Myocarditis Redcap wiki

Charles Dolladille

2022-11-10

Contents

Chapter 1

About

This is a reference document that describes the data structure and variables and the common data management schemes used for the analysis of the Myocarditis Redcap from Vanderbilt University.

1.1 Usage

This document is intended as a cheatsheet, for fast and easy reminder of data structure and variable names, and how to use them.

1.2 Target users

Cardiologist and data manager researchers from Drs Moslehi and Salem teams.

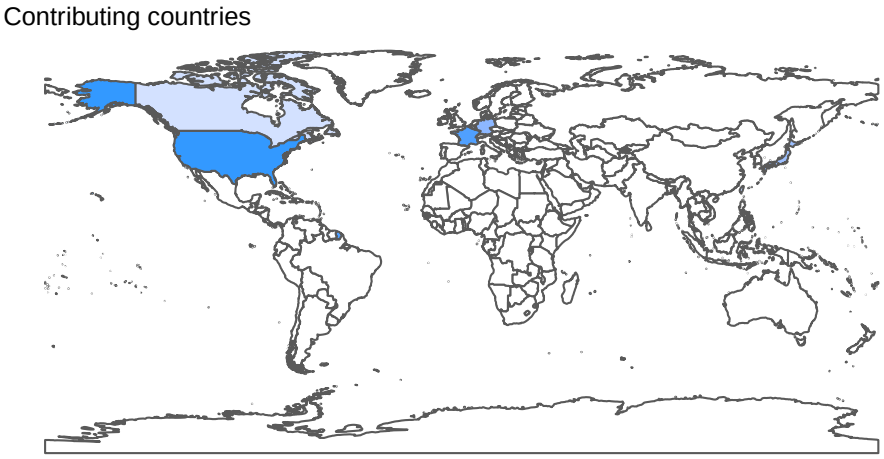
Chapter 2

Introducing the database

This is a VERY brief introduction. The database was created in 2018, and gather more than 700 cases of immune checkpoint inhibitors associated cardiovascular adverse reactions (as of September 2022). Cases are reported from all around the world, and most of them are myocarditis cases.

> October 2022 update. It was decided to remove all non-myocarditis cases from the base. Some cases were removed after a careful semi-automated review, which also tracked duplicates.

If your not familiar with the Redcap data structure and how data can be exported, you should look at the Redcap Help & FAQ. We will not deal with these concepts here.



2.1 What is in the Redcap?

In the Redcap you may find an extensive characterization of the cases, including clinical work-up and outcomes. These features are grouped according to time points (baseline, index date, follow-up…) or critical exams (EKG, biopsy…) that are called **instruments** in Redcap. Each instrument is described in a separate chapter in this document.

2.2 Variables naming rules

Variable names are attributed according to standardized rules as follow:

- A variable name always starts by the instrument identifier (e.g. `p_` for demographic data, see below Data structure), with the exception of date `date_`and time `ti_` variables (see multi-instrument variables).
- In case an instrument is subdivided into multiple instruments, the instrument identifier is followed by the subinstrument identifier (e.g. `ic_ca` for index cardiotoxicity cancer treatment)
- Calculated vars have the same name as manual vars, with an additional suffix `__c`.
- Free text vars have the same name as their branching logic displayer, with an additional suffix `__ft`.

- Descriptive vars have the same name as their branching logic displayer, with an additional suffix `__desc`. Note that these vars do not contain any data and are not usually exported.
- All variables names are lower case, with words separated by underscores (e.g. you will not see upper case vars such as `AGE_VAR`). This is also known as `snake_case` format.
- As much as possible, we use singular rather than plural in the names (e.g. `antimetabolite` rather than `antimetabolites`)
- Variables with the `__old` suffix are former versions that will be integrated in v3 of the database. They contain the same type of information, but often have different levels (e.g. an additional "missing" level).

We translate here a general description of the variables. To get a precise definition of a variable, you will be more inspired looking at the Redcap's codebook.

Please note this document is NOT an exhaustive list of the variables in the redcap and is not intended to be.

2.3 Data structure

The Redcap instruments are:

Instrument	Identifier	Description
Admin	[ad]	Identify the reporting source, the reporter, with mailing contact information.
Demographics	[p]	Patient demographics (age, sex…), medical history (including cardiovascular history), and medications.
Baseline EKG	[be]	If any, description of a baseline Electrocardiogram (prior to cardiotoxicity).
Previous line	[pl]	Features of prior anticancer treatments, including those of older cancers and the current cancer in prior lines of treatment.
Current line	[cl]	Features of the current line of anticancer treatment containing immunotherapy(ies). Also, features of the current cancer (actively treated).
Index cardiotoxicity	[ic]	All clinical work-up of the patient when presenting for cardiotoxicity. This is the main instrument.
Index EKG	[ie]	It is an instrument by itself, as it gathers one or several EKGs, hence there is a large quantity of data here.
Index hospitalization	[ih]	Features that occurred during the hospitalization that followed cardiotoxicity diagnosis.
Follow-up	[fu]	Long term outcomes.

Instrument	Identifier	Description
Biology	[b	All biological features (transversal instrument)

2.4 Multi-instrument variables

Sometime, an event can occur at different times. For example, a patient can experience death during its index hospitalization, or later in follow-up. To ensure a proper identification of events according to the research purpose, a separate chapter is dedicated to theses variables including timings.

2.5 Root variables

When a question can have multiple non-exclusive answers, the Field Type in Redcap is "checklist".

For example: which alkylating agent(s) was(were) used?

- Cisplatin
- Carboplatin

Suppose the dots are checkboxes and you may select either of them.

In this case, Redcap creates subvariables in the data extraction. If the variable name is `p_pl_alkylating_agent`, then you will NOT find it in the extraction. You will have 2 subvariables named `p_pl_alkylating_agent___1` and `p_pl_alkylating_agent___2`.

> Root variables are those variables that are splitted into multiple subvariables, i.e. which have the Field Type argument "checkbox".

You may want to merge back these variables together, to create a global variable: "Has the patient been treated with any alkylating agent?".

Chapter 3

Admin

Admin variable identifier is [ad]

It is the first chapter but actually, administrative variables are quite easy to use.

There is a few logical branching here. If you want to create a country-level variable, you will need to gather back several subvariables including `reporting_region`, (USA are a stand alone in this variable) `european_countries`, `african_countries`, `asian_countries`, `australia_countries`. These variables are mutually exclusive (except for reporting_region).

Useful when you want to draw the world map of contributing countries.

Please look at the contributors and contributing centers section for more details.

Chapter 4

Demographics

Demographics variable identifier is [p] for patient

4.1 Abbreviations

Here is a list of abbreviations used in several variables in the demo instrument.

Abbreviation(s)	Feature
`cancer_hx`	Cancer history: refers to prior cancer(s) the patient had in his/her
`cx`	life before the one actively treated with immune checkpoint inhibitors.
	`cx1`, `2`, `3`… Refer to features for specifics types of cancer (1 is Bladder cancer, 2 is Breast cancer, etc.)
`pl`	Previous line: this is a concept to refer to treatments received either for prior cancer(s) **or the current cancer** in previous lines
	Obviously falls inbetween two instruments (demographics and current cancer)
`autoimmune_hx`	Auto-immune history: prior auto-immune disease
`ai`	
`organ_transplant`	Organ transplantation history
`ot`	

4.2 Calculated variables

You can check the difference between calculated and manual vars here.

manual var	calculated var
p_age	p_age__c
p_bmi	p_bmi__c

4.3 Derived variables

4.3.1 Categorical Body Mass Index and overweight or obese

For the latter, read as is patients BMI>25 or not?

```
p_bmi_cat = expr(
  case_when(
      p_bmi < 18.5 ~ "underweight",
      dplyr::between(p_bmi, 18.5, 25) ~ "normal_weight",
      dplyr::between(p_bmi, 25, 30) ~ "overweight",
      p_bmi > 30 ~ "obese"
    ))

p_overweight_obese = expr(
  if_else(
      p_bmi > 25,
      1,
      0
    )
)
```

4.3.2 Cardiovascular risk factors

At least one traditional cardiovascular risk factor

```
p_cv_risk = expr(
  pmax(
      p_lipid_chol,
      p_tobacco_anytime,
      p_diabetes2,
      p_hypertension,
      na.rm = TRUE
    )
  )
```

4.4 Medications

There is always an ambiguity with checkboxes: if nothing is checked, you can't tell if the answer is negative to your question, or if the data is missing.

For example, if you have two checkboxes for medications

- Beta-blockers
- Aspirin

And say the first is checked and the second is not. In this case, its easy to say the patient is taking betablockers, but what about aspirin? Maybe the patient is not taking aspirin, or maybe data is missing and the user did not checked the box. This could be the case if a patient addressed to the hospital forgets to bring his/her prescription and do not remember precisely the name of the medications. On the other hand, you may often retrieve the complete list by calling the pharmacist or if a related finally brings the prescription during the hospital stay.

> It is assumed that medication list is complete (no missing data) if the `p_meds_any` variable is checked.

Chapter 5

Current line

Current line identifier is [cl]

5.1 Definition

The current line is the most recently administered cure of anticancer drugs that contains immune checkpoint inhibitors. It can be ongoing at the time of index cardiotoxicity, or may have been interrupted.

5.2 Abbreviations

Abbreviations for immune checkpoint inhibitor drugs and classes are the same as in the previous line instrument.

Here is a list of abbreviations used in several variables in the current line instrument.

Abbreviation(s)	Feature
`1stdrug`	Refers to the 1st infusion of the drug
`lastdrug`	Refers to the last infusion of the drug

Chapter 6

Index cardiotoxicity

Index cardiotoxicity identifier is [ic]

There is a little ambiguity here: index cardiotoxicity refers both to the date of cardiotoxicity diagnosis AND date of hospital admission (if any) for this cardiotoxicity.

6.1 Abbreviations

Abbreviation(s)	Feature
Clinical work-up	
`sy`	Symptoms
`ctx`	Cardiotoxicity (used for the description of timings between a ctx symptoms `ctx sy` and index date, and grading).
`pe`	Physical examination features
`irae`	Features of non-cardiac immune related adverse events at the index date
`hyper, hypo, aiha, attp, hus, aa, id, bd, scar`	For iraes subclasses, small adds to refer to the correct subclass (hyperthyroidism, hypothyroidism, auto-immune hemolytic anemia, acquired thrombotic thrombocytopenic purpura, hemolytic and uremic syndrome, aplastic anemia, inflammatory dermatitis, bullous dermatosis, severe cutaneous adverse reaction)
Exams	
`bx`	Biopsy features (muscular, cardiac, other)

Abbreviation (F)	Feature
`lhc`	Left Heart Catheterization
Treatment section	
`tx`	
immsup	Immunosuppressant 4-character abbreviation for immunosuppressants
`ivig`	Intravenous Immunoglobulin
`plas`	Plasmapheresis (PLEX)
`tacy`	Tacrolimus or cyclosporin
`myco`	Mycophenolate mofetil
`othr`	Other immunosuppressant
`athg`	ATG / thymoglobulin
`infl`	Infliximab or other anti-TNFalpha
`abab`	Abatacept or belatacept
`mtor`	mTOR inhibitors (temsirolimus, everolimus, sirolimus)
`cycl`	Cyclophosphamide
`azat`	Azathioprine
`ritu`	Rituximab
`ms`	Mechanical support
`aa`	Anti-arrhythmic drugs or devices

6.2 Biological features

Most normal values of biological parameters are different from one lab to another. Also, users may report bio vars according to different scales.

Examples: troponin upper limit of normal may range from ~ 0.01 to ~ 50. Complete blood count can be reported in G/L or 10^-3/µL (even if the var field asked for reporting with one of them).

It is advised to express biological parameters according to their normal value range if known, or to contrast them into ratios that get rid of the unit (example: neutrophil to lymphocyte ratio)

Chapter 7

Index hospitalization

Index hospitalization identifier is [ih]

Features occurring early after index cardiotoxicity. It encompasses repeated EKG, echos, and complications.

7.1 Abbreviations

Abbreviation(s)	Feature
`wstecho`	The worst echocardiographia (the one with the lowest Left Ventricular Ejection Fraction)
`wstmri`	The worst Cardiac Magnetic Resonance Imaging (note that it might not be the most demonstrative MRI, according to the definition)
Outcome section	
`ou`	
`o1, o2`	Organ 1, 2, 3, affected by sepsis
`rf`	Respiratory failure

Chapter 8

Previous line

Previous line identifier is [pl]

8.1 Definition

Instrument previous line refers to any anticancer treatment received, either for a prior cancer **or for the current cancer**, before the current line.

The current line is the most recently administered one that contains immune checkpoint inhibitors. It can be ongoing at the time of index cardiotoxicity, or may have been interrupted.

8.2 Abbreviations

Here is a list of abbreviations used in several variables in the demo instrument.

Abbreviation(s)	Feature
`atez`	Standardized **4-character** abbreviations of immune checkpoint inhibitors
`avel`	Atezolizumab, avelumab, cemiplimab, durvalumab, nivolumab, ipilimumab, pembrolizumab, tremelimumab
`cemi`	
`durv`	
`nivo`	
`ipil`	
`pemb`	
`trem`	

Abbreviation(s)	Feature
`pd1`	When drug name is not know, standardized abbreviations related to ICI class.
`pdl1`	Anti-PD1, anti-PD-L1, anti-CTLA4
`ctla4`	Also `pd1_other`, `pdl1_other`, and `ctla4_other` to refer to an ICI from a class that is not in the aforementioned list (e.g. a new anti-PD1)

Chapter 9

Common use cases

9.1 Calculated versus manual vars

Many numeric variables can be entered in two ways:

- **calculated var:** The user provides dates (e.g., date of birth, date of index cardiotoxicity), and the variable is automatically calculated from these dates (e.g. age as the difference between index cardiotoxicity date and date of birth).
- **manual var:** The user could not provide dates, but had a free text field to input the value.

As a result, 2 variables contain the same data. They are not mutually exclusive, which means an observation can have a value for the 2 variables (e.g. the user provided both dates, which allows for automated computation in the calculated var, and also entered age in the manual var free text variable). There has to be a rule of thumb to choose which variable is to be used in the analysis. Here is ours:

> Calculated vars are preferred over manual vars

This means that, for a single case:

- If the calculated var is available, it will be retained.
- Else, if the manual var is available, it will be retained.
- Else, if none are available, the value is missing.

Here is an implementation of this simple logic into an R function

```
numvar_uni <- # Numeric variables unifier
  function( # used to organize data entered from 2 variables, currently its a prioriti
    var1, # a quasi quoted name of column from data. Usually, one is the calculated va
    var2 # also a quasi quoted name
    # underlying data.frame data argument is omitted
  ){
    var1 <- rlang::enexpr(var1)
    var2 <- rlang::enexpr(var2)

    ex <- rlang::expr(dplyr::case_when(
      !is.na(!!var1) ~ as.numeric(!!var1),
      !is.na(!!var2) ~ as.numeric(!!var2),
      TRUE ~ NA_real_
    ))

    ex
  }
```

9.2 Calculated and manual vars identifiers

> For a manual var, the associated calculated var has the `__c` suffix.

Example:

manual var	calculated var
`p_age` (patient age, from instrument admin)	`p_age__c`

9.3 Free text variables

Users are often provided additional fields in the case their patient falls out of the checkboxes. For example, a patient may have experienced an auto-immune disease that is not listed in the `p_ai_` vars.

In this case, the user can check the `p_ai_other` box. When this box is checked, the `p_ai_other__ft` variable is displayed. It is a free text field where the user can input additional data (e.g. Bullous pemphigoid).

> Free text vars have the same name as the branching logic displayer, with a `__ft` suffix.

At the moment, data from free text variables is not used to compute additional variables for 2 main reasons:

- There are few data in these vars
- Data quality checking requires additional time consuming ressources that cannot be applied to a general framework

  We recommend you use free text variables only if their data closely match your research question.

9.4 Sex

At this stage, it is a binary variable with value

- 1 for man
- 2 for woman

It is usually easier to use a variable named “man” or “woman”, built from the sex variable, to remember how to interprate model results.

Chapter 10

Multi-instrument variables

10.1 Date and time variables

As dates and times are used across instruments to capture delays between an event and another (e.g. between index cardiotoxicity and death), there are attributed a specific naming scheme:

- They do not start by the instrument identifier
- They have their own identifier `date` for dates and `ti` for times.
- Time variables always have 3 components:
  - The `ti` identifier
  - The first event identifier
  - The second event identifier

There is an exception to this naming rule: time between 2 cures of ICI (either previous line or current line), which has the long suffix `regimen_freq`.

For example, time between index cardiotoxicity and death is `ti_ic_death`. An event identifier can have a middle `_` if its very long (e.g. `cl1stctla4_other` for current line first dose of an unspecified anti-CTLA4, although this is quite a rare case.

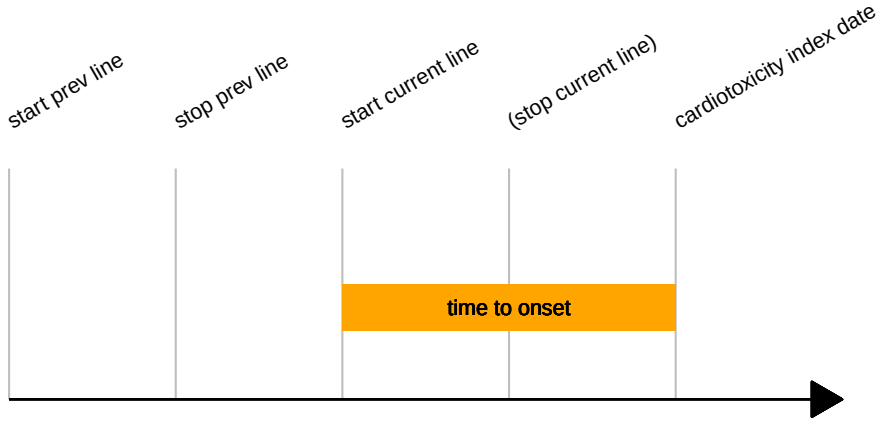
> Time variables represent **delays** between two dates.

Of note, times can be expressed in days, weeks or years.

var	instrument	unit	quick desc
ti_tobaccoquit_ic	demo	years	From tobacco quitting to IC
ti_otx_ic	demo	years	From organ transplant x to IC
ti_plicistop_ic	demo	days	From previous line ICI stop (end of regimen) to IC
pl_drug_regimen_prevq	previous_line	weeks	From one cure to another, during an ICI therapy regimen
ti_ic_1ststeroid	ic	days	From IC to first dose of steroids
ti_cl1stnivo_ic	cl	days	From current line ICI-drug 1st infusion to IC
ti_ic_ihdischarge	ih	days	Length of hospital stay

10.2 Time to cardiotoxicity onset from immune checkpoint inhibitor introduction

As the title suggests, the idea is to collect the time to onset between the **first dose** of immune checkpoint inhibitor received **during the current cycle** and the occurrence of cardiotoxicity.



This data is stored at the immune checkpoint inhibitor level, e.g., a patient who received nivolumab will have this data stored in a nivolumab specific variable. The structure for this data is

Current	Target	Description
`start_days_drugx`	`ti_drugstart_ic_drug`	It is a delay, in days. Drug name is abbreviated here in the current naming (e.g. nivo for nivolumab). Sometimes, the drug name is wrong (see below) Could be : time variable (ti_) from (ici_start) to index cardiotoxicity (ic), for drug (drug complete name)
`start_days_drugx`	`ti_drugstart_ic_drug_comp`	Same but computed from dates (when available). Mind the "x" at the end of the name Could be: same but with a standardized suffix indicating a calculation
`first_doses_drug_mce`		Drug name is complete here. It is an old variable which has now the \@ HIDDEN status, and is similar to start_days_drugx

To capture this time to onset, you have to check every possibilities for a patient, e.g. you must check for all immune checkpoint inhibitor level time variables and choose which of these 3 variables you keep first, if ever more than one is available. This gives a quite heavy piece of code

```
ti_icistart_ic = expr(pmax( # pmax here: maximum delay (in the case multiple ICI were prescribed
   eval(tto_uni_3v(start_days_nivo,
                   start_days_nivox,
                   first_doses_nivolumab_mce)), # identical to start_days_nivox but it is @HIDDEN
   eval(tto_uni_3v(start_days_pem,
                   start_days_pemx,
                   first_doses_pembrolizumab_mce)), # @HIDDEN
   eval(tto_uni_3v(start_days_pem_2, # Other Anti-PD1 regimen
                   start_days_pem_2x,
                   first_doses_pembrolizumab_mce_2)),# @HIDDEN
   eval(tto_uni_3v(start_days_pem_3, # Cemiplimab
                   start_days_pem_3x,
                   first_doses_pembrolizumab_mce_3)),
   eval(tto_uni_3v(start_days_atez,
                   start_days_atezx,
                   first_doses_atezolizumab_mce)),
   eval(tto_uni_3v(start_days_ave,
                   start_days_avex,
                   first_doses_avelumab_mce)),
   eval(tto_uni_3v(start_days_durv,
```

```
                start_days_durvx,
                first_doses_durvalumab_mce)),
  eval(tto_uni_3v(start_days_durv_3,# Ohter Anti-PDL1
                start_days_durv_3x,
                first_doses_durvalumab_mce_3)),
  eval(tto_uni_3v(start_days_ipi,
                start_days_ipix,
                first_doses_ipilimumab_mce)),
  eval(tto_uni_3v(start_days_trem,
                start_days_tremx,
                first_doses_tremelimumab_mce)),
  eval(tto_uni_3v(start_days_trem_2, # Other Anti-CTLA4
                start_days_trem_2x,
                first_doses_tremelimumab_mce_2)),
   na.rm = TRUE
) - # look at the minus!! Shall we keep it? 2022-09-16
   eval(numvar_uni(cardiotox_days_1,
                   cardiotox_days_1x))
   # Number of days before presentation that Myocarditis symptoms began

)
```

Where `eval(tto_univ_3v())` would be a prioritizer function among the 3 variables. Note that we might want to substract the `cardiotox_days_1` delay, if we're interested in the beginning of symptoms rather than index date of referral.

10.3 Death

- `ih_ou_death` : During index hospitalization
- `fu_death` : During follow-up (later than discharge)

10.4 Electrocardiograms

- `be` : Baseline (before immune checkpoint inhibitor current cycle 1st infusion)
- `be2` : The same +++ it is a duplicate in a separate instrument (should be merged in the future). There are more features here.
- `ie` : Index cardiotoxicity (1st EKG at the time of cardiotoxicity diagnosis)
- `ie2` : The same ++ it is a duplicate in a separate instrument (should be merged in the future). There are more features here.

- `ih_ekg` : Additional EKG(s) during index hospitalization
- `fu_ekg` : Most recent EKG during follow-up

10.4.1 Features

Variable	Definition
`exist`	Availability of said EKG
`analysis`	All EKG diagnoses (rhythm, durations, repolarization…)
`cornell_voltage`	Numeric features (only available for `be2` and `ie2`)
`sokolow_lyon_voltage`	
`heart_rate`	
`pr_duration`	
`qrs_duration`	
`qt_duration (also qtcb, qtcf)`	

Also, timings according to index cardiotoxicity, and upload variables.

Chapter 11

Biology

Biology identifier is [bi]

As biological features can be gather multiple times at each time point (i.e. several times during index hospitalization, or follow-up), they are gathered in a separate instrument. Naming process is standard, with

- The second item referring to the biological parameter
- The third item referring to the chronological position of the dosage.

E.g. bi_bnp_ini_value, refers to the BNP initial value

There are some biological features that do not pertain to a timing (e.g. normal value, or ever acquired status). Note that this is a bit twisted, since normal value may change if the patient gets his/her labs in a different place after being discharged.

11.1 Abbreviations

Abbreviation(s)	Feature
`uln`	Upper limit of normal
Biology timings	
`ini`	Initial value (can be slightly before or after the index cardiotoxicity date)
`bimm`	Before first immunosuppressant was introduced during index hospitalization

Abbreviation(s)	Feature
`pk`	Peak value of the parameter during index hospitalization
`mr`	Most recent value (possibly after index hospitalization discharge)
Blood chemistry	
`bnp`	Brain Natriuretic Peptide
`ntpbnp`	Nt-pro-BNP
`crp`	C reactive protein
`cre`	creatinin
`cpk`	CK or CPK
`ckmb`	CK-MB
`trop`	Troponin (either I or T)
`tropi`	Troponin I Be very careful that these parameters are used only if BOTH troponins were used. When only one troponin is used, `unit` and `uln` are flagged with the `trop` abbreviation.
`tropt`	Troponin T Same comment
Blood formula	
`cbc`	Complete blood count
`neu`	Neutrophil count
`lym`	Lymphocyte count
`mon`	Monocyte count
`eos`	Eosinophil count
`bas`	Basophil count
`igc`	Immature granulocyte count
`pla`	Platelet count

11.2 Normal values and ratios

Most normal values of biological parameters are different from one lab to another. Also, users may report bio vars according to different scales.

Examples: troponin upper limit of normal may range from ~ 0.01 to ~ 50. Complete blood count can be reported in G/L or 10^-3/µL (even if the var field asked for reporting with one of them).

It is advised to express biological parameters according to their normal value upper or lower limits if known, or to contrast them into ratios that get rid of the unit (example: neutrophil to lymphocyte ratio)

Chapter 12

Contributors and contributing centers

To ensure proper identification and tracking of contributors and contributing centers, we created separated excel files. We will use dummy contributors to illustrate this page.

Note that Redcap does allow for an easier way to do so, by adding contributors as regular users and using data access groups. At this time, this option was not retained for this project.

12.1 Contributors

Contributors have been standardized on November 2022 so that

- Typos were corrected in email addresses
- Contact with >1 email address were asked to choose one
- Outdated contacts were replaced with current ones, according to institution
- Contributors were assigned to a standardized contributing center (see below)

The excel has 2 sheets: one to tag old erroneous mails to the good one

The second to match contributors to a contributing center (here called an institution)

Table 12.1: A correspondence table between erroneous mails and good ones

old_mail	email_tag
joe1@aphp.fr	salem@aphp.fr
salem@aphp.fr	salem@aphp.fr
john1@ucsf.edu	john_pwr@ucsf.edu
power@ucsf.edu	john_pwr@ucsf.edu

Table 12.2: A correspondence table between contributors and their institution

email_tag	institution_tag
salem@aphp.fr	Sorbonne University
john_pwr@ucsf.edu	UCSF

Special cases: Some contributors might have been contacted for a case far outside of their usual health care perimeter (e.g. Dr Salem for a Belgian patient). In this case, the contributor has an additional special row from the external hospital contact.

12.2 Contributing centers (institutions)

Contributing centers have also been standardized on November 2022

- Duplicates were coerced to a unique name (e.g. UCSF, UC San Francisco… All to UCSF)
- Cities, counties and countries were sought for each contributing center

12.3 Integration to Redcap

These old names were completely removed from the records up to November 2022. Note that geographical data of institution is still located outside of the Redcap, in a separate excel file.

Table 12.3: Special case: this contributor has more than one institution

email_tag	institution_tag
salem@aphp.fr	Sorbonne University
salem@aphp.fr	Belgian Hospital

Table 12.4: A correspondence table between erroneous institution names and good ones

old_institution	institution_tag
APHP Pitié Salpetriere	Sorbonne University
Sorbonne	Sorbonne University
UCSF	UCSF
UC San Francisco	UCSF

Table 12.5: A correspondence table between an institution name and its geographical place

institution_tag	ad_country	ad_admin	ad_city
Sorbonne University	France		Paris
UCSF	United States	California	San Francisco

However, it does not imply that newer records will comply to this standardization. Three options here

- Add contributors as regular users to Redcap, so their mail and institution can be easily constrained
- Add coercion to fields related to emails and institutions (e.g., reporter can only choose in a prespecified list of institutions).
- Keep updating the database by hand periodically

The first two options almost are the same, in the way that new external physicians will have to be registered first, before posting cases. This would add some delay before they can enter their cases, although probably not that much, and could discourage them from contributing. The counterpart is much more easy process to identify and track them thereafter.