Data Understanding Report

Utrecht University

Mike Vink

April 27, 2021

Contents

1 Initial data collection

1.1 Technical description data collection

By following the guide on the FluPrint Github Repository the MySQL server was set up. In this work the FluPrint github was first added as a submodule. This module provides the php scripts to import raw data csv's into the MySQL database. The operating system and versions of php and MySQL used in this work were OSX ”Big Sur” (on Mac Book air 2017), php 7.3.24 (built-in mac version), and MySQL 8.0.23 (homebrew).

In the guide the dependencies to run the php import script were installed first. This was also done in this work, except that the hash-file verification step was skipped.

After the php dependencies were installed the MySQL server was started. By default homebrew recommends to use the `homebrew services [option] [SERVICE]` command to start the MySQL server. However, in this work the server is started using `mysql.server start` which provides a socket that was symlinked using `sudo ln -s /tmp/mysql.sock /var/mysql/mysql.sock`. This was done to prevent an error (StackOverflow: cant connect to local mysql server through socket homebrew) thrown by the php import scripts. Before the import scripts were run a user was added to the MySQL server and a database was created 1, the password type had to be `mysql_native_password` (how to resolve [SQLSTATEHY000] 2054 the server requested authentication method.).

Listing 1: Adding user and database to sql server

```
mysql> CREATE USER 'mike'@'localhost' IDENTIFIED BY ';lkj';
mysql> GRANT ALL PRIVILEGES ON * . * TO 'mike'@'localhost';
mysql> ALTER USER 'mike'@'localhost' IDENTIFIED WITH mysql_native_password BY 'mike';
mysql> CREATE DATABASE fluprint;
```

The databasename, the username, and password were added to the `config/configuration.json` of the FlruPrint github module. At this point the configuration for the php import scripts was finished, and the raw data downloaded in `data/upload` were imported in the MySQL server using `php bin/import.php`.

1.2 Data Requirements

The following subsections will list the information required from the data per data mining goals that are needed to answer the following business questions:

- What kind of studies can be done using the FluPRINT database?
- What immunological factors correlate to a high vaccine responses?

1.2.1 Requirements per data mining goal

"Explore and describe the database and corresponding tables."

Falling under this data mining objective are the outputs and tasks related to data collection and description. These comprise a report on the initial collection of the data, selection of data, and description of properties of the data. The data in this case is in a database format, thus here we describe the tables, keys, and attributes in the database, and provide descriptive analyses where possible. The goal is to replicate the description done in A. Tomic, I. Tomic, Dekker, et al., 2019, and to provide a more detailed explanation of the database from a user perspective. Using these descriptions we provide insight into what kind of studies are possible with the database, and why the initial dataset in A. Tomic, I. Tomic, Rosenberg-Hasson, et al., 2019 was chosen.

"Apply standard feature selection methods to the most interesting datasets."

"Fit classification models to the most interesting datasets."

These two data mining objectives were chosen to comprise the data preparation and modelling phases of this project. The authors of fluprint set up an automated machine learning pipeline to investigate the immunological factors that are correlated with a high vaccine response. In this work we use a conventional data mining modelling process to investigate these results.

2 Data description

2.1 Volumetric analysis

In the work of A. Tomic, I. Tomic, Dekker, et al., 2019 data on indiviuals enrolled in influenza vaccine studies at the Stanford-LPCH Vaccine Program was collected, the data was archived at the Stanford Data Miner. This archive was filtered by assays used in influenza studies, resulting in data from 740 healthy donors, enrolled in influenza vaccine studies conducted by the Stanford-LPCH Vaccine Program from 2007 to 2015. These studies are described in the table accompanying the online publication of the fluprint dataset (Table 2). From those 740 donors a vaccine response classification was only given for 372 donors (Figure 1), by a method that will be described in the section describing the data table containing this attribute. Overall there was no major difference in demographic statistics when stratisfying the data in high or low responder classification (Figure 1).

Importantly, it is reported that in all studies the donors are only vaccinated once, except in the study SLVP015 (Table 2) (A. Tomic, I. Tomic, Dekker, et al., 2019). However, in later work of the same authors it is claimed that vaccines are administered as specified by the study (A. Tomic, I. Tomic, Rosenberg-Hasson, et al., 2019).

The donors for which a vaccine respone classification was available from all clinical studies together span a wide age range (Figure 1)A from 1 - 50 (Table 1), in the original work the demographic statistics include the donors for which no vaccine response classification is given, therefore they report a greater range of 1-90. Stratisfying the donors on vaccine response does not affect the demographic attribute distribution, but the maximum age is lowered in the high responders group (Figure 1)B.

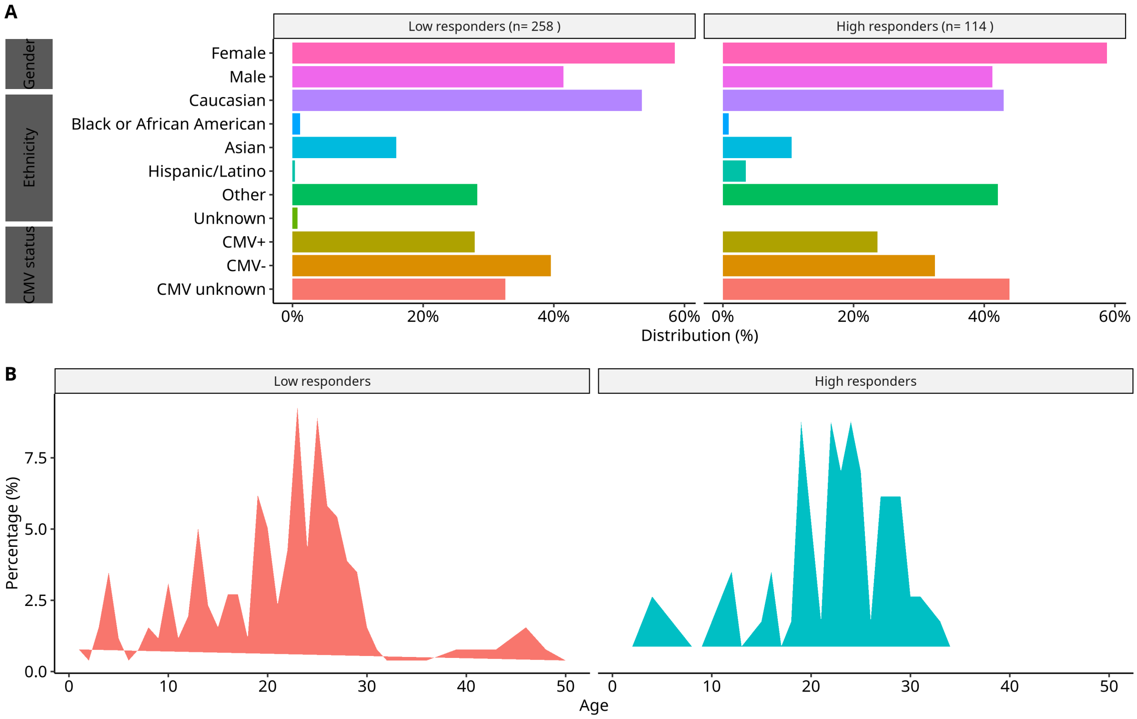

Figure 1: **A.** percentage of donors with factor property within high and low responder groups. Included are sex, race, and CMV status information. **B.** Age distribution of donors with a known response classification.

Age (y)	
Mean ± SD	21.02 ± 8.66
Median (min. to max. range)	22.5 (1 - 50)
Gender	
Male (%)	154 (41.4)
Female	218 (58.6)
Ethnicity	
Caucasian (%)	187 (50.3)
African American (Black) (%)	4 (1.1)
Asian (%)	53 (14.2)
Hispanic/Latino (%)	5 (1.3)
Other (%)	121 (32.5)
Unknown (%)	2 (0.5)

Table 1: **Demographic statistics of donors with known vaccine response classification.**

Stanford study ID	Name	Description	Vaccines	Data in FluPRINT
SLVP015	Comparison of immune responses to influenza vaccine in adults of different ages (2007-2017)	Who: 18-100yo healthy participants How: immunized annually with the seasonal inactivated influenza vaccines from 2007-2017 When: Blood samples acquired before immunization (Day 0), on days 6-8 and 28 after immunization	2007-2013 Seasonal trivalent, inactivated influenza vaccines (Fluzone) 2014-2015 High Dose trivalent Fluzone for participants *geq* 65yo and quadrivalent Fluzone for younger participants	135 donors Assays: 51-plex Luminex 62-plex Luminex MSD 4plex MSD9plex Other Luminex HAI CMV/EBV Hormones CyTOF phenotype Lyoplate Phospho Cytof pheno Phospho cytof phospho Phosphoflow CBCD
SLVP017	B-cell immunity to influenza (2009-2011 and 2013)	Who: 1-2yo (2013), 8-100yo healthy participants who did not receive the seasonal influenza vaccine in previous years (2010, 2011 and 2013) How: immunized with either seasonal inactivated or live, attenuated influenza vaccines in 2009, 2010, 2011 and 2013 When: Blood samples acquired before immunization (Day 0) and on day 28 after immunization	2009-2011 Seasonal trivalent, inactivated influenza vaccines (Fluzone) or seasonal live, attenuated influenza vaccine (FluMist) 2013 Seasonal trivalent inactivated influenza vaccine- (Fluzone) - pediatric formulation for 1-2yo children	153 donors Assays: 51-plex Luminex 62-plex Luminex HAI CMV/EBV CyTOF phenotype CBCD
SLVP018	T-cell and general immune response to seasonal influenza vaccine (2009-2013)	Who: 1-8yo (2013), 8-100yo healthy participants How: immunized with either seasonal inactivated or live, attenuated influenza vaccines from 2009-2013 When: Blood samples acquired before immunization (Day 0), days 7-10 and 28 after immunization	2009-2010 Seasonal trivalent inactivated influenza vaccine (Fluzone) or seasonal trivalent live attenuated influenza vaccine (FluMist) 2010 High Dose trivalent Fluzone for participants *geq* 65yo 2013 Seasonal trivalent, inactivated influenza Pediatric Dose (Fluzone, 0.25 ml) for 1-3yo children	249 donors Assays: 51-plex Luminex 62-plex Luminex MSD 4plex MSD 9plex HAI CMV/EBV Hormones CyTOF phenotype Lyoplate Phospho Cytof pheno Phospho cytof phospho Phosphoflow CBCD
SLVP021	Plasmablast trafficking and antibody response in influenza vaccination (2011-2014)	Who: 8-34yo healthy participants who did not receive the seasonal influenza vaccine in previous years How: immunized with either seasonal inactivated influenza vaccines, given intramuscularly or intradermally, or live, attenuated influenza vaccines from 2011-2014 When: Blood samples acquired before immunization (Day 0), days 6-8 and 24-32 after immunization	2011-2014 Seasonal trivalent inactivated influenza vaccine (Fluzone) given either intramuscularly or intradermally 2011-2012 Seasonal trivalent live attenuated influenza vaccine (FluMist)	84 donors Assays: 51-plex Luminex 62-plex Luminex HAI CMV/EBV Hormones CyTOF phenotype Phospho Cytof pheno Phospho cytof phospho Phosphoflow CBCD
SLVP024	Protective mechanisms against a pandemic respiratory virus (2012)	Who: 2-9yo healthy participants How: immunized with the seasonal live, attenuated influenza vaccine When: Blood samples only from 18-2yo adults acquired before immunization (Day 0), days 7 and 28 after immunization	Seasonal live, attenuated influenza vaccine (FluMist)	Donors: 8 Assays: HAI Phosphoflow
SLVP028	Genetic and environmental factors in the response to influenza vaccination (2014-2018)	Who: 12-9yo healthy participants How: immunized with either seasonal inactivated or live, attenuated influenza vaccines from 2014-2018 When: Blood samples acquired before immunization (Day 0), days 6-8 and 28 + 7 after immunization	Seasonal quadrivalent inactivated influenza vaccine (Fluzone) or seasonal quadrivalent live attenuated influenza vaccine (FluMist)	Donors: 52 Assays: 62-plex Luminex HAI CMV/EBV Hormones CyTOF phenotype
SLVP029	Innate and acquired immunity to influenza infection and immunization (2014-2017)	Who: 6 mo-49yo healthy participants (who did not receive LAIV in the prior season nor received influenza immunizations in two or more prior seasons) How: immunized with either seasonal inactivated or live, attenuated influenza vaccines from 2014-2017 When: Blood samples acquired before immunization (Day 0), days 7 and 28 after immunization. Children ¡9 yrs received 2 immunizations with the second blood samples acquired 28 days after second immunization	Seasonal quadrivalent inactivated influenza vaccine (Fluzone) or seasonal quadrivalent live attenuated influenza vaccine (FluMist)	Donors: 47 Assays: 62-plex Luminex HAI CMV/EBV Hormones CyTOF phenotype
SLVP030	The role of CD4+ memory phenotype, memory, and effector t cells in vaccination and infection (2014-2019)	Who: 6 mo-10yo healthy participants How: immunized annually with either seasonal inactivated or live, attenuated influenza vaccines from 2014-2019 When: Blood samples acquired before immunization (Day 0), days 7 and 60 after immunization. Children with no prior influenza vaccine received 2 immunizations with the second blood sample acquired 60 days after second immunization	Seasonal quadrivalent inactivated influenza vaccine (Fluzone) or seasonal quadrivalent live attenuated influenza vaccine (FluMist) Seasonal trivalent, inactivated influenza Pediatric Dose (Fluzone, 0.25 ml) for 6-35mo children	Donors: 12 Assays: 62-plex Luminex HAI CMV/EBV Hormones CyTOF phenotype

Table 2: **Reference table of clinical studies** Clinical study ID used (but remapped) in the database, age information, vaccine type information, and assay data types of clinical studies are in the rest of the columns.

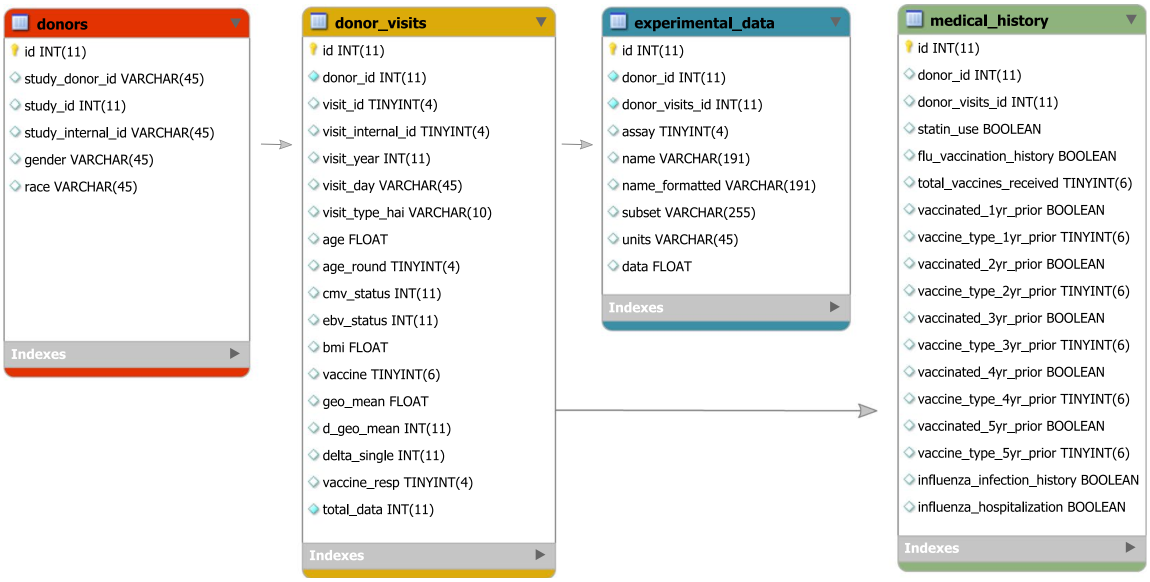

Figure 2: **(taken from original paper)** The FluPRINT database model. The diagram shows a schema of the FluPRINT database. Core tables, donors (red), donor_visits (yellow), experimental_data (blue) and medical_history (green) are interconnected. Tables experimental_data and medical_history are connected to the core table donor_visits. The data fields for each table are listed, including the name and the type of the data. CHAR and VARCHAR, string data as characters; INT, numeric data as integers; FLOAT, approximate numeric data values; DECIMAL, exact numeric data values; DATETIME, temporal data values; TINYINT, numeric data as integers (range 0–255); BOOLEAN, numeric data with Boolean values (zero/one). Maximal number of characters allowed in the data fields is denoted as number in parenthesis.

The data from the clinical studies consisted of 121 CSV files that were imported into the FluPrint database. The data was used to build four tables which will be described in the next sections, but we will not discuss technical validation of the database construction, refer to the original work for that (A. Tomic, I. Tomic, Dekker, et al., 2019). The relation between the tables is best visualised in the original work of (A. Tomic, I. Tomic, Dekker, et al., 2019), it describes the MySql attribute types and columns in the tables (Figure 2) (copied). The volume of the data is also given in the original work, per table the number of rows and columns is reported (Table 3).

2.2 Attribute types and values

Because of the great number of attributes in the database, we discuss them by table starting with the donors (Figure 2).

2.2.1 donors table

The *donors.id* attribute is simply an enumeration of unique donors, importantly, it is used as a key to get attributes from other tables. The column *study_donor_id* is an encrypted identification

Table name	**Rows**	**Columns**
donors	740	6
donor_visits	2,937	18
experimental_data	371,260	9
Medical history	740	18

Table 3: Volume of tables in the Fluprint database.

id	study_donor_id	study_id	study_internal_id	gender	race
1	e27ad74ff9a5f2f32d8e852533f054c0	30	30	Female	Asian
2	4a89ac4d3f4dc869e5c8e8cf862cffda	30	30	Male	Other
3	a2cde6e54dec92422b0427dd49244350	30	30	Female	Caucasian
4	0f7d8d1c13e876017ea465f99d25581f	30	30	Male	Other
5	1ed2f6409584b7b4e9720b28d794fe91	30	30	Female	Caucasian
6	a575678405e9615bfb87eccfa031f7fc	30	30	Male	Other

Table 4: Head of the donors table.

number. Each donor belongs to the study identified by the *study_id*, these are the last two digist of the name code (those starting with SLVP0 ··) in the reference table (Table 2), the *study_internal_id* is either the digit or a string containing the digit in *study_id*. The *gender* and *race* attribute contain the values used in (Figure 1), a minor note is that in the original paper "American Indian or Alaska Native" is listed as one of the *race* values but is not used in the database. There are 5 donors whose race is "NULL", which are mapped to unkown (Figure 1).

2.2.2 donor_visits table

The donor visits table is the core table of the database, it contains donor attributes at visit times during enrolment in clinical studies in rows that are uniquely identified by an *id* integer. Each row also includes the *donor_id* identify the donor that visitted.

The database combines different clinical studies accross years and the data from these studies is incomplete leading to an incomplete and hetergenous database (Table 5). For example some donors might miss their second visit to determine their antibody levels, or the number of parameters measured by an assay changed in the timespan of a clinical study. Unifying these clinical studies in one database resulted in normalised but incomplete data and heterogenous data. More specifically, every attribute in the core table has missing value, which complicates dataset selection. One examples of visit data of a donor is discussed to highlight important attributes and problems in the data: that the number of visits is variable, that all columns are incomplete, and that classification is sometimes based on single visits or inconsistent (Table 6) (Table 5).

Per donor all visits are enumerated in chronological order by *visit_id* (Table 6). Further visit info includes: *visit_internal_id* which is a number that indicates the visit order within an influenza season but this differs per clinical study (e.g. some use 1-2-3, orther use 0-7-28), the *vist_year* is the

stat	age	cmv_status	ebv_status	bmi	vaccine	geo_mean	d_geo_mean	vaccine_resp	total_data
n	2937.0	1081.0	548.0	516.0	2794.0	984.0	1260.0	1206.0	2937.0
na	0.0	1856.0	2389.0	2421.0	143.0	1953.0	1677.0	1731.0	0.0
mean	47.3	0.4	0.8	24.8	3.7	87.6	8.9	0.3	126.4
sd	27.0	0.5	0.4	5.6	1.0	101.7	30.9	0.4	368.4
se_mean	0.5	0.0	0.0	0.2	0.0	3.2	0.9	0.0	6.8
IQR	50.2	1.0	0.0	6.7	0.0	105.4	4.0	1.0	19.0
skewness	0.2	0.3	-1.4	1.0	-1.7	3.6	9.9	1.1	7.1
kurtosis	-1.5	-1.9	-0.1	2.1	3.0	26.6	114.9	-0.9	49.7

Table 5: Descriptive stats of relevant numeric or binary factor columns in the donor visits table. For geo_mean 0 is considered as missing data.

influenza season of the visit, the *visit_day* is the number of days relative to the date of vaccination, *age* and *age_round* indicate the donor's age at time of the visit, and *bmi* gives the donor bmi at visit time, and lastly *visit_type_hai* is the intent of the visit which is either "pre", "post", or "other",

During the "pre" visit a virological assay is performed to determine the CMV and Epstein-Barr virus (EBV) status of the donor, which are indicated by the binary variables *cmv_status* and *ebv_status*.

To measure vaccine response to a vaccine which is indicated by an id (Table 9) in *vaccine*, the hemagglutination inhibition assay (HAI assay) is used. The procedure measures the influenza antibody titers before vaccination during the *visit_type_hai* "pre" visit of a participant, and 28 days after vaccination during a "post" visit. The geometric mean titer (GMT) at each visit is calculated, and a fold change in GMT is calculated as the ratio of the GMT at day 28 (post) and during the first visit (pre). These values are *geo_mean* and *d_geo_mean*, *d_single* is the antibody titer fold-change per strain of virus used in the vaccine, it is unclear how this value is aggregated over different strains and is left out of further analysis. This data was used to classify donors in high or low responders according to FDA guidelines, individuals are high-responders if they seroconverted (4-fold or greater rise in HAI titer) and were seroprotected (GMT HAI $\geq$ 40) after vaccination. The seasonal vaccine response classifications are given by the binary variable *vaccine_resp*.

The assays performed to get a serological/immunlogical profile of the donor before vaccination are described later in the section of the experimental data table, all assays are listed in the original work A. Tomic, I. Tomic, Dekker, et al., 2019 and are summarised here (Table 7), the total rows of assay data is given by *total_data*.

The most important data related to the visits of donor 166 is shown in Table 6. The vaccine response classification is calculated based on the GMT in the "pre" and "post" visits. This classification is done per influenza season, but the HAI assay requires a "pre" visit and a "post" visit 28 days later to measure the difference in GMT. However, sometimes a classification is given when there is only one visit record in a season, like in 2012 for donor 166 (Table 6).

The example of donor 166 contains an inconsistency in the classification, in 2011 the GMT *geo_mean* increases from 25.20 to 160.00, and the *d_geo_mean* is 6, but in this season the donor is wrongly classified as a low responder (Table 6). Because of this the seasonal classification of donors was investigated using the seroprotection and seroconversion criteria **??**, records of incorrectly

visit_id	year	day	type	age	cmv	ebv	bmi	vaccine	geo_mean	d_geo_mean	response	assay_data_rows
1	2011	0	pre	20	1	1	30.31	4	25.20	6	0	343
2	2011	7	other	20	1	1	NULL	4	0.00	6	0	51
3	2011	28	post	20	1	1	NULL	4	160.00	6	0	51
4	2012	0	pre	21	1	1	30.31	4	9.28	4	0	292
6	2013	0	pre	22	1	1	30.31	4	15.91	2	0	2877
7	2013	7	other	22	1	1	NULL	4	0.00	2	0	63
8	2013	28	post	22	1	1	NULL	4	26.75	2	0	82

Table 6: Visit data of donor 166 from study SLVP021 (Table 2), where participants are only vaccinated once. Number of visits and data collected at visit varies, classification is inconsistent with ≥ 40 and 4-fold increase rule in 2011.

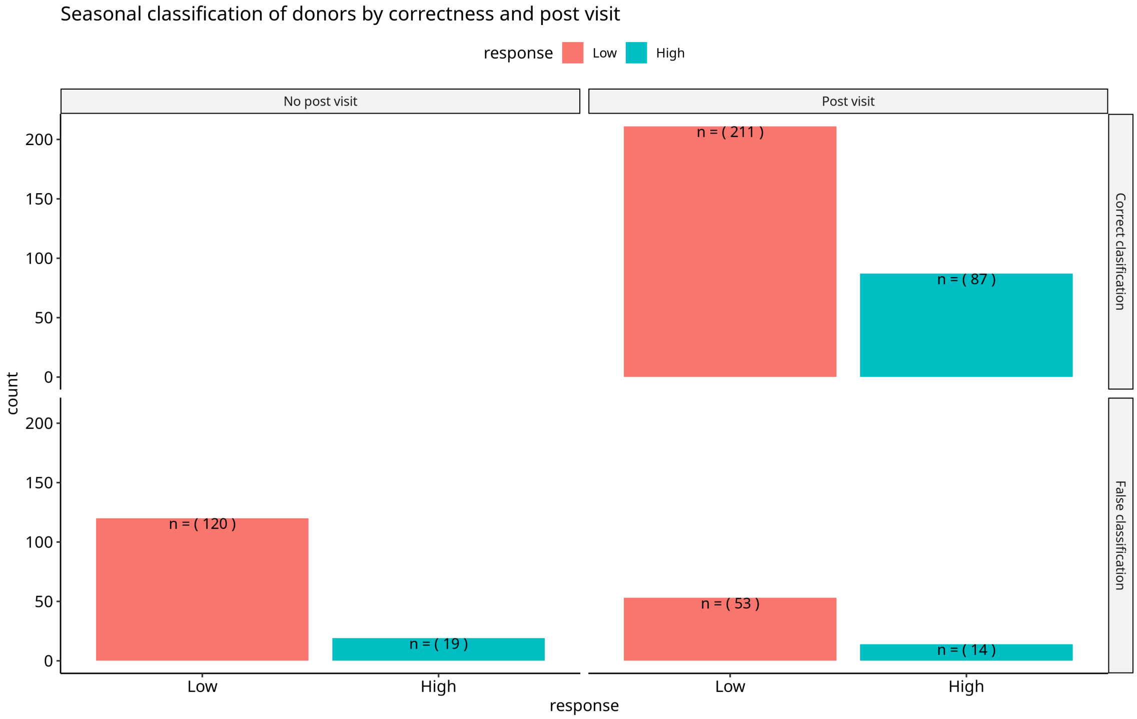

Figure 3

Name	Description	id (*experimental_data.assay*)
(Multiplex) cytokine assays	Multiplex ELISA using Luminex polysterene bead or magnetic bead kits. Measures serum cytokine/hormone level in z.log2 units using fluorescent antibodies.	3, 6, 15, 16
Flow and mass cytometry assays	uses labeled antibodies to detect antigens on a cell surface to identify a subset of a cell population, units are in percentage of parent population.	4, 9, 13, 17
Phosphorylation cytometry assays	Uses antibodies to measure phosphorylation of specific proteins stimulated by an immune system event belonging to cell population subsets. Units are a fold change between stimulated and unstimulated cells, for mass cytometry arcsin readout difference, fold-change of 90th percentile readout values otherwise.	7, 10 (mass cytometry) (flow cytometry)
complete blood count (CBCD)	Different cells are counted using flow cytometry Units are usually in Count/μL	11
meso scale discovery assays (MSD)	A setup where serum cytokines or hormones are captured with antibodies, and then detected by using a detection antibody. Units are arbitrary intensity	2, 12, 14

Table 7: assays table

labelled donors are also saved as a spreadsheet.

2.2.3 Experimental data table

Assays performed in visits are remapped, but the values in the database do not correspond to the reported table (Table 9). Actual assay type, data units, and id in the database are reported here (Table 7).

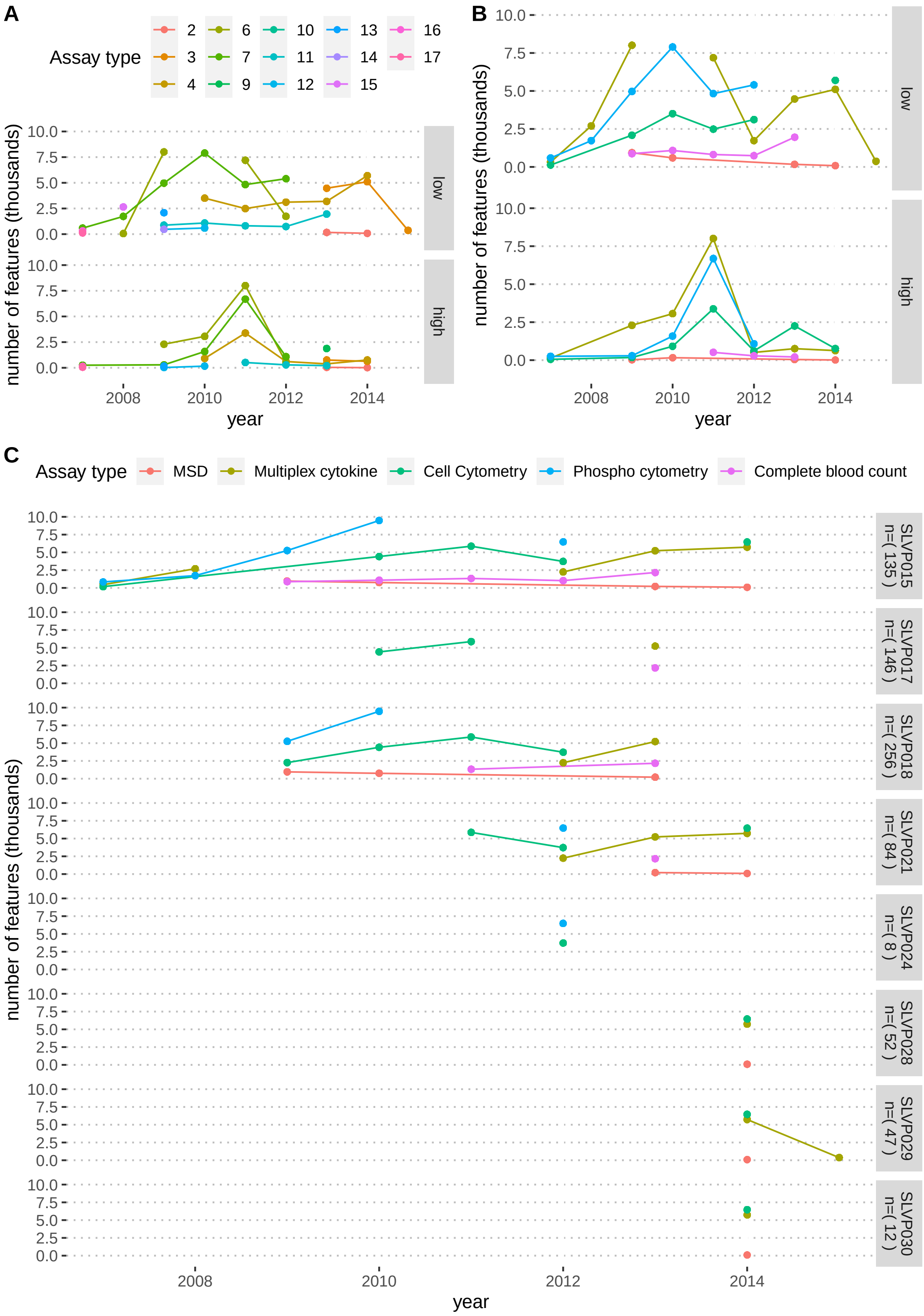

Table 8: **Feature count per individual assay id, assay type, stratified in either response status or study** caption

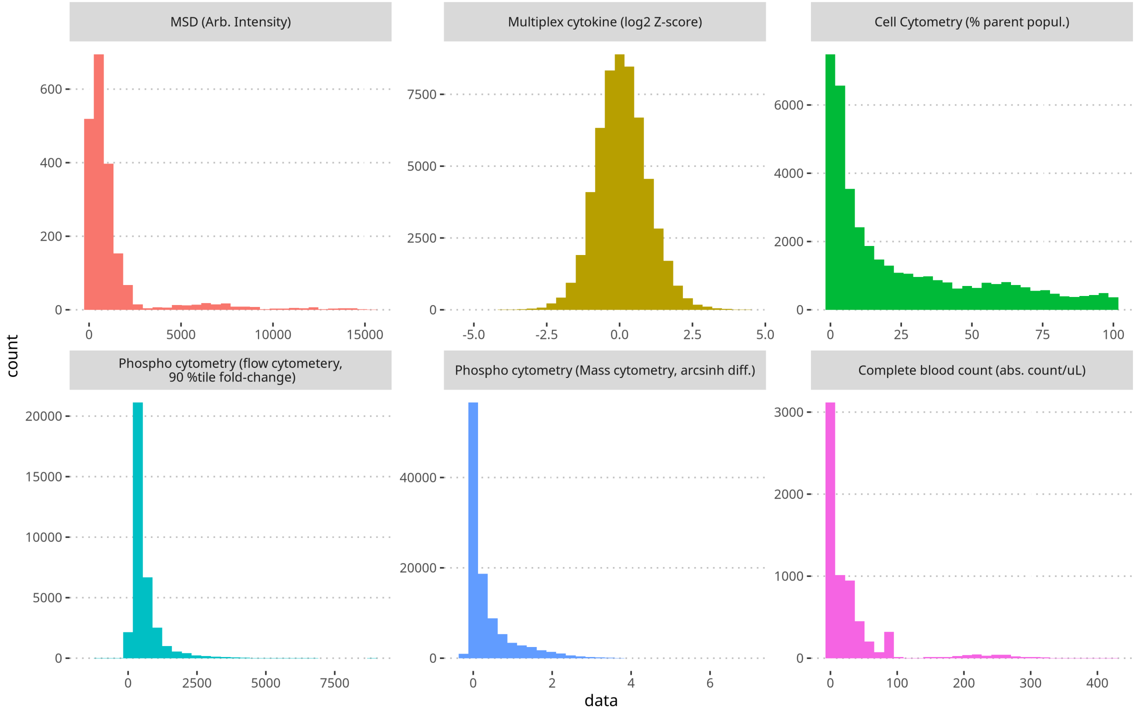

Figure 4

In total there are data from 14 different assays, not counting the virological and HAI antibody assays (Table 7). The virological assays include the cmv virus status and ebv status, and is not used in this work because it is done in a smaller subset of studies. Those 14 assays have been aggregated in this work to 5 different types of experiments: the multiplex assays measure serum molecules such as cytokines and other signaling molecules, flow and mass cell cytometry measure the phenotype of specific immune related cells, phosphorylation flow and mass cytometry measures the phosphorylation signaling pathway activation after an immune stimulation, the blood count measures the count of cells in the blood, and meso scale discovery (MSD) measures hormones or cytokines from the blood.

The experimental data table contains all features recorded for a donor visit. The number of features collected for each visit is large and varies greatly (mean at 126 , ±368 SD) (Table 5), and in total there are 3285 different features measured across all clinical studies. However, not every assay is done in every clinical study (Figure 8) and over the years the data generated by assays has changed, so a table with all features as columns and all donors as rows would be extremely sparse (and crashes R due to RAM limitations). Describing the 3285 different features in this sparse table would be impossible, but assay value distributions across studies are shown to follow normal or power distributions (Figure 4). Global correlation analysis is complicated by the great number of features and sparseness in the data.

What further complicates selecting data is repeat visits of donors, and missing visits. The problem of repeat visits over a span of multiple influenza seasons is that not the same assay types

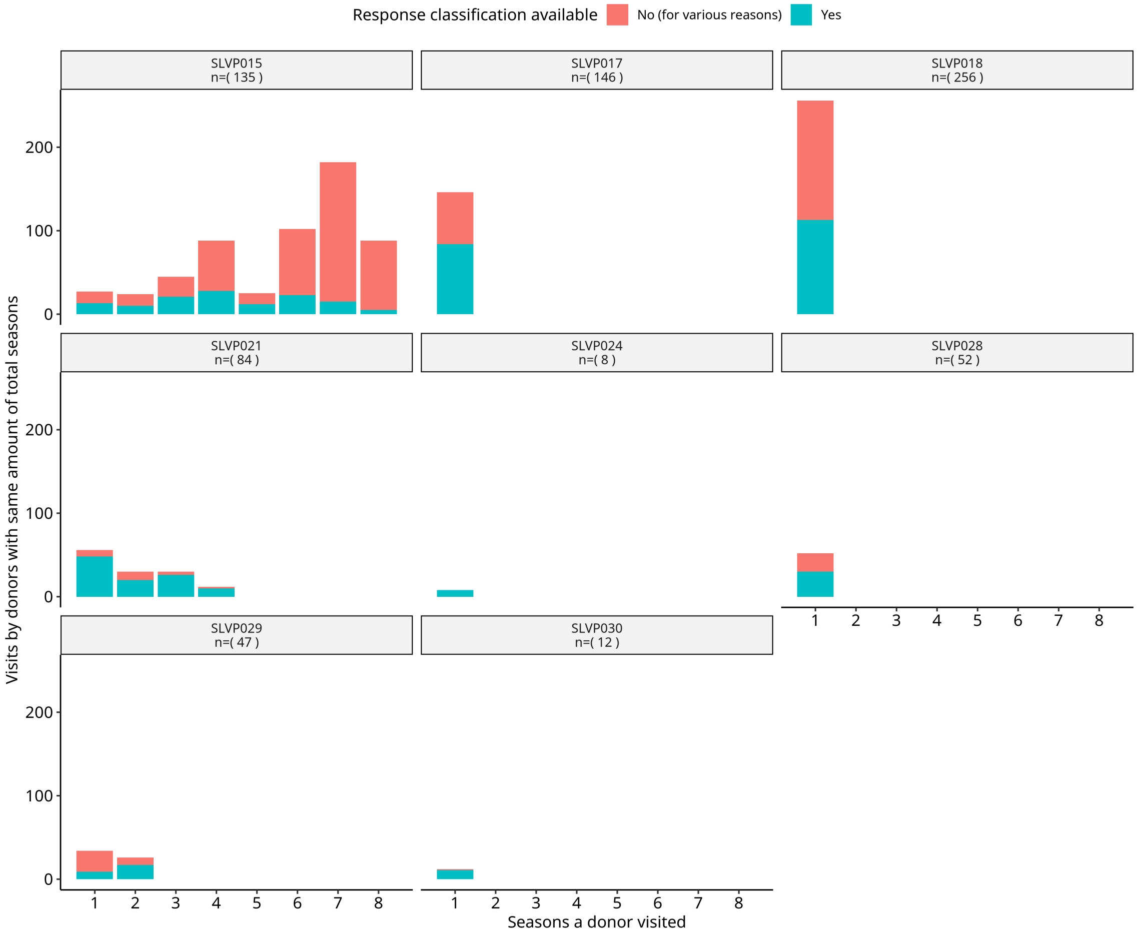

Figure 5: the number of donors that visited per number of influenza seasons they visited (years), per study. The color indicates the number of visits for which a classification was available, counted within the groups of donors that visited the same amount of times.

are done, and that repeat visits are only a small portion of the database. The data is also not suitable right away for studying the effect of repeat vaccination on high versus low vaccine reponse, since the classification in the longitudal study (SLVP015) is mostly not available (Figure 5).

For example exploring the effect repeat vaccination has an response rate would first require manual labelling of high and low responses, at least for the cases where it is possible based on the GMT data. Those cases are when classification is set to a null value even though GMT data is available. The reason for this null value assignment is reported, but the pattern seems to set the vaccine response to null if there is not enough assay data measured.

3 Data quality

The database has issues that are inherent to combining multiple studies and the classification is inconsistent in some cases (Figure 3), or often missing completely because no HAI antibody assay data was available or the classification was set to a null value by the database authors (Figure 5). The main value of the database is the assay data that is fully represented in all studies and across all years, but this information is hard to access since all studies do not use overlapping assays (Figure 8), resulting in high sparsity data. Further, the sample size that can be used for further studies is limitted, since the high versus low vaccine response is only available for a small subset of the data.

Specific attributes that have great amounts of missing values are the virological and HAI assay data, the last is used for the vaccine response classifcation. Potential for Studying the correlation of these values with vaccine response is thus limitted. Nevertheless assay data is often available and could be used to identify immunological factors that correlate with other data, such as repeat vaccination, the exploration of this effect is outside the scope of this work due to the data sparsity issues.

References

Tomic, Adriana, Ivan Tomic, Cornelia L. Dekker, et al. (Oct. 2019). “The FluPRINT Dataset, a Multidimensional Analysis of the Influenza Vaccine Imprint on the Immune System”. English. In: *Scientific Data* 6.1, p. 214. ISSN: 2052-4463. DOI: 10.1038/s41597-019-0213-4.

Tomic, Adriana, Ivan Tomic, Yael Rosenberg-Hasson, et al. (Feb. 2019). “SIMON, an Automated Machine Learning System Reveals Immune Signatures of Influenza Vaccine Responses”. English. In: *bioRxiv*, p. 545186. DOI: 10.1101/545186.

Appendices

A Remaps used in the database

Vaccine received	Vaccine type ID	Vaccine type name
FluMist IIV4 0.2 mL intranasal spray	1	Flumist
FluMist Intranasal spray	1	Flumist
FluMist Intranasal Spray 2009–2010	1	Flumist
FluMist Intranasal Spray	1	Flumist
Flumist	1	Flumist
Fluzone Intradermal-IIV3	2	Fluzone Intradermal
Fluzone Intradermal	2	Fluzone Intradermal
GSK Fluarix IIV3 single-dose syringe	3	Fluarix
Fluzone 0.5 mL IIV4 SD syringe	4	Fluzone
Fluzone 0.25 mL IIV4 SD syringe	5	Paediatric Fluzone
Fluzone IIV3 multi-dose vial	4	Fluzone
Fluzone single-dose syringe	4	Fluzone
Fluzone multi-dose vial	4	Fluzone
Fluzone single-dose syringe 2009–2010	4	Fluzone
Fluzone high-dose syringe	6	High Dose Fluzone
Fluzone 0.5 mL single-dose syringe	4	Fluzone
Fluzone 0.25 mL single-dose syringe	5	Paediatric Fluzone
Fluzone IIV3 High-Dose SDS	6	High Dose Fluzone
Fluzone IIV4 single-dose syringe	4	Fluzone
Fluzone High-Dose syringe	6	High Dose Fluzone

Table 9: Remaps of vaccine type relevant to to the clinical studies reference table (Table 2), and the section on the donor visits table.

Original	Remapped
No	0
Yes	1
IIV injection/im	2
Doesn't know/doesn't remember/na/does not remember	3
LAIV4 intranasal/laiv_std_intranasal/laiv_std_ intranasal/nasal/intranasal	4

Table 10: caption

Original	Remapped
CMV EBV	1
Other immunoassay	2
Human Luminex 62–63 plex	3
CyTOF phenotyping	4
HAI	5
Human Luminex 51 plex	6
Phospho-flow cytokine stim (PBMC)	7
pCyTOF (whole blood) pheno	9
pCyTOF (whole blood) phospho	10
CBCD	11
Human MSD 4 plex	12
Lyoplate 1	13
Human MSD 9 plex	14
Human Luminex 50 plex	15
Other Luminex	16

Table 11: caption