LifeTrack: decades of EHR data in a single view

Harri Siirtola, Roope Raisamo
TAUCHI Research Center
ITC Faculty
Tampere University
Tampere, Finland
firstname.lastname@tuni.fi

Mary Pat Reeve, Javier Gracia-Tabuenca,
Vincent Llorens, Shanmukha Sampath Padmanabhuni
Finnish Institute for Molecular Medicine (FIMM)
University of Helsinki
Helsinki, Finland
firstname.lastname@helsinki.fi

***Abstract*—There are situations where we need to understand a person's health register data, get an overview of it and inspect the details of interest, preferably in a short period of time. Typical situations are a doctor seeing a patient for the very first time, or a researcher trying to understand what makes a person an outlier in a patient cohort.**

We have developed a visualization tool, called LifeTrack, that will show a person's whole health register history in a single view. The structure of the LifeTrack data view is based on concepts familiar to the target audience, and allows one to see both the overview and details of the data.

***Index Terms*—health information management; data visualization**

I. Introduction

Health register data is a valuable resource for clinical work, medical research and the development of new medications. Databases of health data, or Electronic Health Records (EHR), contain longitudinal histories of diagnoses, treatments, medications, and outcomes. Analyzing the health register data allows us to identify patterns and trends in disease prevalence, risk factors, and treatment outcomes. Overall, the health register data can play a crucial role in improving patient care, advancing medical research, and understanding medical care resource use and outcomes.

Health data tends to be voluminous and complex, making identifying patterns and trends daunting. We propose a novel interactive visualization technique for a person's EHR data called *LifeTrack*. It is an overview+detail [1] visualization that allows the recognition of patterns and inspection of details in a single view. LifeTrack lets you explore the available health data without any pre-configuration or assumptions about the available data.

EHR data often originates from separate registers (e.g. hospital, outpatient clinic, death, cancer) and contains a variety of medical vocabularies for diagnoses, procedures and medications. LifeTrack uses the classification systems present in medical code sets to organize the data in a way that clinicians and medical researchers are familiar with.

The main idea is to show all the EHR entries in a (single) view which can be queried or drilled down by mouse-hover, revealing the details (e.g., type of visit or provider, age, etc.) for the health register entry depicted as a point. We use a variant of the dot plot to present EHR entries. The layout is a traditional time vs. register entries, but we bring structure to the visualization by categorizing and classifying the entries according to vocabularies in addition to the source registers.

This work was done in the context of the FinnGen project[1]. The data shown in this article is artificial: any personal IDs visible are fake, and only the data distributions are similar to the actual health register data.

II. Background

The problem of visualizing a person's health history as depicted by EHR entries has been widely studied, as the recent state-of-the-art report shows [2]. There are several systems that focus on the visualization of a person's EHR history, with varying goals. Many of the proposed techniques concentrate on *summarizing* the EHR data, without giving access to the *details*.

Zhang et al. [3], [4] presented *IDMVis*, a decision support tool to visualize temporal event sequences in type 1 diabetes treatment. IDMVis can combine and visualize temporal event sequences for diabetes management tasks, and it was developed in close co-operation with clinicians.

Borland et al. [5] proposed using radial variant of parallel coordinates to present health data. They used the approach to analyze data from the UK's NHS by using an interactive tool, and present an interesting analysis of London practices that appear initially to be outliers.

Gschwandtner et al. [6] presented *CareCruiser*, a prototype system to visualize clinical treatment plans and their effects on the patient's condition. The system focuses on visualizing the lab values and their changes.

Bui et al. [7] created a system called *TimeLine* that provides a rich view into EHR data. Its aim is to create problem-centric temporal visualizations for medical data.

Bade et al. [8] have developed interactive visualization techniques for time-oriented data, which have been applied to the medical and stock exchange domains and should be generally applicable. Their focus is more on the overview than the details.

III. View layout and encoding

EHR data is basically events over time. The main complication is that there are many event types, and they can pile up around the same time points, making the visual presentation difficult to construct and hard to read.

Figure 1 shows our layout idea. Time is on the x-axis, and on the y-axis there are multiple parallel axes defined by the categories and classifications. At the top level, there are categories

[1] https://www.finngen.fi/en

procedures, *diagnoses*, and *medications.* The categories are further divided into classes based on the vocabulary used in the entry, such as ICD (*International Statistical Classification of Diseases and Related Health Problems*[2]) or ATC (*Anatomical Therapeutic Chemical Classification*[3]). ICD and ATC are the two most extensive classification hierarchies used, and only the top levels of them are used in the layout.

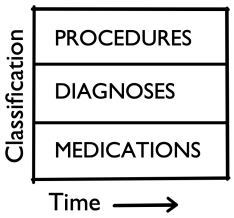

Fig. 1. Dividing EHR events into procedures, diagnoses, and medications. Within each area, the events are placed on distinct x-axes based on the vocabulary, or on the top level of the vocabulary hierarchy.

The *diagnoses* category is further divided into *certified diagnoses* and *diagnoses* (Figure 2), where *certified diagnoses* are diagnoses that have been reviewed by an oversight organization and are often confirmed by more than one specialist. Such diagnoses are needed to get national health (KELA) reimbursement for medication, a diagnosis for cancer, or a cause of death on a death certificate.

The "unclassified" category is for the entries that LifeTrack is unable to place into any of the known classes. At the moment, there are only few such items. Most of them are drug codes that are not yet in the ATC.

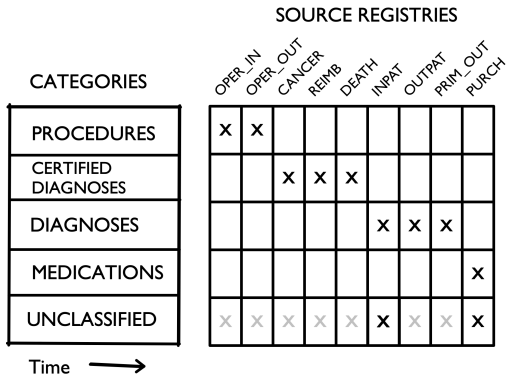

Fig. 2. Categories and source registries.

The EHR entries are color-coded according to their source register. In FinnGen, the main registries are the Finnish Cancer Registry (CANC, since 1953), Causes of Death (DEATH, 1969), Care Register for Health Care inpatient visits (INPAT, OPER_IN for procedures, 1969), Care Register for Health Care specialist outpatient visits (OUTPAT, OPER_OUT, 1998), Register for Primary Health Care Visits (PRIM_OUT, 2011), Drug Purchase (PURCH, 1995), and Kela[4] Reimbursement for medicine expenses (REIMB, 1964). LifeTrack indicates the register's start year with a vertical bar with the same color encoding as the register. In an entry class, there can be events from several registries. Indication of the register start helps users understand that the patient may have had the condition earlier, but we do not have any register data to know for certain.

IV. LifeTrack visualization tool

In the following, we first describe how LifeTrack obtains the data. Then we explain the visualization views, how one can interact with them, and describe how the tool was put together.

A. Input

A person (or a case) to visualize is chosen in LifeTrack from a list of individual IDs provided by a cohort file. The cohort file is a tsv (tab-separated values) text file that must have a column named FINNGENID (other columns may be present, but are ignored). This is the standard method to pass cohort definitions between FinnGen software tools.

Below the case-selector is a checkbox titled "Hide visited". When toggled on, the selector will only show those cases from the cohort that have not been loaded during the current session. This book-keeping feature is meant to a situation where you want to explore every case in the cohort file.

In the FinnGen project, the actual register data is read from a BigQuery database, and the code translations come via the FinnGenUtils.R package.

B. Views

Figure 3 shows the LifeTrack user interface. On the left is the dashboard menu that contains the visualization controls, and on the right is the LifeTrack visualization. From top to bottom, the visualization controls include cohort, individual selector, filtering options, and general settings. On the top of the dot plot view there is a title row that gives some background information about the person: ID, gender, birth year, height, smoking status, the number of offspring, the county of birth, the number of register entries, and the time period we have data for this person. The dashboard menu can be hidden from the hamburger menu on the dashboard header to release more space for the visualization.

The LifeTrack visualization will simplify the view as much as possible by leaving out categories, classes, and source registries that do not have any data. This is the highest abstraction of register entries, and can be acquired simply by reading the y-axis values and the source register legend.

Register entries are depicted as a *dot plot* (or *dot chart*) [9] on separate x-axes according to the class of entry. If there are multiple entries on the same date, or close by, then the points are offset slightly to remain distinct. With this technique, it is easy to see where there are gaps or clusters of entries. Figure 4 shows a simple dotplot.

[2]https://www.who.int/standards/classifications/classification-of-diseases

[3]https://www.who.int/tools/atc-ddd-toolkit/atc-classification

[4]https://www.kela.fi

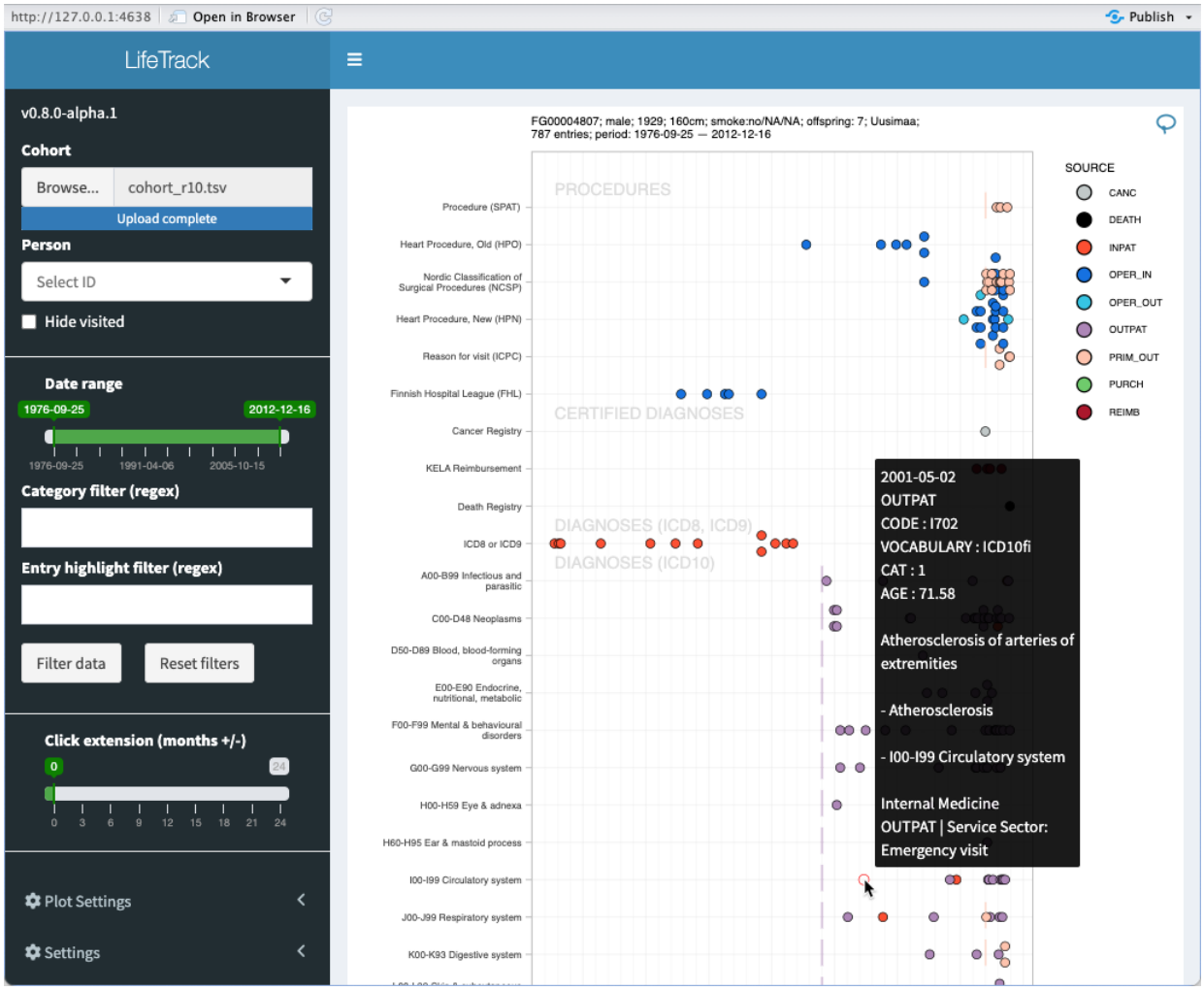

Fig. 3. The user interface of LifeTrack.

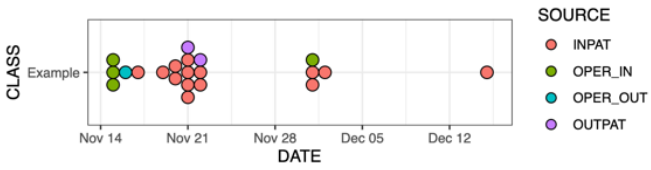

Fig. 4. A simple dot plot illustrating the stacking.

TABLE I
THE DISTRIBUTION OF REGISTER ENTRIES IN FINNGEN DATA.

Entries	N	Percentage
1-100	46848	9.89
101-500	263134	55.60
501-1,000	111703	23.60
1,001-5,000	48796	10.30
5,001-10,000	2109	0.45
10,001-20,000	805	0.17
$>$ 20,000	185	0.04

FinnGen's current "data freeze" (a pseudonymized snapshot of registers) contains genotype and health register data for 473,681 individuals. Typical input data size (the number of health register entries for a person) is shown in Table I. The median for entries is 352, and the mean is 554.3. Over 65% of the individuals have 500 or less register entries, and over 99% have less than 5000 entries. The majority of the cases can be explored with LifeTrack's default settings and by adjusting the plot settings it is possible to process even the cases with the highest number of entries.

C. Interaction

Once an individual ID is chosen, LifeTrack will read all the register data for the person and construct the visualization. It can be then queried by "mouse-hover" or by moving the mouse on top of a register entry (a point). Figure 5 shows how the entry details are presented.

Selecting the desired point can be challenging in some situations. LifeTrack will help by automatically choosing the closest point for the hover. If there are several events on the

same day, they are all selected, but only the details of the closest one are displayed.

Exploring a set of register entries via hover will get tedious when there are many points close to each other. In LifeTrack, it is possible to select a group of points with a lasso tool and see the included register entries as an interactive table (Figure 6). The register entries can then be sorted by the column headings, and the displayed items can be limited by giving a search string.

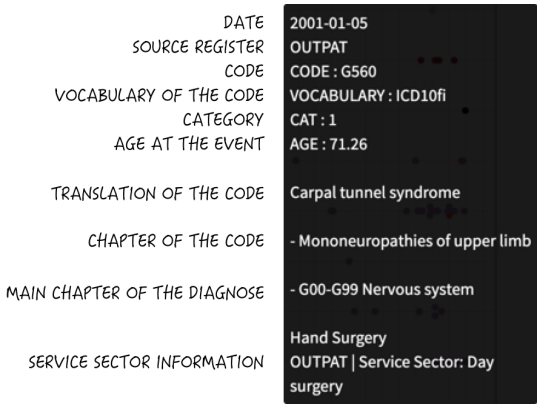

Fig. 5. Details of a health register entry.

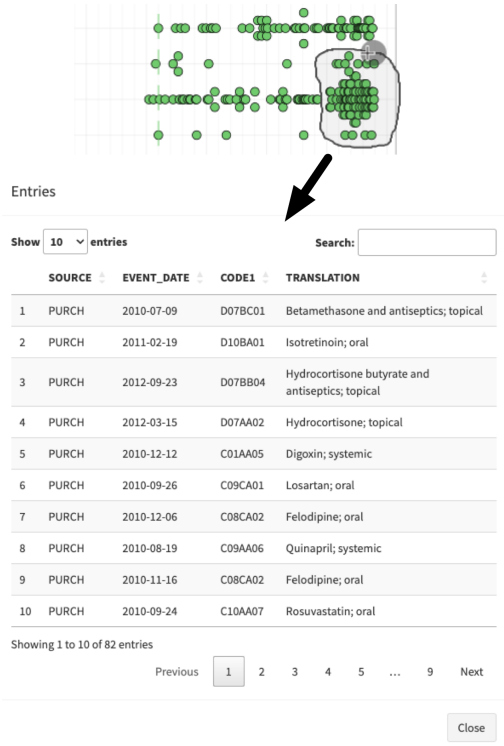

Fig. 6. Selecting a set of cases and viewing them as an interactive table.

If you have found an entry of interest and would like to see the events around it, selecting a single point without using the lasso tool is also possible. This will open the table view for the point (or points, if there are several in that day and entry class). There is then a slider in the user interface that specifies how these single-click selections will be extended by setting the +/- months value. This will give a table of events before and after the entry of interest (even from other parallel classifications on the y-axis that may be out of view), and may help in finding causal relationships.

Another method to simplify the visualization is by horizontal and vertical zooming. The most obvious need is to limit the time span of the displayed register entries. For example, in the FinnGen data set, some registries go back to the 1960s while most registries have data starting from the mid-1990s. This kind of data will waste screen estate and make the dot plots more crowded. Limiting the date range makes it possible to concentrate on a desired period.

The y-axis can be simplified by giving a regular expression of data classes to be viewed. Suppose you are interested in classes related to respiratory diseases. Using "respiratory" as an expression will show only the diagnoses and medications related to it (Figure 7).

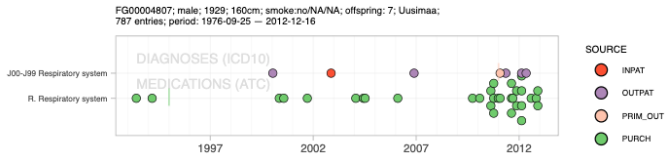

Fig. 7. Limiting LifeTrack into categories containing string "respiratory".

Finally, it is possible to give a regular expression of register entries to be highlighted. The regular expression will be matched both to the entry code and its translation, making it easy to spot desired entries. The entries itself are not visually changed, but the alpha channel value of non-matching entries is lowered, i.e. they fade to the background (Figure 8). While being faded, you can still hover-query them.

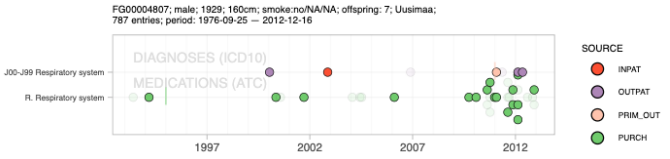

Fig. 8. Highlighting the entries selected by regular expression '^J4|^R03'. This regular expression matches to strings starting either with 'J4' or 'R03'.

In Figure 8, the entry filter is "^R03|^J4" which will choose the ATC code starting with "R03", denoting "drugs for obstructive airway diseases", or "J4", which is the top ICD class for "chronic lower respiratory system diseases". The filter can of course be a simple string, such as "emergency" (highlighting entries entered in emergency services) or "hospital" (highlighting entries recorded in a hospital context). A set of highlighted values can be opened in the table view by clicking any of the points.

D. Implementation

LifeTrack is implemented as a *shiny* web application [10] with the R language [11] by utilizing the *shinydashboard*

framework [12], [13]. Many R packages are used, the most relevant ones being from the *tidyverse* [14] collection, especially *ggplot2* [15] and its interactive SVG extension *ggiraph* [16]. In addition, the *DT* package [17] and *DataTables* JavaScript library [18] are used to implement the interactive table view. LifeTrack is open source and available on GitHub[5].

V. Discussion

LifeTrack complements our earlier *Trajectory Visualization Tool (TVT)* [19]. TVT is limited to six-dimensional data, but it uses OMOP concepts [20] instead of raw register data used by LifeTrack. So with TVT, the OMOP concepts unify register entries making observations over time easier, but the cost is limited dimensionality and additional configuration time. From the data protection perspective, the TVT approach is safer as it masks the underlying raw register data. Since TVT's six dimensions are compressed into one y axis, you can view many individuals at once and look for common patterns across a cohort of individuals, rather than looking in detail at one individual.

Visualization tools which enable drilling down into EHR details are rare because of the data protection issues and the requirement that data viewed outside a Trusted Research Environment (TRE) represent at least 5 individuals. However, as the number of TREs grow, such as the UK Biobank and the All of US TRE environments, so does demand for these types of EHR visualization tools.

LifeTrack has already demonstrated a popular reception from both clinicians and researchers.

Clinicians often face a situation where they need to familiarize themselves with a patient's medical history in a limited time. Patients frequently fail to summarize their health history with adequate detail, and reading several pages of EHR entries is not an option. In this situation, LifeTrack enables a clinician to get an overview and rapidly drill down into relevant details.

Additionally, the ability to see how the medical codes are used in practice helps researchers create better phenotypic categories for genetics analyses such as genome-wide association studies (GWAS). The start and stop times shown vertically for the registries also help researchers understand how age of onset of a condition may be represented in the data.

Researchers are also using LifeTrack to review individuals with rare genetic diseases. For example, a geneticist may be interested in confirming a rare variant as the cause of disease. LifeTrack enables geneticists to easily load and view the individuals with the variant – due to the rareness of the disease and difficulties in arriving at an accurate diagnosis, cases may have different ICD codes – but the researcher can look for additional evidence from treatments, procedures and medications, and basic height and weight statistics. Similarly, for individuals listed as having a rare disease their codes can be reviewd to see if their medical life history supports that diagnosis, or if it is possibly a typo.

The development of LifeTrack still continues. One recognized extension is to add laboratory results, clinical notes, and radiology notes into data. They are also events over time, and would complete the view provided by LifeTrack.

VI. Conclusion

The visualization of electronic health records (EHR) is becoming an essential facility for clinical work and medical research. Understanding the ever-growing content of EHR data provides a reliable record on how the various treatments, procedures, and medications actually perform.

We have designed and implemented an interactive LifeTrack visualization tool for EHR data. It is based on the traditional overview+detail concept and its data representation is immediately understandable to specialists familiar with medical coding systems.

The earlier approaches proposed have usually concentrated on the overview aspect of the problem, mainly for the data security reasons. This is now clearly changing as the pseudonymization and trusted research environments are becoming acceptable.

VII. Acknowledgements

We want to acknowledge the participants and investigators of the FinnGen study[6]. The FinnGen project [21] is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc, Novartis AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Arctic Biobank [22], Auria Biobank [23], THL Biobank [24], Helsinki Biobank [25], Biobank Borealis of Northern Finland [26], Finnish Clinical Biobank Tampere [27], Biobank of Eastern Finland [28], Central Finland Biobank [29], Finnish Red Cross Blood Service Biobank [30], Terveystalo Biobank [31] and The Finnish Hematology Registry and Clinical Biobank [32]. All Finnish Biobanks are members of BBMRI.fi infrastructure [33]. Finnish Biobank Cooperative (FINBB) [34] is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services [35] managed by FINBB.

References

[1] A. Cockburn, A. Karlson, and B. B. Bederson, "A review of overview+ detail, zooming, and focus+ context interfaces," *ACM Computing Surveys (CSUR)*, vol. 41, no. 1, pp. 1–31, 2009.

[2] Q. Wang and R. S. Laramee, "EHR STAR: The state-of-the-art in interactive EHR visualization," in *Computer graphics forum*, 2022, vol. 41, pp. 69–105.

[5]https://github.com/FINNGEN/LifeTrack

[6]https://www.finngen.fi/en

[3] Y. Zhang, K. Chanana, and C. Dunne, "IDMVis: Temporal event sequence visualization for type 1 diabetes treatment decision support," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 512–522, 2018.

[4] Z. Zhang *et al.*, "AnamneVis: A framework for the visualization of patient history and medical diagnostics chains," 2011.

[5] D. Borland, V. L. West, and W. E. Hammond, "Multivariate visualization of system-wide national health service data using radial coordinates," 2014.

[6] T. Gschwandtner, W. Aigner, K. Kaiser, S. Miksch, and A. Seyfang, "CareCruiser: Exploring and visualizing plans, events, and effects interactively," in *2011 IEEE pacific visualization symposium*, 2011, pp. 43–50.

[7] A. A. Bui, D. R. Aberle, and H. Kangarloo, "TimeLine: Visualizing integrated patient records," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 4, pp. 462–473, 2007.

[8] R. Bade, S. Schlechtweg, and S. Miksch, "Connecting time-oriented data and information to a coherent interactive visualization," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2004, pp. 105–112. Available: https://dl.acm.org/doi/pdf/10.1145/985692.985706

[9] L. Wilkinson, "Dot Plots," *The American Statistician*, vol. 53, no. 3, pp. 276–281, 1999, Accessed: Mar. 15, 2023. [Online]. Available: http://www.jstor.org/stable/2686111

[10] W. Chang, J. Cheng, J. Allaire, Y. Xie, J. McPherson, *et al.*, "Shiny: Web application framework for R," *R package version*, vol. 1, no. 5, p. 2017, 2017.

[11] R Core Team, *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2023. Available: http://www.R-project.org/

[12] W. Chang *et al.*, *Shiny: Web application framework for R*. 2021. Available: https://CRAN.R-project.org/package=shiny

[13] W. Chang and B. Borges Ribeiro, *Shinydashboard: Create dashboards with 'shiny'*. 2021. Available: https://CRAN.R-project.org/package=shinydashboard

[14] H. Wickham *et al.*, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019, doi: 10.21105/joss.01686.

[15] H. Wickham and G. Grolemund, *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media, 2016.

[16] D. Gohel and P. Skintzos, *ggiraph: Make 'ggplot2' graphics interactive*. 2023. Available: https://davidgohel.github.io/ggiraph/

[17] Y. Xie, J. Cheng, and X. Tan, *DT: A wrapper of the JavaScript library 'DataTables'*. 2023. Available: https://CRAN.R-project.org/package=DT

[18] A. Jardine, "Datatables (table plug-in for jquery)," *Poslední čtení*, vol. 20, p. 12, 2012, Available: https://datatables.net

[19] H. Siirtola, J. Gracia-Tabuenca, M. Niemi, M. P. Reeve, and T. Laitinen, "Glyph-based visualization of health trajectories," in *2022 26th international conference information visualisation (IV)*, Jul. 2022, pp. 424–431.

[20] G. Hripcsak *et al.*, "Observational health data sciences and informatics (OHDSI): Opportunities for observational researchers," *Studies in health technology and informatics*, vol. 216, p. 574, 2015.

[21] M. I. Kurki *et al.*, "FinnGen provides genetic insights from a well-phenotyped isolated population," *Nature*, vol. 613, no. 7944, pp. 508–518, 2023.

[22] Arctic Biobank, Available: https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank/information-arctic-biobank

[23] Auria Biobank, Available: https://www.auria.fi/biopankki

[24] THL Biobank, Available: https://www.thl.fi/biobank

[25] Helsinki Biobank, Available: http://www.helsinginbiopankki.fi

[26] Biobank Borealis of Northern Finland, Available: https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx

[27] Finnish Clinical Biobank Tampere, Available: https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere

[28] Biobank of Eastern Finland, Available: https://www.ita-suomenbiopankki.fi/en

[29] Central Finland Biobank, Available: https://www.ksshp.fi/fi-FI/Potilaalle/Biopankki

[30] Finnish Red Cross Blood Service Biobank, Available: https://www.veripalvelu.fi/verenluovutus/biopankkitoiminta

[31] Terveystalo Biobank, Available: https://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki

[32] The Finnish Hematology Registry and Clinical Biobank, Available: https://www.fhrb.fi/

[33] BBMRI.fi, Available: https://www.bbmri.fi

[34] Finnish Biobank Cooperative, Available: https://finbb.fi/

[35] Fingenious, Available: https://site.fingenious.fi/en/