No Place to Hide — Reverse Identification of Patients from Published Maps

TO THE EDITOR: The mapping of health data is now widespread in both academic research and public health practice. Although the notion that location influences the risk of disease dates back to the mapping of yellow fever and cholera in the 1800s, research that integrates maps with human health is an emerging field based on the widespread availability of geographic information system (GIS) software. Such systems have broad applicability, and their use has been fueled by the availability of increased computing power, user-friendly software, and large geographic databases. The number of publications that use GIS data for health research has grown by about 26% per year, four times the rate of increase in the number of articles on human health in general. Patients’ addresses are mapped to identify patterns, correlates, and predictors of disease. These maps are then published electronically and in print.¹

Using keyword searches for the terms “geographic” and “map” in the figure legends of articles in five major medical journals published...
between 1994 and 2005, we identified 19 articles (including 5 in the Journal) that included maps with the addresses of patients plotted as individual dots or symbols. In these articles, more than 19,000 such addresses were plotted on maps. Given the potential implications for the privacy of patients, we investigated whether we could use these published maps to reidentify the patients. We created a simulated map of 550 geographically coded addresses of patients in Boston, using the minimum figure resolution required for publication in the Journal (Fig. 1A). We then used standard GIS techniques to determine the accuracy with which such addresses can be identified. Strikingly, the reverse-identification method precisely identified 432 of the addresses (79%) and identified all 550 addresses within 14 m of the correct address (Fig. 1B).

The publication of maps of disease with precise locations of patients jeopardizes patients’ privacy. Guidelines for the display or publication of health data are needed to guarantee patients’ anonymity. A common approach has been to map according to administrative unit rather than home address. However, the aggregation of data in this manner places constraints on the visualization of disease patterns. Another method is spatial skewing, or randomly relocating patients’ addresses within a given distance of their true location. Skewing can allow a visualization that conveys the necessary information while preserving patients’ privacy. Both aggregation and skewing are systematic and reliable means of de-identification that are far safer, in terms of protecting identifiable health information, than simply reducing the resolution of a map. Editors of journals and textbooks should consider implementing such policies to guide the safe reporting of spatial data.

John S. Brownstein, Ph.D.
Children’s Hospital
Boston, MA 02115
john_brownstein@harvard.edu

Christopher A. Cassa, M.Eng.
Harvard–MIT Division of Health Sciences and Technology
Boston, MA 02139

Kenneth D. Mandl, M.D., M.P.H.
Harvard Medical School
Boston, MA 02215


Correspondence Copyright © 2006 Massachusetts Medical Society.