Genetic fingerprinting in the study of tuberculosis transmission

Sophie Kulaga, MSc; Marcel A. Behr, MD, MSc;
Kevin Schwartzman, MD, MPH

The epidemiology of infectious diseases has traditionally relied on observed patterns of occurrence to infer transmission. In the case of a disease with a variable incubation period, such as tuberculosis (TB), a series of overlapping micro-epidemics can appear as a relatively steady notification rate at the population level. Recently, genetic fingerprinting has been used in epidemiologic studies to begin to assess such TB micro-epidemics.

A crucial aspect of any TB control program is the ability to determine where transmission is occurring in order to prevent further spread of infection and prevent active disease by identifying newly infected people and providing them with preventive therapy. Genetic fingerprinting of *Mycobacterium tuberculosis* has vastly improved our ability to observe patterns of transmission in populations. It has helped to establish transmission links between individuals and to demonstrate instances in which related people were infected with unrelated strains.

Since the 1980s the number of TB cases reported annually in Canada has been stable, at about 2000, translating into an incidence rate of about 7 cases per 100 000 annually.[1] During this time, however, the groups at risk have changed. In 1980 the majority of cases of active disease occurred among elderly nonaboriginal people born in Canada whose disease was most likely the result of reactivation of remotely acquired infection. By 1994 the majority of cases involved foreign-born Canadian residents and aboriginal Canadians. The estimated incidence of TB among aboriginals is 80 per 100 000 annually, a rate 40 times that among nonaboriginal people born in Canada.[2] Among foreign-born people the incidence rates vary greatly, usually reflecting rates in the country of origin.[3] A substantial increase in TB risk among HIV-infected people has been well documented.[4] Although the absolute number of HIV-related TB cases has been relatively modest in Canada, pockets of concurrent infection are increasingly observed in urban settings.

Therefore, relatively stable overall rates of TB in Canada mask recent shifts in the epidemiology of the disease. Without an understanding of these changing trends it is unlikely that control programs will be able to address current and future challenges.

In this article we describe the methods and review the progress, advantages and limitations of a powerful application of molecular biology to the investigation of an ancient human disease.

Molecular epidemiology

Classic epidemiologic studies inferred TB transmission on the basis of plausible opportunities for contagion. Two or more cases were considered linked if there was judged to be sufficient sharing of environment for transmission to have occurred.

In the early 1990s laboratory techniques that had been limited to genetic research became available for more widespread epidemiologic and public health investigation. These techniques permit characterization of infecting organisms with "DNA fingerprints," which are then combined with traditional clinical, epidemiologic and public health data.

Restriction fragment length polymorphism (RFLP) analysis is the molecular technique most often used to produce genetic fingerprints of TB isolates. Chromosomal DNA of *M. tuberculosis* contains repetitive sequences of base pairs called "insertion sequences" that are variably distributed throughout the organism's

genome.[5] Laboratory methods use the number and location of these elements to produce a fingerprint for each isolate. The most commonly used insertion sequence in studies of *M. tuberculosis* is known as IS6110.[6]

To produce the fingerprint, the DNA of the organism is cut using a restriction endonuclease, which consistently recognizes and cuts at a specific sequence of base pairs (the restriction site). The result is thousands of DNA fragments of different lengths that are then separated according to size by gel electrophoresis. These separated fragments are transferred to a membrane (Southern blotting) for hybridization with a probe that recognizes the DNA sequence of IS6110. Most isolates of *M. tuberculosis* will produce a DNA fingerprint of 1 to 25 bands of DNA that hybridize IS6110 (Fig. 1). The combination of fragment number and size provides a specific banding pattern characteristic of that isolate: the number of bands corresponds to the number of IS6110 elements, and the distance they travel along the gel reflects the molecular weight of the IS6110-containing fragment (Fig. 2).

When DNA fingerprints are studied across a population, it is not feasible to compare patterns visually because the complexity of different patterns becomes overwhelming after several dozen isolates are studied. Instead, DNA fingerprints are first scanned into a computer for computer-assisted visual comparisons of the banding patterns.

The genetic fingerprints produced before and after repeated laboratory passage of the same isolate have been shown to be identical, indicating that the DNA fingerprint is a stable property of an isolate. As well, in people with active TB from whom multiple samples have been obtained over time, the DNA fingerprints have been found to be relatively stable.[7] Recently, De Boer and associates analysed all repeat isolates from the Netherlands and estimated that the half-life of an IS6110-based DNA fingerprint is about 3 years.[8] Because it is estimated that about 50% of new cases of active TB occur within 2 years of infection, these data suggest that changes in DNA fingerprints occur at a slightly slower rate than secondary cases appear, and therefore most cases in an outbreak should present with very similar, if not identical, DNA fingerprints.

Although the molecular methods for DNA fingerprinting are well understood and standardized,[9] the application of these patterns to epidemiologic analysis remains somewhat uncertain. Among the issues that remain to be resolved are the optimal means of interpreting results and their ultimate utility outside of the research setting. If 2 cases yield completely different fingerprints, we can infer that transmission has not occurred. Conversely, in known outbreaks, the search for identical DNA fingerprints is used to determine the extent of transmission.[10] However, when isolates from 2 patients happen to have identical fingerprints, we cannot necessarily infer a direct transmission link, particularly if others in the community have the same DNA fingerprint or, most important, if there is only a limited diversity of DNA fingerprints in that community.

The idea that identical fingerprints reflect infection by

an organism of the same lineage is grounded in solid biology. However, the application of this information to understanding transmission depends in large part on whether that commonality indeed reflects recent transmission. Molecular methods have added greatly to our understanding of TB, but these results must always be interpreted in the context of epidemiologic information.

Contributions of DNA fingerprinting

Characterization of outbreaks

RFLP techniques for DNA fingerprinting have been used to confirm TB transmission where it was suspected by clinicians or public health authorities. Hospital-based studies have used RFLP results to corroborate the epidemiologic evidence of nosocomial transmission among people hospitalized with AIDS.[11] Although accelerated progression from TB infection to active disease in HIV-infected people has previ-
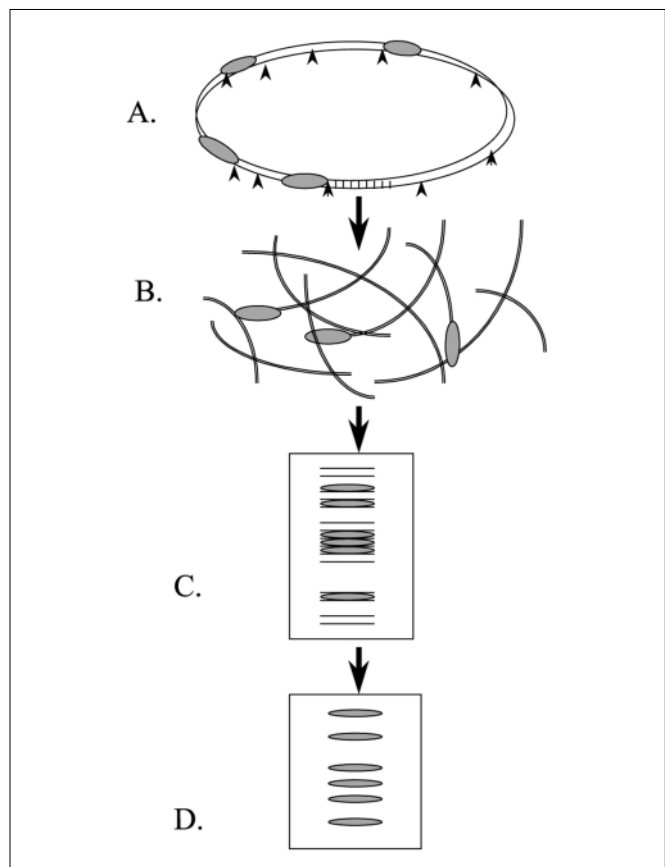

Fig. 1: Genetic fingerprinting of *Mycobacterium tuberculosis* isolates: (A) Restriction enzymes cleave chromosomal DNA at restriction sites (arrowheads). (B) Some DNA fragments contain IS6110 (repetitive sequences of base pairs, represented as shaded ellipses). (C) The fragments are separated according to size by gel electrophoresis. (D) Fragments containing IS6110 hybridize to the specific radioactive probe, which produces a characteristic banding pattern (fingerprint) for each isolate.

ously been reported, RFLP techniques have been used to confirm the route of transmission and the speed of progression.[10] DNA fingerprinting has also repeatedly demonstrated that TB transmission can occur even in the absence of "close personal contact." A good example is an outbreak that was found to be centred in a neighbourhood bar.[12]

Identification of laboratory cross-contamination

A pseudo-outbreak occurs when 2 patients are said to have TB, but spurious "transmission" was in fact the result of contamination of one patient's sample by organisms from a sample from a patient with active TB. False-positive cultures due to spillage during laboratory processing of samples[13] or to transfer of organisms between culture vials via the needles used to monitor growth[14] have been demonstrated by RFLP analysis.

Evaluation of treatment failure

Use of DNA fingerprinting has documented exogenous re-infection during or following successful treatment of active TB in AIDS patients.[15] Patients who had been judged to be cured of fully drug-sensitive active TB subse-
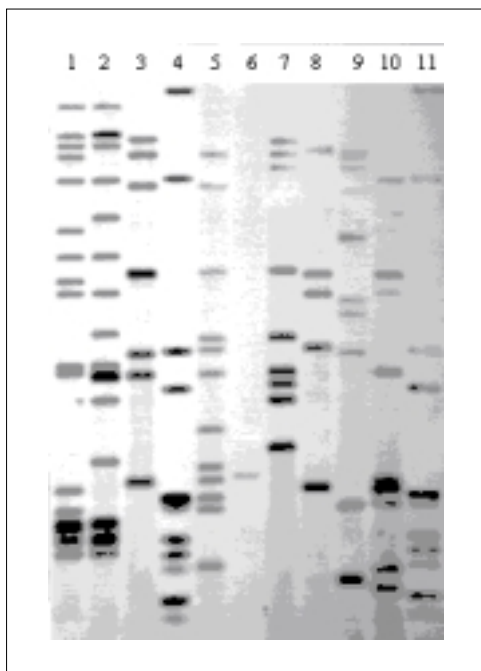

Fig. 2: Autoradiograph, showing IS6110-based DNA fingerprints of *M. tuberculosis* isolates. Lanes 4 and 11 represent molecular weight ladders that provide information about the length of DNA fragments. The other 9 lanes contain whole genomic DNA of different isolates. Lane 6 contains a one-copy strain whose fingerprint is of low resolution; to investigate links between such a strain and others with the same one-copy fingerprint, secondary typing with another marker system is usually performed.

quently became ill with drug-resistant disease. In the past, such apparent relapses were ascribed to acquired drug resistance because of poor treatment compliance. RFLP analysis has shown that some of these "relapses" were in fact the result of newly acquired infection with a different, drug-resistant strain. More recently, such exogenous re-infection has also been observed in patients without HIV infection.[16] Similarly, RFLP analysis has demonstrated a case of active TB that resulted from 2 *M. tuberculosis* strains with distinct fingerprints.[17]

Determination of recent transmission

RFLP surveys have indicated that, in some populations, the proportion of people with active TB felt to be due to recent transmission appears to be much higher than previously estimated.[18–21] Whereas recent transmission had been thought to account for only about 10% of all active TB cases, the high proportion of shared fingerprints in cities such as San Francisco, New York, Bern and Los Angeles suggested that 25%–50% of active TB cases in those areas were the result of recent transmission. These results were supported by post hoc evidence of epidemiologic relatedness based on questionnaires and other traditional methods.

Identification of specific patterns of transmission

Researchers in San Francisco have used DNA fingerprinting to infer that most cases of TB among foreign-born people were due to reactivation of remote infection.[22] Foreign-born people were seldom members of RFLP-defined "clusters," for which there was epidemiologic confirmation of transmission. The same authors identified the single most important weakness of the TB control program to be the failure of contact investigation to identify infected contacts. In nearly 70% of the clusters, the source case did not list the eventual secondary cases as contacts.[22] RFLP analysis has recently been used to demonstrate a decrease in secondary TB cases as a result of intensified control measures in the San Francisco area.[23]

Characterization of communicability and pathogenesis

DNA fingerprinting has demonstrated that people with smear-negative, culture-positive TB are able to transmit infection.[24] Although these patients have usually been considered to be a trivial risk for contagion, and their disease is not diagnosed in most parts of the world, it was estimated that at least 17% of secondary active cases in San Francisco were attributable to smear-negative patients.

TB transmission appears to occur easily in some outbreaks and less so in others. Transmission rates have been thought to vary with host factors and environmental characteristics. RFLP techniques were used to document an

outbreak where unusually high transmission rates appeared to reflect characteristics of the infecting *M. tuberculosis* strain itself.[25] Among HIV-seronegative people who had active TB within 4 years of contact with an infectious index case, DNA fingerprinting suggested a mean interval of 21 weeks between the relevant contact and the development of active disease.[26]

Limitations

In Arkansas, which has a stable, rural population and a low prevalence of HIV infection, no epidemiologic connection could be established for almost 60% of TB patients found to be in RFLP-defined clusters, despite extensive life histories of work, social activities and residence, obtained by personal interview.[27] These results suggest that some RFLP-defined clusters may in fact represent coincidental reactivation of endemic strains. Therefore, one must approach with caution the conclusions of transmission studies that rely on the assumption that shared fingerprints always reflect recent transmission. When 2 people in a community are found to have active TB caused by organisms with the same genetic fingerprint, several explanations must be considered: recent transmission; simultaneous reactivation of remotely acquired infection with the same organism; predominance of a local strain; or laboratory error. In practice, it is often difficult to distinguish between scenarios, and the optimal interpretation of molecular results requires clinical and epidemiologic tools. Moreover, optimal interpretation and application of molecular typing results may vary according to population characteristics — an important and as yet unresolved issue facing researchers and public health professionals who deal with TB.

IS6110-based RFLP analysis is the most widely used technology in molecular epidemiologic studies of TB. Most *M. tuberculosis* isolates contain multiple copies of IS6110; however, some isolates have very few copies, or even none of these fragments.[28] Secondary typing methods, most often using the polymorphic GC-rich sequence, have been used in a number of studies to overcome such technical limitations. A method based on the polymerase chain reaction — spacer oligonucleotide typing (spoligotyping) — has recently been developed for the rapid typing of *M. tuberculosis*[29] and is being evaluated for its usefulness in the differentiation of isolates with few, or no, copies of IS6110.[30]

Genetic fingerprinting can be used only in cases of active TB. It cannot be used to identify people with latent infection (about 90% of infected people), including those at high risk of future reactivation. In fact, use of RFLP-typing in transmission studies introduces bias that tends to highlight the determinants of rapid progression to active disease. For instance, a 2-year study window could easily capture HIV-associated TB outbreaks, but it would capture only a minority of the HIV-negative people who acquire TB infection and ultimately active disease.

In any transmission study, geographic and temporal limits of the study dictate that a certain number of cases with links outside the study confines will appear unique. Recently, concerns have been raised regarding the influence of sampling fractions and small cluster sizes on the proportion of isolates with identical fingerprints.[31] With the establishment of international RFLP and epidemiologic databases spanning many years, researchers may overcome these limitations.

Future directions

In Canada DNA fingerprinting is currently used mainly for research and targeted outbreak investigations. Other countries have begun to use this technique much more extensively for public health practice. For example, in the Netherlands every *M. tuberculosis* isolate undergoes RFLP typing. Identical fingerprints trigger in-depth contact investigation to identify opportunities that existed for transmission between the individuals involved.[26] The cost-effectiveness of such an approach and its impact on TB control, compared with targeted use in public health, have yet to be determined.

Genetic fingerprinting could be used to assist clinicians with diagnostic challenges and treatment decisions. For instance, RFLP results can be used to distinguish relapse from re-infection with a different TB strain. The method also allows the characterization of unusual clinical isolates that would give uncertain results on standard phenotypic testing.[32] In the future RFLP may be used to link newly diagnosed TB cases to drug-resistant isolates already present in the community, and this may allow clinicians to tailor drug therapy more rapidly.

RLFP surveillance databases are currently being established,[33] by which transmission between geographically distant locations can be traced. As political and economic instability continue to fuel large-scale migration from TB-endemic regions, DNA fingerprinting will allow researchers and public health authorities to track the global spread of TB. Containment and ultimate eradication of this ancient disease will only be possible if control efforts target TB where it most frequently develops and spreads. Although prompt diagnosis and appropriate management of active TB cases must remain the priority in international control efforts, it is hoped that this powerful new technology will enhance global control by improving our understanding of the pathogenesis and spread of TB.

References

1. Long R, Njoo H, Hershfield E. Tuberculosis: 3. Epidemiology of the disease in Canada. *CMAJ* 1999;160(8):1185-90.
2. Fitzgerald JM, Black WA, Kunimoto D. Evaluation of non-HIV-related drug-sensitive cluster outbreaks of tuberculosis with PCR-based DNA fingerprinting. *Can Respir J* 1996;3:317-21.
3. Brancker A, Enarson DA, Grzybowski S, Hershfield E, Jeanes CW. A statistical chronicle of tuberculosis in Canada: Part II. Risk today and control. *Health Rep* 1992;4:277-92.
4. Selwyn PA, Hartel D, Lewis VA, Schoenbaum EE, Vermund SH, Klein RS, et al. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. *N Engl J Med* 1989; 320:545-50.
5. Behr MA, Small PM. Molecular fingerprinting of *Mycobacterium tuberculosis*: how can it help the clinician. *Clin Infect Dis* 1997;25:806-10.
6. Yang Z, de Haas PEW, Wachman CH, van Soolingen D, van Embden JDA, Andersen AB. Molecular epidemiology of tuberculosis in Denmark in 1992. *J Clin Microbiol* 1995;33:2077-81.
7. Cave MD, Eisenach KD, McDermott JH, Bates JH, Crawford JT. IS6110 Conservation in the *Mycobacterium tuberculosis* complex and its utilisation in DNA fingerprinting. *Mol Cell Probes* 1991;5:73-80.
8. De Boer AS, Borgdorff MW, de Haas PE, Nagelkerke NJ, van Embden JD, van Soolingen D. Analysis of rate of change of IS6110 RFLP patterns of *Mycobacterium tuberculosis* based on serial patient isolates. *J Infect Dis* 1999;180 (4):1238-44.
9. Van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;31:406-9.
10. Daley CL, Small PM, Schecter GF, Schoolnik GK, McAdam RA, Jacobs WR Jr, et al. An outbreak of tuberculosis with accelerated progression among persons infected with the human immunodeficiency virus. An analysis using restriction-fragment-length polymorphisms. *N Engl J Med* 1992;326:231-5.
11. Edlin BR, Tokars JI, Grieco MH, Crawford JT, Williams J, Sordillo EM, et al. An outbreak of multidrug-resistant tuberculosis among hospitalized patients with the acquired immunodeficiency syndrome. *N Engl J Med* 1992;326:1514-21.
12. Kline SE, Hedemark LL, Davies SF. Outbreak of tuberculosis among regular patrons of a neighborhood bar. *N Engl J Med* 1995;333:222-7.
13. Small P, Singh SP, Schoolnik GK, Tompkins LS, Mickelsen PA. Molecular strain typing of *Mycobacterium tuberculosis* to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. *J Clin Microbiol* 1993;31:1677-82.
14. Wurtz R, Demarais P, Trainor W, McAuley J, Kocka F, Mosher L, et al. Specimen contamination in mycobacteriology laboratory detected by pseudo-outbreak of multidrug-resistant tuberculosis: analysis by routine epidemiology and confirmation by molecular technique. *J Clin Microbiol* 1996;34:1017-9.
15. Small PM, Shafer RW, Hopewell PC, Singh SP, Murphy MJ, Desmond E, et al. Exogenous reinfection with multidrug-resistant *Mycobacterium tuberculosis* in patients with advanced HIV infection. *N Engl J Med* 1993;328:1137-44.
16. Shafer RW, Singh SP, Larkin C, Small P. Exogenous reinfection with multidrug-resistant *Mycobacterium tuberculosis* in an immunocompetent patient. *Tuber Lung Dis* 1995;76:575-7.
17. Yeh RW, Hopewell PC, Daley CL. Simultaneous infection with two strains of *Mycobacterium tuberculosis* identified by restriction fragment length polymorphism analysis. *Int J Tuberc Lung Dis* 1999;3:537-9.
18. Barnes PF, El-Hajj H, Preston-Martin S. Transmission of tuberculosis among urban homeless. *JAMA* 1996;275:305-7.
19. Genewein A, Telenti A, Bernasconi C, Mordasini C, Weiss S, Maurer AM, et al. Molecular approach to identifying route of transmission of tuberculosis in the community. *Lancet* 1993;342:841-4.
20. Small PM, Hopwell PC, Singh SP, Paz A, Parsonnet J, Ruston DC, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med* 1994;330:1703-9.
21. Alland D, Kalkut GE, Moss AR, McAdam RA, Hahn JA, Bosworth W, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med* 1994;330:1710-6.
22. Chin DP, DeRiemer K, Small PM, de Leon AP, Steinhart R, Schecter GF, et al. Difference in contributing factors to tuberculosis incidence in U.S.-born and foreign-born persons. *Am J Respir Crit Care Med* 1998;158:1797-1803.
23. Jasmer RM, Hahn JA, Small PM, Daley CL, Behr MA, Moss AR, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. *Ann Intern Med* 1999;130:971-8.
24. Behr MA, Warren SA, Salamon H, Hopewell PC, Ponce de Leon A, Daley CL, et al. Transmission of *Mycobacterium tuberculosis* from patients smear-negative for acid-fast bacilli [published erratum appears in *Lancet* 1999;353:1714]. *Lancet* 1999;353:444-9.
25. Valway SE, Sanchez MP, Shinnick TF, Orme I, Agerton T, Hoy D, et al. An outbreak involving extensive transmission of a virulent strain of *Mycobacterium tuberculosis* [published erratum appears in *N Engl J Med* 1998;338:1783]. *N Engl J Med* 1998;338:633-9.
26. Ten Asbroek AH, Borgdorff MW, Nagelkerke NJ, Sebek MM, Deville W, van Embden JD, et al. Estimation of serial interval and incubation period of tuberculosis using DNA fingerprinting. *Int J Tuberc Lung Dis* 1999;3:414-20.
27. Braden CR, Templeton GL, Cave MD, Valway S, Onorato IM, Castro KG, et al. Interpretation of restriction fragment length polymorphism analysis of *Myobacterium tuberculosis* isolates from a state with a large rural population. *J Infect Dis* 1997;175:1446-52.
28. Agasino CB, Ponce de Leon A, Jasmer RM, Small PM. Epidemiology of *Mycobacterium tuberculosis* strains in San Francisco that do not contain IS6110. *Int J Tuberc Lung Dis* 1998;2:521-3.
29. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, Kuijper S, et al. Simultaneous detection and strain differentiation of *Myobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* 1997;35:907-14.
30. Cousins D, Williams S, Liebana E, Aranaz A, Bunschoten A, van Embden J, et al. Evaluation of four DNA typing techniques in epidemiological investigations of bovine tuberculosis. *J Clin Microbiol* 1998;36:168-78.
31. Glynn JR, Vynnycky E, Fine PE. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. *Am J Epidemiol* 1999;149:366-71.
32. Gutierrez MC, Galan JC, Blazquez J, Bouvet E, Vincent V. Molecular markers demonstrate that the first described multi-drug resistant *Mycobacterium bovis* outbreak was due to *Mycobacterium tuberculosis*. *J Clin Microbiol* 1999;37:971-5.
33. Casper C, Singh SP, Rave S, Daley CL, Schecter GS, Riley LW, et al. The transcontinental transmission of tuberculosis: a molecular epidemiological assessment. *Am J Public Health* 1996;86:551-3.

Reprint requests to: *Dr. Kevin Schwartzman, Respiratory Epidemiology Unit, Joint Department of Epidemiology and Biostatistics and of Occupational Health, McGill University, 1110 Pine Ave. W, Montreal QC H3A 1A3; fax 514 398-8981*