

New report on creating clinical public use microdata files

September 15 2011

The demand for transparency through publicly available healthcare data is on the rise. This is the case for administrative and clinical data for research, and for clinical trials data used to support new drug approvals. Broad data access has a measurable impact on research and policy making. A new report by Dr. Khaled El Emam, the Canada Research Chair in Electronic Health Information at the University of Ottawa and the Children's Hospital of Eastern Ontario Research Institute, looks at the creation of clinical public use microdata files (PUMFs).

"We have shown how to create useful clinical public data files while providing strong protections for the privacy of individuals," explained Dr. El Emam. "The U.S. and the U.K., as well as other countries are steering an international debate right now about open data, and are leading the way in providing access to detailed health information. But there are very few organizations in Canada, all sitting on gold mines of data, which have made that data publicly available. More Canadian agencies need to step up!" The report demonstrates that making the data anonymous is possible and doing so does not put [patient privacy](#) at risk.

A PUMF can serve multiple purposes, including: confirming published results, providing broader feedback to improve data quality, use it for training students and fellows in data analysis, providing an easily accessible data set for researchers to design studies, examine the feasibility of certain studies, and prepare for analyses on more detailed data that are not publicly available, and be used as a large data set for computer scientists and statisticians to evaluate analysis and data mining

techniques.

The report, available today in the *BMC [Medical Informatics](#) and Decision Making Journal*, looked at a random sample of individual-level data files; part of the discharge abstract database from the Canadian Institute for [Health Information](#) (CIHI). Two different PUMFs were produced; one with geographic information and another without geographic information but containing more clinical information. The findings were clear – the PUMFs ensured that the risk of re-identifying individual patients was very low and after the changes made to the data to protect patient identity the data was still useful for analysis.

To create the PUMF, the paper describes new metrics for measuring re-identification risks; develops a new efficient algorithm for minimizing the amount of information that needs to be removed from the data; explores the different plausible ways that Canadians' health data can be attacked; and demonstrates tactics that can be used to maximize the value of the data that is released. The report provides an example of the steps that need to be followed to create a PUMF.

"Canada has a single payer system, which means that population-level data sets already exist. Currently gaining access to such data is often limited to certain groups, is complex and time-consuming," said Dr. El Emam. "Making data public is not difficult to do; data quality can be maintained; and the economic and health system benefits are substantial! Worries about privacy are simply not a convincing excuse anymore as it's largely a solvable problem."

More information: Link to report:
www.biomedcentral.com/1472-6947/11/53/abstract

Provided by Children's Hospital of Eastern Ontario Research Institute

Citation: New report on creating clinical public use microdata files (2011, September 15)
retrieved 11 June 2026 from <https://medicalxpress.com/news/2011-09-clinical-microdata.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.