sort resp
save hhdata, replace
use M1

* the next couple of lines define family size for the purpose of applying LICOs

g fsize=1
collapse (sum) fsize, by(resp)
replace fsize=7 if fsize==8

sort resp
merge resp using hhdata
keep if _merge==3
drop _merge
sort fsa
merge fsa using csize
drop if _merge==2
drop _merge
replace csize="E" if csize==""
```

```
sort fsize csize
save hhdata, replace
insheet using lico7910.csv, clear
sort fsize csize
merge fsize csize using hhdata
drop if _merge==1
g lowinc=income<lico10 if income+lico10~=.
keep resp dateQ tno lowinc
sort resp
merge resp using M1
keep if _m==3
g age=(dateQ-dob)/365.25
replace age=0 if age<0
replace postsecd=. if age<18
g child=age<18 if age~=.
drop dateQ dob _merge
sort resp subject

* now we retrieve the allergy data that was set up earlier this year.

merge resp subject using allergy1
drop if _merge==2
drop _merge

*Removing implausible "other allergies"

replace srAoth=0 if (resp=="11940s"&subject==1)
replace srAoth=0 if (resp=="11941s"&subject==1)
replace srAoth=0 if (resp=="3994n"&subject==1)
replace srA2=1 if (resp=="3994n"&subject==1)
replace srAoth=0 if (resp=="2928s"&subject==4)
replace srAoth=0 if (resp=="3071s"&subject==1)
replace srAoth=0 if (resp=="13052s"&subject==1)
replace srA3=1 if (resp=="3714s"&subject==1)
replace srA4=1 if (resp=="3714s"&subject==1)
replace srAoth=0 if (resp=="16498s"&subject==1)
replace srAoth=0 if (resp=="3299n"&subject==4)
replace srAoth=0 if (resp=="18888s"&subject==4)
replace srAoth=0 if (resp=="12202s"&subject==1)
replace srAoth=0 if (resp=="6010s"&subject==1)
replace srAoth=0 if (resp=="4184n"&subject==1)
replace srA3=1 if (resp=="4184n"&subject==1)
replace srAoth=0 if (resp=="11740s"&subject==1)
replace srAoth=0 if (resp=="14968s"&subject==4)
```

```
replace srAoth=0 if (resp=="14723s"&subject==1)
replace srAoth=0 if (resp=="14968s"&subject==1)
replace srAoth=0 if (resp=="15436s"&subject==1)
replace srAoth=0 if (resp=="6301s"&subject==1)
replace srAoth=0 if (resp=="6301s"&subject==4)
replace srAoth=0 if (resp=="273n"&subject==1)
replace srAoth=0 if (resp=="14766s"&subject==1)
replace srAoth=0 if (resp=="14766s"&subject==4)
replace srAoth=0 if (resp=="14824s"&subject==1)
replace srAoth=0 if (resp=="18895s"&subject==1)
replace srAoth=0 if (resp=="4293s"&subject==1)
replace srAoth=0 if (resp=="3120n"&subject==1)
replace srAoth=0 if (resp=="3745s"&subject==1)
replace srAoth=0 if (resp=="7947s"&subject==1)
replace srAoth=0 if (resp=="7290s"&subject==1)
replace srA7=1 if (resp=="7290s"&subject==1)
replace srAoth=0 if (resp=="11063s"&subject==2)
replace srAoth=0 if (resp=="10865s"&subject==1)
replace srAoth=0 if (resp=="1140n"&subject==1)
replace srAoth=0 if (resp=="14770s"&subject==1)
replace srAoth=0 if (resp=="3421s"&subject==1)
replace srAoth=0 if (resp=="3479s"&subject==1)
replace srAoth=0 if (resp=="3758n"&subject==1)
replace srAoth=0 if (resp=="14264s"&subject==1)
replace srAoth=0 if (resp=="12113s"&subject==1)
replace srAoth=0 if (resp=="10158s"&subject==1)
replace srAoth=0 if (resp=="11063s"&subject==1)
replace srAoth=0 if (resp=="11269s"&subject==1)
replace srAoth=0 if (resp=="11286s"&subject==4)
replace srAoth=0 if (resp=="12329s"&subject==1)
replace srAoth=0 if (resp=="12574s"&subject==1)
replace srAoth=0 if (resp=="13217s"&subject==1)
replace srAoth=0 if (resp=="14673s"&subject==1)
replace srAoth=0 if (resp=="14833s"&subject==1)
replace srAoth=0 if (resp=="15514s"&subject==4)
replace srAoth=0 if (resp=="17953s"&subject==1)
replace srAoth=0 if (resp=="1868s"&subject==1)
replace srAoth=0 if (resp=="2461n"&subject==1)
replace srAoth=0 if (resp=="2818n"&subject==1)
replace srAoth=0 if (resp=="3135n"&subject==1)
replace srAoth=0 if (resp=="3749n"&subject==1)
replace srAoth=0 if (resp=="4376n"&subject==4)
replace srAoth=0 if (resp=="482n"&subject==1)
replace srAoth=0 if (resp=="2639s"&subject==1)
```

```
replace srAoth=0 if (resp=="12202s"&subject==2)
replace srAoth=0 if (resp=="4250n"&subject==1)
replace srAoth=0 if (resp=="4250n"&subject==2)
replace srAoth=0 if (resp=="11915s"&subject==1)
replace srAoth=0 if (resp=="12919s"&subject==4)
replace srAoth=0 if (resp=="132n"&subject==1)
replace srAoth=0 if (resp=="13703s"&subject==1)
replace srAoth=0 if (resp=="15458s"&subject==4)
replace srAoth=0 if (resp=="16650s"&subject==1)
replace srAoth=0 if (resp=="18980s"&subject==1)
replace srAoth=0 if (resp=="2818s"&subject==1)
replace srAoth=0 if (resp=="9962s"&subject==1)
replace srAoth=0 if (resp=="14270s"&subject==1)
replace srAoth=0 if (resp=="15935s"&subject==2)
replace srAoth=0 if (resp=="5011s"&subject==1)
replace srAoth=0 if (resp=="9668s"&subject==1)
replace srAoth=0 if (resp=="7866s"&subject==1)
replace srA6=1 if (resp=="7866s"&subject==1)
replace srAoth=0 if (resp=="16423s"&subject==1)
replace srAoth=0 if (resp=="12956s"&subject==1)
replace srAoth=0 if (resp=="13007s"&subject==1)
replace srAoth=0 if (resp=="1316n"&subject==1)
replace srAoth=0 if (resp=="16471s"&subject==1)
replace srAoth=0 if (resp=="16845s"&subject==1)
replace srAoth=0 if (resp=="17205s"&subject==2)
replace srAoth=0 if (resp=="1853n"&subject==1)
replace srAoth=0 if (resp=="2626n"&subject==1)
replace srAoth=0 if (resp=="3601s"&subject==4)
replace srAoth=0 if (resp=="9203s"&subject==2)
replace srAoth=0 if (resp=="15466s"&subject==4)
replace srAoth=0 if (resp=="17713s"&subject==1)
replace srAoth=0 if (resp=="613n"&subject==1)
replace srAoth=0 if (resp=="17772s"&subject==4)
replace srAoth=0 if (resp=="20168s"&subject==2)
replace srAoth=0 if (resp=="11286s"&subject==2)
replace srAoth=0 if (resp=="12731s"&subject==3)
replace srAoth=0 if (resp=="16660s"&subject==2)
replace srAoth=0 if (resp=="5784s"&subject==1)
replace srA7=1 if (resp=="5784s"&subject==1)
replace srAoth=0 if (resp=="16101s"&subject==3)
replace srA1=1 if (resp=="16101s"&subject==3)
replace srAoth=0 if (resp=="16900s"&subject==1)
replace srA1=1 if (resp=="16900s"&subject==1)
replace srAoth=0 if (resp=="5216s"&subject==1)
```

```
replace srA1=1 if (resp=="5216s"&subject==1)
replace srAoth=0 if (resp=="10920s"&subject==2)
replace srAoth=0 if (resp=="12626s"&subject==4)
replace srAoth=0 if (resp=="12416s"&subject==1)
replace srAoth=0 if (resp=="17713s"&subject==2)
replace srAoth=0 if (resp=="12776s"&subject==1)
replace srAoth=0 if (resp=="11705s"&subject==1)
replace srA7=1 if (resp=="11705s"&subject==1)
replace srA8=1 if (resp=="11705s"&subject==1)
replace srA9=1 if (resp=="11705s"&subject==1)
replace srAoth=0 if (resp=="4895s"&subject==1)
replace srAoth=0 if (resp=="314s"&subject==1)
replace srAoth=0 if (resp=="10850s"&subject==1)
replace srAoth=0 if (resp=="10687s"&subject==1)
replace srAoth=0 if (resp=="10687s"&subject==2)
replace srAoth=0 if (resp=="11079s"&subject==1)
replace srAoth=0 if (resp=="11239s"&subject==1)
replace srAoth=0 if (resp=="10995s"&subject==1)
replace srAoth=0 if (resp=="10410s"&subject==1)
replace srAoth=0 if (resp=="19776s"&subject==1)
replace srAoth=0 if (resp=="2791s"&subject==1)
replace srA7=1 if (resp=="2791s"&subject==1)
replace srAoth=0 if (resp=="6439s"&subject==1)
replace srAoth=0 if (resp=="17568s"&subject==1)
replace srAoth=0 if (resp=="2436n"&subject==1)

*Generating categorical immigrant variable
g immlt10yc=.
replace immlt10yc=1 if (immlt10y==1 & immlt10y!=.)
replace immlt10yc=2 if (immig==1 & immlt10y==0 & immig!=. & immlt10y!=.)
replace immlt10yc=3 if (immig==0 & immig!=.)

replace srAny=0 if srAny==.
replace srAoth=0 if srAoth==.
local i=1
while `i'<10 {
replace srA`i'=0 if srA`i'==.
replace prC`i'=0 if prC`i'==.
local i=`i'+1
}
sort resp subdemo
compress
save M1, replace
```

```
*Finally, we get the prevalence estimates for each allergen and for each group of interest.
*Note, these are the crude prevalence estimates and do not take into account targeted sampling.
log using prev2012.log
sum srA* prC*
ci srA* prC* if immig==1, binomial
ci srA* prC* if immig==0, binomial
ci srA* prC* if immlt10y==1, binomial
ci srA* prC* if immlt10y==0, binomial
ci srA* prC* if immlt10yc==1, binomial
ci srA* prC* if immlt10yc==2, binomial
ci srA* prC* if immlt10yc==3, binomial
ci srA* prC* if lowinc==1, binomial
ci srA* prC* if lowinc==0, binomial
ci srA* prC* if postsecd==0, binomial
ci srA* prC* if postsecd==1, binomial
ci srA* prC* if aborg==1, binomial
ci srA* prC* if aborg==0, binomial
ci srA* prC* if tno==1, binomial
ci srA* prC* if tno==0, binomial
ci srA* prC* if child==1, binomial
ci srA* prC* if child==0, binomial
ci srA* prC* if male==0, binomial
ci srA* prC* if male==1, binomial
log close

*Now we create a new Stata file
save finalprevalencedata.dta


save finalprevalencedata_weights.dta, replace

log using weightedprev.log
svy: mean srA* prC*
svy: mean srA* prC* if immig==1
svy: mean srA* prC* if immig==0
svy: mean srA* prC* if lowinc==1
svy: mean srA* prC* if lowinc==0
svy: mean srA* prC* if postsecd==1
svy: mean srA* prC* if postsecd==0
svy: mean srA* prC* if aborg==1
svy: mean srA* prC* if aborg==0
svy: mean srA* prC* if tno==1
svy: mean srA* prC* if tno==0
svy: mean srA* prC* if child==1
```

```
svy: mean srA* prC* if child==0
svy: mean srA* prC* if male==1
svy: mean srA* prC* if male==0
log close

*The second option is to use the same categories as above, except for the immigrant variable, where we will now use the cut-off of 10 years.
*This is the option that is used in the manuscript, since we are interested in new Canadians specifically.

g immlt10ycst="1" if immlt10yc==1
replace immlt10ycst="2" if immlt10yc==2
replace immlt10ycst="3" if immlt10yc==3
replace immlt10ycst="." if immlt10yc==.

*We create the string variable again
g categoriesimmig10=territoriesst+immlt10ycst+aboriginalst+postsecst+lowincst if tno==0
replace categoriesimmig10=territoriesst+immlt10ycst+aboriginalst+postsecst+missing if tno==1

*These next strings are for when there is one variable unknown.
*The number of each string corresponds to which variable is missing
g categories5immig10=territoriesst+immlt10ycst+aboriginalst+postsecst+missing
g categories4immig10=territoriesst+immlt10ycst+aboriginalst+missing+lowincst
g categories3immig10=territoriesst+immlt10ycst+missing+postsecst+lowincst
g categories2immig10=territoriesst+missing+aboriginalst+postsecst+lowincst

*These next strings are for when there is more than one variable unknown
g categories345immig10=territoriesst+immlt10ycst+missing+missing+missing
g categories34immig10=territoriesst+immlt10ycst+missing+missing+lowincst
g categories35immig10=territoriesst+immlt10ycst+missing+postsecst+missing
g categories45immig10=territoriesst+immlt10ycst+aboriginalst+missing+missing
g categories245immig10=territoriesst+missing+aboriginalst+missing+missing
g categories25immig10=territoriesst+missing+aboriginalst+postsecst+missing
g categories24immig10=territoriesst+missing+aboriginalst+missing+lowincst

tab categories5immig10
tab categories4immig10
tab categories3immig10
tab categories2immig10
tab categories345immig10
tab categories34immig10
tab categories35immig10
tab categories45immig10
tab categories245immig10
```

```
tab categories25immig10
tab categories24immig10

*Now I need to merge my prevalence data with the weights spreadsheet
sort categoriesimmig10
save finalprevalence.dta, replace
insheet using spaace_categories_10yrcutoff.csv,clear
rename categories categoriesimmig10
sort categoriesimmig10
merge categoriesimmig10 using finalprevalence
drop spaace census
drop categories5immig10-categories24immig10
drop _merge

*Finally, I calculate weighted prevalence estimates
svyset resp [pweight=weight]

save finalprevalencedata_weightsimmig10.dta

log using weightedprev_immig10.log
svy: mean srA* prC*
svy: mean srA* prC* if immlt10yc==1
*Should use the coding below but doesn't make a difference in the results
*svy, subpop(immlt10yc if immlt10yc==1): mean srA* prC*
svy: mean srA* prC* if immlt10yc==2
svy: mean srA* prC* if immlt10yc==3
svy: mean srA* prC* if lowinc==1
svy: mean srA* prC* if lowinc==0
svy: mean srA* prC* if postsecd==1
svy: mean srA* prC* if postsecd==0
svy: mean srA* prC* if aborg==1
svy: mean srA* prC* if aborg==0
svy: mean srA* prC* if tno==1
svy: mean srA* prC* if tno==0
svy: mean srA* prC* if child==1
svy: mean srA* prC* if child==0
svy: mean srA* prC* if male==1
svy: mean srA* prC* if male==0
log close

*Investigating various prevalence estimates for adults and children of different immigration
status
log using immigrantchild.log
svy: mean srA* prC*, over(immlt10yc child)
```

```
log close

*Cut-offs for various ages

*Creating age variable for logistic regression analysis
g agecat=.
replace agecat=1 if age<=2 & age!=.
replace agecat=2 if age>2 & age<=5 & age!=.
replace agecat=3 if age>5 & age<18 & age!=.
replace agecat=4 if age>=18 & age!=.

g agecat2=.
replace agecat2=1 if (age<5 & age!=.)
replace agecat2=2 if (age>=5 & age<18 & age!=.)
replace agecat2=3 if agecat==4

log using weightedprev_agecutoffs.log
svy: mean srA* prC* if agecat2==1
svy: mean srA* prC* if agecat2==2
svy: mean srA* prC* if agecat2==3
log close

*Preparing dataset for multivariate regression analysis

*First, I will create a new education variable with value=1 if postsec education and 0 if no
education or age<18 and . otherwise
gen postsecd2=.
replace postsecd2=1 if postsecd==1
replace postsecd2=0 if postsecd==0
replace postsecd2=0 if child==1

*Then, I will create dummy variables for the 3 age levels (reference category=under5)
g age5to18=.
replace age5to18=1 if agecat2==2 & agecat2!=.
replace age5to18=0 if agecat2!=2 & agecat2!=.

g ageover18=.
replace ageover18=1 if agecat2==3 & agecat2!=.
replace ageover18=0 if agecat2!=3 & agecat2!=.

*I will also create dummy variables for the 3 levels of immigrant (reference category=under 10
years)
g over10immig=.
replace over10immig=1 if immlt10yc==2 & immlt10yc!=.
```

```
replace over10immig=0 if (immlt10yc==1 | immlt10yc==3 & immlt10yc!=.)

g borncan=.
replace borncan=1 if immlt10yc==3 & immlt10yc!=.
replace borncan=0 if immlt10yc<3 & immlt10yc!=.

*Now I will investigate confounding and EMM

*Age and sex
g child_male=child*male
logistic srAny child male child_male

g age5to18_male=age5to18*male
g ageover18_male=ageover18*male
logistic srAny age5to18 ageover18 male age5to18_male ageover18_male

*Education and income
*I will create a variable for non-postsecondary income
g lowed=.
replace lowed=1 if postsecd2==0
replace lowed=0 if postsecd2==1
g lowed_lowinc=lowed*lowinc
logistic srAny lowed lowinc lowed_lowinc

*Education and Aboriginal status
g lowed_aborg=lowed*aborg
logistic srAny lowed aborg lowed_aborg

*Income and Aboriginal status
g lowinc_aborg=lowinc*aborg
logistic srAny lowinc aborg lowinc_aborg

*Territories and Education
g tno_lowed=tno*lowed
logistic srAny tno lowed tno_lowed

*Territories and income
g tno_lowinc=tno*lowinc
logistic srAny tno lowinc tno_lowinc

*Territories and Aboriginal status
g tno_aborg=tno*aborg
logistic srAny tno aborg tno_aborg
```

```
*Immigrant and Education
g immig_lowed=immig*lowed
logistic srAny immig lowed immig_lowed

g over10immig_lowed=over10immig*lowed
g borncan_lowed=borncan*lowed
logistic srAny over10immig borncan lowed over10immig_lowed borncan_lowed

*Immigrant and Income
g immig_lowinc=immig*lowinc
logistic srAny immig lowinc immig_lowinc

g over10immig_lowinc=over10immig*lowinc
g borncan_lowinc=borncan*lowinc
logistic srAny over10immig borncan lowinc over10immig_lowinc borncan_lowinc

*Immigrant and Territories
g immig_tno=immig*tno
logistic srAny immig tno immig_tno

g over10immig_tno=over10immig*tno
g borncan_tno=borncan*tno
logistic srAny over10immig borncan tno over10immig_tno borncan_tno

*Then I need to save the dataset before deleting missing data
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Prevalence
analysis\finalprevalencedata_weightsimmig10.dta", replace

*Now I delete any variables I don't need for the multivariate analysis
drop categoriesimmig10 weight subject subdemo exclE exclI exclG immlt10y postsecd age
territoriesst immigst aboriginalst postsecst lowincst categories missing immlt10ycst agecat
agecat2 adult postsecd2 _Ipostsecd2_1 _Ilowinc_1 _IposXlow_1_1 adult_postsecd2 adult

*Now I drop observations with missing values in preparation for BIC
drop if (tno==. | lowinc==. | male==. | immig==. | aborg==. | child==. | immlt10yc ==. |
age5to18==. | ageover18==. | child_male==. | age5to18_male==. | ageover18_male==. |
postsecd2_lowinc==. | postsecd2_aborg==. | lowed==. | lowed_lowinc==. | lowed_aborg==. |
lowinc_aborg==. | tno_lowed==. | tno_lowinc==. | tno_aborg==. | over10immig==. | borncan==.
| immig_lowed==. | over10immig_lowed==. | borncan_lowed==. | immig_lowinc==. |
over10immig_lowinc==. | borncan_lowinc==. | immig_tno==. | over10immig_tno==. |
borncan_tno==.) )

*Then I save the new dataset to be used for the BIC in R and export the dataset into an excel
spreadsheet
```

```
save completedataset.dta, replace
export excel using "C:\Users\Lianne.Soller\Dropbox\PhD
Thesis\Prevalence\Analysis\Prevalence analysis\completedataset.xls", sheetmodify
firstrow(variables)

*With the results from the BIC, I can now run my logistic model for allergy to any food.

svy: logistic srAny i.child i.male i.child##i.male aborg lowinc lowed over10immig borncan
svy: logistic srAny i.child i.male i.child##i.male aborg lowinc lowed immig

*Prevalence estimates age and sex together

log using agesexprev.log
svy: mean srA* prC* if child==1 & male==0
svy: mean srA* prC* if child==1 & male==1
svy: mean srA* prC* if child==0 & male==0
svy: mean srA* prC* if child==0 & male==1
log close

log using childprev.log
svy: mean srA* prC* if child==1 & male==0
svy: mean srA* prC* if child==1 & male==1
svy: mean srA* prC* if child==1 & lowinc==0
svy: mean srA* prC* if child==1 & lowinc==1
svy: mean srA* prC* if child==1 & immig==0
svy: mean srA* prC* if child==1 & immig==1
svy: mean srA* prC* if child==1 & aborg==0
svy: mean srA* prC* if child==1 & aborg==1
log close

svy: logistic srAny male aborg lowinc immig if child==1

*Prevalence according to highest educational attainment in household
save "C:\Users\Lianne\Desktop\Documents\My Dropbox\PhD
Thesis\Prevalence\Analysis\Prevalence analysis\prevalence_highesteducationalattainment.dta"
drop immig ageround subject tno lowinc exclE exclI exclG male immlt10y aborg age immlt10yc
territoriesst immigst aboriginalst postsecst lowincst categories missing immlt10ycst agecat
agecat2 age5to18 ageover18 postsecd2 adult adult_postsecd2 _Ipostsecd2_1 _Ilowinc_1
_IposXlow_1_1 child_male age5to18_male ageover18_male postsecd2_lowinc postsecd2_aborg
lowed lowed_lowinc lowed_aborg lowinc_aborg tno_lowed tno_lowinc tno_aborg over10immig
borncan immig_lowed over10immig_lowed borncan_lowed immig_lowinc over10immig_lowinc
borncan_lowinc borncan_lowinc immig_tno over10immig_tno borncan_tno
```

```
save "C:\Users\Lianne\Desktop\Documents\My Dropbox\PhD Thesis\Prevalence\Analysis\Prevalence analysis\prevalence_highesteducationalattainment.dta", replace
reshape wide postsecd categoriesimmig10 weight srAoth srA1 srA2 srA3 srA4 srA5 srA6 srA7 srA8 srA9 prC1 prC2 prC3 prC4 prC5 prC6 prC7 prC8 prC9 srAny child , i(resp) j(subdemo)
gen highested=0
replace highested=1 if (postsecd1==1 | postsecd2==1 | postsecd3==1 | postsecd4==1 | postsecd5==1 | postsecd6==1 | postsecd7==1 | postsecd8==1)
replace highested=. if (postsecd1==. & postsecd2==. & postsecd3==. & postsecd4==. & postsecd5==. & postsecd6==. & postsecd7==. & postsecd8==.)
save "C:\Users\Lianne\Desktop\Documents\My Dropbox\PhD Thesis\Prevalence\Analysis\Prevalence analysis\prevalence_highesteducationalattainment.dta", replace
reshape long
save "C:\Users\Lianne\Desktop\Documents\My Dropbox\PhD Thesis\Prevalence\Analysis\Prevalence analysis\prevalence_highesteducationalattainment.dta", replace

log using educattain.log
svy: mean srA* prC* if child==1 & highested==1
svy: mean srA* prC* if child==1 & highested==0
log close

log using educattain2.log, replace
svy: mean srA* prC* if (child==1 & highested==1 | child==0 & postsecd==1)
svy: mean srA* prC* if (child==1 & highested==0 | child==0 & postsecd==0)
log close

log using adulteducation, replace
svy: mean srA* prC* if child==0 & postsecd==0
svy: mean srA* prC* if child==0 & postsecd==1
log close

*Checking to see if missing income data makes a difference

svy: mean srA* prC* if lowinc==.
svy: logistic srAny i.child i.male i.child##i.male aborg lowed over10immig borncan if lowinc==.
svy: logistic srAny i.child i.male i.child##i.male aborg lowed over10immig borncan if lowinc!=.


*Create an interaction for immigrant and number of years since immigration
replace numyears=0 if immig==0

*I am creating a random effects model to account for clustering within households
```

```
egen respid=group(resp)
xtset respid
egen hhweight=mean(weight), by(respid)
xtlogit srAny i.child i.male i.child##i.male aborg lowinc lowed over10immig borncan
[weight=hhweight], or
xtlogit srAny i.child i.male i.child##i.male aborg lowinc lowed immig [weight=hhweight], or
xtlogit srAny i.child i.male i.child##i.male aborg lowinc lowed i.immig c.numyears
i.immig##c.numyears[weight=hhweight], or
```

APPENDIX F: REFUSAL QUESTIONNAIRE

Before you go, could we just take 10 seconds of your time to ask you a few quick questions?

No
Yes

1) Does anyone in your house have a food allergy?

No	Refused
Yes	Don't know

2) How many people live in your house?
Record answer:

3) How many are under 18 years old?
Record answer:

4) What is the highest level of schooling that you have finished?
Record answer:

5) You mentioned a food allergy, could you tell me to which foods?
Record answer:

6) Has the allergy been diagnosed by a doctor?

No	Refused
Yes	Don't know

APPENDIX G: DETAILS ON DEVELOPING NON-RESPONSE BIAS ESTIMATES

To account for missing data in our study, we used multiple imputation for both ignorable and non-ignorable missing data, as proposed by Kmetic et al.[129] We created posterior distributions for the prevalence of food allergy for *Full Participants*, *Refusal Questionnaire Participants, Non-Participants*, 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 *Non-Participants* 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 non-response bias from missing food allergy data within the *Non-Participants* 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.[129] Multiple imputation is the gold standard for adjusting for missing data.[137] It involves filling in missing values for the presence or absence of food allergy with a "best guess" which 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.[137] Point estimates and 95% credible intervals

(CrIs) were estimated. A 95% credible interval 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% credible intervals essentially reflect the information in the data.

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

1) In households who completed the Refusal Questionnaire and indicated that one or more members had a food allergy, the number of allergic individuals was imputed since 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 *Non-Participants*, the *Never Reached Participants*, and the non-allergic *Refusal Questionnaire Participants*, based on the distribution of 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 reporting allergy plus the imputed total number of people in households not reporting allergy (described in step 2).

As detailed in the manuscript, three assumptions regarding the prevalence of food allergy in the *Non-Participants,* and three assumptions regarding the prevalence in the *Never Reached Participants*, were investigated, yielding nine different models, as follows:

1) The prevalence in the *Non-Participants* is half that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Non-Participants*,

2) The prevalence in the *Non-Participants* is the same as in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as in the *Non-Participants*,

3) The prevalence in the *Non-Participants* is twice that of the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is the same as the *Non-Participants*,

4) The prevalence in the *Non-Participants* is half that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is a mixture of the *Full Participants*, *Refusal Questionnaire Participants* and *Non-Participants*,

5) The prevalence in the *Non-Participants* is the same as 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 *Non-Participants*,

6) The prevalence in the *Non-Participants* is twice that in the *Refusal Questionnaire Participants*, and the prevalence in the *Never Reached Participants* is a mixture of the *Full Participants*, *Refusal Questionnaire Participants* and *Non-Participants*,

7) The prevalence in the *Non-Participants* 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 *Non-Participants* 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 *Non-Participants* 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 four levels: individual, household, CT, and province of residence. Each model had the same basic structure, as follows:

logit(prevalence)=intercepti+household effect(number of individuals in household)+assumption about prevalence(1-4, unique for each of the four groups of participants)

where the intercept depended on the census tract and province (represented by “i” in the above equation). There were 13 provinces and 265 census tracts. Province of residence and census tract 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).

APPENDIX H: STATA DO-FILE AND WINBUGS PROGRAM FOR NON-RESPONSE BIAS ANALYSIS

Stata Do-File

```
import excel "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias analysis\Copy of SPAACE by sampling proportions (CT)-1.xlsx", sheet("SPAACE by sampling proportions ") firstrow clear
rename RESPNUM respnum
gen source="s"
tostring respnum, replace
replace respnum=respnum+source
drop source
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias analysis\censusdata.dta", replace
merge 1:1 respnum using "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias analysis\dataset.dta"
drop _merge
replace provcode=13 if provcode==.
replace CTname=999 if CTname==.
replace population=29474 if population==.
replace provcode=1 if provcode==10
replace provcode=2 if provcode==11
replace provcode=3 if provcode==12
replace provcode=4 if provcode==13
replace provcode=5 if provcode==24
replace provcode=6 if provcode==35
replace provcode=7 if provcode==46
replace provcode=8 if provcode==47
replace provcode=9 if provcode==48
replace provcode=10 if provcode==59
replace provcode=11 if provcode==60
replace provcode=12 if provcode==61
replace provcode=13 if CTname==999
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias analysis\censusdata.dta", replace
use "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Prevalence analysis\finalprevalencedata_weightsimmig10.dta", clear
drop exclE- exclG exclG immlt10y postsecd territoriesst- postsecd2_aborg lowed_lowinc-borncan_tno tno- subject male- age srAoth- prC9 immlt10yc-ageround categoriesimmig10 weight child
```

```
rename resp respnum
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\prevalencedataset.dta", replace
reshape wide srAny,  i(respnum) j(subdemo)
egen numppl=rownonmiss( srAny1 srAny2 srAny3 srAny4 srAny5 srAny6 srAny7 srAny8)
egen numallergic=rowtotal( srAny1 srAny2 srAny3 srAny4 srAny5 srAny6 srAny7 srAny8)
drop srAny1- srAny8
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\prevalencedataset.dta", replace
use "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\SPAACE extractions
2011\SPAACE extractions 2011\Refusal Questions.dta", clear
rename resp respnum
drop lang_prob1 lang_prob2 ref_2_opn ref_3_opn ref_4_opn-ltr_received
rename ref_1 refusal
rename ref_2 numadults
replace numadults=. if numadults<0
rename ref_3 numchildren
replace numchildren=. if numchildren<0
egen numpplrefusal=rowtotal(numadults numchildren)
replace numpplrefusal=. if numpplrefusal==0
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\refusalquestions.dta", replace
merge 1:1 respnum using "C:\Users\Lianne.Soller\Dropbox\PhD
Thesis\Prevalence\Analysis\Selection bias analysis\prevalencedataset.dta"
drop _merge
merge 1:1 respnum using "C:\Users\Lianne.Soller\Dropbox\PhD
Thesis\Prevalence\Analysis\Selection bias analysis\censusdata.dta"
drop _merge
drop postalcode censussubdivision CMACtcode
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\selectionbiasdataset_full.dta", replace
replace numppl=numpplrefusal if (numpplrefusal!=. & numppl==.)
gen group=.
replace group=1 if numallergic!=. & numppl!=.
drop if (disp==. & group!=1)
replace group=2 if group!=1 & refusal!=.
replace group=2 if disp==108 & refusal!=.
replace group=3 if group!=1 & group!=2
replace group=3 if disp==116
*Wait to see what Ann says before changing any of the people from one group to another
tab group
*The next lines should be done only for the prevalence analysis, but these people will be
included in the WinBUGS dataset
*drop if disp>=100 & disp<=106
```

```
*drop if disp>=109 & disp<=113
*drop if disp==120
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\selectionbiasdataset_full.dta", replace
*Now I need to replace the missing numppl and numallergic data in R
*number=sample(1:8, 15326,
prob=c(0.253,0.320,0.169,0.144,0.068,0.027,0.009,0.010),replace=T)
*write.table(number,file="C:\Users\Lianne.Soller\Dropbox\PhD
Thesis\Prevalence\Analysis\Selection bias anaylsis\numppl.xls")
*allergic=sample(1:5,23,prob=c(0.899,0.085,0.009,0.005,0.002),replace=T)
*write.table(allergic,"numallergic.csv")
drop numadults numchildren numpplrefusal
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\selectionbiasdataset_full.dta", replace
*Fill in the missing data for numppl and numallergic in excel
*Paste excel file into Stata
*Replace missing data for resp 15211s and 16954s which are missing from prev data but we
have information from demographics
replace provcode = 10 in 21077
replace provcode = 1 in 21078
replace ctname = 21 in 21077
replace ctname = 21 in 21078
*Generate a new ct variable which has continuous values
egen ctnum=group(provcode ctname)
sort ctnum
rename provcode prov
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\selectionbiasdataset_full.dta", replace
*Prepare data for winbugs
sort group numppl ctnum prov
drop respnum refusal population disp ctname group
save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias
analysis\selectionbias_winbugs_july102013.dta", replace

*changes made to reflect discussions with Ann and Lawrence Aug 27th
*add an additional status for those who were never reached
*remove ineligible households
replace group=4 if (group==3 & disp<106)
replace group=4 if (group==3 & disp>109 & disp<=113)
drop if disp==106 & group>2
drop if disp==109 & group>2
drop if disp==120 & group>2
drop if disp==101 & group>2
rename group status
```

save "C:\Users\Lianne.Soller\Dropbox\PhD Thesis\Prevalence\Analysis\Selection bias analysis\selectionbias_aug282013.dta", replace

WinBUGS programs to calculate selection bias adjusted estimates

1. The prevalence in the np group is the same as in the rq group and the prevalence in the nc group is the same as in the fp group

```
model
{
for (i in 1:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <- refusal.effect[2]  # baseline model with no rq to np difference
refusal.effect[4]<- refusal.effect[1] # non-contacts assumed to be the same as full participants
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
```

```
nc<-sum(numallergic[12763:15986])/8419
```

2. The prevalence in the np group is twice that in the rq group and the prevalence in the nc group is the same as in the fp group

```
model
{
for (i in 1:6258)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
for (i in 6259:12762)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect3[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect3[i] <- log(2*exp(refusal.effect[2])/(1-exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) ) # np=2rq
}
for (i in 12763:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect4[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect4[i] <- 0#nc=fp
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <-0  # not used anymore, just a dummy
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
```

```
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```

3. The prevalence in the np group is half that in the rq group and the prevalence in the nc group is the same as in the fp group

```
model
{
for (i in 1:6258)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
for (i in 6259:12762)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect3[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect3[i] <- log(0.5*exp(refusal.effect[2])/(1+0.5*exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) )
# np=0.5rq
}
for (i in 12763:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect4[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect4[i] <- 0#nc=fp
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <-0  # not used anymore, just a dummy
for (j in 1:265) #census tracts
```

```
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```

4. The prevalence in the np group is the same as in the rq group and the prevalence in the nc group is the same as in the np group

```
model
{
for (i in 1:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <- refusal.effect[2]  # baseline model with no rq to np difference
refusal.effect[4]<- refusal.effect[3] # non-contacts assumed to be the same as non-participants
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
```

```
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```

5. The prevalence in the np group is twice that of the rq group and the prevalence in the nc group is the same as the np group

```
model
{
for (i in 1:6258)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
for (i in 6259:12762)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect3[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect3[i] <- log(2*exp(refusal.effect[2])/(1-exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) ) # np=2rq
}
for (i in 12763:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect4[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect4[i] <- log(2*exp(refusal.effect[2])/(1-exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) ) #nc=np
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
```

```
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <-0  # not used anymore, just a dummy
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```

6. The prevalence in the np group is half that in the rq group and the prevalence in the nc group is the same as in the np group

```
model
{
for (i in 1:6258)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
for (i in 6259:12762)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect3[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect3[i] <- log(0.5*exp(refusal.effect[2])/(1+0.5*exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) )
# np=0.5rq
}
for (i in 12763:15986)
```

```
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect4[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect4[i] <-log(0.5*exp(refusal.effect[2])/(1+0.5*exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) ) #nc=np
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <-0  # not used anymore, just a dummy
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```

7. The prevalence in the np group is the same as the rq group and the prevalence in the nc group is a mixture of the prevalence in the fp, rq, and np groups

```
model
{
for (i in 1:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
```

```
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <- refusal.effect[2]  # baseline model with no rq to np difference
refusal.effect[4]<- refusal.effect[1]
refusal.effect[5]<-refusal.effect[2]
refusal.effect[6]<-refusal.effect[2]
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```

8. The prevalence in the np group is twice that in the rq group and the prevalence in the nc group is a mixture of the fp, rq and np group

```
model
{
for (i in 1:6258)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
for (i in 6259:12762)
```

```
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect3[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect3[i] <- log(2*exp(refusal.effect[2])/(1-exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) ) # np=2rq
}
for (i in 12763:14189)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect4[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect4[i] <- 0#nc=fp
}
for (i in 14190:14315)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect5[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect5[i] <- refusal.effect[2]
}
for (i in 14316:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect6[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect6[i] <- log(2*exp(refusal.effect[2])/(1-exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) )
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <-0  # not used anymore, just a dummy
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
```

```
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```

9. The prevalence in the np group is half that in the rq group and the prevalence in the nc group is a mixture of the fp, rq and np group

```
model
{
for (i in 1:6258)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect[status[i]]
numallergic[i]~dbin(p[i] , numppl[i])
}
for (i in 6259:12762)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect3[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect3[i] <- log(0.5*exp(refusal.effect[2])/(1+0.5*exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) )
# np=0.5rq
}
for (i in 12763:14189)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect4[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect4[i] <- 0#nc=fp
}
for (i in 14190:14315)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect5[i]
numallergic[i]~dbin(p[i] , numppl[i])
refusal.effect5[i] <- refusal.effect[2]
}
for (i in 14316:15986)
{
logit(p[i]) <- alpha.ct[ctnum[i]] + house.effect[numppl[i]] + refusal.effect6[i]
numallergic[i]~dbin(p[i] , numppl[i])
```

```
refusal.effect6[i] <- log(0.5*exp(refusal.effect[2])/(1+0.5*exp(alpha.ct[ctnum[i]] +
house.effect[numppl[i]] + refusal.effect[2])) )
}
house.effect[1]  <- 0
for (i in 2:8)
{
house.effect[i] ~ dnorm(0, 2)
}
refusal.effect[1] < - 0  #  no effect for participants, reference category
refusal.effect[2] ~ dnorm(0,2)  # effect for refusals
refusal.effect[3] <-0  # not used anymore, just a dummy
for (j in 1:265) #census tracts
{
alpha.ct[j]~dnorm(mean.prov[prov[j]], tau.ct)
}
for (k in 1:13) #provinces
{
mean.prov[k]~dnorm(mu.prov,tau.prov)
}
#  Priors
mu.prov ~ dnorm(0, 0.05)
sd.ct ~ dunif(0.001, 7)
sd.prov ~ dunif(0.001, 7)
tau.ct <- 1/(sd.ct*sd.ct)
tau.prov <- 1/(sd.prov*sd.prov)

p.overall<-sum(numallergic[1:15986])/41893
fp<-sum(numallergic[1:5734])/15022
rq<-sum(numallergic[5735:6258])/1393
np<-sum(numallergic[6259:12762])/17059
nc<-sum(numallergic[12763:15986])/8419
```