

Canadian Community Health Survey (CCHS) - Annual Component

April 30, 2024



Statistics
Canada Statistique
Canada

Canada



Overview

- Survey Overview
- PUMF overview (from confidential to public)
- Considerations for analysis
- SAS estimation and variance measures

1 – Public Use Microdata File





Survey Overview - Objectives

- Health **surveillance** at the national, provincial and sub-provincial levels
- **Analysis** on small populations and rare characteristics
- **Readily available data** to a diversified user community in a timely fashion
- **Flexible** survey instrument



Survey Overview: Design

- Residents of private households aged 12 and over
- Exclusions:
 - Aboriginal reserves / settlements
 - institutions (health institutions, prisons, religious institutions, convents, etc.)
 - members of Canadian Forces
 - Residents of certain very remote areas
- Coverage (general) ~97% of targeted population
- Designed to provide reliable estimates at the health region level with two years of data

Survey overview: Health Regions

- Administration of health in the provinces is divided into health regions.
 - PEI and territories each are designated as their own single HR
- This is the lowest level of geography on the PUMF, after some grouping*

Survey overview: Collection

- The CCHS is collected by interviewers using a computer based questionnaire:
 - 20% in person, and
 - 80% by telephone*
- Respondents aged 12-15 interviewed alone with parental consent
- 2.5% of cases completed by proxy
- Beginning in 2022, CCHS uses online self-report questionnaire

Survey overview - content

- Three main types of content in survey
 - Core: all years, all respondents
 - Theme: certain years, all respondents
 - Optional: certain years, certain provinces/territories
- Other content not available through DLI:
 - Buy-ins
 - Rapid responses
 - Sub-samples

Content design

Content Type	2015	2016	2017	2018	2019	2020
Core	C	C	C	C	C	C
Theme	T	T			T	T
Optional	Selected by PTs per their data needs (bank of over 80 modules)					
Rapid response	RR1 (Jan-Jun)		RR2 (Jul-Dec)	RR3 (Jan-Mar)		

Content plan, CCHS 2015 and 2016, modified version

	2015	2016
Core Content	Proxy interview (GR) - Age of respondent (ANC1) - Main activity (MAC) Socio-demographic characteristics (SDC) - Labour force (LBF) - Income (INC) Administration information (ADM) - Person most knowledgeable about household situation (PMK) Height and weight - self-reported (HWT) - Fruit and vegetable consumption (FVC) General health (GEN) - Chronic conditions (CCC) - Flu shots (FLU) Smoking (SMK) - Exposure to second-hand smoke (ETS) - Alcohol use (ALC) - Primary health care - (PHC) - Physical activities - adults 18 years and older (PAA) Physical activities for youth (PAY) - Maternal experiences (MEX)	
2-year themes	Alcohol use during the past week (ALW), Drug use methods (DRM), Sexual behaviours (SXB)	
	Health insurance coverage (INS), Patient experiences (PEX), Home care services (HMC)	
	Consultations about mental health (CMH), Suicidal thoughts and attempts (SUI)	
	Chronic conditions - theme 1 (CCC): joint pain, sleep apnea, scoliosis, back pain, osteoporosis, fibromyalgia, migraines, chronic fatigue syndrome, multiple chemical sensitivities	
1-year themes (only in provinces)	Positive mental health (PMH) Functional health (HUI)	Social provisions (SPS) Sources of stress (STS) Canada's food guide use (FGU)
Optional content	Varies by province or territory	

Content plan, CCHS 2017 and 2018, modified version

	PROVINCES		TERRITORIES	
	2017	2018	2017	2018
Core Content	Proxy interview (GR) - Age of respondent (ANC1) - Main activity (MAC) Socio-demographic characteristics (SDC) - Labour force (LBF) - Income (INC) Administration information (ADM) - Person most knowledgeable about household situation (PMK) Height and weight - self-reported (HWT) General health (GEN) - Chronic conditions (CCC) - Flu shots (FLU) Smoking (SMK) - Alcohol use (ALC) - Primary health care - short (PHC) Physical activities - adults 18 years and older (PAA) Physical activities for youth (PAY) Maternal experiences (MEX)		Proxy interview (GR) - Age of respondent (ANC1) - Main activity (MAC) Socio-demographic characteristics (SDC) - Labour force (LBF) - Income (INC) Administration information (ADM) - Person most knowledgeable about household situation (PMK) Height and weight - self-reported (HWT) General health (GEN) - Chronic conditions (CCC) - Flu shots (FLU) Smoking (SMK) - Alcohol use (ALC) - Primary health care (PHC) Physical activities - adults 18 years and older (PAA) Physical activities for youth (PAY) Fruit and vegetable consumption (FVC) Maternal experiences (MEX)	
2-year themes	Sedentary behaviours (SBE), Alcohol use during maternal experience (MXA), Smoking during maternal experience (MXS)		Sedentary behaviours (SBE), Use of protective equipment (UPE), Alcohol use during maternal experience (MXA), Smoking during maternal experience (MXS)	
	Chronic conditions - short (CCC, Theme 2), Injuries (INJ)		Chronic conditions (CCC, Theme 2), Injuries (INJ)	
	Washington Group - Disability measure (WDM)		Washington Group - Disability measure (WDM)	
	Food security (FSC)		Food security (FSC)	
1-year themes	Pap smear test (PAP) Mammography (MAM) Colorectal cancer testing (CCT) Fruit and vegetable consumption (FVC) Use of protective equipment (UPE)	Dental care (DEN) Stress and relationships - STR Pain reliever medication (PRM) Perceived need for care (PNC) Home care services - modified (HMC) Prescriptions - cost-related non-adherence - short (PCN2)	Pap smear test (PAP) Mammography (MAM) Colorectal cancer testing (CCT)	

	PROVINCES		TERRITORIES	
	2019	2020	2019	2020
Core Content	Proxy interview (GR) - Age of respondent (ANC1) - Main activity (MAC) - Socio-demographic characteristics (SDC) - Labour force (LBF) - Income (INC - THI) - Administration information (ADM - HN) - Person most knowledgeable about household situation (PMK) - Height and weight - self-reported (HWT) General health (GEN) - Flu shots (FLU) Smoking (SMK) - Alcohol use (ALC) - Primary health care (PHC) - Cannabis use, (CAN) - Chronic conditions (CCC)		Proxy interview (GR) - Age of respondent (ANC1) - Main activity (MAC) - Socio-demographic characteristics (SDC) - Labour force (LBF) - Income (INC - THI) - Administration information (ADM - HN) - Person most knowledgeable about household situation (PMK) - Height and weight - self-reported (HWT) General health (GEN) - Flu shots (FLU) Smoking (SMK) - Alcohol use (ALC) - Primary health care (PHC) - Cannabis use (CAN) - Chronic conditions (CCC)	
Core selected	(CP3) Contacts with dental professionals (2 Qs from CHP)	Consommation de fruits et légumes (FVC), Physical activities - adults 18 years and older (PAA) - Physical activities for youth (PAY)	(CP3) Contacts with dental professionals (2 Qs from CHP)	
2-Year Theme Content	Services de soins de santé à domicile- avec soins palliatifs (HMC) - Indice de l'état de santé (HUI)- L'échelle de gravité de dépendance au cannabis (SDS) - Suicidal thoughts and attempts (SUI)- Besoins perçus de soins (PNC)		Services de soins de santé à domicile- avec soins palliatifs (HMC) - Indice de l'état de santé (HUI)- L'échelle de gravité de dépendance au cannabis (SDS) - Suicidal thoughts and attempts (SUI) - Besoins perçus de soins (PNC)	
1-Year Theme Content	Childhood experiences (CEX) - Consultations about mental health (CMH) - Couverture d'assurance de santé (INS) - Patient experiences (PEX) - Social provisions (SPS) - Sources of stress (STS)	Covid 19 (COV) - Canada's Food Guide use (FGU) - Loss of productivity (LOP), Sexual behaviours (SXB).	Childhood experiences (CEX) - Consultations about mental health (CMH) - Couverture d'assurance de santé (INS) - Patient experiences (PEX) - Social provisions (SPS) - Sources of stress (STS)	



Survey overview: Data processing

- During the interview, questions get skipped for various reasons and need coding:
 - Non-response
 - Interview incomplete
 - Question not relevant

These cases are coded as “Not stated”

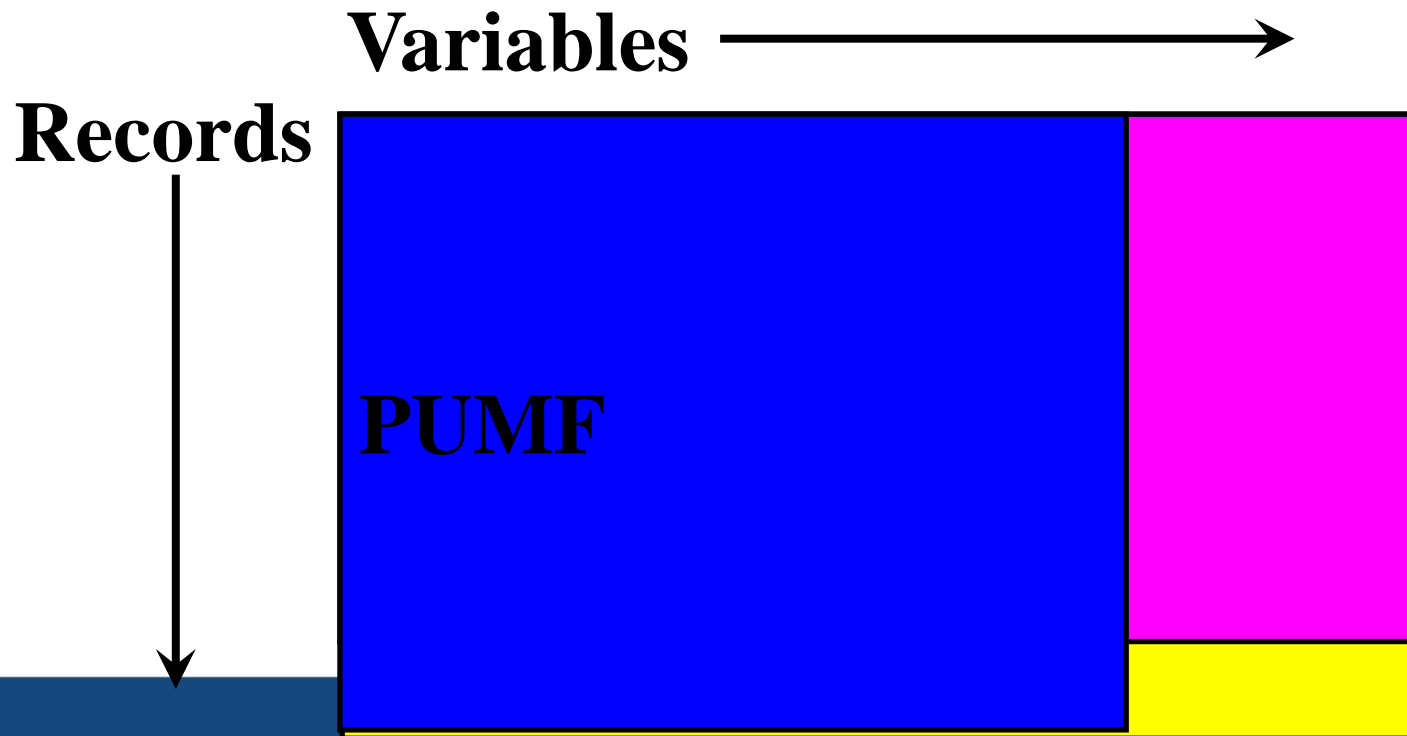
→ Coded as “Not applicable”
- “Other – specify” responses are sometimes recoded to a response category
- Derived variables are added to the file based on responses

Survey overview: Weighting

- Each respondent represents themselves and others in the population
- The sample weight WTS_M indicates the population represented by each respondent
- Weights have been adjusted for non-response and calibrated to match population estimates: $HR * sex * age \text{ group}$

From Master to PUMF

Master, Share, PUMF



Contents of the CCHS-Annual PUMF



100

Reference year	Records	Variables	Geography
2015-2016	109,659	1,199	Province/Territory HR (grouped) Canale
2017-2018	113,735	1,054	Province/Territory HR (grouped)
2019-2020	108,252	691	Province/Territory (grouped) HR (grouped)
*CHSS 2019-2020	41,550	729	Province



CCHS PUMF – disclosure control



- Geography / HR population size
 - Minimum population size of 70,000 per health region
 - 19 HRs grouped into 9 (28 -> 13 for 2019-20)
 - 3 Territories grouped as of 2019-20
- Age of respondent
 - Early PUMFs 12-14, 15-17, 18-19, and five year age groups 20-24 -> 75-79
 - From 2019-20, 12-17, 18-34, 35-49, 50-64, 65+
- Other indirect identifiers grouped or masked to prevent disclosure



CCHS 15-16+ PUMF: Poisson bootstrap weights

- Previous PUMFs did not include bootstrap weights
- 2015-16+ PUMF includes Poisson bootstrap weights
 - Each weight for each replicate is generated independently
 - All bootstrap weights are strictly positive
- The Poisson bootstrap does not pose confidentiality concerns

PUMF Package

- Data
- Bootstrap weights
- Documentation
 - User guides
 - CV tables
 - Data Dictionary
 - Derived Variables
 - Errata (if needed)
 - Questionnaire

Variable Name: SMKG110 **Length:** 1.0 **Position:** 346
Question Name:
Concept: Quit smoking completely - num of years (former daily smoker)
Question Text: How many years ago was it?
Universe: Respondents who answered SMK_100 = 4
Note: This variable has been grouped as a form of disclosure control.

Source:

<u>Answer Categories</u>	<u>Code</u>	<u>Frequency</u>	<u>Weighted Frequency</u>	<u>%</u>	
Stopped smoking between 3 and 5 years ago.	1	102	30,149	0.1	
Stopped smoking between 6 and 10 years ago.	2	134	35,122	0.1	
Stopped smoking 11 or more years ago.	3	316	63,424	0.2	
Valid skip	6	108,746	30,359,278	99.2	
Not stated	9	361	102,806	0.3	
	Total		109,659	30,590,780	100.0



Estimates and variance



Computing point estimates

- Using survey weights, you can calculate many point estimates:
 - Totals
 - Proportions
 - Averages
 - Percentiles
 - Differences in proportions
- Weighted estimates will relate to the population

What are missing data?

- Two types of non-response:
 - Total non-response (Unit nonresponse)
 - No data collected
 - Could be due to: complete refusal, inability to contact the respondent, respondent absent for the duration of the survey (no proxy allowed or available), language barrier
 - Partial non-response (Item nonresponse)
 - At least some variables are collected
 - Could be due to: refusal to answer some questions, respondent doesn't know some answers, partially completed interview

Types of missing data in Statistics Canada surveys

- There are no “holes” in Statistics Canada data files. A specific value identifies the type of missing data.

Type of missing data	Value
Not applicable / Valid Skip	6, 96, 996, ...
Don't know	7, 97, 997, ...
Refusal	8, 98, 998, ...
Not stated	9, 99, 999, ...

INJ_Q005

In the past 12 months, did ^YOU2 have any injuries due to repetitive strain?

1 Yes

2 No

DK, RF

(Go to INJ_D030A)

(Go to INJ_D030A)

INJ_Q010

Were these injuries serious enough to limit ^YOUR1 normal activities?

1 Yes

2 No

DK, RF

(Go to INJ_D030A)

(Go to INJ_D030A)

Variable Name: INJ_005 Length: 1.0 Position: 150

Question Name:

Concept: Repetitive strain injuries - 12 mo

Question Text: In the past 12 months, did you have any injuries due to repetitive strain?

Universe: Respondents with DOINJ = 1

Note:

Source:

<u>Answer Categories</u>	<u>Code</u>	<u>Frequency</u>	<u>Weighted Frequency</u>	<u>%</u>
Yes	1	13,014	3,676,912	11.8
No	2	100,032	27,518,972	88.0
Don't know	7	236	75,164	0.2
Refusal	8	8	3,324	0.0
Total		113,290	31,274,372	100.0

Variable Name: INJ_010 Length: 1.0 Position: 151

Question Name:

Concept: Repetitive strain injuries - limiting normal activities

Question Text: Were these injuries serious enough to limit your normal activities?

Universe: Respondents who answered INJ_005 = 1

Note:

Source:

<u>Answer Categories</u>	<u>Code</u>	<u>Frequency</u>	<u>Weighted Frequency</u>	<u>%</u>
Yes	1	7,370	2,045,223	6.5
No	2	5,629	1,626,612	5.2
Valid skip	6	100,032	27,518,972	88.0
Don't know	7	15	5,077	0.0
Not stated	9	244	78,488	0.3
Total		113,290	31,274,372	100.0

How to handle missing data?

- Example:

Respondent	Smoking
1	Yes
2	No
4	No
5	No

Conclusions about the population:

1) Keep the missing values:

20% Smokers

60% Non-smokers

20% Don't know or refuse to answer

2) Ignore the missing values:

25% Smokers

75% Non-smokers

Example

Daily fruit & vegetable (FVCDGTOT) consumption in Nova Scotia

FVCDGTOT Before Recoding	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1 (<5 times)	482,679	60.27	482,679	60.27
2 (5-10 times)	250,105	31.23	732,784	91.50
3 (>10 times)	29,147	3.64	761,931	95.14
9 (not stated)	38,890	4.86	800,821	100.00

FVCDGTOT After Recoding	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1 (<5 times)	482,679	63.35	482,679	63.35
2 (5-10 times)	250,105	32.83	732,784	96.17
3 (>10 times)	29,147	3.83	761,931	100.00

Estimation of the Precision

- Measures of precision include:
 - Variance, Standard error
 - Coefficient of variation (CV)
 - Margin of error
 - Confidence Interval (CI)
- Precision is a function of:
 - Sample and population size
 - Sampling design used (design effect)
 - Magnitude of the proportion estimated

CV Lookup tables

FVCDGTOT After Recoding	Frequency	Percent
1 (<5 times)	482,679	63.35
2 (5-10 times)	250,105	32.83
3 (>10 times)	29,147	3.83

Nova Scotia

ESTIMATED PERCENTAGE

NUMERATOR
OF
PERCENTAGE
('000)

	0.1%	1.0%	2.0%	5.0%	10.0%	15.0%	20.0%	25.0%	30.0%	35.0%	40.0%	50.0%	70.0%	90.0%
1	56.2	55.9	55.6	54.8	53.3	51.8	50.3	48.7	47.0	45.3	43.5	39.7	30.8	17.8
2	39.7	39.6	39.3	38.7	37.7	36.6	35.6	34.4	33.3	32.0	30.8	28.1	21.8	12.6
3	32.4	32.3	32.1	31.6	30.8	29.9	29.0	28.1	27.2	26.2	25.1	22.9	17.8	10.3
4	28.1	28.0	27.8	27.4	26.7	25.9	25.1	24.3	23.5	22.7	21.8	19.9	15.4	8.9
5	25.1	25.0	24.9	24.5	23.8	23.2	22.5	21.8	21.0	20.3	19.5	17.8	13.8	7.9
↓														
200	*****	*****	*****	*****	*****	*****	3.6	3.4	3.3	3.2	3.1	2.8	2.2	1.3
250	*****	*****	*****	*****	*****	*****	*****	3.0	2.9	2.8	2.5	1.9	1.1	
300	*****	*****	*****	*****	*****	*****	*****	*****	2.6	2.5	2.3	1.8	1.0	
350	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.3	2.1	1.6	1.0	
400	*****	*****	*****	*****	*****	*****	*****	*****	*****	2.2	2.0	1.5	0.9	
450	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.9	1.5	0.8	
500	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	1.8	1.4	0.8	

SAS coding example

```
5 data PUMF;
6 merge cchs.hs cchs.bsw;
7 by adm_rno;
8
9 if INJ_010=1 then ActLimRepStr=1; /*Has repetitive strain injury, activity limiting*/
10 else if INJ_010=2 then ActLimRepStr=0; /*Has repetitive strain injury, not activity limiting*/
11 else if INJ_005=2 then ActLimRepStr=0; /*No repetitive strain injury*/
12 else ActLimRepStr=.; /*For non-response to INJ_005 or INJ_010*/
13
14 if DHHGAGE in (1:3) then Agegrp=1; /*12 to 19*/
15 else if DHHGAGE in (4:7) then Agegrp=2; /*20 to 39*/
16 else if DHHGAGE in (8:12) then Agegrp=3; /*40 to 64*/
17 else if DHHGAGE in (13:16) then Agegrp=4; /*65 and older*/
18
19
20
21 keep adm_rno FWGT BSW: ActLimRepStr Agegrp;
22 run;
```

SAS coding example

```
26 □ proc surveyfreq data=PUMF
27     varmethod=BRR nosummary;
28     tables Agegrp*ActLimRepStr / cl(type=wilson truncate=yes ADJUST=NO) clwt row;
29     weight FWGT;
30     repweight BSW1-BSW1000;
31     ods output crosstabs=temp_estimate;
32 run;
```

SAS coding example

Table of Agegrp by ActLimRepSt

Agegrp	ActLimRepSt	Frequency	Weighted Frequency	Std Err of Wgt Freq	95% Confidence Limits for Wgt Freq		Percent	Std Err of Percent	95% Confidence Limits for Percent		CV for Percent	Row Percent	Std Err of Row Percent	95% Confidence Limits for Row Percent		CV for Row Percent
1	0	10156	3012559	1042763	966304	5058814	9.6585	3.3501	4.8688	18.2557	0.3469	96.6831	7.8376	53.8718	99.8627	0.0811
	1	374	103353	241202	0	576672	0.3314	0.7732	0.0144	7.1211	2.3334	3.3169	7.8376	0.1373	46.1282	2.3629
	Total	10530	3115912	1016944	1120323	5111501	9.9898	3.2669	5.2302	18.2469	0.3270	100.0000				
2	0	25925	9215193	1714372	5851014	12579372	29.5446	5.4676	20.1173	41.1157	0.1851	93.7812	7.5626	64.3482	99.2126	0.0806
	1	1849	611078	755992	0	2094591	1.9592	2.4238	0.2489	13.7936	1.2372	6.2188	7.5626	0.7874	35.6518	1.2161
	Total	27774	9826271	1680876	6527822	13124720	31.5037	5.3589	22.1095	42.7014	0.1701	100.0000				
3	0	38523	11145329	1433189	8332926	13957732	35.7327	4.5894	27.3551	45.0838	0.1284	91.0338	7.6546	66.1559	98.1391	0.0841
	1	3888	1097732	949059	0	2960107	3.5194	3.0405	0.7415	15.1186	0.8639	8.9662	7.6546	1.8609	33.8441	0.8537
	Total	42411	12243061	1246607	9796796	14689326	39.2521	3.9714	31.8186	47.2196	0.1012	100.0000				
4	0	31057	5772501	340543	5104240	6440762	18.5071	1.0920	16.4609	20.7444	0.0590	96.1193	5.3936	71.3401	99.5959	0.0561
	1	1259	233060	323935	0	868731	0.7472	1.0385	0.0796	6.6385	1.3899	3.8807	5.3936	0.4041	28.6599	1.3898
	Total	32316	6005561	108098	5793437	6217685	19.2543	0.3438	18.5887	19.9378	0.0179	100.0000				
Total	0	105661	29145582	1267116	26659071	31632093	93.4429	3.9810	81.0826	97.9330	0.0426					
	1	7370	2045223	1242758	0	4483935	6.5571	3.9810	2.0670	18.9174	0.6071					
	Total	113031	31190804	327050	30549022	31832587	100.0000									

Wilson confidence limits are computed for percents.

Frequency Missing = 259



Sampling variability guidelines

- For totals, means, ratios and proportions

TYPE OF ESTIMATE	CV	CCHS INDICATOR	QUALITY GUIDELINE
Acceptable	0.0-15.0	A	General unrestricted release
Marginal	15.1-35.0	E	General unrestricted release but with warning cautioning users of the high sampling variability. Should be identified by letter E.
Unacceptable	> 35.0	F	StatCan recommends no release. Should be suppressed and replaced by letter F.

Future CCHS PUMFs

- Next two-year cycle 2021-2022 broken, due to changes in 2022
 - Transition to Electronic Questionnaire
 - Revisit content and develop a new content plan
 - Revisited methodology
 - Non-response bias difference, due to response rates
- => TBD if 2021 and 2022 will be released as PUMFs, and what order