Health Canada Santé Canada

Canada

Early Experience analyzing Dietary Intake Data using the National Cancer Institute (NCI) method:
Canadian Community Health Survey - Nutrition

Karelyn A. Davis PhD, P.Stat.
Alejandro Gonzalez M.Sc.
Kuan Chiao Wang M.Sc.
Lidia Loukine M.Sc.
Cunye Qiao M.Sc.
Michel Vigneault M.Sc.
Alireza Sadeghpour M.Sc.
Dominique Ibanez M.Sc.

Bureau of Food Surveillance and Science Integration
Office of Nutrition Policy and Promotion
Health Canada

University of Waterloo
June 19, 2018

YOUR HEALTH AND SAFETY... OUR PRIORITY.

Background

- Health Canada statisticians working with nutrition data often collaborate with researchers and policy analysts on a variety of scientific questions
  - Our mandate is to provide in-depth analysis of nutrition data in support of research, regulatory and policy objectives

Statistics Canada

- Disseminate official statistics on a wide variety of topics, including nutrition, by providing data products and tools to users
- Provide expertise on survey design and methodology

Collaborate: Technical working groups

Health Canada

- Expertise in analysis of a wide variety of nutrition studies
- Conduct in-depth statistical analysis of nutrition data in collaboration with government research scientists, policy analysts and academics

Background

- Current Projects
  - Analyze data on food consumption and nutrient intake among the Canadian population to help establish policies and set standards and advice on the safety and nutritional value of food
  - Contribute to sampling plans and analysis evaluating the safety of novel foods (including genetically modified foods) and food additives
  - Participate in risk assessments to human health from food-borne pathogens
  - Modeling of healthy eating patterns for *Canada's Food Guide*

- Publish results in academic journals and technical reports

Background

- Data is collected and analyzed from a variety of sources including
  - Population health surveys
  - Laboratory analysis of human and animal specimens
  - Industry submissions

- We are located in two directorates:
  - Food Directorate: 10 statisticians
  - Office of Nutrition Policy and Promotion: 2 statisticians

Overall Objectives

- Describe our experience analysing dietary usual intakes related to current projects
  - Compendium of Nutrient Intakes from Food
  - Modelling for Revision of *Canada's Food Guide*

- Analysis from foods and nutrients will be presented
  - Canadian Community Health Survey (CCHS) – Nutrition both cycles 2004 and 2015

Presentation Overview

- Section 1 – CCHS- Nutrition
- Section 2 – Current Projects for analysis of Usual Dietary Intake
  - Compendium of Nutrient Intakes from Food
  - Analysis of Episodically Consumed Foods
- Section 3 – NCI Testing
  - Model Choice
  - Covariates
  - Pooling vs Strata
  - Outliers
  - Computational time
- Section 4 – Statistical Considerations and Recommendations
- References

Section 1

CANADIAN CONTEXT – CCHS – NUTRITION

2015 CCHS Nutrition Objectives

- To collect detailed data on the **consumption of foods and dietary supplements** among a representative sample of Canadians at national and provincial levels;
- To estimate the distribution of **usual dietary intake** in terms of nutrients from foods, food groups, dietary supplements and eating patterns;
- To gather **anthropometric** (physical) **measurements** for accurate body weight and height assessment to interpret dietary intake;
- To support the interpretation and analysis of dietary intake data by collecting data on selected **health conditions and socio-economic and demographic characteristics** and;
- To evaluate **changes in dietary intake** from the previous 2004 CCHS – Nutrition.

2004 CCHS vs 2015 CCHS Nutrition

Parameters	CCHS 2004	CCHS 2015
Sample	Ages 0+; 10 provinces; general households	Ages 1+; 10 provinces; general households
Sample Size	35,107 respondents with 10,927 repeats	20,487 respondents with 7,608 repeats
Survey Mode	1st interview; 60 to 65 minutes - CAPI	1st interview; 60 minutes CAPI
	2nd interview; 35 minutes CATI	2nd interview; 35 minutes CATI
Response rate	76.5% - first recall 72.8% - second recall	61.6% - first recall 68.6% - second recall

Section 2

CURRENT PROJECTS– ANALYSIS OF USUAL DIETARY INTAKE

Analysis of Usual Dietary Intake

- The NCI method is implemented in three separate collections of SAS macros:
  - 1) modeling a single dietary component, which may be episodically consumed;
  - 2) modeling two components only one of which is permitted to be episodically consumed;
  - 3) modeling many components simultaneously.

- So far, we have focused on modeling a single dietary component as used to describe usual intakes for
  - **Nutrients:** Compendium of Nutrient Intakes from Food
  - **Foods:** Episodically consumed foods to support revision of *Canada's Food Guide*

Compendium of Nutrient Intakes from Food

- An update to the Compendium using 2015 CCHS data is ongoing
  - Descriptive statistics are compiled in summary data tables (by age and sex and province/region) of usual intakes from food for 40 nutrients and other dietary components
  - For nutrients with a Dietary Reference Intakes (DRIs), the tables also provide estimates of % below or above a DRI threshold

- Previous Nutrient Compendium using data from 2004 CCHS is available on the Open Data Portal

Compendium of Nutrient Intakes from Food - 2004

Table 14.13 Folate (DFE/d): Usual intakes from food, by DRI age–sex group, household population, Canada excluding territories, 2004[1]

Sex	Age (years)	n	Mean (SE)	Percentiles (*and SE*) of usual intake							EAR[2]	% <EAR (SE)
				5th (SE)	10th (SE)	25th (SE)	50th (SE)	75th (SE)	90th (SE)	95th (SE)		
Both												
	1-3	2117	283 (6)	137 (8)	164 (8)	213 (7)	274 (7)	345 (8)	420 (12)	472 (15)	120	2.9 (0.8) [E]
	4-8	3235	396 (6)	250 (10)	277 (9)	327 (8)	390 (7)	462 (10)	537 (15)	587 (19)	160	<3
Male												
	9-13	2080	490 (10)	344 (18)	372 (16)	424 (13)	488 (11)	562 (14)	637 (21)	687 (27)	250	<3
	14-18	2288	571 (11)	328 (17)	373 (16)	459 (15)	570 (14)	697 (18)	827 (24)	915 (30)	330	5.2 (1.4) [E]
	19-30	1804	587 (14)	398 (29)	434 (25)	500 (20)	580 (17)	669 (22)	759 (34)	816 (44)	320	<3
	31-50	2596	528 (11)	325 (18)	363 (17)	433 (15)	521 (13)	622 (17)	726 (25)	793 (31)	320	[F]
	51-70	2550	485 (8)	274 (12)	311 (11)	380 (10)	471 (10)	577 (12)	690 (19)	768 (25)	320	11.5 (1.9)
	>70	1520	425 (11)	235 (11)	267 (11)	326 (11)	403 (14)	494 (19)	593 (27)	664 (35)	320	23.1 (3.2)
	19+	8470	520 (6)	304 (8)	342 (8)	414 (7)	508 (7)	615 (10)	730 (14)	807 (17)	320	6.8 (1.0)
Female												
	9-13	1980	425 (8)	276 (13)	304 (12)	356 (11)	420 (10)	493 (12)	565 (17)	613 (21)	250	[F]
	14-18	2256	445 (8)	246 (12)	282 (11)	349 (10)	434 (10)	532 (13)	635 (19)	705 (24)	330	20.1 (2.6)
	19-30	1854	415 (10)	260 (13)	287 (12)	338 (11)	401 (12)	472 (15)	542 (21)	588 (25)	320	18.8 (3.7) [E]
	31-50	2686	423 (9)	250 (12)	281 (12)	338 (11)	411 (11)	496 (13)	586 (18)	646 (23)	320	19.6 (3.3) [E]
	51-70	3200	400 (7)	236 (9)	265 (9)	320 (8)	390 (8)	472 (10)	555 (13)	610 (17)	320	25.0 (2.6)
	>70	2610	340 (6)	192 (7)	218 (6)	265 (7)	328 (8)	404 (10)	486 (13)	543 (16)	320	47.0 (2.9)
	19+	10350	405 (5)	235 (5)	265 (5)	321 (5)	393 (5)	477 (7)	566 (10)	625 (12)	320	24.6 (1.6)

Data source: Statistics Canada, Canadian Community Health Survey, Cycle 2.2, Nutrition (2004) - Share File

Symbol Legend

[E] Data with a coefficient of variation (CV) from 16.6% to 33.3%; interpret with caution.

<3 Data with a coefficient of variation (CV) greater than 33.3% with a 95% confidence interval entirely between 0 and 3%; interpret with caution.

[F] Data with a coefficient of variation (CV) greater than 33.3% with a 95% confidence interval not entirely between 0 and 3%; suppressed due to extreme sampling variability.

Footnotes

[1] Intakes are based on food consumption only. For additional detail, see footnote 4 in Appendix A.

[2] EAR is the Estimated Average Requirement. For additional detail, see footnote 9 in Appendix A.

For additional footnotes common to all tables, see Appendix A.

Compendium of Nutrient Intakes from Food 2015

- Similar nutrients as in the previous Compendium, with the addition of Caffeine and Alcohol
    - 23 nutrients have been completed to date, including:

Calcium (mg/d)	Cholesterol (mg/d)	Folacin (mcg/d)	Folate (DFE/d)
Linoleic fatty acid (g/d)	Naturally occurring folate (mcg/d)	Niacin (NE/d)	Energy (kcal/d)
Monounsaturated fat (%E)	Polyunsaturated fat (%E)	Saturated fats (g/d, %E)	Total sugars (g/d, %E)
Potassium (mg/d)	Riboflavin (mg/d)	Total fats (g/d)	Sodium (mg/d)
Thiamin (mg/d)	Total dietary fibre (g/d)	Vitamin B12 (µg/d)	Vitamin C (mg/d)
Vitamin D (µg/d)			

- Results from 2015 CCHS will also be published on the Open Data Portal in late 2018/early 2019

Compendium of Nutrient Intakes from Food 2015

- Comparing results from 2004 to 2015
  - Usual intakes re-run for 2004 using the NCI method to allow comparisons
    - Compendium will highlight statistically significant differences in % above or below DRI threshold
  - Caution when comparing estimates from 2004 and 2015
    - Changes to food booklet for some models (e.g. bowl, mug, cup, wine glass size)
    - Slightly different Canadian Nutrient Files
    - Upcoming Webinar to discuss comparing results
- Usual intakes for most nutrients use the NCI one-part (amount-only) model
    - Probability of inadequacy for Iron is calculated using the full probability method
    - **Exceptions**: Alcohol and Caffeine for which the two-part model is used

Analysis of Episodically Consumed Foods

- As part of the revision to Canada's Food Guide, a modeling approach will be used to determine the amounts and types of foods in a healthy eating pattern
  - Two step iterative approach:
    - STEP 1: uses food composites in mathematical optimization
    - STEP 2: uses individual foods in simulation of individual diets
  - Mathematical optimization requires an objective function and a set of **nutrient** (DRIs) and **food-based contraints** (convincing evidence between food and health)
  - Usual intakes will be used to help inform food based constraints used in the mathematical optimization to ensure healthy eating pattern includes realistic amounts of foods in relation to Canadian eating patterns

Analysis of Episodically Consumed Foods

- To support the CFG modelling, analysis of usual intakes of certain episodically consumed foods is underway by DRI age/sex group:
  - Red meat
  - Fish and shellfish
  - Legumes
  - Whole grains
  - Other foods to be considered

- The two-part NCI model is used to compute usual intakes for episodically consumed foods

Section 3

CANADIAN CONTEXT – NCI TESTING

Statistical Considerations –NCI Testing Questions

1. Given the NCI method uses different models for nutrients and foods, which model should be implemented?

2. Which covariates should be included in the model?

3. Should groups be pooled to provide better estimates?

4. How should outliers be identified?

5. What is the computational time to obtain usual dietary intake for nutrients and foods using the NCI method?

Statistical Considerations –NCI Testing

- Analysis of usual intakes:
  - Descriptive statistics including mean, median, percentiles, % above/below DRI threshold (for nutrients)
  - Bootstrap method used to estimate standard errors

- **Datasets:**
  - 2004 CCHS (prior to 2015 CCHS release):
    - Nutrients: Calcium from Food, Dietary Folate Equivalents
    - Foods: Orange vegetables, Green vegetables
  - 2015 CCHS
    - Nutrients: Compendium nutrients
    - Episodically consumed foods: Red meat

Statistical Considerations – NCI Testing

- Age/Sex groupings from the Institute of Medicine's Dietary Reference Intakes:
  - 1-3 years (both genders combined)
  - 4-8 years (both genders combined)
- For males and females separately:
  - 9-13 years
  - 14-18 years
  - 19-30 years
  - 31-50 years
  - 51-70 years
  - 71 years and older

Statistical Considerations – NCI Testing

- For nutrients and foods, the model estimates the amount consumed using a non-linear mixed model.
  - Choice between a 1- or 2-part model depends on whether or not the nutrient/food was consumed daily by almost everyone

- Episodically consumed foods: a 2-part model is used
  - Along with the non-linear mixed model we must also estimate the probability of consuming the food through a logistic regression model
- Most nutrients and ubiquitously consumed foods: the 1-part model is used
  - The probability part of the model is not needed

1. Which model to use?

- Adapted from Krebs-Smith et. al. (2010)[2]
  - if < 5% of the 24 hr recalls (unweighted) had zero intake of food, then the 1-part model was used
  - if >10% of the 24 hr recalls had zero intake of a food, the two-part model was fit twice
    - Correlated and uncorrelated model
  - if between 5% and 10% of the 24 hr recalls had zero intake of a food, then all three models were fit
    - Best-fitting model was chosen by examining the significance of the correlation coefficient through Fisher's z-transformation.

Frequencies of 24hr Recalls

DRI Age/Sex Group	Red meats (CCHS 2004 and 2015)	
	Unweighted	
	No. of consumers (Day 1/Day 2)	% with zero intake (Day 1/Day 2)
Children (Male & Female)		
2 - 3y	1023/419	**57.4 / 55.4**
4 - 8y	1996/540	**55.0 / 55.6**
Female		
9 - 13y	1336/553	**54.2 / 54.4**
14 - 18y	1387/521	**57.5 / 57.7**
19 - 30y	1309/406	**56.5 / 60.4**
31 - 50y	2254/598	**56.1 / 55.9**
51 - 70y	2451/699	**55.6 / 58.2**
71+y	1719/655	**57.6 / 55.5**
Male		
9 - 13y	1473/608	**52.2 / 52.7**
14 - 18y	1722/615	**46.0 / 48.0**
19 - 30y	1455/442	**45.3 / 42.7**
31 - 50y	2511/621	**45.4 / 45.0**
51 - 70y	2423/678	**48.4 / 45.0**
71+	1264/538	**53.2 / 51.4**

Source: 2004 Canadian Community Health Survey – Nutrition (cycle 2.2); CCHS 2015 Canadian Community Health Survey - Nutrition

1. Which model to use?

- **Compendium Nutrients**
  - Vast majority used the amount-only model
    - % with zero intake <1% in most cases
  - Exceptions: Alcohol and Caffeine
    - % with zero intake varies in different DRI age/sex groups
    - **Caffeine:** 19% in Males 19 – 30 to 6% in Males 71+ among first recalls
- **Episodically Consumed Foods**
  - Most foods analyzed to date used two-part model
    - % with zero intake > 10% in most cases
  - Analyses are still ongoing…

2. Which covariates should be included?

- NCI permits incorporation of important covariates to improve model fit and obtain usual intake estimates

  - For 2004 CCHS episodically consumed foods, both parts of the correlated model (and uncorrelated model, where applicable) were adjusted by weekend/weekday and by frequency of fruit and vegetable consumption.
    - Food frequency variables not available for 2015 CCHS

  - Compendium nutrients included cycle, weekday/weekend, order of recall and province within each DRI age/sex stratum
    - Vitamin C also included smoking status as a covariate
  - Episodically consumed foods included cycle, weekday/weekend, gender where appropriate

3. Should groups be pooled to provide better estimates?

- Related to covariates, the NCI macros offer the ability to pool different groups (e.g. age/sex groups, province) to increase sample size
  - Alternative is to compute analysis by separate age/sex groups (**Strata**)
  - Pooling assumes the variability is equal within each level of a covariate
    - Variance components are estimated under this assumption

- **Advantages:**
  - Pooling often reduces standard errors
  - For the 2-part model, pooling is sometimes the only model which converged

Pooling vs. Strata

- **Disadvantages:**
  - If the equal variance assumption is not true, the adjustment from one-day to usual intake distribution may be over- or under-estimated
    - Estimates at the distribution tails (e.g. 5th, 10th, 90th, 95th percentiles) may be different while mean and median are relatively unaffected
    - Examine the ratio of the within- to between- variance components to determine which groups may be most affected

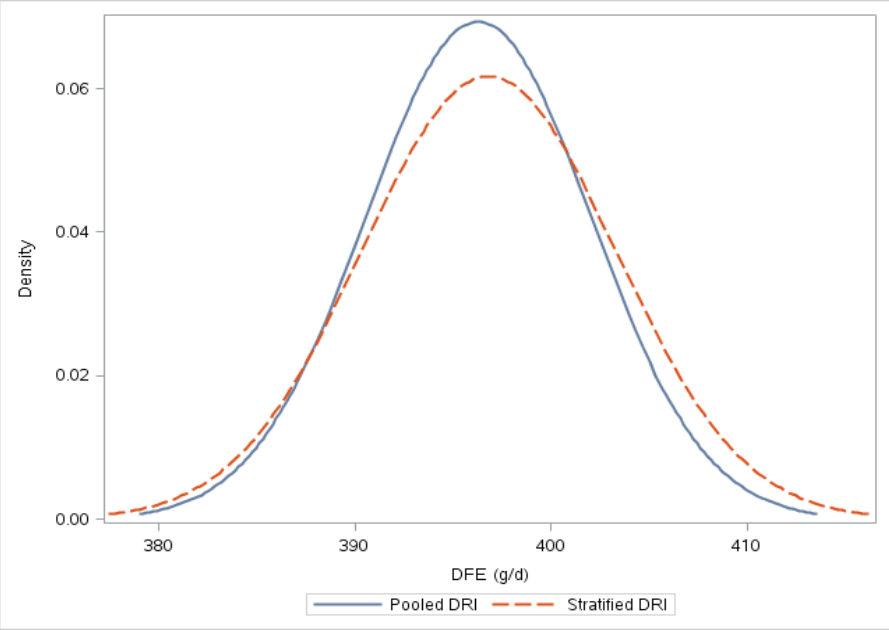

Figure 1: Usual intake distribution of Folate (DFE/d) for children 4 – 8 years

Source: 2004 Canadian Community Health Survey – Nutrition (cycle 2.2)

Pooling vs. Strata: Folate (DFE/d)

DRI Group	Method	Variance components: Between	Variance components: Within	R	Ratio of Within/ Between	n	Mean	Percentiles: 5th	Percentiles: 50th	Percentiles: 90th	Percentiles: 95th	% < EAR
Children 1 - 3	NCI - Pooled	4.16	16.59	0.80	3.99	2093	286.36	181.93	280.81	378.79	410.17	<3
	NCI - Stratified	23.27	36.29	0.61	1.56	2093	285.75	144.12	276.53	414.36	459.39	F
						% Diff	0.21	20.78	1.52	9.39	12.00	F
Children 4 - 8	NCI - Pooled	4.16	16.59	0.80	3.99	3235	396.61	265.72[E]	390.78	510.05	547.97	<3
	NCI - Stratified	0.66	6.08	0.90	9.15	3235	397.44	303.72[E]	393.99	477.54	503.38	<3
						% Diff	0.21	14.30	0.82	6.37	8.14	<3

Source: 2004 Canadian Community Health Survey – Nutrition (cycle 2.2)

Pooling vs. Strata: Folate (DFE/d)

		Variance components						Percentiles				
DRI Group	Method	Between	Within	R	Ratio of Within/ Between	n	Mean	5th	50th	90th	95th	% < EAR
Females 9 - 13	NCI - Pooled	2.87	7.69	0.73	2.68	1980	426.70	257.03	415.59	579.65	633.67	4.2[E]
	NCI - Stratified	2.92	10.89	0.79	3.72	1980	426.03	281.57	418.63	553.41	596.51	F
						% Diff	0.16	9.55	0.73	4.53	5.86	F
Females 14 - 18	NCI - Pooled	2.87	7.69	0.73	2.68	2254	439.17	266.27	427.86	595.51	649.82	17.6
	NCI - Stratified	7.23	18.52	0.72	2.56	2254	446.23	259.06	434.26	614.95	673.89	18.1[E]
						% Diff	1.61	2.71	1.50	3.26	3.70	-0.5
Females 19 - 30	NCI - Pooled	2.87	7.69	0.73	2.68	1853	410.92	246.04	399.91	560.32	612.90	21.5
	NCI - Stratified	6.28	17.07	0.73	2.72	1853	413.27	238.51	402.20	571.38	627.14	22.7[E]
						% Diff	0.57	3.06	0.57	1.97	2.32	-1.2
Females 31 - 50	NCI - Pooled	2.87	7.69	0.73	2.68	2686	421.06	253.15	409.84	573.12	627.60	19.1
	NCI - Stratified	1.39	6.11	0.81	4.38	2686	423.24	278.62	415.14	551.00	595.68	F
						% Diff	0.52	10.06	1.29	3.86	5.09	F
Females 51 - 70	NCI - Pooled	2.87	7.69	0.73	2.68	3200	401.68	240.33	390.81	548.68	600.94	24.2
	NCI - Stratified	6.13	9.83	0.62	1.60	3200	399.51	218.88	386.51	565.91	625.90	28.2
						% Diff	0.54	8.93	1.10	3.14	4.15	-4
Females 71+	NCI - Pooled	2.87	7.69	0.73	2.68	2609	349.45	204.21	339.17	482.33	530.33	42.1
	NCI - Stratified	1.88	3.43	0.65	1.83	2609	338.77	197.80	328.21	468.25	515.72	46.5
						% Diff	3.06	3.14	3.23	2.92	2.75	-4.4

Source: 2004 Canadian Community Health Survey – Nutrition (cycle 2.2)

Pooling vs. Strata

- Similar results found for episodically consumed foods
  - Stratified analysis often not possible for many DRI age/sex groups due to non-convergence (e.g. Red meat)

- **Recommendation:**
  - For nutrients and foods, pool data from 2004 and 2015 CCHS to increase sample size and enable variance components to be estimated accurately
  - Verify assumption that within-between variation does not vary significantly between cycles

Pooling vs. Strata: Standard Errors: DFE – Female

		Pooled				Stratified			
Age Group	No. of Bootstrap replicates	Mean	5th percentile	95th percentile	% < EAR	Mean	5th percentile	95th percentile	% < EAR
Females 9 - 13	200	7.95	10.72	17.28	1.21	8.26	19.47	28.23	1.61
	300	7.77	10.44	16.6	1.19	8.22	20.12	28.29	1.66
	400	7.84	10.25	16.87	1.17	8.11	19.89	28.48	1.62
	500	7.81	10.28	16.91	1.16	8.03	19.55	27.69	1.58
Females 14 - 18	200	7.89	10.88	17.33	2.52	8.22	18.82	31.63	3.57
	300	7.74	10.22	17.44	2.37	7.99	18.92	31.36	3.58
	400	7.68	10.15	17.33	2.35	8.01	19.29	31.12	3.66
	500	7.83	10.11	17.7	2.34	8.07	19.54	32.15	3.69
Females 19 - 30	200	8.65	10.08	18.73	2.76	9.79	22.76	41.51	4.64
	300	8.73	9.86	18.49	2.73	9.96	22.04	40.82	4.50
	400	8.6	9.66	18.43	2.64	9.87	21.22	39.25	4.33
	500	8.66	9.59	18.67	2.62	9.94	21.33	39.76	4.36

Source: 2004 Canadian Community Health Survey – Nutrition (cycle 2.2)

4. How should outliers be determined?

- Outliers affect model estimates and may lead to non-convergence or nonsensical estimates

- Two methods for outlier detection were considered:

Method A – Within/Between Variance:

  - When ratio of within/between variation is greater than 10, consider the mean distribution of the difference between Day 1 and Day 2 recall.
  - Values were identified as possible outliers if they fell ±3, ±2.5 or ±2 SD away from the mean distribution of difference between Day 1 and Day 2 values
  - Day 2 recalls were removed as Day 1 recalls are considered to be less biased
  - The scenario which first resulted in the within-between variance ratio less than 10 and excluded the fewest second 24hr recalls was retained

Outlier Detection

- **Method B: Normality Violation**
  - Second method considered raw values for all intake variables which may influence the data
  - Box-Cox transformation of the raw non-zero values to approximate normality
  - Extreme values identified as points either
    - below the 25th percentile minus 2.5*IQR of the transformed distribution
    - above the 75th percentile plus 2.5*IQR of the transformed distribution
  - Analysis with and without extreme values conducted and compared

- Based on method suggested by Krebs-Smith et. al. (2010)[2]

Outlier Detection – Compendium 2015

Nutrient	DRI Age-Sex Group	Threshold	No. of recalls removed
%total energy intake from fats	19 to 30 years, females	3 SD	4
% total energy intake from monounsaturated fats	19 to 30 years, females	3 SD	8
Sodium (mg/d)	19 to 30 years, males	2 SD	39
Potassium (mg/d)	31 to 50 years, males	2 SD	63
% total energy intake from linolenic fatty acid	9 to 13 years, females	3 SD	11
% total energy intake from linolenic fatty acid	14 to 18 years, males	3 SD	18

Source: 2004 Canadian Community Health Survey – Nutrition (cycle 2.2); 2015 CCHS Canadian Community Health Survey - Nutrition

5. What is the computational time for the NCI method?

Approximate Time to convergence (Hours)			
Females	**Total sugar**	**Red meats**	
	Stratified+Pooled	Stratified	Pooled
14-18	7.19	**2.87**	**12.07**
19-30	6.26	**8.28**	
31-50	9.34		
51-70	10.58	**5.08**	
71+	8.81	**4.12**	
All DRI Age/Sex Groups	**102.07 (4.25 days)**	**48.2 (2.01 days)**	**66.06 (2.75 days)**

Source: 2004 Canadian Community Health Survey – Nutrition (cycle 2.2); 2015 CCHS Canadian Community Health Survey - Nutrition

- **Total sugar**
  - Time for national, provincial and regional results presented
- **Red meats**
  - National results only with three strata: Children < 14 years; Females 14+, Males 14+

Section 4

CANADIAN CONTEXT – STATISTICAL CONSIDERATIONS

Summary

- The use of the NCI macros requires statistical considerations in addition to SAS programming
  - Macro options incorporate different statistical assumptions in fitting the measurement error model
  - Care must be taken in evaluating outliers, equal variance assumption and inclusion of covariates

- Recommend using MIXTRAN with root survey weight to evaluate statistical considerations

- Once final model is found, DISTRIB macro can be used to determine estimates (mean, percentiles, % above or below a DRI threshold)

- VBRR macro may be used to compute bootstrap standard errors
  - Note that for 2004 and 2015 CCHS, ensure option `f=1` (no Fay Adjustment)

Summary – Statistical Considerations

- **Choice of 1- part or 2-part model**
  - Nutrient or Food? Or better Episodically or Ubiquitously consumed?
  - Obtain frequencies of 24hr recalls to evaluate number of consumers and percentage with zero intake
  - Test for correlation using Fisher's z-transformation
    - Consider 50 to 100 bootstrap replicates for hypothesis testing

- **Covariates of interest**
  - Choose important covariates for analysis and evaluate coefficients using t-tests
  - For Compendium 2015 nutrients, covariates include:
    - Cycle
    - Province
    - Weekend/Weekday
    - Smoking status (Vitamin C)

Summary – Statistical Considerations

- **Pooling versus strata**
  - Using MIXTRAN, fit pooled and stratified analysis and evaluate ratio of within-between variance components
  - If one or more groups have much larger ratio (30% difference or more) consider fitting stratified analysis, if convergent
  - Evaluate coefficient of variation, particularly for foods, to determine reliability
  - NOTE: For most analyses completed at Health Canada, both cycles were pooled in order to ensure reliable analysis by DRI age/sex group
    - Ensure survey weights and bootstrap weights are rounded prior to running NCI macros

Summary – Statistical Considerations

- **Pooling versus strata**
  - **Compendium:** Nutrients stratified by DRI age-sex group; cycles and provinces pooled
    - Within-between variation not similar by DRI age-sex group, but similar among province and cycle
  - **Episodically consumed foods:** Cycles pooled; Three strata considered for variance component estimation at the national level:
    - Children < 19
    - Females 19 and older
    - Males 19 and older

Summary – Statistical Considerations

- **Outliers**
  - Consider two approaches to identify outliers
    A. For groups with within-between variance ratio above 10, remove recalls which are within 3, 2.5 or 2 standard deviations from the mean and re-fit
    B. Consider departures from normality by analyzing residuals once Box-Cox transformation is fit on raw nonzero data
  - Caution should be used in removal of outliers
    - Removal of a small amount of extreme values which lead to violation of model assumptions and subsequent change in results

Analysis Flowchart

- Flow-Chart for Univariate Analysis of Nutrients and Foods

REFERENCES

References – Websites

- *Health Canada Reference Guide to Understanding and Using the Data*

- *Statistics Canada CCHS-2015 website: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5049*

- NHANES Dietary Web Tutorial on the CDC's website: https://www.cdc.gov/nchs/tutorials/dietary/advanced/index.htm

References – Journal Articles

1. Katamay, S.W., Esslinger, KA., Vigneault, M., Johnston, J.L., Junkins, E.A., Robbins, L.G., Sirois, I.V., Jones-McLean, E.M., Kennedy, A.F., Bush, M.A.A., Brulé, D., Martineau, C. Eating Well with Canada's Food Guide (2007): Development of the Food Intake Pattern. *Nutrition Reviews* 2007. **65 (4)**: 155-166.

2. Krebs-Smith SM, Guenther PM, Subar AM, Kirkpatrick SI, Dodd KW. Americans do not meet federal dietary recommendations. *The Journal of Nutrition* 2010; 140, 1832-1838.

3. Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ, Kipnis V. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *Journal of the American Dietetic Association* 2006; 1575-1587.

4. Tooze JA, Kipnis V, Buckman DW, Carroll RJ, Freedman LS, Guenther PM, Krebs-Smith SM, Subar AF, Dodd KW. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: The NCI method. *Statistics in Medicine* 2010; 29: 2857-2868.

Thank you / Merci

karelyn.davis@canada.ca

cunye.qiao@canada.ca