National Library of Canada
Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch
Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

THE UTILITY OF INDIRECT STANDARDIZATION FOR ESTIMATING COMMUNITY HEALTH STATUS

Carol Janice Strike
057224

Thesis submitted to
the School of Graduate Studies and Research
in partial fulfilment of the requirements for
the M.Sc. in Epidemiology

University of Ottawa

© Carol Janice Strike, Ottawa, Canada, 1994

National Library of Canada

Bibliothèque nationale du Canada

Acquisitions and Bibliographic Services Branch

Direction des acquisitions et des services bibliographiques

395 Wellington Street
Ottawa, Ontario
K1A 0N4

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

THE AUTHOR HAS GRANTED AN IRREVOCABLE NON-EXCLUSIVE LICENCE ALLOWING THE NATIONAL LIBRARY OF CANADA TO REPRODUCE, LOAN, DISTRIBUTE OR SELL COPIES OF HIS/HER THESIS BY ANY MEANS AND IN ANY FORM OR FORMAT, MAKING THIS THESIS AVAILABLE TO INTERESTED PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE IRREVOCABLE ET NON EXCLUSIVE PERMETTANT A LA BIBLIOTHEQUE NATIONALE DU CANADA DE REPRODUIRE, PRETER, DISTRIBUER OU VENDRE DES COPIES DE SA THESE DE QUELQUE MANIERE ET SOUS QUELQUE FORME QUE CE SOIT POUR METTRE DES EXEMPLAIRES DE CETTE THESE A LA DISPOSITION DES PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP OF THE COPYRIGHT IN HIS/HER THESIS. NEITHER THE THESIS NOR SUBSTANTIAL EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT HIS/HER PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE DU DROIT D'AUTEUR QUI PROTEGE SA THESE. NI LA THESE NI DES EXTRAITS SUBSTANTIELS DE CELLE-CI NE DOIVENT ETRE IMPRIMES OU AUTREMENT REPRODUITS SANS SON AUTORISATION.

ISBN 0-315-95980-0

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

Indirect standardization was used to produce estimates of the prevalence of selected risk factors and health status indicators for all health units in Ontario, based on their socio-demographic characteristics and on the relationships between these socio-demographic variables and the health variables of interest, as measured by the Ontario Health Survey (OHS). Socio-demographic variables were drawn from both the OHS and the 1986 Census of Canada. In addition, the validity of the estimates was assessed. A total of 13 combinations of predictor variables were used to estimate 18 selected health status indicators and risk factors.

Results showed that the choice of predictor variables used for estimation can have a substantial impact on the error associated with a set of estimates. As well, it was found that the dispersion of the estimates was reduced in comparison to the true values. The ranking of the estimates was found to vary from poor to good. In addition, estimates were shown to be better than using the provincial value as the estimate for each health unit value. Comparison by region showed that estimates for northern regions had somewhat more error than estimates for other regions and that addition of a geographic identifier reduced the amount of error among the estimates.

Given all these results, it was concluded that this type of methodology does not appear to be adequate to meet the information needs for health units but that other methods such as regression might produce better results.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. R.A. Spasoff and Dr. R. Nair for all their help, guidance and support. I would also like to thank Jack Boulet for his guidance and support. Finally, I would like to thank my family and friends for their support during thesis distress.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Health information takes many forms, from mortality and morbidity statistics to risk assessments and factors underlying behaviour, to name but a few. As the types of health information available are diverse, so are the needs they meet. In addition, strategies for health planning use many types of information for various groups and geographic locations. Under the Health Protection and Promotion Act, 1983, all Boards of Health are required to address the health needs of the residents of a community[1]. Among other requirements, Boards of Health are required to produce a health status report at least every five years including demographics, mortality and morbidity figures, reproductive outcomes, and risk factor prevalence. To meet these requirements, Boards of Health need health information.

While much information is available at the national/ provincial level from sources such as surveillance systems and national/provincial surveys, equivalent information for smaller geographic regions is limited or non-existent. Mortality data, which are available for small areas, provide only a limited assessment of health status[2].

Health surveys are conducted at the national or provincial level but have insufficient sample sizes to provide valid estimates for smaller populations (e.g., census tracts). The obvious solution to this lack of data would be to conduct surveys for small geographic areas. However, surveys are costly and budget restraints often preclude such endeavours. Therefore we must use existing data to the maximum extent[3].

This project uses indirect standardization to produce estimates of the prevalence of selected risk factors and health status indicators for all health units in Ontario, based on their socio-demographic characteristics and on the relationships between these socio-demographic variables and the health variables of interest, as measured by the Ontario Health Survey (OHS). In addition, the validity of the estimates is assessed. If this methodology produces accurate results, the need for expensive health surveys will diminish. Health units will be able to use census data to derive reasonably accurate estimates.

BACKGROUND: STANDARDIZATION

Throughout this thesis the term "indirect standardization" is used instead of the more accurate term "synthetic estimation". Synthetic estimation uses only part of the process of indirect standardization. The former was chosen because it is a term more familiar to epidemiologists.

Using indirect standardization, we can select socio-demographic information (e.g., from the Census) to adjust risk factor and health status data from large surveys (e.g., OHS) to estimate the levels of certain health related variables in small population groups. Throughout this thesis the term "indirect standardization" will be used instead of the more accurate term "synthetic estimation". The former was chosen because it is a term more familiar to epidemiologists.

To produce estimates, indirect standardization requires data on the target variables (e.g., smoking status, heart disease etc.,) at the national or provincial level, broken down by all the predictor variables (e.g., age, gender, home language etc.,) simultaneously, to produce mutually exclusive and exhaustive strata.

The basis for this approach is that there are strong and stable associations between certain socio-demographic factors (i.e., predictor variables) and the prevalence of risk factors[2,4] and health status indicators (i.e., target variables). As well, it is assumed that the relationships among the variables concur at both the national/ provincial and the small population level.

Stratum-specific summary proportions for the entire survey sample are calculated and then weighted according to the composition of the small area population using the following formula:

$$ESTIMATE=\frac{\sum_{k=1}^{K} p_k * Y_k}{\sum_{k=1}^{K} p_k}$$

Where:

estimate =the target variable, summarized as a proportion, mean or rate, summed across all strata for a particular small area (e.g., proportion of the small area which smokes)

p_k =the number of people in the k-th stratum of the small area population (e.g., 35-39 year old males).

Y_k =for a target variable, the proportion of people in the k-th stratum of the reference population who have the target variable (e.g., proportion of males aged 35-39 who smoke)

This process is analogous to calculating the expected number in the target area divided by the target population.

The estimates produced are based on the assumption that the distribution of the target variables varies only to the extent that the small areas differ in terms of the distribution of the demographic variables (predictors). Some error in the estimates will be introduced if the assumption is not met.

LITERATURE

1. Previous work

Known in the survey literature as synthetic estimation, early work on indirect standardization began in the 1950's at the National Center for Health Statistics (NCHS) in the United States[5]. Since then, much work has been done to evaluate the indirect standardization procedure from both a demographic and public health perspective. However, most studies of indirect standardization were done during the 1970's, with the exception of Mackenzie et al. Some recent studies have also focused on regression techniques to produce small area statistics[2,6].

Several studies in the United States have produced synthetic estimates of unemployment at the county level[7,8,9]. Gonzalez and Waksberg used Census data to produce and validate estimates of unemployment at the county level.

Comparison of the estimates produced by a variety of stratification factors was undertaken. These strata included (1) occupation, sex, race; (2) marital status, sex and race; (3) occupation, income, race and sex; (4) industry, race and sex. The best estimates were based on (1) occupation, race and sex and (2) occupation, income, race and sex. The addition of income improved the correlation of census values versus estimates of unemployment. However, the improvement was small, from 0.682 to 0.685. Most estimates were found to be a within relative 10% of the Census estimate. As well, the results showed that the estimates for areas with larger populations were associated with less error than those for smaller populations.

In a similar study, Schaible, Brock and Schnack[8] compared the errors of direct and synthetic estimates of unemployment rates for counties in Texas. Direct estimates refer to estimates derived from survey data. They found that when the sample sizes for the direct estimates were small, synthetic estimates outperformed direct estimates.

Synthetic estimates were used by the US Bureau of the Census to revise the count of the 1970 Census of Population and Housing for the population of housing units reported as vacant but actually occupied[6]. Another study by Gonzalez and Hoza[10] focussed on the estimation of dilapidated housing units for 15 selected Standard Metropolitan Statistical Area (SMSA) and for four geographic regions. Their estimation process used Census counts of tenure, race of head of household and other unspecified characteristics said to be related to the quality of the housing unit. Census estimates were used for validation. It was noted that the ability to estimate a particular characteristic increased with the size of the small area. Specifically, as the size of the area (i.e., number of housing units) increased, the amount of error diminished.

Another study by Schaible, Brock, Casady and Schnack[11] estimated five demographic variables using the Census and the United States Health Interview Survey. These included (1) percent of the population under age 1 (2) percent of the population married (3) percent of the population separated (4) percent of the population completing high school (5) percent completing college. The stratification factors included race, gender, age group, family size, industry of head of household.

Schaible et al compared the results of direct and indirect methods of estimating demographic variables at the state level. Estimates were produced using both methods. For both methods, weighted/ unweighted data, ratio adjustment for regions and alteration of the number of stratification factors (i.e., collapsing categorical variables into fewer categories) were used to produce 16 sets of estimates. Ratio adjustment refers to the correction of the state-specific estimates (i.e., population counts) to agree with the total regional population totals (i.e., states are grouped into regions). Specifically, each state estimate is weighted by the proportion of the regional population in that state to force the weighted sum of the state estimates to be consistent with the regional population.

They found that estimates for states with high proportions were too low, while estimates for low proportions were too high. As well, it was noted that indirect estimates produced by 64 strata based on age, sex and race outperformed those produced by a collapsed 16 strata categorization of the same variables. Adjustment for region was found to be beneficial for only 1 of the 5 variables, percent completing high school. For some variables, synthetic estimates were closer to the census values than the comparable direct estimates. Using correlation coefficients, it was found that estimates closest to the census values did not necessarily produce the highest r values. Unfortunately, a discussion of this was not provided.

From a public health perspective, the National Centre for Health Statistics (NCHS) conducted an investigation to compare three methods of obtaining estimates of long and short-term disability for each state for 1962-1964[12]. These methods included (1) a regression technique originally developed to estimate retail trade; (2) synthetic estimation (i.e., indirect standardization); and (3) a nearly unbiased estimator.

The nearly unbiased estimator uses the same methodology as that of indirect standardization. However, it assumed that the ratio of the stratum specific population (e.g., number of males) to the total small area population is the same or very nearly the same across all small areas. When this assumption is true, the bias reduces to zero.

Of the three methods explored, synthetic estimation produced the most promising results. The nearly unbiased estimator was found to be subject to a small amount of bias and produced unstable results. Comments concerning the regression method were not provided. Synthetic estimates were based on race, sex, age group, residence (SMSA), family income, family size, and industry of head of family (Standard Industrial Code - SIC) collapsed to 78 strata. Analysis of the validity of the results was not published.

In 1977, the NCHS investigated synthetic estimation again, producing estimates of disability and utilization of medical services for the 50 states and the District of Columbia [13]. For this investigation, race, sex, age group, family income, family size and industry of head of the family were used to produce 50 strata. Results were ratio adjusted to agree with regional estimates from the US Health Interview Survey. The purpose of the investigation was to examine the impact of altering the number of strata on the error associated with the estimates.

Estimates from the 50 strata were compared with those obtained from (1) two strata - sex; (2) four strata - age; (3) eight strata - age and sex; (4) 16 strata - race, sex and age; and (5) 16 strata - family income, sex and age. Since the purpose of this investigation was to study the impact of altering the number of stratification factors, estimates produced from the 50 strata were considered gold standards. No discussion of actual error was provided. They found that with the exception of estimates based on 2 strata (i.e., gender alone) most of the estimates for the states were within 5% of the estimates based on the 50 cell grid. In addition, they found high correlations between the estimates based on 4,8, and 16 strata and 50 strata It was also noted that the performance of some of the groupings of predictor variables varied by state.

Levy[14] used indirect standardization to produce state estimates of death rates from four causes (motor vehicle accidents, major cardiovascular-renal diseases, suicide and tuberculosis) and compared them with official state mortality statistics. The results of this investigation revealed that the validity of synthetic estimates differed among the causes of death investigated. Specifically, the investigators found that there was good agreement between the estimates and true rates of major cardio-vascular-renal diseases but poorer agreement for suicides, motor vehicle accidents and tuberculosis.

Namekata et al estimated complete and partial work disability using the 1970 US Census[15,16]. Assessments of validity were done using absolute and relative percentage differences, mean square error and Pearson correlations. The overall conclusion of the investigation was that estimation produced fairly good estimates for partial work disability, but poor results for complete disability. As well, it was shown that the dispersion of estimates was not as large as that for the direct estimates (which were considered the gold standards). In fact, the standard deviations of the estimates were between a fifth and a third the size of the direct estimates. Although correlations between indirect estimates and direct estimates were good (e.g., 0.70), the ranking of the estimates was not preserved. Namekata also found that for small areas (i.e., counties) whose true values were farther away from the mean than others, the amount of error detected was larger.

Mackenzie et al[2] evaluated the utility of synthetic estimation and regression techniques to produce estimates of 37 health status variables. Age, sex and race, drawn from the Health Interview Survey, were used to define the strata for the synthetic estimates. Gold standards were taken from a telephone survey conducted in the Baltimore Statistical Metropolitan Area.

Results showed that the dispersion of the estimates for sub-areas of Baltimore was not as great as that for the gold standards. As well, it was found that the proportional difference between the gold standards and the synthetic estimates was between 5% and 25%. The authors suggested that using regional data may produce better results than using national data as some health variables vary widely between regions. This suggests that the relationships between the socio-demographic and target variables may vary by region, depending on the homogeneity of the population.

2. Summary

Past work has shown that indirect estimation may produce poorer results for areas with smaller populations[6,7] under/over-estimation of high and low prevalence variables[10,14] and estimates with less dispersion than the true values[2,14,15] Additionally, research has shown that using the maximum number of strata available from the predictor variables produces results closer to the true values than a reduced number of strata[10,11]. Other findings indicate that results for regions may vary by the type of predictor variables used[2,11]. Schaible et al also noted that estimates with the least amount of error do not necessarily produce the highest correlation coefficient and that the ranking of the estimates from lowest to highest prevalence may not be preserved. As well, unless small areas vary in terms of the socio-demographic variables used, the estimates themselves will not vary significantly between the areas[2,11].

3. Error and Bias

Bias is defined as a "systematic deviation from the correct value of a particular variable."[17]. This may result in a consistent under/over estimation of rates, proportions etc. This is in contrast to random error wherein estimates differ from the truth unpredictably so that estimates are sometimes inflated and sometimes deflated and they average out.

Many authors note that the synthetic estimator is biased toward the mean in that the dispersion of the estimates is smaller than the dispersion of the true values. This occurs because the actual stratum-specific rates in the small areas differ from those predicted from the reference population[9,14,15]. This may occur when the data for the target and predictor variables are not drawn from the same year. As such, any relationship between target and predictor variables which varies over time will introduce bias. As well, Namekata et al have suggested that bias will be introduced when demographic characteristics and target variables have different associations among regions. Also there is always unexplained variation that cannot be systematically accounted for.

To evaluate error and bias, measures such as the mean square error can be used. The mean square error relates an estimate for each small area to its true value as given by[7]:

MSE = SUM OF THE VARIANCE + $BIAS^2$

which is equivalent to:

$$MSE(X^*) = \sum p_i^2 \sigma^2{}_{x^*} + (X_i^* - X_i)^2$$

Where:

$\sigma^2_{x^*}$ = sampling variance of the estimate

p_i = stratum specific proportion

X^* = estimate

X = true value

given that:

(a) p_i's are fixed and measured without error

(b) the covariance $(X_j, X_k)=0$, for j not equal to k

Thus, the MSE includes both chance and systematic error, and if the estimate of X is unbiased, the MSE reduces to the variance. But, the synthetic estimator has been shown to be biased in that the p_i's are subject to some error[7]. Furthermore, the available "gold standards" are often themselves subject to sampling error because they are taken from surveys. However, when estimates are derived from large surveys, sampling errors will likely be small[15]. As a result, the main sources of error will be from systematic sources. Nonetheless, the bias and small amounts of error make the MSE difficult to interpret.

To evaluate the amount of error associated with a set of estimates produced by indirect standardization, Gonzalez and Waksberg[6,7] at the United States Bureau of the Census developed a method for estimating the average mean square error (AMSE) across small areas. The AMSE was derived from the MSE. This measure has been used to compare various estimates in several investigations[6,7,9,15].

The AMSE is calculated using the following formula;

$$AMSE = \frac{\sum_{K=1}^{K} (ESTIMATE - TRUE)^2}{N}$$

The AMSE measures the average squared deviation of a set of estimates from their true values and is somewhat analogous to a variance. However, it does not provide an evaluation of the error for a particular estimate but only for a group of estimates.

By taking the square root of the AMSE, the units are returned to the same scale as that from which they were drawn (e.g., proportions). This makes the figure easier to comprehend as it is similar to a standard deviation. Herein this value will be referred to as a Root Mean Square Error (RMSE). One of the features of the RMSE it that it provides a means of identifying a set of estimates with the smallest amount of error for a particular target variable. For example, if the RMSE for set a=.03 and for set b=.04, then set a is associated with a smaller amount of error.

One of the drawbacks of the RMSE is that comparison between target variables is not possible as the RMSE reflects the underlying magnitude of the units (e.g., proportions) from which it is calculated. To standardize the RMSE, such that comparisons across target variables can be made, the RMSE can be divided by the mean of the estimates. This creates an index similar to a coefficient of variation, which gives some sense of the error across variables. This type of measure was used by Namekata et al to assess the accuracy of results.

OBJECTIVE

The objective of this project was to test the ability of indirect standardization to produce accurate estimates of the prevalence of risk factors and health status indicators in small populations.

PROJECT DESIGN

To meet the objectives of this project, six phases were undertaken (Figure 1). The first phase was to select target and predictor variables for the estimation process. All predictor variables selected had to be available in both the Ontario Health Survey data set (OHS) and the 1986 Census of Canada.

In Phase 2, all variables were operationalized. This involved derivation of some variables, locating and downloading appropriate Census files and ensuring the comparability of OHS based and Census based variables (i.e., predictor variables).

For Phase 3, two types of standardization programs were developed: (1) using only OHS data; and (2) integrating both OHS and Census data into the standardization procedure. All programs were written in SAS[18] and run on a Sun SPARCstation1.

Estimation of all target variables using <u>only</u> OHS data was undertaken in Phase 4. Gold standards, targets and predictors were drawn from this data set. The RMSE was calculated for each set of estimates in order to identify the set associated with the least amount of error. Phase 4 was exploratory in nature as all reasonable combinations of predictor variables were utilized for estimation.

Phase 5 replicated Phase 4 but with predictor variables drawn from the 1986 Census of Canada and gold standards and target variables drawn from the OHS data set. As well, only predictor variables and/or sets of predictor variables which produced the best estimates in Phase 4 were employed in Phase 5. While Phase 4 was exploratory, Phase 5 produced estimates using data which would be available to a public health unit should this methodology be used.

In Phase 6, all estimates were validated using a variety of indices. A discussion of the results follows these six phases.

Figure 1 Project Design

PHASE 1 ↓	Target and predictor variable selection
PHASE 2 ↓	Operationalization of all variables
PHASE 3 ↓	Development of standardization and validation programs
PHASE 4 ↓	Estimation of targets (OHS data)
PHASE 5 ↓	Estimation of targets (OHS and Census data)
PHASE 6	Validation

METHODS

1. Level of Analysis

Estimates were developed at the Health Unit level for Ontario for individuals aged 15 and over. Estimates were produced for all 42 health units. The health unit was selected as the level of analysis because it reflects the level at which validation was possible.

2. Sources of Data

For this project, the 1986 Census of Canada and the Ontario Health Survey (1990) were used. The Ontario Health Survey was conducted across Ontario in 1990 to collect information concerning health status and the social, economic, physical, behavioural and nutritional factors related to health and health care utilization[19]. A multi-stage stratified cluster design was used to draw the sample of Ontario residents. Sample sizes were designed to produce stable estimates at the health unit level (i.e., 42 health units) for all variables. The sample (approximately= 1100 per health unit) was surveyed across the 4 quarters of the year in order to account for seasonal differences in health related variables.

Three data collection forms were used in the survey. The first (Form 3) collected household information (e.g., age, sex, and marital status) from a household representative. Responses for the second form (Form 4) were also taken from a household representative and included questions such as health care contacts, two week disability, and hearing. For Form 4, the household representative provided proxy responses for all other household members. The third form (Form 5) was completed by each household member above age 15 and included questions about smoking, alcohol and food consumption, exercise, etc.

The response rates for the OHS differed by form. For Forms 3 and 4, the response rate was 88% and for Form 5 the response rate was 76% of individuals in those households responding to Form 4[18].

The final data set used in this investigation consisted of 46,583 respondents aged 15 and over, which represents a total of 7,774,450 in the province as a whole. All OHS data used in this investigation were weighted.

Census data available for this project were limited to the data tapes available on the mainframe computer at the University of Ottawa. Unfortunately, the 1991 Census data were not yet available on the mainframe at the time of this analysis. Accordingly, the 1986 Census data were used for Phase 5 of the project as it is the only source of data for small areas which provides all predictor variables.

In phase 4, the OHS was used for all purposes. Specifically, gold standards, stratum specific proportions of the target variables and stratum specific population counts for each health unit were taken from this data set. The data for predictor variables are much less precise than those from the Census. On the other hand, use of the same data source for all requirements eliminates errors due to temporal variation or question wording.

In phase 5, gold standards and stratum specific proportions of target variables were taken from the OHS. Stratum specific population counts were drawn from the 1986 Census. The implicit assumption of this phase is that there will be stability of predictor variables across time (i.e., 1986 to 1990).

3. Validation Indices

The validation indices used in the investigation included:

(a) Root mean square error
(b) Correlation coefficients
(c) Quintiles
(d) Provincial Index

In addition the dispersion of the estimates and the health unit specific error was investigated. The health unit specific error was calculated as follows:

$$\textit{Mean Proportion Deviation} = \frac{\sum_{j=1}^{18} \frac{|\textit{Estimate} - \textit{True Value}|}{\textit{True value}}}{18}$$

where j=18 is the number of target variables.

(A) Root mean square error

As discussed earlier, this measure was used to identify the best set of estimates for each target variable and also to determine for which types of variables estimation was most successful.

(B) Correlation Coefficients

When interest lies in assessing the association between the true values and the estimates, use of the Pearson correlation coefficient as a measure of validity is particularly appropriate. The Pearson correlation coefficient measures the amount of linear association between two variables (i.e., gold standards and estimates). Use of the Spearman rank correlation coefficient results in an examination of only the ranks.

Selection of the correlation coefficients places the absolute value of the prevalence as a secondary interest. If two variables are associated in a linear fashion, be it positive or negative, the correlations will approach -1.0 or 1.0. This measure was used by Levy and French and by Namekata.

Unfortunately, if the estimates are uniformly under- or over-estimated, r will not reflect this. As such, r alone is not a good indicator of estimation accuracy. This investigation will determine if any set of estimates is uniformly under- or over- estimated to avoid mis-interpretation.

(C) Predictive accuracy: quintiles

Another feature of the standardization process examined was the extent to which the ordering of the PHU's from lowest to highest prevalence was preserved. For this assessment, quintiles were used as a measure of predictive accuracy. At the cost of loss of some information, health units were categorized on the basis of the gold standards into quintiles for each target variable, and the ability of the estimation procedure to place each health Unit into the correct quintile was assessed.

(D) Provincial Index

A characteristic of synthetic estimates is that unless small areas vary in terms of the independent variables of interest (i.e., predictor variables), the estimates themselves will not vary significantly across the areas[2,11]. If a target variable does not vary much between health units, it would appear better to adopt the provincial value. To test whether a set of estimates was closer to the true value as opposed to the provincial value the following index was considered.

$$INDEX=\frac{\sum_{k=1}^{k} \mid \mathit{Estimate} - \mathit{Gold\ Standard} \mid}{\sum_{k=1}^{K} \mid \mathit{Provincial\ proportion} - \mathit{Gold\ Standard} \mid}$$

If the index is less than one, this indicates that there is less error when using the estimates than in assuming that the provincial value is an adequate substitute for the health unit value.

(E) Summary

Each validation index was selected to investigate a different facet of the estimation process. The RMSE was selected to identify the set of predictor variables which produced estimates with the least amount of error. The RMSE divided by the mean of the estimates was selected to identify which target variable was best estimated. To assess the linear association and the preservation of ranks, Pearson and Spearman rank correlations were used. To examine the agreement in the ranking of health units from lowest to highest prevalence quintiles were employed. To determine if the estimates were better than employing an estimate from a larger area, the index comparing the error of the estimates versus the error of the provincial value was developed.

4. Development of Estimates

All target and predictor variables were operationalized, using micro-data tapes from the Ontario Health Survey and the 1986 Census of Canada. For some variables, operationalization required aggregation, recoding, and/or construction of secondary variables. Comparison between Census variables and OHS variables was done to identify those variables common to both data sets which could be used as predictor variables.

All data were subject to many types of error including:

(a) under-coverage (missed dwellings)
(b) over-coverage (inclusion of non-residents)
(c) non-response
(d) response errors (incorrect responses)
(e) coding errors
(f) data capture errors
(g) imputation errors
(h) weighting

Precision of the predictor variables was not a factor for Census variables collected from 100% of the population, and was a relatively small problem for variables collected from 20%. The stratum-specific provincial proportions came from the OHS which had a large sample and fairly small standard errors. For example, the standard error[20] associated with an estimate of the number of smokers was 0.0008 for the province (estimated proportion=.29), 0.00465 for Algoma (estimated proportion=.31) and .0043 for Ottawa-Carleton (estimated proportion=.31). The "gold standards" were based on the OHS, and thus were actually direct or post-stratified estimates. Nevertheless, these will be referred to as the gold standards to avoid confusion with the estimates that will be calculated. Gold standards are based on samples of about 1100 per Health Unit, which produced a standard error of 0.9% for a variable with a prevalence of 10% or 90% and a standard error of 1.4% for a variable with a prevalence of 50%, for 95% confidence intervals of ± 1.8% and ± 2.9%, respectively.

5. Selection of Variables

(A) Risk factors

In order to be selected for estimation, risk factors had to be (1) important in terms of proportion of disease burden "explained"; (2) susceptible to local public health intervention; and (3) included in the OHS. As all variables were needed in dichotomous form, cut-off points used were those specified by nutritional[21] and weight[22] recommendations and Mandatory Health Programs and Service Guidelines[1]. Variables selected were:

smoking
alcohol consumption
exercise
blood pressure
Body mass index
total caloric intake
percent calories from fat
total fibre intake
use of seatbelts: driver
use of seatbelts: passenger

The percentages quoted refer to the hazardous condition, e.g., smoking, not using a seatbelt.

Specifics regarding the operationalization of each of these risk factors are presented below.

(a) Cigarette smoking

Any persons who indicated that they were occasional or current smokers were included as participating in a hazardous behaviour. The association between smoking and cancer, heart disease and other diseases is well documented.

hazard = daily or occasional smoker

(b) Number of drinks per week for all current drinkers

Nutritional recommendations from Health and Welfare Canada[20] suggest that alcohol consumption should not exceed 2 drinks per day.

hazard = greater than 14 drinks per week

(c) Physical activity status

Identifying the appropriate level of exercise for an individual is dependent on many factors such as age and health status. As well, capturing an individual's level of physical activity from a survey is problematic for although an individual may not have participated in a sport or other exercise activity, he/she may be very physically active at work or at home. Unfortunately, the OHS does not provide an opportunity to identify the overall level of physical exercise. As such, hazardous behaviour will be identified as sedentary lifestyle, which is associated with many chronic diseases.

hazard = sedentary lifestyle

(d) Self-reported hypertension status

Hypertension has been identified as a risk factor for many chronic health problems (e.g., heart disease) and life threatening medical events (e.g., stroke). As such, people who identified themselves as hypertensive will be considered to have a hazardous health status.

hazard = hypertension

(e) Body mass index

Being overweight is recognized as a risk factor for many diseases including hypertension, diabetes, cardiovascular diseases and hyperlipidemias. Guidelines set by Health and Welfare suggest that a healthy BMI lies between 20-25, with 26-27 borderline[21]. BMI's above or below 20-27 may be associated with health problems for some people. For the purpose of this investigation, BMI's exceeding 27 will be considered as hazardous.

hazard = exceeds 27

(f) Daily caloric intake

Nutritional recommendations from Health and Welfare Canada[20] for each age/sex group are shown below:

16-18	M	3200
	F	2100
19-24	M	3000
	F	2100
25-49	M	2700
	F	2000
50-74	M	2300
	F	1800
75+	M	2000
	F	1700

A variable was created to determine if an individual's daily intake was in excess of the amount suggested.

hazard = daily intake exceeds recommendation

(g) Percent calories from fat per day

Consumption of excess fat, saturated fat in particular, has been implicated in the etiology of cardiovascular disease. Nutritional recommendations from Health and Welfare Canada suggest that only 30% of all energy intake per day should be drawn from fat[20].

hazard = daily intake exceeds recommendation

(h) Seatbelt usage: Driver

A hazardous behaviour was identified as consistent failure to use a seatbelt while a driver.

hazard = does not use seatbelt every time

(i) Seatbelt usage: Passenger

A hazardous behaviour was identified as consistent failure to use a seatbelt while a passenger.

hazard = does not use seatbelt every time

(B) Health Status Indicators

To be selected for estimation, a health status indicator had to be: (1) responsible for a significant burden of ill health on the community; (2) susceptible to public health intervention, in terms of health promotion, disease prevention or provision of support services; and (3) included in the OHS. Indicators selected include:

two week disability
heart disease
presence of one or more chronic conditions
inability to perform ADL: personal care activities
inability to perform ADL: personal affairs
presence of hearing problems
self-rated health
happiness with life
activity limitation

Specifics regarding the operationalization of each of these indicators are presented below.

(a) Two week disability: Sum of cutdown/bed-days

Any reported days spent in bed which resulted in an inability to participate in usual daily activities or reductions in daily activities in the two weeks prior to the survey were coded as a poor health status indicator.

poor status = disability day in the previous two weeks

(b) Self-reported heart disease

Heart disease is the primary cause of death in Canada. As such, it is a serious health concern.

poor status = reported heart disease

(c) Number of chronic conditions

Any reported chronic condition (e.g., asthma, chronic bronchitis, heart disease, allergies) was coded as indicative of a poor health status. The OHS allowed for reports of a maximum of 8 chronic health problems.

poor status = at least one chronic health problem

(d) Activities of daily living: requires help with personal care

Any reported difficulties with performing activities of related to personal care (e.g., bathing) were coded as a poor health status indicator.

poor status = difficulty with personal care

(e) Activities of daily living: requires help with personal affairs

There are also known as "instrumental" activities of daily living. Any reported difficulties with performing activities related to personal affairs (e.g., taking care of personal finances) were coded as a poor health status indicator.

poor status = difficulty with personal affairs

(f) Activity limitation

Anyone who reported that he/she was limited at home, school, work or leisure by a health problem was included in the poor status category.

poor status = any reported limitations

(g) Hearing problems

Hearing was coded as problematic when an individual reported he/she could not hear the spoken word. Three variables were combined to create a new variable which consisted of all people with any type of self-reported hearing problem.

poor status = any hearing problem

(h) Self-rated health

Studies have shown that self-rated health is a fairly valid indicator of an individual's health status. As such, responses of fair or poor self-rated health were coded as poor health status.

poor status = self-reported fair or poor health

(i) Self-rated happiness

Responses of unhappy or very unhappy were coded as poor health status.

poor status = self-reported unhappy or very unhappy

(C) Predictor Variables

To be included in the study, predictor variables had to be (1) known to be associated with one or more target variables (risk factors or health status indicators); (2) included in the Census; and (3) included in the Ontario Health Survey. The last two sources had to be compatible in terms of question wording and coding.

In all cases, the joint distribution of predictor variables was needed. Unfortunately, as the number of variables increases, the number of combinations became large. Especially when using survey data, the number of predictor variables that can be used at any one time is limited. When proportions are based on a small number of respondents, the precision of the estimates is diminished. As a result, the number of combinations of predictor variables which can be used is limited. When Census data were used, the range of predictor variables was limited by the public use tables available from Statistics Canada. The lack of microdata at the health unit level (i.e., individual responses to items) precluded the construction of additional tables beyond those provided. All of these practical limitations were used to guide the selection of predictor variables.

Comparison of the socio-demographic variables available from the OHS and the Census resulted in the selection of following variables for this investigation:

age
gender
marital status
home language
level of completed education
occupation
household composition

As all Census data were aggregated, operationalization of these variables was limited by the number of categories available from both the Census tables and the OHS. Where possible, the maximum number of categories available from both sources were used. The specifics regarding each variable are presented below.

(a) Age: 15-24
25-34
35-44
45-54
55-64
65 and over

(b) gender: male
female

(c) marital status:
married
single
widowed
separated/divorced

(d) home language:
English
French
other European
Asian
other

(e) level of completed education:
primary only
some secondary
completed secondary
college
some university
completed university

(f) occupation:
managerial/professional
clerical
sales
services
primary occupations
processing
construction
transportation
materials handling
other
not applicable (i.e., not in paid labour force)

(g) household composition:
couple with or without children
single parent
individual alone
individual and others
other and non-family households

The Basic Summary Tables were used as the source of Census data. The crosstabulations they offered included age and/or gender with many of the other predictor variables identified above. As such, the types of distributions and joint distributions used in this investigation included:

(1) age
(2) gender
(3) age and gender

(4) marital status
(5) age and marital status
(6) age, gender and marital status

(7) home language

(8) completed education
(9) age and completed education
(10) age, gender and completed education

(11) occupation
(12) gender and occupation

(13) household composition

To each Census table used in this investigation a health unit identifier was added. This was done using a list provided by the Ontario Ministry of Health specifying the Census divisions and sub-divisions included in each public health unit.

(D) Missing Values

Census data did not have missing values (i.e., missing values were already imputed). The Census uses "donor" records for imputation[23]. Donor records are selected at random from among a group of consistent records which are similar to the record with a missing value in terms of a number of related characteristics. Given the large number of records the error from imputation is thought to extremely small.

For all variables, records from the OHS with a missing value were excluded from the calculations. Table 1 shows the level of non-response for each variable. A comparison of the responders versus the non-responders was done using a number of demographic variables (i.e., age, gender, marital status, home language and education). For all high non-response variables, results showed that non-responders tended to be older and have completed less education than responders.

The options for dealing with missing values for the targets include:

(1) delete a record with missing values completely from analysis involving that variable

(2) assign the same response to each record with a missing value (i.e., create a new value or assign a pre-existing value)

(3) impute a value for the missing records

The assumption for option 2 is that each respondent with a missing value for a particular variable is the same as all others with missing values and also the same as all respondents who actually gave the response chosen for imputation.

The assumptions for option 3 would depend on the type of imputation employed. For instance, if imputation was based on the age-sex-household composition of a person, this would assume that all people with a certain socio-demographic background would likely show the same behaviour/ characteristic. Options 2 and 3 were thought to introduce additional error on top of the error from the missing values alone and as such, option 1 was utilized.

Through exclusion, the standardization procedure will produce results which assume that the association between the target and predictor variables for cases with missing variables is the same as that found for the records with complete data. For each estimation procedure, only those records with missing values for the variables being used were excluded. For example, the number of records used to estimate sedentary lifestyle was smaller than the number used to estimate chronic health problems because the number of missing values differed.

Table 1 : Percent missing values, target and predictor variables, ages 15 and over, OHS

	percent
age	0.00
sex	0.00
marital status	0.40
education	0.80
home language	0.00
occupation	0.02
household composition	0.00
Excess consumption of fat	7.50
Chronic health problems	0.00
Sedentary lifestyle	18.80
Excess consumption of calories	7.50
Smoking status	3.30
High BMI	11.30
Hypertension	1.00
Two-week disability	1.20
Excess consumption of alcohol	15.30
Self-rated health	1.00
Seatbelt usage: passenger	7.30
Activity limitation	0.40
Heart disease	1.00
Hearing problems	0.10
Seatbelt usage: driver	15.90
ADL: personal affairs	0.40
ADL: personal care	0.30
Self-rated happiness	0.10

RESULTS

1. Root Mean Square Error

Estimates for each health unit (total=42), target variable (total=18), and combination of predictor variables (total=12) are not presented because the total number of estimates (8568) was thought to be too lengthy. However, an example of the estimates and true values for excess consumption of fat predicted by home language (OHS) is presented in Table 2.

Initially, predictor variables were drawn from the OHS and all selected combinations of variables were used. Table 3 presents the "best" (smallest RMSE) and "worst" (largest RMSE) estimates for each target variables. For all targets the error associated with the largest RMSE was at least 5% and no more than 67% greater than the smallest RMSE (Table 3). For 10 of the target variables, there was less than a 25% improvement in the error from worst to best sets of estimates. For variables with a small ratio, the different types and combinations of predictor variables resulted in few differences in the magnitude of the error.

Table 2: True, provincial and estimated percentage of excess consumption of fat (OHS)

PHU		TRUE	PROV	ESTIMATE
1	ALGOMA	93.4	86.2	87.6
2	BRANT COUNTY	88.5	86.2	87.3
3	BRUCE-GREY-OWEN	93.4	86.2	87.8
4	DURHAM REGIONAL	91.2	86.2	87.2
5	EASTERN ONTARIO	90.8	86.2	88.8
6	ELGIN-ST. THOMAS	90.9	86.2	87.5
7	WINDSOR-ESSEX	83.9	86.2	86.5
8	HALDIMAND-NORFOLK	91.6	86.2	87.7
9	HALIBURTON-KAWARTHA	94.4	86.2	87.9
10	HALTON REGIONAL	88.5	86.2	87.5
11	HAMILTON-WENTWORTH	87.4	86.2	86.3
12	HASTINGS +	91.0	86.2	87.9
13	HURON COUNTY	93.3	86.2	87.7
14	KENT-CHATHAM	92.7	86.2	87.4
15	KINGSTON +	88.3	86.2	87.3
16	SARNIA-LAMBTON	86.2	86.2	87.7
17	LEEDS, GRENVILLE +	91.6	86.2	87.9
18	MIDDLESEX-LONDON	86.5	86.2	86.7
19	MUSKOKA +	91.8	86.2	87.9
20	NIAGARA REGIONAL	89.5	86.2	87.1
21	NORTH BAY	90.9	86.2	88.3
22	NORTHWESTERN	90.2	86.2	87.6
23	OTTAWA-CARLETON	86.8	86.2	88.0
24	OXFORD COUNTY	90.8	86.2	87.8
25	PEEL REGIONAL	83.5	86.2	84.5
26	PERTH DISTRICT	91.4	86.2	87.7
27	PETERBOROUGH	93.0	86.2	87.9
28	PORCUPINE	92.1	86.2	89.5
29	RENFREW COUNTY	92.6	86.2	87.9
30	SIMCOE COUNTY	90.3	86.2	87.9
31	SUDBURY	91.0	86.2	88.3
32	THUNDER BAY	87.8	86.2	87.4
33	TIMISKAMING	91.6	86.2	88.5
34	EAST YORK	82.2	86.2	85.4
35	ETOBIKOKE	80.7	86.2	85.5
36	NORTH YORK	71.9	86.2	83.1
37	SCARBOROUGH	81.7	86.2	83.1
38	TORONTO	80.5	86.2	83.5
39	YORK	79.2	86.2	84.0
40	WATERLOO	88.4	86.2	86.3
41	WELLINGTON +	90.3	86.2	87.4
42	YORK REGIONAL	85.3	86.2	85.8

Conversely, for target variables with larger ratios, the choice of predictor variables had a more substantial impact on the error and hence the estimates themselves.

Of the predictor variables, age, gender and education as a group produced 5 of the best sets of estimates, followed by home language (4 sets) and education (4 sets). Age, household composition and gender each produced 3 of the worst sets of estimates. Only the "best" estimates are presented in the remainder of the thesis. The best set refers to those shown to have the smallest RMSE. These "best" estimates are repeated drawing predictors from the Census.

As stated earlier, the RMSE/$X_{est(mean)}$ allows for comparison across variables. The lowest RMSE/$X_{est(mean)}$ was obtained for excess consumption of fat for both OHS (.043) and Census (.044) estimates (Table 4). Self-rated happiness, for both OHS (.478) and Census (.444), showed the largest amount of error.

Table 3: Comparison of OHS estimates with smallest and largest RMSE for each target variable

	Predictor variables	RMSE smallest (s) largest (l)	Ratio (l) / (s)
Excess consumption of fat	Home language Occupation	3.770 5.661	1.50
Chronic health problems	Home language Education	3.623 4.139	1.14
Sedentary lifestyle	Age,gender,education Household composition	2.511 3.690	1.47
Excess consumption of calories	Gender,occupation Household composition	4.032 5.250	1.30
Smoking	Language Age	3.024 3.891	1.28
High BMI	Age,gender,education Age	2.513 3.314	1.31
Hypertension	Age, education Home language	1.097 1.473	1.35
Two-week disability	Home language Age,gender,education	2.102 2.246	1.07
Excess consumption of alcohol	Age, marital status Education	1.546 2.361	1.54
Self-rated health	Education Marital status	2.037 3.385	1.67
Seatbelt usage: passenger	Occupation Marital status	7.486 8.429	1.12
Activity limitation	Education Gender	1.960 2.312	1.18
Heart disease	Age,gender,education Gender	1.059 1.301	1.23
Hearing problems	Age,gender,education Age,gender	1.322 1.600	1.20
Seatbelt usage: driver	Occupation Age	7.879 8.915	1.14
ADL: personal affairs	Education Gender	0.987 1.079	1.10
ADL: personal care	Age,gender,education Home language	0.565 0.602	1.06
Self-rated happiness	Education Household composition	0.292 0.306	1.05

For eight variables (i.e., alcohol, self-rated health, seatbelt usage:passenger, activity limitation, adl:personal affairs, adl:personal care, seatbelt usage: driver and self-rated happiness) the RMSE/$X_{est(mean)}$ among the Census estimates was smaller than that for the comparable OHS set of estimates. For the other 10 variables the Census estimate was associated with greater error than the OHS. Examination of the ratio of the OHS RMSE/$X_{est(mean)}$ to the Census RMSE/$X_{est(mean)}$ showed that for most variables, the differences in the RMSE/$X_{est(mean)}$ between the OHS and Census were small. For variables such as excess consumption of calories the ratio was larger than others. In this case, OHS estimates of excess consumption of calories were shown to be appreciably better than Census estimates.

Comparison between the OHS and Census estimates revealed that for 9 variables the ranking of the RMSE/$X_{est(mean)}$, from lowest to highest was not the same for the Census as that for the OHS (Table 4). The difference in rank was as small as ±1 and as large as ±3 (e.g.,calories, BMI).

Correlations between the OHS estimates and Census estimates (Table 4) were high with Pearson correlations ranging from .663 (seatbelt usage:passenger) to .980 (excess consumption of calories). Spearman correlations ranged from .668 (seatbelt usage:passenger) to .965 (Hypertension). Overall, Pearson correlations tended to be higher than the comparable Spearman correlations.

Table 4: Root mean square error divided by the mean of the estimates corresponding to the smallest RMSE (OHS) for each target variable, and Pearson and Spearman correlations, OHS best estimates and Census estimates

	OHS RMSE/ $X_{est(mean)}$ (a)	Census RMSE/ $X_{est(mean)}$ (b)	Ratio (a/b)	Pearson	Spearman
Excess consumption of fat	.043*	.044	.98	.980	.964
Chronic health problems	.052*	.053	.98	.909	.844
Sedentary lifestyle	.095*	.116	.82	.888	.861
Excess consumption of cal	.097*	.140	.69	.919	.876
Smoking	.102*	.108	.94	.922	.874
High BMI	.104*	.107	.97	.914	.901
Hypertension	.110*	.128	.86	.959	.965
Two-week disability	.158*	.159	.99	.910	.909
Excess consumption of alcohol	.180	.178*	1.01	.810	.842
Self-rated health	.181	.175*	1.03	.871	.860
Seatbelt usage: passenger	.198	.168*	1.18	.663	.668
Activity limitation	.223	.213*	1.05	.885	.880
Heart disease	.238*	.250	.95	.963	.948

* Flags the better estimates

Table 4: Root mean square error divided by the mean of the estimates corresponding to the smallest RMSE (OHS) for each target variable and Pearson and Spearman correlations, OHS best estimates and Census estimates (continued)

	OHS RMSE/ $X_{est(mean)}$ (a)	Census RMSE/ $X_{est(mean)}$ (b)	Ratio (a/b)	Pearson	Spearman
Hearing problems	.315*	.325	.97	.960	.929
Seatbelt usage: driver	.344	.271*	1.27	.696	.680
ADL: personal affairs	.375	.361*	1.04	.863	.851
ADL: personal care	.429	.423*	1.01	.895	.851
Self-rated happiness	.478	.444*	1.08	.898	.872
Total *	10	8			

* Flags the better estimates

2. Dispersion

Scatterplots for each target variable are not presented. However, the example of excess consumption of fat (OHS) is presented in Figure 2. The reader should note the smaller dispersion for the estimates in comparison to the true values.

As expected, not one of the ranges of the best estimates was as large as those for the true values (Table 5). With one exception (hypertension), OHS estimates showed a dispersion less than 50% that of the true values. Hypertension estimates (.84) showed the greatest amount of dispersion in comparison to the true values while seatbelt usage: passenger (.12) had the smallest relative dispersion.

Among the Census estimates, hypertension (.94) showed the greatest dispersion in comparison to the true values. The least amount of relative dispersion was found for self-rated happiness (.14). Ten of the Census estimates were more greatly dispersed than the OHS.

For the OHS estimates, the smallest amounts of relative dispersion were shown for the 2 variables with the largest true dispersion. This suggests that for target variables with the greatest dispersion, estimation produces a substantially reduced dispersions. However, the same was not true among the Census estimates. Closer inspection shows that the variables with the largest absolute dispersion (i.e., seatbelt usage: passenger and seatbelt usage: driver) did not have the smallest relative dispersions. As well, for variables such as fat and calories, which also had large absolute dispersions, the reduction in the dispersion among the estimates was not as large as that for seatbelt usage (passenger or driver).

Figure 2 True and estimated percent excess consumption of fat, OHS

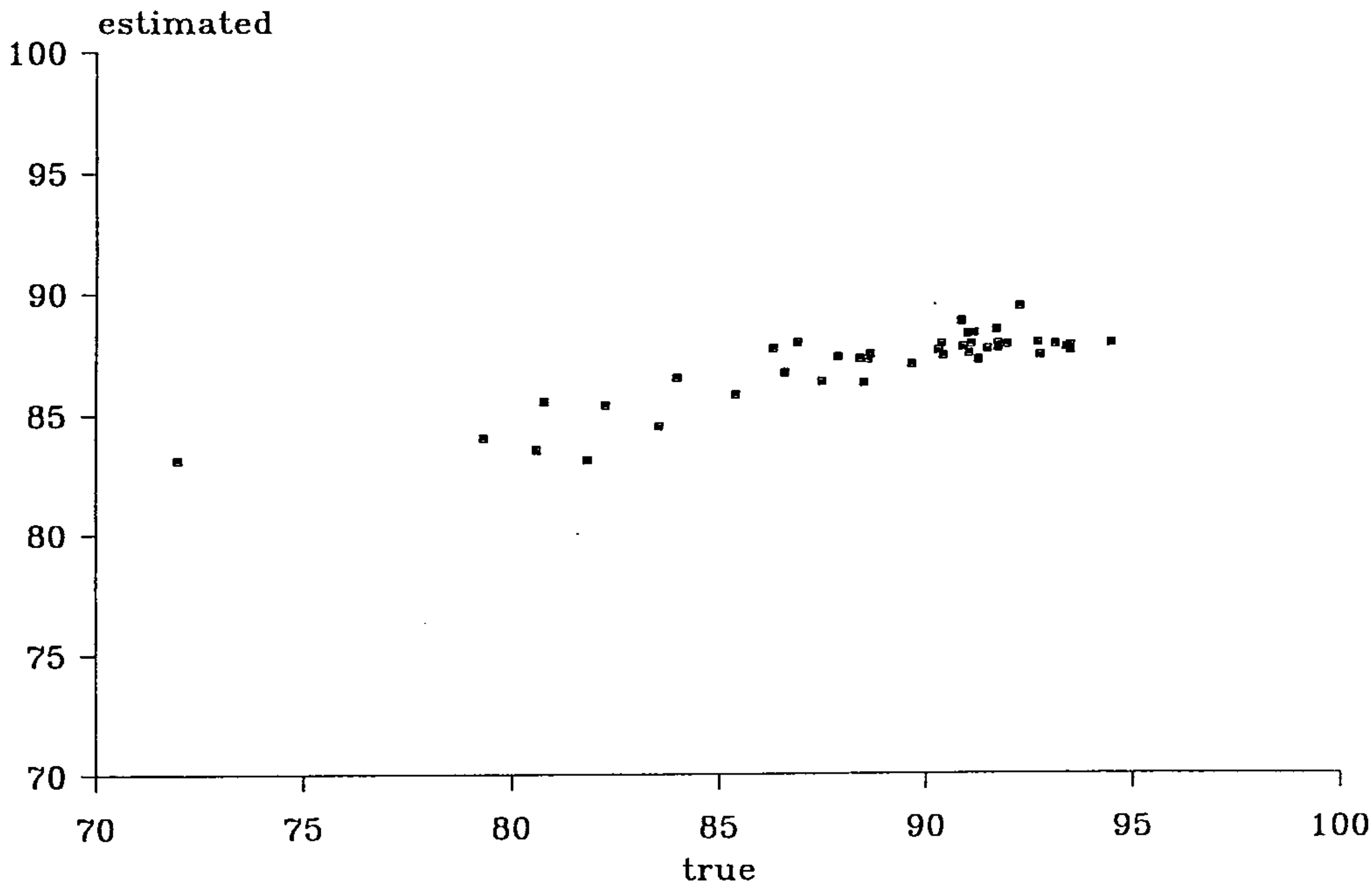

Table 5: Comparison of the dispersion of true values, and OHS best estimates and Census estimates

	True			OHS: estimates				Census			
	Min	Max	Range (T)	Min	Max	Range (O)	Ratio (O/T)	Min	Max	Range (C)	Ratio (C/T)
Excess consumption of fat	71.96	94.43	22.46	82.88	89.49	6.61	.29	75.49	83.27	7.78	.35
Chronic health problems	62.69	79.79	17.10	66.60	70.03	3.43	.20	66.51	70.14	3.63	.21
Sedentary lifestyle	21.49	40.79	19.30	29.43	34.73	5.30	.27	30.85	36.48	5.63	.29
Excess consumption of calories	33.16	53.37	20.21	39.92	44.97	5.04	.25	36.84	42.20	5.64	.28
Smoking	21.35	39.66	18.31	26.57	32.67	6.10	.33	25.69	32.46	6.77	.37
High BMI	18.76	30.56	11.80	21.78	25.39	3.61	.31	22.08	25.57	3.49	.29
Hypertension	7.59	12.60	5.01	7.55	11.77	4.22	.84	7.17	11.86	4.69	.94
Two-week disability	9.24	18.70	9.46	11.97	13.98	2.01	.21	11.96	13.94	1.98	.21
Excess intake of alcohol	4.66	11.57	6.91	8.16	9.24	1.08	.16	8.31	9.37	1.06	.15
Self-rated health	6.04	16.68	10.64	8.61	13.29	4.68	.44	9.49	13.58	4.09	.38
Seatbelt usage: passenger	24.19	58.42	34.23	35.40	39.48	4.08	.12	30.61	40.72	10.11	.29
Activity limitation	4.64	14.81	10.17	7.25	9.96	2.71	.27	7.79	10.12	2.32	.23
Heart disease	1.76	8.00	6.24	3.10	5.49	2.39	.38	2.96	5.61	2.65	.42
Hearing problems	1.74	7.71	5.97	2.96	5.18	2.22	.37	2.75	5.19	2.44	.41
Seatbelt usage: driver	12.29	42.88	30.59	20.34	24.90	4.56	.15	20.34	27.71	7.08	.23
ADL: personal affairs	1.08	5.82	4.74	1.98	3.14	1.16	.24	2.22	3.22	1.00	.21
ADL: personal care	.43	3.64	3.21	.96	1.59	.63	.20	.96	1.61	.65	.20
Self-rated happiness	.07	1.92	1.85	.43	.74	.31	.17	.49	.76	.27	.14

3. Correlation Coefficients

Pearson correlation coefficients between true values and OHS best estimates (Table 6) ranged from .264 (self-rated happiness) to .885 (excess consumption of fat). A total of 11 target variables had r values in excess of .500. For Census estimates, the Pearson correlations ranged from .153 (ADL:personal care) to .804 (fat). In total, 10 target variables had a correlation of .500 or better. Comparison between the OHS and Census estimates showed that for 4 variables (alcohol, hearing, seatbelt usage:driver, and self-rated happiness), the Census Pearson correlation was greater than the comparable OHS correlation.

Spearman correlations showed similar but slightly lower results ranging from .186 (excess consumption of alcohol) to .787 (excess consumption of calories) among the OHS estimates, with eight in excess of .500. For the Census estimates, the correlation ranged from .215 (ADL:personal care) to .762 (excess consumption of fat) with nine in excess of .500. Six Census Spearman correlations were greater than those for the OHS.

Overall, the correlations reveals mixed results. The linear association between the true values and the estimates is quite good for some variables, but quite poor for others. Spearman correlations show that the ranking of the estimates is good for some variables, but quite poor for others.

Table 6 : Pearson and Spearman correlation coefficients for each target variable, between true value and OHS best estimates and Census estimates

	Pearson OHS	Census	Spearman OHS	Census
Excess consumption of fat	.885*	.804	.742	.762*
Chronic health problems	.494*	.439	.408*	.386
Sedentary lifestyle	.470*	.351	.490*	.316
Excess consumption of calories	.803*	.745	.787*	.647
Smoking	.650*	.565	.571*	.523
High BMI	.712*	.607	.681*	.587
Two-week disability	.374*	.372	.482*	.364
Excess consumption of alcohol	.302	.316*	.186	.288*
Self-rated health	.531*	.524	.402	.497*
Hypertension	.591*	.503	.577*	.500
Seatbelt usage: passenger	.663*	.639	.668*	.604
Activity limitation	.591*	.568	.564*	.541
Heart disease	.517*	.467	.432*	.347
Hearing problems	.523	.543*	.460	.509*
Seatbelt usage: driver	.650	.666*	.602	.611*
ADL: personal affairs	.363	.471*	.446*	.443
ADL: personal care	.302*	.153	.282*	.215
Self-rated happiness	.264	.309*	.264	.320*
Total *	13	5	12	6

* Flags better value

4. Quintiles

The number of estimates placed in the correct quintile was low (Table 7). In fact, for both OHS and Census estimates, fewer than half were placed in the correct quintile. The largest number of correct classifications was 20 (calories) for the OHS and also 20 (hypertension) for the Census, out of 42 Health units. Both alcohol and adl:personal care from the OHS had only 11 of 42 estimates placed in the correct quintile. Among Census estimates, adl:personal affairs and ADL:personal care placed 6 health units in the correct quintile.

For six variables (i.e., hypertension, alcohol, self-rated health, activity limitation, hearing problems and seatbelt usage:driver), the number of Census estimates in the correct quintile exceeded the number for the OHS.

By chance alone, 8 of 42 estimates would be expected to be in the correct quintile. From the table, it can be seen that overall, the correct classification was sometimes less than what would be expected from chance and at best not much better than chance alone.

Comparison with Pearson correlation coefficients shows some consistency between r and the number in the correct quintile, for variables with the highest and lowest values. Specifically, variables with the highest and lowest r values tended to have the highest and lowest number of estimates in the correct quintile. However, for mid-range results, the results are less consistent.

Comparison between the quintiles and the Spearman correlations shows similar results, in that higher Spearman correlations are associated with a larger number of correct classifications. However, examination of the quintiles reveals an overall poor classification into ranks, while the Spearman correlations reveal from poor to good preservation of the ranks. The quintiles and Spearman measure different facets of estimation. While quintiles examine the estimates in terms of 5 groups, Spearman correlations examine the estimates as a whole group.

Table 7 : Number of estimates in the correct quintile for each target variable, OHS best estimates and Census estimates

	OHS (#)	Census (#)	
Excess consumption of fat	19	19	
Chronic health problems	16*	14	
Sedentary lifestyle	13*	11	
Excess consumption of calories	20*	18	
Smoking status	13*	11	
High BMI	13*	10	
Hypertension	16	20*	
Two week disability	14*	8	
Excess consumption of alcohol	11	12*	
Self-rated health	14	17*	
Seatbelt use:passenger	18*	13	
Activity limitation	12	15*	
Heart Disease	17*	16	
Hearing problems	13	16*	
Seatbelt use:driver	15	16*	
ADL:personal affairs	15*	6	
ADL:personal care	11*	6	
Self-rated happiness	14*	11	
Total *	11	6	1 tie

* Flags greater number in correct quintile

5. Provincial Index

Comparison of the error associated with estimates to the error associated with using the provincial value as a substitute for a particular health unit value, showed that for all OHS estimates, the error from using the estimates was less than that from using the provincial value (Table 8). The index values for the OHS estimates ranged from .70 (fat) to .98 (alcohol). This indicates that for fat there was a 30% reduction, on average, in the error when using the estimates for excess fat consumption, but only a 2% improvement, on average, when using the alcohol estimates.

While 15 of the index values for the Census estimates were less than 1, for 3 variables (i.e., sedentary lifestyle, calories and seatbelt usage:passenger) the index exceeded 1. The index ranged from .74 (fat) to 1.12 (calories). This shows that there was, on average, a 26% improvement in error when using the census estimates of fat. However, for calories, the index value shows that the provincial value is, on average, closer to the true values than are the estimates.

For 3 variables (i.e., two-week disability, alcohol, and seatbelt usage:driver), the Census index was smaller than the comparable OHS index. In many cases, the improvement was less than 10% (8 cases).

For the most part, the estimates were shown to be better than using the provincial value as a substitute. However, the improvement, at best, was only 30%.

Table 8 : Estimates versus provincial value index for each target variable, OHS best estimates and Census estimates

	OHS	Census	
Excess consumption of fat	.70*	.74	
Chronic health problems	.88*	.90	
Sedentary lifestyle	.90*	1.10	
Excess consumption of calories	.75*	1.12	
Smoking status	.78*	.82	
High BMI	.77*	.80	
Hypertension	.71*	.81	
Two-week disability	.97	.96*	
Excess consumption of alcohol	.98	.97*	
Self-rated health	.89	.89	
Seatbelt usage: passenger	.90*	1.04	
Activity limitation	.84*	.86	
Heart disease	.86*	.89	
Hearing problems	.85	.85	
Seatbelt usage: driver	.92	.82*	
ADL: personal affairs	.90*	.92	
ADL: personal care	.94*	.97	
Self-rated happiness	.90*	.93	
Total *	13	3	2 ties

* Flags smaller index value

6. Public Health Unit Specific Error

The results of this investigation have been somewhat poor. Previous research has shown that estimates for areas with smaller populations tend to be associated with more error than estimates for areas with larger populations. Table 9 presents for each target variable the correlation between the mean deviation and the health unit population. With one exception (Census calories) all the OHS and Census correlations were low. As well, for some targets, the correlations were in the opposite direction from that expected. This suggests that for this investigation, estimates for small populations do not tend to be associated with more error.

A reason for poor results may be that the estimates for some health units or regions were substantially poorer than for other health units or regions. This hypothesis is examined in Tables 10 to 13, which present results across target variables for each health unit.

Inspection of Tables 10 and 11 reveals that no one health unit had consistently the best or the worst estimates. Comparison between the OHS and Census reveals that for two-week disability and excess consumption of alcohol, the same health units (Sarnia-Lambton and Kent-Chatham, respectively) had the smallest mean percent value. Comparison of the largest values shows that for 12 of 18 variables, the same health unit had the largest mean proportion deviation for both OHS and Census estimates.

The mean relative deviation for the OHS estimates (Table 12) ranged from .092 (Elgin St-Thomas) to .509 (Etobicoke). For most health units mean deviations were between .100 and .190 of the size of the true value. Among the Census estimates, the mean deviation ranged from .091 (Kent-Chatham) to .549 (Etobicoke) with most having a mean deviation of .100 to .190 (Table 13). Most medians were smaller than the means, suggesting a positive skewness in the distribution of the deviations. The larger means are in fact explained by some unusually large deviations. For example, the largest mean, obtained for Etobicoke, is explained by the large deviation of the estimate for self-rated happiness. This is due to the low prevalence of this variable. The true value was .070, the estimate .549 and the provincial value .578. In this case, it may be that the gold standard was poor, contributing to the larger deviation. Because the dispersion of the estimates is not as great as that of the true values, for variables with a low prevalence, this tendency can result in an unusually high mean deviation value.

Comparison between the OHS and Census revealed that the Census estimates tended to have larger mean deviations than the comparable OHS estimates. However, for 16 of the 42 health units, the Census mean deviation was smaller than that for the OHS. This finding is consistent with that for the RMSE, in that Census estimates tended to show more error than did OHS estimates.

Examination of the mean deviations does not show a striking pattern by provincial region. However, the northern health units (i.e., Algoma, Muskoka, North Bay, North Western, Porcupine, Sudbury, Thunder Bay and Timiskaming) tended to have mid to high range mean deviation values. This suggests that for these health units the relationship between the target and predictor variables may be different from the relationships for the province as a whole.

Table 9 : Correlation between the mean relative deviation of estimates and PHU populations, for each target variable, OHS best estimates and Census estimates

	OHS	Census
Excess consumption of fat	-.0391	.0674
Chronic health problems	-.0596	-.0133
Sedentary lifestyle	-.0684	-.2280
Excess consumption of calories	-.0339	-.5364
Smoking status	.0578	.0954
High BMI	-.1129	-.0967
Hypertension	-.0476	-.1206
Two-week disability	.0781	.1008
Excess consumption of alcohol	.0605	.0670
Self-rated health	.0031	-.0183
Seatbelt usage: passenger	-.0003	.0244
Activity limitation	.0491	.0829
Heart disease	.0560	-.0100
Hearing problems	.2297	.1791
Seatbelt usage: driver	.2672	.1869
ADL: personal affairs	-.0055	.0295
ADL: personal care	-.0692	-.0568
Self-rated happiness	.0457	.0655

Table 10 : Smallest and largest mean relative deviation of the estimates from the true values and corresponding health units, OHS best estimates

	Smallest		Largest	
	%	PHU	%	PHU
Excess consumption of fat	.003	Middlesex-London	.152	Kingston
Chronic health problems	.002	Scarborough	.130	Niagara Regional
Sedentary lifestyle	.006	Hamilton-Wentworth	.224	Thunder Bay
Excess consumption of calories	.001	Toronto	.260	Scarborough
Smoking status	.002	Ottawa-Carleton	.329	York Regional
High BMI	.000	Wellington	.218	Peterborough
Hypertension	.019	Hamilton-Wentworth	.405	Haliburton-Kawartha
Two-week disability	.003	Sarnia-Lambton	.460	Hastings
Excess consumption of alcohol	.003	Kent-Chatham	.842	North York
Self-rated health	.008	Peel Regional	.616	York Regional
Seatbelt usage: passenger	.007	Windsor-Essex	.469	Halton Regional
Activity limitation	.002	Durham Region	.601	Halton Regional
Heart disease	.003	York	.954	York Regional
Hearing problems	.009	Algoma	.701	Scarborough
Seatbelt usage: driver	.040	Peterborough	.741	Toronto
ADL: personal affairs	.001	Haliburton-Kawartha	1.039	Hastings
ADL: personal care	.000	Kent-Chatham	1.051	Brant County
Self-rated happiness	.000	Haldimand-Norfolk	6.802	Etobicoke

Table 11 : Smallest and largest mean relative deviation of the estimates from the true values and corresponding health units, Census

	Smallest %	PHU	Largest %	PHU
Excess consumption of fat	.004	York Regional	.262	Peterborough
Chronic health problems	.004	Algoma	.132	Niagara Regional
Sedentary lifestyle	.006	Brant	.296	Thunder Bay
Excess consumption of calories	.006	Windsor-Essex	.223	Windsor-Essex
Smoking status	.001	Peterborough	.301	York Regional
High BMI	.002	Durham	.230	Peterborough
Hypertension	.008	Niagara Regional	.457	Haliburton-Kawartha
Two-week disability	.002	Sarnia-Lambton	.466	Hastings
Excess consumption of alcohol	.006	Kent-Chatham	.865	North York
Self-rated health	.001	Hamilton-Regional	.582	Kingston
Seatbelt usage: passenger	.011	Peel Region	.525	Ottawa-Carleton
Activity limitation	.012	Toronto	.742	Halton Region
Heart disease	.013	North Western	.857	York Regional
Hearing problems	.006	North Bay	.650	Scarborough
Seatbelt usage: driver	.002	Haliburton-Kawartha	.518	Halton Region
ADL: personal affairs	.000	Etobicoke	1.145	Halton Region
ADL: personal care	.024	Scarborough	1.657	Brant
Self-rated happiness	.005	Oxford County	7.528	Etobicoke

Table 12 : Mean and median proportion deviation of estimates from true values, for each PHU, OHS best estimates

	Mean	Median
Elgin-St.Thomas	.092	.096
Kent-Chatham	.098	.052
Renfrew County	.101	.083
Wellington	.104	.092
Leeds,Grenville	.108	.091
Eastern Ontario	.109	.059
Peel Regional	.117	.062
Simcoe County	.122	.085
Niagara Regional	.124	.102
East York	.125	.081
Hamilton Regional	.126	.063
Middlesex-London	.131	.042
Windsor-Essex	.134	.094
Oxford County	.135	.081
Sarnia-Lambton	.136	.094
North Bay	.136	.112
Kingston	.138	.109
Muskoka	.141	.115
Peterborough	.145	.130
Timiskaming	.146	.102
Bruce-Grey-Owen	.148	.114
Haliburton-Kawartha	.150	.089
Haldimand-Norfolk	.153	.076
Durham Regional	.153	.098
Toronto	.155	.089
York	.155	.164
Algoma	.158	.123
Waterloo	.164	.077
North Western	.172	.112
Ottawa-Carleton	.176	.106
North York	.186	.121
Porcupine	.187	.174
Sudbury	.187	.184
Hastings	.191	.105
Thunder Bay	.221	.125
Scarborough	.222	.143
York Regional	.224	.098
Brant County	.229	.080
Perth District	.239	.134
Halton Regional	.296	.135
Huron County	.301	.168
Etobicoke	.509	.077

Table 13 : Mean and median proportion deviation of estimates from true values, for each PHU, Census estimates

	Mean	Median
Elgin-St.Thomas	.103	.094
Kent-Chatham	.091	.063
Renfrew County	.116	.125
Wellington	.111	.091
Leeds,Grenville	.131	.110
Eastern Ontario	.114	.053
Peel Regional	.114	.062
Simcoe County	.112	.083
Niagara Regional	.130	.100
East York	.129	.109
Hamilton Regional	.142	.085
Middlesex-London	.154	.085
Windsor-Essex	.164	.101
Oxford County	.131	.094
Sarnia-Lambton	.126	.095
North Bay	.122	.097
Kingston	.147	.103
Muskoka	.127	.073
Peterborough	.154	.160
Timiskaming	.146	.141
Bruce-Grey-Owen	.158	.118
Haliburton-Kawartha	.171	.094
Haldimand-Norfolk	.150	.101
Durham Regional	.164	.063
Toronto	.117	.086
York	.130	.087
Algoma	.183	.154
Waterloo	.175	.081
North Western	.193	.149
Ottawa-Carleton	.185	.098
North York	.172	.102
Porcupine	.202	.201
Sudbury	.178	.149
Hastings	.236	.152
Thunder Bay	.210	.139
Scarborough	.226	.147
York Regional	.198	.071
Brant County	.241	.087
Perth District	.244	.144
Halton Regional	.346	.181
Huron County	.287	.165
Etobicoke	.549	.113

7. Addition of north/south distinction

In view of the foregoing, each target variable was estimated again using the predictor variables associated with the smallest RMSE, but adding a north/south identifier. The purpose of this procedure was to determine the impact of regional differences on the estimation, given that the northern health units had somewhat poorer results than did other regions of Ontario.

Mean relative deviations of estimates from true values were reduced for 5 of 7 northern health units and for 22 of the 35 remaining health units (Table 14). The addition of north/south to the estimation scheme resulted in a reduction of the RMSE for all target variables with the exception of BMI and hypertension (Table 15) but the reduction in error was usually small. Only 10 of 18 variables had an increased correlation coefficient (Table 16). As well, the number of estimates in the correct quintile was improved for 9 of 18 variables. However, the total correctly classified was still low, suggesting that the placement into the correct quintile was not much better than chance alone.

Table 14 : Mean and median relative deviation of estimates from true values including and excluding north/south distinctions, for each PHU, OHS best estimates

	Mean	Median	Mean (n/s)	Median (n/s)
Elgin-St.Thomas	.092	.096	.107	.101
Kent-Chatham	.091	.063	.105	.062
Renfrew County	.116	.125	.121	.114
Wellington	.111	.091	.114	.107
Leeds,Grenville	.131	.110	.112	.101
Eastern Ontario	.114	.053	.102	.063
Peel Regional	.114	.062	.105	.069
Simcoe County	.112	.083	.115	.089
Niagara Regional	.130	.100	.134	.107
East York	.129	.109	.114	.066
Hamilton Regional	.142	.085	.130	.083
Middlesex-London	.154	.085	.128	.048
Windsor-Essex	.164	.101	.138	.075
Oxford County	.131	.094	.152	.098
Sarnia-Lambton	.126	.095	.139	.096
North Bay	.122	.097	.102	.085
Kingston	.147	.103	.145	.110
Muskoka	.127	.073	.082	.071
Peterborough	.154	.160	.155	.141
Timiskaming	.146	.141	.125	.075
Bruce-Grey-Owen	.158	.118	.144	.095
Haliburton-Kawartha	.171	.094	.141	.083
Haldimand-Norfolk	.150	.101	.164	.104
Durham Regional	.164	.063	.152	.116
Toronto	.117	.086	.149	.088
York	.130	.087	.154	.059
Algoma	.183	.154	.241	.092
Waterloo	.175	.081	.150	.063
North Western	.193	.149	.198	.101
Ottawa-Carleton	.185	.098	.169	.120
North York	.172	.102	.165	.095
Porcupine	.202	.201	.072	.056
Sudbury	.178	.149	.117	.067
Hastings	.236	.152	.186	.091
Thunder Bay	.210	.139	.141	.079
Scarborough	.226	.147	.211	.155
York Regional	.198	.071	.219	.109
Brant County	.241	.087	.224	.086
Perth District	.244	.144	.221	.104
Halton Regional	.346	.181	.295	.116
Huron County	.287	.165	.286	.157
Etobicoke	.549	.113	.498	.087

Table 15: Root mean square error and root mean square error divided by the mean of the estimates corresponding to the OHS best estimate for each target variable, including and excluding a north/south distinction

	RMSE	RMSE (n/s)	Ratio RMSE(n/s)/ RMSE
Excess consumption of fat	3.770	3.636	.96
Chronic health problems	3.623	3.601	.99
Sedentary lifestyle	2.511	2.341	.93
Excess consumption of calories	4.032	3.889	.96
Smoking status	3.024	2.609	.86
High BMI	2.513	2.296	1.18
Hypertension	1.097	1.133	1.03
Two-week disability	2.102	2.039	.97
Excess consumption of alcohol	1.546	1.538	.99
Self-rated health	2.037	1.890	.92
Seatbelt usage: passenger	7.486	5.947	.79
Activity limitation	1.960	1.690	.86
Heart disease	1.059	0.759	.72
Hearing problems	1.322	1.174	.88
Seatbelt usage: driver	7.879	6.698	.85
ADL: personal affairs	0.987	0.919	.93
ADL: personal care	0.565	0.550	.97
Self-rated happiness	0.292	0.290	.99

Table 16 : Pearson correlation coefficient, number of estimates in the correct quintile, for each target variable, including and excluding north/south distinction, OHS best estimates

	r	r (n/s)	quintile	quintile (n/s)
Excess consumption of fat	.885*	.700	19	21*
Chronic health problems	.494	.561*	16*	9
Sedentary lifestyle	.470	.639*	13	15*
Excess consumption of calories	.830*	.720	20	24*
Smoking status	.650	.700*	13	15*
High BMI	.712*	.639	13	17*
Hypertension	.374	.554*	16*	13
Two week disability	.302	.419*	14*	12
Excess consumption of alcohol	.531*	.302	11	13*
Self-rated health	.591	.623*	14	14
Seatbelt use:passenger	.663*	.652	18	18
Activity limitation	.591	.647*	12	15*
Heart Disease	.517	.779*	17*	16
Hearing problems	.523	.622*	13	14*
Seatbelt use:driver	.650*	.580	15	16*
ADL:personal affairs	.363	.484*	15*	14
ADL:personal care	.302	.311*	11*	10
Self-rated happiness	.264	.271*	14*	12
Total *	6	12	7	9

* Flags larger number

DISCUSSION

This study has shown that the choice of predictor variables used for estimation can have a substantial impact on the error associated with a set of estimates. As a consequence, the ratio of largest to smallest error detected ranged from a low of 1.05 to a high of 1.67. However, the magnitude of the impact was shown to vary between predictor variables.

As well, the dispersion of the estimates was shown to be reduced in comparison to the true values, and the validity of the ranking of the estimates from lowest to highest prevalence was shown to be poor and not much better than chance. Both OHS and Census estimates had greatly reduced dispersions in comparison to the true values. Specifically, health units with high and low prevalences tended to be under and over-estimated, respectively. This is consistent with the results presented by Mackenzie et al, Schaible et al and Namekata et al. Unusually low and high prevalences may indicate that for these units, the nature of the association between target and predictor variables is different from that for other health units (i.e., that other factors may be acting). It may also be due to a large amount of unexplained variation. Differences in population structures may also account for variations in prevalence, however, given that indirect standardization takes these variations into account, differences in the associations may be a better explanation of this tendency.

Overall, the estimates produced were shown to be better than assuming that the provincial value is close enough to the health unit value. However the estimates are not overwhelmingly better. In fact, for the best set of estimates, there was only 30% improvement in error over using the provincial value alone.

Comparison of the results from all validation indices showed that Census estimates were usually poorer than OHS estimates. None of the sets of Census estimates showed better results for all validation indices; however, Census estimates of alcohol and seatbelt utilization: driver did show better results for 4 of 5 indices. Census estimates of hearing outperformed OHS estimates for 3 of 5 indices.

It was expected that Census estimates would be more valid, given that population counts for the health units would be more accurate. However, this expectation was based on the assumption that there would be stability in the associations over time. That OHS estimates were for the most part better than Census estimates calls this assumption into question. Changes in patterns of behaviour (e.g., reduced smoking rates) and changes in the population structure (e.g., aging of the population) may account for instability in the associations. Thus, the validity of the assumption is questionable, and we may need to use data sources from the same year or closer to each other in time.

Another explanation for the poorer quality of the Census estimates may be that the gold standards were poor, so that the Census estimates are actually of better quality than shown. This may be the case, but it is difficult to prove since these are the only estimates for the target variables available at the health unit level.

For the most part, consistent differences in the mean deviations for health units were not shown for regions. However, estimates for northern regions had somewhat more error than estimates for other regions. Addition of north/south into the estimation procedure resulted in a reduction of the RMSE for most northern health units for most target variables. As well, r values were improved for some variables, as were the numbers of estimates in the correct quintile. These results show that the nature of the association between target and predictor is somewhat different in the northern health units than in other regions of Ontario.

Although results were improved, using a regional identifier assumes that the data set from which the associations of target and predictor variables are drawn will also be able to provide the geographic distinction. This was the case for the OHS, but the sample sizes and types of data available from other surveys (e.g., Health Promotion Survey, Health and Welfare Canada) may prove insufficient and this type of distinction may not be available again.

Gonzalez and Waksberg found that counties with smaller populations had poorer estimates than counties with large populations. Comparison of the deviations by population size in this investigation did not show poorer estimates for smaller health units. The reasons for this are not clear. One difference between the investigations was the reliance on only one set of predictor variables by Gonzalez and Waksberg. It may be that this finding is due to the types of predictor variables employed and not a property of indirect standardization.

Some of the correlations were low and the number of estimates placed in the correct quintile was low. This reveals that the procedure does not preserve the ranking of the health units from lowest to highest prevalence for these target variables. As such, if the primary interest of an investigation is to identify areas of particularly high/low prevalence of a particular factor, indirect standardization has been shown to produce inaccurate results. In addition, this investigation also found that variables with the least amount of error do not necessarily produce the highest correlation coefficients. This finding is consistent with that of Schaible et al.

All validation indices utilized the point estimate as the true value, but the point estimate is subject to error. If the point estimate was under or over-estimated, the amount of error detected will be under/over-estimated for each health unit. As such, some variables may actually have been more poorly or better estimated than shown by this investigation.

Limitations and Error

This study was limited by the types of predictor variables available. It may be that other predictor variables would be more closely associated with the target variables, and would produce estimates with less error and a more accurate ranking of health units.

The error and bias introduced into the estimates came from many sources. It was expected that the error introduced into the estimates would be low for the OHS and almost or completely negligible for the Census. The tendency of the estimates to gravitate toward the mean, together with any error from the data sources would further increase the discrepancy between the estimates and true values. As well, this procedure was based on the assumption that there would be strong and stable associations between the predictors chosen and the target variables. The extent to which this assumption is not valid is difficult to quantify. Nonetheless, given the poor quality of the estimates, the validity of this assumption may be questionable.

As stated earlier, the project was limited by the types and combinations of predictor variables available. This limitation is likely to be the most important area where error was introduced. The amount of variation in the target variables explained by the predictor variables was obviously low given the reduced dispersion of the estimates.

CONCLUSIONS

Overall, the validity of the estimates has been shown to be questionable. This is particularly true for the Census estimates. Dispersions were greatly reduced. Consequently, for areas with prevalences similar to that of the larger area from which stratum specific proportions are drawn, the estimates tended to be closer to the true values than for other areas. Thus, this type of methodology is limited to small areas with prevalences similar to the larger area. This being the case, the utility of indirect standardization is limited as previous knowledge of the prevalence of a particular factor is necessary in order to determine whether or not this type of methodology would be useful to a particular health unit. Attempts to estimate prevalences for risk factors or health status indicators without a prior estimate would result in estimates for which the user would be uncertain of their validity. As such, this type of methodology does not appear to be adequate to meet the information needs for health units. Use of regression[2] may prove to be a more valid method of estimation for small areas.

REFERENCES

1. Ontario. Ministry of Health. Mandatory Health Programs and Service Guidelines. Ontario, April 1989.

2. Kleinman JC. Age-adjusted mortality indexes for small areas: Applications to health planning. AJPH 1977,67,9,834.

3. Mackenzie EJ, Shapiro S, Yaffe R. The utility of synthetic and regression estimation: Estimates for local health planning. Medical Care 1985;23,1,1.

4. Purcell N, Kish L. Estimation for Small Domains. Biometrics 1979;6,365.

5. National Centre for Health Statistics. Synthetic estimates of disability. PHS Publication No. 1759. Public Health Service. Washington DC. US Government Printing Office 1968.

6. Elston J, Koch G, Wiessert G. Regression-adjusted small area estimates of functional dependency in the non-institutionalized American population age 65 and over. Am J Pub Health 1991;81,3,335.

7. Gonzalez ME, Waksberg JE. Estimation of the error of synthetic estimates. Presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria, August, 1973.

8. Gonzalez ME. Use and evaluation of synthetic estimates. Proceedings of the American Statistical Association, Social Statistics Section, 1017, 1977, 33-36.

9. Schaible WL, Brock DB, Schnack GA. An empirical comparison of the simple inflation, synthetic and composite estimators for small area statistics. Proceedings of the American Statistical Association, Social Statistics Section, 1017, 1977.

10. Gonzalez ME, Hoza C. Small area estimation with application to unemployment and housing estimates. J AM Stat Assoc 1978;73,7.

11. National Centre for Health Statistics. Schaible WL, Brock DB, Casady RJ, Schnack GA. Small area estimation: An empirical comparison of conventional and synthetic estimators for states. Public Health Service. Vital and Health Statistics. Series 2, no.82. DHEW Pub. No. (PHS) 80-1356. Washington DC. US Government Printing Office, 1979.

12. National Centre for Health Statistics. Levy PS, French DK. Synthetic estimates of state health characteristics based on the Health Interview Survey. Public Health Service. Vital and Health Statistics, Series 2, no.75. DHEW Pub. No. (PHS) 78-1349.

Washington DC. US Government Printing Office, 1977.

13. National Centre for Health Statistics. State estimates of disability and utilization of medical services, United States, 1969-71. Public Health Service. DHEW Pub. No. (HRA) 77-1241. Health Resources Administration. Washington DC. US Government Printing Office, 1977.

14. Levy PS. The use of mortality data in evaluating synthetic estimates. Proceedings of the American Statistical Association 1971, Social Statistics Section. Washington. American Statistical Association, 1974;328.

15. Namekata T. Synthetic estimates of work disability. Unpublished Ph.D. thesis. University of Illinois, Champagne, Illinois, 1974.

16. Namekata T, Levy PS, O'Rourke TW. Synthetic estimates of work loss disability for each state and the District of Columbia. Pub Health Rep 1975;90,532.

17. Norman GR, Streiner DL. PDQ Statistics. B.C. Decker, Toronto, 1986.

18. SAS, Version 6.07, SAS Institute Inc.

19. Ontario. Ministry of Health. Ontario Health Survey 1990: User's Guide Vol. 1. Documentation. 1993.

20. Hunter, Lecily. Guidelines for the analysis and release of data from the Ontario Health Survey. Ontario Ministry of Health.1992

21. Canada. Health and Welfare. Nutrition Recommendations: The report of the Scientific Review Committee. 1990.

22. Canada. Health and Welfare. Canadian Guidelines for Healthy Weights. Report of an expert group convened by Health Promotion Directorate, Health Services and Promotion Branch, Ottawa, 1988.

23. Statistics Canada. Census of Canada, 1986: User summary tapes. Documentation. 1989.