# Privacy Interests in Prescription Data, Part 2

## Patient Privacy

KHALED EL EMAM
*Children's Hospital of Eastern Ontario, University of Ottawa*

PATRICIA KOSSEIM
*Genome Canada*

For several years, concern has been mounting about the privacy implications of using and disclosing prescription data. Several transactions forming part of a growing chain of valuable health information have come under scrutiny, including

- the sale or transfer of prescription data from pharmacies to commercial data brokers;
- processing of the data to analyze physicians' prescribing patterns; and
- the subsequent sale of these prescribing patterns to pharmaceutical companies, among others, that use this information to customize their marketing strategies aimed at physicians.

In part one of this two-part series, we discussed privacy concerns with respect to prescribers. In this second installment, we examine the privacy risks to patients from Canadian and US perspectives.

### Identifiability

In privacy terms, we generally understand personal information to be *identifiable* information *about* an individual. Although prescription data in relation to a patient is clearly information about him or her, the privacy issue that arises is whether prescription records contain fields that—taken together or in combination with other publicly available data—could identify the individual patient concerned. If so, pharmacies would have a clear privacy obligation to obtain prior consent from patients before disclosing any prescription data about them to commercial data brokers.

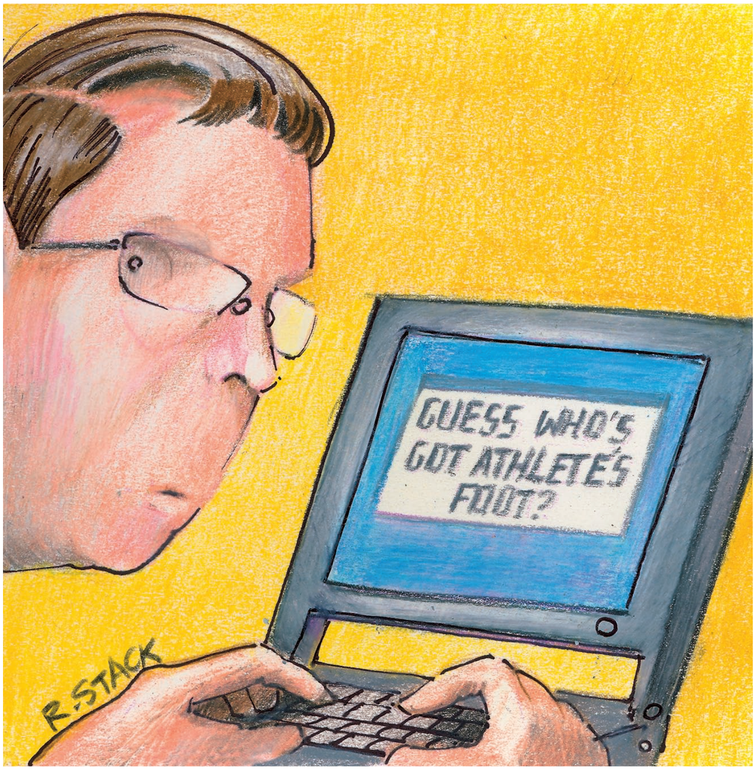
Prescription data disclosed by pharmacies to commercial data brokers don't typically contain any directly identifiable information about the patient involved, but do contain fields such as the patient's age and gender. Sometimes, pharmacies also disclose, directly or inferentially, geographic information about the patient's residence, such as the first three characters of the Canadian postal code (otherwise known as the *forward sortation area*, or FSA). The privacy question here is whether these three fields, limited as they are, can still be used, either alone or in combination with other available information, to re-identify individual patients, thereby jeopardizing their right to confidentiality.

Privacy laws generally apply only to identifiable information. These laws might either explicitly or implicitly provide some kind of threshold for determining when information ceases to be de-identified and becomes identifiable. Where these do exist, legal thresholds for identifiability in Canada and the US differ, both in substance and approach.

### Canadian Approach

Although most, if not all, Canadian privacy laws require personal information to be identifiable,[1] few explicitly define what "identifiable information" actually means (see Table 1). Statutory tests for identifiability range from lower thresholds of what is "reasonably foreseeable" or can be "reasonably expected" to identify individuals, to higher thresholds of what is "readily ascertainable" or "obvious." Unlike some tests we see emerging in other jurisdictions, Canadian thresholds don't require application from any particular perspective (for example, a highly sophisticated expert, a motivated intruder, or an average layperson). Nor do Canadian thresholds expressly reference the level of resources, time, or effort necessary to re-identify individuals.

The Federal Court of Canada recently interpreted the meaning of "identifiable" in *Gordon v. Canada (Health)*.[2] This case involved an access-to-information request for the "province" field of the Canadian Adverse Drug Reaction Information System, housed at Health Canada. The issue was whether the province field, when combined with all the other data elements already publicly released, could possibly identify affected patients. The court adopted the



test put forward by the Privacy Commissioner of Canada (who was intervenor in this case), holding that "information will be about an identifiable individual where there is a serious possibility that an individual could be identified through the use of that information, alone or in combination with other available information."[2] In Canadian common law, the concept of "serious possibility" means something more than a frivolous chance and less than a balance of probabilities. Based on substantive evidence adduced, the court found that disclosing the province field—particularly in cases of unique or quasi-unique reports in smaller provinces—would substantially increase the possibility that a patient could be identified. Hence, the court found that the province field did constitute identifiable personal information in this case and upheld Health Canada's refusal to release the requested information.

The practical difficulty with these qualitative approaches to defining identifiability is that a typical privacy officer working in, say, a hospital, wouldn't necessarily be able to assess, on his or her own, whether a particular prescription record contained "identifiable" information without the assistance of a sophisticated, external expert. Such expert advice would be critical to the privacy officer's assessment about whether disclosure is permissible, absent individual consent.

## US Approach

In contrast with the qualitative tests used in Canada, the US Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule draws a clearer line between de-identified and identifiable data (see www.hhs.gov/ocr/privacy/index.html). Under HIPAA, data are deemed to be de-identified if they do not contain any of the 18 data elements expressly specified in the law. The US National Committee on Vital and Health Statistics (NCVHS) has estimated that applying this objective HIPAA standard results in a 0.04 percent chance of re-identification in the US,[3] but other evidence also suggests that it results in significant information loss.[4]

Another de-identification method set out in HIPAA refers to a less deterministic, user-type test. HIPAA considers data to be de-identified if an individual with knowledge and expertise in de-identification techniques (such as an expert statistician) can certify that only a "very small risk" of re-identification exists. In contrast to HIPAA's objective list of 18 identifying elements, this alternative method of determining identifiability leaves relatively more "gray" room, and some have expressed concern about the potential liability of these experts should an intruder later succeed in re-identifying the data.[5] Consequently, the NCVHS has called for more concrete guidance on thresholds for statistical de-identification.[3]

A recent freedom of information case in the US, somewhat equivalent to the Gordon case in Canada, is also instructive in helping to understand the concept of identifiability. *Southern Illinoisan vs. The Illinois Department of Public Health*[6] involved a request for information from the State Cancer Registry by cancer type, diagnosis date, and patients' ZIP codes. The issue here was whether the requested information should be exempt from disclosure because it "tends to lead to the identity of any person." Despite expert testimony on behalf of the department demonstrating that it was empirically possible to re-identify affected patients, the Illinois Supreme Court held that the "tends to lead" test must be applied from the general public's perspective rather than that of an information technology expert with the requisite knowledge, skills, and expertise. By this interpretation, a data custodian must show that nonexperts could achieve the same re-identification outcome before courts will

Table 1. How some Canadian provincial privacy laws define identifiable information.

| PRIVACY LAW | DEFINITION |
|---|---|
| Personal Health Information Protection Act (Ontario) | "Identifying information" means information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify an individual. |
| Protection of Personal Health Information (Newfoundland & Labrador) | "Identifying information" means information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized either alone or together with other information to identify an individual. |
| Health Information Protection Act (Saskatchewan) | "De-identified personal health information" means personal health information from which any information that may reasonably be expected to identify an individual has been removed. |
| Health Information Act, (Alberta) | "Individually identifying" means that the identity of the individual who is the subject of the information can be readily ascertained from the information; "nonidentifying" means that the identity of the individual who is the subject of the information cannot be readily ascertained from the information. |
| Protection of Personal Information Act (New Brunswick) | "Identifiable individual" means an individual can be identified by the contents of the information because the information includes the individual's name, makes the individual's identity obvious, or is likely in the circumstances to be combined with other information that includes the individual's name or makes the individual's identity obvious. |

consider a data set to constitute personal information.

## Patient Identifiability in Prescription Records

Few empirical studies exist on the risks of patient re-identification from prescription records. One way to evaluate patient identifiability is to examine the extent to which certain fields included in a prescription record make individuals unique in the population. The basic premise is that the higher the degree of population uniqueness, the easier it is to re-identify individuals.

Using this method in a recent study, one of the authors (Khaled El Emam) reviewed prescription records disclosed by retail pharmacies in Canada. These contain the patient's age, gender, and, in some cases, the FSA (either directly or inferentially). El Emam and his colleagues analyzed Canadian census data and concluded that, except for New Brunswick, the vast majority of the population wouldn't be unique on age, gender, and FSA,[7] making it very difficult for any member of the general public to re-identify most Canadians using just those three variables.

However, a further question is whether the re-identification risk becomes more elevated when examined from the perspective of a determined intruder who is highly motivated to find a specific patient. As part of an ethics-board-approved research project on de-identification techniques,[8] El Emam and his colleagues went on to examine six months of pharmacy data from the Children's Hospital of Eastern Ontario and created the same record layout used by one of Canada's commercial data brokers. An important difference between retail and hospital pharmacy data is that the latter includes admission and discharge dates, as well as diagnosis information. According to the data broker's literature at the time, 100 hospital pharmacies across the country already disclose this same data. An examination of these pharmacy data indicated that out of 3,510 patient visits during which the hospital pharmacy dispensed a prescription, 99.6 percent were unique on age, gender, FSA, admission date, and discharge date. Hence, a patient's nosy neighbor who is determined to find out information about the patient's health status, for example, would have sufficient background information to identify the unique prescription record and determine the drugs prescribed to the patient as well as the patient's diagnosis.

When El Emam removed admission and discharge dates from the prescription record, only 52 percent of the records were unique. Although this is a significant reduction, it still represents a significant percentage of the records in the data set and poses a serious possibility that a motivated intruder with basic background information about an individual could re-identify a specific patient's record.

To demonstrate this risk scenario more concretely, El Emam and his colleagues, as part of the same investigation, asked colleagues who weren't working at the hospital if they knew any patient who was admitted during the six-month period under examination. One colleague knew that his neighbor's child was admitted with a serious infection in January 200X and also knew the child's gender, age range, and FSA. A database search using these criteria produced a single record that matched. Subsequent verification confirmed the patient's identity. The remainder of the prescription record revealed all drugs taken during the patient's stay and the patient's diagnosis. This is a concrete example of how a neighbor or other motivated individual could use background knowledge about a patient, match it against variables in the prescription record to successfully re-identify the patient, and obtain sensitive health information about him or her.

Ultimately, the level of re-identification risk associated with prescription records will depend on several variables—not least of which is the conceptual threshold of identifiability in accordance with applicable legal requirements, policy guidelines, and best practices. Whether the trigger used to evaluate re-identification risk is reasonableness on a balance of probabilities, a serious possibility, readily ascertainable, or obviousness will clearly affect the outcome of any risk assessment.

Re-identification risk will further vary across retail and hospital pharmacies depending on how much data the pharmacies agree to disclose to commercial data brokers and, more specifically, whether the FSA, admission dates, or discharge dates are disclosed in addition to patient age and gender. In the case of retail pharmacies, which typically disclose fewer data elements, El Emam found the re-identifiability risk to be significantly lower than with hospital pharmacies, which typically release more fields.[7,8]

Finally, the level of re-identification risk will vary significantly depending on the subjective users' viewpoint—that is, whether the user views the risk from a disinterested perspective (as with the general public) or from the perspective of a motivated intruder who has additional background knowledge of the specific patient or patients they're looking for. Although the latter scenario isn't necessarily commonplace, it nonetheless represents a serious possibility of a confidentiality breach and should be factored into custodians' risk assessments, particularly when dealing with highly sensitive data.

We've based these conclusions on what we know about prescription data and clinical practice today. However, with the growing field of pharmacogenomics and its promising potential for personalized medicine and targeted therapeutics, the level of re-identification risk associated with prescription data might have to eventually be reassessed. □

## Acknowledgments

The views expressed in this article are the authors' own and do not represent the official positions of their employer organization.

## References

1. P. Kosseim, A. Kardash, and A. Penta, eds., *Compendium of Canadian Legislation Respecting the Protection of Personal Information in Health Research*, 2nd ed., Public Works and Government Services Canada, 2005; www.cihr-irsc.gc.ca/e/documents/ethics_privacy_compendium_june2005_e.pdf.
2. *Gordon v. Canada (Health)*, *Federal Trial Reports*, vol. 324, 2008, p. 94; www.canlii.org/en/ca/fct/doc/2008/2008fc258/2008fc258.html.
3. "Report to the Secretary of the US Department of Health and Human Services on Enhanced Protections for Uses of Health Data: A Stewardship Framework for 'Secondary Uses' of Electronically Collected and Transmitted Health Data," Nat'l Committee on Vital and Health Statistics, 2007; www.ncvhs.hhs.gov/071221lt.pdf.
4. S. Clause et al., "Conforming to HIPAA Regulations and Compilation of Research Data," *Am. J. Health-System Pharmacy*, vol. 61, no. 10, 2004. pp. 1025–1031.
5. "Statisticians and De-Identifying Protected Health Information for HIPAA," Am. Public Health Assoc. newsletter, 2005; www.apha.org/membergroups/newsletters/sectionnewsletters/statis/fall05/2121.htm.
6. *Southern Illinoisan vs. The Illinois Department of Public Health*, Supreme Court of the State of Illinois, docket no. 98712, 2006; www.state.il.us/court/Opinions/SupremeCourt/2006/February/Opinions/Html/98712.htm.
7. K. El Emam, A. Brown, and P. Abdelmalik, "Evaluating Predictors of Geographic Area Population Size Cutoffs to Manage Re-identification Risk," *J. Am. Medical Informatics Association*, vol. 16, no. 2, 2009, pp. 256–266.
8. K. El Emam et al., "Evaluating Patient Re-Identification Risk from Hospital Prescription Records," to be published in *Canadian J. Hospital Pharmacy*, June 2009.

*__Khaled El Emam__ is a senior scientist at the Children's Hospital of Eastern Ontario Research Institute, and is a Canada research chair and associate professor in the Faculty of Medicine at the University of Ottawa. His research interests include re-identification risk assessment and developing practical de-identification techniques for health information. El Emam has a PhD in electrical and electronic engineering from King's College, University of London. Contact him at kelemam@uottawa.ca; www.ehealthinformation.ca.*

*__Patricia Kosseim__ is Chief GE$^3$LS Officer at Genome Canada, responsible for leading a national strategy to address ethical, economic, environmental, legal, and social issues related to genomics research and its applications. She is on a two-year executive leave from the Office of the Privacy Commissioner of Canada, where she holds the position of General Counsel. Kosseim has degrees in business and laws from McGill University and a master's degree in medical law and ethics from King's College, University of London. She is a member of the Barreau du Québec and the Canadian Bar Association. Contact her at pkosseim@genomecanada.ca.*