Appendix E

A Note on the Record Linkage

As noted in Chapters 2, 3, and 4, techniques of nominal record-linkage were employed in this study to join records of individuals from several distinct sources. The linkage conducted for this research was totally manual, as opposed to computer-based procedures. To proceed by hand was a requirement rather than a choice, for my search for individuals identified in the 1861 manuscript census involved a relatively small proportion of the population of any given place, and not entire lists of inhabitants. (The census-to-assessment linkage of literates, however, had been accomplished through use of a semiautomated system by the Canadian Social History Project.) Consequently, the linkage was based on an examination of complete lists of census-enumerated and assessed individuals in each of the three cities in 1861 and 1871. (For references to the literature on nominal record-linkage in history, see the footnotes to Chapter 3.) My system was based upon the rules for computer-assisted linkage developed by Ian Winchester and the Canadian Social History Project; however, I did not compute weighted scores or additive totals to judge each potentially linked pair of records.

Two types of linkage formed the structure of data-collecting for this study: that over a short time span (census-to-assessment, 1861, 1870–1871–1872) and decennial searches (census-to-census, 1861–1871, four-way census-to-assessment linkages, 1861–1871). While both were conducted manually, the basis of the search and the rules for acceptance of records as truly linked differed.

The criteria for judging possible records as linked and selection or rejection were based upon a hierarchy of individual characteristics. For census-to-assessment linkage, the surname of the individual (heads of

household only in this case) was obviously the most important identifiable variable. For a record to be accepted, then, the name had to be either the same or a close approximation in spelling or pronunciation. Initials, or in some cases, first names, were important too, although as expected they were a less reliable indicator. Age, when given in both sources, was important as well. For a link to be made, age had to be approximately the same in both sources (with some allowance for age-heaping at 5-year intervals; i.e., 30, 35, 40) unless all other variables were the same. Occupation and ward of residence followed in importance, as I assumed that there would be little residential or occupational mobility within a short time span. Generous allowance, however, was made for cases of occupational equivalence, and if name and age were the same in both sources, a link was made, regardless of occupation or residence. Moreover, the 1871 assessments included family size and school support (public or separate: an indication of Protestantism or Catholicism) adding important variables to the roster of comparison. If it was impossible to choose between two cases, especially in those involving a common name like Sullivan, none were selected; this rule, however, was seldom applied. In all cases, these criteria were stringently employed; consequently, all links were made with a relatively large degree of confidence, and the rates of persistence must be considered conservative ones.

The census-to-census search involved all illiterates, regardless of household status, and all children present in their families in 1861. With these links, name was also the key variable for selection, followed by sex, birthplace, religion, age, and initials. More liberal allowance was made for aging, as far from all "aged" a full 9 or 10 years over the decade, as was reported by them. Nevertheless, when considered as a whole, those linked did age between 9 and 10 years from 1861 to 1871. Sex, however, had to be the same for a link to be accepted. Birthplace and religion, importantly, were the same in virtually all cases, excepting some children who were reported as native-born Canadians by 1871 when they had been foreign-born a decade earlier. These involved few cases. If religion and birthplace were not identical, a match was not made unless all other variables agreed. For heads of households, and for most children, information on the family was important. For example, a head of household's match was supported by the presence of the spouse, unless marital status had changed from married to widowed. Names and ages of spouses and children as well as their birthplaces were useful in selection, if children aged 9 years or older remained at home, which of course the majority did. Similarly, the decision to link a child depended upon his or her parents' or parent's identity in both years as

well as that of siblings. If, however, household or marital status changed (and these could not be used as independent criteria for judgment), name, age, sex, initials, birthplace, religion, and other family data (such as for a woman who became a widow and head of household) were the only judgments for a link. Therefore, the fewer the variables, the more rigorously the rules for equivalence and exactness were applied.

In census-to-census linkages, neither occupation nor place of residence could be used in making the decision to link. Comparison of change in these variables was the object of the linkage, and therefore they were ignored in judgments to accept or reject a potentially linked pair. As with the other linkage, these rules were rigorously applied, and the resulting set of matched pairs of records must be considered conservative. If biased, then, the linkage results would be in favor of greater transiency rather than persistence, a choice for caution and not greater numbers.

Finally, the four-way linkage was in many respects the simplest, involving the fewest possible matches, the shortest lists, and the greatest number of useful variables. The rules were much the same as those employed above; however, as with occupation and residence in census-to-census linkages, neither wealth nor homeownership was used in accepting or rejecting a record: their changes were the object of study. I refer the concerned reader to the work of Ian Winchester, the files of the *Historical Methods Newsletter,* and to E. A. Wrigley, ed., *Identifying People in the Past* (cited in Chapter 3), for both theoretical and practical discussions of nominal record-linkage and the problems encountered by researchers.