MoBa Explorer

Enabling the navigation of data from the Norwegian Mother, Father, and Child Cohort Study (MoBa)

A design study to understand the different and intersecting expert and lay user needs when visually exploring MoBa data.

Hanna Balaka
University of Bergen
Hanna.Balaka@student.uib.no

Ragnhild Valen
Norwegian Institute of Public Health
ragnhild.valen@fhi.no

Laura Ann Garrison
University of Bergen
laura.garrison@uib.no

Marc Vaudel
University of Bergen
marc.vaudel@uib.no

MOTIVATION

The Norwegian Mother, Father and Child Cohort Study (MoBa) is a study in public heath with information about pregnancy and childhood [3]. However, the volume of the data and its sensitive nature make its examination challenging. This is a **work-in-progress design study** and accompanying web application, the MoBa Explorer, which presents MoBa study data genotypes and phenotypes. Our research explores how to serve two distinct purposes in one application:

(1) allow **researchers** to interactively explore MoBa data for further study, and

(2) provide MoBa study details to an interested **general public**.

We draw inspiration from visualization research in genomic [4] and phenotypic data [5], and similar resources such as UK BioBank [6] and FinnGen [2].

TASKS

R: Researcher; GP: General public

- Quickly identify MoBa study variables of interest (R)
- Identify MoBa study variables with high missingness (R)
- Identify and compare total participants of a given variable (R, GP)
- Create subsets of MoBa variables according to demographic (R, GP)
- Observe associations between genotype and phenotype data (R, GP)

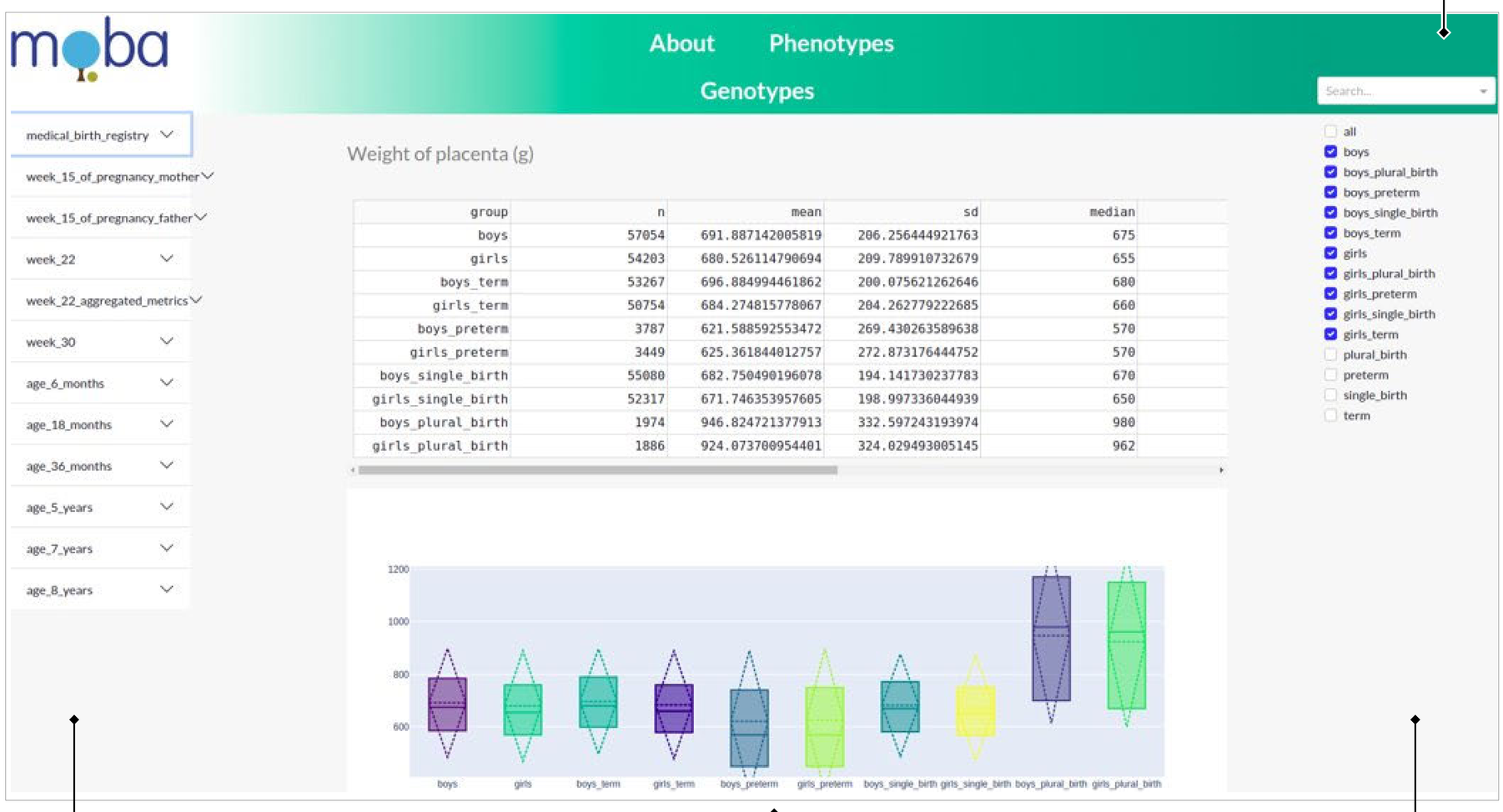
REQUIREMENTS

- Unify disparate MoBa data sources
- Ability to subset according to key demographic features of interest
- Display changing values of variables over time
- Present data in an **aggregate manner**
- Ability to show missingness for a given variable
- Meet accessibility standards [1]

MOBA EXPLORER

We have developed a **proof-of-concept MoBa Explorer** addressing some of the tasks and requirements that we have identified. The interface provides a list of time points and corresponding variables which a user can select and explore in more detail. Variables may be further subset according to user interests, and a table supports visualization accessibility.

Top panel: Consists of three tabs and a search box. Users can select About to learn more about the MoBa study, Phenotypes or Genotypes to browse variables, or search via the search box



group	n	mean	sd	median
boys	57054	691.887142005819	206.256444921763	675
girls	54203	680.526114790694	209.789910732679	655
boys_term	53267	696.884994461862	200.075621262646	680
girls_term	50754	684.274815778067	204.262779222685	660
boys_preterm	3787	621.588592553472	269.430263589638	570
girls_preterm	3449	625.361844012757	272.873176444752	570
boys_single_birth	55080	682.750490196078	194.141730237783	670
girls_single_birth	52317	671.746353957605	198.997336044939	650
boys_plural_birth	1974	946.824721377913	332.597243193974	980
girls_plural_birth	1886	924.073700954401	324.029493005145	962

Left panel: Accordion dropdown that allows to choose a time point for exploration. After its selection, the accordion dropdown lays out the list of variables measured at the chosen time point

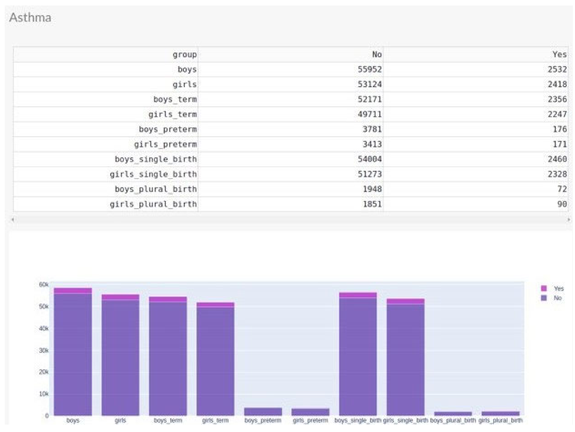
Middle panel: Descriptive statistics within a selected variable for each category. Categories represent participants grouped based on a child's sex, number of children born at a time (single or plural births), and labor conditions (preterm or normal)

Right panel: List of checkboxes to filter categories: groups of participants stratified depending of a child's sex, number of children produced by the same pregnancy, and whether a birth was premature or normal

FIGURE 1: After clicking on "Phenotype" tab in the top panel, selecting "medical birth registry" time point and choosing "Weight of placenta" variable in the left panel, the MoBa explorer interface displays information on this variable showing the corresponding table and chart with box plots in the middle panel for each of the categories checked in the right panel.

NEXT STEPS

- Conduct interviews with users from the general public to understand their needs
- Explore alternatives that may show these data in an equally or more expressive and engaging manner
- Explore intuitive ways to make the data **visually accessible**
- Conduct pilot evaluations of the tool to assess its success across expert and lay audience groups



group	No	Yes
boys	55952	2532
girls	53124	2418
boys_term	52171	2356
girls_term	49711	2247
boys_preterm	3781	176
girls_preterm	3413	171
boys_single_birth	54004	2460
girls_single_birth	51273	2328
boys_plural_birth	1948	72
girls_plural_birth	1851	90

FIGURE 2: Representation of the "Asthma" variable shown in the middle panel as a table and a respective stacked bar chart.

REFERENCES

[1] ELAVSKY F., BENNETT C., MORITZ D.: How accessible ismy visualization? evaluating visualization accessibility with chartabil-ity. In Computer Graphics Forum (2022), vol. 41, Wiley Online Library,pp. 57–70. doi:10.1111/cgf.14522.; [2] FinnGen results — r7.finngen.fi. https://r7.finngen.fi/. [Accessed 07-09-2023]; [3] MAGNUS P., BIRKE C., VEJRUP K., HAUGAN A., AL-SAKER E., DALTVEIT A. K., HANDAL M., HAUGEN M., HØISETH G.,KNUDSEN G. P., ET AL.: Cohort profile update: the Norwegian motherand child cohort study (MoBa). International journal of epidemiology45, 2 (2016), 382–388. doi:10.1093/ije/dyw029.; [4] NUSRAT S., HARBIG T., GEHLENBORG N.: Tasks, tech-niques, and tools for genomic data visualization. In Computer GraphicsForum (2019), vol. 38, Wiley Online Library, pp. 781–805; [5] PENG J., XUE H., HUI W., LU J., CHEN B., JIANG Q.,SHANG X., WANG Y.: An online tool for measuring and visualizing phenotype similarities using hpo. BMC genomics 19, 6 (2018), 89–97.; [6] UK Biobank - UK Biobank — ukbiobank.ac.uk. https://www.ukbiobank.ac.uk/. [Accessed 07-09-2023].

UNIVERSITY OF BERGEN

NIPH
Norwegian Institute of Public Health

EXPLORE OUR PROTOTYPE

ACKNOWLEDGEMENTS

This work is supported in part by the University of Bergen and the Trond Mohn Foundation in Bergen (#813558, Visualizing Data Science for Large Scale Hypothesis Management in Imaging Biomarker Discovery (VIDI) Project).