C.A.N
CENTRALFÖRBUNDET
FÖR ALKOHOL- OCH
NARKOTIKAUPPLYSNING

Estimating total alcohol consumption based on the Monitor survey

An updated technical description of the estimation method

CAN Rapport 207

Mats Nyfjäll & Björn Trolldal

Centralförbundet för alkohol- och narkotikaupplysning, CAN

ISSN 0283-1198
ISBN 978-91-7278-336-2
urn urn:nbn:se:can-2022-1

Preface

The Swedish Council for Information on Alcohol and Other Drugs (CAN) is an independent national competence center. Our foremost task is to increase and disseminate knowledge about trends in the consumption of alcohol and other drugs, and related harm. We do this mainly by conducting research, by publishing articles and reports as well as by arranging courses and conferences. Our main national surveys are: Alcohol and Drug Use Among Students; the Monitor Study; and Habits and Consequences.

CAN is part of Swedish civil society and includes 50 member organizations. The members of our board are appointed by the Swedish Research Council for Health, Working Life and Welfare (Forte), the Swedish Research Council, the Public Health Agency of Sweden, the National Board of Health and Welfare and at the CAN's annual meeting. The Swedish Government appoints the board's chairman and deputy chairman.

The main purpose of the Monitor Study is to calculate the total alcohol consumption in Sweden. This is done by adding the amount of alcohol procured from unregistered sources to the amount shown in published data on registered sales. The amount of alcohol that comes from unregistered sources is captured by means of a continuous survey among Sweden's inhabitants. Alcohol from these sources is measured at the acquisition level and not at the consumption level. Measuring acquisition from unregistered sources requires that we make certain assumptions and do some calculations. These procedures are described in detail, and in plain language, in our published reports in Swedish. However, in the present report the methodology is described in technical terms and with statistical/mathematical notation. The parameters estimated in the study are formalized, as is the sampling design. In addition, the method used to compensate for non-response is described.

In 2020 a new weighting procedure was introduced in the study. In addition, the respondents were given the opportunity to answer the questions in the survey via a self-administered questionnaire in parallel to interviews by telephone. In this updated version of the report, the new weighting procedure is described as well as the way in which the estimates from 2019 are linked to the estimates based on the new methodology in 2020.

The report was written by Mats Nyfjäll at Statisticon in Uppsala, with support from Björn Trolldal at CAN. Gösta Forsman at Statisticon reviewed the report. The part of the Monitor study that covers alcohol is funded by Systembolaget.

Stockholm February, 2022

Charlotta Rehnman Wigstad, Director, CAN

Index

1 BACKGROUND

The Swedish Council for Information on Alcohol and Other Drugs (CAN) is a non-governmental organization. CAN's main tasks are to follow the alcohol and drug trends in Sweden and to inform the general public and educate professionals on alcohol and other drugs. This is e.g. done by publishing national reports and performing surveys. One such important survey is the Monitor survey.

The main purpose of the Monitor survey is to calculate the total quantity of alcohol and tobacco consumed, or more precisely acquired, in Sweden. The acquisition of alcohol can be divided into (i) registered acquisition and (ii) unregistered acquisition. The registered acquisition is alcohol bought from Systembolaget (The retail sales monopoly in Sweden), restaurants and grocery stores[1]. Quantities for registered acquisition is available from registers whereas unregistered acquisition is not.

The unregistered alcohol acquisition consists primarily of travelers' import, but also of purchases of alcohol that has been smuggled into the country, home production and purchases via the internet (from other sources than Systembolaget). For the acquisition of tobacco, the travelers' import also constitutes the largest unregistered source.

In order to calculate the quantities of the unregistered parts of acquisition, interviews with a random sample of people aged 17-84 are carried out on an ongoing basis (via telephone or since 2020 also via a self-administered digital questionnaire).

Over a year, more than 18 000 respondents participate in the Monitor survey. In addition to questions about the travelers' import of alcohol and tobacco, questions about of consumption patterns are also included.

The Monitor survey started in mid-2000 and CAN has been responsible for the study since 2013. Systembolaget (alcohol) and the Ministry of Health and Social Affairs (tobacco) are financing the survey.

Several reports describe results from the survey, see e.g. Trolldal (2020). Appendix 2 in that report describes the methodology in detail. However, the description has no statistical/mathematical notation. The present report is a step towards filling that gap. By describing the methodology in technical terms, procedures and methods can become more transparent, at least for those who are familiar with reading mathematical text.

The present report only deals with the acquisition of alcohol, not tobacco. Another delimitation is that only the technical aspects of the Monitor survey are treated, not conceptual parts. For example, in this report we do not repeat the four starting points in appendix 2 in Trolldal (2020). The *purpose of this report* is thus to *describe* the estimation procedure in technical terms. We will formalize the parameter that the survey is

[1] Only low alcohol beer at grocery stores (2.8 % to 3.5 % alcohol by volume).

estimating as well as the sampling design and estimator(s). Moreover, the report will also describe the method used to compensate for non-response. However, we will not discuss uncertainty due to frame imperfections and measurement errors. This report is an updated version of an earlier technical report published in 2019, see Nyfjäll & Trolldal (2019).

In 2020 a new weighting procedure was introduced in the study. In addition, the respondents were given the opportunity to answer the questions in the survey via a self-administered questionnaire, in parallel to interviews made by telephone. In this version of the report, the new weighting procedure is described as well as the way in which the estimates from 2019 are linked to the estimates based on the new methodology in 2020.

The disposition of the report is that we start by formalizing the parameters in chapter 2. In chapter 3, we describe the sampling design, and in chapter 4, the estimators are given and the new weighting procedure is described. The purpose of chapter 5 in to help the reader connect estimates presented in Trolldal (2020) with the estimators in chapter 4. In chapter 6, there is a description of the way in which the estimates from 2019 are linked to the estimates based on the new methodology in 2020.

2 PARAMETERS

The acquisition of alcohol comes from two sources: (i) registered and (ii) unregistered acquisition. In Sweden, the registered acquisition is mainly done at Systembolaget. In addition, the alcohol sold at restaurants and in grocery stores contributes to the registered acquisition. The unregistered acquisition has four main sources:

- Travelers' import
- Purchases of smuggled alcohol
- Internet acquisition (from other sources than Systembolaget)
- Home production

Based on the Monitor survey, the total quantity of unregistered acquisition of alcohol in the country is estimated. The registered acquisition is obtained from register data, mainly from Systembolaget and the Public Health Agency of Sweden (Folkhälsomyndigheten).

To describe the parameters that the Monitor survey is estimating, we start by introducing some notation. Additional notation will be introduced later. The notation is somewhat extensive. To facilitate a quick reference, we summarized the notation in appendix 1.

2.1 Basic notation

The target population for the Monitor survey is all registered Swedish citizens in aged 17 to 84. The reference period for the survey is calendar year, e.g. 2020. Let $U = (1,2,\dots,k,\dots,N)$ denote the target population where N is the population size of individuals and k is a running index for individuals in the population[2]. Since each individual k can acquire alcohol several times during the reference period (a year), an individual can be seen as a cluster where the cluster size is the number of times an acquisition is made. Let N_k denote the number of acquisitions that individual k does during the reference period. The population of acquisitions is thus of size $\sum_{k\in U} N_k$.

We introduce variables y and z that are associated with unregistered acquisition of alcohol. Let z be associated with each separate acquisition of unregistered acquisition that individual k does. That means that summing each acquisition for individual k over the reference period gives the total acquisition done by individual k, that is[3]

$$y_k = \sum_{\iota=1}^{N_k} z_\iota \qquad (1)$$

[2] During a whole year, the population size varies every day due to deaths, births and migration. We do not take that complicated aspect into account in the notation.
[3] The running index is Greek letter iota ι.

Correspondingly, we introduce variable x associated with registered acquisition of alcohol. Let z' be associated with each separate acquisition of registered acquisition that individual k does. Hence,

$$x_k = \sum_{\iota=1}^{N_k} z'_\iota \qquad (2)$$

is the total acquisition of registered alcohol for individual k over the reference period[4].

The construction of y_k and x_k might seem unconventional, but the purpose is to avoid (the somewhat burdensome) cluster notation, so that the parameters (and estimators) can be expressed in terms of $y{:}\,s$ and $x{:}\,s$.

The notation y and x follows the survey sample notation tradition that y is a study variable and x is an auxiliary variable used (in this case) in the estimation.

Both y and x are being defined in a conceptual manner of unregistered and registered acquisition of alcohol. In appendix 2 a more operational definition is presented.

Since there are different types of alcohol and different acquisition modes, we need subindexes to keep track. We use subindex i for type of alcohol and j for mode of acquisition, see appendix 1 for an explanation of the indexes. Hence, y_{ijk} denotes the (total) unregistered acquisition of type of alcohol i from acquisition mode j for individual k during the reference period. This variable is expressed in volume of liters, not in pure alcohol. Correspondingly, x_{ijk} is interpreted as the (total) registered acquisition of type of alcohol i from acquisition mode j for individual k during a year.

Please note that unregistered alcohol is associated with subindex $j = 1,2,3,4$ and registered alcohol is associated with subindex $j = 5,6,7$. See appendix 1 for an overview of the notation.

2.2 Parameters

Now the parameters can be defined. We follow the notation used in Särndal et al. (1992). Summing all y_{ijk} in the population U, i.e.

$$t_{y_{ij}U} = \sum\nolimits_{k \in U} y_{ijk} \qquad (3)$$

gives the parameter *total* quantity of unregistered acquired alcohol for type of alcohol i and acquisition mode j. Correspondingly,

[4] Please note that N_k in (1) and (2) usually differs.

$$t_{x_{ijU}} = \sum_{k \in U} x_{ijk} \quad (4)$$

is the parameter *total* quantity of registered acquired alcohol for alcohol type i and acquisition mode j. Both $t_{y_{ijU}}$ and $t_{x_{ijU}}$ are expressed in volume liters of alcohol, i.e. not in pure alcohol.

Another important parameter is the quantity of acquired pure alcohol expressed as per capita. Let α_{ij} denote the (average) alcohol content, i.e. the percentage of pure alcohol by weight, for type of alcohol i and acquisition mode j. For the majority of acquisition modes, the alcohol content according to Systembolaget ($j = 5$) is used. For example, for spirits ($i = 4$) the alcohol content is $\alpha_{i=4,j=5} = 0{,}373$ according to Systembolaget. See table 13 in appendix 1 for the (average) alcohol content[5]. Then

$$Q_{ij}^{UNREG} = t_{y_{ijU}} \times \alpha_{ij} \quad (5)$$

is the *total* quantity of unregistered acquired pure alcohol for type of alcohol i and acquisition mode j. Please note that $\alpha_{i,j=5}$, i.e. the alcohol content according to Systembolaget, is used for all acquisition modes except restaurants ($j = 6$) and grocery stores ($j = 7$).

Dividing (5) by the population size gives the per capita measure. However, the division is not done by N, the population size of individuals 17-84 years old, but rather by the number of individuals aged 15 and older, which is a national and international standard procedure. Denote this number by N_{15+}. The per capita parameter of unregistered acquisition of pure alcohol (for type of alcohol i and acquisition mode j) is given by

$$\frac{Q_{ij}^{UNREG}}{N_{15+}} = \frac{t_{y_{ijU}} \times \alpha_{ij}}{N_{15+}} \quad (6)$$

Correspondingly, the parameter of registered acquisition of pure alcohol is given by

$$\frac{Q_{ij}^{REG}}{N_{15+}} = \frac{t_{x_{ijU}} \times \alpha_{ij}}{N_{15+}} \quad (7)$$

[5] Year-specific values for alcohol content are used. The difference between the years is small.

3 SAMPLING DESIGN

The sampling frame in the survey is PAR Konsument, which is based on the (larger) frame named SPAR, which in turn is a complete (highly accurate) register of all Swedish citizens. The survey is done by telephone interviews and via a self-administered questionnaire sent to the respondents' mobile telephones, so a sampling frame that contains telephone numbers is necessary. Now, SPAR does not contain telephone numbers, but PAR Konsument does. Trolldal (2020) indicates that PAR Konsument contains approximately 70 percent of all individuals in the population, so there is undercoverage in the frame.

We first describe the sampling design in words, then in a more formal way. Each month, a large sample is drawn from the frame. The sample is stratified by gender and age groups and the sample size is one million individuals. A simple random sample[6] is drawn within each stratum. The allocation is proportional, which means that the sample is self-weighted, i.e. all individuals have the same inclusion probability. From this large base sample, a simple random sample is drawn every week[7]. Each sample is in use for four weeks, which means that four different samples are in use in parallel. During a year, 52 different week samples are drawn from 12 different base samples. The 52-week samples are not coordinated, so a selected individual can be selected twice. If this occurs within the same base sample, the duplicate is removed from the sample. This is the case even if duplicates occur in different base samples during a calendar year.

An aspect regarding the data collection might be mentioned. For a given week-sample, the sampled individuals are contacted by telephone during a four-week period at most (if they do not accept the opportunity to answer the questions via the self-administered questionnaire). For example, an individual in the week-3-sample is contacted during week 3 to 6 (at most). If no reply is obtained during that period, no further contact is made, and the individual is classified as a non-respondent. If contact is obtained and an interview is done, the respondent is asked about acquisition during the last 30 days. The date that the interview is made determines to which period (see below) an individual is allocated in the estimation process.

The sampling design with a base sample from the frame and then weekly samples from the base sample is in fact a *two-phase* sampling procedure (with stratification by gender and age in the first phase). In the second phase, independent weekly samples of individuals are drawn by a simple random sampling procedure from the first phase base sample. The estimation is done on a monthly basis and then estimates are added together to form a yearly estimate. This means that from a randomization-based

[6] Without replacement.
[7] If an individual is selected week 1 and then selected again week 2, the individual is removed from the sample. Hence, we can say that the sampling procedure is without replacement.

perspective, the design, as well as the estimator, should formally encompass both the first and second phase. However, since the first phase is proportionally allocated, this sample can be regarded (from a practical viewpoint) as a simple random sample from the frame[8]. Moreover, together, the weekly samples from the base sample can (for practical purposes) also be regarded as a simple random sample from the base sample[9]. Hence, in the technical description below, we regard the weekly samples during a month from one base sample as one simple random sample from the frame.

In general terms, let $s = \{1,2, \dots, k, \dots, n\}$ denote the (whole) sample over a year and n its size. The base sample is drawn each month. Let $p = 1,2, \dots, P$ be an index for *period* (month). There are $P = 12$ periods (base samples) during a year. Since the base sampling procedure is repeated each month, the monthly P samples can be regarded as stratified samples from the population. With an index for period the sample is s_p and the size n_p and the corresponding population notation is U_p with size N_p.

The sample s_p is considered a simple random sample from the population $U_p = \{1, \dots, k, \dots, N_p\}$, rather than from the frame (PAR Konsument), which is approximately of size $0.7 \times N_p$.

Note that the population size in period p is approximately equal to the population size over the whole year, i.e. $N_p \approx N$.

Summing the P samples give $s = \cup_{p=1}^{P} s_p$ and $n = \sum_{p=1}^{P} n_p$.

Non-response occurs. Let $r_p = \{1,2, \dots, k, \dots, m_p\}$ denote the response set and m_p its size[10] in period p. Correspondingly, summing the P samples give $r = \cup_{p=1}^{P} r_p$ and $m = \sum_{p=1}^{P} m_p$. During a typical year m is approximately 18 000 individuals.

[8] The (first order) inclusion probability in a proportionally allocated stratified simple random sample is approximately equal to the inclusion probability in simple random sampling.
[9] Drawing four independent simple random samples from a population and joining them together is formally not equivalent to one simple random sample (four times as large as the individual samples) from the population, but the difference is for practical purposes negligible.
[10] It is the date that an individual responds to the survey that determines which period he or she is allocated to.

4 ESTIMATION

4.1 Calibration and weighting

To account for non-response, a calibration procedure is used. Earlier a post-stratification procedure was used, see Nyfjäll and Trolldal (2019), but in 2020 the weighting procedure was changed into a calibration approach. The new weighting procedure is used for 2020 and onward but estimates for 2019 have also been revised based on the new calibrated weight.

Briefly we can mention that, earlier, the post-stratification was done by cross classification of age (4 groups), gender (2 groups) and regions (with 3 groups). This means $4 \times 2 \times 3 = 24$ cells used for calculating weights[11]. Post-stratification is in fact a special case of calibration, so by calibration we here mean that the auxiliary information is used more extensively in a way that goes beyond post-stratification. The reason for changing the weighting procedure was twofold:

1. One of the poststratification variables, H-region, which is geographic regions in Sweden, is outdated. Newer and better regional divisions exist nowadays.
2. Other register variables, besides gender, age and region, are now available in the weighting procedure.

The change in weighting system was anticipated by several analyses, for example a nonresponse analysis based on data from 2019 as well as a study carried out by Källebring (2021). These analyses revealed discrepancies between the population distribution and distribution in the response set regarding the following variables:

- Age groups: younger persons are underrepresented, and older persons are overrepresented in the response set compared to the population.
- Geographic regions[12]: persons living in larger cities are underrepresented.
- Education: persons with low education are underrepresented.
- Foreign background: persons born outside Sweden are underrepresented.

The variable gender was also analyzed, but the response set consists of approximately 50 percent men and women respectively, which agrees with the population distribution. The variables listed above were the potential candidates to use in a calibration procedure. Different numbers of categories were tried for each variable. Variables were

[11] In each cell, the weight was calculation by the ratio of population number to number of responding individuals (i.e. “straight expansion” within post-strata).
[12] The geographical variable is the geographic division according to SKR (Sveriges Kommuner och Regioner).

also cross classified in the analysis. The final choice of variables to be used in the calibration weighting procedure each month (period p) was:

- Age groups (4) cross classified with gender (2) into $4 \times 2 = 8$ groups.
- Geographic regions into 3 groups.
- Education into 2 groups.
- Foreign background into 2 groups.

Table 1 shows the number of individuals in each age group cross classified with gender in the population 2019. Table 2, 3 and 4 show the number of individuals in the population divided into: (i) main regional divisions according to SKR, (ii) education level and (iii) foreign background, respectively.

Table 1. Number of individuals in the population 2019 divided into age groups and gender

		Individuals in the population	
Age	Gender	Number	Percent
17–29	Man	866 797	10,8
	Woman	797 040	10,0
30–49	Man	1 361 226	17,0
	Woman	1 303 618	16,3
50–64	Man	942 363	11,8
	Woman	923 890	11,6
65–84	Man	867 331	10,8
	Woman	934 780	11,7
	Sum	7 997 045	100

Table 2. Number of individuals in the population 2019 divided into geographic regions[13]

	Individuals in the population	
Geographic region	Number	Percent
A - Big cities and municipalities close to big cities	2 951 611	36,9
B - Larger cities and municipalities near larger cities	3 051 308	38,2
C - Smaller cities / urban areas and rural municipalities	1 994 126	24,9
Sum	7 997 045	100

Table 3. Number of individuals in the population 2019 divided into education[14] groups

	Individuals in the population	
Education	Number	Percent
University	2 472 585	30,9
Not so	5 524 460	69,1
Sum	7 997 045	100

[13] The geographic division is according to municipality division (“kommungruppsindelning”) by SKR, see https://skr.se/skr/tjanster/kommunerochregioner/faktakommunerochregioner/kommungruppsindelning.2051.html

[14] University means degree from either university or “högskola” (approximately same as college).

Table 4. Number of individuals in the population 2019 divided into foreign background

	Individuals in the population	
Foreign background	Number	Percent
Born in SE and at least one parent born in SE	5 885 218	73,6
Not so	2 111 827	26,4
Sum	7 997 045	100

Below, the calibration weighting is described in a more technical way. For a reference, see for example Särndal and Lundström (2005). In summary, the vector used for calibration each month can be described according to

- **Age group × gender (8 groups) + geography (3 groups) + education (2 groups) + foreign background (2 groups)**

The multiplication sign between age group and gender indicates that the two variables are cross classified. The plus sign between the other variables indicates that the calibration is done according to the marginal distribution of each variable. Hence, there are $8 + 3 + 2 + 2 = 15$ totals in the calibration.

Let $\mathbf{x}_k$ denote the auxiliary vector that represents which combination of the categories individual k belongs to. The calibrated weight w_k is then created each period p according to

$$w_k = d_k \times v_k \tag{8}$$

where

$$d_k = \frac{N_p}{m_p} \tag{9}$$

and the adjusting weight v_k is given by

$$v_k = 1 + \left(\sum_{k \in U_p} \mathbf{x}_k - \sum_{k \in r_p} d_k \mathbf{x}_k\right)' \left(\sum_{k \in r_p} \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \mathbf{x}_k \tag{10}$$

The calibrated weight w_k has property

$$\sum_{k \in r_p} w_k \mathbf{x}_k = \sum_{k \in r_p} d_k v_k \mathbf{x}_k = \sum_{k \in U_p} \mathbf{x}_k \tag{11}$$

Summing the calibrated weight w_k equals the population size each period p, that is

$$\sum_{k \in r_p} w_k = N_p \tag{12}$$

From table 1 to 4 we see that $N_p = 7\,997\,045$ in 2019.

Remark: The number of individuals in tables 1 to 4 is used in the calibration each period p (i.e. month). The number of individuals varies each day over the year (and hence each

period p), but the register information is only available on a yearly basis, so all N_p are equal for each period p within a year.

Remark: w_k in (8) has the property that, when summing all of them within a period p (a month), they sum to the population total N_p, which is stated in (12). CAN prefer the weights scaled so that they sum to the number of respondents m_p within a period. Therefore, a scaled calibrated weight is calculated according to

$$w'_k = \frac{m_p}{N_p} \times w_k \tag{13}$$

The scaled calibrated weight sum within a period p to the number of respondents, i.e.

$$\sum_{k \in r_p} w'_k = m_p \tag{14}$$

which should be apparent. Moreover, the mean of w'_k within a period p equals unity.

4.2 Point estimators

In expression (5) $t_{y_{ijU}}$ is the total quantity of unregistered alcohol in volume liters for type of alcohol i and acquisition mode j. Estimating this total is central in the estimation of the parameter Q_{ij}^{UNREG} in (5) and in the per capita parameter in (6). We start by summarizing the form of the estimator for $t_{y_{ijU}}$. Then, in section 4.2.1 below, the estimator $\hat{t}_{y_{ijU}}$ is explicitly stated.

To summarize, the estimator for $t_{y_{ijU}}$ can be written as

$$\hat{t}_{y_{ijU}} \times R_{ij} \tag{15}$$

where $\hat{t}_{y_{ijU}}$ is an estimator using weights w_k in (8) and R_{ij} is (an inflating) ratio[15] that accounts for several aspects. For example:

- It is well known, see Trolldal (2020), that self-reported acquisition of alcohol (during the last month) tends to be under-estimated. This is true for acquisition of both unregistered and registered alcohol. Since the quantity of registered alcohol can be obtained from reliable register data (mainly Systembolaget) and the survey asks questions regarding acquisition of unregistered as well as alcohol bought at Systembolaget, the amount of under-reporting can be estimated. This is encapsulated into R_{ij}, as will be shown later, in a sort of ratio type of estimation.

[15] We could have used a "hat" over R_{ij}, i.e. $\hat{R}_{ij}$, but since the ratio does not estimate anything, we omitted it, not to clutter the notation unnecessarily.

- Earlier studies have shown that consumers of large amounts of alcohol are under-represented in the response set. Hence, an extra effort is made to account for this under-representation in the estimation process. This aspect is also encapsulated into R_{ij}.

When $t_{y_{ijU}}$ is estimated, the parameter of quantity of pure alcohol Q_{ij}^{UNREG} in (5) and in the per capita parameter in (6) can be estimated as

$$\hat{Q}_{ij}^{UNREG} = \hat{t}_{y_{ijU}} \times R_{ij} \times \alpha_{ij} \tag{16}$$

and

$$\frac{\hat{Q}_{ij}^{UNREG}}{N_{15+}} = \frac{\hat{t}_{y_{ijU}} \times R_{ij} \times \alpha_{ij}}{N_{15+}} \tag{17}$$

Below, we explicitly describe the construction of $\hat{t}_{y_{ijU}}$ and R_{ij}.

4.2.1 The estimator $\hat{t}_{y_{ij}U}$

As a first step in the estimation procedure, a mean per period (month) p is calculated according to[16]

$$\hat{\bar{y}}_{ijU_p} = \frac{\sum_{k \in r_p} w_k y_{ijk}}{\sum_{k \in r_p} w_k} \tag{18}$$

A couple of aspects are worth mentioning:

- In this expression y_{ijk} is the total acquisition of type of alcohol i by mode of acquisition j for individual k during period p, i.e. not the whole year as was stated earlier in section 2.1. That means that the definition of y_k in (1), summing all individual acquisitions for individual k during a year now means summing all individual acquisitions for individual k during a month. However, the questions asked in the survey are only about the volume of a beverage at the *last* acquisition. This is then multiplied by the number of acquisitions (for each mode j)[17]. In appendix 2 the operational construction of y_{ijk} is described in more detail, see e.g. expression (60) regarding travelers' import. One important aspect of this

[16] Please note that the weight w_k according to (8) and w_k' according to (13) will give identical results. The factor m_p/N_p in the nominator and denominator will cancel out.
[17] To ask for only the last acquisition makes the questionnaire less burdensome.

procedure is that the total acquisition for individual k during period p is *estimated* rather than registered[18].

- The estimator $\hat{\bar{y}}_{ijU_p}$ can be seen as an estimator of the parameter $\bar{y}_{ijU_p} = (1/N_p)\sum_{k\in U_p} y_{ijk}$ the true mean of the acquisition of type of alcohol i for acquisition mode j (expressed in volume liters[19]). The estimates obtained by (18) are not of any particular interest other than as an intermediate calculation step and are not presented in any of CAN:s publications.

In an additional intermediate calculation, the means $\hat{\bar{y}}_{ijU_p}$ in (18) are multiplied by $N \times 0.01$. The multiplication by 0.01 converts the unit from centiliters to liters.

Next, the values are summed over all P periods giving

$$\hat{t}_{y_{ijU}} = \sum_{p=1}^{P} \hat{\bar{y}}_{ijU_p} \times N \times 0.01 \qquad (19)$$

Note that although this is an estimator, it is not the estimator of $t_{y_{ijU}}$ in (3), since R_{ij} in (15) needs to be multiplied with $\hat{t}_{y_{ijU}}$ to form the (used) estimator.

4.2.2 The estimator $\hat{t}_{x_{ij}U}$

In the survey, the respondents are asked about their acquisition of registered alcohol at Systembolaget too. Variable x, instead of y, is used to indicate acquisition of all registered alcohol. In section 2.1 we defined x_{ijk} as the total acquisition of registered alcohol for individual k over the reference period (for alcohol type i and mode of acquisition j). Since the period in the survey is a month, x_{ijk} is the total quantity acquired per month. This is equivalent to the construction of variable y described in section 4.2.1, see also appendix 2 and expression (64). Since the respondents are only asked about their acquisition of registered alcohol at Systembolaget ($j = 5$), the appropriate variable is $x_{i,j=5,k}$.

The estimator $\hat{t}_{x_{i,j=5}U}$ is formed identically as $\hat{t}_{y_{ijU}}$ in (19), i.e.

$$\hat{t}_{x_{i,j=5}U} = \sum_{p=1}^{P} \hat{\bar{x}}_{ijU_p} \times N \times 0.01 \qquad (20)$$

where $\hat{\bar{x}}_{ijU_p}$ mimics the expression (18), but with x instead of y.

---

[18] It might have been more appropriate to account for this estimation by the notation $\hat{y}_{ijk}$ instead of y_{ijk}. However, this is not done in order keep the notation as simple as possible.

[19] At this point in the calculation process, the unit is centiliters rather than liters. That does not affect the principle.

Remark 1: $\hat{t}_{x_{i,j=5}U}$ in (20) is not used to estimate $t_{x_{i,j=5}U}$, the true quantity of registered alcohol at Systembolaget (for type of alcohol i). Rather, $t_{x_{i,j=5}U}$ is obtained from reliable transaction registers at Systembolaget. $\hat{t}_{x_{ij}U}$ is used to form a ratio type of estimator in R_{ij} in (15).

Remark 2: Acquisitions from restaurants ($j = 6$) and grocery stores ($j = 7$) are not estimated based on the Monitor survey. The true parameter values for $t_{x_{ij}U}$ for $j = 6,7$ are obtained from external sources[20], see section 4.2.5.

4.2.3 The construction of the ratio R_{ij}

The construction of R_{ij} in (15) is somewhat complex and is done in a slightly different way depending on type of alcohol i and acquisition mode j. In words the procedure can be described as:

1. First heavy consumers[21] are empirically identified based on responses to specific questions about consumption of alcohol in the survey, and estimates for this domain are done.
2. Secondly, a preliminary adjustment ratio is calculated. This preliminary adjustment (ratio) is calculated in two versions.
3. Thirdly, the preliminary adjustment ratio is "fine-tuned" leading to a final R_{ij}

Below, we describe the procedure chronologically.

4.2.3.1 Definition of heavy consumers, and estimator

The definition of a heavy consumer is data driven, i.e. depends on collected data. The quantity of pure alcohol consumed during the last month (period p) is estimated based on answers to questions about alcohol consumption in the survey. The answers to the questions are not in terms of pure alcohol but the units of each beverage, which are converted into pure alcohol and summed over all types of alcohol. Let c_{100k} denote the consumed quantity of 100 percent pure alcohol for individual k in period p. Let $Percentile_{99}$ denote the 99:th empirical (unweighted) percentile in the response set r_p. We introduce an indicator for the heavy consumers according to

$$Ind_k = \begin{cases} 1 & \text{if } c_{100k} \geq Percentile_{99} \\ 0 & \text{if } c_{100k} < Percentile_{99} \end{cases} \quad (21)$$

Since the threshold $Percentile_{99}$ is data driven it varies between periods.

---

[20] Note that for grocery stores ($j = 7$) only alcohol type $i = 7$, low alcohol beer, applies.
[21] The identification of a heavy consumer is based on specific question about *consumption* of alcohol, not *acquisition* patterns.

The next step is to apply a mean estimator similar to (18), but only for the heavy consumers and summing over all periods, to get a mean value for the whole year. This is equivalent to a domain estimator where the domain[22] is defined by the indicator variable (21). For domain estimation, it is convenient to introduce a domain variable y_{dijk} defined as

$$y_{dijk} = \begin{cases} y_{ijk} & \text{if } Ind_k = 1 \\ 0 & \text{if } Ind_k = 0 \end{cases} \quad (22)$$

In words, variable y_{dijk} takes the value of y_{ijk} within the domain (i.e. the heavy consumers) and zero outside the domain. The domain variable is used in the estimator

$$\hat{\bar{y}}_{ijU_d} = \sum_{p=1}^{P} \hat{\bar{y}}_{ijU_{dp}} \quad (23)$$

where $\hat{\bar{y}}_{ijU_{dp}}$ is the same as (18) but with y_{ijk} replaced by y_{dijk}. The estimator $\hat{\bar{y}}_{ijU_d}$ gives an estimate of *average* quantity of acquired unregistered alcohol during a year (for type of alcohol i and acquisition mode j) in domain d.

If variable y_{dijk} in (23) is replaced by y_{ijk}, i.e. the ordinary $y-$ variable without a domain indicator, we obtain

$$\hat{\bar{y}}_{ijU} = \sum_{p=1}^{P} \hat{\bar{y}}_{ijU_p} \quad (24)$$

It is apparent that the only difference between (23) and (24) is what variable, y_{dijk} or y_{ijk}, is plugged into the estimator.

Estimators (23) and (24) are used to calculate the preliminary adjustment ratio. This is described in the next section.

4.2.3.2 Preliminary adjustment ratio

The adjustment ratio is *principally* of the form[23]

$$\frac{\text{True value of acquisition from Systembolaget}}{\text{Estimated value of acquisition from Systembolaget based on Monitor survey}} \quad (25)$$

[22] A domain (or domain of study) is a subpopulation of interest e.g. “men” and “women” as two subpopulations of “all persons”. To belong to the domain “men” or “women” does not depend on data, but is fixed in advance, as well as the size of the domain. To belong to the domain “heavy consumers” does depend on data. Hence, the domain heavy consumers is not a domain in a traditional meaning. However, it is convenient to borrow notation from domain theory to describe the procedure.

[23] The denominator is thus based on question posed in the Monitor survey.

Multiplying this ratio with the estimator $\hat{t}_{y_{ijU}}$ in (19) gives (in spirit) the well-known ratio estimator, see for example Särndal et al. (1992). However, the final (fine-tuned) version of the ratio is not calculated according to the schoolbook. Let us look at the steps involved in the process.

Version 1 – <u>no</u> heavy consumer adjustment

Since the ratio (25) is based on the acquisition from Systembolaget, we use estimator (20) with subindex $j = 5$ for the denominator

$$\hat{t}_{x_{i,j=5}U} = \sum_{p=1}^{P} \hat{\bar{x}}_{ijU_p} \times N \times 0.01 \tag{26}$$

This is an estimator based on questions asked in the survey about registered acquisition of alcohol (for type of alcohol i and acquisition mode from Systembolaget, $j = 5$). The true value of registered acquisition of alcohol from Systembolaget in the numerator of (25) is $t_{x_{i,j=5}U}$. However, the true value $t_{x_{i,j=5}U}$ is adjusted to account for the fact that many Norwegians acquire alcohol at Systembolaget, especially close to the border in the northern part of Swedens's west coast. Hence, the true value is adjusted down with 5 percent; $t_{x_{i,j=5}U} \times 0.95$.

Version 1 of the adjustment ratio is then given by

$$v_{i,j=5,l=1} = \frac{t_{x_{i,j=5}U} \times 0.95}{\hat{t}_{x_{i,j=5}U}} \tag{27}$$

The new subindex $l = 1$ indicates that this is the first version of the preliminary adjustment ratio. Note also that subindex $j = 5$ indicates that this concerns acquisition at Systembolaget only. Version 1 does not take the heavy consumers into any special consideration, which version 2 does.

Version 2 –heavy consumer adjustment

The adjustment for heavy consumers involves a couple of steps. The idea behind the steps is to account for the fact that heavy consumers are believed to be under-represented in the response set.

First, from the expression (23) and (24), calculate the ratio

$$prop_{di,=5} = \frac{\hat{\bar{y}}_{i,j=5,U_d}}{\hat{\bar{y}}_{i,j=5,U} \times 100} \tag{28}$$

This is an estimator of the proportion[24] of acquisition for heavy consumers (the subindex d indicates domain) in relation to the total population at Systembolaget $j = 5$ (for type of alcohol i). The rationale behind this interpretation is provided in appendix 3.

For example, regarding spirits ($i = 4$) the estimates for 2019 are

$$prop_{d,i=4,j=5} = \frac{\hat{\bar{y}}_{i=4,j=5,U_d}}{\hat{\bar{y}}_{i=4,j=5,U} \times 100} = \frac{1413.7}{153.6 \times 100} \approx 0.092 \quad (29)$$

The numerator indicates that the heavy consumers on average acquired 1413.7 centiliters of spirits from Systembolaget during 2019. The denominator indicates that the whole population on average acquired 153.6 centiliters of spirits from Systembolaget during 2019. The ratio indicates that the heavy consumers stand for approximately 9.2 percent of all acquisition of spirits from Systembolaget.

Remark: Both the numerator and the denominators underestimate their true population counterparts. It is believed that the underestimation is roughly equal in both groups, so the ratio is believed to be more accurate.

Secondly, multiply the estimated proportion with the estimator for acquisition from Systembolaget (20) which gives

$$\hat{t}_{x_{i,j=5,U_d}} = \hat{t}_{x_{i,j=5,U}} \times prop_{di,j=5} \quad (30)$$

i.e. an estimator of the total quantity in volume liters of acquired alcohol from Systembolaget by heavy consumers (for alcohol type i).

This estimator does not take into consideration that the heavy consumers are underrepresented in the response set. The estimator (30) is based on the definition that the heavy consumers are the consumers with the one percent largest consumption of alcohol. The estimator (30) can be interpreted as the quantity in volume liters of acquired alcohol from Systembolaget by heavy consumers (for alcohol type i) if they constitute *one percent* of the population. Based on results from earlier studies (Kühlhorn et al., 1999), this group of individuals are believed to be underrepresented in the response set. Therefore, the estimator is multiplied with a factor 3, i.e. $\hat{t}_{x_{i,j=5,U_d}} \times 3$.

At the same time the respondents that are not considered heavy consumers are "weighted down" with a factor 97/99 according to

[24] The notation $prop$ is used for proportion, since the more common (and short) notation p already is used to denote period.

$$\hat{t}_{x_{i,j=5,}U_{d=not\ HC}} \times \frac{97}{99} \tag{31}$$

Please note the notation $U_{d=not\ HC}$ which indicates the domain "not" heavy consumers (HC), that is the complement to the domain heavy consumers.

Thirdly, combining these gives the second version ($l = 2$) of the preliminary adjustment factor

$$v_{i,j=5,l=2} = \frac{t_{x_{i,j=5}U} \times 0.95}{\hat{t}_{x_{i,j=5,}U_{d=HC}} \times 3 + \hat{t}_{x_{i,j=5,}U_{d=not\ HC}} \times \frac{97}{99}} \tag{32}$$

Please note the resemblance between (27) and (32). The numerator is the same, but the denominators differ. The denominator is a sum of two domain estimators; heavy consumers ($d = HC$) and not heavy consumers ($d = not\ HC$) which combined constitutes the whole population. But multiplying the estimators with factors 3 and $\frac{97}{99}$ gives a sum in the denominator that is larger than the denominator in (27). In (32) the heavy consumers are given a larger weight. Since the denominator in version $l = 2$ in (32) is larger than the denominator in version $l = 1$ in (27), $v_{i,j=5,l=2}$ will be smaller than $v_{i,j=5,l=1}$.

Numerical examples for version 1 and 2

It might be helpful for interpretation with some numerical examples. Continuing with spirits ($i = 4$) and 2019, the version 1 preliminary adjustment ratio given by (27) is given by

$$v_{i=4,j=5,l=1} = \frac{t_{x_{i=4,j=5}U} \times 0.95}{\hat{t}_{x_{i=4,j=5}U}} = \frac{20\ 635\ 543 \times 0.95}{11\ 956\ 262} \approx 1.64 \tag{33}$$

This preliminary adjustment ratio can be interpreted. Regarding spirits ($i = 4$) there is a self-reported underestimation regarding the acquisition at Systembolaget approximately equal to $\frac{1}{1.64} \approx\ 0.61$, i.e., 61 percent, giving rise to a preliminary adjustment ratio that augments the estimate of acquired unregistered alcohol by a factor of 1.64.

Version 2 is given by first calculating (28), which is numerically given in (29). Then secondly by calculating (30)

$$\hat{t}_{x_{i=4,j=5,}U_d} = \hat{t}_{x_{i=4,j=5,}U} \times prop_{di,j=5} = 11\ 956\ 262 \times 0.092 \approx 1\ 099\ 976 \tag{34}$$

Thirdly, calculate

$$v_{i,j=5,l=2} = \frac{20\,635\,543 \times 0.95}{1\,099\,976 \times 3 + (11\,956\,262 - 1\,099\,976) \times \frac{97}{99}} \approx 1.41 \qquad (35)$$

We note that when the heavy consumers are given a larger weight, the version 2 of the preliminary adjustment ratio is 1.41, compared to 1.64 for version 1.

The preliminary adjustment ratio is calculated in two versions for all types of alcohol.

4.2.3.3 Fine-tuning of the preliminary adjustment ratio - final adjustment ratio

The fine-tuning of the preliminary adjustment ratio involves calculating moving averages. This is done to stabilize the ratio from temporary fluctuations. Regarding travelers' import of spirits, wine and beer as well as purchases of smuggled spirits and beer, the fine-tuning involves calculating moving averages as well as some other adjustments. For the other beverages and acquisition modes, the preliminary adjustment ratio is used as the final ratio and the fine-tuning only involves calculating moving averages. Table 5 provides a summary.

Acquisition mode travelers' import $(j = 1)$ **and wine, beer and spirits** $(i = 1,2,4)$

In (28) the proportion of acquisition of alcohol from Systembolaget ($j = 5$) for heavy consumers (domain d) was defined. Replacing the specific index $j = 5$ with a general j gives the proportion for other acquisition modes

$$prop_{dij} = \frac{\hat{\bar{y}}_{ijU_d}}{\hat{\bar{y}}_{ijU} \times 100} \qquad (36)$$

where the numerator is given by (23) and the denominator by (24). For example, $prop_{d,i=4,j=1} = \frac{307.6}{103.4 \times 100} \approx 0.030$, indicating that the heavy consumers stand for 3.0 percent of the of acquisition of spirits from travelers' import. This can be compared with their share of the acquisition of spirits from Systembolaget which was 9.2 percent, see expression (29).

A moving average of (36) is formed. We introduce subindex t for time[25] (year), hence

[25] Please note the distinction between subindex t for time and t_{yU} the parameter total quantity of unregistered acquired alcohol. The latter has a subindex which can vary; t_{yU}, t_{xU}, $t_{y_{ij}U}$ and can also contain a "hat" for estimation $\hat{t}_{y_{ij}U}$.

$$prop_{dij}^{MOV} = prop_{dijt}^{MOV} = \frac{p_{dijt} + p_{dij,t-1} + p_{dij,t-2}}{3} \quad (37)$$

is a moving average[26] over three years. A remark regarding the notation:

- $prop_{dijt}^{MOV}$ in the middle equality contains subindex t for time, which is natural and indicates that it is connected to time period t. In the leftmost expression in (37), i.e., $prop_{dij}^{MOV}$, the subindex t is omitted. There has not been a need for an index for time anywhere else in the report, it is only in connection to the moving averages. Since we do not want to burden the notation with a time-index in every notation, we omit the t in $prop_{dij}^{MOV}$. Thus, $prop_{dij}^{MOV}$ always means the latest year t. This applies for all moving averages in the report.

A similar moving average is formed for $prop_{di,j=5}$, the acquisition from Systembolaget. Since (37) encompasses j in general, we need not explicitly state the moving average for $prop_{di,j=5}$.

A moving average is also formed regarding the preliminary adjustment ratio $v_{i,j=5,l}$ in (27) and (32). Adding a subindex t for time gives

$$v_{i,j=5,l}^{MOV} = \frac{v_{i,j=5,lt} + v_{i,j=5,l,t-1} + v_{i,j=5,l,t-2}}{3} \quad (38)$$

Note that this moving average is only done for $j = 5$, i.e. the acquisition from Systembolaget.

Now, the final (fine-tuned) adjustment ratio can be stated as

$$R_{ij} = \frac{prop_{dij}^{MOV}}{prop_{di,j=5}^{MOV}} \times \left(v_{i,j=5,l=1}^{MOV} - v_{i,j=5,l=2}^{MOV}\right) + v_{i,j=5,l=2}^{MOV} \quad (39)$$

Note that this adjustment ratio applies for travelers' import ($j = 1$) and wine, beer and spirits ($i = 1,2,4$). Below, we give some comments on the interpretation and the rationale for the ratio.

A numerical example might be illustrative and help the interpretation. For spirits ($i = 4$) and acquisition mode ($j = 1$) and $t = 2019$ we have

$$prop_{d,i=4,j=1}^{MOV} = \frac{0.030 + 0.027 + 0.037}{3} \approx 0.031 \quad (40)$$

[26] Note that the moving average is not centered around the middle value.

This shows that in 2019 the heavy consumers acquired 3.0 percent ($i = 4$) of all travelers' import of spirits ($j = 1$). Corresponding estimates for 2018 and 2017 are 2.7 and 3.7 respectively giving a moving average of 3.1 percent.

The corresponding estimate for Systembolaget ($j = 5$) is

$$prop_{d,i=4,j=5}^{MOV} = \frac{0.092 + 0.080 + 0.090}{3} \approx 0.087 \quad (41)$$

It can be noted that heavy consumers acquired 9.2 percent ($i = 4$) of all acquisitions of spirits at Systembolaget ($j = 5$) in the year 2019, a much bigger percentage than the share of travelers' import (3.0 percent). The moving average is 8.7 percent.

The moving average for (38) version $l = 1$ is

$$v_{i,j=5,l=1}^{MOV} = \frac{1.64 + 1.59 + 1.55}{3} \approx 1.59 \quad (42)$$

whereas (38) for version $l = 2$ is

$$v_{i,j=5,l=2}^{MOV} = \frac{1.41 + 1.40 + 1.34}{3} \approx 1.38 \quad (43)$$

Plugging this into (39) gives

$$R_{i=4,j=1} = \frac{0.03}{0.09} \times (1.59 - 1.38) + 1.38 \approx 0.07 + 1.38 \approx 1.46 \quad (44)$$

Taking a closer look at (39), it turns out that in many cases it can be characterized as

$$R_{ij} = small\_number + v_{i,j=5,l=2}^{MOV} \quad (45)$$

where $small\_number$ is the first term in (39). Hence, R_{ij} depends mostly on $v_{i,j=5,l=2}^{MOV}$. But $small\_number$ has the function of adding a smaller or larger value to $v_{i,j=5,l=2}^{MOV}$. Thus, R_{ij} is "stretched" by a small or large amount depending on $prop_{dij}^{MOV}$, $prop_{di,j=5}^{MOV}$ and $v_{i,j=5,l=1}^{MOV}$. To understand how this stretching is done, we will look at some examples.

Example 1: if $prop_{dij}^{MOV} \approx prop_{di,j=5}^{MOV}$ then their ratio is 1. $prop_{dij}^{MOV}$ is the (moving average of the) proportion of acquired alcohol for heavy consumers compared to the general public (for alcohol type i and acquisition mode j). $p_{di,j=5}^{MOV}$ is the same proportion but regarding acquisition at Systembolaget. If the proportions are the same, the acquisition pattern for heavy consumers is the same when comparing acquisition mode j with Systembolaget. This gives

$$R_{ij} = 1 \times \left(v^{MOV}_{i,j=5,l=1} - v^{MOV}_{i,j=5,l=2}\right) + v^{MOV}_{i,j=5,l=2} = v^{MOV}_{i,j=5,l=1} \quad (46)$$

i.e. version $l = 1$ of the preliminary adjustment factor with no special account for heavy consumers. This is reasonable, because if $prop^{MOV}_{dij} \approx prop^{MOV}_{di,j=5}$ we do not need to do any special adjustment for heavy consumers, which is what $v^{MOV}_{i,j=5,l=1}$ does (see e.g. expression (27), which is the building block in the moving average).

Example 2: if $prop^{MOV}_{dij} < prop^{MOV}_{di,j=5}$, as in (44), then their ratio is small, giving rise to expression (45). If the proportion that heavy consumers acquire from e.g. travelers' import is small compared to their proportion acquired from Systembolaget, the adjustment factor depends (almost) solely on version $l = 2$ of the adjustment factor, i.e. $v^{MOV}_{i,j=5,l=2}$. This is also reasonable, because if the acquisition made by heavy consumers for alcohol type i for travelers' import is relatively modest, then we do not need to have the larger expansion factor that version $l = 1$ gives. It suffices with the expansion that version $l = 2$ gives. In other words, the underrepresentation of heavy consumers in the response set does not matter that much since they do not acquire that particular alcohol type from this acquisition mode (travelers' import in the example).

Example 3: if $prop^{MOV}_{dij} > prop^{MOV}_{di,j=5}$, then their ratio is large. If e.g. the ratio is 2.5, then

$$R_{ij} = 2.5 \times \left(v^{MOV}_{i,j=5,l=1} - v^{MOV}_{i,j=5,l=2}\right) + v^{MOV}_{i,j=5,l=2} \quad (47)$$

Studying the expression, we see that $v^{MOV}_{i,j=5,l=2}$ is the starting point and then we add 2.5 times the difference between version 1 and 2 of v. In (44) the difference is $1.59 - 1.38 = 0.21$ so this difference is multiplied with a factor 2.5 that stretches R_{ij} to become larger, in this case even larger than $v^{MOV}_{i,j=5,l=1} = 1.59$. This is also reasonable, because if the acquisition made by heavy consumers for alcohol type i for travelers' import is relatively large, then we want to compensate for them being underrepresented in the response set by the larger expansion factor that version $l = 1$ gives.

Acquisition mode travelers' import ($j = 1$), **cider and fortified wine** ($i = 3{,}5$)

As regards cider and fortified wine, the final adjustment ratio takes a simpler form than for wine, beer and spirits, namely

$$R_{ij} = v^{MOV}_{i,j=5,l=1} \quad (48)$$

i.e. the moving average of version 1 of the preliminary adjusting ratio according to (27). This means that the preliminary adjustment ratio also is the final adjustment ratio in this

case. This is done in this way because the number of heavy consumers that acquire cider and fortified wine is too small for further calculations.

All unregistered acquisition modes ($j = 1,2,3,4$) **and all types of alcohol** ($i = 1,2,3,4,5$)

Above, acquisition mode $j = 1$ and type of alcohol $i = 1,2,3,4,5$ was described in detail. The other acquisition modes are treated similarly. Table 5 summarizes the calculation of R_{ij} for all unregistered acquisition modes and types of alcohol.

Table 5. Summary of calculation of final adjustment ratio for all unregistered acquisition modes and types of alcohol

Acquisition mode	*j*	Type of alcohol	i	Final adjustment ratio	See
Travelers' import	1	Wine, beer and spirits	1,2,4	$R_{ij} = \frac{prop_{dij}^{MOV}}{prop_{di,j=5}^{MOV}} \times \left(v_{i,j=5,l=1}^{MOV} - v_{i,j=5,l=2}^{MOV}\right) + v_{i,j=5,l=2}^{MOV}$	(39)
Travelers' import	1	Cider, fortified wine	3,5	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(48)
Smuggled	2	Beer, spirits	2,4	$R_{ij} = \frac{prop_{dij}^{MOV}}{prop_{di,j=5}^{MOV}} \times \left(v_{i,j=5,l=1}^{MOV} - v_{i,j=5,l=2}^{MOV}\right) + v_{i,j=5,l=2}^{MOV}$	(39)
Smuggled[27]	2	Wine, cider	1,3	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(48)
Internet	3	Wine, beer, spirits, cider, fortified wine	1,2,3,4,5	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(48)
Home production	4	Wine, beer	1,2	$R_{ij} = v_{i,j=5,l=1}^{MOV}$	(48)
Home production[28]	4	Spirits	4	Special estimator, see below	(53)

4.2.4 The estimator for Q_{ij}^{UNREG} and per capita estimator

In section 4.2.1 to 4.2.3 all building blocks for estimating Q_{ij}^{UNREG} according to (5) have been formalized. The estimator for all types of alcohol and unregistered acquisition modes, except for home production ($j = 4$) and spirits ($i = 4$), is done according to

$$\hat{Q}_{ij}^{UNREG} = \hat{t}_{y_{ijU}} \times R_{ij} \times \alpha_{ij} \qquad (49)$$

where $\hat{t}_{y_{ijU}}$, the estimator of total unregistered alcohol (for type of alcohol i and acquisition mode j), is given by (19), R_{ij} is given in table 5 and α_{ij} is given in table 13 in

[27] Smuggled fortified wine is not estimated.
[28] Cider and fortified wine are not estimated as regards home production.

appendix 1. Please not that $\alpha_{i,j=5}$, i.e. the alcohol content each year, according to Systembolaget, is used for all acquisition modes except restaurants ($j = 6$) and grocery stores ($j = 7$).

The estimation for home production of spirits is done in a different way which is described below, but first the per capita estimator is described.

The per capita parameter is given by (6), so replacing Q_{ij}^{UNREG} by its estimator $\hat{Q}_{ij}^{UNREG}$ from (49) gives

$$\frac{\hat{Q}_{ij}^{UNREG}}{N_{15+}} = \frac{\hat{t}_{y_{ij}U} \times R_{ij} \times \alpha_{ij}}{N_{15+}} \tag{50}$$

Special calculation of $\hat{Q}_{i=4,j=4}^{UNREG}$ for home production ($j = 4$) and spirits ($i = 4$)

The procedure is described in words in Trolldal (2020), page 45, and is summarized here. Let pH denote the (estimated) proportion of consumed spirits from home production compared to all consumption of spirits. Please note that pH concerns consumption rather than acquisition. A usual estimate of pH is around 2 to 3 percent. Below, we make some comments on how pH is calculated.

The procedure involves a couple of steps. First, calculate an estimate of total acquisition[29] of pure alcohol as regards spirits ($i = 4$). This is done by summing the acquisition of pure alcohol per capita over all acquisition modes, except home production[30]. Denote this quantity ÖS as in Trolldal (2020), i.e.

$$\ddot{O}S = \frac{\hat{Q}_{i=4,j=1}^{UNREG}}{N_{15+}} + \frac{\hat{Q}_{i=4,j=2}^{UNREG}}{N_{15+}} + \frac{\hat{Q}_{i=4,j=3}^{UNREG}}{N_{15+}} + \frac{Q_{i=4,j=5}^{REG}}{N_{15+}} + \frac{Q_{i=4,j=6}^{REG}}{N_{15+}} \tag{51}$$

Secondly, inflate this quantity by (one minus) the proportion of consumed spirits from home production compared to all consumption of spirits p according to

$$\frac{\ddot{O}S}{1 - pH} \tag{52}$$

The difference is the estimator for $Q_{i=4,j=4}^{UNREG}/N_{15+}$ the per capita measure, i.e.

$$\frac{\hat{Q}_{i=4,j=4}^{UNREG}}{N_{15+}} = \frac{\ddot{O}S}{1 - pH} - \ddot{O}S = \ddot{O}S \times \frac{pH}{1 - pH} \tag{53}$$

A numeric example as regards 2019 might be helpful. We have

[29] Not consumption in this case, which p is.
[30] Grocery stores are also excluded since they are not allowed to sell spirits in Sweden.

$$\text{ÖS} = 0.45 + 0.09 + 0.02 + 0.90 + 0.11 = 1.57 \quad (54)$$

These estimates can be found in Trolldal (2020) in table 14[31]. The estimate pH for 2019 is 0.022, which gives

$$\frac{\hat{Q}_{i=4,j=4}^{UNREG}}{N_{15+}} = \frac{1.57}{1 - 0.022} - 1.57 = 1.61 - 1.57 \approx 0.04 \quad (55)$$

This estimate concerns consumption of home-produced spirits as opposed to all other statistics, which concern acquisition of alcohol. See Trolldal (2020) for a discussion on this topic.

Remark: The estimator pH is formed by a moving average over the three last years similar to $v_{i,j=5,l}^{MOV}$ in (38). Before the moving average is calculated, both the total consumption of sprits and consumption of homemade spirits are corrected for over-representation in the non-response set among the heavy consumers. Since the per capita consumption of spirits from home production is relatively small compared to the total acquisition, we omit the (lengthy) technical details concerning the construction of pH.

4.2.5 The true parameter for Q_{ij}^{REG} and per capita

The parameter Q_{ij}^{REG}/N_{15+} for registered acquisition modes and types of alcohol per capita was earlier stated in (7) and needs no additional explanation. We can underline that this estimator concerns registered acquisition from Systembolaget ($j = 5$), restaurants ($j = 6$) and grocery stores ($j = 7$). In Trolldal (2020), page 43, it is stated that the registered acquisition as regards Systembolaget comes directly from Systembolaget. The acquisition from restaurants is based on wholesalers' reported information published by the Public Health Agency of Sweden (Folkhälsomyndigheten). The acquisition from grocery stores is calculated by the company Delfi, on behalf of the Swedish Brewers Association (Sveriges Bryggerier) and this concerns only low alcohol beer with between 2.8 % and 3.5 % alcohol by volume.

4.3 Variance estimators

In Trolldal (2020) the uncertainty in the estimates for unregistered acquisition is not calculated. Due to the construction of the estimators, especially the construction of R_{ij} in table 5, deriving analytic expressions for the variance for e.g. $\hat{Q}_{ij}^{UNREG}$ in (49) is a complex

[31] Swedish wording: ”Tabell 14. Den totala alkoholanskaffningen uppdelad på anskaffningskälla och dryck, i liter ren alkohol per invånare 15 år och äldre, 2001–2019”.

task. If variance estimators are sought, perhaps a bootstrap procedure can be used. This report does not go any further into this matter.

5 CONNECTION BETWEEN ESTIMATORS AND ESTIMATES

In Appendix 1 (Bilaga 1) in Trolldal (2020) there are several tables with estimates. To facilitate the transition between estimators and their corresponding estimates we provide here some guidance in table 6.

Table 6. Connection between estimators and estimates in Trolldal (2020). Table references are from Trolldal (2020)

Table	Estimates based on estimator (or parameter)	Expression
7, 10, 11, 12, 13, 14	$\frac{\hat{Q}_{ij}^{UNREG}}{N_{15+}}$ and $\frac{Q_{ij}^{REG}}{N_{15+}}$	(50) and (7) and summations
8	Q_{ij}^{REG}	Numerator in (7)
9	$\hat{t}_{y_{ij}U} \times R_{ij}$	Part of numerator in (50)
17	$\hat{t}_{y_{ij}U}$	(19) (No adjustment with R_{ij})
19	R_{ij}	See table 5

We give some examples. In table 7 (Trolldal, 2020) the total quantity of pure alcohol for both unregistered and registered acquisition for all types of alcohol is given. This unregistered acquisition is obtained by summing $\hat{Q}_{ij}^{UNREG}/N_{15+}$ over all i and j, i.e.

$$\frac{\hat{Q}^{UNREG}}{N_{15+}} = \frac{1}{N_{15+}} \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{Q}_{ij}^{UNREG} \tag{56}$$

In a similar manner the registered acquisition is obtained

$$\frac{Q^{REG}}{N_{15+}} = \frac{1}{N_{15+}} \sum_{i=1}^{I} \sum_{j=1}^{J} Q_{ij}^{REG} \tag{57}$$

Remark 1: All indexes j for acquisition do not apply in (56) nor in (57).

Adding them together gives the total acquisition regardless of acquisition mode, i.e.

$$\frac{\hat{Q}}{N_{15+}} = \frac{\hat{Q}^{UNREG}}{N_{15+}} + \frac{Q^{REG}}{N_{15+}} \tag{58}$$

Remark 2: Since $\hat{Q}/N_{15+}$ contains one part that is estimated and one part that is a registered based (true) parameter value, their sum is an estimate and we use a hat in $\hat{Q}/N_{15+}$ to indicate that.

For example, the year 2019 $\hat{Q}^{UNREG}/N_{15+}$ is estimated to 1.54 and Q^{REG}/N_{15+} is 7.13 which gives the sum $\hat{Q}/N_{15+} = 8.67$ (Trolldal, 2020). In a similar manner all estimates in table 7, and the other tables, can be obtained by summing over appropriate indexes.

One more example: Acquisition of unregistered wine is obtained by summing over the unregistered acquisition modes ($j = 1,2,3,4$) for wine ($i = 1$)

$$\frac{\hat{Q}_{i=1}^{UNREG}}{N_{15+}} = \frac{1}{N_{15+}} \sum_{j=1}^{4} \hat{Q}_{i=1,j}^{UNREG} \quad (59)$$

From table 7 in the report $\hat{Q}_{i=1}^{UNREG}/N_{15+} = 0.37$.

A couple of more remarks: In Trolldal (2020), table 17 contains the unadjusted quantities of acquisition of unregistered alcohol (for type of alcohol i and acquisition mode j) in volume liters (not pure alcohol). These estimates are based on the estimator $\hat{t}_{y_{ij}U}$ in (19) without the adjustment ratio R_{ij}. These estimates are believed to underestimate the true parameter value. For example, for spirits and travelers' import $\hat{t}_{y_{i=4,j=1}U} = 7.1$. From table 19 the adjustment factor $R_{i=4,j=1} = 1.46$ and multiplying them gives $9.0 \times 1.55 = 10.4$ which is equal to the adjusted estimate published in table 9.

6 HOW THE ESTIMATES FROM 2019 ARE LINKED TO THE ESTIMATES BASED ON THE NEW METHODOLOGY IN 2020

From the beginning of 2020, the respondents are given the opportunity to answer the questions in the survey via a self-administered questionnaire. The persons in the sample who have a mobile phone receive a text-message including a link to the self-administered questionnaire. The persons who do not answer the survey that way, or have a landline only, are called up for a personal interview on the telephone instead, in the same way as before 2020. In addition, a new weighting procedure was introduced in the study at the same time, see section 4.1 above.

In order to measure the difference between the estimates based on the self-administered questionnaire and the ones based on the standard survey, a test survey was carried out during 2019. In this test, the new methodology was used in a separate survey. This survey was carried out in parallel to the on-going standard survey in 2019. The test encompassed approximately 9.000 respondents, and 18.000 respondents were included in the standard survey.

The difference between the estimates calculated according to the test and the ones calculated according to the standard procedure was assumed to be the effect of the changes in methodology.

Furthermore, in order to measure the effect of the new weighting procedure on the estimates, both the new and the old weights were used in the dataset from the test. The difference between the two was assumed to be explained by the difference between the old and the new weight used. For some beverages and acquisition modes the new weight resulted in an increased estimate and in some cases a decrease. However, the effect of the new weight was in general small, see table 7.

The share of respondents who chose to answer the self-administered questionnaire in the test survey in 2019 was 53 percent. This share increased to 63 percent during 2020.

This increase during 2020 is included in the calculated total link-values as well, see table 7. In the table, the calculated total link-values are divided into two columns, the effect of the new weight and the effect of the new methodology. The total link-values have been used when the estimates from 2020 are compared with 2019. In the column to the right, there are some figures which refer to comments that are discussed below.

Table 7. The effect of the new weight and the new methodology summed up in total link-values for different types of beverages and acquisition modes.

Mode and beverage	New weight	New methodology	Total link-value*	Comments
Travelers' import				
Spirits	0,04	1,27	1,32	
Wine	-0,01	1,64	1,62	
Beer	0,00	1,21	1,20	
Fortified wine			1,37	1
Cider			1,37	1
Smuggled alcohol				
Spirits	0,16	1,90	2,05	
Beer	0,04	2,27	2,31	
Wine			2,23	2
Cider			2,23	2
Internet:				
Spirits			1,37	1
Fortified wine			1,37	1
Wine			1,37	1
Beer			1,37	1
Cider			1,37	1
Production at home:				
Spirits			1,00	3
Wine			1,37	1
Beer			1,37	1
Systembolaget:				
Spirits	-0,02	1,29	1,27	
Fortified wine	0,01	1,47	1,47	
Wine	-0,11	1,36	1,25	
Beer	-0,02	1,27	1,26	
Cider	-0,06	1,59	1,52	

* The total link-value may differ from the sum of the new weight and the new methodology due to rounding

The major part of the calculated link-values are due to the new methodology, i.e. the use of the new way of answering the questions in the survey. The new weight has a minor effect on the estimates. That effect of the new weight on the estimates ranges from -0,11 to +0,16.

However, the number of persons in the test who acquired alcohol via some modes was too small to be used as a basis for the calculations of any link-values. In these cases, different approaches were used to calculate these values.

1. When it comes to travelers' import, the link-value used for acquisition of fortified wine and cider is the weighted mean value for the travelers' import of spirits, wine and beer. This mean value is also used for acquisition via internet and home production of wine and beer.
2. The weighted mean value of smuggled beer and spirits has been used as the link-value for smuggled wine and cider.
3. The estimated acquisition of spirits produced at home is based on questions concerning consumption of such spirits, and no specific link-values have been used.

On the whole, the new methodology, with the self-administered questionnaire, captures larger volumes of beverages in all the acquisition modes. Consequently, in order to compare the estimates from 2020 with those from 2019, the latter estimates have to be recalculated upwards. However, since purchases at Systembolaget are compared with actual sales and the difference is used as a part of the correction factor, see section 4.2 above, these correction factors become lower when the new methodology is used.

Consequently, the total effect of the new methodology on the estimated total consumption is limited. In 2019, the total consumption of alcohol was estimated at 8.66 liters of pure alcohol per inhabitant aged 15 and above according to the old methodology, but 8.98 when the new methodology was used. The difference was 3,7 % (Trolldal & Åström, 2021).

REFERENCES

Kühlhorn E. Hibell B. Larsson S. Ramstedt M. & Zetterberg HL. (1999). *Alkoholkonsumtionen i Sverige under 1990-talet.* Stockholm: OAS, Socialdepartementet.

Källebring N. (2021). *Bortfall i Monitormätningarna*. CAN report 201.

Nyfjäll M. & Trolldal B. (2019). *Estimating total alcohol consumption in the Monitor survey – a technical description of the estimation method*. CAN report 182.

Särndal C.E. Swensson B. & Wretman J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Särndal C.E. & Lundström S. (2005). *Estimation in Surveys with Nonresponse*. Wiley.

Trolldal B. (2020). *Alkoholkonsumtionen i Sverige 2019*. CAN report 193.

Trolldal B. & Åström V. (2021). *Alkoholkonsumtionen i Sverige 2001-2020*. CAN report 202.

APPENDIX 1 – SUMMARY OF NOTATION

Table 8 shows a summary of notation used. In the table, we do not attach subindex to variables, since it would mean that the table would be even longer. For example, m_p is not in the table. However, m and index p are listed in the table, so by combining the listed variables and indexes all versions of variables used in the report can be obtained.

Table 8. Summary of notation

d	=	Index for domain; in this report the only domain is heavy consumers
i	=	Index for type of alcohol
I	=	Number of types of alcohol
j	=	Index for mode of acquisition
J	=	Number of acquisition modes
k	=	Index for individuals in target population
l	=	Index for version of handling heavy consumers
p	=	Index for period
P	=	Number of periods
ι	=	Index for acquisitions *within* individual k (Greek letter iota)
y	=	Variable for unregistered acquisition of alcohol (sum for an individual over all acquisitions in a period)
y_d	=	Domain variable, equals y in the domain and zero outside the domain
z	=	Variable for unregistered acquisition of alcohol (each acquisition)
x	=	Variable for registered acquisition of alcohol
z'	=	Variable for registered acquisition of alcohol (each acquisition)
U	=	Target population for the Monitor survey (set of individuals)
N	=	Population size of target population U
N_k	=	Number of acquisitions for individual k during the reference period (year or month[32])
t_{yU}	=	Parameter total quantity of unregistered acquired alcohol (volume liters, not pure alcohol)
t_{xU}	=	Parameter total quantity of registered acquired alcohol (volume liters, not pure alcohol)
Q^{UNREG}	=	Parameter total quantity of unregistered acquired pure alcohol
Q^{REG}	=	Parameter total quantity of registered acquired pure alcohol
α	=	Average alcohol content, i.e. the percentage of pure alcohol by weight
N_{15+}	=	Population size of age 15 and older in Sweden
s	=	Sample from the population (set of individuals)
n	=	Sample size
r	=	Response set
m	=	Response set size
w	=	Weight variable
$\hat{t}_{yU}$	=	Estimator for t_{yU}
R	=	Inflating adjustment ratio; account for several aspects, primarily under-reporting of acquisition
$\hat{t}_{xU}$	=	Estimator for t_{xU}
$\hat{Q}^{UNREG}$	=	Estimator for Q^{UNREG}

[32] The length of the reference period should be clear from the context.

Ind = Indicator for the heavy consumers
c_{100} = Consumed quantity of 100 percent pure alcohol
$Percentile_{99}$ = The 99:th percentile
$\hat{\bar{y}}_U$ = Mean estimator
$\hat{\bar{y}}_{U_d}$ = Mean estimator for domain
$prop_d$ = Proportion of acquisition for heavy consumers compared to the total population
$prop_d^{MOV}$ = Three year moving average of $prop_d$
v = Preliminary adjustment factor, used in calculating R
v^{MOV} = Three year moving average of v
$\ddot{O}S$ = Estimate of total acquisition except home production of pure alcohol as regards spirits
pH = Proportion of consumed spirits from home production compared to all consumption of spirits

In appendix 2, the operational definition of variables $\boldsymbol{y}$ and $\boldsymbol{x}$ is made. This requires some additional variables which are listed in table 9.

Table 9. Additional variables in appendix 2

g = number of individuals travelling together on most recent trip (travelers' import)
a = number of trips crossing the Swedish border last 30 days (travelers' import)
q = quantity of peddled alcohol (smuggling acquisition)
b = number of times smuggled alcohol is acquired last 30 days (smuggling acquisition)
f = number of times of internet acquisitions during the last 30 days (internet acquisition)
e = number of times acquiring alcohol from Systembolaget during the last 30 days

Table 10 shows an explanation of index $\boldsymbol{i}$.

Table 10. Categories for index $\boldsymbol{i}$

i	Type of alcohol
1	Wine
2	Beer
3	Cider
4	Spirits
5	Fortified wine
6	Low alcohol beer (folköl)

Table 11 explains index $\boldsymbol{j}$.

Table 11. Categories for index $\boldsymbol{j}$

j	Mode of acquisition
1	Travelers' import (resandeinförsel)
2	Smuggled
3	Internet
4	Home production (hemtillverkning)
5	Systembolaget
6	Restaurants
7	Grocery stores (only $i = 6$, low alcohol beer)

Table 12 explains index $\boldsymbol{l}$.

Table 12. Categories for index $\boldsymbol{l}$

l	Ways of handling heavy consumers
1	No special nonresponse compensation for heavy consumers
2	Special nonresponse compensation for heavy consumers

Table 13 shows numeric values for variable $\boldsymbol{\alpha}$. Each year, the values are revised (often only by a small amount). Note that $\boldsymbol{\alpha_{i,j=5}}$, i.e. the alcohol content according to Systembolaget, is used in the estimation for all acquisition modes except restaurants ($\boldsymbol{j = 6}$) and grocery stores ($\boldsymbol{j = 7}$).

Table 13. Average alcohol content by weight 2019, $\boldsymbol{\alpha_{ij}}$

j	Acquisition mode	i	Type of alcohol	α_{ij}
5	Systembolaget	4	Spirits	0,3697
5	Systembolaget	1	Wine	0,1275
5	Systembolaget	5	Fortified wine	0,1605
5	Systembolaget	2	Beer	0,0556
5	Systembolaget	3	Cider	0,0510
7	Grocery stores	7	Low alcohol beer	0,0330
6	Restaurants	4	Spirits	0,3043
6	Restaurants	1	Wine	0,1126
6	Restaurants	2	Beer	0,0525

APPENDIX 2 – OPERATIONAL DEFINITION OF STUDY VARIABLES

The operational definition of y_{ijk} and z_{ijk} (unregistrated alcohol) and x_{ijk} and z'_{ijk} (registered alcohol) is done slightly different for different acquisition modes. In section 2.1 it was stated that since each individual k can acquire alcohol several times during the reference period (year), an individual can be seen as a cluster where the cluster size is the number of times an acquisition is made. Let N_k denote the number of acquisitions that individual k makes during a year and N_{kp} the number of acquisitions during a period (month).

In section 2.1 we used ι (Greek iota) as a running index for different acquisitions. In the survey only questions about the most recent acquisition are made. This means that we do not have to use the running index ι. Instead, we use the ordinary index k.

We divide the description by acquisition mode.

Travelers' import ($j = 1$)

In the survey, the respondent is asked about the most recent trip from abroad to Sweden. This constitutes travelers' import. For the most recent trip, define the following variables asked in the survey:

- $z_{i,j=1,k}$ is the quantity (in centiliters[33]) acquired of unregistered alcohol at the most recent trip k from abroad to Sweden[34], for type of alcohol i and travelers' import ($j = 1$).
- g_k is the number of individuals travelling together (as a group) on the most recent trip. For example, the respondent and the partner gives $g_k = 2$.
- a_k is the number of trips crossing the Swedish border during the last 30 days.

The total acquisition for unregistered type of alcohol i and travelers' import ($j = 1$) that individual k has made during the period, i.e. the y –value, is calculated as

$$y_{i,j=1,k} = \frac{z_{i,j=1,k} \times a_k}{g_k} \tag{60}$$

Note that if $a_k > 1$ this is an *estimate* of all travelers' import acquisitions made by individual k during period p. However, we omit the "hat" over y to facilitate the notation.

Remark: CAN has implemented rules if the y – variable in (60) takes too large values. We do not describe these rules in this report.

[33] The respondents can actually answer in any unit, e.g. number of bottles, but this is converted into centiliters.
[34] Note that an answer of zero acquired alcohol is a valid answer.

Smuggled alcohol ($j = 2$) - ***Purchases of alcohol that has been smuggled into the country***

In the survey the respondent is asked about the most recent purchase of alcohol, that has been smuggled into the country, and if the respondent in turn has sold any part of it. For the most recent acquisition, define the following variables asked in the survey:

- $z_{i,j=2,k}$ is the quantity (in centiliters) acquired of unregistered smuggled alcohol at the most recent acquisition k.
- $q_{i,j=2,k}$ quantity of peddled alcohol in combination with the most recent acquisition for type of alcohol i and acquisition mode j.
- b_k the number of times that individual k has acquired smuggled alcohol in the last 30 days.

The total acquisition for unregistered type of alcohol i as regards smuggled alcohol ($j = 2$) individual k has made during the period, i.e. the y −value, is calculated as

$$y_{i,j=2,k} = (z_{i,j=2,k} - q_{i,j=2,k}) \times b_k \qquad (61)$$

This is an *estimate* of all smuggled acquisitions made by individual k during period p. Note that if $z_{i,j=2,k} = q_{i,j=2,k}$ all acquired smuggled alcohol is peddled.

Remark: Expression (61) is not calculated for fortified wine since it is smuggled to an extent that is negligible.

Remark: CAN has implemented rules if variable y in (61) takes too large values. We do not describe these rules in this report.

Internet ($j = 3$)

In the survey the respondent is asked about the most recent internet acquisition, with the exception of Systembolaget. For the most recent acquisition, define the following variables asked in the survey:

- $z_{i,j=3,k}$ is the quantity (in centiliters) acquired of unregistered alcohol from the internet (most recent acquisition).
- $f_{i,j=3,k}$ is the number of times of internet acquisitions during the last 30 days.

The total acquisition for unregistered type of alcohol i and internet ($j = 3$) that individual k has made during the period, i.e. the y −value, is calculated as

$$y_{i,j=3,k} = z_{i,j=3,k} \times f_{i,j=3,k} \quad (62)$$

Remark: CAN has implemented rules if variable y in (62) takes too large values. We do not describe these rules in this report.

Home production ($j = 4$)

In the survey the respondent is asked about home production during the last 30 days. A criterion is that the home-produced alcohol should have been completed during the last 30 days. The total quantity of completed alcohol is asked. This means that we do not have to distinguish between individual k and all possible acquisitions. Define the following variables asked in the survey:

- $z_{i,j=4,k}$ is the quantity (in liters) of home-produced (completed) alcohol during the last 30 days.

The total acquisition for unregistered type of alcohol i and home production ($j = 4$) for individual k during the period, i.e. the y −value, is calculated as

$$y_{i,j=4,k} = z_{i,j=4,k} \times 100 \quad (63)$$

The factor 100 converts the unit liters in the question into centiliters to harmonize with other types of alcohol.

Remark 1: CAN has implemented rules if variable y in (63) takes too large values. We do not describe these rules in this report.

Remark 2: Expression (63) only applies to wine ($i = 1$) and beer ($i = 2$). Regarding spirits ($i = 4$) a (completely) different procedure is used. See expression (53). No questions concerning home production of cider nor fortified wine are asked.

Registered alcohol ($j = 5$)

In the survey the respondent is asked about the most recent acquisition from Systembolaget ($j = 5$). As regards acquisition from restaurants and grocery stores, no questions are asked in the survey. For the most recent acquisition, define the following variables asked in the survey:

- $z'_{i,j=5,k}$ is the quantity (in centiliters) acquired registered alcohol as regards the last acquisition from Systembolaget.

- e_k is the number of times acquiring alcohol from Systembolaget during the last 30 days.

The total acquisition of registered alcohol i at Systembolaget ($j = 5$) that individual k has made during the period, i.e. the y −value, is calculated as

$$x_{i,j=5,k} = z'_{i,j=5,k} \times e_k \quad (64)$$

This is an *estimate* of all acquisitions at Systembolaget made by individual k during period p.

APPENDIX 3 – RATIONALE FOR INTERPRETATION OF (28)

Expression (28) is interpreted as estimator of the percentage of acquisition for heavy consumers in relation to the total population at Systembolaget $j = 5$ (for type of alcohol i). The rationale for this interpretation is motivated here.

The heavy consumers constitute a domain in the population. The domain is defined as the individuals in the response set with the one percent largest consumption of pure alcohol (see section 4.2.3.1 and the use of the 99:th percentile $Percentile_{99}$). The *total* acquisition of alcohol (not pure alcohol) for this domain (heavy consumers) at Systembolaget divided by total acquisition of alcohol for the whole population ought to give the sought percentage. Both these parameters are estimated from the survey, so we write

$$\frac{\hat{t}_{i,j=5,U_d}}{\hat{t}_{i,j=5,U}} = \frac{\hat{\bar{y}}_{i,j=5,U_d} \times N_d}{\hat{\bar{y}}_{i,j=5,U} \times N} \approx \frac{\hat{\bar{y}}_{i,j=5,U_d} \times \widehat{N}_d}{\hat{\bar{y}}_{i,j=5,U} \times \widehat{N}} \tag{65}$$

Since the domain is defined by the 99:th percentile $Percentile_{99}$ in the response set, the following equality holds approximately

$$\frac{\widehat{N}_d}{\widehat{N}} \approx \frac{1}{100} \tag{66}$$

Which motivates the multiplication of $1/100$ in expression (28).

C.A.N

CAN – Nationellt kompetenscentrum inom ANDTS (alkohol, narkotika, dopning, tobak och spel). Vårt fokus är undersökningar, forskning och kommunikation. Vi genomför även lokala och regionala undersökningar och kartläggningar på uppdrag av kommuner och län.

CAN är en unik aktör med en fot i civilsamhället och en i det offentliga. Styrelsen utses till hälften av regeringen och till hälften av de cirka 50 medlemsorganisationerna vid årsmötet. CAN står för Centralförbundet för alkohol- och narkotikaupplysning.

Läs mer om oss på www.can.se