

# Package ‘UKB.COVID19’

January 25, 2024

**Type** Package

**Title** UK Biobank COVID-19 Data Processing and Risk Factor Association Tests

**Version** 0.1.4

**Date** 2024-01-25

**Author** Longfei Wang [aut, cre] (<<https://orcid.org/0000-0002-5143-4146>>)

**Maintainer** Longfei Wang <wang.lo@wehi.edu.au>

**Description** Process UK Biobank COVID-19 test result data for susceptibility, severity and mortality analyses, perform potential non-genetic COVID-19 risk factor and co-morbidity association tests. Wang et al. (2021) <[doi:10.5281/zenodo.5174381](https://doi.org/10.5281/zenodo.5174381)>.

**Imports** questionr, data.table, tidyverse, magrittr, here, dplyr

**License** MIT + file LICENSE

**URL** <https://github.com/bahlolab/UKB.COVID19>

**Encoding** UTF-8

**RoxygenNote** 7.1.1.9001

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2024-01-25 00:40:02 UTC

## R topics documented:

comorbidity_asso . . . . .	2
comorbidity_summary . . . . .	3
covid_example . . . . .	4
data_reform . . . . .	5
log_cov . . . . .	5
makeGWASFiles . . . . .	6

makePhenotypes . . . . .	7
risk_factor . . . . .	9
sampleQC . . . . .	10
variantQC . . . . .	11

<b>Index</b>	<b>12</b>
--------------	-----------

---

comorbidity_asso	<i>Generate comorbidity association result file</i>
------------------	---

---

## Description

Association tests between each co-morbidity and given phenotype (susceptibility, mortality or severity) with the adjustment of covariates.

## Usage

```
comorbidity_asso(
  pheno,
  covariates,
  cormorbidity,
  population = "all",
  cov.name = c("sex", "age", "bmi"),
  phe.name,
  ICD10.file
)
```

## Arguments

pheno	phenotype dataframe - output from makePheno function
covariates	covariate dataframe - output from risk.factor function. Optional.
cormorbidity	Comorbidity summary generated from comorbidity.summary.
population	Choose self-report population/ethnic background group from "all", "white", "black", "asian", "mixed", or "other". By default, population="all", include all ethnic groups.
cov.name	Selected covariates names. By default, cov.name=c("sex","age","bmi"), covariates are sex age and BMI.
phe.name	Phenotype name.
ICD10.file	The ICD10 code file, which is included in the package.

## Value

Outputs a comorbidity association test result with OR, 95% CI and p-value.

## Examples

```
## Not run:
comorb.asso <- comorbidity_asso(pheno=phe,
covariates=covar,
cormorbidity=comorb,
population="white",
cov.name=c("sex","age","bmi","SES","smoke","inAgedCare"),
phe.name="hospitalisation",
ICD10.file=covid_example("ICD10.coding19.txt.gz"))

## End(Not run)
```

---

comorbidity\_summary    *Create comorbidity summary file*

---

## Description

summarise disease history records of each individual from the hospital inpatient diagnosis data.

## Usage

```
comorbidity_summary(
  ukb.data,
  hesin.file,
  hesin_diag.file,
  primary = FALSE,
  ICD10.file,
  Date.start = NULL,
  Date.end = NULL
)
```

## Arguments

ukb.data	tab delimited UK Biobank phenotype file, containing sample qc fields (with default UKBiobank codes as column names)
hesin.file	Latest hospital inpatient master file.
hesin_diag.file	Latest hospital inpatient diagnosis file.
primary	TRUE: include primary diagnosis only; FALSE: include all diagnoses.
ICD10.file	The ICD10 code file, which is included in the package.
Date.start	Date, dd/mm/yyyy, select the start date of hospital inpatient record period.
Date.end	Date, dd/mm/yyyy, select the end date of hospital inpatient record period.

**Value**

Outputs comorbidity summary table, named `comorbidity_<Date.start>_<Date.end>.RData`, including phenotype, non-genetic risk factors and all comorbidities, which will be used in the comorbidity association tests.

**Examples**

```
## Not run:
comorb <- comorbidity_summary(ukb.data=covid_example("sim_ukb.tab.gz"),
hesin.file=covid_example("sim_hesin.txt.gz"),
hesin_diag.file=covid_example("sim_hesin_diag.txt.gz"),
ICD10.file=covid_example("ICD10.coding19.txt.gz"),
primary = FALSE,
Date.start = "16/03/2020")

## End(Not run)
```

---

`covid_example`*Provide working directory for UKB.COVID19 example files*

---

**Description**

Provide working directory for UKB.COVID19 example files

**Usage**

```
covid_example(path)
```

**Arguments**

`path`                    path to file

**Value**

Outputs the working directory for UKB.COVID19 example files.

**Examples**

```
covid_example('results/covariate.txt')
```

---

data_reform	<i>Reform variables</i>
-------------	-------------------------

---

**Description**

Reform variables

**Usage**

```
data_reform(res, type)
```

**Arguments**

res	Merged data of phenotype from makePhenotypes or comorbidity_summary and covariates from risk_factor.
type	Data type: susceptibility, severity, mortality or comorbidity.

**Value**

Reformed data for association tests using logistic regression models.

---

log_cov	<i>Perform association tests between phenotype and covariates</i>
---------	---

---

**Description**

Perform association tests between phenotype and covariates

**Usage**

```
log_cov(pheno, covariates, phe.name, cov.name = c("sex", "age", "bmi"))
```

**Arguments**

pheno	phenotype dataframe - output from makePhenotypes function
covariates	covariate dataframe - output from risk_factor function.
phe.name	Phenotype name in the data.
cov.name	Selected covariate names in the data. By default, cov.name=c("sex","age","bmi"), covariates include sex, age and BMI.

**Value**

Outputs association test results with OR, 95% CI, and p-value.

**Examples**

```
## Not run:
log_cov(pheno=phe, covariates=covar, phe.name="hospitalisation", cov.name=c("sex","age","bmi"))

## End(Not run)
```

---

makeGWASFiles	<i>Generate files for GWAS Software. SAIGE and Plink currently supported.</i>
---------------	---

---

**Description**

Generate files for GWAS Software. SAIGE and Plink currently supported.

**Usage**

```
makeGWASFiles(
  ukb.data,
  pheno,
  covariates,
  phe.name,
  cov.name = NULL,
  includeSampsFile = NULL,
  software = "SAIGE",
  outDir = "",
  prefix
)
```

**Arguments**

ukb.data	tab delimited UK Biobank phenotype file, containing sample qc fields (with default UKBiobank codes as column names)
pheno	phenotype dataframe - output from makePhenotype function
covariates	covariate dataframe - output from risk.factor function. Optional.
phe.name	phenotypes to be included in outputted data. multiple phenotypes can be specified as a vector. if null, all phenotypes will be outputted.
cov.name	covariates to be included in outputted data. Optional. multiple covariates can be specified as a vector. if null, all covariates in file will be outputted
includeSampsFile	list of samples to be included GWAS. File with the first column containing sample IDs to be kept. Can contain other columns. output from sampleQC function may be used. Optional - if null, all samples will be outputted.
software	specify "SAIGE" or "plink" - defaults to "SAIGE"
outDir	specify directory to output file
prefix	prefix for file - optional

**Value**

outputs file, suitable for reading by chosen GWAS software

**Examples**

```
## Not run:
makeGWASFiles(ukb.data=covid_example("sim_ukb.tab.gz"),
pheno=phe,
covariates=covar,
phe.name="hospitalisation",
cov.name=NULL,
includeSampsFile=NULL,
software="SAIGE",
outDir=covid_example("results"),
prefix="hospitalisation")

## End(Not run)
```

---

makePhenotypes

*Generate COVID-19 phenotypes*

---

**Description**

Generate COVID-19 phenotypes

**Usage**

```
makePhenotypes(
  ukb.data,
  res.eng,
  res.wal = NULL,
  res.sco = NULL,
  death.file,
  death.cause.file,
  hesin.file,
  hesin_diag.file,
  hesin_oper.file,
  hesin_critical.file,
  code.file,
  pheno.type = "severity",
  Date = NULL
)
```

**Arguments**

ukb.data	tab delimited UK Biobank phenotype file.
res.eng	Latest covid result file/files for England.
res.wal	Latest covid result file/files for Wales. Only available for downloads after April 2021.
res.sco	Latest covid result file/files for Scotland. Only available for downloads after April 2021.
death.file	Latest death register file.
death.cause.file	Latest death cause file.
hesin.file	Latest hospital inpatient master file.
hesin_diag.file	Latest hospital inpatient diagnosis file.
hesin_oper.file	Latest hospital inpatient operation file.
hesin_critical.file	Latest hospital inpatient critical care file.
code.file	The operation code file, which is included in the package.
pheno.type	The phenotype options, which include "susceptibility", "severity", and "mortality".
Date	Date, ddmmyyyy, select the results until a certain date. By default, Date = NULL, the latest hospitalization date.

**Value**

Returns a data.frame with phenotypes for COVID-19 susceptibility, severity and mortality.

**Examples**

```
## Not run:
pheno <- makePhenotypes(ukb.data=covid_example("sim_ukb.tab.gz"),
  res.eng=covid_example("sim_result_england.txt.gz"),
  death.file=covid_example("sim_death.txt.gz"),
  death.cause.file=covid_example("sim_death_cause.txt.gz"),
  hesin.file=covid_example("sim_hesin.txt.gz"),
  hesin_diag.file=covid_example("sim_hesin_diag.txt.gz"),
  hesin_oper.file=covid_example("sim_hesin_oper.txt.gz"),
  hesin_critical.file=covid_example("sim_hesin_critical.txt.gz"),
  code.file=covid_example("coding240.txt.gz"),
  pheno.type = "severity")

## End(Not run)
```



---

risk_factor	<i>Generate covariate file</i>
-------------	--------------------------------

---

## Description

This function formats and outputs a covariate table, used for input for other functions.

## Usage

```
risk_factor(  
  ukb.data,  
  ABO.data = NULL,  
  hesin.file,  
  res.eng,  
  res.wal = NULL,  
  res.sco = NULL,  
  fields = NULL,  
  field.names = NULL  
)
```

## Arguments

ukb.data	tab delimited UK Biobank phenotype file. The file should include fields of gender, year of birth, BMI, ethnic background, SES, and smoking.
ABO.data	Latest yyyyymmdd_covid19_misc.txt file.
hesin.file	Latest yyyyymmdd_hesin.txt file.
res.eng	Latest covid result file/files for England.
res.wal	Latest covid result file/files for Wales. Only available for downloads after April 2021.
res.sco	Latest covid result file/files for Scotland. Only available for downloads after April 2021.
fields	User specified field codes from ukb.data file.
field.names	User specified field names.

## Value

Outputs a covariate table, used for input for other functions. Automatically returns sex, age at birthday in 2020, SES, self-reported ethnicity, most recently reported BMI, most recently reported pack-years, whether they reside in aged care (based on hospital admissions data, and covid test data) and blood type. Function also allows user to specify fields of interest (field codes, provided by UK Biobank), and allows the users to specify more intuitive names, for selected fields.

## Examples

```
## Not run:
covars <- risk_factor(ukb.data=covid_example("sim_ukb.tab.gz"),
ABO.data=covid_example("sim_covid19_misc.txt.gz"),
hesin.file=covid_example("sim_hesin.txt.gz"),
res.eng=covid_example("sim_result_england.txt.gz"))

## End(Not run)
```

---

sampleQC

*Sample QC for genetic analyses*

---

## Description

Sample QC for genetic analyses

## Usage

```
sampleQC(ukb.data, withdrawnFile, ancestry = "all", software = "SAIGE", outDir)
```

## Arguments

ukb.data	tab delimited UK Biobank phenotype file, containing sample qc fields (with default UKBiobank codes as column names)
withdrawnFile	csv file with withdrawn IDs from UK Biobank
ancestry	specify "WhiteBritish" or "all" - defaults to "all"
software	specify "SAIGE" or "plink" - defaults to "SAIGE"
outDir	specify directory for sample QC file and inclusion/exclusion lists

## Value

outputs sample QC file, and sample inclusion / exclusion lists for specified software

## Examples

```
## Not run:
sampleQC(ukb.data=covid_example("sim_ukb.tab.gz"),
withdrawnFile=covid_example("sim_withdrawn.csv.gz"),
ancestry="all",
software="SAIGE",
outDir=covid_example("results"))

## End(Not run)
```

---

variantQC	<i>Variant QC for Genetic Analyses</i>
-----------	--

---

**Description**

Variant QC for Genetic Analyses

**Usage**

```
variantQC(snpQcFile, mfiDir, mafFilt = 0.001, infoFilt = 0.5, outDir)
```

**Arguments**

snpQcFile	file containing SNP QC info (ukb_snp_qc.txt)
mfiDir	directory where the per chromosome UKBiobank MAF/INFO files (ukb_mfi_chr*_v3.txt) are located
mafFilt	minor allele frequency filter - default 0.001
infoFilt	imputation quality (INFO) score filter - default 0.5
outDir	output directory

**Value**

outputs SNP inclusion lists (SNPID and rsID formats) for given MAF/INFO filters. Also outputs list of SNPs to be used for genetic Relatedness Matrix (GRM) calculations.

**Examples**

```
## Not run:  
variantQC(snpQcFile=covid_example("sim_ukb_snp_qc.txt.gz"),  
mfiDir=covid_example("alleleFreqs"),  
mafFilt=0.001,  
infoFilt=0.5,  
outDir=covid_example("results"))  
  
## End(Not run)
```

# Index

## \* **reform**

data\_reform, 5

comorbidity\_asso, 2

comorbidity\_summary, 3

covid\_example, 4

data\_reform, 5

log\_cov, 5

makeGWASFiles, 6

makePhenotypes, 7

risk\_factor, 9

sampleQC, 10

variantQC, 11