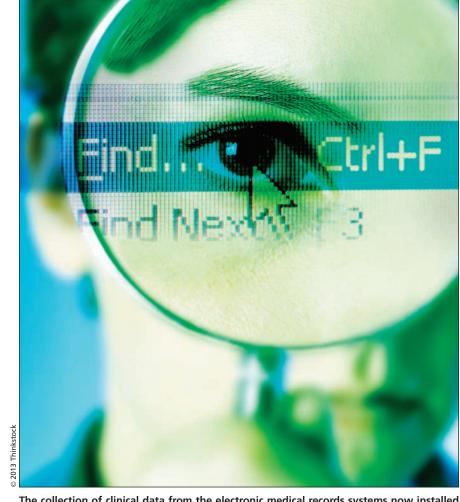# Big data's dirty secret

The era of "big data" is dawning, with the convergence of electronic medical records (EMRs) systems, mobile health care monitors, genetic sequencing and predictive analytics. And with it, hopes are growing that health researchers are poised to use the vast torrents of precise patient data about disease prevalence, quality of care and treatment outcomes to deliver bold new insights that could transform health delivery.

But according to many experts at a conference on e-health in Ottawa, Ontario, May 26–29, technological limitations stemming from mismanagement by government e-health agencies and commercial turf battles are preventing researchers and patients from realizing many of the rewards that the "big data" era could deliver.

"We have not been able to reap research benefits from the hundreds of millions spent on EMRs," lamented Dr. Karim Keshavjee, a Toronto, Ontario, physician and e-health consultant who analyzed EMR data from 450 physicians gathered by the Canadian Primary Care Sentinel Surveillance Network. Keshavjee assessed network data gathered from 10 types of EMRs and found that information on prescribing and diagnostics is high quality, but the data on disease risk factors, such as tobacco usage, "tend to be of lower quality, in part because they are documented less frequently and with less specificity, but also because they are often recorded in noncoded or free text fields" (*Can Fam Physician* 2013;59:108).

"We have not been able to achieve the patient improvements promised by EMR data analysis," Keshavjee argued at the Ottawa conference, e-Health 2013: Accelerating Change, "mostly because the data is dirty. It seems as though there are 500 ways to say diabetes. There's a lot of spurious data, and you have to check it all manually, which is a major effort."

To resolve the issue, Keshavjee and network coauthors in another paper

The collection of clinical data from the electronic medical records systems now installed in an estimated 70% of Canadian physicians' offices has been stymied by a proliferation of different commercial systems that produce incompatible data.

(*Can Fam Physician* 2012;58:1168) suggest that physicians get data training and that EMR vendors be "implored" to create "structured text fields for collecting comprehensive and specific risk factor data." They also called for "standardization across the different EMR systems to ensure consistent data elements."

Data standardization among different EMR systems is a central problem, agrees Patricia Sullivan-Taylor, manager of primary health care information at the Canadian Institute for Health Information. She notes that the collection of clinical data from the EMR systems now installed in an estimated 70% of Canadian physicians' offices has been stymied by a proliferation of different commercial systems that produce incompatible data.

In Ontario, where 11 different vendors are selling systems to clinicians, she warned, "you're going to have 11 different types of data." Sullivan-Taylor added that "it would have been nice" if data collection issues had been resolved before EMRs were installed in clinicians' offices."

Jim Kavanagh, medical director at Telus Physician Solutions, noted that although steps were taken to avert the

problem with EMR data incompatibility, provincial and federal agencies charged with establishing data standards have failed to resolve it.

Alex Mair, director of health system use for Canada Health Infoway, the agency managing $2.1 billion in federal e-health projects, acknowledged progress has been slow on using big data, and says the failure of e-health investments to yield insights for clinicians and patients is a "key gap." Very few examples exist in Canada in which EMR investment is yielding research insights, he says. "We looked, and nothing really stood out."

Mair presented a complex scheme for addressing the problem, all tied to Infoway's complex master plan, which has not been updated to reflect technological developments since 2006. "There has been a lot of discussion around why health care has not adopted better data usage. A lot of it has to do with [health care] managers and providers not understanding analytics."

Martha Burd, executive director of modelling and analysis for the British Columbia Ministry of Health, pointed to her province's successes in using the country's most advanced data collection systems to model diseases as evidence that the promises of the "big data" era are real and achievable.

Burd said the BC Ministry of Health hopes to reform payments to clinicians, hospitals and other health service providers by using data to measure performance and then benchmarking payments to performance measurements.

But she noted that this may deter accurate reporting. "Do you want to report data knowing it's going to have a financial impact?" she asked. — Paul Christopher Webster, Toronto, Ont.

Editor's note: This is the second of a two-part series at cmaj.ca on the e-Health 2013: Accelerating Change conference.