

Verification Report: data_clean

Ellis Hughes

2021-06-15

Description

This is the verification report for the `data_clean` folder of the `correlates_reporting` project for CoVPN.

In this document, the output of `make_data_proc.R` is compared against the output of `process_data_raw.R`. The two scripts use the same base mock data. The two datasets generated by each of the two scripts will be compared with each other to confirm they contain the same values.

`process_data_raw.R` was independently double programmed based on the specifications found in `dat_clean_specifications.pdf`. This script is run to execute outputs for verification for both Moderna and Janssen protocol outputs.

The file `data_clean/verification/verification_input/moderna_mock_data_processed.csv` was provided by the original programmer to the tester for verification purposes of `make_data_proc.R` for the moderna protocol output. Its md5 hash is 33f197dd588d7e45fd18b5d2c4f5bc2a.

The file `data_clean/verification/verification_input/janssen_pooled_mock_data_processed.csv` was provided by the original programmer to the tester for verification purposes of `make_data_proc.R` for the Janssen protocol output. Its md5 hash is 1a0b0b9ba9d11064e67b7a7f39a2dac3.

The file `data_clean/verification/verification_output/COVID_MODERNA_verification_output.csv` was created by the tester for verification using the `process_data_Raw.R` script and using all time points available. Its md5 hash is e0b9226eef1c0148fd79ed84d4d224bd.

The file `data_clean/verification/verification_output/COVID_ENSEMBLE_verification_output.csv` was created by the tester for verification of Janssen Ensembl data processing using the `process_data_Raw.R` script and using only the time points Baseline and Day 29. Its md5 hash is 9657910a60629115bb015502add9d99e.

Load Data

```
original_data_moderna <- read_csv(  
  here("data_clean/verification/verification_input", "moderna_mock_data_processed.csv"),  
  guess_max = 30000)  
  
original_data_janssen <- read_csv(  
  here("data_clean/verification/verification_input/", "janssen_pooled_mock_data_processed.csv"),  
  guess_max = 30000)  
  
verification_data_moderna <- read_csv(  
  here("data_clean/verification/verification_output/COVID_MODERNA_verification_output.csv"),  
  guess_max = 30000)  
  
verification_data_janssen <- read_csv(  
  here("data_clean/verification/verification_output/COVID_ENSEMBLE_verification_output.csv"),  
  guess_max = 30000)
```

Verification

```
data_clean_comparison_moderna <- compare_datasets(  
  cols = colnames(original_data_moderna), index = "Ptid",  
  ds1 = original_data_moderna, ds2 = verification_data_moderna  
)
```

```
## There are 0 mismatched fields of 104.
```

```
data_clean_comparison_janssen <- compare_datasets(  
  cols = colnames(original_data_janssen), index = "Ptid",  
  ds1 = original_data_janssen, ds2 = verification_data_janssen  
)
```

```
## There are 0 mismatched fields of 69.
```

Output of `make_data_proc.R` is equivalent to the output of `process_data_raw.R` for the Moderna Protocol.
Output of `make_data_proc.R` is equivalent to the output of `process_data_raw.R` for the Janssen protocol.
`make_data_proc.R` passes verification.

Signatures

Role	Name	Signature	Date
Tester	Ellis Hughes		