

Using Administrative Data to Analyze the Prevalence and Distribution of Schizophrenic Disorders

Elliot M. Goldner, M.D., M.H.Sc.

Wayne Jones, M.A., M.Sc.

Paul Waraich, M.D., M.H.Sc.

Objective: In order to effectively plan and implement psychiatric services, a clear estimate of the prevalence and distribution of the population in need is required. The authors examined the use of administrative data as a means of estimating the prevalence and distribution of schizophrenic disorders. **Methods:** Administrative health services data for residents of the Canadian province of British Columbia in the age range 15 to 65 years (total population in 1997–1998 of 2,703,588) were examined over a three-year period. Potential cases of schizophrenic disorder were identified on the basis of the presence of a diagnostic code of 295 in one or more of three databases. One-year prevalence rates were estimated for each of the province's geographic regions, and associations with low income and unemployment were examined. **Results:** One-year prevalence rate estimates were .45 cases per 100 population in 1996–1997 and 1997–1998 and .42 cases per 100 in 1998–1999. The prevalence estimates of all 88 local health areas in the province were consistent across the three-year period; Pearson correlations were determined to be approximately .9. One-year contact prevalence rates for schizophrenic disorders were significantly correlated in all three years to the percentage of persons with low income in the individual geographic regions but were not correlated with unemployment rates. **Conclusions:** In areas with well-developed health services, analyses of administrative data appear to provide cost-effective means of examining the prevalence and distribution of schizophrenic disorders. (*Psychiatric Services* 54:1017–1021, 2003)

An important foundation for planning and improving mental health services is a clear estimate of need (1). Although it has been well established that schizophrenic disorders produce a substantial need for mental health services (2), communities require accurate measures of the extent and distribution of such need in order to correct-

ly allocate services. An underestimate of need could result in the tragic neglect of individuals who are suffering, and an overestimate could result in the misallocation of precious health resources.

One approach might be to extrapolate prevalence estimates for schizophrenic disorders by applying the findings of large-scale epidemiologic

studies arithmetically to local populations. However, the results of such formulaic calculations are unlikely to provide accurate estimates. As seems to be the case for all health problems studied, schizophrenic disorders are unequally distributed in communities (3). A recent systematic review of high-quality, large-scale epidemiologic studies that used similar methods to determine the incidence and prevalence of schizophrenic disorders produced varying rates (4). Furthermore, such differences in the aggregated data of large-scale studies, which often sample entire cities, states, provinces, or nations, do not demonstrate the considerable variability in rates across individual towns, communities, or neighborhoods—the so-called small-area variations. Small-area variations in the prevalence rates of schizophrenic disorders have been found to be substantial (5–7).

Given the variation in the prevalence rates of schizophrenic disorders at the level of individual communities, how can planners estimate the appropriate level of resources to be allocated in a community? In this article we examine the use of administrative data—information routinely collected by health providers and governments—for the purpose of estimating the prevalence, incidence, and distribution of schizophrenic disorders. We also discuss the relative benefits and limitations of this method and discuss its utility in complementing rigorous epidemiologic surveys.

The authors are affiliated with the mental health evaluation and community consultation unit of the department of psychiatry at the University of British Columbia, 2250 Wesbrook Mall, Vancouver, British Columbia, Canada V6T 1W6 (e-mail, goldnee@interchange.ubc.ca).

Methods

British Columbia population data were obtained from yearly population estimates provided by BC Statistics (8). Unemployment and low-income data for the province of British Columbia were obtained from national census tabulations (9). Administrative data on the number of persons with diagnoses of schizophrenic disorders were obtained through collaboration with staff at the British Columbia Ministry of Health Services.

In accordance with established policies and principles for protection of privacy and confidentiality (10), administrative data from the British Columbia Ministry of Health Services were given to the researchers with all personal identifiers removed. A formal data request was reviewed by the Ministry of Health Services and approved. Informed consent was not required, because only aggregated population data were provided. Administrative data from three sources were examined over a three-year period—April 1, 1996, to March 31, 1999—and potential cases of schizophrenic disorder were identified on the basis of the presence of an *ICD-9* or *DSM-IV* code of 295 in at least one of three databases—physician services, hospital discharge abstracts, or the community mental health information management system.

Hospital discharge abstracts contain up to 16 *ICD-9* codes related to the patient's hospital stay. The provincial community mental health information system contains room for diagnostic information, and, in most cases, a diagnosis is documented at the time the record is closed. Record selection was limited to patients between the ages of 15 and 64 years and those with a valid personal health number. Virtually all residents of British Columbia are served by the public health system and are issued a personal health number, and all three databases record this identifier. Thus the personal health number provides the link that allows identification of individuals and avoids double counting of records or individuals.

The province of British Columbia is subdivided into 88 different geographic regions called local health areas (LHAs). These geographic units are used in the analysis of a wide vari-

ety of provincial data. Thus census data on unemployment and low income as well as population estimates are readily available at this level. All three administrative data sources contain information on the patient's residence, usually in the form of a postal code, and this information can be assigned to an LHA. In cases in which a patient has more than one LHA registered as the place of residence within the year—for example, when a patient has had more than one contact and has moved to a different residence between contacts—the LHA of the patient as of October 1 of the year in question (the midpoint for the fiscal years April 1 to March 31) is designated as the patient's LHA for that year.

The LHAs can be aggregated to form the province's 15 larger health service delivery areas (HSDAs) and further aggregated to form the five parent health authorities. We calculated whether rates in the individual HSDAs within the health authority were significantly lower or higher than the overall rate for the province by using a method that has been described by Cain and Diehr (11). The individual HSDAs were examined by means of a 2×2 table analysis, whereby one row in the table is the observed and expected (on the basis of the provincial rate) number of schizophrenia cases for the HSDA in question, and the second row is the observed and expected number of cases for the remainder of the province.

A significant chi square value was interpreted as indicating that the particular HSDA rate is higher—or lower—than that of the remainder of the province. All HSDAs can be examined in this manner to determine which ones have rates that are significantly different from those in the remainder of the province. The chi square value was compared with the critical value with one degree of freedom adjusted for the planned number of multiple tests in the year (using the Bonferroni adjustment for multiple comparisons; here 15 tests were planned, so the significance level used was .05 divided by 15, or .003).

In undertaking the main analyses in this study, we used data from 1997–1998, because this was the midpoint year and was also the one that

we judged to have the most reliable documentation in all administrative databases. Associations among prevalence and low income and between prevalence and unemployment were examined by calculating Pearson correlation coefficients. Low income and unemployment are two measures of social deprivation that are routinely collected in Canadian census data. Some degree of association between social deprivation and the prevalence of schizophrenia would further strengthen the validity of the use of administrative data to provide prevalence estimates in this population.

Results

The province of British Columbia has a number of conurbations, often comprising a number of LHAs, and large rural and sparsely populated areas. The population of individual LHAs is highly variable, ranging in 1997 from 397 persons in the relatively remote community of Telegraph Creek to 229,868 persons in the urban community of Surrey. The mean adult population (persons aged 15 to 64 years) of the LHAs in 1997 was $30,723 \pm 39,656$, and the median was 13,466. The figures for 1996 and 1998 were similar, because British Columbia's adult population grew by 1 to 2 percent each year during that period. When the five larger health authorities were used as the geographic units of analysis, the mean population was 553,118 and the median was 463,637, with a maximum of 893,889 and a minimum of 221,530.

A total of 12,087 persons were identified as receiving some sort of assessment or treatment and being given a diagnosis of schizophrenia in 1997–1998. On the basis of the province's adult population of 2,703,588, this treatment and diagnosis rate translates to a one-year contact prevalence rate of .45 cases per 100 population. These estimates were relatively stable over the three years: 11,929 persons or .45 cases per 100 in 1996–1997 and 11,516 persons or .42 cases per 100 in 1998–1999.

Within the LHAs, the median one-year prevalence rate was .38 per 100 in 1997–1998, and rates ranged from lows of zero per 100 in the LHA with the smallest population to 1.51 per

100 in one of the LHAs in the city of Vancouver. Similar results were observed for the other two years. When data were aggregated to the level of the 15 larger HSDAs, rates for 1997–1998 ranged from .28 per 100 to .61 per 100, with a median rate of .43 per 100. Aggregation to the even larger health authorities produced ranges for 1997–1998 of .39 to .49 per 100.

An analysis of the prevalence estimates for all 88 LHAs showed consistency across the three-year period; Pearson correlations were determined to be approximately .9 (Table 1).

Table 2 presents the one-year prevalence rates for schizophrenic disorders in each of the three years studied for each of British Columbia's 15 HSDAs and their respective parent health authorities. Table 3 provides one-year prevalence rates for the LHAs within one HSDA (Vancouver).

Using 1996 census data available for the 88 LHAs, we determined the

Table 1

Pearson correlations among estimates of one-year prevalence rates of schizophrenic disorders for local health areas in British Columbia, Canada, across three years

Estimate period	1998–1999	1997–1998	1996–1997
1998–1999	1.0		
1997–1998	.91	1.0	
1996–1997	.91	.89	1.0

median percentage of persons with low income to be 16.1 percent (range, 51.8 to 6.1 percent). Census data also determined that the median unemployment rate in 1996 was 11 percent (range, 31.2 to 4.7 percent).

The one-year contact prevalence rates for schizophrenic disorders for the 88 LHAs estimated in this study were significantly correlated in all three years with the percentage of persons with low income in the LHAs: $r=.61$ for 1996–1997, $r=.39$ for

1997–1998, and $r=.46$ for 1998–1999. In contrast, no significant correlations were found between prevalence and unemployment.

Discussion and conclusions

Given that the prevalence rates estimated in this study are based on data obtained from contact with health service providers—contact prevalence—might they be an underestimate of true prevalence? (The term “true prevalence” is sometimes used,

Table 2

One-year prevalence rates (1996–1999) of schizophrenic disorders in health service delivery areas (HSDAs) in British Columbia, Canada, as estimated by using administrative data and indication of magnitude relative to the province's average^a

Health authority and HSDA	Cases			Rates (per 100 population)			Relative to province's average ^b		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
1. Interior									
Okanagan	954	988	935	.45	.46	.43			
East Kootenay	184	211	172	.35	.39	.32	Low		Low
Kootenay Boundary	270	294	292	.5	.54	.53		High	High
Thompson–Cariboo	518	567	532	.43	.46	.43			
Total	1,926	2,060	1,931	.44	.46	.43			
2. Fraser									
Fraser Valley	711	748	702	.48	.5	.46		High	
South Fraser	1,367	1,374	1,282	.37	.37	.34	Low	Low	Low
Simon Fraser	1,524	1,484	1,389	.44	.42	.39			Low
Total	3,602	3,606	3,373	.42	.41	.38	Low	Low	Low
3. Vancouver Coastal									
North Shore–Coast									
Garibaldi	524	498	425	.3	.28	.23	Low	Low	Low
Vancouver	2,588	2,517	2,513	.64	.61	.6	High	High	High
Richmond	305	333	294	.28	.29	.25	Low	Low	Low
Total	3,417	3,348	3,232	.5	.47	.45	High	High	High
4. Vancouver Island									
Capital	1,124	1,202	1,155	.51	.55	.52	High	High	High
Central–North Island	978	1,013	967	.43	.43	.41			
Total	2,102	2,215	2,122	.47	.49	.46		High	High
5. Northern									
Northern Interior	431	404	426	.4	.36	.38		Low	
North West	297	297	309	.48	.47	.49			
Peace Liard	154	157	123	.35	.35	.27	Low	Low	Low
Total	882	858	858	.41	.39	.39	Low	Low	Low
British Columbia	11,929	12,087	11,516	.45	.45	.42			

^a Analysis based on the method of Cain and Diehr (11)

^b “High” and “low” indicate rates that are significantly different from the provincial rate.

Table 3

One-year prevalence rates (1996–1999) of schizophrenic disorders in one urban health service delivery area (Vancouver), as estimated by using administrative data and indication of magnitude relative to the province's average^a

Local health area	Cases			Rates (per 100 population)			Relative to province's average ^b		
	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3	Year 1	Year 2	Year 3
City Centre	425	435	427	.62	.6	.56	High	High	High
Mid East	691	612	646	1.85	1.57	1.62	High	High	High
North East	306	312	305	.47	.47	.46			
Westside	297	299	283	.34	.33	.31	Low	Low	Low
Midtown	437	419	440	.71	.66	.69	High	High	High
South	432	440	412	.52	.52	.49			
Total	2,588	2,517	2,513	.64	.61	.6	High	High	High

^a Analysis based on the method of Cain and Diehr (11)

^b "High" and "low" indicate rates that are significantly different from the provincial rate.

in contrast with treatment prevalence—the prevalence of a disorder as determined by examining only individuals who are receiving treatment. However, a significant proportion of persons with schizophrenic disorders come into contact with health providers without being in treatment—for example, during an emergency assessment or a visit to a physician for a different concern. Thus we use the term contact prevalence to more clearly describe this rate and to distinguish it from the lower rates that would be obtained from treatment records.)

In jurisdictions in which most people have good access to medical care it is more likely that an individual with an illness will be detected in administrative data. However, the concordance between contact prevalence and true prevalence varies from disorder to disorder. For persons with depressive disorders, it has been found that although 90 percent have had contact with their family doctors for a medical problem in the previous year, only 50 percent have received a mental health diagnosis (12).

In contrast, most studies have found that only 10 percent of persons with schizophrenic disorders do not present for mental health treatment during the course of a year (5,6,13). A single exception was a finding of the U.S. Epidemiologic Catchment Area study that only 60 percent of persons with schizophrenia obtained services (14). However, this finding may be an artifact of the exceptionally high prevalence rates estimated in that study (4).

For schizophrenic disorders, a high ratio of health-system contact to prevalence is likely the result of a number of factors, including relatively high diagnostic visibility (the symptoms are relatively distinctive and frequently lead to at least brief receipt of health services), a policy-based emphasis on provision of services to persons with severe mental disorders, proliferation of assertive community treatment programs, and efforts to increase the early diagnosis and treatment of psychotic disorders. Furthermore, the method of case identification that we used required that only one diagnostic notation be made by any service provider in the system.

Nevertheless, it is inevitable that some proportion of persons with schizophrenic disorders will not be seen by any health professional in a one-year period, even though that proportion may be small. If the proportion were sizeable, the prevalence findings based on the method we used would likely produce significant underestimates of need.

Conversely, it is possible that our analysis artificially inflated prevalence rates by identifying false-positives as a result of misdiagnoses of schizophrenic disorders. Yet physicians and mental health professionals are generally conservative in assigning such a diagnosis because of the gravity of the diagnosis, and particular effort is made to avoid false-positives. Nevertheless, if the proportion of false-positives were sizeable, the prevalence rates based on the method we used would likely produce significant overestimates of need.

cant overestimates of need.

An epidemiologic survey in the neighboring province of Alberta yielded one-year prevalence rates of .4 per 100 population for schizophrenic disorders and .3 per 100 population for the more narrowly defined diagnosis of schizophrenia (15), whereas a recent systematic review of international prevalence studies of schizophrenic disorders found that averaging one-year prevalence rates produces a figure of .66 per 100 population for schizophrenic disorders and .41 per 100 population for schizophrenia (4). These findings from large-scale epidemiologic surveys are remarkably similar to the province's average rate of .46 per 100 population estimated in this study on the basis of administrative data. Thus, although we could not demonstrate criterion validity for the prevalence rates we found, there is an indication of face validity when these are compared with findings of community surveys.

It should be noted that large-scale epidemiologic studies are not immune to limitations that threaten validity—for example, incomplete case findings and suboptimal sensitivity and specificity in diagnosis. Thus they share some of the problems associated with the methods we used. In some respects, our methods may offer benefits over large-scale surveys. One advantage is that all diagnoses are registered by persons who, by definition, have received training in the identification and diagnosis of mental illness, in contrast with large-scale surveys that most commonly

have used lay interviewers.

A significant advantage of the method we used is in its recognition of every case diagnosed within the year. Consequently, this method obviates the need for sampling methods and eliminates the margin of error that is inevitable with even the most sophisticated sampling techniques. Furthermore, our method may be more likely to identify persons with schizophrenic disorders who are homeless or who cannot be contacted through telephone surveys and other traditional epidemiologic survey methods.

Two additional sets of findings support the likelihood that the administrative data supplied accurate information about the distribution of schizophrenic disorders across the geographic regions studied. First, the findings were highly consistent across each of three years in the 88 geographic regions, and this consistency is strengthened by the fact that data were not carried over from year to year. Rather, data were individually entered each year on the basis of diagnosis by physicians during outpatient visits or in hospital assessments and by mental health professionals during mental health clinic assessments. The level of correlation observed within all regions across time indicates that the data were highly reliable.

Second, we found significant correlations between the one-year prevalence of schizophrenic disorders and low income in the 88 LHAs. Thus the LHA with the highest prevalence of schizophrenic disorders in the province was the one comprising the Vancouver urban core, which had the greatest degree of poverty and social deprivation. In contrast, in high-income LHAs, rates of schizophrenia were very low. This relationship has been robust in other studies (5,7,16), and our replication of this finding suggests that the estimated distribution of prevalence rates is valid. No significant correlations were found between the one-year prevalence of schizophrenic disorders and unemployment, an indicator of social deprivation that on its own is a poor correlate of the distribution of schizophrenic disorders (16).

In jurisdictions that routinely collect and store administrative data, the costs associated with prevalence analyses such as those presented here

are minute compared with the costs of epidemiologic surveys. However, many jurisdictions do not have systems in place to record and maintain such databases or are unable to link data gathered in various settings. For example, applying this methodology to some regions in the United States may be challenging because of the existence of multiple independent provider agencies with differing data systems. These various administrative sources would need to be pooled within a defined geographic area to ensure that cases are not missed.

Initiatives are already under way in the United States to improve the integration of mental health information (17,18), which should make the methods described here more applicable to these regions in the future. Although it may not currently be feasible to link population data for an entire state in the United States, similar studies may be possible through health organizations that provide services to a large number of patients in a defined region.

In areas in which services for persons with schizophrenic disorders are not well developed or are in short supply, contact prevalence—whether derived from administrative data or by other means—is likely to be substantially lower than true prevalence and would underestimate service need. Notwithstanding these limitations, contact prevalence rates based on administrative data in areas with well-developed health services appear to provide a cost-effective means of examining the prevalence and distribution of schizophrenic disorders. This method constitutes an inexpensive means of complementing the findings of large-scale epidemiologic surveys with locally derived information relevant for resource planning and allocation. ♦

Acknowledgment

Funding support for this project was provided by the British Columbia Ministry of Health Services.

References

1. Wing J, Brewin CR, Thornicroft G: Defining mental health needs, in *Measuring Mental Health Needs*. Edited by Thornicroft G, Brewin CR, Wing J. London, Gaskell, 1992
2. McCrone P, Leese M, Thornicroft G, et al: Comparison of needs of patients with schizo-

phrenia in five European countries: the EP-SILON Study. *Acta Psychiatrica Scandinavica* 103:370–379, 2001

3. Torrey EF: Prevalence studies in schizophrenia. *British Journal of Psychiatry* 150:598–608, 1987
4. Goldner EM, Hsu L, Waraich P, et al: Prevalence and incidence studies of schizophrenic disorders: a systematic review of the literature. *Canadian Journal of Psychiatry* 47:833–843, 2002
5. Harvey CA, Pantelis C, Taylor J, et al: The Camden Schizophrenia Surveys: II. high prevalence of schizophrenia in an inner London borough and its relationship to socio-demographic factors. *British Journal of Psychiatry* 168:418–426, 1996
6. McCreadie RG, Leese M, Tilak Singh D, et al: Nithsdale, Nunhead, and Norwood: similarities and differences in prevalence of schizophrenia and utilization of services in rural and urban areas. *British Journal of Psychiatry* 170:31–36, 1997
7. Glover GR, Leese M, McCrone P: More severe mental illness is more concentrated in deprived areas. *British Journal of Psychiatry* 175:544–548, 1999
8. BC Statistics: People 25 Population Projections. Victoria, British Columbia, Ministry of Finance and Corporate Relations, July 2000
9. Statistics Canada: Data obtained from the Health Data Warehouse, Victoria, British Columbia Ministry of Health Services, Mar 1999
10. Access Policy for Research Uses of Linked Health Data. Vancouver, British Columbia, Center for Health Services and Policy Research, Health Information Development Unit, University of British Columbia, 2002. Available at www.chspr.ubc.ca/hidu/codetables/linkdap.pdf
11. Cain KC, Diehr P: Testing the null hypothesis in small area analysis. *Health Services Research* 27:267–294, 1992
12. Parikh SV, Lesage AD, Kennedy SH, et al: Depression in Ontario: under-treatment and factors related to antidepressant use. *Journal of Affective Disorders* 52:67–76, 1999
13. Crocetti GM, Lemkau PV, Kulcar Z, et al: Selected aspects of the epidemiology of psychoses in Croatia, Yugoslavia: 3. the cluster sample and the results of the pilot survey. *American Journal of Epidemiology* 94:126–134, 1971
14. Regier DA, Narrow WE, Rae DS, et al: The de-facto United States mental and addictive disorders service system. *Archives of General Psychiatry* 50:85–94, 1993
15. Bland RC, Newman SC, Orn H: Period prevalence of psychiatric disorders in Edmonton. *Acta Psychiatrica Scandinavica* 77:33–42, 1988
16. Glover GR, Robin E, Emami J, et al: A needs index for mental health care. *Social Psychiatry and Psychiatric Epidemiology* 33:89–96, 1998
17. National Committee for Quality Assurance: Measuring the quality of America's health-care. Available at www.nccqa.org, 2003
18. Mental Health Statistics Improvement Program. Available at www.mhsip.org, 2003