

## Sampling Weights and Design of the International Tobacco Control (ITC) 4CV Wave 1 Survey

C. Boudreau<sup>1,2</sup>, M.E. Thompson<sup>1,2</sup> and Y. Li<sup>2,3</sup>

This short document describes the various longitudinal weights (section 2.1) and cross-sectional weights (section 2.2) available for the ITC 4CV Wave 1 Survey (labeled as wave V1 in this document to prevent confusion with previous waves of the ITC 4C Survey). It also provides some guidance on which set of weights should be used depending on the analysis they are performing, as well as some basic characteristics of the sampling design and protocol.

### Contents

<b>1</b>	<b>Important remark about quitters</b>	<b>2</b>
<b>2</b>	<b>Sampling weights</b>	<b>2</b>
2.1	Longitudinal sampling weights . . . . .	5
2.2	Cross-sectional sampling weights . . . . .	6
2.2.1	Cross-sectional weights for the main sample . . . . .	6
2.2.2	Cross-sectional weights for the reduced US sample . . . . .	9
2.2.3	Cross-sectional weights for the Australian e-cig users . . . . .	10
2.3	Inflation versus rescaled weights . . . . .	11
2.4	Additional remarks . . . . .	12
2.4.1	Cautionary note about quitters . . . . .	12
2.4.2	Cautionary note about e-cigarette users . . . . .	12
2.4.3	Covariates to include in statistical modelling . . . . .	13
<b>3</b>	<b>Data Collection &amp; Sampling Design</b>	<b>14</b>
3.1	ITC 4C Survey . . . . .	14
3.2	Data Collection and Recruitment . . . . .	14
3.3	Sample Sizes and Inclusion Criteria . . . . .	14
3.4	Sample Sources . . . . .	15
3.5	Sampling Protocols . . . . .	16
3.6	Cautionary notes . . . . .	18

---

<sup>1</sup>Dept. of Statistics & Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada.

<sup>2</sup>Data Management Core (DMC) – ITC Project, University of Waterloo.

<sup>3</sup>Dept. of Psychology, University of Waterloo.

<sup>4</sup>This document was created using L<sup>A</sup>T<sub>E</sub>X, and last updated on Mar. 21, 2018

## 1 Important remark about quitters

After careful consideration, it was decided not to compute cross-sectional weights for long-term quitters (i.e., those who had quit smoking more than five years ago at the time of wave V1 data collection) in Canada, the US and England. This resulted in a total of 224 long-term quitters (79 from the US, 95 from Canada and 50 from England) not receiving any cross-sectional weights. Note that this only applies to cross-sectional weights, and that longitudinal weights were computed for those quitters; see section 2.1.

As mentioned in section 2.4.1, the vast majority of quitters in the ITC 4CV sample were initially recruited as smokers. Since quitting is one of the reasons for dropping out of the survey, quitters in the 4CV sample should not be viewed as representative of quitters in the general population. This lack of representativeness becomes more and more important when length of quit gets longer. Hence the decision to remove those who have quit more than five years ago. This has also the advantage of making the quitters in Canadian, US and English samples more comparable to those in the Australian sample, where it was decided that respondents who had quit more than two years ago were not to be recontacted at wave V1. Last but not least, since this sample of 224 long-term quitters would have had “represented” millions of quitters in those three countries (e.g., about 20% of Canadians and about 17% of Americans are long-term quitters), this would have resulted in very high sampling weights for those respondents. Consequently, analyses including all quitters (i.e., both those who have quit more than five years ago and those who have quit within the last five years) would have been dominated by those long-term quitters. Likewise, analyses involving all respondents (i.e., cigarette smokers, e-cigarette users and quitters) would have been greatly biased towards the behaviours of those long-term quitters.

## 2 Sampling weights

Four sets of longitudinal weights (section 2.1) and 18 sets of cross-sectional weights (section 2.2) were computed at wave 1 of the ITC 4CV Survey; see tables 1 and 2 below. The 18 sets of cross-sectional weights can be divided into 3 categories:

1. The first 7 sets of cross-sectional weights (see section 2.2.1 and column 2 of table 2) were computed for the main sample. That sample consists of 12295 respondents (2733 from the US, 3734 from Canada, 4324 from England, and 1504 from Australia). It includes all respondents except the 581 Australian respondents from the dedicated CCV sample, the 224 long-term quitters (79 from the US, 95 from Canada, 50 from England, and 0 from Australia) mentioned in section 1 and the 87 respondents (26 from the US, 13 from Canada, 48 from England, and 0 from Australia) deemed to be fraudulent (also referred to as speeders in some 4CV documentation).
2. The next 7 sets of cross-sectional weights (see section 2.2.2 and column 3 of table 2) were computed for the reduced US sample. This sample is the main US sample of

2733 respondents (mentioned above) minus the 494 US e-cigarette users recruited by Ipsos; hence, the reduced US sample consists of 2239 respondents. Those e-cigarette users were recruited via a non-probability based panel, and those sets of cross-sectional weights were thus computed to give data users the flexibility to remove those individuals if they so desired for their specific analyses. Since this only applies to US respondents, no such weights were computed for respondents from Canada, England or Australia.

3. The last 4 sets of cross-sectional weights (see section 2.2.3 and column 4 of table 2) were computed for Australian respondents using e-cigarettes. This sample contains all 816 Australian respondents that were using e-cigarettes at the time of wave V1. This was done to allow for weighted analyses including the 581 respondents recruited via the dedicated CCV sample. This concerns Australian respondents only, and no such weights were computed for respondents from the US, Canada or England.

Weight	Variable Name
Rescaled waves 8–V1 longitudinal weights	kWTS967v
Rescaled waves 8.5–V1 longitudinal weights (Australia only)	kWTS969v
Rescaled waves 9–V1 longitudinal weights	kWTS971v
Rescaled waves 10–V1 longitudinal weights (England and Australia only)	kWTS973v

Table 1: List of the available longitudinal sampling weights for wave V1 of the ITC 4CV Survey

Weight	Variable Names		
	Main sample*	reduced US sample <sup>†</sup>	AU e-cig users <sup>‡</sup>
Wave V1 cross-sectional inflation weights	kWTS100v	kWTS102v	n/a
Wave V1 cross-sectional inflation weights for e-cigarette users	n/a	n/a	kWTS304v
Rescaled wave V1 cross-sectional weights for cigarette smokers	kWTS201v	kWTS203v	n/a
Rescaled wave V1 cross-sectional weights for e-cigarette users	kWTS301v	kWTS303v	kWTS305v
Rescaled wave V1 cross-sectional weights for dual users	kWTS401v	kWTS403v	kWTS405v
Rescaled wave V1 cross-sectional weights for quitters	kWTS501v	kWTS503v	n/a
Rescaled wave V1 cross-sectional weights for quitters who use e-cigarettes	n/a	n/a	kWTS505v
Rescaled wave V1 cross-sectional weights for all tobacco users	kWTS601v	kWTS603v	n/a
Rescaled wave V1 cross-sectional weights for all respondents (i.e., all tobacco users and quitters)	kWTS101v	kWTS103v	n/a

\* The main sample consists of 12295 respondents (2733 from the US, 3734 from Canada, 4324 from England, and 1504 from Australia). It includes all respondents except the 581 Australian respondents from the dedicated CCV sample, the 224 long-term quitters (79 from the US, 95 from Canada, 50 from England, and 0 from Australia) mentioned in the remark section of page 1 and the 87 respondents (26 from the US, 13 from Canada, 48 from England, and 0 from Australia) deemed to be fraudulent (also referred to as speeders in some 4CV documentation).

<sup>†</sup> The reduced sample is the main US sample of 2733 respondents minus the 494 US e-cigarette users recruited by Ipsos; hence, the reduced US sample consists of 2239 respondents. Those e-cigarette users were recruited via a non-probability based panel, and those sets of cross-sectional weights were thus computed to give data users the flexibility to remove those individuals if they so desired for their specific analyses.

<sup>‡</sup> This sample contains all 816 Australian respondents that were using e-cigarettes at the time of wave V1. As detail in section 2.2.3, this was done to allow for weighted analyses including the 581 respondents recruited via the dedicated CCV sample.

Table 2: List of the available cross-sectional sampling weights for wave V1 of the ITC 4CV Survey

## 2.1 Longitudinal sampling weights

- 1- Variable `kWTS967v` contains the waves 8–V1 longitudinal weights for the 1202 respondents (258 from the US, 467 from Canada, 240 from England, and 237 from Australia) who completed the wave 8 survey, and were successfully retained and interviewed at wave V1 (and all the waves in between). These weights are designed to make these 467 Canadian smokers (and quitters) representative of the Canadian population of smokers at the time of wave 8 data collection; likewise for the US, England and Australia.
- 2- Variable `kWTS969v` contains the waves 8.5–V1 longitudinal weights for the 276 Australian respondents who completed the wave 8.5 survey, and were successfully retained and interviewed at wave V1 (and all the waves in between). These weights are designed to make these 276 smokers (and quitters) representative of the Australian population of smokers at the time of wave 8.5 data collection.
- 3- Variable `kWTS971v` contains the waves 9–V1 longitudinal weights for the 2662 respondents (1400 from the US, 661 from Canada, 269 from England, and 332 from Australia) who completed the wave 9 survey, and were successfully retained and interviewed at wave V1 (and all the waves in between). These weights are designed to make these 1400 US smokers (and quitters) representative of the American population of smokers at the time of wave 9 data collection; likewise for the Canada, England and Australia.
- 4- Variable `kWTS973v` contains the waves 10–V1 longitudinal weights for the 819 respondents (304 from England, and 515 from Australia) who completed the wave 10 survey, and were successfully retained and interviewed at wave V1. These weights are designed to make these 304 English smokers (and quitters) representative of the English population of smokers at the time of wave 10 data collection; likewise for Australia.

**Note:** It was decided to no longer compute waves 1–V1, waves 2–V1, . . . , waves 7–V1 longitudinal weights. There are essentially 3 reasons for this: 1) before wave 8, all interviews were conducted by phone, whereas (practically<sup>1</sup>) all interviews at wave V1 were done online,

2) the tobacco landscape in the 4 countries has changed quite a lot since waves 1–7 took place, and 3) given the high level of attrition of cohorts 1–7 respondents between waves 8 and V1, the wisdom of conducting longitudinal analyses of those cohorts is becoming more and more questionable. Note that this is different from conducting cross-sectional analyses, or waves 8–V1, waves 8.5–V1, waves 9–V1 or waves 10–V1 longitudinal analyses. Though cohorts 1–7 respondents (successfully retained and interviewed at wave V1) will be part of the sample when conducted such cross-sectional or longitudinal analyses, they are only a fraction of the sample, as opposed to the whole sample when conducting waves 1–V1, waves 2–V1, etc. longitudinal analyses. It is thus perfectly acceptable to include cohorts 1–7 respondents with cohorts 8–V1 respondents in cross-sectional and longitudinal analyses.

<sup>1</sup>46 US cohort respondents were exceptionally allowed to complete the V1 survey by phone.

## 2.2 Cross-sectional sampling weights

User group <sup>†</sup>	US		Canada	England	Australia	
	Main <sup>‡</sup>	Reduced <sup>‡</sup>			Main <sup>‡</sup>	E-cig <sup>‡</sup>
Cigarette only	1347	1347	2180	2664	1147	0
Dual users	980	542	1036	1222	192	292
Pure e-cigarette users	37	20	71	17	5	29
Quitters						
E-cig users	126	87	103	181	38	495
Do not use e-cig	243	243	344	240	122	0
Total	369	330	447	421	160	495
Total	2733	2239	3734	4324	1504	816

<sup>†</sup> Variable `kUserGroup` in the dataset

<sup>‡</sup> See notes below Table 2

Table 3: 4CV wave V1 respondents by country and user group.

### 2.2.1 Cross-sectional weights for the main sample

1- Variable `kWTS100v` contains the wave V1 cross-sectional inflation weights for the main sample of 12295 respondents (2733 from the US, 3734 from Canada, 4324 from England, and 1504 from Australia).

These weights were computed by dividing the respondents into four broad user groups (variable `kUserGroup`<sup>2</sup> in the dataset): i) cigarette only users, ii) dual users, iii) pure e-cigarette users and iv) quitters; see table 3. For the US, Canada and England, the quitter group consists of individuals that have quit cigarette smoking within the last 5 years; whereas in Australia it consists of individuals that have quit cigarette smoking within the last 2 years. Not that some of those quitters were using e-cigarettes at the time of wave V1 data collection. The pure e-cigarette users group consists of individuals who have smoked less than 100 cigarettes in their lifetime and were using e-cigarettes at the time of data collection, as well as who have quit more than five years ago (more than 2 years ago in Australia) and using e-cigarettes. In addition to those 4 groups, quitters were further divided into 4 sub-groups (variable `kQuitGroup`<sup>3</sup> in the dataset): iv.a) those who had quit within the last year, but were using e-cigarettes at the time of wave V1 data collection, iv.b) those who had quit 1–5 years ago (1–2 years in Australia), but were

<sup>2</sup> Where 1 = cigarette only, 2 = dual users, 3 = pure e-cigarette users, and 4 = quitters; see [appendix](#)

<sup>3</sup> Where 1 = quit within the last year and uses e-cigarettes, 2 = quit more than 1 year ago and uses e-cigarettes, 3 = quit within the last year and does not use e-cigarettes, and 4 = quit more than 1 year ago and does not use e-cigarettes; see [appendix](#)

using e-cigarettes at the time of data collection, iv.c) those who had quit within the last year and were not using e-cigarettes at the time of data collection, and iv.d) those who had quit 1–5 years ago (1–2 years in Australia) and were not using e-cigarettes at the time of data collection; see table 3.

Calibration/target figures (e.g., estimated number of individuals that are dual users and estimated number of individuals that are e-cigarette users) were then obtained for each of the four groups and four subgroups. In the US, those calibration figures were obtained from the 2016 [National Health Interview Survey \(NHIS\)](#); whereas as the 2015 [Canadian Tobacco Alcohol and Drugs Survey \(CTADS\)](#), the 2015 [Opinions and Lifestyle Survey \(OPN\)](#), and the 2016 [National Drug Strategy Household Survey \(NDSHS\)](#) were used for Canada, England and Australia, respectively. For groups i, ii and iv, estimated number of individuals in each cells of the following cross-tabs were obtained: user group  $\times$  gender, user group  $\times$  age group, user group  $\times$  geographic region, user group  $\times$  ethnicity (US only), user group  $\times$  education (except for Canada) and user group  $\times$  language (Canada only). For quitters (group iv), estimates were also obtained for the 4 subgroups mentioned above. Since the pure e-cigarette groups are fairly small (see table 3), it was not possible/practical to use as many cross-tabs as for the other three groups. Hence, for group iii, separate estimates were obtained for gender (male vs. female) and age group (18–24 vs.  $> 24$ ) in the US, Canada and England (since there are only 5 such respondents in Australia, they were combined into a single group for weight calculation purposes). A raking procedure was then applied to calibrate the weights using the above mentioned cross-tabs; this was done separately for each country.

These weights are designed to make respondents in each of the four groups representative of the corresponding population at the time of wave V1 data collection. For example, the `kWTS100v` weights of the 1036 Canadian dual users (i.e., individuals who smoked traditional cigarettes and also use e-cigarettes) are designed to make them representative of the Canadian population of dual users at the time of data collection; likewise for the other countries and the other groups. If interests lie in a target population that consists of two or more of the four user groups, the `kWTS100v` weights are still appropriate. For example, when studying Canadian cigarette smokers, one can simply combine the `kWTS100v` weights of the 2180 cigarette only users with those of the 1036 dual users (for a total of 3216 respondents in the analysis), and assigned a weight of 0 to respondents in the other two user groups.

Last but not least, since these are inflation/un-rescaled weights, they should not be used in analyses involving two or more countries. The various rescaled weights (i.e., variables `kWTS101v` to `kWTS601v`) described below were created especially for such multi-country analyses; see section 2.3 for more information on inflation versus rescaled weights.

2- Variable `kWTS201v` contains the rescaled wave V1 cross-sectional weights for the 10768 (2327 from the US, 3216 from Canada, 3886 from England and 1339 from Australia; see table 3) respondents who were cigarette smokers at the time of wave E1 data collection. These are simply the wave V1 cross-sectional inflation weights (variable `kWTS100v`) of

those 10768 respondents rescaled to sum to sample size in each country (i.e., 2327 in the US, 3216 in Canada, 3886 in England and 1339 in Australia). These weights are designed to make these 3216 Canadian cigarette smokers representative of the Canadian population of cigarette smokers at the time of wave V1 data collection; likewise for the US, England and Australia.

- 3- Variable `kWTS301v` contains the rescaled wave V1 cross-sectional weights for the 4589 (1143 from the US, 1210 from Canada, 1420 from England and 816 from Australia; see table 3) respondents who were e-cigarette users at the time of wave V1 data collection. These are simply the wave V1 cross-sectional inflation weights (variable `kWTS100v`) of those 4589 respondents rescaled to sum to sample size in each country. These weights are designed to make these 1210 Canadian e-cigarette users representative of the Canadian population of e-cigarette users at the time of wave V1 data collection; likewise for the US, England and Australia.
- 4- Variable `kWTS401v` contains the rescaled wave V1 cross-sectional weights for the 3530 (980 from the US, 1036 from Canada, 1222 from England and 292 from Australia; see table 3) respondents who were dual users at the time of wave V1 data collection. These are simply the wave V1 cross-sectional inflation weights (variable `kWTS100v`) of those 3530 respondents rescaled to sum to sample size in each country. These weights are designed to make these 1036 Canadian dual users representative of the Canadian population of dual users at the time of wave V1 data collection; likewise for the US, England and Australia.
- 5- Variable `kWTS501v` contains the rescaled wave V1 cross-sectional weights for the 1397 (369 from the US, 447 from Canada, 421 from England and 160 from Australia; see table 3) quitters at the time of wave V1 data collection. These are simply the wave V1 cross-sectional inflation weights (variable `kWTS100v`) of those 1397 respondents rescaled to sum to sample size in each country (i.e., 369 in the US, 447 in Canada, 421 in England and 160 in Australia).  
Though these weights are designed to make these 447 Canadian quitters as representative as possible of the Canadian population of quitters at the time of wave V1 data collection, the vast majority of those quitters were initially recruited as smokers. Consequently, they are an imperfect sample when it comes to be representative of the whole population of quitters; see section 2.4.1 on the representativeness of quitters in the ITC 4CV sample. The same cautionary note also applies to the US, England and Australia. Furthermore, the US, Canadian and English samples consist of those who have quit smoking within the last five years, whereas the Australian sample consists of those who have quit smoking within the last two years. Since the target populations are different, care must be taken when comparing Australian quitters with those of the other three countries.
- 6- Variable `kWTS601v` contains the rescaled wave V1 cross-sectional weights for the 11346 (2490 from the US, 3390 from Canada, 4084 from England and 1382 from Australia; see table 3) respondents who were tobacco users (i.e., cigarette smoker and/or e-cigarette

users) at the time of wave V1 data collection. These are simply the wave V1 cross-sectional inflation weights (variable `kWTS100v`) of those 11346 respondents rescaled to sum to sample size in each country. These weights are designed to make these 3390 Canadian tobacco users representative of the Canadian population of tobacco users (i.e., cigarette smoker and/or e-cigarette users) at the time of wave V1 data collection; likewise for the US, England and Australia.

- 7- Variable `kWTS101v` contains the rescaled wave V1 cross-sectional weights for the main sample of 12295 respondents (2733 from the US, 3734 from Canada, 4324 from England, and 1504 from Australia; see table 3). These are simply the wave V1 cross-sectional inflation weights (variable `kWTS100v`) of those 12295 respondents rescaled to sum to sample size in each country (i.e., 2733 in the US, 3734 in Canada, 4324 in England and 1504 in Australia). These weights are designed to make these 3734 Canadian tobacco users and quitters representative of the Canadian population of tobacco users and quitters at the time of wave V1 data collection; likewise for the US, England and Australia.

In addition to the warning about the representativeness of quitters in the ITC 4CV sample (see variable `kWTS501v` above), it should be noted that tobacco users and quitters are ultimately two distinct populations. Hence, great care must be taken when deciding to analyse them together using the `kWTS101v` weights. This is probably fine when the goal is to carry out descriptive inference about the joint population of tobacco users and quitters. However, carrying out analytical inference (e.g., linear regression and logistic regression) from that same joint population is probably much more questionable.

### 2.2.2 Cross-sectional weights for the reduced US sample

As mentioned at the beginning of section 2.2, the following 7 sets of weights only applies to the US sample.

- 1- Variable `kWTS102v` contains the wave V1 cross-sectional inflation weights for the reduced US sample of 2239 respondents. As noted below table 2, the reduced sample is the main sample of 2733 respondents minus the 494 US e-cigarette users recruited by Ipsos (identifiable via the variables `kOWNERID` and `kEC309v`<sup>4</sup> in the dataset). Those e-cigarette users were recruited via a non-probability based panel, and the 7 sets of cross-sectional weights were thus computed to give data users the flexibility to remove those individuals if they so desired for their specific analyses.

These weights were computed the same way as the wave V1 cross-sectional inflation weights for the main sample (see variable `kWTS100v` in section 2.2.1). Consequently, respondents were divided into four broad user groups (cigarette only users, dual users, e-cigarette only users and quitters), and the weights were then calibrated (user group  $\times$  gender, user group  $\times$  age group, user group  $\times$  4 Census Regions, user group  $\times$  ethnicity,

---

<sup>4</sup>Use  $!(\text{kOWNERID}=5 \text{ and } \text{kEC309v} \in \{1,2\})$  to exclude the 494 US e-cigarette users recruited by Ipsos, where  $!$  is the logic symbol for not.

user group  $\times$  education and the 4 quitter sub-groups) based on figures from the 2016 NHIS.

- 2- Variable `kWTS203v` contains the rescaled wave V1 cross-sectional weights for the 1889 respondents (see table 3) from the reduced US sample who were cigarette smokers at the time of wave V1 data collection. These are simply the wave V1 cross-sectional inflation weights (variable `kWTS102v`) of those 1889 respondents rescaled to sum to sample size (i.e., 1889). These weights are designed to make these 1889 US cigarette smokers representative of the US population of cigarette smokers at the time of wave V1 data collection. In other words, variable `kWTS203v` is the analogue of variable `kWTS201v`, but for the reduced sample.
- 3- Variable `kWTS303v` is the analogue of variable `kWTS301v`, but for the reduced US sample of 649 respondents who were e-cigarette users at the time of wave V1 data collection.
- 4- Variable `kWTS403v` is the analogue of variable `kWTS401v`, but for the reduced US sample of 542 respondents who were dual users at the time of wave V1 data collection.
- 5- Variable `kWTS503v` is the analogue of variable `kWTS501v`, but for the reduced US sample of 330 respondents who were quitters at the time of wave V1 data collection.
- 6- Variable `kWTS603v` is the analogue of variable `kWTS601v`, but for the reduced US sample of 1996 respondents who were tobacco users (i.e., cigarette smoker and/or e-cigarette users) at the time of wave V1 data collection.
- 7- Variable `kWTS103v` is the analogue of variable `kWTS101v`, but for the reduced US sample of 2239 respondents. Hence, the same cautionary notes listed for variable `kWTS101v` also applies here.

### 2.2.3 Cross-sectional weights for the Australian e-cig users

As mentioned at the beginning of section 2.2, the following 4 sets of weights only applies to the Australian sample of 816 respondents (235 from the main sample and 581 from the CCV sample; see table 3) that were using e-cigarettes at the time of wave V1. No such weights were computed for respondents from the US, Canada, England or for Australian respondents that are not e-cigarette users.

It should first be mentioned that CCV vaper sample (identifiable via the variable `kOWNERID`<sup>5</sup> in the dataset) consists of self-selected individuals recruited mainly through on-line sites. They are a group of vaper activists and early adopters. They are thus not a representative sample of vaper users in Australia. This is why those respondents are excluded from the 7 sets of weights computed in section 2.2.1. Nevertheless, respondents from the CCV sample can be of scientific interests, and some data users might want to include them in their analyses. This is why the following 4 sets of weights were computed.

---

<sup>5</sup>Use `kOWNERID=7` to select the 581 Australian respondents from the dedicated CCV sample.

The initial plan was to compute weights for the entire/full sample of 2085 Australian respondents. Though this was technically possible, those weights would have been unstable. It was thus decided to instead compute weights for the 816 respondents that were using e-cigarettes at the time of wave V1. There are 617 quitters (individuals who have quit cigarette smoking within the last 2 years) in the full sample. Of those, only 122 (or 19.7%) are not e-cigarette users. However, according to the 2016 NDSHS data, 96.4% of those who quit within the last two years are not using e-cigarettes. This huge difference between the ITC sample of quitters and the 2016 NDSHS target data results in the weights of quitters who do not use e-cigarettes being extremely large (the average weight of a quitter who does not use e-cigarette is about 200 times that of a quitter who uses e-cigarettes); making the weights unstable.

- 1- Variable `kWTS304v` contains the wave V1 cross-sectional inflation weights for the Australian sample of 816 respondents (235 from the main sample and 581 from the CCV sample; see table 3) that were using e-cigarettes at the time of wave V1. These weights were computed in a very similar way as the wave V1 cross-sectional inflation weights for the main sample (see variable `kWTS100v` in section 2.2.1). However, respondents were first divided into three (dual users, e-cigarette only users and quitters using e-cigarettes) instead of for user groups. The weights were then calibrated (user group  $\times$  gender, user group  $\times$  age group, and the 2 quitter sub-groups) based on figures from the 2016 NDSHS.
- 2- Variable `kWTS305v` contains the rescaled wave V1 cross-sectional weights for the 816 (235 from the main sample and 581 from the CCV sample; see table 3) Australian respondents who were e-cigarette users at the time of wave V1 data collection. These are simply the wave V1 cross-sectional inflation weights (variable `kWTS304v`) of those 816 respondents rescaled to sum to sample size (i.e, 816). These weights are designed to make these 816 e-cigarette users representative of the Australian population of e-cigarette users at the time of wave V1 data collection. In other words, variable `kWTS305v` is the analogue of variable `kWTS301v`, but for all 816 Australian e-cigarette users.
- 3- Variable `kWTS405v` is the analogue of variable `kWTS401v`, but for the 292 (192 from the main sample and 100 from the CCV sample; see table 3) Australian respondents who were dual users at the time of wave V1 data collection.
- 4- Variable `kWTS505v` contains the rescaled wave V1 cross-sectional weights for the 495 (38 from the main sample and 457 from the CCV sample; see table 3) Australian respondents who had quit cigarette smoking but were using e-cigarettes at the time of wave V1 data collection.

### 2.3 Inflation versus rescaled weights

The main reason for rescaling the weights is to facilitate joint analyses involving data from multiple ITC countries. Taking wave V1 as an example, from the data used to calibrate the

weights, there were about 39.8 million cigarette smokers in the United States at the time of wave V1 data collection; whereas there were only 3.6 millions such individuals in Canada, 7.4 millions in England and 2.7 millions in Australia. Hence, any joint analysis using data from all four countries will be dominated by the US if the inflation weights (i.e., variables [kWTS100v](#) or [kWTS102v](#)) are used.

On the other hand, the various rescaled weights sum to the sample size, as described above. Hence, if the rescaled weights are used, England and Canada will have a slightly greater impact on the results (since the Canadian and British sample sizes are larger than the Australian and US samples; see table 3), but no country will dominate the analysis. In summary, rescaling the weights to sum to the sample size is a simple and efficient way to make countries with different population sizes comparable. This also holds true when comparing 4CV data to other ITC countries; for example, ITC Netherlands and ITC 6E.

Last but not least, it should be mentioned that rescaling the weights will not affect the results when estimating population means and proportions/percentages, as well as when fitting various statistical models (e.g., logistic and linear regressions). However, the rescaled weights should not be used to estimate population totals (e.g., the total number of daily smokers or e-cigarette users).

## 2.4 Additional remarks

### 2.4.1 Cautionary note about quitters

As mentioned above (see variable [kWTS501v](#)), the vast majority of quitters in the ITC 4CV sample were initially recruited as smokers. Since quitting is one of the reasons for dropping out of the ITC Survey, quitting experience of our cohort respondents could well be affected by being in the sample and because of the sampling design itself, quitters in the ITC 4CV sample should not be considered to be representative of quitters in the population. For example, comparisons between the quitters in the ITC sample and quitters in the cross-sectional [Smoking Toolkit Study](#) showed an important discrepancy in distribution of length of time quit.

As described in the section detailing the construction of the wave V1 cross-sectional inflation weights (see variable [kWTS100v](#)), the sampling weights of quitters were calibrated on gender, age group, geographic region, ethnicity (US only), education (except for Canada), language (Canada only) and use of e-cigarettes  $\times$  length of quit ( $\leq 1$  year vs.  $> 1$  year). Quitters in the ITC 4CV sample should thus be representative of the population in terms of those variables, but not in terms of other related attributes. Again, they are an imperfect sample when it comes to be representative of the whole population of quitters.

### 2.4.2 Cautionary note about e-cigarette users

Comparisons on measures related to e-cigarettes between the 4CV1 Survey and other ITC countries must be viewed with caution, as weight construction for the 4CV1 Survey was done in a different fashion than of other ITC countries.

As described in the section detailing the construction of the wave V1 cross-sectional inflation weights (see variable `kWTS100v`), weights were computed by first dividing respondents into four broad user groups: cigarette only users, dual users, pure e-cigarette users and quitters. The sampling weights of dual users and of quitters were then calibrated on gender, age group, geographic region, ethnicity (US only), education (except for Canada) and language (Canada only), whereas those of pure e-cigarette users were calibrated on gender and age group. The sampling weights of quitters were also calibrated on use of e-cigarettes and length of quit ( $\leq 1$  year vs.  $> 1$  year). In other ITC countries, separate estimates for the number of individuals that are dual users and individuals that only smoke cigarettes were not available at the time of weight calculation. Hence, the weights were calibrated using smoking prevalences (often by age/gender groups and/or geographic regions), and thus have no special adjustment for e-cigarette usage.

### 2.4.3 Covariates to include in statistical modelling

As with other surveys, it is good practice to include the survey design variables and the variables used in the weight construction, when fitting statistical models (e.g, linear or logistic regression models) using ITC 4CV data. Hence, we highly recommend that any statistical model includes the following covariates:

- gender (labelled `sex` in the dataset)
- age (labelled `kAGE`, continuous, and `kageGroup`, categorical, in the dataset)
- user group (labelled `kUserGroup` in the dataset); see description of the `kWTS100v` weight variable.

The geographic region (labelled `kStrata` in the dataset) should also be used as the stratification variable in the statistical software. Though somewhat less essential, users should also strongly consider adding the following covariates:

- ethnicity (labelled `ethnic` in the dataset), when fitting models using US data
- education (labelled `kDE312v` in the dataset), when fitting models using US, English and/or Australian data
- language (labelled `kcaFrench` in the dataset), when fitting models using Canadian data
- use of e-cigarettes and/or length of quit (labelled `kQuitGroup` in the dataset), when fitting data using quitters; see description of the `kWTS100v` weight variable
- frequency of use of e-cigarettes/vaping (labelled `kEC309v` in the dataset), when fitting models with e-cigarettes users.

## 3 Data Collection & Sampling Design

### 3.1 ITC 4C Survey

Respondents in these countries had previously been surveyed in the ITC Four Country (4C) Survey, as described at <http://www.itcproject.org/methods>. The respondents of the first few waves of the ITC 4C Survey were recruited by telephone. Telephone recruitment continued in later waves, but was eventually supplemented by recruitment from commercial panels in Canada (wave 9), the UK (waves 9 and 10) and the US (wave 9), as more and more of the sample (including telephone recruits) were responding online. Respondents (mainly the telephone recruits) for whom contact information has been retained by the ITC Project are referred to as the ITC-owned in this document.

### 3.2 Data Collection and Recruitment

- During the design stage of ITC 4CV, it was decided that all survey response data would be collected through an online questionnaire, hosted by the [Survey Research Centre](#) at the University of Waterloo. This requirement was later relaxed for 46 respondents in the US, who were members of the ITC-owned cohort and who agreed to participate by telephone, but not online.
- All members of the ITC-owned cohort were to be invited to the new study. (Exception: In Australia, those who had been quit for more than 24 months at Wave 10 of the ITC 4C Survey were not contacted for the new study.)
- New respondents in each country were to be obtained from the best possible survey firms and sources, subject to budgetary constraints. In Australia, the main part of the sample was a custom-recruited by telephone and in part obtained from a partner (SSI) by the firm [Roy Morgan Research](#); in the US, Canada and England, firms with online panels supplied the recruits.

### 3.3 Sample Sizes and Inclusion Criteria

- US, Canada and England
  - In the US, Canada and England, the sample consisted primarily of cigarette smokers or recent (within past 24 months) quitters of cigarette smoking, aged 18 and over.
  - Because of the importance of studying younger smokers, it was decided to obtain altogether 500 such individuals in the 18–24 age group in the US and 900 in England, and 750 in Canada. The sample sizes in Canada and England are larger because of the expected lower retention rates between waves 1 and 2.

- The wave 1 sizes of the samples aged 25 & older were chosen so as to provide at least 1400 who would be present in both waves 1 and 2 (about 18 months later), according to retention rate estimates provided by the firms at the design stage.
  - As well, an additional sample of 500 at-least-weekly users of e-cigarettes were to be recruited in the US and England, and 715 in Canada.
  - In Canada, geographic quotas crossed with language (i.e., French vs English) were applied to each of the three groups: 18–24 smokers (and recent quitters), 25 & older smokers (and recent quitters), and the sample of 715 e-cigarette users.
  - In the US, geographic quotas were used for the sample of 18–24 smokers (and recent quitters) and for the sample of e-cigarette users. No quotas were used for the sample of 25 & older smokers (and recent quitters), as the majority of those individuals were cohort members from the ITC 4C Survey.
  - In England, geographic quotas crossed with sex were applied to the 18–24 age group of smokers (and recent quitters). For the sample of e-cigarette users, only geographic quotas were used; whereas, geographic quotas crossed with sex and age group (i.e., 25–34 vs 35–49 vs 50 & older) were used for the sample of 25 & older smokers (and recent quitters).
- Australia
    - In Australia, the sample was to consist mainly of cigarette smokers or recent (within past 24 months) quitters of cigarette smoking, aged 18 and over.
    - The sample size was to be approximately 1500, determined by the available budget.
    - Male/female crossed with geographic quotas were applied.
    - As well, an additional sample of users of e-cigarettes (vapers) was to be recruited by [Cancer Council Victoria](#) by referral sampling from vaper sites. This technique eventually yielded 581 respondents.

### 3.4 Sample Sources

- US
  - The ITC-owned members invited were wave 9 respondents who had been recruited as smokers by telephone in waves 1 through 8, and recruited as smokers or recent quitters by telephone or by the firm [GfK](#) (probability-based online [KnowledgePanel](#)) in wave 9.
  - ITC 4CV new recruits who were being recruited as smokers or recent quitters in the 25 & older age group were supplied by the GfK KnowledgePanel.
  - ITC 4CV new recruits who were being recruited as smokers or recent quitters in the 18–24 age group were supplied from an opt-in panel by the GfK partner [Lucid](#).

- The sample of 500 who were being recruited as e-cigarette users (but who could also be smokers) was supplied by [Ipsos US](#).
- There were also 238 additional smokers supplied by Ipsos US because of a quota programming issue, and retained in the sample.
- Canada
  - The ITC cohort members invited were wave 9 respondents who had been recruited as smokers by telephone in waves 1 through 8, and recruited as smokers by telephone or by the firm [Léger](#) (probability-based online panel) in wave 9.
  - ITC 4CV new recruits aged 18 & older who were being recruited as smokers or recent quitters were supplied by Léger, where possible from their probability-based panel.
  - The sample of approximately 715 who were being recruited as e-cigarette users (but who could also be smokers) was also supplied by Léger.
- England
  - The ITC cohort members invited were wave 10 respondents who had been recruited as smokers by telephone in waves 1 through 10. Those supplied by the online panel firm EMI in waves 9 and 10 were not included.
  - ITC 4CV new recruits aged 18 & older who were being recruited as smokers or recent quitters were supplied by [Ipsos UK](#), where possible from their own panel.
  - The sample of approximately 500 who were being recruited as e-cigarette users (but who could also be smokers) was also supplied by Ipsos UK.
- Australia
  - The ITC cohort members invited were wave 10 respondents who had been recruited as smokers by telephone in waves 1 through 10, and had not been quit by more than 24 months at wave 10.
  - ITC 4CV new recruits aged 18 & older who were being recruited as smokers or recent quitters were custom recruited by Roy Morgan Research.
  - The additional sample of 581 users of e-cigarettes (vapers) was recruited by [Cancer Council Victoria](#) by referral sampling from vaper sites.

### 3.5 Sampling Protocols

- Canada and England

In Canada and England, all new respondents were being recruited from the same sources, regardless of smoking and e-cigarette use status. Both Léger and Ipsos found it most economical to give priority to filling the e-cigarette user quotas, and thus to begin with, dual users or recent quitters who used e-cigarettes were assigned to the

e-cigarette user quotas; as these were filled, dual users and recent quitters who used e-cigarettes were assigned to the smoker/recent quitter quotas. However, the e-cigarette user quotas were not filled until relatively late in the fieldwork period. Thus, in each of Canada and England, the overall new sample in each age-sex-geography group should be approximately representative of the corresponding Léger or Ipsos population which is the union of smokers, recent quitters and e-cigarette users.

When the ITC cohort respondents are put together with the new recruits, discrepancies from the overall population in proportions of cigarette only users, dual users, and e-cigarette only users (and, to some extent, quitters who use neither) can be corrected in the combined sample by weighting; the weights are expected to be fairly smooth within age groups (18–24 and 25 & older).

- US

In the US, in each age group, the new recruit sample had two separate pieces: one part consisted of those recruited as smokers or recent quitters (who could also be e-cigarette users), and one part consisted of those recruited as e-cigarette users (who could also be smokers or recent quitters). An unweighted combination of the two samples would be expected to depart from representativity. For example:

- The relative proportions of the group who use cigarettes only and the group who use e-cigarettes only would not reflect their proportions in the population, because these sample groups come from different sources;
- There are two groups of recent quitters: those recruited with smokers and recent quitters and those recruited as e-cigarette users. The overall sample is thus likely to over-sample recent quitters who are e-cigarette users and under-sample those who are not.

When the ITC cohort respondents are put together with the new recruits, discrepancies from the overall population in proportions of cigarette only users, dual users, and e-cigarette only users are corrected in the combined sample by weighting.

- Australia

The situation in Australia resembles that in the US, with the new part of the sample being recruited in two pieces, and the added difficulty that the CCV vapor sample is quite different from the subsample of e-cigarette users in the main sample, in terms of age-sex composition, socioeconomic status (SES), and frequency of vaping.

When the ITC cohort respondents are put together with all the new recruits, discrepancies from the overall population in proportions of cigarette only users, dual users, and e-cigarette only users will be corrected to a large extent in the combined sample by weighting according to demographics and geography.

### 3.6 Cautionary notes

- Although it offers rich possibilities for analysis, the ITC 4CV Survey is not particularly well suited for estimation of prevalences. The weighted data will give prevalences of cigarette, e-cigarette and dual use that match those of the national health surveys used to calibrate the weights. However, because of rapid change in patterns of use, the calibration figures may not correspond well to the pattern which existed at the time of data collection.
- If the desired analysis involves quitters or comparison of quitters and smokers, please contact the DMC for advice.
- Cross country analyses, even among the four countries of the survey, should take into account differences in sample composition with respect to demographics, time-in-sample, and perhaps other characteristics.
- When comparing characteristics of e-cigarette users across groups or countries, controlling for frequency of use is advised.

## Appendix: pseudo code

Pseudo code detailing how variables `kUserGroup` and `kQuitGroup` were created (for US, Canada and England). The code for Australia is identical, with the exception that cut-off for quitters is 2 years instead of 5.

```
if {smokes cigarettes at least monthly (ie, kFR309v in (1,2,3))}
  and {uses e-cig less than once a month or not at all
      (ie, kEC309v in (4,5,6,7,8) or kNC304=4)}
  then kUserGroup=1; /* cig only */;

else if {smokes cigarettes at least monthly (ie, kFR309v in (1,2,3))}
  and {uses e-cig at least monthly (ie, kEC309v in (1,2) or kNC304=3)}
  then kUserGroup=2; /* dual user */;

else if {never smoked cigarette or quit over 5 years
        (ie, kFR309v=9 or kQA439=10)}
  and {uses e-cig at least monthly (ie, kEC309v in (1,2) or kNC304=3)}
  then kUserGroup=3; /* pure ecig */;

else if {quit cigarette smoking (ie, kFR309v in (4,8))}
  then kUserGroup=4; /* quitter */;

if {quitters (ie, kUserGroup=4)} then do;
  if {quit within the last 12 months (ie, kQA439<=6)}
    and {uses e-cig at least monthly (ie, kEC309v in (1,2) or kNC304=3)}
    then kQuitGroup=1;
  else if {quit 1-5 years ago (ie, 6 < kQA439 <= 9)}
    and {uses e-cig at least monthly (ie, kEC309v in (1,2) or kNC304=3)}
    then kQuitGroup=2;
  else if {quit within the last 12 months (ie, kQA439<=6)}
    and {uses e-cig less than once a month or not at all
        (ie, kEC309v in (4,5,6,7,8) or kNC304=4)}
    then kQuitGroup=3;
  else if {quit 1-5 years ago (ie, 6 < kQA439 <= 9)}
    and {uses e-cig less than once a month or not at all
        (ie, kEC309v in (4,5,6,7,8) or kNC304=4)}
    then kQuitGroup=4;
end;
```