Health Canada
Santé Canada
Statistics Canada
Statistique Canada

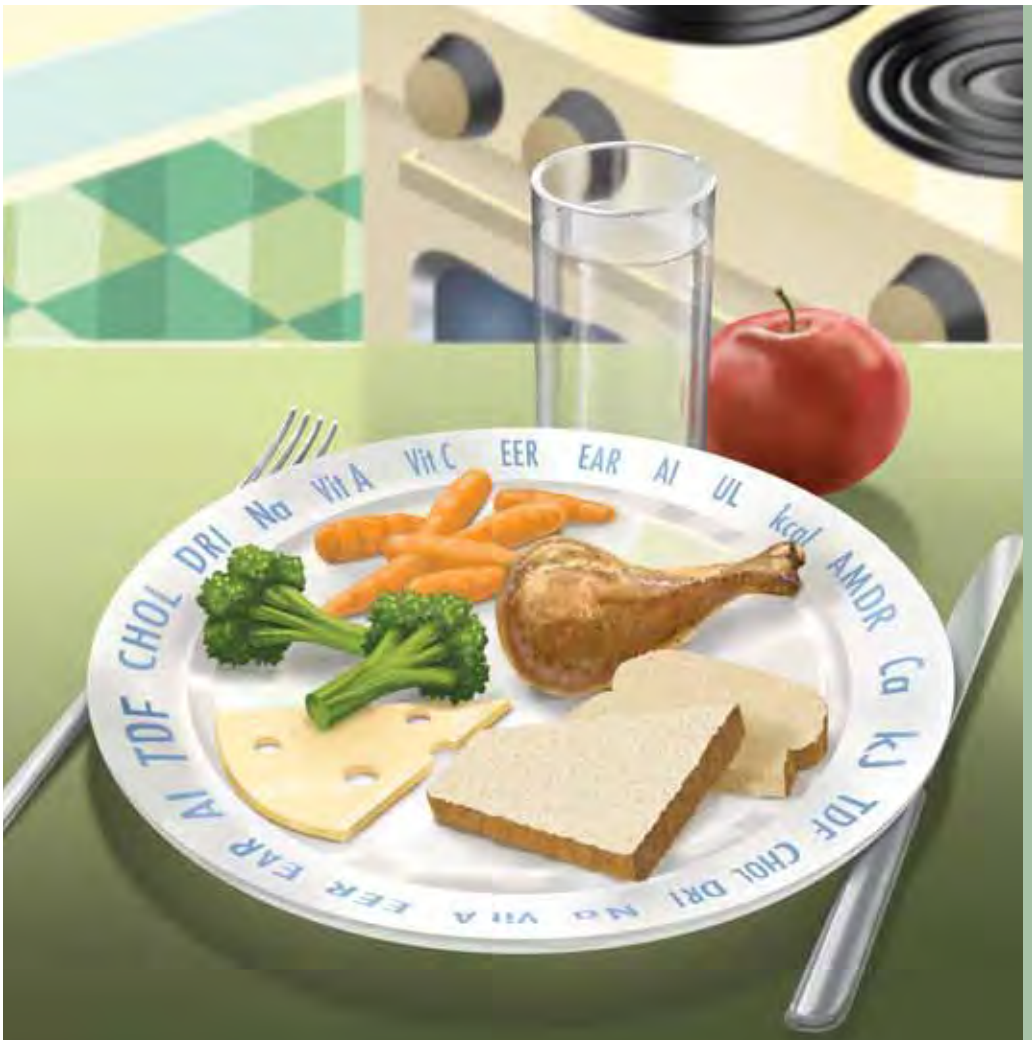
Canadian Community Health Survey Cycle 2.2, Nutrition (2004)

Nutrient Intakes from Food

Provincial, Regional and National Summary Data Tables
Volume 1

Note:

This PDF contains the descriptive text. The full report, which also includes summary data tables and appendices, is available at:
www.hc-sc.gc.ca/fn-an/surveill/nutrition/commun/index_e.html



Canada

Health Canada is the federal department responsible for helping the people of Canada maintain and improve their health.
We assess the safety of drugs and many consumer products, help improve the safety of food, and provide information to Canadians to help them make healthy decisions. We provide health services to First Nations people and to Inuit communities. We work with the provinces to ensure our health care system serves the needs of Canadians.

Published by authority of the Minister of Health.

Canadian Community Health Survey, Cycle 2.2, Nutrition (2004)—Nutrient Intakes from Food
Provincial, Regional and National Summary Data Tables, Volume 1
is available on Internet at the following address:
www.hc-sc.gc.ca/fn-an/surveill/nutrition/commun/index_e.html

Également disponible en français sous le titre :
Enquête sur la santé dans les collectivités canadiennes, cycle 2.2, Nutrition (2004)—Apports nutritionnels provenant des aliments
Tableaux sommaires provinciaux, régionaux et nationaux, Volume 1

This publication can be made available on request on diskette, large print, audio-cassette and braille.

For further information or to obtain additional copies, please contact:
Publications
Health Canada
Ottawa, Ontario K1A 0K9
Tel.: (613) 954-5995
Fax: (613) 941-5366
E-Mail: info@hc-sc.gc.ca

Cat.: H164-45/1-2007E-PDF
ISBN: 978-0-662-45760-2

Acknowledgements

Health Canada would like to acknowledge and thank the individuals who have contributed to this report. This publication was produced as a joint venture between Health Canada and Statistics Canada. Technical experts from Health Canada's Food Directorate and Office of Nutrition Policy and Promotion and from Statistics Canada's Health Statistics Division wrote the report. Project guidance was provided by Health Canada's Food and Nutrition Surveillance System Working Group, which consists of individuals from Health Canada, the Public Health Agency of Canada and the Federal/Provincial/Territorial Group on Nutrition. External experts have advised both departments, through the CCHS 2.2 Users Group, on the need for a consistent approach to analyzing and reporting these data. The draft report was reviewed by internal experts from both departments and underwent external peer review.

We gratefully acknowledge the contribution to the completion of this publication by external reviewers Nancy Yu (Manitoba Health); Nathalie Plante (Institut de la statistique du Québec); Alanna Moshfegh (United States Department of Agriculture); and Susan Barr (University of British Columbia).

Table of Contents (for the full report)

List of Tables

List of Appendices

List of Abbreviations

Abbreviation	Meaning
AI	Adequate Intake
AMDR	Acceptable Macronutrient Distribution Range
CCHS	Canadian Community Health Survey
CV	coefficient of variation
d	day
DRI	Dietary Reference Intake
EAR	Estimated Average Requirement
g	gram
kcal	kilocalories
mg	milligram
n	sample size
RAE	Retinol Activity Equivalent
SE	standard error
SIDE	Software for Intake Distribution Estimation
UL	Tolerable Upper Intake Level

Glossary

A complete glossary of terms appears in *Canadian Community Health Survey, Cycle 2.2, Nutrition (2004)—A Guide to Accessing and Interpreting the Data* starting on page xiii. Additional technical terms found in the Methodology section of the present document (Section II) are described below.

BOOTVAR

A program available in SAS or SPSS, which was developed by Statistics Canada, that uses the bootstrap method to estimate variance from complex survey sampling designs.

Centered Fourth Moment

The Kurtosis statistic that measures the heaviness or thickness of the tails of a distribution. Observations that are normally distributed should have a Kurtosis near three.

LINFRAC

SIDE option that specifies the proportion of data points to treat as linear at the tails of the distribution when attempting a semi-parametric transformation to normality (Default = 0.0 or two points).

MAXJP

SIDE option that specifies the maximum number of join points allowed when fitting a grafted polynomial function to a normal probability plot (Default = 12).

PEVCR

SIDE dataset that instructs SIDE to force the within-individual variance and centered fourth moment to the specified amount.

NPEVCR

SIDE dataset (used with PEVCR) that contains the relative weight that SIDE should use to determine when to force the variance from PEVCR (name = 9999 instructs SIDE to give full weight to PEVCR and not to attempt to calculate the within-individual variance).

I Introduction

This publication provides summary data tables about the nutrient intakes from food obtained by Canadians in 2004, using data from the Canadian Community Health Survey (CCHS), Cycle 2.2, Nutrition (2004). Data are provided for 14 Dietary Reference Intake (DRI) age–sex groups. For nutrients that have DRIs, the tables also compare usual intakes of selected nutrients to the DRIs. Data used for producing the tables in this report were obtained from the CCHS 2.2 Share File. The nutrient intakes represent food consumption; data on nutrient intakes from vitamin and mineral supplements were still being validated at the time these tables were compiled.[1]

This document is a reference for those who will use the CCHS Cycle 2.2 data and its findings to guide nutrition-related program and policy decisions. It will be of particular benefit to provincial ministries of health, researchers and graduate students, policy makers and analysts, public health professionals, epidemiologists, dietitians, the food industry, and the health media.

This report is the third in a series of products released by Health Canada as part of its ongoing support to users of the CCHS 2.2 data. It has been undertaken as a joint venture with Statistics Canada. To optimize its usage, we recommend that it be read in concert with the first report, *Canadian Community Health Survey, Cycle 2.2, Nutrition (2004)—A Guide to Accessing and Interpreting the Data* (available at **www.hc-sc.gc.ca/fn-an/surveill/nutrition/commun/cchs_focus-volet_escc_e.html**), published by Health Canada in 2006. That report includes an overview of the CCHS 2.2, including a description of the survey sample, how the survey was conducted, survey components and an introduction to DRIs.

[1] Because supplements may make meaningful contributions to nutrient intakes, inferences about the prevalence of nutrient excess or inadequacy based on intakes from food alone may respectively underestimate or overestimate the prevalences based on total nutrient intakes from both food and supplements.

This document consists primarily of data tables. Section II describes the methodology used to produce the tables, including the process used and the means of addressing problems encountered. The report does not provide any interpretation or draw conclusions. Readers are encouraged to consult *Canadian Community Health Survey, Cycle 2.2, Nutrition (2004)—A Guide to Accessing and Interpreting the Data* for examples of how to interpret the CCHS 2.2 data.

This document is Volume 1 of what is expected to be a three-volume set of summary tables comparing the usual intake of 30 nutrients and other dietary components to the DRIs. The 13 sets of tables created for Volume 1 are based on the core nutrients that must be listed on the Nutrition Facts table;[2] they include energy, 6 nutrients presented as percent of total energy intake and 6 as absolute amounts. Results are presented for 13 geographical areas: the 10 provinces, the Atlantic Region, the Prairie Region, and Canada excluding the territories. Data from the four Atlantic provinces and the three Prairie provinces were combined into the Atlantic Region and the Prairie Region, respectively, given the small sample sizes in these provinces.

Volumes 2 and 3 will present data tables related to the remaining nutrients, two as percent of energy intake (linoleic acid, linolenic acid) and the remainder in absolute amounts. The latter include the vitamins folate (Dietary Folate Equivalents, folic acid, naturally occuring folate and total folacin), niacin, riboflavin, thiamin, and vitamins A, B_6, B_{12} and D; the minerals iron, magnesium, phosphorus, potassium and zinc; the macronutrients linoleic acid, linolenic acid, monounsaturated fats, polyunsaturated fats, saturated fats, total fats, protein, carbohydrates and total sugars; and the other dietary components water and caffeine. Recognizing that smoking status affects vitamin C requirements, additional tables will be provided on the intake of vitamin C by smoking status.

[2] There are two exceptions. Because iron needs the full probability method for the EAR comparisons, it will be included in a subsequent Summary Data Tables volume. (**Note**: Explanation of the full probability method can be found in *Canadian Community Health Survey, Cycle 2.2, Nutrition (2004)—A Guide to Accessing and Interpreting the Data*.) Trans fat intake data cannot be obtained from the CCHS 2.2 dataset and therefore are not reported.

Work is underway to complete the data tables for the remaining nutrients. The table below provides a tentative outline of which components are intended to be included in Volume 2 and Volume 3.

Tentative Release Schedule for Remaining Data Tables	
Volume 2	**Volume 3**
Dietary Folate Equivalents	Caffeine
Folic acid	Carbohydrates
Iron	Linolenic acid (g, % energy)
Linoleic acid (g, % energy)	Moisture (water)
Magnesium	Monounsaturated fats
Niacin	Naturally occuring folate
Phosphorus	Polyunsaturated fats
Potassium	Protein
Riboflavin	Saturated fats
Thiamin	Total fats
Vitamin B_6	Total folacin
Vitamin B_{12}	Total sugars
Vitamin C intake by smoking status*	
Vitamin D	
Zinc	

* For domains with adequate sample

II Methodology: Estimation with Software for Intake Distribution Estimation (SIDE)

II.1 Introduction

One of the goals of the Canadian Community Health Survey (CCHS), Cycle 2.2, Nutrition (2004) is to estimate distributions of usual intake from food for several nutrients at the provincial level for 15 Dietary Reference Intake (DRI) age–sex groups. To accomplish this, recalls of what respondents ate in the 24 hours preceding the interview were collected; a second recall was obtained from a representative subsample of the group. Using data from only the first dietary recall produces a measure of daily intake, while data from both the first and second recalls can be used to produce an estimate of usual intake.

The *daily intake* of an individual is the quantity of nutrients or food eaten in one day, whereas an individual's *usual intake* is the long-term average of the daily intake. Similarly, for a population, daily intake data reflect intakes of a large number of people on a given day. To obtain an estimate of a population's *usual intake* distribution from daily intake data, one must fit a measurement error model. The Software for Intake Distribution Estimation (SIDE) is a program that is able to do this.

In general, it is more informative to study the usual intake distribution of a population than its daily intake distribution. For example, knowing the fraction of the population with low intakes of a particular nutrient over a long period of time is more important than knowing the fraction of the population with low intakes of a particular nutrient on a given day. It is important to note that SIDE does not compute whether a single respondent does or does not have low (or high) intakes over the long or short term, but rather computes only the proportion of the population as a whole. The variability in intakes among a group on a given day reflects both variability in intake *within* specific individuals (who may have eaten more or less than usual on that day) as well as *between* different individuals (who habitually have higher or lower intakes). One of the main reasons that SIDE is used as a measurement

error model is to reduce the effect of the within-individual variance while measuring the between-individual variance.

It is not necessary to use SIDE to calculate estimates that have no link to the usual intake distribution. For example, to study daily intake distributions, one can simply use any preferred statistical software. Practically, this means SIDE is only needed for analyses requiring both the first and second recalls.

Three main types of estimates can be obtained from studying usual intake distributions (listed from the simplest to the most complex):

1. the usual intake mean;
2. the percentage of the population having a usual intake under (or over) a given threshold (cut-off); and
3. a percentile of the distribution.

Under the fitted measurement error model, the mean of the usual intakes is equal to the mean of the daily intakes. Consequently, it is not necessary to use SIDE (i.e. adjust the model) to obtain this type of estimate. For the estimate of the mean, a standard technique is used. However, SIDE is needed for estimating a percentage or a percentile of a usual intake distribution.

II.2 SIDE

SIDE uses the method described by Nusser et al. (1996).[3] As indicated in Section II.1, this method is based on a measurement error model that makes the link between daily intake measurements and estimates of usual intake distributions. The methodology of the software is divided into four steps:

1. preliminary adjustments;
2. a semi-parametric transformation to normality;
3. estimation of the within-individual and between-individual variances of the daily intake, which is needed to estimate the usual intake distribution in the normal scale; and

[3] Nusser SM, Carriquiry AL, Dodd KW, Fuller WA: A semiparametric transformation approach to estimating usual daily intake distributions. *J Am Stat Assoc* 1996; 91: 1440–1449

4. a transformation of the usual intake distribution from the normal scale to the original scale.

Statistics Canada recommends the use of SIDE to estimate usual intake distribution characteristics from the CCHS Cycle 2.2 data. This software is complex, complete and precise for this kind of estimation. The version of SIDE used was Version 1.11, written in the SAS/IML language.

The use of SIDE is laborious, especially when confidence intervals of the studied characteristics are needed.

- Statistics Canada documentation (available at **www.statcan.ca/english/sdds/document/5049_D22_T9_V1_E.pdf**) and the SIDE user guide explain the process for using the software.
- The official SIDE user guide can be obtained from the Department of Statistics and Center for Agricultural and Rural Development, Iowa State University. The guide for the SAS/IML version of SIDE can be found at: **www.card.iastate.edu/publications/DBS/PDFFiles/96tr30.pdf**.
- For more information on the mathematics behind the software, consult the technical guide at: **www.card.iastate.edu/publications/DBS/PDFFiles/96tr32.pdf**.

11.3 Using SIDE to Produce Tables from the CCHS Share File

The dataset used in this compendium was the CCHS 2.2 Share File, which consists of all respondents who agreed to share their responses with Health Canada, provincial ministries of health and l'Institut de la Statistique du Quebec. Fewer than 5% of respondents refused to share their responses. Excluded from the dataset were respondents with null intakes (zero total intake from food) or invalid intakes, breastfed children and pregnant or breastfeeding women. Day one and day two recalls were used, although respondents with day two recalls who did not have a corresponding day one recall were excluded. Analysis was performed on provincial, regional (Atlantic and Prairies) and national levels for all DRI age–sex groups other

than children aged between 0 and 1 year. Analysis was also performed on the aggregated age–sex groups: males 19+ years and females 19+ years.

SIDE is a generalized software product that offers the option of specifying certain restraints or implementing options different than the default. The following settings were used in the production of this compendium:[4]

CLASSVAR

CLASSVAR specifies the list of variables to ratio-adjust the analysis variables for the effects of classification variables. Because the variables included in this dataset must not reduce the estimate of the between-individual variance, only day of the week (ADMDDD) was used in the CLASSVAR dataset.

NPTS

This option controls the number of percentiles to output. If NPTS is fixed at 9999, the software will output all percentiles from 0.0001 to 0.9999 (0.0001, 0.0002, ..., 0.9998, 0.9999). By default NPTS is equal to 41, but NPTS = 1000 was used sometimes to obtain better graphical assessments of the model.

DESC

The DESC dataset includes the LINFRAC and MAXJP options which are useful for solving some problems with the model fitting. When SIDE attempts a semi-parametric transformation to normality, LINFRAC specifies the proportion of data points to treat as linear at the tails of the distribution. By default, LINFRAC uses two points in each tail: the highest two points and the lowest two points. For the tables in this compendium, LINFRAC was set to 1% of the data points then adjusted if needed to a value between 0 and 5%. Adjustments were made based on SIDE's ability to produce normality with its transformation. Occasionally the nature of the data required a different value of LINFRAC to satisfy normality of the transformation. Similarly, MAXJP specifies the maximum number of join points allowed when fitting a

[4] The document *Software for Intake Distribution (SIDE) Non-default Model Options* presents all of the non-default model options that were used for each summary data table. Available from the Statistics Canada website at: **http://cansim2.statcan.ca/cgi-win/cnsmcgi.exe?LANG=e&ResultTemplate=OLC&CORCMD=GETEXT&CORTYP=1&CORRELTYP=2&CORID=5049**

grafted polynomial function to a normal probability plot. The default value for MAXJP is 12 but occasionally the nature of the data required an adjustment to get an acceptable semi-parametric transformation.

PEVCR and NPEVCR

These datasets enable the user to force values of the measurement error variance (within-individual variance) and the centered fourth moment for the procedure. Forcing the within-individual variance is sometimes required when SIDE is unable to produce an estimate because it calculates a negative value for the variance of usual intake (Warning 65). In these cases the within-individual variance and centered fourth moment of a higher level domain (regional or national level) must be used to allow SIDE to produce a valid estimate. When forcing the variance it is assumed that the within-individual variance at the higher level domain is similar to that of the lower level domain.

WTVAR

The WTVAR dataset assigns the weights to be used for each record. Each respondent in the CCHS 2.2 Share File dataset has an assigned weight representing the Canadian population. For the tables the sampling weight called WTSD_S was used. For bootstrap estimation of the variance the bootstrap weights BSW1–BSW500 were used.

II.4 Measuring Sampling Variability with Bootstrap Replication

For surveys with simple sampling designs (e.g. simple random sampling or stratified sampling), mathematical formulas exist to estimate the sampling variance. The CCHS 2.2 has a complex design, implying that no mathematical formula exists to calculate the sampling variability. It is necessary to use a replication method to estimate this variance, and the most convenient method is bootstrap replication. BOOTVAR is a program available in SAS or SPSS, developed by Statistics Canada, that uses the bootstrap method to estimate variance from complex survey sampling designs.

For simple estimates such as totals, ratios or regression parameters, it is possible to estimate the sampling variability by using BOOTVAR directly with the bootstrap weights. To obtain this estimate, the macro calculates the parameter of interest (e.g. total, ratio) for each of the 500 replicates and then calculates the variance between the 500 values. This is the method used to estimate the nutrient intake average using day one recalls only. For estimates related to distributions of usual intake, this process must be imitated when estimating with SIDE. Thus, it is necessary to estimate the parameters of interest with SIDE for each replicate (using each bootstrap weight) and then calculate the variance between each of the 500 estimates. It is important to note that for distribution estimates other than the mean, the method used for calculating the variance among the 500 SIDE estimates is slightly different than that used by BOOTVAR.

When BOOTVAR calculates the variance of the 500 replicates it compares each estimate to the mean of the 500 bootstraps (the bootstrap mean). But we also have the root estimate (the estimate calculated using the design weights) from the data. Under normal circumstances, because the number of replicates is large (500) the bootstrap mean will converge to the root estimate. However, since SIDE may fail for some of the 500 replicates, we cannot be certain that all 500 distribution estimates will be available to calculate the mean of the bootstrap estimates. For this reason, when calculating the variance from the bootstrap estimates we compare each replicate to the root estimate, not to the mean of the bootstrap estimates. This allows us to account for some of the bias caused by failing replicates.

For a description of the BOOTVAR programs consult the documentation located at: **www.statcan.ca/english/rdc/whatdata.htm#tools**.

Because SIDE requires steps including adjustments, transformations to normality and estimates of within-individual and between-individual variance, it is not always possible to produce 500 error-free replicates when using bootstrap replication to estimate sampling variability. For each domain studied, decisions had to be made as to whether the estimates produced by

SIDE were acceptable. Often, changing the default values such as LINFRAC and MAXJP produced more reliable results. Since SIDE stops processing when it calculates a negative value for the variance of usual intake (Warning 65) it is possible that some or many of the 500 replicates fail. When some replicates fail, accepting the estimate of sampling variability means accepting a certain level of bias. In general, a small number of failing replicates represents a small bias, while a large number of failing replicates represents a large bias. It should be noted that point estimates and bootstrap estimates are produced using the same SIDE options.

When using bootstrap replication to estimate the sampling variability, it is important to ensure that the estimate converges to a value over the 500 replicates. Ideally, 1000 or more bootstrap replicates would be used to ensure that the estimate converges to a value. However, by convention and due to processing time, 500 replicates are considered sufficient.

Considering the convergence of the bootstrap estimate of sampling variability and the number of failing replicates due to SIDE errors, we adopted the following general guidelines to decide when a bootstrap estimate may be accepted:

1. If the bootstrap estimate does not converge to a value then the estimate is not accepted.
2. If the number of failed replicates is 100 or less then the level of convergence must be medium to strong, otherwise the estimate is not accepted.
3. If the number of failed replicates is 100 to 150 then the level of convergence must be strong, otherwise the estimate is not accepted.
4. If the number of failed replicates is 150 to 200 then the level of convergence must be strong to very strong, otherwise the estimate is not accepted.
5. If the number of failed replicates is greater than 200 then the estimate is not accepted regardless of the strength of convergence.

The thresholds listed above are overlapping because each decision is made separately in a subjective manner, based on a visual interpretation of the plot

of variance estimate over the 500 replicates. The assessment of whether a particular plot represents a very strong, strong or medium convergence is at the discretion of the analyst. The plot of a very strong convergence will show a wave that decreases its width as the number of replicates increases. The tail end of a very strong convergence plot should resemble a straight line of slope 0. Conversely, a weaker convergence plot may show a wave that has one or many spikes and valleys with no tendency of approaching a slope of 0 at the tail.

One has to balance the risk of accepting bias with the ability of the bootstrap method to estimate the sampling variability. For cases in which the estimate is not accepted, the SIDE option of forcing the within-individual variance is used. The within-individual variance and centered fourth moment of the next highest geographical domain are used, whether it is at the regional or national level. For cases in which the regional level variance also fails to produce an acceptable estimate, forcing the variance from the national level is used. When using the bootstrap method it is important to note that the centered fourth moment and the within-individual variance are forced for each of the 500 replicates. So for replicate one, the centered fourth moment and the within-individual variance of the higher domain from replicate one are used; for replicate two, the centered fourth moment and the within-individual variance of the higher domain from replicate two are used; and so forth.

In some cases, one or possibly two replicates are too influential in the bootstrapping method. When these one or two influential replicates cause a bootstrap estimate not to converge, and when the exclusion of these one or two replicates allows the bootstrap estimate to converge, it is an option to exclude them from the bootstrap estimate.