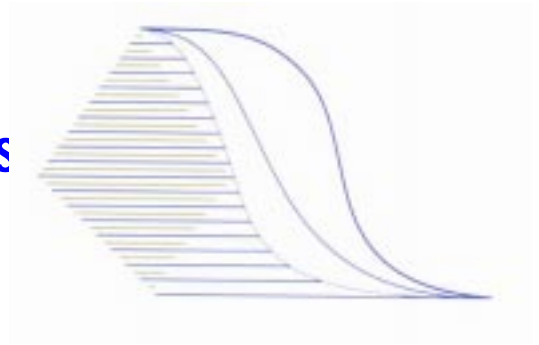


Centre for Health Services
and Policy Research



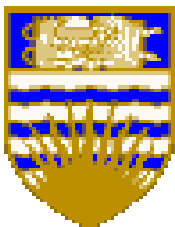
Where are we?
Individual geographic location
using two different sources of
postal code information in BC

Kimberlyn McGrail
and
Patrick Wong Fung

Health Information Development Unit
Research Reports

HIDU 99:01

December, 1999



THE UNIVERSITY OF BRITISH COLUMBIA

Introduction

Any regional analysis of health services utilisation is only as accurate as the patient geographic location information on which it is based. In BC, such analyses have until recently relied on the postal code(s) recorded on the MSP Registration and Premium Billing file (R&PB). In the past several years, Ministry of Health analyses have used postal codes from the Client Registry (CR), but these data have not been available to the research community outside the Ministry.

The CR began as a simple copy of the R&PB. This means that when the CR was first established, the postal codes recorded in it were exactly the same as those recorded in the R&PB. As time proceeded, the two information sources were expected to diverge because of different procedures for updating the postal codes recorded in them. The same postal code is recorded in both the CR and the R&PB for new Medical Services Plan registrants, but address changes are handled differently. Any changes made to the R&PB are reported to the CR, but the CR also draws on address changes reported from hospital separations, from the PharmaNet system, and from a number of other sources. The R&PB is, however, the single most important source of change in address information, accounting for more than half of all changes made to the CR.

The reason for the difference in updating procedures is the administrative responsibility for each of the systems. Billing and collection of premiums is the main function that the R&PB data system supports, so it is contract address (billing address) rather than home address that is of concern in that system. Home address is recorded at initial enrollment, and may be updated when there is a contract change (e.g. a change in employment), but otherwise, there is no motivation (or necessity) for the R&PB to keep up-to-date home address information. The CR, on the other hand, was designed to be a registry of all people who have had contact with the BC health care system, and was intended, in part, to create an accurate account of geographic location.

As a result, to the extent the CR and the R&PB are now different, the CR is assumed to be more accurate because of the multiple sources for address updating. In addition, the discrepancy between the two is commonly assumed to result in relatively large inconsistencies in regional analyses. Ultimately, the validity of research with a geographic (or socioeconomic) component using R&PB postal codes is questioned.

The purpose of this project was to compare, on an age-specific basis, the postal codes from these two sources as recorded in 1999. They have been compared at three levels: six-digit postal code, local health area (LHA) and Health Region (HR). The results should provide researchers with information on the reliability of geographic location based on the different sources and therefore about potential difficulties with regional analyses. The results will also provide an empirical basis for the decision about which source of postal codes ought to be used for research purposes.

Methods

The Ministry of Health provides the Centre for Health Services and Policy Research (CHSPR) with an annual 'snapshot' of the R&PB. This file is used to update the 'Linkage Coordinating File' (LCF), an amalgamation of yearly snapshots from 1986 through 1999 which forms the backbone of the BC Linked Health Data set (BCLHD). The LCF is the registry to which all administrative data sets are probabilistically linked (see: Chamberlayne R et al., (1998) *Creating a Population-based Linked Health Database: A New Resource for Health Services Research*. Canadian Journal of Public Health July/Aug 89(4):270-73). A file of PHNs¹ and postal codes as recorded in the June 1999 'snapshot' was created for this project.

¹ All PHNs in the BCLHD are scrambled. For ease of reading, however, they are referred to as 'PHNs' rather than 'scrambled PHNs'.

In October 1999, the Ministry of Health provided CHSPR with a file derived from the CR that contains PHN and (then) current postal code for all 'active' (i.e. non-retired) PHNs.² PHNs are only 'retired' when the individuals to whom they are assigned die. Thus, active implies neither residence in the province nor registration with MSP. We know, in fact, that the list provided by the CR will contain many people not in the LCF. For example, residents of Alberta or other provinces who fill a prescription in BC are assigned a PHN through the CR, based on the policy that everyone who receives any kind of health care service in BC be given one.

The first step was to merge the CR data with the extract of 1999 postal codes from the LCF. The second was to retain all PHNs that appeared in both, along with LCF postal code and CR postal code. Next, postal codes from both sources were converted, using the latest version of the BC Stats Translation Master File (TMF), to LHA and HR. Finally, age was calculated as of 30 June 1999, based on the birthdate recorded in the LCF. Individuals were assigned to age groups 0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, or 80+.

There were 209,077 PHNs with postal codes that were not in the TMF, and thus could not be assigned a LHA or HR. This can occur when a postal code has been introduced subsequent to the yearly publication of the TMF. Affected postal codes were included in the analyses, since comparison at the 6-digit level was clearly possible. In all cases when 6-digit postal code agreed, LHA and HR were automatically set to agreement, minimising the effect of postal codes missing from the TMF. Missing LHA and HR values will still occur, however, when either the LCF or CR postal code cannot be converted; this may result in a slight underestimate of agreement at these levels.

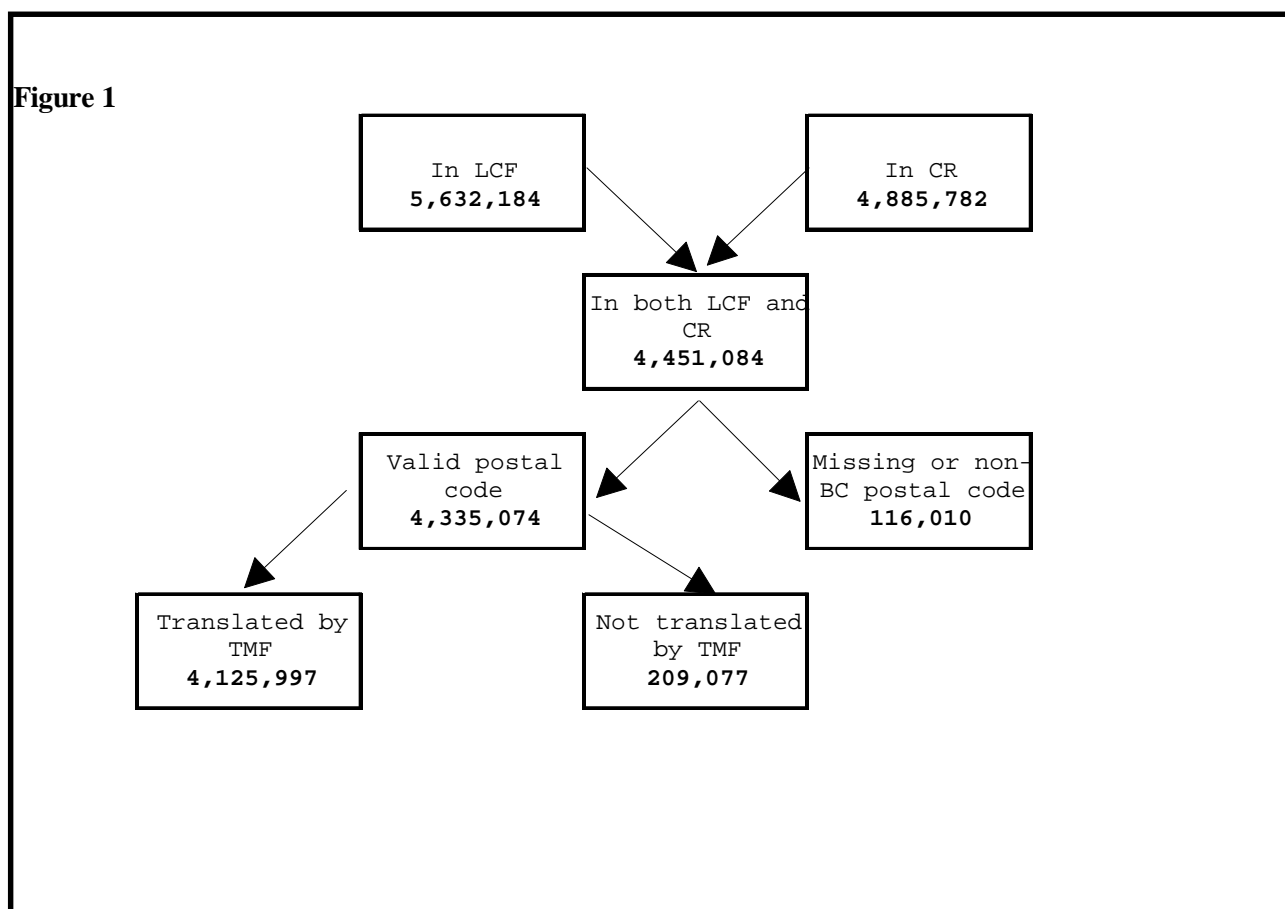
The following information was produced:

1. A frequency of agreement (and disagreement) between the two six-digit postal codes, by age group.
2. A frequency of agreement (and disagreement) between the two LHA codes, by age group.
3. A frequency of agreement (and disagreement) between the two HR codes, by age group.
4. 1, 2 and 3 above, excluding blank and non-BC postal codes.

Results

The 1999 LCF contains 5,632,184 unique PHNs, and the CR contains 4,885,782 (Figure 1). There are 1,181,100 unique PHNs on the LCF that are not on the CR. This occurs both because the LCF has records of people registered as early as 1986, even if that registration is no longer valid (compared to the early 1990s for the CR) and because it includes people who have died. The CR, on the other hand, contains 434,698 unique PHNs not included in the LCF. This occurs because of out-of-province residents who receive health services in BC (emergency hospital use, filling of a prescription) and people who are BC residents and received services, but who are not registered with the province (either because of eligibility status or because of non-payment of premiums).

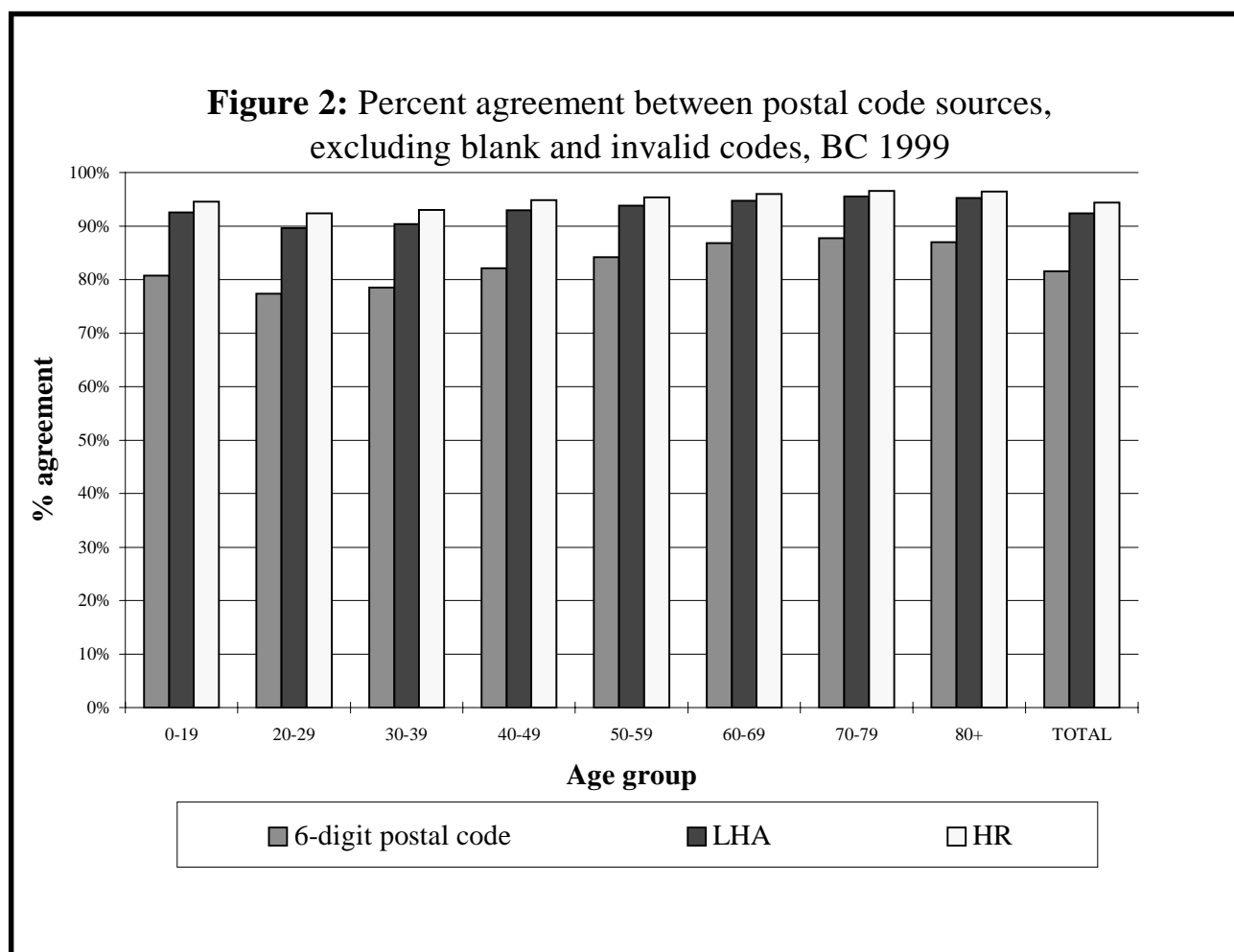
² The time delay between the R&PB snapshot (created in June) and the CR extract (created in October) may explain a portion of any discrepancy in postal codes. The effect of this, however, is assumed to be small.



There are 4,451,084 unique PHNs that appear on both the LCF and the CR. There was no attempt made to see whether these are current registrants, but we know some must not be because that number is about 10% larger than the estimated BC population in 1999. Of these 4,451,084 there were 116,010 that did not have a valid BC postal code on either the CR or the LCF (blank or beginning with something other than 'V'). We calculated comparison percentages including and excluding these PHNs. The differences were less than 1% in all cases, so the numbers reported here exclude these, leaving 4,335,074 unique scrambled PHNs with postal codes to be compared.

As expected, there is an increase in agreement as the level of aggregation increases from 6-digit postal code to LHA to HR (Figure 2). There is also a trend with age, with the 20-39 age groups showing the lowest level of agreement. This is expected because of higher intra-provincial migration in these age groups. The level of agreement even in these age groups, however, is still relatively high; above 75% at the 6-digit postal code level, and 90% or higher for LHA and HR

At the highest level of aggregation - the Health Region - agreement in the other age groups ranges from about 95 to 97%.



Discussion

The CR and R&PB have behaved as expected, and in 1999 contain different postal codes for about one in five registrants. Because the CR started as a copy of the R&PB, we can infer that if the same comparison were made between the 1994 CR and 1994 LCF (for example), there would be a higher level of agreement than in 1999. Nevertheless, these results indicate that even several years from the start of the CR, there is still a high degree of concordance between the two sources, and that agreement is increased further when the postal codes are aggregated to LHA and HR. The most likely explanation for this is that people often change residences within the same community. In this case, if the address was updated on the CR and not the R&PB, the postal code would disagree, but the LHA and HR would be unaffected.

The preceding analysis does not answer all questions that might be of interest. For example, there are other geographic boundaries used for research and planning purposes, such as census tract or census sub-division. Such comparisons between the CR and LCF are possible, but the result would likely fall somewhere between the 6-digit postal code and LHA comparisons.

We also did not address the ecologic construction of individual or family socioeconomic status. Socioeconomic status might, for example, be calculated based on enumeration area– a very small geographic location where difference in six-digit postal code might have a large impact – or census tract – a larger area where differences in 6-digit postal code might matter less. Because of the small number of categories (either five for SES quintiles or 10 for deciles), we would expect some agreement based on chance alone³. Our best guess is that SES comparisons would probably have agreement in the same range as HR comparisons.

Even though the difference between the CR and LCF – especially at the LHA and HR levels – is relatively small, it still may be worth addressing. There is, for example, differential agreement among the age groups, probably because of patterns of mobility. This mobility may be of some interest, so capturing as much as possible of it (by using a full ‘string’ of postal codes from the CR) is potentially important. For example, if ‘high mobility’ people use health care services more frequently than the average similarly-aged BC resident, the impact on geographic analyses may be greater than expected based on the magnitude of disagreement.

Another consideration is the difference in trends or projections that might occur in Ministry of Health reports vs. reports produced by external researchers. If these differences (whether they are in fact large, or are only perceived to be large) decrease the credibility of analyses using geographic location drawn from the LCF, then they might be of concern to the research community, or to the Ministry. There is something to be said for consistent results emerging from similar analyses of common databases. This will become even more important as the Ministry proceeds with building a Health Data Warehouse and increases the number of people accessing data that way.

For these reasons, it makes sense to consider providing historical postal code information from the CR to researchers accessing data through the BCLHD. These postal codes would be considered additional to those recorded in the LCF, and would provide researchers the opportunity to do ‘sensitivity analyses’ comparing the results of one source vs. the other. Since neither can truly be considered a ‘gold standard’, it is most likely that the ‘truth’ would be found in the range produced by such analyses.

Acknowledgments:

Thanks to Morris Barer and Clyde Hertzman for comments on an earlier draft. Funding and data for this work were provided by the Ministry of Health and Ministry Responsible for Seniors.

³ A smaller number of possible values implies a higher level of assignment of the ‘correct’ purely due to luck.