



Why Health Services Research Needs Geoinformatics: Rationale and Case Example

Tracy Onega^{1,2,3}, Jennifer Alford-Teaster^{1,2*}, Steven Andrews⁴, Craig Ganoë⁴, Mike Perez JD⁵, David King BS⁵ and Xun Shi⁶

¹Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

²Norris Cotton Cancer Center, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

³The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA

⁴Collaboratory for Healthcare and Biomedical Informatics, Geisel School of Medicine at Dartmouth, Hanover, NH, USA

⁵Exaptive, Inc., Cambridge, MA, USA

⁶The Geography Department, Dartmouth College, Hanover, NH, USA

Introduction

Delivery of health care in the United States has become increasingly complex over the past 50 years, as health care markets have evolved, technology has diffused, population demographics have shifted, and cultural expectations of health and health care have been transformed. Identifying and understanding important patterns of health care services, accessibility, utilization, and outcomes can best be accomplished by combining data from all of these dimensions in near-real time. The Big Data paradigm provides a new framework to bring together very large volumes of data from a variety of sources and formats, with computing capacity to derive new information, hypotheses, and inferences [1,2]. The complementary fields of genomics and bioinformatics have already made great advances only made possible by Big Data approaches. Similar gains can be made by pairing health services research with geoinformatics -- defined as "the science and technology dealing with the structure and character of spatial information, its capture, its classification and qualification, its storage, processing, portrayal and dissemination, including the infrastructure necessary to secure optimal use of this information" [3]. Integrating geospatial technologies with health services research brings informatics approaches, data sciences, and spatial theories of health and healthcare together to explore relationships among geography, health, and delivery of care in novel ways made possible through geoinformatics. synergy between the two disciplines will enhance our ability to discover how health care is delivered most effectively for the greatest health benefits across populations.

Shared History of Geography and Health Services Research: Successes and Limitations

Health services research and geography have intersected under the rubrics of medical geography, epidemiology of health care, small area analysis, and public health since ancient Roman times, but more formally since the 1930s. Then, James Glover--an English physician -- noticed that tonsillectomies were occurring at highly variable rates across school districts, which could not be explained by geographic, socio-demographic, or clinical factors, and the most likely explanation suggested being differences in how physicians practice [4]. Similar work, begun in the 1970s, became the impetus for the Dartmouth Atlas of Health Care [5], which used health care utilization data to develop geographic units representing health care markets. These health care-based spatial units can be directly compared to identify patterns associated with both effective and ineffective care. Other notable examples of health services research linked to a geospatial framework can be found in the public health arena, with planning of population-based vaccination programs and designation of federally-qualified health centers. Despite these tremendous contributions, limitations exist that geoinformatics and health service research are now poised to overcome as information technology and the digital era expands data

availability, accessibility, usability, and timely knowledge generation.

Understanding spatiotemporal distributions of health services is a fundamental aspect of health services research upon which studies of utilization, outcomes, comparative effectiveness, resource allocation, and others are based. Four key limitations can be found in typical approaches to measuring health services distributions: 1. retrospective methods; 2. limited geographic extents; 3. ascertainment challenges; and 4. structured data only; which will be addressed below (Table 1). By combining geoinformatics with health services research, an important domain of questions within the field of medical informatics can be addressed, such as is illustrated with a case example.

A Health Services Research Problem: Diffusion of Medical Technology

Technological innovation is a hallmark of the U.S. health care system, and relies on the backbone of translational research to evaluate effectiveness after efficacy has been established. The full potential of a new technology is determined as clinical improvements and impacts on population health are assessed. Yet typically diffusion and the research to establish effectiveness occur asynchronously. This limits the potential for timely assessment of broad clinical impact and leads to two concerns: 1. overuse of technologies with unproven or minimal benefits in the general population and concomitant unwarranted costs; 2. underuse of technologies that have beneficial effects and improve outcomes for particular patient populations. As medical technology diffuses into generalizable practice, data and research needs arise to address these concerns from a host of stakeholders perspectives, including patients, clinicians, researchers, health care facilities, commercial vendors, payers, health care systems, communities, and regulatory bodies. A prerequisite for assessing new technologies in community practice is knowledge of the locations or *extent* of diffusion and the populations reached. Data availability and timeliness are critical barriers to establishing this knowledge, contributing to spotty information that is retrospective, often with notable lags. For example, The Dartmouth Atlas of Healthcare [5] and other work [6-8] has

***Corresponding author:** Jennifer Alford-Teaster, Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA, Tel: 603-653-6088; E-mail: jennifer.a.alford-teaster@dartmouth.edu

Received October 08, 2014; **Accepted** November 28, 2014; **Published** December 01, 2014

Citation: Onega T, Alford-Teaster J, Andrews S, Ganoë C, Mike Perez JD, et al. (2015) Why Health Services Research Needs Geoinformatics: Rationale and Case Example. J Health Med Informat 5: 176. doi:10.4172/2157-7420.1000176

Copyright: © 2015 Onega T, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Key Current Limitations in Geographic Based Health Services Research	Example Geoinformatics Approaches to Address Current Limitations
Retrospective methods available data is usually retrospective with lag in timeliness, so pattern described are rarely current.	Using web content mining and mobile technology feeds, near –real time locational data can be obtained.
Limited geographic extents Typically a tradeoff exists between rich/granular data available for small geographic extents, or coarse/ broad data available for large geographic extents.	Use of internet, and mobile- technology based locational data mining, allows for very broad geographic capture, without spatial scale limitations.
Ascertainment Challenges Locational/spatial data is often not available at a point location (address) for health services and/ or patients. Locational data may be only at an area level (e.g. ZIP code), or not known with certainty or completeness at all.	Use of internet, and mobile- technology based locational data mining, is based on either point locations (latitude-longitude) from IP address, or address mining that can be automatically geocoded to point location.
Structured data only typically only spatial data in structured form –such as databases or files-is available. Manual abstraction of spatial data is possible, but limits the scale of examination.	Content Mixing (text and images) allows the use of unstructured data from the internet and mobile technology feeds to Obtain locational information that is not captured explicitly in an existing database.
Asynchronous evaluation of technology with outcomes because evaluation of new technologies is typically limited by the factors above, evaluation of technology occurs <i>after</i> it is already being used in actual practice, and often in <i>small or non representative areas and or populations</i> . This asynchronicity creates the potential for detrimental outcomes to occur prior to establishing outcomes to be unavailable to populations who may benefit if known that a technology should be available.	By addressing the limitations noted above, evaluation of technology is more likely to occur in near real time, allowing for outcomes to be determined in a more timely manner, rather than with a temporal lag during which patients and populations may be impacted by negative outcomes not previously understood, or missing out on positive outcomes if the technology is revealed to not be located where a population may benefit.

Table 1: Summary of major limitations in geographical based health services research and how geoinformatics approaches can be used to address them.

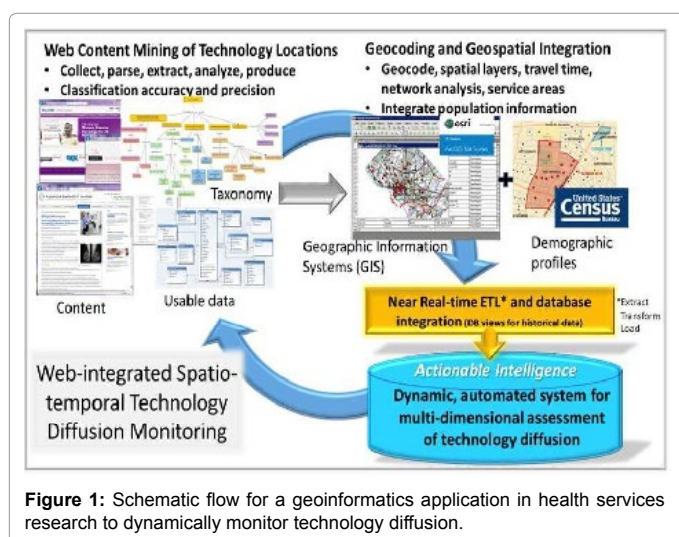


Figure 1: Schematic flow for a geoinformatics application in health services research to dynamically monitor technology diffusion.

described geographic variation of medical technologies at a national level, but relied on Medicare data, which are typically available for research with a lag of 2-3 years. Further, Medicare, as well as private insurers, are dependent on the new technology being approved for reimbursement since they rely on billing data (claims), and coverage of the technology may take years following FDA approval and commercial dissemination, thus not able to be ascertained until unique billing codes are implemented. Some data resources may be timely for monitoring diffusion—such as registries like the Breast Cancer Surveillance Consortium (BCSC) and the HMO Research Network (HMORN) [9,10], and other clinical/provider networks with health information exchanges and/or robust electronic health records. However, these data resources are limited in geographic extent, are often not population-based, and may not capture use of new technologies until they are uniquely coded (e.g. CPT–Common Procedural Terminology), or are reimbursable. Further, most data sources require existing structured data, although natural language processing (NLP) is increasingly applied to unstructured medical record information, which can be powerful, but also time-consuming, complicated and variable across settings [11-13]. Achieving national extent, fully ascertained, and timely data capture to characterize geographic and sociodemographic diffusion of new technologies is a critical need, particularly as the U.S. seeks to improve health and healthcare and limit health care costs to effective use within populations.

Measuring Near-Real Time Diffusion of Breast Imaging Technology

To more fully understand dissemination of new technologies—we need to be able to capture the occurrences of the technology in large geographic areas, and in a dynamic way that reflects the dynamic process that dissemination is. Figure 1 presents a schematic approach to do this, using a breast imaging technology diffusion case example (Figure 1). We can address the existing limitations in measuring, monitoring, and characterizing health care diffusion—with breast imaging as a timely example, given new technologies, such as digital breast tomosynthesis (DBT) [14-16], and legislation related to breast density notification [17]. With a geospatial semantic web [18-22], which combines web mining techniques with geographic information systems and census data, one can ascertain geographic uptake of DBT nationally, estimate potential access overall and by population subgroups, and identify correlates of dissemination patterns.

Web content mining is used in this project to identify instances of DBT based on taxonomy of terms. Using associated web pages from these instances, street address information is captured for the DBT instance. These addresses and related attributes (facility name, date of data capture, etc.) are brought into a GIS for geocoding, spatial joins with other layers (e.g. road network), travel time analysis, and service area creation. Population demographics and other census-derived data are attributed to DBT locations and service areas. This constitutes the database containing geographic extent of DBT diffusion and population characteristics served by DBT facilities, which is refreshed at regular intervals following external validation. This application characterizes heterogeneous diffusion both geographically and socio-demographically, which also serves as a proof of concept for higher-dimension data which incorporates spatial layers, as well as other technologies.

Conclusion

The application of geoinformatics to health services research has high potential for advancing currently used research methods to monitor and evaluate new technologies as they are translated from experimental settings into communities and populations. A successful application of this approach will yield a validated tool to dynamically integrate geospatial data, population data, and web content for automated discovery and monitoring of technology diffusion. For example, a user interface will provide static functions, such as maps

of service locations, derived service areas, density of services available, and populations that are in service catchment areas. Interactive and near-real time functions could produce for user-defined areas, time, and populations: video/time trends of service locations, derive-time defined service areas, time trends of populations coincident with services, and projected population coverage for actual or potential service locations. Such tools can be readily scalable, applicable to other new technologies, and foundational for further capabilities, such as using geostatistical methods for predictive modeling, visualization, and other “Big Data” analytics.

Acknowledgments

Research reported in this (publication/press release) was supported by The Dartmouth Clinical and Translational Science Institute, under award number UL1TR001086 from the National Center for Advancing Translational Sciences (NCATS) of the National Institutes of Health (NIH). The content is solely the responsibility of the author(s) and does not necessarily represent the official views of the NIH

References

1. Philip CCL, Zhang CY (2014) Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences* 275: 314-347.
2. Kambatla K, Kollias G, Kumar V, Grama A (2014) Trends in big data analytics. *Journal of Parallel Distributed Computing* 74: 2561-2573.
3. Raju PLN (2003) Fundamentals of geographic information systems. In Workshop: satellite remote sensing and GIS applications in agricultural meteorology, India.
4. Glover JA (1938) The incidence of tonsillectomy in school children. *Proceedings of the Royal Society of Medicine* 31: 95-113.
5. The Dartmouth Atlas of Health Care.
6. Onega T, Toseston TD, Wang Q, Hillner BE, Song Y, et al. (2012) Geographic and sociodemographic variation of PET use in Medicare beneficiaries with cancer. *J Am College Rad* 9: 635-642.
7. Onega T, Hubbard R, Hill D, Lee CI, Haas JS, et al. (2014) Geographic Access to Breast Imaging for U.S. Women *J Am College. Rad* 11: 874-882
8. Onega T, Duell EJ, Goodman DC, Demidenko E, Shi X, et al. (2008) Geographic access to cancer care in the U.S. *Cancer* 112: 909-918.
9. Breast Cancer Surveillance Consortium (BCSC).
10. HMO Research Network (HMORN).
11. Xu R, Garten Y, Superkar K, Das AK, Altman RB, et al. (2007) Extracting subject demographic information from abstracts of randomized clinical trial reports. *Proceedings of the 12th World Congress on Medical Informatics Medinfo2007*: 550-554
12. Xu R, Superkar K, Huang Y, Das AK, Garber AM (2006) Combining text classification and Hidden Markov Modeling techniques for structuring randomized clinical trial abstracts. *Annual Symposium*: 824-828
13. Embley DW (2004) Toward semantic understanding: An approach based on information extraction ontologies. *ACM International Conference Proceeding Series. Australian Computer Society* 52: 3-12.
14. Dobbins JT (2009) Tomosynthesis imaging: at a translational crossroads. *Medical Physics*. 36: 1956-1967.
15. Kopans DB (2014) Digital breast tomosynthesis from concept to clinical care. *Am J Roentgenol* 202: 299-308.
16. Lee CI, LC (2013) Digital breast tomosynthesis and the challenges of implementing an emerging breast cancer screening technology into clinical practice. *Journal of the American College of Radiology* 10: 913-917.
17. Brower V (2013) Breast density legislation fueling controversy. *Journal of the National Cancer Institute* 105: 510-511.
18. Roman D, Klien E (2007) SWING—A semantic framework for geospatial services. In A. Scharl & K. Tochtermann, *The geospatial web: How geobrowsers, social software and the web 2.0 are shaping the network society*, pp. 229-234
19. Yue P, Di L, Yang W, Yu G, Zhao P (2007) Semantics-based automatic composition of geospatial Web service chains. *Computers and Geoscience* 33: 649–665.
20. Brodaric B, Fox P, McGuinness DL (2009) Geoscience knowledge representation in cyberinfrastructure. *Computers and Geosciences*. 35: 697-699.
21. Egenhofer MJ (2002) Toward the Semantic Geospatial Web. In *Proceedings of the 10th ACM international symposium on advances in geographic information systems*, McLean, Virginia.
22. Yang C, Raskin R, Goodchild M, Gahegan M (2010) Geospatial Cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems* 34: 264–277.