

## ***Sampling Methodology and its Application: An Illustration of the Tension Between Theory and Practice***

by EVELYN KOLISH\*

### ***Résumé***

Cet article étudie les tensions entre la théorie et la pratique de l'échantillonnage des dossiers de l'arriéré du projet d'échantillonnage des dossiers judiciaires du Québec. L'article examine d'abord les raisons qui ont motivé le choix de l'échantillonnage statistique des dossiers des cours de justice comme instrument de tri tel qu'énoncé dans le *Report of the Interministerial Committee on Court Records*. Une seconde partie explique la méthode spécifique d'échantillonnage développée par le Comité et se penche sur les diverses difficultés d'application rencontrées durant les deux dernières années: système de classement maison hétérodoxe, classement alphabétique, juridictions partagées, et l'impact des pardons. L'article termine en suggérant que malgré le fait que plusieurs problèmes peuvent être résolus sans compromettre la justesse de cette méthode, que quelques obstacles exigent un tel investissement de ressources humaines à être surmonté, que l'échantillonnage ne représente plus une solution rentable. Les archivistes et les gestionnaires devront faire la part des choses entre les exigences de la représentativité et la précision et leurs coûts en ressources humaines et matérielles.

### ***Abstract***

In the context of a conference on archivists caught between "the rock and a hard place," sampling offers several ways of illustrating the compromises that inevitably take place in the dialectic between theory and practice. This article examines briefly two of the ways in which sampling confronts us with the tension between theory and practice: first, the choice of sampling as a method of selection, in and of itself; second, the practical problems of applying a specific sampling methodology to the untidy reality of a mass of accumulated documents. Both these reflections are set in the specific context of Quebec's reten-

tion decisions with regards to court records and the Archives nationales du Québec's recent experience in applying a particular sampling method to those records.

### *Appraisal and Sampling*

In some ways, statistical sampling can be perceived as a desperate reaction to unmanageable masses of records, and as a way of reducing the size of the archival haystacks through which archivists must provide signposts for researchers pursuing their individual intellectual needles. The use of sampling techniques pushes the appraisal process back from the level of the record to that of the record series. Generally, it is easiest to contemplate sampling when dealing with series that are routine or repetitious in nature, where the informational value of individual records is both relatively consistent and not extremely high—yet not low enough to merit either outright elimination or the taking of simple specimens.

Sometimes, however, the lack of resources for appraisal, in the face of vast and ever-growing quantities of documents, leads archivists to propose sampling for series where the value of records can vary quite dramatically. Court records—most particularly, case files—are one example of this situation. It is not feasible to pay the cost of appraisal at the level of either the file or its component items; sampling, however, will also undoubtedly result in the loss of some extremely interesting historical documents.

How have such appraisal decisions been justified? Let us look a bit more closely at the arguments put forth with regard to case files in the case of Quebec's Interministerial Committee on Court Records.<sup>1</sup>

The decision to sample case files was made in the context of exponentially increasing quantities of case files—a reality visible in courts around the world. The wealth of information on daily life that case files offer prior to the twentieth century, when there is a relative lack of sources for social history, easily justify permanent retention of all case files—in spite of their uneven quality—for earlier periods. In Quebec's case, it was the volume of twentieth-century files, and particularly the rapid growth from 1920 on, that led to a consensus in the Committee that there would be no choice but to sample after some specified cut-off date.<sup>2</sup>

<b>Table 1</b>	
<b>CASE FILES PRODUCED BY QUEBEC COURTS</b>	
<b>Period in which files produced</b>	<b>Quantity of files in linear metres</b>
From the beginnings to 1920	6 264
From 1921 to 1950	10 463
From 1951 to 1966	15 478
From 1967 to 1982	41 661
<b>TOTAL</b>	<b>73 866</b>

SOURCE: **Table 1** (p.8) of the *Report of the Interministerial Committee on Court Records*.

The Committee eventually decided to sample files produced on or after 1 January 1920. This decision limited the loss of information involved in sampling to a time period when alternative sources were available and becoming more and more abundant.<sup>3</sup>

In addition, other court records contain some of the information to be found in case files, notably the various docket books, judgment books, indexes, and so on produced by the court registries. The judges, lawyers, notaries, and jurists on the Committee<sup>4</sup> were in fact in agreement that the essential memory of the courts, necessary to live up to their obligation as "courts of record," is contained not in the files, but in the docket books and judgment books. Hence a recommendation for permanent preservation of these records, along with their corresponding indexes. It must be noted that the Committee felt that sampling was allowable only when these other key documentary series were in existence. Case files will thus be kept in their entirety for any years in which the appropriate registers are missing. Thus we are assured that there will always remain some information about every court action, even after sampling.

It cannot be denied, however, that the information in docket and judgment books is sometimes a mere skeleton of data and that, especially in the case of the lower courts, it can even be impossible to determine the grounds of an action on the sole basis of the docket or judgment book. Before opting for permanent retention of files because of the insufficiency of detail in some registry books, one must consider the degree to which average files contain significant *additional* information. In the study carried out by Michael Hindus *et al.* on the case files of the Massachusetts Supreme Court of Judicature, detailed analysis of files at ten year intervals made clear that a small minority of files actually contain significant information that is not already provided by registry records.<sup>5</sup> In other words, while some files are veritable treasure troves of data, the informational value of the "average" case file is not very high. This situation is compounded by the fact that there is an enormous amount of repetition in the case load of the courts, in terms of the types of cases encountered. This is increasingly true in the twentieth century, where a marked increase in recourse to the courts stems in part from the growth of government regulation.

Unfortunately, as the Massachusetts study indicated and the work of the support teams of Quebec's Interministerial Committee confirmed, there is no ready way to separate out the "interesting" files from those of little value, unless the courts themselves have kept them in distinct categories. In fact, one would need to open and read each file in order to apply any qualitative criteria. The mass of documentation prevents qualitative appraisal decisions, unless a feasible way of identifying valuable files can be found.<sup>6</sup>

Another method has been followed in Quebec, where the files of all cases that climb the ladder all the way to the provincial Court of Appeals or on to the Supreme Court are being retained, not only because of their value for jurisprudence but also because they contain transcripts of testimony, in contrast to the vast majority of files for cases that have not been appealed. These records could be readily identified by using the docket books and files of the Court of Appeals

(although the operation of creating a complete and verified list of the files to select, involved over a year of effort in over seventy locations throughout the province).<sup>7</sup>

Ironically, the selection of files going on to appeal—a selection decision based on the role of appeals in the creation of jurisprudence—results in the retention of a considerable proportion of “fat” files. During the period when files were folded (the pre-file folder era), very large files, which could not be successfully folded because of their size, were often stored separately, between cardboard covers secured with ribbons. In the selection process carried out at the Court services’ intermediate record centre in Montreal, about fifty per cent of these files turned out to be appeal files. Lest those who favour the fat file method interpret this as proof of the validity of such a standard, I might add that the retention of appeal files is a compromise solution. All such files are retained because some of them might document precedents. As long as an appeal court accepts appeals automatically upon demand for cases involving amounts over a fixed level (which is the general rule for appeal courts, and was the situation for civil suits even for the Supreme Court of Canada, up until 1974), rather than because an important point of law is in question, one might well argue that the files of reported cases would have a greater chance of being genuinely “interesting.” Not everyone, however, has the greatest confidence in the uniform quality of Canada’s system of court reporting. Besides, drawing up lists of reported cases and identifying them sufficiently to locate them and pull them out would be even more time-consuming than using the criterion of cases going on to appeals.

The mind boggles to contemplate the human resources and time necessary to appraise individual files amongst the ninety kilometres of such records accumulated in Quebec between 1920 and 1992.

In cases where files have been grouped by category, obviously the informational value of the whole category can be used in making appraisal decisions for each distinct grouping of files. In Quebec, relatively few such distinct groupings—or jurisdictions, as the registries call them—existed prior to the 1970s. Most older files fell into huge grab-bag groupings such as “general criminal matters” or “general civil matters,” which offer no help for selection decisions. Only a very few of these jurisdictions have been used for appraisal decisions in Quebec—notably for permanent retention, as in the case of adoption, guardianship, and other non-contentious matters. Generally, a conscious decision was made to treat the files within the majority of jurisdictions on an equal footing; that decision is linked to the positive side of using sampling—for there is indeed a positive side.

Sampling eliminates two of the possible pitfalls of “qualitative” appraisal. First, a statistical sample provides a faithful microcosm of the original population, so that the information necessary to study the overall patterns of litigation and the functioning of the administration of justice is in no danger of disappearing. In fact, the inescapable reality that a major part of the courts’ time and effort is spent on fairly routine cases can scarcely be documented better than by a statistical sample: a qualitative approach could well serve to distort future understanding of the functioning of the administration of justice, by emphasizing the caviar and champagne at the expense of the bread-and-butter.

Second, sampling allows archivists to avoid the dangers of irreparable loss in eliminating whole series currently judged to be of lesser value (such as the files of inferior courts). In Quebec, the perils of subjectivity were felt to be too great to authorize the complete elimination of any but the most ephemeral and administrative series (such as court rolls): all the others would be sampled or, in a few special cases, preserved in their entirety. The Committee's decision rested on the conviction that future research could best be served by preserving a faithful documentary record of the activity of the entire system—rather than sacrificing the records of some courts in their entirety, in order to preserve more records from courts whose work currently seems to be of more interest to the historical community.

Finally, the decision to sample was facilitated by the confidence that the Committee had in the quality of the sampling method. Care was taken to ensure that the resulting sample would take into account regional and diachronic variations in court activities. The use of a variable sampling grid—where the percentage of files retained varied according to production—combined with the decision to sample every jurisdiction in every court registry separately and annually, meant that an accuracy level of ninety per cent and a confidence level of ninety-five per cent could be expected.<sup>8</sup>

In summary, before plunging into a discussion of the problems of application, the recourse to statistical sampling offers the assurance of a documentary heritage that accurately reflects the character of the original population of files, while avoiding the dangers of subjectivity and the difficulty of agreeing upon selection criteria. The risk of the loss of valuable individual files can be considered acceptable if it is counter-balanced by the retention of complementary record series containing overlapping information, the preservation of series or sub-series of particularly and consistently high value, limiting sampling to time periods when alternative sources are available, and using a rigorous sampling method with a high level of statistical precision.

### ***Problems of Application of a Given Sampling Method***

The Interministerial Committee's sub-committee on sampling developed a method of systematic sampling based on the case file numbers. Case files are particularly well suited to systematic sampling, because case numbers are generally assigned in simple chronological order, on a first-come first-served basis. The sampling grid provides for the retention of seven different percentages, varying from five to one hundred per cent, according to the number of files produced.<sup>9</sup> In the planning process that followed the official adoption of the Committee's report and the adjoined retention schedules, a manual of procedures was produced to ensure a uniform application of the method. This manual established the specific terminal numbers to be pulled according to the percentage retained, and required any anomalies to be reported to a provincial coordinator, to prevent unacceptable deviations from the sampling method when local work teams encounter unexpected problems.

As might be expected, sampling is easy and rapid in those cases where the files being sampled come from a registry that kept its documents in an orderly manner. The lack of order or, in some cases, the peculiar in-house systems used make the process laborious and slow in other instances. In yet other instances, the sampling

method cannot be applied, or may give end results rather different from those predicted by statistical theory. Let us look at a few of the problems encountered, to see the full range of problems, from annoying through to prohibitive.

A fairly simple problem exists when, through neglect, case files are found in a state of total disorder—a situation that the court services encountered in Sorel for the 1930s and 1940s. The files had to be sorted in numerical order before sampling could be carried out. While costly in terms of human resources, this imposed no threat to the sampling method *per se*.

Peculiar in-house classification or filing systems pose somewhat greater problems. In Quebec—and, I suspect, in other provinces as well—registry offices had a high degree of local autonomy until some years after World War II.<sup>10</sup> Although the laws on the administration of justice, the rules of practice, and the codes of procedure require the retention of certain kinds of records, containing certain types of information, they generally say nothing about *how* this is to be achieved. Specific internal directives on records management that would compensate for this lack of legislative precision are quite recent. As a result, processing older court records from a variety of registries presents some interesting challenges.

For example, in Montreal a very peculiar numbering and filing system was in use from 1926 through to the mid-1940s for the files of the Superior Court. The yearly docket books were organized in sets of six huge, fat tomes each holding one thousand numbers: volume A ran from 60,001 to 61,000; volume B from 61,001 to 62,000; volume C from 62,001 to 63,000; and so on. The first case of the year received the first number in volume A, the second case that of the first number in volume B, and so on, with the seventh case being the second number in Volume A, the eighth being the second number in B, and so on. When all the numbers in a six-volume set were assigned, six new volumes, with the six succeeding “thousands,” were begun.

The sampling method posits consecutive numbers, assigned on a purely chronological basis. In this case, a mathematical formula had to be used to generate a list of the numbers that were the equivalents of those that would have been chosen for the sample if a consecutive numbering system had been used. Moreover, the staff doing the sampling had to keep on hand a table of equivalents to help them determine the “following” case file, whenever the target number was missing (see **Table 2**).

The actual process of pulling the sample files was further complicated by the filing system used. Case files were physically grouped in ribbon-tied packets according to the last four digits of their file number—up to 4,999. When the last four digits were 5,000 or more, one has to subtract 5,000 in order to get the number of the packet. What does this mean in fact? #2,174, #22,174, #32,174 and so on, are cheek-by-jowl with #7,174, #27,174, #37,174, etc.. Needless to say, this kind of sampling requires personnel that are fairly intelligent, meticulous, and motivated. Nonetheless, we are still operating in the realm of problems that, while they may complicate the sampling process, do not undermine its theoretical foundations.

<b>Table 2</b>					
<b>REPLACEMENT GRID</b>					
Volume A	Volume B	Volume C	Volume D	Volume E	Volume F
000	001	002	003	004	005
006	007	008	009	010	011
012	013	014	015	016	017
018	019	020	021	022	023
024	025	026	027	028	029
030	031	032	033	034	035
036	037	038	039	040	041
042	043	044	045	046	047
048	049	050	051	052	053
054	055	056	057	058	059
060	061	062	063	064	065
066	067	068	069	070	071
072	073	074	075	076	077
078	079	080	081	082	083
084	085	086	087	088	089
090	091	092	093	094	095
096	097	098	099	100	101

*The replacement for a missing case file has the same terminal numbers but will begin with the “thousand” in the column to the right, except when the missing file is in the thousands found in the extreme right-hand column (Volume F), in which case you add “1” to the terminal digits and use the thousand in the extreme left-hand column of the same row (Volume A). Example: the replacement for 69,230 is 70,230 but the replacement for 71,230 is 66,231.*

Let us look briefly at three more serious challenges to sampling theory: alphabetical filing, interlocking jurisdictions, and the impact of pardons.

Alphabetical filing provides the advantage of rapid access, without the need for an index. It is particularly useful when dealing directly with clients, who will likely not be aware of the number of their file. This may explain why alphabetical filing was the norm throughout the province for the case files for voluntary judicial deposit—a procedure used for the gradual repayment of debt as a way for creditor and debtor to avoid seizure or insolvency proceedings. There were file numbers—or rather alphanumerical codes: over a period of some sixty years, files were numbered consecutively under the letter of the plaintiff’s family name. However, the

docket books were in binder form, a separate leaf being used for each entry; the leaves were organized within the binder in pure alphabetical order, and not by the alphanumeric code, doubtless in order to avoid having to create an index. Without reorganizing the dockets by file number, there is no way of determining the annual number of files, nor the approximate chronological sequence of their creation, short of transcribing the basic entry for each case in the docket books into a computer file sorted alphanumerically and by date of file opening. There is, of course, no way that we can determine the exact chronological sequence of entry for files opened on the same day but under different letters. An arbitrary order would have to be established to allow the production of a list for sampling; that list could then be re-sorted in alphabetical order to facilitate the pulling of the sample files.

On the one hand, the sequential order that is the very basis of the sampling method cannot be guaranteed. On the other, the cost of human resources needed for such a process may well be too high in relation to the economic benefits of sampling. While one can justify the investment in transcribing files that went on to the Court of Appeals, taking into account their higher informational content, it is much less evident that voluntary deposit files merit such special treatment.

What options remain? One could decide, as in the case of missing docket books, to preserve the entire series. This is the preferred option of the justice department, as it can then transfer the records to the Archives nationales with a minimum of effort. Alternatively, since the case files in this particular series have very little information that is not contained in the docket books, one could develop a less statistically rigorous sampling method specifically for this series. An estimate of the overall number of files, with an average annual production, could be used to determine a percentage to retain. Provided that the number of files produced annually was fairly high, systematic pulling of files at fixed intervals for each of the letters of the alphabet, would then be feasible (every fifth or tenth or twentieth—for 20, 10, or 5 per cent retention—for example). Clearly, however, the result would not attain the same levels of precision and confidence as to representativity as the original sampling method. No decision has yet been made concerning this series. However, as there are over 230 linear metres of files in question for the District of Montreal alone, more investigation will undoubtedly follow when sampling has been completed for the simpler series.

Interlocking jurisdictions provide another problem. How does one accommodate a file that starts before one court, continues before another, and ends in yet a third, all at the trial court level? This is the reality of many criminal and penal case files—where the accused's option to choose to be judged by a magistrate alone or by a judge and jury mix with the peculiar role of the justice of the peace to create an archivist's nightmare. In today's era of computerized docket books, a single number is assigned for each criminal case, regardless of its trajectory across the various criminal courts. In the past, unfortunately, this was not the case.

All criminal and penal cases began, and still do begin, before a justice of the peace—that heritage of the English legal tradition. Clerks of the peace enter each such case in a docket book. Many cases end right there, as the accused confesses and is sentenced on the spot for many summary types of offenses. Many more go on to the Sessions of the Peace (formerly Quarter Sessions), where a specific range of crimes and statutory offenses must be heard, in accordance with the criminal



code and provincial statutes and regulations. The Court of the Sessions of the Peace sometimes had its own docket books, separate from those of the J.P.'s; sometimes, however, in some judicial districts, the clerk of the peace kept the records for both these judicial instances together in the same book. When the docket books were distinct, so was the numbering of the case files. A smaller group of criminal cases went on, either through the choice of the accused or obligatorily because of the nature of the offence, to the criminal assizes, or the Court of King's Bench, Crown side, as it was known in Quebec right up until the mid-1970s. Here the clerk of the Crown entered the case in yet another distinct docket book, with yet another distinct file number.

This is not too hard to deal with, in terms of the sampling method, if a separate physical file was kept at each level—provided one accepts that each file contains only a portion of the overall trial documents. To the extent that each jurisdiction referred to above has exclusive competencies, sampling each separately is important in order to ensure that the sample will adequately reflect the different kinds of cases according to the trial court. Researchers would undoubtedly prefer us to keep all the documents generated by the same case, regardless of how many and which courts dealt with it. Doing so, however, would not only complicate sampling and go against the concept of original order, it would also result in over-sampling of all three levels and risk distorting the sample base in statistical extrapolations. This distortion would arise at both the level of the volume of each jurisdiction, and that of annual populations, for the file concerning a specific offense was often opened in different years at different levels.

In some court registries, files were indeed kept physically apart. In others, two or three separate series were kept; in those cases that moved from one level to another, the documents followed the accused and were integrated into the file of the final trial court. In yet other registries, local staff decided after the fact to reduce their efforts at file retrieval by fusing the previous series, so that a single physical file now exists, with two or three distinct case file numbers on it. In this situation, the whole series is kept under the number assigned by the clerk at the first trial level.

The provincial coordinators of our project had to decide whether to adjust the method to the existing situation in each registry or archives centre, or to maintain a general method and require personnel to invest the time to compensate for local deviations in classification and filing. The decision has been to do a bit of both. On the one hand, all registries must respect the principle of provenance, as embodied in the existence of distinct docket books with distinct file numbers. Wherever the original docket books exist, the files will be sampled on the basis of those books. Where staff later reorganized files, we shall have to find a way to track down the targeted files of each of the trial courts that had been integrated into a file identified by the number and the year of another judicial instance. This is no small task, involving considerable research and cross-referencing in the docket books of the series involved. In the intermediate record centre for court records in Montreal, for example, this decision has tripled or quadrupled the time required to sample the criminal and statutory files from 1920 to 1960.

On the other hand, there will be no effort to reconstitute the original separate series by sorting the contents of the files and keeping only the documents produced

by the trial court for which the sample is being taken. While this will cause a risk of distortion of the statistical base of the sample, the task would be too labour-intensive and costly to carry out. This decision will have the added benefit, from the researcher's point of view, of providing a "complete" file for each case sampled. The risk of distortion of the sample's statistical base will be countered by warnings to the researcher that only the files identified as parts of the sample of a given court should be used in any quantitative analysis.

My final example of problems in the application of the sampling method is the impact of pardons. As you may know, when a pardon is granted, the case file must be removed from its series and sent to the registry where the trial took place, there to be disposed of by the appropriate authorities.<sup>11</sup> A definite but as yet unmeasured potential for distortion exists here. Given the realities of aging, most requests for pardons touch files produced in the last fifty years, and most especially, in the last thirty years, during the case file's semi-active stage. Some, however, do occur after the files have become inactive and have been sampled.

There are two major dangers to the sampling method here. First, we have no assurance that pardons will be randomly distributed across the full range of types of offenses and types of accused individuals. If pardons are not randomly distributed, the population of files remaining at the end of the semi-active stage will be skewed before sampling can be carried out. Ideally, efforts should be made to determine the nature of the distortion; at the very least, researchers must be warned of the extent to which target files have had to be replaced because the individuals involved were pardoned. The spirit of the *Criminal Records Act*<sup>12</sup> requires the guardians of court records to eliminate every trace of the existence of the record of a pardoned individual and to pretend, if any inquiry for the record is made, that it never existed. This means that archivists can only inform researchers that a given percentage of files in the original population were missing because of pardons (without providing any information as to which files fell into this category).<sup>13</sup> Fortunately, current information from the Judicial Services indicates that the percentage of files removed for pardons is not large.<sup>14</sup>

The second danger comes from the loss of files after sampling. If a pardon is granted to an individual whose file was not retained in the sample anyway, no harm will be done to the representativity or level of accuracy of the sample. However, whenever sampled files must be removed because of a pardon, there will be no way to replace them, and the sample's precision—and possibly its representativity—will diminish over time, to a degree that we cannot determine beforehand. It will be necessary to keep a record of how many files are removed from the sample in this way, in order to allow researchers to take the erosion of the sample into account.

Unfortunately, given that the sampling method has been published and will be explained in finding aids to court records, it will not take a genius to realize that missing criminal files that have not been replaced must pertain to cases involving pardons. In other words, lists of the file numbers in criminal file samples should not be made accessible to the public, but should be used only by reference room staff to determine whether a requested file is still preserved by the archives. It is not immediately evident how to ensure that researchers using entire samples of

criminal series for quantitative research do not compile their own lists of file numbers and make deductions as to which missing files might be pardons. A restriction on publication by such researchers would appear to be necessary.

The gradual erosion of the annual samples will presumably halt with the deaths of the individuals involved. Provided the numbers of pardons granted after sampling are not too high, and that researchers are informed of the number of files removed from any given sample, the representativity of the samples should not be seriously compromised. A more radical attack on documentary heritage of the criminal courts, coming from a Charter of Rights challenge concerning acquittals, may well diminish the relative impact of pardons. If the courts decide that individuals acquitted in criminal cases should benefit from the same treatment as convicted but pardoned criminals, the entire sampling method will become unworkable; any sample or selection eventually made could only represent the population of cases in which convictions were obtained and no pardons were granted.<sup>15</sup>

### **Conclusion**

The road to perfect samples of case files is not always straight and well-paved. Sampling aims at reducing the volume of records retained in a systematic way, guaranteeing a representative documentary heritage while simultaneously reducing the labour involved in selection. The variations in the condition of records—their classification, filing, inter-relations, and their being subject to external rules for elimination, foreign to the logic of the sampling process—all may have a considerable effect on the end results. Rigorous and consistent application of a given sampling method, where files need extensive sorting or re-organizing for the method to be effective, may actually prove to be more expensive than permanent retention for certain record series. Archivists and administrators need to be aware of this, and must think carefully about when and how much they are prepared to yield in the constant tension between the requirements of representativity and precision, and the cost of the human and material resources needed to attain them.

### **Notes**

\* Based on a paper presented at the annual conference of the Association of Canadian Archivists, St. John's, Newfoundland, 23 July 1993.

1 For a more detailed explanation, see Québec, Ministère des affaires culturelles, Ministère de la Justice, *Report of the Interministerial Committee on Court Records* (1991) [hereafter referred to as the *Report*] (or the longer, three volume French-language original, *Rapport du comité interministériel sur les archives judiciaires* (1989), which also includes, in Annex 1, the detailed retention schedule produced by the committee). An entire issue of the journal *Archives* 22, no. 4 (1991), was also devoted to the presentation and analysis of the *Report* as well as to the reactions of historians and to reports on the approach to court records adopted elsewhere in Canada.

As might be expected, many historians find the very notion of sampling unacceptable, and either want to move the cut-off date forward closer to the present, or at the least save those types of files that are of most interest to their own research. Others, however, such as J.-C. Robert (who sat on the Interministerial Committee and was put in charge of its sub-committee on sampling), urge their colleagues to admit that exhaustive research techniques are neither always necessary nor invariably appropriate in the context of the twentieth century's information explosion. As one might expect, those historians who favour quantitative methods are more open to sampling, and those in

- need of information on specific cases are opposed. In a series of public information meetings held after the *Report* of the CIAJ was submitted, most members of the public—generally historians and genealogists—who were present were willing to accept sampling, once they were made aware of the quantities of court records involved.
- 2 The problem posed by the volume of files was quite concrete and immediate: the Court services, under pressure through lack of space in the court houses and intermediate record centres, was finally willing by the beginning of the 1980s to transfer older records (prior to the mid-1950s) to the ANQ; in Montreal this would have meant the transfer of over ten kilometres of files and registers and the Montreal centre simply did not have enough space.
  - 3 For the historical arguments put forth as to the validity of the 1920 cut-off date, see the *Report*, pp. 36-40.
  - 4 The committee included two judges (one for the Superior Court—Mr. Hon. Justice Roland Durand—and one for the unified Court of Quebec—Mr. Justice Raymond Boily, replaced after his death during the committee's mandate by Mr. Justice Raymond Boucher)—named by their respective chief justices, a representative of the bar (Me Pierre Gauthier, the then Director General of the Quebec Bar Association), a representative of the notariat (Me René Marcoux, named by the Chamber of Notaries), one legal scholar (Jacques Boucher, Dean of the Faculty of Graduate Studies at the Université de Montréal, a law professor not unfamiliar with the writing of legal history, who was the president of the committee), one professional historian (Jean-Claude Robert, of the Department of History of the Université du Québec à Montréal, a noted nineteenth-century Quebec historian and previous president of the Canadian Historical Association, named by the Institut d'histoire de l'amérique française), one genealogist and amateur historian (Jacques Lemieux, a teacher at Collège Mérici), one representative of the tax-paying public (Lucien Saulnier, an important figure in the Montreal urban community municipal scene), one representative of the justice department's Court services (Me André-Gaetan Corneau), and one representative of the Archives nationales du Québec (Jacques Grimard, Deputy Keeper of the ANQ; on his departure from the ANQ, Mr. Grimard was officially replaced on the committee by Normand Gouger, Deputy Keeper of the ANQ for Montreal, but Mr. Grimard remained with the committee as a resource person for the rest of its deliberations). The secretary of the committee was Jacques Ducharme, an archivist and former Deputy Keeper of the ANQ, who had been one of the major proponents of a broad-based interministerial committee, and who wrote the committee's report just before his tragic death from cancer in 1989.
  - 5 Between six and eight per cent. See Michael Stephen Hindus, T.M. Hammett, and B.M. Hobson, *The Files of the Massachusetts Superior Court, 1859-1959: An Analysis and a Plan for Action* (Boston, 1979), p. 74.
  - 6 One method suggested by the Massachusetts study, and fairly commonly used for other kinds of files as well, is the "fat" file concept. The Interministerial Committee considered and explicitly rejected this method, jurists and historians being unanimous in their rejection of the equation between historical or legal "interest" and file thickness. In fact, this rather standard approach appears to me to be lacking a solid scientific base, and reposes on the rather fragile tautological foundation that whatever is more abundantly documented is, because of its more abundant documentation, of greater historical interest. Anyone who has had to wade through the routine procedural documents and the mountains of marginally literate verbiage produced by zealous lawyers with wealthy clients willing to pay for the privilege of dragging out the final reckoning will be aware of the fact that many fat files are appallingly uninteresting. Conversely, historians will tell you that a very slim file may often be of great interest. The fat file method would be more acceptable if it were based on studies clearly showing the link between the width of the file and the quality of the information it contains. To my knowledge, however, this has not yet been done.
  - 7 Readers may be interested to learn that for the Superior Court files, transfers of files after selection and sampling have revealed that the files that went on to the Court of Appeals represent one and a half times the volume of the sample files.
  - 8 For the general outline of the sampling method, see the *Report*, Appendix, Part Two, pp. 99-109. For a somewhat more detailed discussion, see Michel De Grand'maison, "La méthode d'échantillonnage," *Archives* 22, no. 4 (1991), pp. 23-48.
  - 9 1-60: 100 per cent; 61-100: 60 per cent; 101-200: 50 per cent; 201-300: 30 per cent; 301-800: 20 per cent; 801-2,000: 10 per cent; 2,001- : 5 per cent.
  - 10 On the evolution of the role of court registry clerks in Quebec, see Pierre-E. Audet, *Les officiers de*

*justice des origines de la colonie à nos jours* (Montreal, 1986).

- 11 This is a touchy and confidential subject which court authorities are reluctant to discuss because of the tension between the obligation to confidentiality, which the public often equates with the physical destruction of the records, and the need to be able to resuscitate the files in the case of a pardoned individual committing another crime and thereby invalidating his/her pardon.
- 12 *Criminal Records Act*, R.S.C. (1985), c. C-47.
- 13 Another result of the spirit of the *Criminal Records Act* is to make court authorities very reluctant to allow any access to the files of pardoned individuals, even for the purpose of analyzing the range of types of offenses and individuals involved. In fact, so strong is the instinct for confidentiality, that most registries do not even keep a list of the files locked away after pardon, as such a list in and of itself would present a risk of leaking confidential information.
- 14 Apparently, between one and five per cent of criminal files have been removed as a result of pardons, according to unpublished information from the Judicial Services branch of the Justice Department.
- 15 Indeed, this would appear to be the general idea behind the Supreme Court ruling in the McIntyre affair in 1982; however, few concrete steps appear to have been taken as yet, at least in Quebec, to actually segregate and render inaccessible the records of those who have been acquitted.