ANALYSIS OF HYPERTENSION PREVALENCE DATA FOR THE CANADIAN POPULATION

Jane F. Gentleman and Monica Tomiak, Statistics Canada
Jane F. Gentleman, Statistics Canada, Ottawa, Canada K1A OT6

KEY WORDS: Prevalence, hypertension, high blood pressure, Canada Health Survey

1. INTRODUCTION

Hypertension (high blood pressure) is a major risk factor for cardiovascular disease, which is the leading cause of death in Canada. The 1978/79 Canada Health Survey (CHS) therefore devoted a significant portion of its resources to gathering data about blood pressure in the Canadian population. Conclusions from an initial analysis of the results were as follows:

> Nearly 200,000 Canadians have blood pressure elevated to such a degree that (additional) treatment would almost certainly be beneficial. An additional 2.6 million persons might benefit from having their blood pressure lowered. Two thirds of Canadians who have elevated blood pressure are unaware of the fact. Even among those who do know that their blood pressure is elevated, approximately one in five is not taking medication.

(See Statistics Canada and National Health and Welfare, 1981, p. 143.)

More recently, the 1985 Canadian Blood Pressure Survey (CBPS) concluded, similarly, that

> A large number of Canadians are at increased risk of cardiovascular disease due to high blood pressure. Of those, many are unaware of their condition or do not have their blood pressure under control.

(See National Health and Welfare (1989), p. 3.)

We are currently engaged in an analysis, using CHS data, of the prevalence of high blood pressure, and of the association of certain demographic, socio-economic, and risk factors with high blood pressure. In order to carry out this analysis, we endeavored to select a measure of prevalence from among several which were available from the CHS. This paper describes the various prevalence measures and makes comparisons among them. The general problems associated with measuring the prevalence of a health characteristic are discussed. We also present some results from an exploratory analysis of the statistical distributions of systolic and diastolic blood pressure measurements.

Data from the CHS are being analyzed rather than those from the CBPS because the former provides a larger number of respondents and a fuller set of covariates. The CHS gathered information on the health status of 31,668 persons, while the CBPS, which had the more specific objective of assessing the problem of high blood pressure in Canada, had only 3092 respondents. The CHS obtained smoking and drinking histories, information on income and education, and various other covariates which are being analyzed in the multivariate regression models of our study.

Because the intent of the CHS was to examine the health status of Canadians at the time of the survey, CHS data are more suitable for measuring the prevalence of health characteristics than for measuring their incidence. The incidence of a characteristic is the probability of (or proportion) acquiring it; prevalence is the probability of (or proportion) having it.

2. THE DATA

The Canada Health Survey (CHS) was conducted jointly by Statistics Canada and Health and Welfare Canada over a period of several months during 1978 and 1979. The area frame for the survey was stratified by province. The provinces of Quebec and Ontario each contained further strata defined according to provincial health region and according to rural/urban characteristics. This was a household survey to which a total of 31,668 individuals (15,655 males and 16,013 females) responded. The sample did not include institutionalized persons, or persons residing in the Yukon, Northwest Territories, on Indian reserves, or in certain geographically remote areas. The exclusion of institutions implies that hospitalized persons are not represented in CHS data; this must be taken into account in analyzing the health data.

Five vehicles for collecting and recording data were used:

* Household Record Card (HRC)
  Identifying and demographic data (e.g., sex, age, geographical location) for all household members of all ages were recorded. The number of respondents was 31,668.

* Interviewer Administered Questionnaire (IAQ)
  Data (e.g., economic family income, education, marital status, labour force status) were collected about all household members of all ages by personally interviewing a "suitable member" of the household. The number of respondents was 31,668. (The term

"respondents" will be used here to refer to a person about whom data were gathered, whether or not the data were supplied by proxy.)

* Lifestyle and Your Health Questionnaire (LHQ)
  Personal, possibly sensitive data (e.g., smoking habit, drinking habit, genetic history) were collected for household members of ages 15 and above. Respondents filled out their own individual questionnaires which had been left behind and were picked up later by the IAQ interviewer. The number of respondents was 23,791.

* Physical Measures Questionnaire (PMQ)
  For a subset of the households, a nurse accompanied the IAQ interviewer on the return visit to the household. The nurse took physical measurements of blood pressure, cardiorespiratory fitness, height, weight, and skinfold measurements (after screening out certain types of people, as described in Statistics Canada and National Health and Welfare (1981), p. 71). Because of time and cost limitations, PMQ sample sizes were relatively small; for example, the number of respondents between ages 15 and 64 was 5765.

* Blood Sample
  During the household visit, the nurse also took blood samples from persons age three years and over.

All data described in this paper refer to persons of age 15 or more. For an extensive discussion of the intent and design of the CHS, and an initial analysis of the data, see Statistics Canada and National Health and Welfare (1981).

3. THE DISTRIBUTIONS OF SYSTOLIC AND DIASTOLIC BLOOD PRESSURE

Fig. 1 is a scatter plot of systolic versus diastolic blood pressure (in millimeters of mercury - mmHg). Although a large number of the 4653 points in Fig. 1 are superimposed on each other, the plot reveals the positive correlation (.63) between the two measures. There are no outlying points in the upper left or lower right portions of the scatter plot, showing that no respondent had an extremely high systolic and extremely low diastolic blood pressure (or vice versa). Inspection of data for the slightly remote points elsewhere on the graph did not reveal them to be particularly unusual.

Superimposed on the scatter plot are lines showing the boundaries of the three categories used by the World Health Organization to categorize blood pressure:

<u>Normal</u>: Diastolic < 90 and Systolic < 140
<u>Borderline</u>: Not Normal or Elevated
<u>Elevated</u>: Diastolic $\geq$ 95 or Systolic $\geq$ 160

In Fig. 1, the percentages of persons in these categories were 73.7%, 15.9%, and 10.4%, respectively.

Fig. 2 and Fig. 3 are also scatter plots of systolic versus diastolic blood pressure, for just the lowest (15-19) and highest (65+) ages observed. These graphs vividly demonstrate the tendency of blood pressure to increase with age. Examination of such plots for five intermediate age groups showed that the slope of a regression line through the points increases with age from .56 at age 15-19 to 1.24 at age 60-64, and then decreases to 1.10 at age 65+. (The slope of a line fitted to Fig. 1 is 1.06.) This indicates that systolic blood pressure increases faster with age than diastolic blood pressure.

A graph (not shown here) was produced of the empirical cumulative distribution function (ecdf) of the systolic blood pressure measurements obtained from 4677 PMQ respondents. On it was superimposed the cdf for the Normal distribution with the same mean and standard deviation (125.5 and 19.4) as the data. The graph showed that the distribution of systolic blood pressure is non-normal and has a long right tail. Thus, it is easier to find a person with an unusually high systolic blood pressure than one with an unusually low systolic blood pressure. (This is true for diastolic blood pressure as well.) This is because people with exceptionally low blood pressure either die, or they are likely to be in the hospital (perhaps because they have recently had a heart attack). In the latter case, the CHS did not survey them.

The fact that blood pressure measurements were supposed to be recorded to the nearest even digit was evident from the ecdf graph; measurements were more clustered at even blood pressure values (although the presence of a few lone points at odd blood pressure values indicated that some interviewer/nurse teams did not adhere to these instructions). Furthermore, vertical point clusters were relatively large at multiples of ten, indicating that some measurements were rounded to the nearest multiple of ten, rather than to the nearest even blood pressure value (as previously noted in Statistics Canada and National Health and Welfare, 1981, p. 143). Some clustering occurred at multiples of five as well. This type of clustering may be hidden by a histogram, and may, in fact, distort the impression of the data given by the histogram, but it is easily observed in ecdf graphs. The same rounding tendencies were noted for diastolic blood pressure.

A Q-Q plot (not shown here) comparing the distribution of male systolic blood pressure to that of females had a configuration of points which was quite straight for such plots. This indicates that the distribution (whatever it is) of systolic blood pressure is the same for males and females, except for scale and location differences: the means were 128.7 and 122.9 for 2101 males and 2576 females, respectively, and the standard deviations were 17.9 and 20.2. Similar results were found for diastolic blood pressure.

A Q-Q plot (not shown here) was produced comparing those being medicinally treated for high blood pressure to those not being treated (based on IAQ responses). The 4349 untreated persons had a much lower average systolic blood pressure (123.7) than the 328 treated persons (who averaged 149.9). The standard deviations were 17.9 and 21.7, respectively. The plot showed that the underlying distributions were otherwise the same. The minimum systolic blood pressure for those not being treated was 80, while the minimum for those being treated was 106; people with very low blood pressure do not take medication for high blood pressure, and people who do take medicine apparently do not respond so much as to achieve extremely low systolic blood pressure. Similar patterns were noted for diastolic blood pressure.

4. INDICATORS OF HIGH BLOOD PRESSURE AND THEIR USEFULNESS FOR MEASURING PREVALENCE

Five CHS variables can be used to determine whether a person has or has had high blood pressure. Two of these are obtained from IAQ responses, one from LHQ responses, and two from PMQ data:

* Prevalence of High Blood Pressure - I (HBP-I)
  Response to IAQ question, "Does (person in the family) presently have high blood pressure?"

* Prevalence of High Blood Pressure - II (HBP-II)
  Response to LHQ question, "Have you ever had high blood pressure?"

* Prevalence of High Blood Pressure - III (HBP-III)
  Response to PMQ question, "Has your doctor ever said your blood pressure was too high?"

* Blood Pressure Measurements: Systolic and Diastolic (BPM).
  Results from PMQ, categorized as Normal, Borderline, or Elevated (as defined in Section 2).

* Use of Drugs for Hypertension (UDH)
  Response to the last of a series of IAQ questions: "Yesterday, or the day before, did (person in the family) take or use (list of types of medications, including 'Medicine for the heart or blood pressure')? (If so) Over the past month was this medication taken at least once every week? (If so, or if unsure) What is the main health problem for which (person in the family) took this medication?" (Possible responses to the last question include 'hypertension'.)

Table 1 summarizes the major advantages and disadvantages of using each of the five indicators to measure the prevalence of high blood pressure. Seven properties are compared: (1) the number of respondents of age 15 and above; (2) the upper age limit of respondents; (3) whether or not a proxy could have supplied the information; (4) what the time frame of the question was (i.e., whether the question referred to the present or to anytime in the past); (5) whether there could have been a perceptual problem regarding how to answer the question for a respondent who was receiving medicinal treatment for high blood pressure; (6) what the possible responses to the question were; and (7) whether the PMQ was also administered to that set of respondents.

If used alone as a measure of prevalence of high blood pressure, UDH would underestimate it, since not all of those suffering from high blood pressure are being treated for it. Some analysts, however, use such an indicator in conjunction with others to define high blood pressure. For example, one could categorize a person as having high blood pressure if the response to either HBP-I or UDH were yes. This would help to alleviate the perceptual problem re medicine which exists for HBP-I; the question did not specify whether a person being (successfully) treated for high blood pressure would or would not be considered as having high blood pressure. UDH could also be used to "correct" a prevalence measure derived from BPM in cases where medication has reduced what would have been high blood pressure to the normal range.

The UDH indicator is, however, somewhat uncertain due to the time frame to which it refers. To yield a yes answer to this question, the respondent must have taken medication for high blood pressure within the last two days, as well as at least once per week over the past month. Since most blood pressure medication is taken daily, this may not cause a serious analytical problem (although any newly-treated patients will not be counted as having high blood pressure by this indicator).

The time frame is also a problem for those questions which ask whether the person ever had high blood pressure or whether the doctor ever said so, unless

TABLE 1: Pros and Cons of 5 Indicators of High Blood Pressure

	HPB-1	HPB-II	HPB-III	BPM	UDH
Source	IAQ	LHQ	PMQ	PMQ	IAQ
# Respondents ≥ 15	23971	23791	5765	6604	23791
Upper Age Limit	none	none	64	none	none
Proxy Possible?	yes	no	no	no	yes
Time Frame	now	ever	ever	now	last 2 days/ month
Perceptual Problem re Medication	yes	no	no	no	no
Possible Responses	yes no	yes no not sure no resp	yes no no resp	norm bord elev unav	yes no or not asked
Physical Measures Also Available for All?	no	no	yes	yes	no

one considers high blood pressure to be incurable. In fact, some doctors do consider ordinary high blood pressure to be incurable, but some occurrences of it are indisputably temporary - e.g., when it occurs as a result of pregnancy. For measuring prevalence of high blood pressure, therefore, the HBP-I and BPM indicators are preferable with respect to time frame.

Three of the indicators have a significantly larger number of respondents than the other two, which is a clear advantage when performing a univariate analysis of the indicator. However, if a multivariate analysis requires covariates which are available for only a smaller subset of the data, the extra observations must be discarded and the advantage is lost. For example, we were particularly anxious to include Body Mass Index (BMI) as a covariate in our regression analysis, and BMI is available only from the PMQ.

A large number of observations may also be lost when persons for whom the responses were "not sure", "no response", or "unavailable" must be discarded from the analysis. This would suggest that HBP-I is a preferred indicator, except that it is difficult to believe that every person who was asked that question was able to provide a definitive yes or no answer, especially since HBP-I is one of the two indicators for which a proxy answer was possible. We therefore suspect that it would be overly optimistic to believe that HBP-I clearly distinguishes between these two categories.

Another consideration is the age range of respondents. HBP-III respondents were limited to those under age 65, which excluded an important part of the population from the study, especially since blood pressure tends to increase with age.

Thus, there are pros and cons regarding the use of each of the indicators. Rather than study a single favored measure of prevalence of high blood pressure, we therefore analyzed the consistency among the different indicators, and we repeated our multivariate regressions (to be described elsewhere) using different prevalence measures as dependent variables.

The top section of Table 2 gives the proportion of persons who would be deemed to have high blood pressure using each of the indicators in Table 1 (including UDH as a crude prevalence measure). Each indicator is treated as a dichotomous variable, with all uncertain responses excluded from the calculations. The BPM indicator is used as a prevalence measure in two ways: with high blood pressure defined as having an elevated BPM, and with high blood pressure defined as having either an elevated or a borderline BPM.

TABLE 2: Comparison of Indicators of High Blood Pressure

CHARACTERISTIC		MALES	ALL	FEMALES
Pr(HBP-I=Yes)		.069	.091	.112
Pr(HBP-II=Yes)		.125	.160	.190
Pr(HBP-III=Yes)		.110	.138	.161
Pr(BPM=Elevated)		.116	.104	.094
Pr(BPM=Elevated+Borderline)		.304	.264	.231
Pr(UDH=Yes)		.045	.064	.081
Pr(HBP-II=Yes	|HBP-I=Yes)	.949	.965	.973
Pr(HBP-III=Yes	|HBP-I=Yes)	.970	.960	.954
Pr(BPM=Elevated	|HBP-I=Yes)	.406	.406	.405
Pr(BPM=Elev+Border	|HBP-I=Yes)	.769	.721	.696
Pr(UDH=Yes	|HBP-I=Yes)	.637	.686	.714
Pr(HBP-I=Yes	|HBP-II=Yes)	.563	.585	.598
Pr(HBP-III=Yes	|HBP-II=Yes)	.784	.825	.847
Pr(BPM=Elevated	|HBP-II=Yes)	.344	.322	.310
Pr(BPM=Elev+Border	|HBP-II=Yes)	.660	.600	.569
Pr(UDH=Yes	|HBP-II=Yes)	.379	.422	.447
Pr(HBP-I=Yes	|HBP-III=Yes	.495	.473	.461
Pr(HBP-II=Yes	|HBP-III=Yes)	.802	.837	.855
Pr(BPM=Elevated	|HBP-III=Yes)	.337	.281	.250
Pr(BPM=Elev+Border	|HBP-III=Yes)	.653	.538	.475
Pr(UDH=Yes	|HBP-III=Yes)	.288	.308	.319
Pr(HBP-I=Yes	|BPM=Elevated)	.268	.389	.510
Pr(HBP-II=Yes	|BPM=Elevated)	.449	.576	.689
Pr(HBP-III=Yes	|BPM=Elevated)	.426	.531	.648
Pr(UDH=Yes	|BPM=Elevated)	.165	.270	.375
Pr(HBP-I=Yes	|BPM=Elev+Border)	.193	.273	.358
Pr(HBP-II=Yes	|BPM=Elev+Border)	.313	.417	.522
Pr(HBP-III=Yes	|BPM=Elev+Border)	.277	.361	.466
Pr(BPM=Elevated	|BPM=Elev+Border)	.381	.394	.408
Pr(UDH=Yes	|BPM=Elev+Border)	.133	.196	.264
Pr(HBP-I=Yes	|UDH=Yes)	.973	.980	.984
Pr(HBP-II=Yes	|UDH=Yes)	.961	.975	.982
Pr(HBP-III=Yes	|UDH=Yes)	.983	.972	.966
Pr(BPM=Elevated	|UDH=Yes)	.367	.399	.416
Pr(BPM=Elev+Border	|UDH=Yes)	.780	.738	.717

Table 2 shows a wide range of possible values for the prevalence of blood pressure, both among the different indicators and between the sexes. For example, as measured by HBP-I, the prevalence of high blood pressure in the overall population is .091, but it is much lower for males (.069) than for females (.112). This male/female pattern is reversed, however, when using either of the two BPM prevalence measures. Clearly, the results of an analysis of prevalence data will be profoundly affected by the analyst's choice of a prevalence measure.

Since BPM is the more objective of the two measures, these results indicate that a higher proportion of males than females have high blood pressure. Table 2 also indicates that about twice as many females as males were taking medication for high blood pressure. This may be due in part to better self health care on the part of females.

The subsequent six sections of Table 2 give pairwise comparisons among the six resulting measures of prevalence. For example, the probability of answering yes to HBP-II, given that the answer to HBP-I was yes, was .965 for the overall population, .949 for males, and .973 for females. Because these three values are near 1.0 and are among the highest in Table 2, one might conclude that there is a high degree of consistency between HBP-I and HBP-II as prevalence measures. However, it is somewhat disconcerting to note that 3.5% of those who have high blood pressure now have never had it! This discrepancy, which is larger for males (5.1%) than for females (2.7%), may be due to incorrect proxy answers to HBP-I and/or to unwillingness to admit, in responding to HBP-II, that a health problem exists.

There are large differences between sexes in the probability of a person with a high BPM also having a yes response to HBP-I. For example, 51.0% of females with elevated BPM had a yes answer to HBP-I, while only 26.8% of males did so. This suggests that more males than females have undiagnosed high blood pressure, and/or that males are less likely than females to admit to or tell their families about their having high blood pressure. For either sex, the consistency between HBP-I and BPM is very low, which is unfortunate, since the former type of data are much less expensive to collect.

The low degree of consistency between HBP-I and HBP-II suggests that a large number of those who formerly had high blood pressure had later been cured (although some of the difference may be due to incorrect proxy answers). For example, the probability of having high blood pressure now (as measured by HBP-I), having ever had it (as measured by HBP-III), is only .473 in the overall population.

The proportion of persons taking medication for hypertension who nevertheless had an elevated BPM is .399, indicating the failure of medication to correct the problem in these cases.

5. CONCLUDING REMARKS

Considering the above results, we reluctantly conclude that while physical measurements are extremely expensive to obtain, and while they impose a heavy response burden, they are far more accurate than self-report or proxy health data. Administrative health data are often satisfactory, but only if answers to the desired questions can be obtained and if the required covariates are available. We note that in study after study, analysts have discarded large numbers of observations of self-report variables in favor of utilizing smaller sample sizes of more objective, direct physical measurements, and/or because other physical measurements were needed as covariates. The cost of collecting physical measurements may therefore be considerably offset if one subtracts from it the cost of wasting data.

6. REFERENCES

National Health and Welfare (1989). Main Findings Report of the Canadian Blood Pressure Survey. Catalogue H39-143/1989E.

Statistics Canada and National Health and Welfare (1981). The Health of Canadians. Report of the Canada Health Survey. Catalogue 82-538E. Ottawa, Canada.

FIG 1 : SYSTOLIC VS DIASTOLIC BLOOD PRESSURE
MALES AND FEMALES AGED 15+

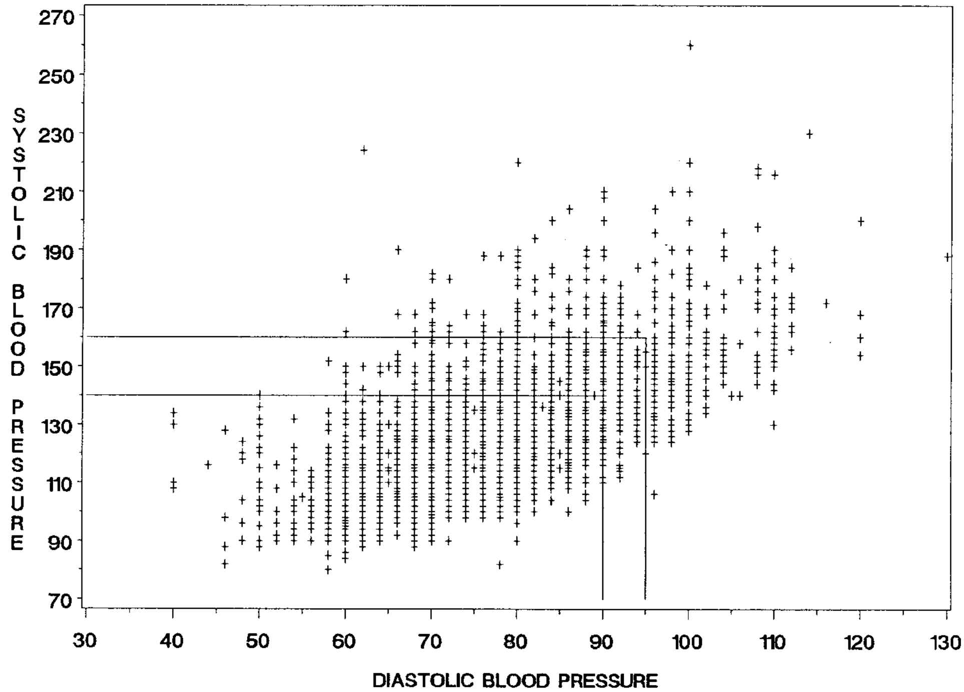

FIG 2 : SYSTOLIC VS DIASTOLIC BLOOD PRESSURE
MALES AND FEMALES AGED 15-19

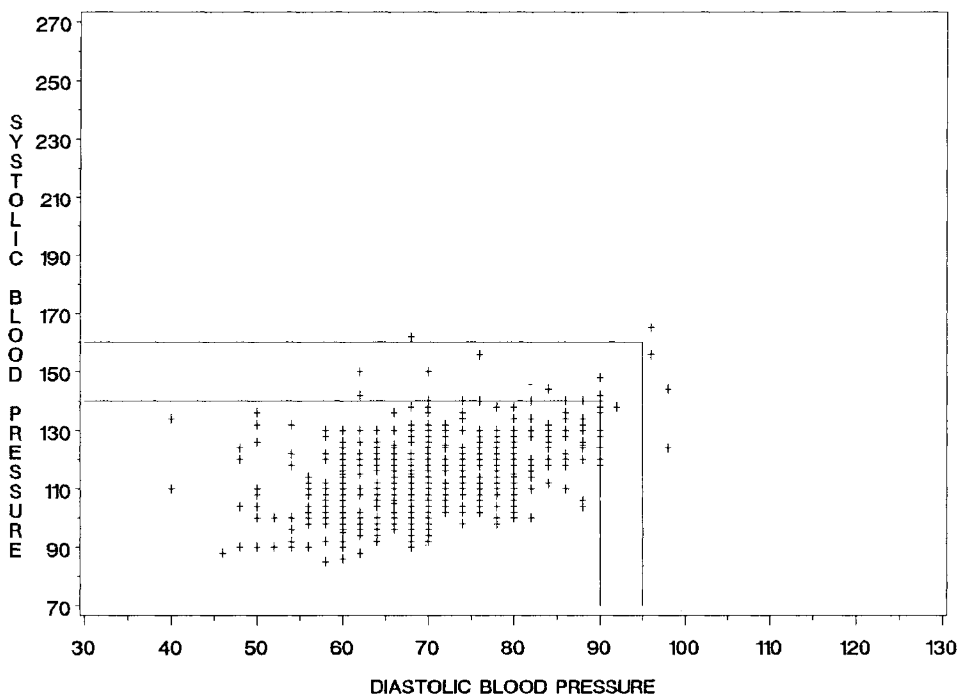

FIG 3 : SYSTOLIC VS DIASTOLIC BLOOD PRESSURE
MALES AND FEMALES AGED 65+