

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.¹ 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.² 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.² 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

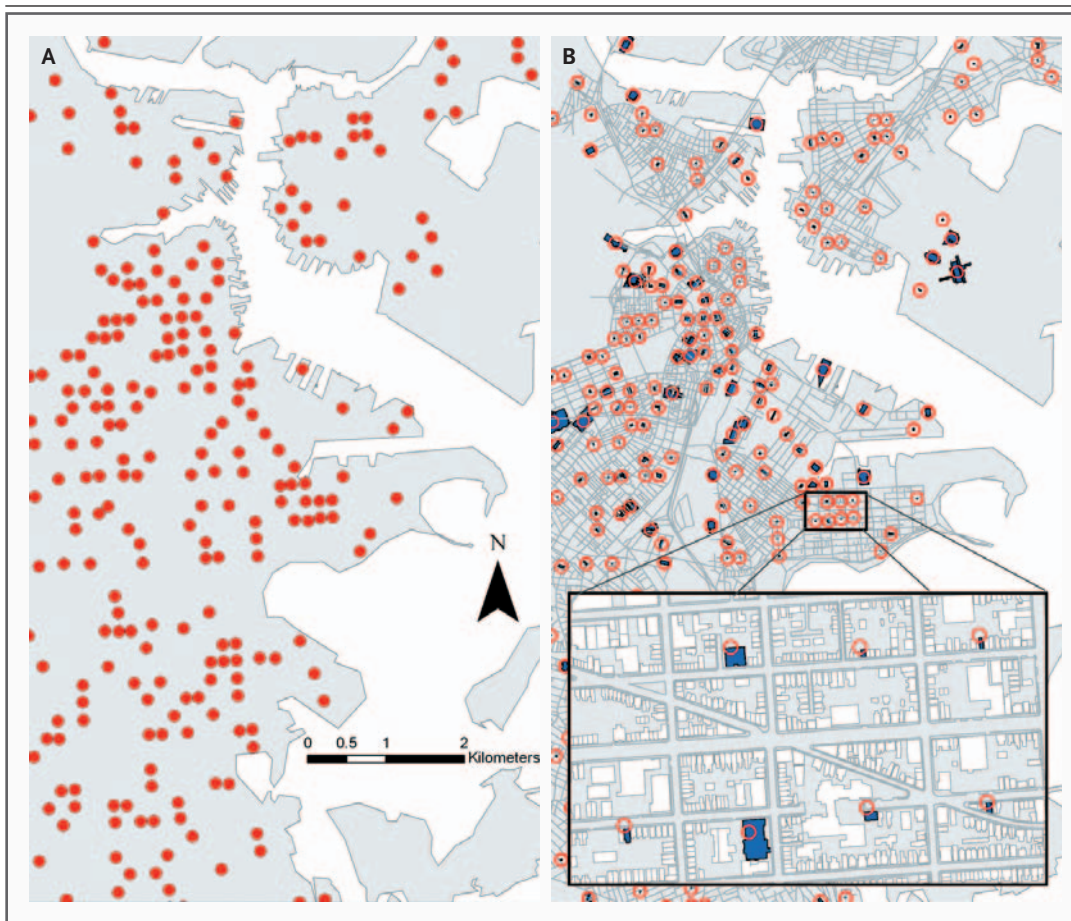


Figure 1. Reverse Identification of Patients from a Simulated Health-Data Map of Boston.

Panel A shows a section of a map with the address locations of 550 patients (circles) selected according to a stratified random-sampling design. The original JPEG image that was used in the analysis had a resolution of 266 dots per inch (the minimum resolution required by the *Journal*), a file size of 712 kb, and a scale of 1:100,000. Panel B shows the results of reverse identification of the patients' addresses. The circles indicate the predicted locations of the patients' homes according to the reverse-identification method, and the blue shapes outline the patients' actual homes (with a portion of a neighborhood shown in detail in the inset).

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.³ 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.⁴ 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 preserv-

ing patients' privacy.⁵ 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 02115

1. Croner CM, Sperling J, Broome FR. Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships. *Stat Med* 1996;15:1961-77.
2. Pickle LW, Waller LA, Lawson AB. Current practices in cancer spatial data analysis: a call for guidance. *Int J Health Geogr* 2005;4:3.
3. Brownstein JS, Cassa CA, Kohane IS, Mandl KD. Reverse geocoding: concerns about patient confidentiality in the display of geospatial health data. *AMIA Annu Symp Proc* 2005:905.
4. Rushton G, Armstrong MP, Gittler J, et al. Geocoding in cancer research: a review. *Am J Prev Med* 2006;30:Suppl:S16-S24.
5. Cassa CA, Grannis SJ, Overhage JM, Mandl KD. A context-sensitive approach to anonymizing spatial surveillance data: impact on outbreak detection. *J Am Med Inform Assoc* 2006;13:160-5.

Correspondence Copyright © 2006 Massachusetts Medical Society.

INSTRUCTIONS FOR LETTERS TO THE EDITOR

Letters to the Editor are considered for publication, subject to editing and abridgment, provided they do not contain material that has been submitted or published elsewhere. Please note the following: •Letters in reference to a *Journal* article must not exceed 175 words (excluding references) and must be received within three weeks after publication of the article. Letters not related to a *Journal* article must not exceed 400 words. All letters must be submitted over the Internet at <http://authors.nejm.org>. •A letter can have no more than five references and one figure or table. •A letter can be signed by no more than three authors. •Financial associations or other possible conflicts of interest must be disclosed. (Such disclosures will be published with the letters. For authors of *Journal* articles who are responding to letters, this information appears in the original articles.) •Include your full mailing address, telephone number, fax number, and e-mail address with your letter.

Our Web site: <http://authors.nejm.org>

We cannot acknowledge receipt of your letter, but we will notify you when we have made a decision about publication. Letters that do not adhere to these instructions will not be considered. Rejected letters and figures will not be returned. We are unable to provide prepublication proofs. Submission of a letter constitutes permission for the Massachusetts Medical Society, its licensees, and its assignees to use it in the *Journal's* various print and electronic publications and in collections, revisions, and any other form or medium.