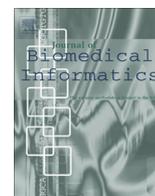


Contents lists available at [ScienceDirect](http://ScienceDirect.com)

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Methodological Review

Visualization and analytics tools for infectious disease epidemiology: A systematic review



Lauren N. Carroll^a, Alan P. Au^a, Landon Todd Detwiler^b, Tsung-chieh Fu^c, Ian S. Painter^d,
Neil F. Abernethy^{a,d,*}

^a Department of Biomedical Informatics and Medical Education, University of Washington, 850 Republican St., Box 358047, Seattle, WA 98109, United States

^b Department of Biological Structure, University of Washington, 1959 NE Pacific St., Box 357420, United States

^c Department of Epidemiology, University of Washington, 1959 NE Pacific St., Box 357236, Seattle, WA 98195, United States

^d Department of Health Services, University of Washington, 1959 NE Pacific St., Box 359442, Seattle, WA 98195, United States

ARTICLE INFO

Article history:

Received 13 September 2013

Accepted 3 April 2014

Available online 16 April 2014

Keywords:

Visualization

Infectious disease

Public health

Disease surveillance

GIS

Social network analysis

ABSTRACT

Background: A myriad of new tools and algorithms have been developed to help public health professionals analyze and visualize the complex data used in infectious disease control. To better understand approaches to meet these users' information needs, we conducted a systematic literature review focused on the landscape of infectious disease visualization tools for public health professionals, with a special emphasis on geographic information systems (GIS), molecular epidemiology, and social network analysis. The objectives of this review are to: (1) identify public health user needs and preferences for infectious disease information visualization tools; (2) identify existing infectious disease information visualization tools and characterize their architecture and features; (3) identify commonalities among approaches applied to different data types; and (4) describe tool usability evaluation efforts and barriers to the adoption of such tools.

Methods: We identified articles published in English from January 1, 1980 to June 30, 2013 from five bibliographic databases. Articles with a primary focus on infectious disease visualization tools, needs of public health users, or usability of information visualizations were included in the review.

Results: A total of 88 articles met our inclusion criteria. Users were found to have diverse needs, preferences and uses for infectious disease visualization tools, and the existing tools are correspondingly diverse. The architecture of the tools was inconsistently described, and few tools in the review discussed the incorporation of usability studies or plans for dissemination. Many studies identified concerns regarding data sharing, confidentiality and quality. Existing tools offer a range of features and functions that allow users to explore, analyze, and visualize their data, but the tools are often for siloed applications. Commonly cited barriers to widespread adoption included lack of organizational support, access issues, and misconceptions about tool use.

Discussion and conclusion: As the volume and complexity of infectious disease data increases, public health professionals must synthesize highly disparate data to facilitate communication with the public and inform decisions regarding measures to protect the public's health. Our review identified several themes: consideration of users' needs, preferences, and computer literacy; integration of tools into routine workflow; complications associated with understanding and use of visualizations; and the role of user trust and organizational support in the adoption of these tools. Interoperability also emerged as a prominent theme, highlighting challenges associated with the increasingly collaborative and interdisciplinary nature of infectious disease control and prevention. Future work should address methods for representing uncertainty and missing data to avoid misleading users as well as strategies to minimize cognitive overload.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

Abbreviations: GIS, geographic information systems; PHIN, Public Health Information Network; SVG, scalable vector graphics; DHTML, dynamic HTML.

* Corresponding author at: Department of Biomedical Informatics and Medical Education, University of Washington, 850 Republican St., Box 358047, Seattle, WA 98109, United States.

E-mail addresses: ln carr@uw.edu (L.N. Carroll), aa u@uw.edu (A.P. Au), det@uw.edu (L.T. Detwiler), tfu@uw.edu (T.-c. Fu), ipainter@uw.edu (I.S. Painter), neila@uw.edu (N.F. Abernethy).

1. Introduction

In the last 20 years, an increasing focus on the need for informatics and analytics in public health has resulted in a growing investment in information systems [1–7]. This investment has generated a myriad of new tools for different public health activities and jurisdictions, including tools and systems developed by

<http://dx.doi.org/10.1016/j.jbi.2014.04.006>

1532-0464/© 2014 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

federal, state and local governments, as well as research organizations [8–12]. Advances in electronic reporting and interoperability, computer technology, biotechnology (e.g. genetic sequencing), and other methods (e.g. social network analysis and geographic information systems) have put pressure on the informatics discipline and public health practitioners alike to translate these advances into common practice [1,7,13,14]. This pressure has been particularly acute for the surveillance and management of infectious diseases with pandemic or bioterrorism potential [7,15–17].

To characterize the variety of tools and analytical approaches developed for infectious disease control, we conducted a systematic literature review of informatics tools for infectious diseases, with a focus on platforms for information visualization. In this review, we assessed the current landscape of these tools in terms of information needs and user preferences, features and system architectures of existing tools, as well as usability and adoption considerations. Due to the challenges of integrating, analyzing, and displaying public health data, particularly new types of data encountered in public health, this review places a special emphasis on efforts to visualize geographic information systems (GIS), molecular epidemiology, and social networks.

1.1. Background

Since John Snow first plotted cholera cases on a map of London, graphs and visualizations have played important roles in epidemiology, supporting communication, aggregation, analysis, and use of data for hypothesis testing and decision making [18,19]. In the electronic age, computer-aided generation of charts, maps, and reports have enabled a further increase in the use of visualization tools to supplement individual-level clinical data and population-level statistics [7,15]. Infectious disease burden in the population, whether measured for programmatic or outbreak management purposes, is now commonly analyzed in terms of geographic distribution, clinical risk factors, demographics, molecular and phylogenetic features, or sources of exposure such as social networks [20–23]. While routine features of public health reports include epidemic curves and choropleth maps, new visualization motifs such as social network graphs and phylogenetic trees have increasingly been used to characterize disease outbreaks [24,25]. Indeed, a keyword search by year in PubMed highlights the increased reference to GIS, molecular epidemiology, and social network analysis in publications relative to all indexed PubMed publication (Fig. 1).

Tools for these three types of complex data allow public health professionals and researchers to integrate, synthesize, and visualize information pertaining to disease surveillance, prevention, and control. The ability to track disease distribution with GIS tools has helped public health professionals and researchers alike to detect disease clustering, analyze spread of disease in communities and across territories, and to predict outbreaks [26–30]. Surveillance of different strains of tuberculosis, influenza, and other diseases via characterization of molecular markers is commonly used to identify potential risk factors, pathogenicity, potential outbreaks, and prepare adequate interventions [31–36]. With the growth of network theory and the availability of modern computing, social network analysis and network-based epidemic models have been increasingly used to depict outbreaks and disease dynamics [37–40], identify potential cases and focus control efforts by prioritizing contacts [24], and evaluate strategies to interrupt transmission [40–42]. Together these data types can tell a compelling story about disease risk factors, spread and transmission, and can lead to more effective control measures and interventions.

However, this surge in surveillance capacity has produced more complex and disparate data, leading to new discussions about data sharing and interoperability, data confidentiality, and strategies for managing redundancies as well as incomplete data

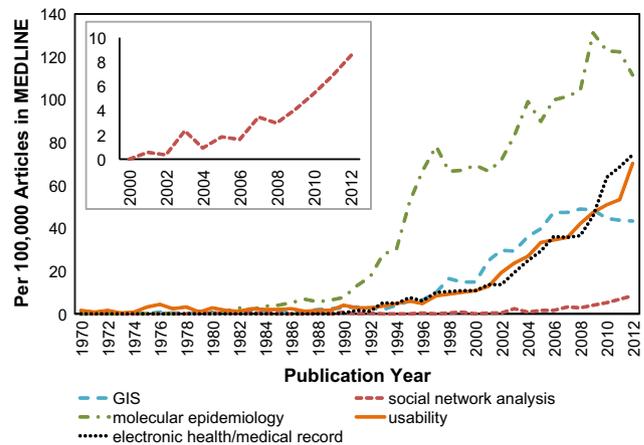


Fig. 1. Increased reference to common complex data types. Keyword search for GIS, molecular epidemiology, and social network analysis in PubMed highlights the increase in these terms relative to all PubMed index articles. The frequency of other biomedical informatics terms (usability, electronic health record) is shown for comparison. Although the growth of social network analysis has been more recent, the inset shows that this concept has also experienced rapid growth in the published literature.

[1,17,29,43–46]. For example, public health practitioners and researchers are faced with integrating diverse data sources such as mortality data (e.g. autopsy reports), clinical data (e.g. laboratory reports, immunization records), geographical data (e.g. address of work, residence, preschool), relationships (e.g. names of family, friends, partners), patient and pathogen genetics, medical imaging, travel plans, and timelines. Each of these types of information can be recorded, stored, accessed, evaluated, and displayed in many different systems and formats. Organizations are therefore challenged to maximize the potential of this flood of data to impact public health practice. Visualization tools have the potential to improve comprehension of this data by increasing the memory and processing resources available to users, reducing the search for information, enhancing the detection of patterns, and providing mechanisms for inference [47]. However, visualization tools also risk misleading users due to misinterpretation or cognitive overload [48,49].

As such, funders and developers of visualization tools encounter a range of challenges when designing new tools for public health data, generating a growing collection of tools as new ideas and approaches are explored. However, these tools are often developed in silos, limiting their use in practice [50]. And despite the advances in public health informatics, many public health professionals still use visualization tools and data management systems that may no longer suit their current needs [6,7,51]. The unique focus of this systematic review on visualizing GIS, molecular epidemiology, and social network data for infectious diseases highlights the progress to date in public health informatics for infectious disease by identifying information needs and user preferences, characterizing features and system architectures of existing visualization tools, as well as identifying usability and adoption considerations. Finally, we explore commonalities among complex data types and underscore some of the challenges that lie ahead for novel visualization tool development.

2. Methods

This review explored the lifecycle of development and adoption of infectious disease visualization tools from conception to evaluation in practice. Infectious disease surveillance and control efforts encompass a wide variety of fields and require integration, synthesis, and analysis of information [21,52,53]. Consequently, we

employed a sensitive search strategy for this review in hopes of capturing relevant literature from diverse fields. This review focused on the following objectives:

1. Identify public health user needs and preferences for infectious disease information visualization tools.
2. Identify existing infectious disease information visualization tools and characterize their architecture and features.
3. Identify commonalities among complex data types.
4. Describe tool utility and usability evaluation efforts, and characterize barriers to the adoption of such tools.

To gain a comprehensive understanding of the current landscape of these tools, we identified articles published in English from January 1, 1980 to June 30, 2013 from the following bibliographic databases: National Library of Medicine's MEDLINE through PubMed, Cochrane Library, New York Academy of Medicine's Grey Literature, Web of Science, and IEEE Digital Library. Articles identified through additional manual searches were subject to the same inclusion criteria.

2.1. Selection criteria

The final search strings were (A and B), (A and C and D), and (C and E and F), where each chain is defined as the following: (A) infectious disease OR public health data; (B) information needs OR task analysis; (C) visualization OR visualisation OR mapping; (D) molecular epidemiology OR social network analysis OR geospatial OR geographic OR GIS OR adoption OR utility OR outbreak OR surveillance OR disease mapping OR contact investigation OR usability OR functional requirements OR interactive OR real time OR needs assessment; (E) software; (F) usability OR adoption OR functional requirements.

To ensure a focus on the current landscape of infectious disease visualization tools, articles were excluded if the primary focus of the article was: clinical trials, decision-making aids, learning behavior, cognitive behavioral theory, disease or outbreak case studies, (health) information networks, data mining, concept mapping, systems mapping, programming language, ontologies and taxonomies, software methodology or framework, and resource mapping. Moreover, studies were excluded if the primary emphasis was: laboratory methodology, epidemic modeling or statistics, risk mapping, public health interventions, non-human infectious disease, architecture or system visualization, software case studies, and healthcare or medical treatment. The final set of articles were assessed for quality, with a focus on methods and risk of bias (e.g. selection, detection, reporting).

3. Results

Of the 247 articles we screened, a total of 88 articles are included in this review (Appendix A) and the process is described in Fig. 2 per the PRISMA guidelines [54]. The articles primarily included descriptive reports, qualitative studies (e.g. interviews, focus groups), and usability studies. None were excluded due to methodological deficiencies. The literature included in this review is comprised of articles from both US and non-US journals. The content was abstracted and the articles were organized into the following categories based on the primary topics discussed in the articles: information needs and learning behavior ($n = 18$); architecture of existing tools ($n = 22$); user preferences ($n = 20$); features of existing tools ($n = 54$); usability and evaluation ($n = 15$); and implementation and adoption ($n = 27$). These categories highlight the logical progression of novel tool development. Note that these

categories are not mutually exclusive. Summaries of findings in each category are described in the sections below.

3.1. Information needs and learning behavior

The types of information required by public health professionals have been studied in many contexts. The studies meeting our inclusion criteria offered several insights about information seeking behavior among public health professionals. While the public health workforce is extremely diverse [55–60] and public health information sources are often disparate and unstandardized [46,55,56,60], several themes held constant. Public health professionals need timely access to current data from reliable, high quality sources [9,55,58,59,61,62]. Furthermore, public health professionals need synthesized and collated data on relevant information such as best practices, effective prevention strategies or interventions, and evidence-based research, to name a few [9,55,59,61,63]. Public health professionals gather information from colleagues, literature and health departments [55,57,58,62,64]. However, multiple studies suggested that public health professionals are still often unaware of available information resources, and emphasized collaboration to improve search outcomes [56,60,63,65]. Additional challenges associated with meeting information needs include external barriers (e.g. lack of time, sufficient staff), technological barriers (e.g. inadequate equipment, lack of internet access), internal barriers (e.g. stress, lack of confidence in ability to complete task, lack of training), and lack of trust in the information source [9,55–58,60,62–65]. These studies suggested centralized access to reliable resources, as well as improved access to and delivery of timely information, as key to overcoming these barriers.

However, information needs specifically pertaining to information visualization tools have not been as well explored. Two studies explored the context in which participants learned about, used, and synthesized information from visualization tools through interviews and questionnaires with public health professionals [62,66]. The first highlighted the importance of prior knowledge and intuition to give context to the results, and demonstrated participants' frustration with tools that were not intuitive or were too awkward for regular use [66]. The second study indicated that public health professionals spend less than 10 hours per month learning about new tools or methods for work and primarily learn about them from internet, literature, conferences and colleagues [62]. Participants wanted to know how the tool was developed and by whom (e.g. author's name(s), fields of expertise, credentials, affiliations) [62] as well as how the tool provided its results [66]. Additional studies also highlighted the importance of the users' perception of, and trust in, the tool's reliability as a potential learning barrier to new visualization tools [9,49,67]. In a study of user needs and preferences for visualization tools, sixty percent of users indicated they typically use more than one visualization tool for their visualization and analysis needs in a recent questionnaire [67]. This finding was supported in multiple studies wherein users indicated that no one existing tool or system met all their data needs [49,55,58,66]. Further, studies indicated that users most commonly created static graphics, and many users relied on Microsoft Office suite [49,62,66,67]. Collectively, these findings indicate many users are interested in learning about new tools in a time-efficient manner, and support an important relationship between user trust, tool credibility, and transparency.

Multiple articles raised concerns regarding interpretation of graphics, specifically misinterpretation of results and cognitive overload. For example, some users voiced concerns that data can be manipulated or unintentionally misrepresented due to confusion about how the tool works, or what type of graphic to use if given options [9,49,68]. Cognitive overload, wherein a user is presented with more information than they are able to successfully

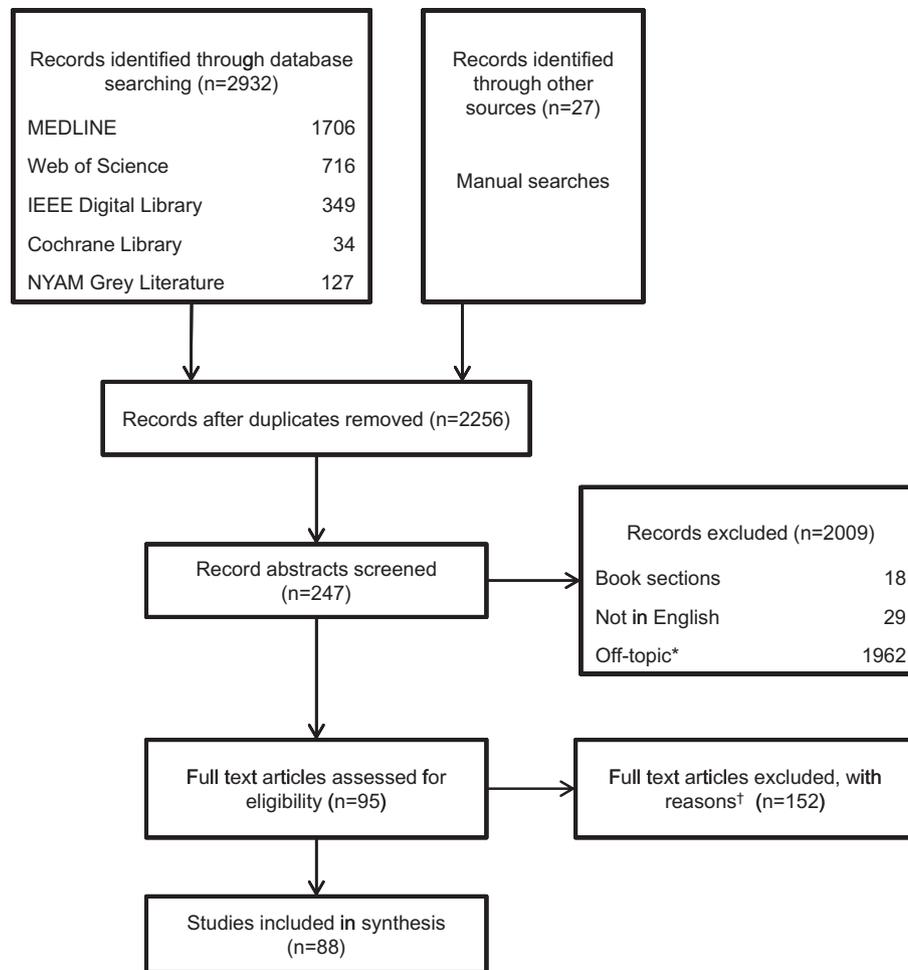


Fig. 2. Flowchart of literature review process. *Off-topic exclusions: clinical trials, decision-making aids, learning behavior, cognitive behavioral theory, disease or outbreak case studies, (health) information networks, data mining, concept mapping, systems mapping, programming language, ontologies and taxonomies, software methodology or framework, and resource mapping. †Full text exclusions: laboratory methodology, epidemic modeling or statistics, risk mapping, public health interventions, genome mapping, not human infectious disease, architecture or system visualization, software case studies, and healthcare or medical treatment.

process, was addressed in several studies. This highlights the unique challenge of displaying complex and large datasets without reducing usability reaching the technical limits of the platform or the cognitive limits of the user [69,70]. Strategies to minimize cognitive overload were less defined, although Herman et al. [69] suggested human-centered design as a means of improving data visualization interpretation.

The data sources available for a given target end-user influence the architecture of the visualization tool. The next section explores common architectures reported by the articles included in this review.

3.2. Architecture of existing tools

We considered *architecture* to address the means by which a system was constructed in the software design sense, referring to the way in which system components fit together. Components may be individual classes in a software program or larger components, like a database management system, a web service, and the connections in between these components. Other features, such as interface design, operation workflow, functionality, features, visualization layouts, and analysis algorithms are often independent of underlying system architecture. These are covered in later sections of this review.

Several articles in this review made only cursory reference to system architecture. For example, some papers referenced use of

specific components such as a particular database, management system, GIS, or statistics package [71–76]. Others alluded to particular architectural choices through discussion of other technical issues, for example the computational complexity of a statistical routine [73,77]. However, these references alone gave little insight into the structure of the system as a whole. Some of articles in this review contained more significant coverage of system architecture, including a discussion of the general architectural design in terms of the number and function of system tiers [28,29,78,79]. This may reflect the purpose behind many such publications, which typically focused on the utility of design features for public health purposes or the challenges inherent in linking data to visualization tools. One publication, however, explicitly described the *structured application framework for Epi Info (SAFE)*, a set of application development guidelines to improve the software design and modularity of public health information systems developed using components from the Epi Info tool provided by the Centers for Disease Control (CDC) [80].

Web-based systems, or systems having some web accessible components, were the delivery platform of choice in many cases [28,29,46,70,72,74,77–79,81–86]. These were often intended to permit distributed access by public health staff, reduce software implementation costs, or expose public health information for public dissemination. As such security and privacy was a frequently noted concern. Although privacy of health data was mentioned as a concern in many studies, only one article specifically

discussed implementation of security protocols [84]. Others discussed methods for aggregating or otherwise de-identifying data [74,79].

Total data volume, size of data transfer packets, or processing complexity in time or space were cited in a few studies [28,29,82,87]. These articles suggested the use of data warehousing and caching as possible approaches to address processing time related issues, noting that it takes time to calculate statistical values for use in infectious disease mapping. Several studies also mentioned cost as a major factor affecting architectural component choices [28,29,79,81,82]. Presented solutions included using open source or free proprietary software, using free web resources like the Google Maps API [88], and building modular reusable components such as web services [28,29,79,80,82,89]. Overall, there appears to be a trend away from standalone visualization systems, and toward modular, service-oriented architectures and web-based user interfaces.

3.3. User preferences

User preferences highlight how users prefer to interact with a tool or system, and can provide insights into possible sources of usability issues or adoption barriers. Studies of academic researchers and public health professionals indicated a preference for tools that help users evaluate disparate and complex high-quality data [44,46,49,62,90,91], with the goal of improving comprehension and communication, as well as facilitating decision-making [19,44,46,49,62,66,67,90–92]. Additionally, participants in qualitative and quantitative studies emphasized the importance of user-friendly, reliable tools, with high-quality online documentation, and easy access to the source code [9,46,62,67,68,93]. Users in a variety of settings raised concerns regarding interoperability of new and existing tools, data sharing, and data confidentiality [46,49,66,68,93]. Additionally, analysis of a survey conducted by Bassil and Keller [67] indicated that users in academic settings are nearly twice as sensitive to the cost of a new tool as are users in industry. This finding is consistent with many studies exploring or advocating for open-source and web-based infectious disease visualization tools to overcoming cost and resource barriers [62,70,73,77,79,84–86]. Moreover, these preferences mirror key themes from Section 3.1, namely user trust, tool credibility and transparency.

A host of studies highlighted user preferences for data abstraction, each with the underlying theme of making complex data digestible and useful for users. Users expressed a strong interest in dynamic, interactive graphics that allow them to review their data at different levels (e.g. population or individual level) [19,44,46,66,67,69,90–92]. With such a function, users felt they could incrementally explore the data to evaluate both the big picture and the finer details. In addition, users valued common interface features such as zoom, pan, search, filter, save, undo, and work history [9,46,67–69,91,93]. Users also showed interest in high-quality automated layouts and customizable features (e.g. color, size, shape) to facilitate understanding of the data [67–69,90]. Furthermore, some users demonstrated high interest in tools with multiple views or panels, enabling them to review their data from different perspectives [44,46,67–69,91–95]. In concert, users preferred easy navigation between views and synchronized browsing (e.g. monitor the same variable across panels) [67,68]. The ability to layer data, particularly among GIS users, was a common request to facilitate understanding of interactions or risk factors that overlap with disease outcomes [9,46,49,93]. Overall, these preferences emphasize the importance of information discovery and synthesis through iterative data exploration.

Such preferences guide the development of infectious disease visualization tools, and can inform strategies for incorporating

the tools into routine practice. The corresponding features and functions have the potential to help users discover complex or hidden patterns [96].

3.4. Features of existing tools

Having identified common information needs, system architectures, and user preferences, the following subsections explore existing tools and applications in more depth as they pertain to GIS, molecular epidemiology, and social network analyses. Each section also provides examples of common representations of GIS, molecular epidemiology, and social network data, respectively.

3.4.1. GIS

The development of increasingly sophisticated geographic information systems (GIS) has provided a new set of tools for public health professionals to monitor and respond to health challenges. These systems can help pinpoint cases and exposures, identify spatial trends, identify disease clusters, correlate different sets of spatial data, and test statistical hypotheses. Often, these analyses are aided by visualization and mapping of data, provided via web services or a user interface. Our review identified many approaches to delivering GIS functions based on various sources of public health data. Common functions among these studies and systems were geocoding [8,72,73,79,97], integrating data sources [72,73,98,99], and cluster detection [84,97]. Mapping of data was commonly achieved through dot maps (Fig. 3A) [8,46,72,73,75,76,81,84–86,96,99–106], choropleth maps (Fig. 3B) [8,44,71,75,78,95,96,99–102,104,105], and isopleth or gradient maps (Fig. 3C) [8,76,81,87,107]. Recurrent considerations cited within these papers included the privacy of public health data [29,75,79,82,101], the alignment of GIS analytics to users' needs [76,96,101,103,105,108], the motivations to make analysis services accessible, and the interoperability of data or system [8,29,79,101]. Since many GIS analytical services and geographic data are available through providers such as ESRI, Google, or the U.S. Census [88,109], GIS systems in our review often utilize an architecture based on these services and map data.

The systems reviewed were designed with various targeted users in mind. Two broad divisions of these were systems intended for public access using publicly available data, and restricted systems intended for users with access to private public health data. In many cases, these systems cited the use of publicly available maps and cartographic data as a basis for spatial integration of other information [72–74,98,101]. Many systems utilize administrative geographic units as a basis to merge data across different health and population databases, for example to calculate incidence rates based surveillance data and a population census. Other approaches may either map other sources into an internal data model [77] or to an ontology that supports data integration [29].

Visualization methods for GIS in public health in our review focused on functions geared toward simplifying, integrating, or analyzing data in a spatial context. The simplest visualizations plot or aggregate spatial data to deliver static point or choropleth maps of individual or aggregate data, respectively. Many systems incorporated a temporal component, enabling either animation of data through time or restriction of the data displayed to a time window of interest [77,79,84,110]. A step beyond mere display of information, some GIS or spatial statistical methods seek to perform kernel-based smoothing to estimate risk maps [87,97,107], visualize disease risk according to a statistical model [29,46,76,81,85,86,107,111,112], or compare one feature to another [71,84,97,100,101]. While the ability to zoom and pan to navigate maps [79,96,105] is a common interactive feature enjoyed by users, more advanced systems contain interactive controls to enable users to



Fig. 3. Common geographic (GIS) visualizations. A dot map (left) uses dots to represent a certain measure or feature displayed over a geographical map. They are often used to present the geographical distribution of various disease cases in infectious disease surveillance. This figure represents hypothetical infectious disease cases in the state of California. Each dot represents a specific disease case. These maps may help identify clusters of disease. In interactive tools, users may click individual cases or select subsets of cases to obtain further information. Individual level data is often aggregated in a choropleth map (middle), which uses graded colors or shades to indicate the values of some aggregate measure in specified areas. This figure shows the incidence rate per 100,000 persons of cases from map (left). Differences in the incidence rates by county are indicated with different shades, with a darker color indicating a higher rate. Interactive choropleth maps allow selection of regions to obtain additional information. Individual or aggregate level data may be used to statistically derive a spatial risk gradient (right). Other visualization features may allow zooming/panning of maps, introduction of other map layers such as roads, or selection of color scales.

retrieve information about selected items or regions, visualize the results of arbitrary queries [79], control visualization options, control temporal ranges of data returned [77,79], or link displays of data with alternate or comparative visualizations [78,79].

3.4.2. Molecular epidemiology

Molecular epidemiology is concerned with understanding the distribution or clustering of genetic variants, strains, serotypes, or other molecular groupings of pathogens. In molecular epidemiology, relationships between isolates are often calculated and conveyed through phylogenetic trees or *dendrograms* (Fig. 4). Visualization tools for molecular epidemiology often included phylogenetic analysis and visualization capabilities [113–115], and visualization of contextual data using connected graphs [70,113]. The tools we reviewed were primarily designed to be accessed through the internet [70,113,114,116]. Most studies in our review included the capability to integrate GIS or location-based data with genetic or serotype visualizations [11,70,84,113–117]. Two of the tools were designed to produce visualization (KML) files for display in other GIS packages [11,115], while other web based tools made use of external GIS services embedded within the website, primarily Google maps, ESRI/ArcGIS or HealthMap [70,116,117].

Some tools were designed with specific organisms in mind, for example staphylococcal [117] or influenza [116] infections. Driscoll et al. [70] developed Disease View, a set of tools to understand host-pathogen molecular epidemiology. They demonstrated the use of this tool to analyze aspects of the *Vibrio cholera* outbreak that occurred in the aftermath of the 2010 Haiti earthquake. These tools allow spatial views of molecular epidemiological properties associated with outbreaks, for example showing sequence variation of genes associated with disease virulence between outbreak locations. Other tools were designed to accommodate multiple organisms or user-specified organisms [113–115]. One such tool, designed specifically for geospatial surveillance of genomic characteristics of NIAID category A–C viral and bacterial pathogens, is GeMIna [11]. This tool collects curated metadata relating to the diseases. Other views of the distribution of genotypes across a large geographic scale help to understand the relationship between the population biology and geography of a pathogen species [118]. This is sometimes known as *phylogeography*.

As with GIS systems, data integration was a key component of the web based tools, with all web based tools incorporating access to or prepopulated with existing sets of data or meta-data, including pathogen, isolate and sequence data. Several studies discussed approaches for integration of genetic and social network data

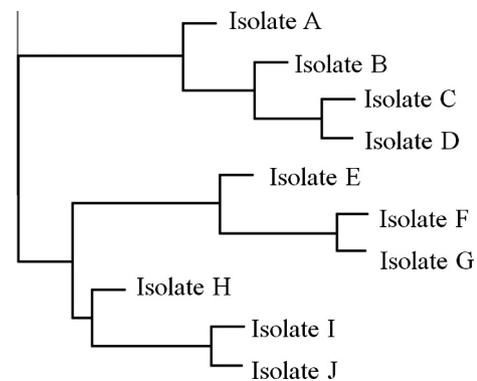


Fig. 4. Dendrogram. A dendrogram, or phylogenetic tree, is a branching diagram or “tree” showing the evolutionary history between biological species or other entities based on their genetic characteristics. Species or entities joined together by nodes represent descendants from a common ancestor and are more similar genetically. This figure shows a hypothetical example of a rooted dendrogram, wherein the horizontal position of individuals represents the genetic distance from a specific progenitor. With the advancement of DNA sequencing technologies, phylogenetic trees have been used widely in infectious disease control to depict the genetic similarities and differences between strains and variants of a certain disease pathogen. Knowing whether infectious diseases occurring in different areas are from the same strain provides key information on the source of infection and how the disease may be transmitted. Interactive features of these visualizations may include the ability to collapse or color and label branches.

[35,36,38,89,119]. In the absence of known exposures between cases, or in the case of ineffective contact investigations, molecular epidemiology or genomic approaches can identify potential members of an outbreak cluster. These studies showed social network data alongside genetic data using custom visualizations, but tools with the capacity to visualize the interplay of these data types systematically are still being developed.

3.4.3. Social network analysis

In addition to geographic and molecular epidemiologic data, networks of social contact or disease exposure are a third type of complex data that are increasingly being used to understand disease outbreaks. As shown in Fig. 1, social network analysis as a field is growing relative to health literature as a whole; however, it is at an earlier stage than for the other two topics. In order to describe the use of social network visualizations for public health, we therefore considered a broader set of publications that often

described visualizations of single outbreaks or analyses, in addition to those directly describing tools used to visualize outbreak networks. Nevertheless, these publications inform desiderata for visualizations of these networks, which in turn inform the features or design requirements such systems should consider. Applications of social network analysis in public health typically focus on routes of infection in communicable disease contact investigation; hence, most of the publications in our survey address this topic.

Although only eight articles were directly pertinent to social network analysis, these papers did address a variety of uses of and challenges for the application of network analysis for infectious disease control. Common purposes of network analysis included risk stratification of contacts, identifying common characteristics of those infected, visually communicating or mapping cases for improved understanding of outbreaks, and identifying potential pathways of transmission [24,38,120]. Among the considerations for data visualization addressed by these studies, several common features were observed: use of shape, color, and graph position to convey information [24,38,89,120,121]; display of individual case features or identity; and identification of important clusters or paths in the network [24,89,120]. In more advanced analyses, studies may seek to compare or estimate networks across other variables, such as including a temporal dimension in the study [38,83,89,120,121] (Fig. 5); integrating geographic or location features [83,89,110,120,121], or identifying exposures via molecular epidemiology as discussed in the previous section.

Consistent with our findings in other sections of the study, other important considerations recognized within the network analysis studies focused on the importance of designing network visualizations that provide the right information to users without confusing them. These considerations took the form of discussions about information overload from complex graphs [120], the inclusion of diverse user preferences for visualization [120,121], and the importance of training to help users understand and utilize these graphics [24,89]. Viegas and Donath [121] and Hansen et al. [120] studied non-standard network layouts, and included user assessments to help evaluate how these could best be used. Although most publications discussed the use of networks in a disease control context, Andre et al. [24], Cook et al. [89], and McElroy et al. [38] explicitly described how network visualizations could be used to aid decision-making via prioritization of resources or investigations. Other less common considerations for network analyses described in our review include the use of repeated contacts as a heuristic for risk, studies of population mixing [24], the use of touch-screen interfaces to navigate networks [120], the

importance of aggregated data visualization options to prevent information overload [120], and the use of simulation to augment missing data [83].

Mostly absent from these studies were visualization methods to help users understand network structures at an aggregate or summarized level, comparable to the choropleth map in GIS. Although visualizations like collapsed nodes, flow diagrams, and network metric distributions (such as node degree distribution) [122] have been used in other domains, these techniques may not yet be familiar interfaces for lay users, and hence have not been widely employed in tools for public health. As network data become increasingly integrated with GIS, molecular epidemiology, and other health indicators, evaluation of more diverse methods of network visualization consistent with end-user preferences, training level, statistical literacy, and cognitive ability will be needed.

3.5. Usability and evaluation

The usefulness of a system is often used to describe a system's overall effectiveness. The concept of usefulness can be measured as a combination of utility and usability. Traditional system evaluation has focused on utility, determining whether an information system is able to meet the functional requirements of a user who wants to accomplish a specific set of work tasks. This is demonstrated in studies which evaluate information systems based on a strict set of functional metrics, such as accuracy and efficiency [77].

In addition to evaluating system functionality, it is becoming increasingly important to evaluate system usability. Some researchers have conducted usability evaluations to provide justification for the time and effort spent developing and deploying these complex tools [123]. In addition, the intended benefit of many information systems is to facilitate interaction between users and data, and so usability itself is the primary measure of system usefulness [124]. However, the articles we examined have revealed that features which improve the usability of one system cannot always be generalized to other systems, since different users may have different task-specific system requirements [62,93].

Even though specific design recommendations may not apply broadly across systems, studies cited common methods to reliably evaluate the usability of a system. These methods include the use of qualitative investigation techniques, such as participant observations, interviews, and workflow analysis [51,125,126]. Participant observations involve watching users as they perform their work, during which researchers have encouraged users to “talk

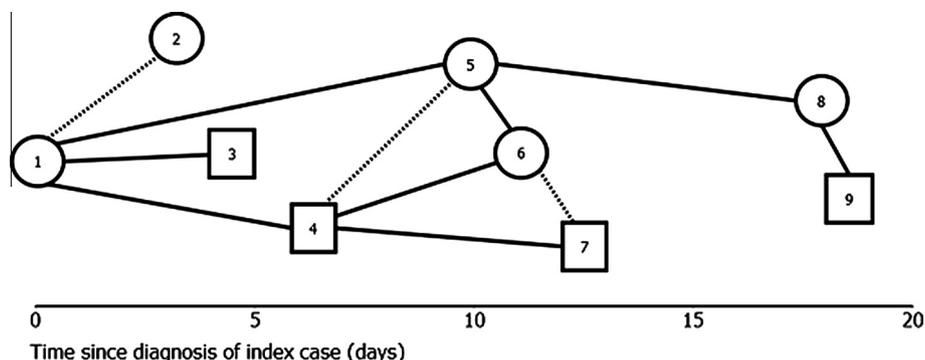


Fig. 5. Social network diagram. A social network is a graphical representation of social relations or exposures consisting of nodes (individuals within the network) and ties (relationships between individuals). Nodes are usually represented as points or other shapes while ties are represented by lines between the nodes. Differences in the shapes or lines of the diagram may be used to represent different characteristics of the individuals or the relationships. This figure shows a hypothetical example of a force-directed social networks diagram. Social networks analyses in infectious disease control have been gaining importance in the past decade. Examining these social relationships between disease cases and their secondary contacts may be beneficial to tracking the spread of infectious diseases within interconnected social networks. It is especially useful in identifying the index or source case and predicting which individuals are more likely to become infected and further infect others.

aloud” during interactions with information systems. These observations have served as the basis for semi-structured interviews and focus groups used to obtain in-depth descriptions of user behavior [51,125,126]. Published studies also describe the use of interviews to highlight areas for further investigation, either by pinpointing particular aspects that a user does not like, or by uncovering new interactions that a user would like to see added [125]. In addition, observations and interviews have been combined with questionnaires containing Likert scale questions, asking users to rate their satisfaction with information systems [46].

However, researchers acknowledged that efforts to simplify interactions between users and data may have the unintended consequence of limiting functionality [62]. For this reason, some researchers found it important to engage users in the design and development processes. This was accomplished by employing usability evaluation techniques in conjunction with participatory design methods, allowing feedback to be incorporated into the system throughout the development process [8,70,125,126]. Researchers also expressed interest in studying user work behaviors over longer time periods [93], an aspect which might be addressed by soliciting feedback during an ongoing participatory design process [91,124].

3.6. Implementation and adoption

Barriers to adoption vary widely and are not mutually exclusive within a given organization or individual. System-level barriers, such as access issues (e.g. lack of internet or finances) and lack of organizational support were significant barriers in organizations worldwide [8,9,47,49,57,58,79,99,101,127,128]. Jurisdictions often struggle to share data due to lack of data standardization (e.g. data heterogeneity, missing data, lack of interoperability) and face data confidentiality concerns which collectively compound the already-complex task of monitoring diseases [8,9,29,46,49,73,99,127,128]. Furthermore, user-level concerns may also result in adoption barriers. Confusion regarding how to create or use effective graphics, and a lack of familiarity with the concepts in the tool could be substantial learning barriers [47,58,99]. Fear of change and an interest in staying within one’s comfort zone, in addition to a lack of trust and misconceptions about the use of the tool, may also prevent adoption of a valuable tool [47,49,68]. Indeed, studies indicated that many users relied on other tools (e.g. Microsoft Office suite) because they felt that many existing tools were too complex and had a substantial learning curve [8,9,47,68,79,85,96,99,128].

Despite the potential for data visualization tools to monitor and aid control efforts for infectious diseases, such tools have had only limited adoption [49,97] and only one system was assessed for distribution [129]. Usability studies and implementation projects are remarkably interdependent, as successful adoption often requires developers to re-design elements of the tool to further address the users’ needs [8,70,125]. The resulting iterative design process often helps users identify previously unexpressed or unknown information needs [93], resulting in the need for subsequent usability studies. However, this process can be time consuming, and users may find alternative systems that meet their current needs before the tool is completed [8,127]. Moreover, existing tools are largely isolated to the jurisdictions and organizations that developed them and may be based on proprietary systems [8,46]. Such silos could prevent the widespread adoption of tools by other agencies or organizations.

While the specifics of adoption strategies may vary depending on the particular organization or agency and their needs, some common strategies emerged from the literature review. Several studies recommended ongoing user collaboration with the tool developers to ensure that the users’ needs were heard early on in the project, and to create the opportunity for regular feedback

[8,9,68,79,125,127]. Further, studies advocated for open source tools to reduce access barriers, particularly in low-resource settings [71,79,99,127]. Integrating the tool into existing workflow was also recommended as a strategy to encourage users to regularly utilize the tool [8,47,51,68,99,124]. Additionally, providing adequate user training and education, as well as ongoing technical support, for staff was considered essential for successful adoption of a novel tool in many studies [8,46,47,49,62,79,80,92,96,99,101,124,127,128]; effective user training may build the users’ self-confidence in the use of the tool and encourage them to try the tool [125,127]. In concert, these strategies may create an environment for sustained use.

4. Discussion and conclusion

In this review, we assessed the current landscape of visualization tools developed for infectious disease epidemiology. We characterized these tools in terms of information needs and user preferences, features and system architectures of existing tools, as well as usability and adoption considerations. By focusing on visualizations of GIS, molecular epidemiological, and social network data, we also explored similarities among these three types of increasingly common data types. The richness of the information offered by these data for communication and decision making are counterbalanced by difficulties in displaying, interpreting, and trusting these data sources. In our review of tools throughout their lifecycle from conception to development to adoption, several themes and challenges emerged pertaining to both individual stages as well as broader topics. Despite the different scholarly approaches of the included articles, the following themes emerged: (1) importance of knowledge regarding user needs and preferences; (2) importance of user training; integration of the tool into routine work practices; (3) complications associated with understanding and use of visualizations; and (4) the role of user trust and organizational support in the ultimate usability and uptake of these tools. Another broader theme that became apparent is that individual tools and datasets are rarely sufficient, even for local decision making. Therefore, interoperability of tools and the importance of data sharing and integration were important goals that should factor into the design of visualization tools.

The utility of visualization tools is constrained by the extent to which they address the information needs of users. Information needs are as complex and varied as the tasks performed by public health professionals. Consequently, developing information visualization tools to meet these needs is correspondingly complex. Indeed, developers have addressed information needs in a multitude of ways, resulting in the current diversity of data visualization tools, each serving as a case study for one approach to resolve these needs. Regardless of the study population, users indicated that they needed timely access to reliable, high-quality information to perform their duties. Efforts to map users’ queries of common data types (e.g. GIS, molecular epidemiology, and social networks) to meaningful visualizations have raised concerns regarding the potential for misinterpretation and cognitive overload due to the complexity of infectious disease data [130].

Despite results from studies with users emphasizing the value of dynamic, interactive graphics to facilitate data exploration and abstraction, existing tools are largely still static. And while static graphics are extremely useful, pairing them with interactive features may give users more freedom to explore and learn from their data. Sophisticated data analysis and visualization systems, such as R [131], SAS [132] and Matlab [133] have traditionally enabled expert users to create hard coded (but rapidly adjustable) graphics using code. The increasing use of these platforms to create user-friendly, interactive, web-based versions of these visualizations

through technologies such as scalable vector graphics (SVG), dynamic HTML (DHTML), and Shiny [134] has the potential to greatly simplify users' access to interactive, web-based visualizations. The distinction between visualization tools requiring coding and online visualization tools is also somewhat blurred by the ability to embed fully functional data analysis and visualization within web applications, as has been done using RStudio [135] to allow the use of R within the Centers for Disease Control and Prevention's BioSense surveillance system [136].

In addition to BioSense, several well-known surveillance systems are not included in this review, including the Centers for Disease Control's EARS (Early Aberration Reporting System) and Johns Hopkins's ESSENCE (Electronic Surveillance System for the Early Notification of Community-based Epidemics). These systems have a limited representation in scholarly literature (as they are commonly developed and evaluated internally) and are not discussed in terms of visualization features. For example, a published evaluation of the EARS system focused chiefly on its aberration detection algorithms [137]. Consequently, they were not captured in the scope of this review. However, these systems face many of the same constraints as those discussed in this review: data standardization and quality in diverse jurisdictions, limitations of user knowledge and organizational capacity to implement the tool, as well as generation of accurate and easy-to-understand visualizations.

Visualizations with interactive features or sophisticated visual elements may require sufficient rendering capability and user experience to maximize their potential. For example, to access an area of interest in a 3D representation, users will typically need to adjust other visual cues (e.g. rotate the graphic, change transparency or depth queuing) [69]. Koenig et al. [92] explored visual perceptions among public health users in GIS environments and demonstrated a preference for a blue and red color scheme to represent health and morbidity, respectively. However, studies emphasized that color schemes and visual elements should be sensitive to multi-cultural users, users with color-blindness, and rendering limitations of existing systems [46,67,92]. These visual elements also contribute to data (mis)interpretation. Consequently, guidelines have been proposed for color schemes and visual elements to minimize the risk of misinterpreting the data. For example, use of single-hue color progression (e.g. white to dark blue) to show sequential data is more intuitive than spectral schemes (e.g. rainbow) that force users to assign arbitrary magnitudes to rainbow colors [138].

Together with utility (functional effectiveness), usability (perceived ease of use) is sometimes considered to be a core component of determining the overall usefulness of a system. This makes usability one of the dimensions that can contribute to the adoption of a new information system [139]. Usability has been assessed by examining several dimensions including learnability, memorability, error prevention/recovery, efficiency, and user satisfaction [140]. However, usability also varies depending on the specific information needs of an individual user, particularly because efficiency depends on the task being performed. This presents an interesting problem when trying to highlight best practices with regards to usability. After a system has been developed, usability evaluation techniques can be used to assess its overall usability. The evaluation can contain quantitative assessment of accuracy and time efficiency as compared to a previous system or suitable alternative, such as a spreadsheet or database. With a sufficient pool of users and clearly defined metrics, a usability evaluation can yield statistically significant results, although this is not necessarily meaningful when assessing qualitative aspects, such as user satisfaction and perceived learnability.

Further, there was little discussion in the included literature about how to organize and sustain the implementation phase.

The literature highlighted minimal success in widespread implementation and adoption of data visualization tools. While substantial barriers exist, there are strategies to address many of them, including obtaining management support, providing ongoing user training support, and starting a pilot program to integrate the tool into existing workflow. However, with extensive variability in data management systems, needs, and attitudes, widespread adoption of a given tool is difficult task. For example, integrating the novel tool into a given workflow requires collaboration between agencies and organizations, qualified staff for observation and interview studies, and time. Due to the variability of site structure, and thus workflow, the optimal implementation strategy may vary, limiting the desired widespread adoption. Consequently, implementation becomes a site-specific endeavor, rather than a one-size-fits all task. The participatory design approach can increase the amount of exposure that users have to the system, allowing for a better approximation of usage habits over time and understanding of the users' needs. Obtaining management support and creating a pilot implementation project may benefit from theory-driven communication campaigns to raise interest and support. For example, the literature supports a highly variable knowledge of and support for data visualization tools among management and staff. Behavior change theories, such as the Stages of Change Model [141] or the Diffusion of Innovations [142], may improve adoption rates by targeting messages to different populations based on their readiness and interest in adopting the novel tool.

Many studies highlighted the importance of adequate and ongoing training for users, providing a possible avenue to explore in more depth to minimize the risk of misinterpretation as well as improve adoption. In a recent study, more than half of the participating public health professionals indicated they were likely to seek training in a variety of tasks, including data visualization, epidemic modeling, GIS, cluster analysis and statistical modeling [62]. They also preferred a variety of training styles (e.g. task-oriented tutorials, user guides, and hands-on training). Such training opportunities may also improve the perceived transparency of the tool. Further, integrating user training time, cost of the tool, and support staff into site budgets may also encourage more consistent, trained use of the tool. An atmosphere supporting regular use of the tool can encourage users to spend more time learning about the tool's features and functions while helping them become more savvy, creative, and comfortable with the tool [96,125]. However, few of these studies addressed the growing need for enhanced statistical education to enable users to better understand their data in more depth. For many non-expert users, a trade-off is often made in favor of easy-to-use, "black box" programs instead of in-depth understanding of the limitations of data analysis. The desire for a system that allows users to query the data and receive results in plain language may undermine the very nature of complex data. Future research should endeavor to help users strike a balance between in-depth understanding of data and system usability.

Lastly, pragmatic constraints of widespread tool adoption, including funding considerations, jurisdictional constraints, as well as data sharing and confidentiality concerns, may prove more difficult to overcome. Public health organizations worldwide face technological and financial access barriers preventing them maximizing the potential of visualization tools for epidemiology. Finite funding streams often force organizations to adapt existing systems that may not best serve their needs. Further, jurisdictional constraints and data sharing concerns create information silos, leading to reduced data potential. Infectious diseases do not follow jurisdictional boundaries, and new policies are needed to increase secure data sharing across organizations to facilitate decision making and improve distribution of resources. Public health organizations need more funding to explore customizable visualization tools for infectious disease that include public health users

throughout development. Such work would further inform best practices for visualization tools of the complex data types public health professionals are expected to synthesize and act upon.

4.1. Future directions

As data types and sources become increasingly large and complex, so too should the strategies to integrate disparate and often incomplete data into novel visualization tools. Concerns regarding data quality and accuracy are particularly relevant for visualization tools as these tools can be limited by the inputted data. Discussions of current data limitations highlight issues of scale and uncertainty, accuracy of datasets for spatial and epidemic models used in tools, and the impact of residential address errors in geocoding, to name a few [143–147]. In order to draw meaningful and accurate conclusions from the data, visualization tools should represent missingness and uncertainty clearly. For instance, a recent study demonstrated that participants interpreting graphics with missing data tended to misinterpret results, but with equal confidence in their interpretations as those viewing more complete graphics [148]. Similarly, geographic analyses are known to be sensitive to overestimation of rates in small populations, which often correspond to large, sparsely populated regions, resulting in visual biases in interpreting choropleth maps [130]. These results suggest that users may not be aware of the need for better representation of missingness and uncertainty, and studies to evaluate the best means of doing so are still in their infancy. Continuing research on visualization algorithms that account for missing and uncertain data is needed to overcome these hurdles.

Another important challenge for future developers of information visualization tools for public health is to focus not only on individual user needs and comprehension of graphics, but also to plan and develop these tools in the broader contexts of available data, existing algorithms/services, team collaboration, and inter-organizational and interdisciplinary needs. Too many software projects are developed as new information silos, resulting in redundancy of effort, failure to integrate data and tools, and challenges to training and adoption. Further, many existing systems (e.g. BioSense) are access-restricted, limiting their use in infectious disease epidemiology, and may not have completed (or shared) evaluations of their visualization features. Visualization tools of the future should be developed to be compatible with existing data formats and standards, and interoperable with each other. Future tools should also adapt to the increasing pressure to be open-access, allowing users from low-resource settings, academia, and industry to capitalize on the advances in surveillance and visualization technology. This level of interoperability could support more advanced features such as phylogeography (the study of genetic variation across geographic space), inference of person-to-person contact from molecular epidemiology, statistical cluster detection based on joint spatiotemporal and genomic data, integration of remote sensing and environmental data, and other tasks as users become increasingly savvy in their use of advanced analytical and visualization tools for public health.

4.2. Limitations

Although this systematic review covers a wide range of visualization tools for infectious diseases, there are three main limitations. First, while the scope and search terms of the review were purposefully broad, we likely missed relevant articles. Second, many public health informatics tools, if described in manuscripts at all, may be published in non-indexed conference proceedings, and thus more recent or undersold tools may not have been retrieved. Further, systems or informatics needs assessments that have no associated publications on their visualization features

(e.g. developed and used in practice only) were not readily available for our study, and systems with access-controlled content could not be assessed in context with the other tools identified here. For example, the Centers for Disease Control and Prevention's Public Health Information Network (PHIN) and BioSense as well as the International Society for Disease Surveillance have non-indexed content that the review did not capture. Lastly, we focused on English articles for practical reasons, but by doing so we may have excluded valuable contributions from teams around the world. However, our review included English articles in journals worldwide.

Acknowledgments

The authors gratefully acknowledge Margo W. Bergman, Ph.D, MPH and Kyle M. Jacoby, Ph.D for thoughtful review and comment. Research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health under award number R01LM011180. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jbi.2014.04.006>.

References

- [1] Bos L, Blobel B. Semantic interoperability between clinical and public health information systems for improving public health services. *Med Care Comput Informatics* 2007;4:256.
- [2] Friede A, Reid JA, Ory HW. CDC WONDER: a comprehensive on-line public health information system of the Centers for Disease Control and Prevention. *Am J Pub Health* 1993;83(9):1289–94.
- [3] Friede A, Blum HL, McDonald M. Public health informatics: how information-age technology can strengthen public health. *Annu Rev Pub Health* 1995;16(1):239–52.
- [4] Baker EL et al. CDC's Information Network for Public Health Officials (INPHO): a framework for integrated public health information and practice. *J Pub Health Manage Pract* 1995;1(1):43–7.
- [5] Victor LY, Edberg SC. Global Infectious Diseases and Epidemiology Network (GIDEON): a world wide Web-based program for diagnosis and informatics in infectious diseases. *Clin Infect Dis* 2005;40(1):123–6.
- [6] Reeder B et al. Public health practice within a health information exchange: information needs and barriers to disease surveillance. *Online J Pub Health Inform* 2012;4(3).
- [7] Khan AS et al. The next public health revolution: public health information fusion and social networks. *Am J Pub Health* 2010;100(7):1237–42.
- [8] Driedger SM et al. Correction: using participatory design to develop (public) health decision support systems through GIS. *Int J Health Geogr* 2007;6.
- [9] Kothari A et al. Mapping as a knowledge translation tool for Ontario Early Years Centres: views from data analysts and managers. *Implement Sci* 2008;3:4.
- [10] Robertson C, Nelson TA. Review of software for space-time disease surveillance. *Int J Health Geogr* 2010;9:16.
- [11] Schriml LM et al. GeMInA, genomic metadata for infectious agents, a geospatial surveillance pathogen database. *Nucl Acids Res* 2010;38(Suppl. 1):D754–64.
- [12] Lopes CT et al. Cytoscape Web: an interactive web-based network browser. *Bioinformatics* 2010;26(18):2347–8.
- [13] Heymann DL, Brilliant L. Surveillance in eradication and elimination of infectious diseases: a progression through the years. *Vaccine* 2011;29(Suppl. 4):D141–4.
- [14] Klompas M et al. Harnessing electronic health records for public health surveillance. *Online J Pub Health Inform* 2011;3(3).
- [15] Hills RA, Lober WB, Painter IS. Biosurveillance, case reporting, and decision support: public health interactions with a health information exchange. In: *Biosurveillance and biosecurity*. Springer; 2008. p. 10–21.
- [16] Reis B, Kohane I, Mandl K. An epidemiological network model for disease outbreak detection. *PLoS Med* 2007;4(6).
- [17] Chen H, Zeng D, Yan P. Data visualization, information dissemination, and alerting. In: *Infectious disease informatics*. Springer; 2010. p. 73–87.
- [18] Koch T, Denike K. Crediting his critics' concerns: Remaking John Snow's map of Broad Street cholera, 1854. *Soc Sci Med* 2009;69(8):1246–51.
- [19] Karlsson D et al. Visualization of infectious disease outbreaks in routine practice. *Stud Health Technol Inform* 2013;192:697–701.

- [20] Hay SI et al. Global mapping of infectious disease. *Philos Trans Roy Soc B Biol Sci* 2013;368(1614).
- [21] Thacker SB, Qualters JR, Lee LM. Public health surveillance in the United States: evolution and challenges. *MMWR* 2012;61:3–9.
- [22] Eubank S. Network based models of infectious disease spread. *Jpn J Infect Dis* 2005;58(6):S.
- [23] Holmes EC. Molecular epidemiology and evolution of emerging infectious diseases. *Brit Med Bull* 1998;54(3):533–43.
- [24] Andre M et al. Transmission network analysis to complement routine tuberculosis contact investigations. *Am J Pub Health* 2007;97(3):470–7.
- [25] Abernethy NF. Automating social network models for tuberculosis contact investigation. Stanford University; 2005.
- [26] AvRuskin GA et al. Visualization and exploratory analysis of epidemiologic data using a novel space time information system. *Int J Health Geogr* 2004;3(1):26.
- [27] Castronovo DA, Chui KKH, Naumova EN. Dynamic maps: a visual-analytic methodology for exploring spatio-temporal disease patterns. *Environ Health* 2009;8.
- [28] Dominkovics P et al. Development of spatial density maps based on geoprocessing web services: application to tuberculosis incidence in Barcelona, Spain. *Int J Health Geogr* 2011;10.
- [29] Gao S et al. Online GIS services for mapping and sharing disease information. *Int J Health Geogr* 2008;7.
- [30] Carpenter TE. The spatial epidemiologic (r)evolution: a look back in time and forward to the future. *Spat Spatiotemporal Epidemiol* 2011;2(3):119–24.
- [31] Foxman B, Riley L. Molecular epidemiology: focus on infection. *Am J Epidemiol* 2001;153(12):1135–41.
- [32] Maccannell D. Bacterial strain typing. *Clin Lab Med* 2013;33(3):629–50.
- [33] Arts IC, Weijnenberg MP. New training tools for new epidemiologists. *Environ Mol Mutagen* 2013;54(7):611–5.
- [34] Carrico JA et al. Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution. *Eurosurveillance* 2013;18(4):32–40.
- [35] Hollm-Delgado M-G. Molecular epidemiology of tuberculosis transmission: contextualizing the evidence through social network theory. *Soc Sci Med* 2009;69(5):747–53.
- [36] Lowy FD, Miller M. New methods to investigate infectious disease transmission and pathogenesis—*Staphylococcus aureus* disease in drug users. *Lancet Infect Dis* 2002;2(10):605–12.
- [37] Fitzpatrick LK et al. A preventable outbreak of tuberculosis investigated through an intricate social network. *Clin Infect Dis* 2001;33(11):1801–6.
- [38] McElroy P et al. A network-informed approach to investigating tuberculosis outbreak: implications for enhancing contact investigations. *Int J Tuberc Lung Disease* 2003;7(12 Suppl. 3). p. S486–93.
- [39] Dewan PK et al. A tuberculosis outbreak in a private-home family child care center in San Francisco, 2002 to 2004. *Pediatrics* 2006;117(3):863–9.
- [40] Morris M, Goodreau S, Moody J. Sexual networks, concurrency, and STD/HIV. *Sexy transmitted diseases*. McGraw-Hill: New York; 2007.
- [41] Polgreen PM et al. Prioritizing healthcare worker vaccinations on the basis of social network analysis. *Infect Control Hosp Epidemiol Off J Soc Hosp Epidemiol Am* 2010;31(9):893.
- [42] Basta NE et al. Strategies for pandemic and seasonal influenza vaccination of schoolchildren in the United States. *Am J Epidemiol* 2009;170(6):679–86.
- [43] O'Carroll PW et al. Information needs in public health and health policy: results of recent studies. *J Urban Health-Bull N Y Acad Med* 1998;75(4):785–93.
- [44] Gesteland PH et al. The EpiCanvas infectious disease weather map: an interactive visual exploration of temporal and spatial correlations. *J Am Med Inf Assoc* 2012;19(6):954–9.
- [45] Bishr Y. Overcoming the semantic and other barriers to GIS interoperability. *Int J Geogr Inf Sci* 1998;12(4):299–314.
- [46] Hu PJH et al. System for infectious disease information sharing and analysis: design and evaluation. *IEEE Trans Inf Technol Biomed* 2007;11(4):483–92.
- [47] Folorunso O, Ogunseye O. Challenges in the adoption of visualization system: a survey. *Kybernetes* 2008;37(9/10):1530–41.
- [48] Lê M-L. Information needs of public health staff in a knowledge translation setting in Canada 1. *J Can Health Libr Assoc* 2013;34(01):3–11.
- [49] Joyce K. "To me it's just another tool to help understand the evidence": public health decision-makers' perceptions of the value of geographical information systems (GIS). *Health Place* 2009;15(3):801–10.
- [50] Fuller S. Tracking the global express: new tools addressing disease threats across the world. *Epidemiology* 2010;21(6):769–71. 10.1097/EDE.0b013e3181f56757.
- [51] Pina J et al. Task analysis in action: the role of information systems in communicable disease reporting. *AMIA Annu Symp Proc* 2009;2009:531–5.
- [52] Gorman S. How can we improve global infectious disease surveillance and prevent the next outbreak? *Scand J Infect Dis* 2013.
- [53] Rolka H et al. Analytical challenges for emerging public health surveillance. *CDC's Vision Pub Health Surveillance 21st Century* 2012;61:35.
- [54] Moher D et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Int Med* 2009;151(4):264–9. w64.
- [55] Revere D et al. Understanding the information needs of public health practitioners: a literature review to inform design of an interactive digital knowledge management system. *J Biomed Inf* 2007;40(4):410–21.
- [56] Humphreys BL. Meeting information needs in health policy and public health: priorities for the National Library of Medicine and the National Network of Libraries of Medicine. *J Urban Health-Bull N Y Acad Med* 1998;75(4):878–83.
- [57] Lee P, Giuse NB, Sathe NA. Benchmarking information needs and use in the Tennessee public health community. *J Med Libr Assoc* 2003;91(3):322–36.
- [58] Turner AM et al. From the ground up: nurses information needs in a public health department in Oregon. *J Med Libr Assoc* 2008;96(4):335–42.
- [59] LaPelle NR et al. Identifying strategies to improve access to credible and relevant information for public health professionals: a qualitative study. *Bmc Public Health* 2006;6.
- [60] Walton LJ et al. Outreach to public health professionals: lessons learned from a collaborative Iowa public health project. *Bull Med Libr Assoc* 2000;88(2):165–71.
- [61] Centers for Disease and Prevention. Information needs and uses of the public health workforce—Washington, 1997–1998. In: *MMWR morb mortal wkly rep*, vol. 49(6); 2000. p. 118–20.
- [62] Robinson AC, MacEachren AM, Roth RE. Designing a web-based learning portal for geographic visualization and analysis in public health. *Health Inf J* 2011;17(3):191–208.
- [63] Ford J, Korjonen H. Information needs of public health practitioners: a review of the literature. *Health Inf Libr J* 2012;29(4):260–73.
- [64] Twose C et al. Public health practitioners' information access and use patterns in the Maryland (USA) public health departments of Anne Arundel and Wicomico Counties. *Health Inf Libr J* 2008;25(1):13–22.
- [65] Fourie I. Learning from research on the information behaviour of healthcare professionals: a review of the literature 2004–2008 with a focus on emotion. *Health Inf Libr J* 2009;26(3):171–86.
- [66] Robinson A. Needs assessment for the design of information synthesis visual analytics tools. *IEEE Inf Vis* 2009;353–60.
- [67] Bassil S, Keller R. Software visualization tools: survey and analysis. In: *Program comprehension, 2001. IWPC 2001. Proceedings. 9th international workshop on IEEE*; 2001.
- [68] Kienle H, Müller H. Requirements of software visualization tools: a literature survey. In: *4th IEEE international workshop on visualizing software for understanding and analysis (VISSOFT)*. IEEE; 2007.
- [69] Herman I, Melançon G, Marshall M. Graph visualization and navigation in information visualization: a survey. *IEEE T Vis Comput Gr* 2000;6(1):24–43.
- [70] Driscoll T et al. Integration and visualization of host-pathogen data related to infectious diseases. *Bioinformatics* 2011;27(16):2279–87.
- [71] Anselin L, Syabri I, Kho Y. GeoDa: an introduction to spatial data analysis. *Geogr Anal* 2006;38(1):5–22.
- [72] Blanton JD et al. Development of a GIS-based, real-time Internet mapping tool for rabies surveillance. *Int J Health Geogr* 2006;5:47.
- [73] Hurlimann E et al. Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. *PLoS Negl Trop Dis* 2011;5(12):e1404.
- [74] Sopan A et al. Community Health Map: a geospatial, and multivariate data visualization tool for public health datasets. *Gov Inf Q* 2012;29(2):223–34.
- [75] Green J, et al. Design and implementation of a geographic information system for the general practice sector in Victoria, Australia. In: *3rd National conference GIS in public health*; 1998.
- [76] Porcasi X et al. An operative dengue risk stratification system in Argentina based on geospatial technology. *Geospat Health* 2012;6(3):S31–42.
- [77] Freifeld CC et al. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. *J Am Med Inf Assoc* 2008;15(2):150–7.
- [78] Heitgerd JL et al. Community health status indicators: adding a geospatial component. *Prev Chronic Dis* 2008;5(3):A96.
- [79] Yi Q et al. Integrating open-source technologies to build low-cost information systems for improved access to public health data. *Int J Health Geogr* 2008;7.
- [80] Ma J et al. New frontiers for health information systems using Epi Info in developing countries: structured application framework for Epi Info (SAFE). *Int J Med Inform* 2008;77(4):219–25.
- [81] da Silva FA et al. IntegraEPI: a Grid-based epidemic surveillance system. *Stud Health Technol Inf* 2007;126:197–206.
- [82] Gao S et al. Towards Web-based representation and processing of health information. *Int J Health Geogr* 2009;8.
- [83] Ramirez-Ramirez LL et al. A new surveillance and spatio-temporal visualization tool SIMID: SIMulation of Infectious Diseases using random networks and GIS. *Comput Meth Prog Biomed* 2013;110(3):455–70.
- [84] Reinhardt M et al. EpiScanGIS: an online geographic surveillance system for meningococcal disease. *Int J Health Geogr* 2008;7.
- [85] Alonso WJ, McCormick BJ. EPIPO: a user-friendly analytical tool for the extraction and visualization of temporal parameters from epidemiological time series. *Bmc Public Health* 2012;12.
- [86] Lewis SL et al. SAGES: a suite of freely-available software tools for electronic disease surveillance in resource-limited settings. *Plos One* 2011;6(5).
- [87] Atkinson PJ, Unwin DJ. Density and local attribute estimation of an infectious disease using MapInfo. *Comput Geosci* 2002;28(9):1095–105.
- [88] Google. Google Maps API. Available from: <<https://www.developers.google.com/maps/>>, [cited 09.09.13].
- [89] Cook V et al. Transmission network analysis in tuberculosis contact investigations. *J Infect Dis* 2007;196(10):1517–27.
- [90] McGrath C, Krackhardt D, Blythe J. Visualizing complexity in networks: seeing both the forest and the trees. *Connections* 2003;25(1):37–47.
- [91] Shneiderman B, Plaisant C, Hesse B. Improving health and healthcare with interactive visualization methods. *Computer* 2013;46(5):58–66.

- [92] Koenig A, Samarasundera E, Cheng T. Interactive map communication: pilot study of the visual perceptions and preferences of public health practitioners. *Public Health* 2011;125(8):554–60.
- [93] Plaisant C. The challenge of information visualization evaluation. In: *Proceedings of the working conference on advanced visual interfaces*. ACM; 2004.
- [94] Chui KKH et al. Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs. *PLoS One* 2011;6(2).
- [95] Upadhyayula SM et al. Filaria monitoring visualization system: a geographical information system-based application to manage lymphatic filariasis in Andhra Pradesh, India. *Vector-Borne Zoonotic Dis* 2012;12(5):418–27.
- [96] Thew SL et al. Supporting creativity and appreciation of uncertainty in exploring geo-coded public health data. *Meth Inf Med* 2011;50(2):158–65.
- [97] Rushton G. Public health, GIS, and spatial analytic tools. *Ann Rev Pub Health* 2003;24:43–56.
- [98] Chen J, MacEachren AM. Resolution control for balancing overview and detail in multivariate spatial analysis. *Cartogr J* 2008;45(4):261–73.
- [99] Fisher RP, Myers BA. Free and simple GIS as appropriate for health mapping in a low resource setting: a case study in eastern Indonesia. *Int J Health Geogr* 2011;10.
- [100] Aimone AM, Perumal N, Cole DC. A systematic review of the application and utility of geographical information systems for exploring disease–disease relationships in paediatric global health research: the case of anaemia and malaria. *Int J Health Geogr* 2013;12:1.
- [101] Geanuracos CG et al. Use of geographic information systems for planning HIV prevention interventions for high-risk youths. *Am J Pub Health* 2007;97(11):1974–81.
- [102] Hadjichristodoulou C et al. Surveillance of brucellosis in a rural area of Greece: application of the computerized mapping programme. *Eur J Epidemiol* 1999;15(3):277–83.
- [103] Buckeridge D, Purdon L. The South East Toronto Urban Health Research Group. Health data mapping in Southeast Toronto: a collaborative project. In: *3rd National conference GIS in public health*; 1998.
- [104] Nobre FF et al. GISEpi: a simple geographical information system to support public health surveillance and epidemiological investigations. *Comput Meth Prog Biomed* 1997;53(1):33–45.
- [105] Joshi A et al. Designing human centered GeoVisualization application—the SanaViz—for telehealth users: a case study. *Technol Health Care* 2012;20(6):473–88.
- [106] ter Waarbeek H et al. Strengthening infectious disease surveillance in a Dutch–German crossborder area using a real-time information exchange system. *J Bus Contin Emer Plan* 2011;5(2):173–84.
- [107] Stevens KB, Pfeiffer DU. Spatial modelling of disease using data- and knowledge-driven approaches. *Spat Spatiotemporal Epidemiol* 2011;2(3):125–33.
- [108] Joshi A et al. A human centered GeoVisualization framework to facilitate visual exploration of telehealth data: a case study. *Technol Health Care* 2012;20(6):457–71.
- [109] ESRI. *ArcGIS Desktop*. Redlands, CA; 2013.
- [110] Benavides J et al. Smartphone technologies for social network data generation and infectious disease modeling. *J Med Biol Eng* 2012;32(4):235–44.
- [111] Maciejewski R et al. A pandemic influenza modeling and visualization tool. *J Vis Lang Comput* 2011;22(4):268–78.
- [112] Moore KM, Edge G, Kurc AR. Visualization techniques and graphical user interfaces in syndromic surveillance systems. Summary from the disease surveillance workshop, September 11–12, 2007, Bangkok, Thailand. *BMC Proc* 2008;2(Suppl. 3):S6.
- [113] Parks DH et al. GenGIS: a geospatial information system for genomic data. *Genome Res* 2009;19(10):1896–904.
- [114] Gopinath G et al. The Pathogen-annotated Tracking Resource Network (PATRN) system: a web-based resource to aid food safety, regulatory science, and investigations of foodborne pathogens and disease. *Food Microbiol* 2013;34(2):303–18.
- [115] Janies DA et al. The Supramap project: linking pathogen genomes with geography to fight emergent infectious diseases. *Cladistics* 2011;27(1):61–6.
- [116] Macdonald N, Parks D, Beiko R. SeqMonitor: influenza analysis pipeline and visualization. *PLoS Curr* 2009;1:Rrn 1040.
- [117] Grundmann H et al. Geographic distribution of *Staphylococcus aureus* causing invasive infections in Europe: a molecular-epidemiological analysis. *PLoS Med* 2010;7(1):e1000215.
- [118] He X et al. A comprehensive mapping of HIV-1 genotypes in various risk groups and regions across China based on a nationwide molecular epidemiologic survey. *PLoS One* 2012;7(10):e47289.
- [119] Zarrabi N et al. Combining epidemiological and genetic networks signifies the importance of early treatment in HIV-1 transmission. *PLoS One* 2012;7(9):e46156.
- [120] Hansen T, et al. Interactive visualization of hospital contact network data on multi-touch displays. In: *Proceedings of the 2010 mexican workshop on human–computer interaction*, vol. 1; 2010. p. 15–22.
- [121] Viégas F, Donath J. Social network visualization: can we go beyond the graph. In: *Workshop on social networks*, CSCW; 2004.
- [122] Xia H, et al. Evaluating strategies for pandemic response in Delhi using realistic social networks. In: *NDSSL technical report*; 2013.
- [123] Ives B, Olson M, Baroudi J. The measurement of user information satisfaction. *Commun ACM* 1983;26(10):785–93.
- [124] Roth R, MacEachren A, McCabe C. A workflow learning model to improve geovisual analytics utility. In: *Proc int cartogr conf. NIH Public Access*; 2009.
- [125] Robinson A et al. Combining usability techniques to design geovisualization tools for epidemiology. *Cartogr Geogr Inf Sci* 2005;32(4):243–55.
- [126] Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. *Comput Biol Med* 2002;32(3):141–9.
- [127] Driedger SM et al. If you build it, they still may not come: outcomes and process of implementing a community-based integrated knowledge translation mapping innovation. *Implement Sci* 2010;5.
- [128] Yan WR et al. ISS - an electronic syndromic surveillance system for infectious disease in Rural China. *PLoS One* 8 2013(4).
- [129] Harbage B, Dean AG. Distribution of Epi Info software: an evaluation using the Internet. *Am J Prev Med* 1999;16(4):314–7.
- [130] Olsen SF, Martuzzi M, Elliott P. Cluster analysis and disease mapping – why, when, and how? A step by step guide. *Brit Med J* 1996;313(7061):863–6.
- [131] Core Team R. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
- [132] SAS Institute. SAS. Cary, NC: SAS Institute; 2013.
- [133] MathWorks. MATLAB. Natick, MA: The MathWorks Inc.; 2013.
- [134] RStudio and Inc. Shiny: Web application framework for R; 2013.
- [135] RStudio. RStudio: Integrated development environment for R. Boston, MA; 2013.
- [136] Chester KG. BioSense 2.0. *Online J Pub Health Inf* 2013;5(1).
- [137] Light A, Bartlein PJ. The end of the rainbow? Color schemes for improved data graphics. *Eos, Trans Am Geophys Union* 2004;85(40):385–91.
- [138] Zhu, Yiliang, et al. Initial evaluation of the early aberration reporting system—Florida. In: *MMWR morb mortal wkly rep* 54; 2005. p. 123–30.
- [139] Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. *Manage Sci* 1989;35(8):982–1003.
- [140] Nielson J. *Introduction to Usability*; 2012. Available from: <<http://www.nngroup.com/articles/usability-101-introduction-to-usability/>>, [cited 10.09.13].
- [141] Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol* 1983;51(3):390.
- [142] Rogers EM. *Diffusion of innovations*. Simon and Schuster; 2010.
- [143] Atkinson PM, Graham AJ. Issues of scale and uncertainty in the global remote sensing of disease. *Adv Parasitol* 2006;62:79–118.
- [144] Linard C, Tatem AJ. Large-scale spatial population databases in infectious disease research. *Int J Health Geogr* 2012;11:7.
- [145] Tatem AJ et al. The effects of spatial population dataset choice on estimates of population at risk of disease. *Popul Health Metr* 2011;9:4.
- [146] Oppong JR et al. Mapping late testers for HIV in Texas. *Health Place* 2012;18(3):568–75.
- [147] Zinszer K et al. Residential address errors in public health surveillance data: a description and analysis of the impact on geocoding. *Spat Spatio-temporal Epidemiol* 2010;1(2–3):163–8.
- [148] Eaton C, Plaisant C, Drizd T. Visualizing missing data: classification and empirical study. In: *IFIP international conference on human-computer interaction*, September 12–16, 2005, Rome, Italy; 2005.