Example R and Stata code for a Mendelian randomization analysis

Sanderson et al.

24 September 2021

Example R and Stata code to perform a multivariable Mendelian randomization (MVMR) analysis

R. code

• Read in the data

```
suppressPackageStartupMessages({
   library(tidyverse)
   library(haven)
   library(remotes)
})
dat <- read_dta("https://raw.github.com/remlapmot/mrrobust/master/dodata.dta")
dat <- dat %>% filter(ldlcp2 < 1e-8)</pre>
```

Example code using the MendelianRandomization package

• Install the package and load it into memory

library(MendelianRandomization)

• Convert our data frame to the required class

```
datfmt <- mr_mvinput(
  bx = as.matrix(cbind(dat$ldlcbeta, dat$hdlcbeta, dat$tgbeta)),
  bxse = as.matrix(cbind(dat$ldlcse, dat$hdlcse, dat$tgse)),
  by = dat$chdbeta,
  byse = dat$chdse,
  exposure = "exposure",
  outcome = "outcome",
  snps = "snp",
  effect_allele = dat$a1,
  other_allele = dat$a2,
  eaf = NA
)</pre>
```

• Fit an MVMR/MVIVW model

```
mvivwfit <- mr_mvivw(datfmt)</pre>
mvivwfit
## Multivariable inverse-variance weighted method
## (variants uncorrelated, random-effect model)
## Number of Variants : 73
##
   ______
##
##
     Exposure Estimate Std Error 95% CI p-value
## exposure_1 0.429 0.061 0.309, 0.548 0.000
## exposure_2 -0.194 0.131 -0.451, 0.062 0.138
## exposure_3 0.226 0.123 -0.016, 0.468 0.067
## Residual standard error = 1.490
## Heterogeneity test statistic = 155.3766 on 70 degrees of freedom, (p-value = 0.0000)
  • Fit an MVMR-Egger model
mvmreggerfit <- mr_mvegger(datfmt)</pre>
mvmreggerfit
## Multivariable MR-Egger method
## (variants uncorrelated, random-effect model)
##
## Orientated to exposure : 1
## Number of Variants : 73
##
      Exposure Estimate Std Error 95% CI
                                              p-value
##
    exposure_1 0.567 0.100 0.371, 0.764 0.000
##
    exposure_2 -0.136 0.133 -0.398, 0.125 0.306
##
    exposure_3 0.274 0.125 0.030, 0.518
                                                0.028
    (intercept) -0.009 0.005 -0.020, 0.001
##
                                                0.084
## Residual standard error = 1.469
## Heterogeneity test statistic = 148.9290 on 69 degrees of freedom, (p-value = 0.0000)
Example code using the MVMR and RMVMR packages
  • Install the package and load it into memory
if (!requireNamespace("MVMR", quietly = TRUE)) {
```

• Create a data object of the required structure

remotes::install_github("WSpiller/MVMR")

}

library(MVMR)

```
r_input <- format_mvmr(
    BXGs = dat[,c("ldlcbeta","hdlcbeta","tgbeta")],
    BYG = dat$chdbeta,
    seBXGs = dat[,c("ldlcse","hdlcse","tgse")],
    seBYG = dat$chdse,
    RSID = dat$rsid
)</pre>
```

• Fit an MVMR model

##

```
mvmrfit <- ivw_mvmr(r_input)

##

## Multivariable MR

##

## Estimate Std. Error t value Pr(>|t|)

## exposure1 0.4286200 0.0609661 7.030464 1.099077e-09

## exposure2 -0.1941989 0.1308289 -1.484372 1.421994e-01

## exposure3 0.2260456 0.1232828 1.833554 7.097168e-02
```

• Conditional F-statistics for instrument strength (Sanderson, Spiller, and Bowden 2021)

Residual standard error: 1.49 on 70 degrees of freedom

```
## Warning in strength_mvmr(r_input): Covariance between effect of genetic variants on each exposure
## not specified. Fixing covariance at 0.

##
## Conditional F-statistics for instrument strength
##
## exposure1 exposure2 exposure3
## F-statistic 126.7447 35.29937 39.32731

##
## exposure1 exposure2 exposure3
## F-statistic 126.7447 35.29937 39.32731
```

• Fit a radial MVMR model

```
if (!requireNamespace("RMVMR", quietly = TRUE)) {
    remotes::install_github("WSpiller/RMVMR")
}
library(RMVMR)

rmvmr_input <- mrmvinput_to_rmvmr_format(datfmt)

rmvmr_fit <- ivw_rmvmr(rmvmr_input, summary = TRUE)</pre>
```

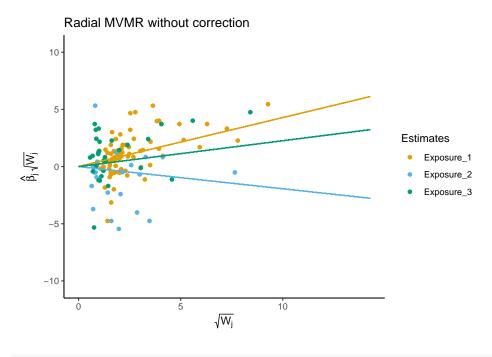
```
##
## Radial Multivariable MR
##
## Estimate Std. Error t value Pr(>|t|)
## exposure1 0.4286200 0.0609661 7.030464 1.099077e-09
## exposure2 -0.1941989 0.1308289 -1.484372 1.421994e-01
## exposure3 0.2260456 0.1232828 1.833554 7.097168e-02
##
## Residual standard error: 1.49 on 70 degrees of freedom
```

rmvmr_fit\$coef

```
## Estimate Std. Error t value Pr(>|t|)
## exposure1 0.4286200 0.0609661 7.030464 1.099077e-09
## exposure2 -0.1941989 0.1308289 -1.484372 1.421994e-01
## exposure3 0.2260456 0.1232828 1.833554 7.097168e-02
```

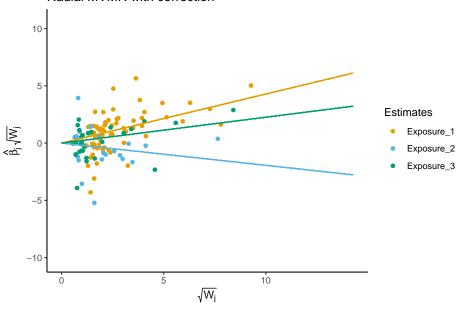
• Plot the radial MVMR models

```
plt_rmvmr <- plot_rmvmr(rmvmr_input, rmvmr = rmvmr_fit)
plt_rmvmr$p1</pre>
```



plt_rmvmr\$p2

Radial MVMR with correction



• Heterogeneity statistics

```
pleio_rmvmr <- pleiotropy_rmvmr(rmvmr_input, rmvmr = rmvmr_fit)
pleio_rmvmr$gq</pre>
```

```
## Exposure_1 76.37356 2.812609e-01
## Exposure_2 59.58894 8.243650e-06
## Exposure_3 45.88633 1.308596e-02
```

head(pleio_rmvmr\$qdat)

```
##
                 wj corrected_beta
                                                      qj_p ref_exposure
                                             qj
## 1 snp_1 2.414215
                         0.3141338 0.031643343 0.85881269
                                                             Exposure_1
## 2 snp_2 3.938760
                         0.3821443 0.008507667 0.92650973
                                                             Exposure_1
## 3 snp 3 3.246657
                        -0.6057993 3.473998077 0.06234046
                                                             Exposure_1
## 4 snp_4 2.001191
                         0.3273484 0.020524059 0.88608310
                                                             Exposure_1
## 5 snp_5 9.278826
                         0.5414293 0.118081748 0.73112437
                                                             Exposure_1
## 6 snp_6 1.317671
                         1.0713541 0.544339092 0.46064001
                                                             Exposure_1
```

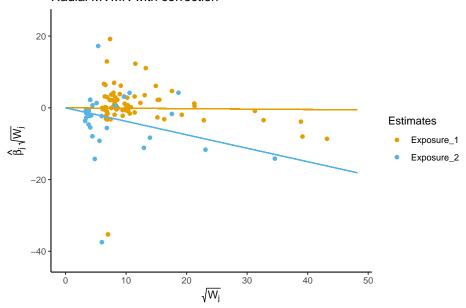
• Conditional F-statistics for instrument strength (Sanderson, Spiller, and Bowden 2021)

```
str_rmvmr <- strength_rmvmr(rmvmr_input)</pre>
```

Warning in MVMR::strength_mvmr(r_input, gencov): Covariance between effect of genetic variants on ## each exposure not specified. Fixing covariance at 0.

```
str_rmvmr$plot[[2]]
```

Radial MVMR with correction



str_rmvmr\$qstat[[2]]

Stata code

• Load the Statamarkdown package to enable Stata code chunks in an R Markdown file

```
if (!requireNamespace("Statamarkdown", quietly = TRUE)) {
   remotes::install_github("Hemken/Statamarkdown")
}
library(Statamarkdown)
```

• Read in the data and create an indicator variable to select observations with p-value between the genotype and LDL-C < 10-8

```
use https://raw.github.com/remlapmot/mrrobust/master/dodata, clear
gen byte sel1 = (ldlcp2 < 1e-8)</pre>
```

Example code using the mrrobust package

• Install the mrrobust package using the github package

```
// Note: output suppressed
net install mrrobust, from("https://raw.github.com/remlapmot/mrrobust/master/")
mrdeps
```

• Fit IVW (with fixed effect standard errors)

mregger chdbeta ldlcbeta [aw=1/(chdse^2)] if sel1==1, ivw fe

			Residual		umber of geno error constr	
		Std. Err.				Interval]
chdbeta	. 4815055			0.000	. 4065938	.5564173

• Fit MR-Egger reporting I2GX statistic and heterogeneity Q-test

```
mregger chdbeta ldlcbeta [aw=1/(chdse^2)] if sel1==1, gxse(ldlcse) heterogi
```

```
Q_GX statistic (weighted) = 3454.26

I^2_GX statistic (weighted) = 97.92%

Number of genotypes = 73

Residual standard error = 1.548

Ruecker's Q for heterogeneity; chi2(71) = 170.11 (p = 0.0000)

I-squared statistic = 58.3% (95% CI 45.8%, 67.8%)

| Coef. Std. Err. z P>|z| [95% Conf. Interval]

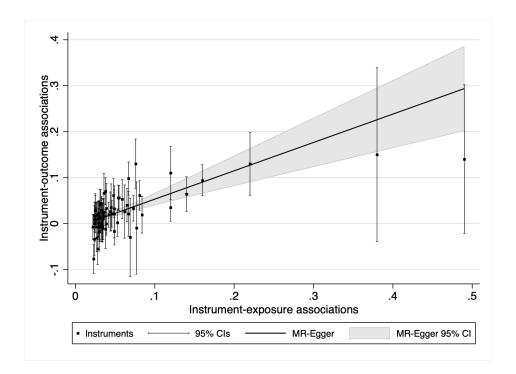
chdbeta |

slope | .6173131 .1034573 5.97 0.000 .4145405 .8200858

_cons | -.0087706 .0054812 -1.60 0.110 -.0195136 .0019723
```

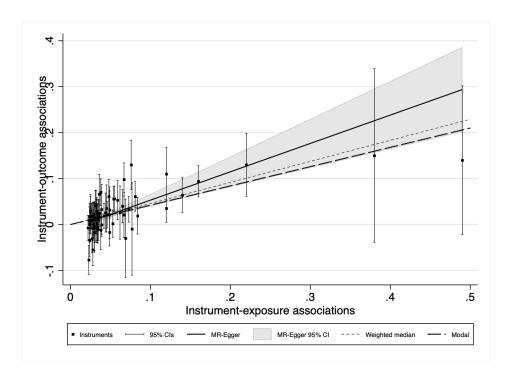
• Simple plot of IVW and MR-Egger univariate estimates

```
mreggerplot chdbeta chdse ldlcbeta ldlcse if sel1==1
qui gr export mreggerplot-1.png, replace
```



• Adding the modal and median estimates onto the plot

				Number of genotypes = 73 Replications = 1000
	Coef.			[95% Conf. Interval]
				.3341648 .5823499
				N
				Number of genotypes = 73 Replications = 1000 Phi = .25
		Std. Err.	P> z	Replications = 1000 Phi = .25 [95% Conf. Interval]



• Fit and MVMR model with phenotypes LDL-c and HDL-c (Burgess, Dudbridge, and Thompson 2015).

mvmr chdbeta ldlcbeta hdlcbeta [aw=1/(chdse^2)] if sel1==1

Number of genotypes = 73 Number of phenotypes = 2 Standard errors: Random effect Residual standard error = 1.514

	Coef.					Interval]
chdbeta ldlcbeta	.4670719	.0581901	8.03	0.000	.3530214 5305175	

• Additionally include a third phenotype – trigly cerides.

mvmr chdbeta ldlcbeta hdlcbeta tgbeta [aw=1/(chdse^2)] if sel1==1

Number of genotypes = 73 Number of phenotypes = 3 Standard errors: Random effect Residual standard error = 1.490

				[95% Conf.	
chdbeta ldlcbeta	.0609661	7.03	0.000	.3091286	.5481113

tgbeta | .2260456 .1232828 1.83 0.067 -.0155842 .4676755

• Report the QA statistic for instrument validity and the conditional F-statistics for instrument strength for each phenotype (Sanderson et al. 2019; Sanderson, Spiller, and Bowden 2021).

mvmr chdbeta ldlcbeta hdlcbeta tgbeta [aw=1/(chdse^2)] if sel1==1, gxse(ldlcse hdlcse tgse)

> hdlcse tgse)

Number of genotypes = 73 Number of phenotypes = 3 Standard errors: Random effect Residual standard error = 1.490

		Std. Err.				Interval]
chdbeta						
ldlcbeta	.42862	.0609661	7.03	0.000	.3091286	.5481113
hdlcbeta	1941989	.1308289	-1.48	0.138	4506189	.0622211
tgbeta	.2260456	.1232828	1.83	0.067	0155842	.4676755

 Q_A statistic for instrument validity; chi2(70) = 152.88 (p = 0.0000)

Conditional F-statistics for instrument strength:

 $F_x1 = 130.31$ (ldlcbeta)

 $F \times 2 = 36.29$ (hdlcbeta)

 $F_x3 = 40.44$ (tgbeta)

• Fit an MVMR-Egger regression (Rees, Wood, and Burgess 2017), orienting the model with respect to the first phenotype in the main *varlist*.

mrmvegger chdbeta ldlcbeta hdlcbeta tgbeta [aw=1/(chdse^2)] if sel1==1

MVMR-Egger model oriented wrt: ldlcbeta

Number of genotypes = 73

Number of phenotypes = 3

Residual standard error = 1.469

	Coef.	Std. Err.	z	P> z	[95% Conf.	Interval]
chdbeta ldlcbeta hdlcbeta tgbeta _cons	.5672993 1364113 .2739803	.1002611 .1332727 .1246927 .0054187	5.66 -1.02 2.20 -1.73	0.000 0.306 0.028 0.084	.370791 3976209 .0295871 019986	.7638075 .1247983 .5183735 .001255

We can also orient the model with respect to HDL-C.

mrmvegger chdbeta ldlcbeta hdlcbeta tgbeta [aw=1/(chdse^2)] if sel1==1, orient(2)

> (2)

MVMR-Egger model oriented wrt: hdlcbeta

Number of genotypes = 73

Number of phenotypes = 3

Residual standard error = 1.501

 	Coef.	 Std. Err.	z	P> z	[95% Conf.	Interval]
chdbeta						
ldlcbeta	.4286398	.0614056	6.98	0.000	.308287	.5489926
hdlcbeta	1989637	.1541909	-1.29	0.197	5011723	.1032449
tgbeta	.2256794	.1243221	1.82	0.069	0179875	.4693463
_cons	.0002155	.0036218	0.06	0.953	006883	.0073141

Or we can orient the model with respect to triglycerides.

```
mrmvegger chdbeta ldlcbeta hdlcbeta tgbeta [aw=1/(chdse^2)] if sel1==1, orient(3)
```

> (3)

MVMR-Egger model oriented wrt: tgbeta Number of genotypes = 73 Number of phenotypes = 3

Residual standard error = 1.499

	 	Coef.	Std. Err.	z	P> z	[95% Conf.	Interval]
chdbeta							
ldlcbeta		.4203073	.0660026	6.37	0.000	.2909447	.54967
hdlcbeta		1903089	.1321536	-1.44	0.150	4493252	.0687075
tgbeta		.2065651	.1365427	1.51	0.130	0610537	.474184
_cons	I	.0013499	.003951	0.34	0.733	0063939	.0090936

R session information for reproducibility

```
library(sessioninfo)
session_info()
```

- Session info ----setting value

version R version 4.1.1 (2021-08-10)

macOS Big Sur 11.6 system aarch64, darwin20

ui X11 language (EN)

collate en_GB.UTF-8 ctype en_GB.UTF-8 Europe/London tz

package	* version	date	lib	sourc	:e		
arrangements	1.1.9	2020-09-13				4.1.0)	
assertthat	0.2.1	2019-03-21					
backports	1.2.1	2020-12-09					
broom	0.7.9	2021-07-27					
cellranger	1.1.0	2016-07-27					
cli	3.0.1	2021-07-17					
codetools	0.2-18	2020-11-04					
colorspace	2.0-2	2021-06-24					
conquer	1.0.2	2020-08-27					
crayon	1.4.1	2021-02-08					
curl	4.3.2	2021-06-23					
data.table	1.14.0	2021-02-21					
DBI	1.1.1	2021-01-15					
dbplyr	2.1.1	2021-04-06					
DEoptimR	1.0-9	2021-05-24					
digest	0.6.28	2021-09-23					
_	* 1.0.7	2021-06-18					
ellipsis	0.3.2	2021-04-29					
evaluate	0.3.2	2019-05-28					
fansi	0.14	2019 05 26					
farver	2.1.0	2021-03-23					
	1.1.0	2021-02-28					
fastmap forcats	* 0.5.1	2021-01-25					
foreach		2021-01-27					
	1.5.1						
fs	1.5.0	2020-07-31					
generics	0.1.0	2020-10-31					
001	* 3.3.5	2021-06-25					
glmnet	4.1-2	2021-06-24					
glue	1.4.2	2020-08-27					
gmp	0.6-2	2021-01-07					
gtable	0.3.0	2019-03-25					
	* 2.4.3	2021-08-04					
hms	1.1.0	2021-05-17					
htmltools	0.5.2	2021-08-25					
htmlwidgets	1.5.4	2021-09-08					
httr	1.4.2	2020-07-20					
iterators	1.0.13	2020-10-15					
iterpc	0.4.2	2020-01-10					
jsonlite	1.7.2	2020-12-09					
knitr	1.34	2021-09-09					
labeling	0.4.2	2020-10-20					
lattice		2021-09-22					
lazyeval	0.2.2	2019-03-15					
lifecycle	1.0.0	2021-02-15					
lubridate	1.7.10	2021-02-26					
magrittr	2.0.1	2020-11-17					
Matrix	1.3-4	2021-06-01					
MatrixModels	0.5-0	2021-03-02	[1]	CRAN	(R	4.1.0)	
matrixStats MendelianRandomization	0.61.0	2021-09-17 2021-04-16					

```
2020-05-19 [1] CRAN (R 4.1.0)
modelr
                         0.1.8
munsell
                         0.5.0
                                 2018-06-12 [1] CRAN (R 4.1.0)
MVMR
                       * 0.3
                                 2021-09-24 [1] Github (wspiller/mvmr@3a0b54b)
pillar
                         1.6.2
                                 2021-07-29 [1] CRAN (R 4.1.0)
pkgconfig
                         2.0.3
                                 2019-09-22 [1] CRAN (R 4.1.0)
                         4.9.4.1 2021-06-18 [1] CRAN (R 4.1.0)
plotly
                       * 0.3.4
                                 2020-04-17 [1] CRAN (R 4.1.0)
purrr
                         5.86
                                 2021-06-06 [1] CRAN (R 4.1.0)
quantreg
R6
                         2.5.1
                                 2021-08-19 [1] CRAN (R 4.1.1)
RadialMR
                         1.0
                                 2021-08-13 [1] Github (WSpiller/RadialMR@d63d3fc)
Rcpp
                         1.0.7
                                 2021-07-07 [1] CRAN (R 4.1.0)
                       * 2.0.1
                                 2021-08-10 [1] CRAN (R 4.1.1)
readr
readxl
                         1.3.1
                                 2019-03-13 [1] CRAN (R 4.1.0)
                       * 2.4.0
                                 2021-06-02 [1] CRAN (R 4.1.0)
remotes
                         2.0.1
                                 2021-08-05 [1] CRAN (R 4.1.1)
reprex
rjson
                         0.2.20
                                 2018-06-08 [1] CRAN (R 4.1.0)
                         0.4.11
                                 2021-04-30 [1] CRAN (R 4.1.0)
rlang
rmarkdown
                         2.11
                                 2021-09-14 [1] CRAN (R 4.1.1)
RMVMR
                       * 0.2
                                 2021-09-24 [1] Github (WSpiller/RMVMR@a162f9e)
robustbase
                         0.93-8
                                 2021-06-02 [1] CRAN (R 4.1.0)
rstudioapi
                         0.13
                                 2020-11-12 [1] CRAN (R 4.1.0)
rvest
                         1.0.1
                                 2021-07-26 [1] CRAN (R 4.1.0)
                                 2020-05-11 [1] CRAN (R 4.1.0)
scales
                         1.1.1
                       * 1.1.1
                                 2018-11-05 [1] CRAN (R 4.1.0)
sessioninfo
                         1.4.6
                                 2021-05-19 [1] CRAN (R 4.1.0)
shape
SparseM
                         1.81
                                 2021-02-18 [1] CRAN (R 4.1.0)
Statamarkdown
                       * 0.7.0
                                 2021-08-26 [1] Github (Hemken/Statamarkdown@a68a8b9)
                         1.7.4
                                 2021-08-25 [1] CRAN (R 4.1.1)
stringi
                       * 1.4.0
                                 2019-02-10 [1] CRAN (R 4.1.1)
stringr
survival
                         3.2-13 2021-08-24 [1] CRAN (R 4.1.1)
tibble
                       * 3.1.4
                                 2021-08-25 [1] CRAN (R 4.1.1)
tidyr
                       * 1.1.3
                                 2021-03-03 [1] CRAN (R 4.1.0)
tidyselect
                         1.1.1
                                 2021-04-30 [1] CRAN (R 4.1.0)
                       * 1.3.1
                                 2021-04-15 [1] CRAN (R 4.1.0)
tidyverse
tzdb
                         0.1.2
                                 2021-07-20 [1] CRAN (R 4.1.0)
utf8
                         1.2.2
                                 2021-07-24 [1] CRAN (R 4.1.0)
vctrs
                         0.3.8
                                 2021-04-29 [1] CRAN (R 4.1.0)
viridisLite
                         0.4.0
                                 2021-04-13 [1] CRAN (R 4.1.0)
withr
                         2.4.2
                                 2021-04-18 [1] CRAN (R 4.1.0)
xfun
                         0.26
                                 2021-09-14 [1] CRAN (R 4.1.1)
xm12
                         1.3.2
                                 2020-04-23 [1] CRAN (R 4.1.0)
yaml
                         2.2.1
                                 2020-02-01 [1] CRAN (R 4.1.0)
```

[1] /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/library

Stata session information for reproducibility

```
about
ado describe mrrobust
```

Stata/MP 16.1 for Mac (Apple Silicon) Revision 08 Jul 2021

```
Copyright 1985-2019 StataCorp LLC
Total physical memory: 8.01 GB
Stata license: Unlimited-user 2-core network, expiring 21 Jan 2022
Serial number: 501609352178
 Licensed to: Tom Palmer
             University of Bristol
[15] package mrrobust from https://raw.github.com/remlapmot/mrrobust/master
_____
TITLE
     'mrrobust': Stata package for two-sample Mendelian randomization analyses
DESCRIPTION/AUTHOR(S)
     Author: Tom Palmer
     Distribution-Date: 20210917
INSTALLATION FILES
     m/mrmedian.ado
     m/mrmedian.sthlp
     m/mrmedianobs.ado
     m/mrmedianobs work.ado
     m/mrmedianobs.sthlp
     m/mregger.ado
     m/mregger.sthlp
     m/mrrobust.sthlp
     m/mreggerplot.ado
     m/mreggerplot.sthlp
     m/mrmodal.ado
     m/mrmodal.sthlp
     m/mrratio.ado
     m/mrratio.sthlp
     m/mrivests.ado
     m/mrivests.sthlp
     m/mrforest.ado
     m/mrforest.sthlp
     m/mreggersimex.ado
     m/mreggersimex.sthlp
     m/mreggersimexonce.ado
     m/mrmodalplot.ado
     m/mrmodalplot.sthlp
     m/mrfunnel.ado
     m/mrfunnel.sthlp
     m/mrdeps.ado
     m/mrdeps.sthlp
```

m/mr.ado
m/mr.sthlp
m/mrmvivw.ado
m/mrmvivw.sthlp
m/mvivw.ado

```
m/mvivw.sthlp
m/mvmr.ado
m/mvmr.sthlp
m/mrmvegger.ado
m/mrmvegger.sthlp
m/mrleaveoneout.ado
m/mrleaveoneout.sthlp
m/mrrobust-author.ihlp
```

INSTALLED ON

24 Sep 2021

References

Burgess, S, F Dudbridge, and SG Thompson. 2015. "Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects." *American Journal of Epidemiology* 181: 251--260. https://doi.org/10.1093/aje/kwu283.

Rees, J, A Wood, and S Burgess. 2017. "Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy." Statistics in Medicine 36: 4705–18. https://doi.org/10.1002/sim.7492.

Sanderson, E, G Davey Smith, F Windmeijer, and J Bowden. 2019. "An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings." *International Journal of Epidemiology* 48: 713–27. https://doi.org/10.1093/ije/dyy262.

Sanderson, E, W Spiller, and J Bowden. 2021. "Testing and Correcting for Weak and Pleiotropic Instruments in Two-Sample Multivariable Mendelian Randomization." Statistics in Medicine. https://doi.org/10.1002/sim.9133.