

Probing interactivity in open data for General Practice. An evidence-based approach

Federico Cabitza^{1,2}, Francesco Del Zotti³, and Angela Locoro²

¹ IRCCS Istituto Ortopedico Galeazzi, Milano, Italy

² Università degli Studi di Milano-Bicocca, Milano, Italy
{cabitza,angela.locoro}@disco.unimib.it

³ Centro Studi FIMMG Verona and NetAudit, Verona, Italy
delzotti@libero.it

Abstract. We undertook a user study to evaluate whether the perceived utility of some common open data sets for family doctors would increase if they are rendered in interactive heat maps. We also investigated whether Parallel Coordinates (PC) are perceived as a convenient diagram to summarize multiple patients data; and whether making PCs interactive would increase their informativity. We interviewed 29 expert family doctors through a questionnaire to find out that: interactive maps make health datasets be perceived as more useful, especially in regard to registers on exposure to carcinogens. PCs were found to be informative visualizations but making them interactive increases their perceived informativity. In light of this evidence, designing for better interactivity is worthy of further efforts in Human-Data Interaction research.

Keywords: evidence-based design, open data, general practice, data visualization

1 Introduction

An evidence-based approach to IT design⁴ [1] means to back up any claim of utility or usability of a particular solution or tool with empirical studies that collect data from real users (including the perceptions of domain experts) and undertake standard statistical tests and group comparisons [11]. This approach is typical in the Human-Computer Interaction field and it has also been recently proposed for studies accomplished in the cognate research area denoted as Human-data Interaction (HDI) [5]. HDI regards the design and study of innovative means by which users can retrieve and explore great amounts of data and the optimal ways by which these data can be represented conveniently and conveyed to users so as to make them valuable and informative [6].

⁴ Of course, the parallel with the approach to medical practice known with the phrase “evidence-based medicine” is not coincidental at all. Indeed, it constitutes a sort of natural extension of that methodological attitude to the IT-based tools that can have an impact on medical decision and care quality much similarly to the knowledge retained by doctors and nurses about clinical conditions and treatments.

As the name suggests, HDI lies at the intersection of the fields of human-computer interaction, data and information visualization and cognitive ergonomics. While other fields, which focus on very large databases, data warehouses and cooperative information systems, are concerned with the collection of increasingly complex and vast datasets, on the improvement and management of their information quality, HDI studies focus on how to improve the user experience of the information services that can be enabled by these datasets, and on the assessment of dimensions of this experience like informativity, intuitiveness, clarity, interactivity, confidence in decision making [4].

In this paper we focus on health datasets and how these can positively inform general practice of family doctors as consumers of the above datasets. In particular, we will report the results of an exploratory user study that we aimed at evaluating whether the perceived utility of some common open data sets and visualization tool would increase if the data consumers were allowed to interact with the data and explore them. This study is then to be related to similar initiatives aimed at evaluating the value of data visualization tools in the health-care sector. To our knowledge the most relevant of the recent initiatives is the “Visualizing Health” project⁵, developed by the University of Michigan and the Robert Wood Johnson Foundation. In this project, which is subtitled “a scientifically vetted style guide for communicating health data”, the involved researchers compared tens of different styles by which to communicate risk-related information in order to select the best ones for the general public according to a series of empirical tests involving real users. The main finding, which we also back up, is that when it comes to presenting health information, there is no single “best” graphic, as this depends on the intended goal and, in a subtler way, on what matters most to the intended users of the diagrams and infographics [4]. This calls for an evidence-based approach to this kind of evaluations, which must be tailored to specific groups of people (e.g., practitioners, patients) and contexts of use [13].

In this paper we focus on the context of use of health open data, from the perspective of the health practitioners. Nowadays, an increasing number of data are collected by health agencies and hospital institutions, often with the necessary collaboration of the General Practitioners (GPs). Building health datasets, keeping them up-to-date and making them open, freely accessible, and above all, usable in everyday medical practice (i.e., the primary HDI concern) require skills and resources, in particular the time of doctors and the money of citizens and tax payers [8]. For this reason, there is a need for further studies that involve potential users and data consumers in the early identification of useful datasets, in prioritizing their digitization and publication in open data platforms, and in exploring novel means that could improve the end-user experience. Our study is a preliminary, but yet evidence-based [11] contribution in this strand of research.

⁵ <http://www.vizhealth.org/>

2 Method

This study addresses two related research questions on open data in healthcare. First: What health open data are perceived as *more useful* in general practice (if any)? Second: Would visualizing these open data by means of *interactive* (that is browsable, zoomable, etc.) maps make these data to be perceived as *more useful and usable*? In addition, we also address the question whether diagrams displaying *parallel coordinates* [9] (see Figure 2) are effective means to represent patient data at various levels of description and to help GPs understand population trends, discover (frequency-based) normal and abnormal conditions, and detect correlations between multiple conditions.

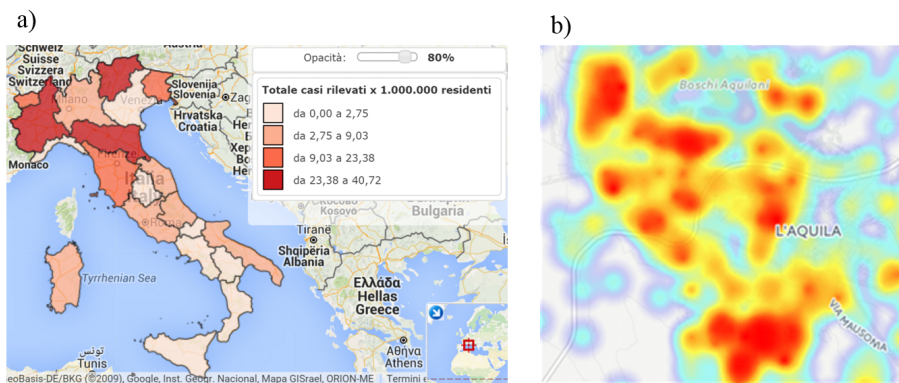


Fig. 1. The static map (on the left, a); the heat map (on the right, b) resulting from interacting with the map depicted on the left.

To these aims, the authors conceived a short questionnaire to be administered as a Computer-Aided Web Self-Interview (CAWI). This questionnaire consisted of four sections, which were to be rendered each in a different Web page: in the first page, the platform was to present a list of 13 datasets that the authors had previously selected among the health open datasets made available on the DatiOpen.it⁶ platform, which is one of the biggest Italian open data portals collecting more than 2,000 thousand datasets. In particular the respondents were asked to choose at least one dataset (and at most three) that they could want to consult in an online and customized “visual dashboard” supporting their own professional practice.

In the second page we further selected 6 datasets for their suitability to be mapped into a georeferenced map like the one depicted in Figure 1.a, and asked the doctors to evaluate the degree of *utility for their profession* of each dataset on a six-value ordinal scale, from 1 (not useful at all) to 6 (very useful).

⁶ <http://www.datiopen.it/en>

Then, we asked the doctors to imagine to be able to interact with the map, so that they could browse the map by dragging, and zooming in and out at the desired level of visualization (e.g., province, city, neighborhood if possible) and get a *heat map* like the one depicted in Figure 1.b. We then asked whether visualizing the above datasets on such a browsable map would make the data more useful or not (on a 5-value scale ranging from 1 “the data would be much *less* useful” to 5 “the data would be much *more* useful”, passing through a middle value 3 “it wouldn’t make almost any difference”).

In the third section, the questionnaire displayed a parallel coordinates diagram like the one depicted in Figure 2. We asked to indicate on a six-value ordinal scale the degree of informativity of such a way to represent patient data (from 1 not informative at all, to 6, very informative).

Then we asked to imagine to be able to interact with the diagram, so as: 1) to select one specific line (that is a patient) and see that line highlighted while the other lines would fade slightly in the background; 2) to draw small rectangles above one or more vertical axis, in order to have the system highlight only the patient-lines intersecting the axes within that data ranges (see Figure 3). Similarly to the case of dataset utility, we asked whether the interactive diagram would be more or less informative than the static version of it.

Finally, the last page of the questionnaire contained two profile-related questions: one regarding the work experience of the respondent; the other one regarding the perceived familiarity with IT.

We invited the potential respondents by sending an electronic mail to the family doctors registered to the NetAudit⁷ mailing list; this is a list of approximately 120 Italian GPs who are interested in medical audit and research initiatives regarding the assessment of performance in general practice and the continuous improvement of the related medical quality and outcome.

⁷ <http://www.netaudit.org/>

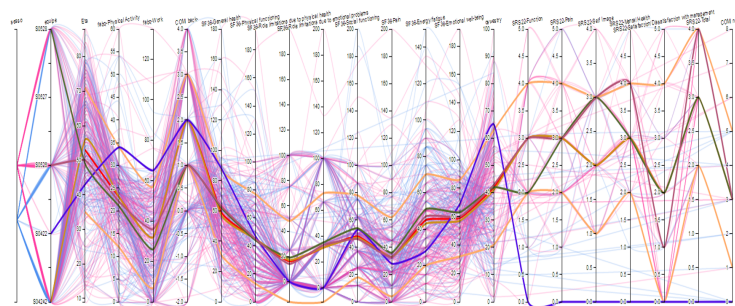


Fig. 2. Parallel coordinates to visualize patient data on multiple dimensions and see both population-level tendencies and inter-dimensional correlations (inspired by [7]).

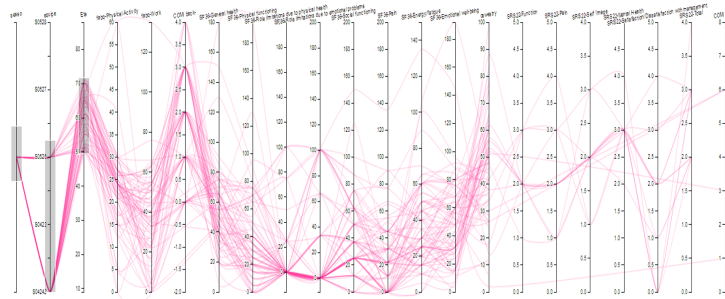


Fig. 3. The same diagram depicted in Figure 2 after that three specific ranges of data have been specified to filter the patient data to be highlighted: more precisely, female patients treated by two specific medical teams (equipe) and between 70 and 50 years of age.

3 Results

The questionnaire was left open for two weeks after the invitation had been sent to potential respondents. In this time lapse no reminder was sent. When we closed the survey, 29 GPs had accessed the questionnaire; 23 of them had completely filled in every item of the questionnaire in an average time of 5.9 minutes (SD=2.6 minutes). In this respondent sample, 52% have been family doctors for more than 30 years, 91% for more than 20 years; no respondent had less than 10 years of work experience.

The majority of the respondents (65%) claimed to be competent in Information Technology (IT) but neither experts nor enthusiasts. These latter ones were, respectively, approximately one third and one fifth of the sample. We detected a statistically significant and moderately negative correlation between age and IT expertise ($-0.49, p=0.018$), as it could be expected.

As said in the previous section, respondents could choose the three most useful datasets to be visualized on a personal dashboard for their practice, out of a list of 13 data sets. The sets chosen more frequently by the respondents were:

1. Regional distribution of the percentage of elderly people⁸ involved in the program of Integrated Home Care (*Assistenza Domiciliare Integrata*).
2. Number (and details) of the Rehabilitation facilities for the elderly (*Istituti di Riabilitazione Extraospedaliera per Anziani*).
3. Inpatient care average length of stay.

These datasets were selected by approximately one third of the sample (38%, 33% and 30% respectively). The other datasets collected much fewer preferences, with “number of cases affected by pneumococcal disease by year” that was not chosen by any respondent.

⁸ That is, people older than 65 years.

In regard to the utility of data sets to be displayed on a georeferenced map, we got statistical significance for two sets only, which were considered useful for general practice, namely:

1. Number of registers on exposure to carcinogens (.5 vs. .95, $p=.000$).
2. Hospitalization rates, by regime, patient genre and region (.26 vs. .74, $p=.035$).

For the other data sets we did not get statistical significance: however, the “number of medical prescriptions delivered” can be considered the least appreciated information set (.64 vs. .36).

We also asked whether putting all of the above datasets into an interactive map, instead of a static one, would change the perceived utility of the datasets themselves. The majority claimed mapping data would make a difference, and that in so doing the datasets would be more useful (.83 vs. .17, $p=.003$, one third of the GPs even claiming that the data “would be much more useful”). The ranking of relative utility of mapping open data sets is reported in Table 1.

Table 1. Open datasets to be visualized in maps ranked by perceived utility.

Dataset	Priority level ^(sig.)	Perceived utility ^(sig.)	Median
Carcinogen exposure registers	Higher ^(*)	Positive ^(***)	5
National Drug Consumption	Uncertain ^(NS)	Positive ^(NS)	4
Home care patients	Uncertain ^(NS)	Positive ^(NS)	4
Hospitalization rate	Uncertain ^(NS)	Positive ^(*)	4
Incidence of invasive diseases	Uncertain ^(NS)	Positive ^(NS)	4
Incidence of occupational diseases	Uncertain ^(NS)	Positive ^(NS)	4
Number of prescriptions	Uncertain ^(NS)	Negative ^(NS)	3

In regard to the last research question addressed in this short study, parallel coordinates diagrams (see Figures 2 and 3) were found to be an informative visualization to render patient data on multiple dimensions. The responses were positive for the majority (.64 vs. .46, $p=.286$). Furthermore, almost every GP involved in the user study claimed that making such a diagram interactive would make it even *more* informative (.96 vs. .04, $p=.000$, 60% “much more informative”).

4 Discussion and Conclusions

The characteristics of the respondent sample, in terms of work experience and attitude towards Health IT suggest that the respondents can be considered representatives of a population of real domain experts and that the potential (positive) bias introduced by the IT enthusiasts should be low. Besides the results reported

above, the questionnaire also allowed the respondents leave free text comments in order to suggest datasets and applications to be considered in further studies, give advice on potential improvements to the visual tools presented, and more generally share remarks with the research team.

Through these items, three respondents suggested to perform a similar usability study on one of the most adopted GP governance tools in Italy, i.e., MilleGPG ⁹, by also noticing an almost total lack of this kind of studies, with a few exceptions (e.g.,[3]), in regard to a class of applications that are used (or must be used) by thousands of GPs in Italy, and in regard to applications that can foster collaboration between healthcare operators by means of free text comments, tags labelling and semantically aware classifications [10], for example for transforming tabular data into valuable visualizations.

Moreover, we report two comments that suggested relevant improvements to the Parallel Coordinate (PC) dashboard. One GP suggested to implement PCs indicating explicitly central tendency parameters on their axes (wherever applicable), like means and medians, standard deviations and interquartile ranges, both at the level of the single GP patient set, and (if available) at the regional and national level. This doctor noticed how the bundle of lines, in virtue of their variable thickness at the intersection with the different coordinates, would nicely and visually represent the frequency distribution of the values within the subject population. Another GP suggested to allow the user to switch from a cross-sectional view (i.e., the default one) where each line represents the current conditions of a single patient, to a longitudinal view. In such a view, the current conditions would be represented as a highlighted line, while the bundle of the other lines would represent the different conditions that the same patient exhibited in the past, as these are recorded on the medical record of the GP or in the patient's health record.

We are aware of the limitations of this study. For instance, the fact that the involved doctors had to imagine the novel interaction, by looking at static pictures of how the visualization tools would change according to their actions. However, giving doctors real dashboards with which to interact and ask about their perceptions would have been problematic for the impossibility to give them a uniform training and to check remotely their interaction with the tool to evaluate. For this reason, we believe that studies like the present one, where interactivity is simulated in terms of its effects, could be useful anyway in a preliminary phase of development, where to understand if endowing visual tools with some kind of interactivity would be worth the effort.

In light of the evidence collected in this study, we can conclude that designing for a better interactivity of open data platforms for general practice is worth further efforts. From the HDI perspective, this conclusion was desirable; although it might also seem easily predictable, nevertheless that conclusion could not be taken for granted in a committed *evidence-based approach* to Health IT design [12, 11, 6]. For this reason, this study is a small but necessary contribution in the research agenda outlined in Section 1, which is aimed at gaining feedback

⁹ <https://www.millegpg.it/>

and design-oriented indications from both caregivers and patients to improve their experience in interacting with the next data visualization tools to come.

References

1. Ammenwerth, E., & Rigby, M. (Eds.). (2016) Evidence-Based Health Informatics: Promoting Safety and Efficiency Through Scientific Methods and Ethical Policy (Vol. 222). IOS Press.
2. Cabitza, F., & Simone, C. 2010. WOAD: a framework to enable the end-user development of coordination-oriented functionalities. *Journal of Organizational and End User Computing (JOEUC)*, 22(2), 1-20.
3. Cabitza, F., Del Zotti, F., & Misericordia, P. 2014. Electronic Records for General Practice – Where we Are, Where we should Head to Improve Them. In *Healthinf'14: Proceedings of the International Conference on Health Informatics*, (pp. 535-542). INSTICC.
4. Cabitza, F., Locoro, A., & Batini, C. 2015. A User Study to Assess the Situated Social Value of Open Data in Healthcare. *Procedia Computer Science*, 64, 306-313.
5. Cabitza, F., Locoro, A. 2016. Human-Data Interaction in Healthcare: Acknowledging Use-related Chasms to Design for a Better Health Information. In *the Procs of the 8th IADIS International Conference on e-Health 2016, Part of the Multi Conference on Computer Science and Information Systems, MCCSIS 2016* .
6. Cabitza, F., Fogli D., Giacomini, M. & Locoro, A. 2016. Valuable Visualization of Healthcare Information: from the quantified self data to conversations. In *AVI 2016: Proceedings of the International Working Conference on Advanced Visual Interfaces*, Bari, Italy, 7-10 June 2016. pp. 376-380. ACM.
7. Croon, R. D., Klerkx, J., & Duval, E. 2015. Design and evaluation of an interactive proof-of-concept dashboard for General Practitioners. In *ICHI 2015: Proceedings of the International Conference on Healthcare Informatics*, 2015. pp. 150-159. IEEE.
8. Eysenbach, G. 2008. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. *Journal of medical Internet research*, 10(3).
9. Inselberg, A. 2009. Parallel coordinates. In Ling Liu, Tamer Ozsu (Eds.) *Encyclopedia of Database Systems*. pp. 2018-2024. Springer.
10. Locoro, A., Grignani, D. and Mascardi, V. 2011. MANENT: An infrastructure for integrating, structuring and searching digital libraries. *Studies in Computational Intelligence*, Springer, 375, 315-341.
11. Longhurst, C. A., Palma, J. P., Grisim, L. M., Widen, E., Chan, M., & Sharek, P. J. 2013. Using an Evidence-Based Approach to EMR Implementation to Optimize Outcomes and Avoid Unintended Consequences. *Journal of Healthcare Information Management (JHIM)*, 27(3), 79.
12. Rigby, Michael, Elske Ammenwerth, M. Beuscart-Zephir, Jytte Brender, Hannele Hypponen, Siobhan Melia, Pirkko Nykanen, Jan Talmon, and Nicolette de Keizer. 2013 Evidence Based Health Informatics: 10 years of efforts to promote the principle. *Yearb Med Inform* 8(1), 34-46.
13. Solomon, J., Scherer, A. M., Exe, N. L., Witteman, H. O., Fagerlin, A., & Zikmund-Fisher, B. J. 2016. Is This Good or Bad?: Redesigning Visual Displays of Medical Test Results in Patient Portals to Provide Context and Meaning. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. pp. 2314-2320. ACM.