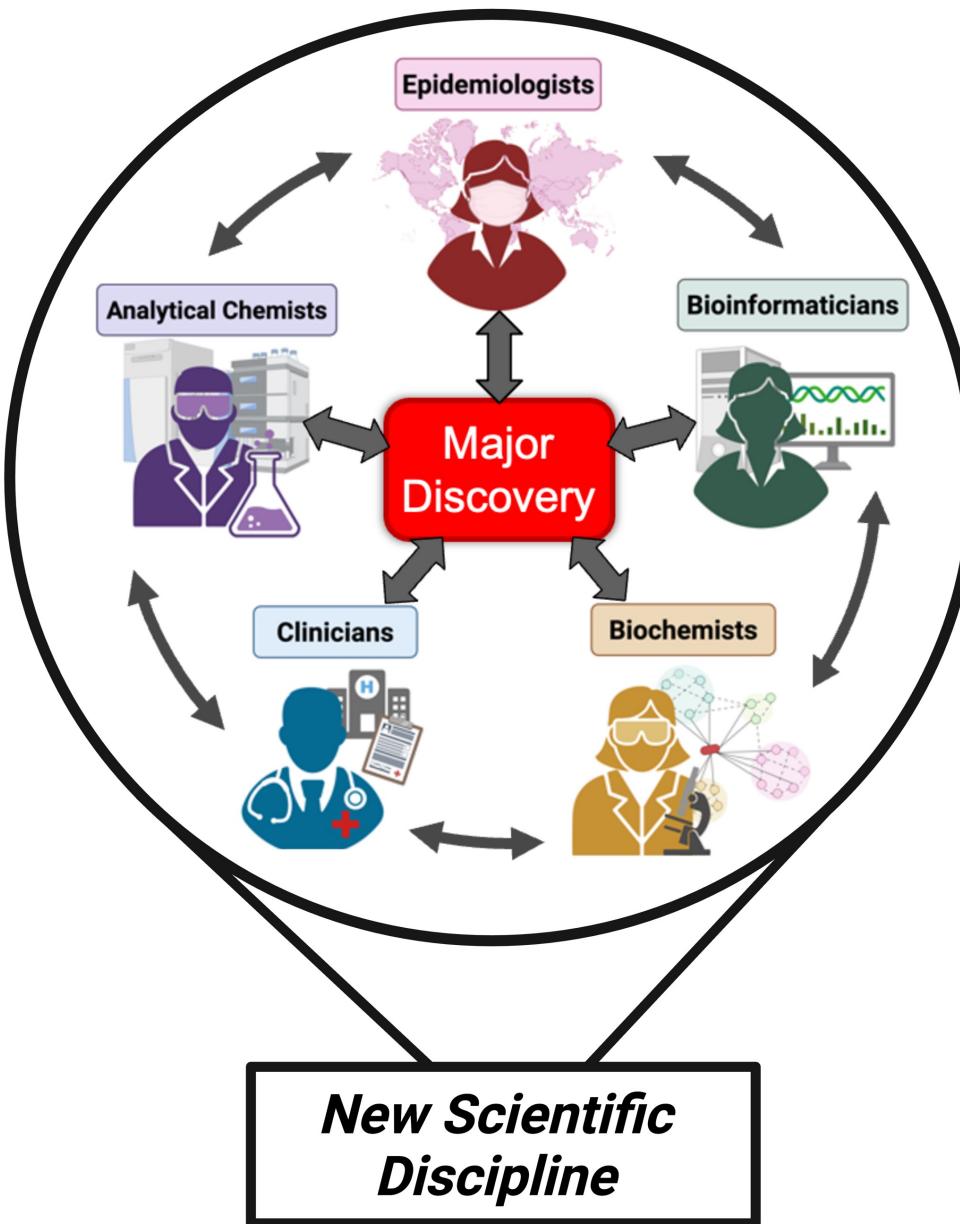


# *Conducting Single Cohort Analyses using COMETS Analytics*

Nicole Prince, Ph.D.

Postdoctoral Research Fellow, Harvard Medical School

May 15, 2024



# Why conduct single cohort analysis using COMETS Analytics?

- COMETS Analytics provides a streamlined approach to conducting metabolomics investigations
- Enhances **reproducibility** through standardized input sheet
- Enhances **rigor** by providing format by encouraging deliberate planning across interdisciplinary collaborators

Same process and input template for each cohort

Standardized  
Data and  
Models  
(user input)

Input Integrity  
Checks and  
Harmonization  
(automated)

Cohort  
analyses with  
Model Validity  
checks

Standardized  
Results

Meta-Analyses



## American Journal of Epidemiology



Society for  
Epidemiologic  
Research



JOHNS HOPKINS  
BLOOMBERG SCHOOL  
of PUBLIC HEALTH

### COMETS Analytics: An online tool for analyzing and meta-analyzing metabolomics data in large research consortia

Marinella Temprosa, Steven C Moore, Krista A Zanetti, Nathan Appel, David Ruggieri, Kaitlyn M Mazzilli, Kai-ling Chen, Rachel S Kelly, Jessica A Lasky-Su, Erikka Loftfield ... Show more

Author Notes

American Journal of Epidemiology, kwab120, <https://doi.org/10.1093/aje/kwab120>

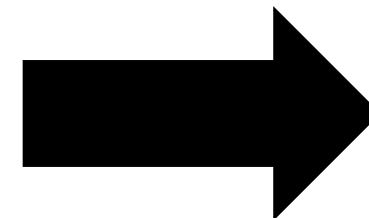
Published: 22 April 2021 Article history ▾

# *Single cohort analyses using COMETS Analytics GUI*

# Single Cohort Analysis using GUI

- After formatting the input sheet, investigators can perform analysis using GUI
- This may be a good option for those who are not comfortable coding or using the R package

*Scan the QR code or visit:*  
<https://www.comets-analytics.org/>



<https://bit.ly/3wt0aQ>

# Upload Input Sheet to GUI

COMETS Analytics Web Tool

HOME ANALYSIS ABOUT

**Cohort-Specific Analyses**

COMETS Cohort \*  
Other/Undefined

If not COMETS-specific, choose Other/Undefined

Custom Cohort \*  
test

If there are multiple datasets to be meta-analyzed from a single cohort, be sure to use a unique custom name for each dataset

Input Data File \*  
Choose File cometsInputAge.xlsx  
Download Sample Input

Reset Check Integrity

**Method of Analyses**

All models from the input file  
 Pre-specified models from the input file  
 Custom model

Integrity Check Results Heatmap

Integrity Check Successful  
Passed all integrity checks, analyses can proceed. If you are part of COMETS, please download metabolite list below and submit to the COMETS harmonization group.

**1. Upload input sheet**

The variable(s) multivitamin.2 have been removed from adjvars because of: correlated with another predictor

**Input Data Summary**

Total Metabolites 611 Subjects 1,000 Subject Covariates 19 Subject Metabolites 611  
Download Results

**Harmonization Summary**

N Metabolites 611 N Harmonized 611 N Non-Harmonized N with zero variance N with > 25% at min



<https://bit.ly/3wt0aQ>

# Upload Input Sheet to GUI

COMETS Analytics Web Tool

HOME ANALYSIS ABOUT

Cohort-Specific Analyses

COMETS Cohort \*

Other/Undefined

If not COMETS-specific, choose Other/Undefined

Custom Cohort \*

test

If there are multiple datasets to be meta-analyzed from a single cohort, be sure to use a unique custom name for each dataset

Input Data File \*

Choose File cometsInputAge.xlsx

Download Sample Input

Reset Check Integrity

Method of Analyses

All models from the input file

Pre-specified models from the input file

Custom model

Integrity Check Results Heatmap

Integrity Check Successful

Passed all integrity checks, analyses can proceed. If you are part of COMETS, please download metabolite list below and submit to the COMETS harmonization group.

2 Warnings

- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) multivitamin.2 have been removed from adjvars because of: correlated with another predictor

Input Data Summary

Total Metabolites 611 Subjects 1,000 Subject Covariates 19 Subject Metabolites 611

Harmonization

Download Results

611

th > 25% in

2. Choose models to run



<https://bit.ly/3wt0aQ>

# Upload Input Sheet to GUI

COMETS Analytics Web Tool

HOME ANALYSIS ABOUT

Cohort-Specific Analyses

COMETS Cohort \* Other/Defined If not COMETS-specific, choose Other/Defined

Custom Cohort \* test If there are multiple datasets to be meta-analyzed from a single cohort, be sure to use a unique custom name for each dataset

Input Data File \* Choose File cometsInputAge.xlsx Download Sample Input

Reset Check Integrity

Method of Analyses

All models from the input file  
 Pre-specified models from the input file  
 Custom model

A B C

MODEL	OUTCOMES	EXPOSURE	EXP
1 Age	All metabolites	age	
2 Age Spearman Min 10 Subjects	All metabolites	age	
3 Age Multivariable adjusted	All metabolites	age	
4 Age Multivariable adjusted stratified	All metabolites	age	
5 Age Multivariable adjusted stratified subset	All metabolites	age	
6 All pairwise metabolites	All metabolites	All metabolites	
7 Poisson regression	n_visits	All metabolites	
8 Logistic regression	nested_case	All metabolites	
9 Survival model	event	All metabolites	
10 Conditional logistic	nested_case	All metabolites	

Harmonization 611 1,000 19 611 th > 25% in

2. Choose models to run



<https://bit.ly/3wt0aQ>

# Integrity Check in GUI

COMETS Analytics Web Tool

HOME ANALYSIS ABOUT

Cohort-Specific Analyses

COMETS Cohort \*

Other/Undefined

If not COMETS-specific, choose Other/Undefined

Custom Cohort \*

test

If there are multiple datasets to be meta-analyzed from a single cohort, be sure to use a unique custom name for each dataset

Input Data File \*

Choose File cometsInputAge.xlsx

Download Sample Input

Reset Check Integrity

Method of Analyses

All models from the input file

Pre-specified models from the input file

Custom model

Integrity Check Results Heatmap

**Integrity Check Successful**

Passed all integrity checks, analyses can proceed. If you are part of COMETS, please download metabolite list below and submit to the COMETS harmonization group.

**2 Warnings**

- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) multivitamin.2 have been removed from adjvars because of: correlated with another predictor

**Input Data Summary**

Total Metabolites 611 Subjects 1,000 Subject Covariates 19 Subject Metabolites 611

**Harmonization Summary**

N Metabolites 611	N Harmonized 611	N Non-Harmonized	N with zero variance	N with > 25% at min
-------------------	------------------	------------------	----------------------	---------------------



<https://bit.ly/3wt0aQ>

# Obtaining Results with GUI

COMETS Analytics Web Tool

HOME ANALYSIS ABOUT

Cohort-Specific Analyses

COMETS Cohort \*

Other/Undefined

If not COMETS-specific, choose Other/Undefined

Custom Cohort \*

test

If there are multiple datasets to be meta-analyzed from a single cohort, be sure to use a unique custom name for each dataset

Input Data File \*

Choose File cometsInputAge.xlsx

Download Sample Input

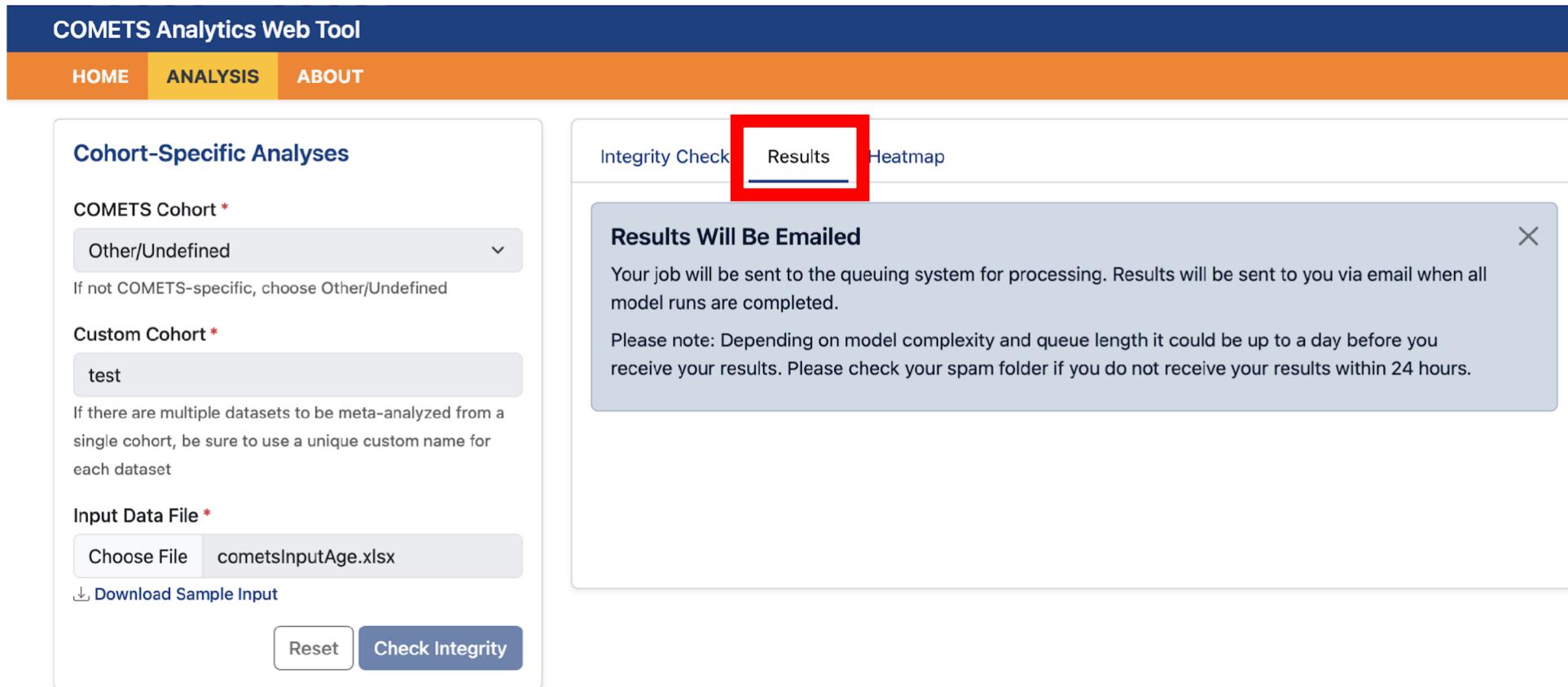
Reset Check Integrity

Integrity Check Results Heatmap

Results Will Be Emailed

Your job will be sent to the queuing system for processing. Results will be sent to you via email when all model runs are completed.

Please note: Depending on model complexity and queue length it could be up to a day before you receive your results. Please check your spam folder if you do not receive your results within 24 hours.



<https://bit.ly/3wt0aQ>

# Obtaining Results with GUI

**1 Age** (1.53 seconds)

**2 Age Spearman Min 10 Subjects** (1.07 seconds)

**3 Age Multivariable adjusted** (5.11 seconds)

**4 Age Multivariable adjusted stratified** (8.63 seconds)

Warnings:

- The variable(s) race\_grp.2 have been removed from adjvars because of: linearly dependent
- The variable(s) race\_grp.2 have been removed from adjvars because of: linearly dependent

**5 Age Multivariable adjusted stratified subset** (9.51 seconds)

Warnings:

- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) race\_grp.2, multivitamin.2 have been removed from adjvars because of: linearly dependent
- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) multivitamin.2 have been removed from adjvars because of: correlated with another predictor
- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) race\_grp.2, multivitamin.2 have been removed from adjvars because of: linearly dependent

**6 All pairwise metabolites** (2.66 seconds)

**7 Poisson regression** (3.89 seconds)

**8 Logistic regression** (3.39 seconds)

**9 Survival model** (5.92 seconds)

**10 Conditional logistic** (7.54 seconds)



<https://bit.ly/3wt0aQ>

# Obtaining Results with GUI

1 Age (1.53 seconds)

2 Age Spearman Min 10 Subjects (1.07 seconds)

3 Age Multivariable adjusted (5.11 seconds)

4 Age Multivariable adjusted stratified (8.63 se

Warnings:

- The variable(s) race\_grp.2 have been re
- The variable(s) race\_grp.2 have been re

5 Age Multivariable adjusted stratified subset (

Warnings:

- The variable(s) female, fasted have bee
- The variable(s) race\_grp.2, multivitamin
- The variable(s) female, fasted have bee
- The variable(s) multivitamin.2 have bee
- The variable(s) female, fasted have bee
- The variable(s) race\_grp.2, multivitamin

6 All pairwise metabolites (2.66 seconds)

7 Poisson regression (3.89 seconds)

8 Logistic regression (3.39 seconds)

9 Survival model (5.92 seconds)

10 Conditional logistic (7.54 seconds)

MODEL	OUTCOMES	EXPOSURE	A	B	C
			EXPOSURE	OUTCOMES	EXPOSURE
1 Age	All metabolites	age			
2 Age Spearman Min 10 Subjects	All metabolites	age			
3 Age Multivariable adjusted	All metabolites	age			
4 Age Multivariable adjusted stratified	All metabolites	age			
5 Age Multivariable adjusted stratified subset	All metabolites	age			
6 All pairwise metabolites	All metabolites	All metabolites			
7 Poisson regression	n_visits	All metabolites			
8 Logistic regression	nested_case	All metabolites			
9 Survival model	event	All metabolites			
10 Conditional logistic	nested_case	All metabolites			



<https://bit.ly/3wt0aQ>

# Obtaining Results with GUI

**1 Age** (1.53 seconds)

**2 Age Spearman Min 10 Subjects** (1.07 seconds)

**3 Age Multivariable adjusted** (5.11 seconds)

**4 Age Multivariable adjusted stratified** (8.63 seconds)

Warnings:

- The variable(s) race\_grp.2 have been removed from adjvars because of: linearly dependent
- The variable(s) race\_grp.2 have been removed from adjvars because of: linearly dependent

**5 Age Multivariable adjusted stratified subset** (9.51 seconds)

Warnings:

- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) race\_grp.2, multivitamin.2 have been removed from adjvars because of: linearly dependent
- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) multivitamin.2 have been removed from adjvars because of: correlated with another predictor
- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) race\_grp.2, multivitamin.2 have been removed from adjvars because of: linearly dependent

**6 All pairwise metabolites** (2.66 seconds)

**7 Poisson regression** (3.89 seconds)

**8 Logistic regression** (3.39 seconds)

**9 Survival model** (5.92 seconds)

**10 Conditional logistic** (7.54 seconds)



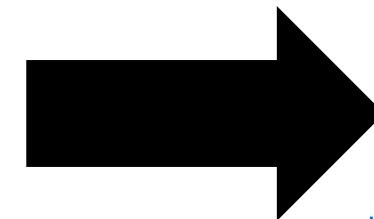
<https://bit.ly/3wt0aQ>

# *Single cohort analyses using COMETS Analytics R Package*

# Single Cohort Analysis using R Package

- Great for situations in which cohort analyst wants to optimize models
- Allows exploration of models outside of pre-specified parameters in input sheet

***Want to follow along? Scan QR code to access sample code and resources***



[https://bit.ly/NIH\\_NCATS\\_Met\\_Epi\\_workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

# Single Cohort Analysis using R Package

**Prespecified Models:** Already defined in input sheet

**Interactive Models:** useful for finalizing which models should be run

- Model is not predefined in the input sheet
- Model type and parameters specified in R

**Super-batch Mode:** quickest option for cohort analysts

- Very useful for running many models at once
- Results are standardized (e.g., one Excel file per model)

# Single Cohort Analysis using R Package

## *RcometsAnalytics Workflow*

### Step 1:

`readCOMETSinput()`

Read in input sheet  
with all data

### Step 2:

`getModelData()`

Prepares models to  
be run

### Step 3:

`runModel()`

Runs model as  
specified by  
`getModelData()`

# Single Cohort Analysis using R Package

## *RcometsAnalytics Workflow*

### Step 1:

**readCOMETSinput()**

**Read in input sheet  
with all data**

### Step 2:

**getModelData()**

**Prepares models to  
be run**

### Step 3:

**runModel()**

**Runs model as  
specified by  
getModelData()**



# Step 1: readCOMETSinput()

## Read input sheet and perform integrity checks

```
# Retrieve the full path of the input data  
dir <- system.file("extdata", package="RcometsAnalytics", mustWork=TRUE)  
csvfile <- file.path(dir, "cometsInputAge.xlsx")  
  
# Read in and process the input data  
exmetabdata <- RcometsAnalytics::readCOMETSinput(csvfile)
```

```
## VarMap sheet is read in.  
## Metabolites sheet is read in.  
## SubjectMetabolites sheet is read in.  
## SubjectData sheet is read in.  
## Models sheet is read in.  
## Model_Types sheet is read in.  
## There are 16 categorical variables.  
## Running Integrity Check...
```

```
## Joining with `by = join_by(hmdb_id)`
```



# Step 1: readCOMETSinput()

Read input sheet and perform integrity checks

```
# Retrieve the full path of the input data  
dir <- system.file("extdata", package="RcometsAnalytics", mustWork=TRUE)  
csvfile <- file.path(dir, "cometsInputAge.xlsx")  
  
# Read in and process the input data  
exmetabdata <- RcometsAnalytics::readCOMETSinput(csvfile)
```

```
## VarMap sheet is read in.  
## Metabolites sheet is read in.  
## SubjectMetabolites sheet is read in.  
## SubjectData sheet is read in.  
## Models sheet is read in.  
## Model_Types sheet is read in.  
## There are 16 categorical variables.  
## Running Integrity Check...
```

```
## Joining with `by = join_by(hmdb_id)`
```

[https://bit.ly/NIH\\_NCATS](https://bit.ly/NIH_NCATS)  
[Met Epi workshop](#)



# Step 1: readCOMETSinput()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

```
## Begin testing models in Models sheet...
```

```
## Filtering subjects according to the rule(s) age< 70. 836 of 1000 are retained.
```

```
## Warning in runModel.addRemVars(rem.obj, vars[oneVal], vars.type, "too few  
## unique non-missing values", : The variable(s) female, fasted have been removed  
## from adjvars because of: too few unique non-missing values
```

```
## Warning in runModel.addRemVars(rem.obj, tmp[rem], varSet, "correlated with  
## another predictor", : The variable(s) multivitamin.2 have been removed from  
## adjvars because of: correlated with another predictor
```

```
## Finished testing models in Models sheet.
```



# Step 1: readCOMETSinput()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

```
## Begin testing models in Models sheet...
## Filtering subjects according to the rule(s) age< 70. 836 of 1000 are retained.
```

```
## Warning in runModel.addRemVars(rem.obj, vars[oneVal], vars.type, "too few
## unique non-missing values", : The variable(s) female, fasted have been removed
## from adjvars because of: too few unique non-missing values
```

```
## Warning in runModel.addRemVars(rem.obj, tmp[rem], varSet, "correlated with
## another predictor", : The variable(s) multivitamin.2 have been removed from
## adjvars because of: correlated with another predictor
```

```
## Finished testing models in Models sheet.
```



# Step 1: readCOMETSinput()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

```
## Begin testing models in Models sheet...
## Filtering subjects according to the rule(s) age< 70. 836 of 1000 are retained.
```

```
## Warning in runModel.addRemVars(rem.obj, vars[oneVal], vars.type, "too few
## unique non-missing values", : The variable(s) female, fasted have been removed
## from adjvars because of: too few unique non-missing values
```

```
## Warning in runModel.addRemVars(rem.obj, tmp[rem], v
## another predictor", : The variable(s) multivitamin.
## adjvars because of: correlated with another predict
```

```
## Finished testing models in Models sheet.
```

From GUI →

#### Integrity Check Successful

Passed all integrity checks, analyses can proceed. If you are part of COMETS, please download metabolite list below and submit to the COMETS harmonization group.

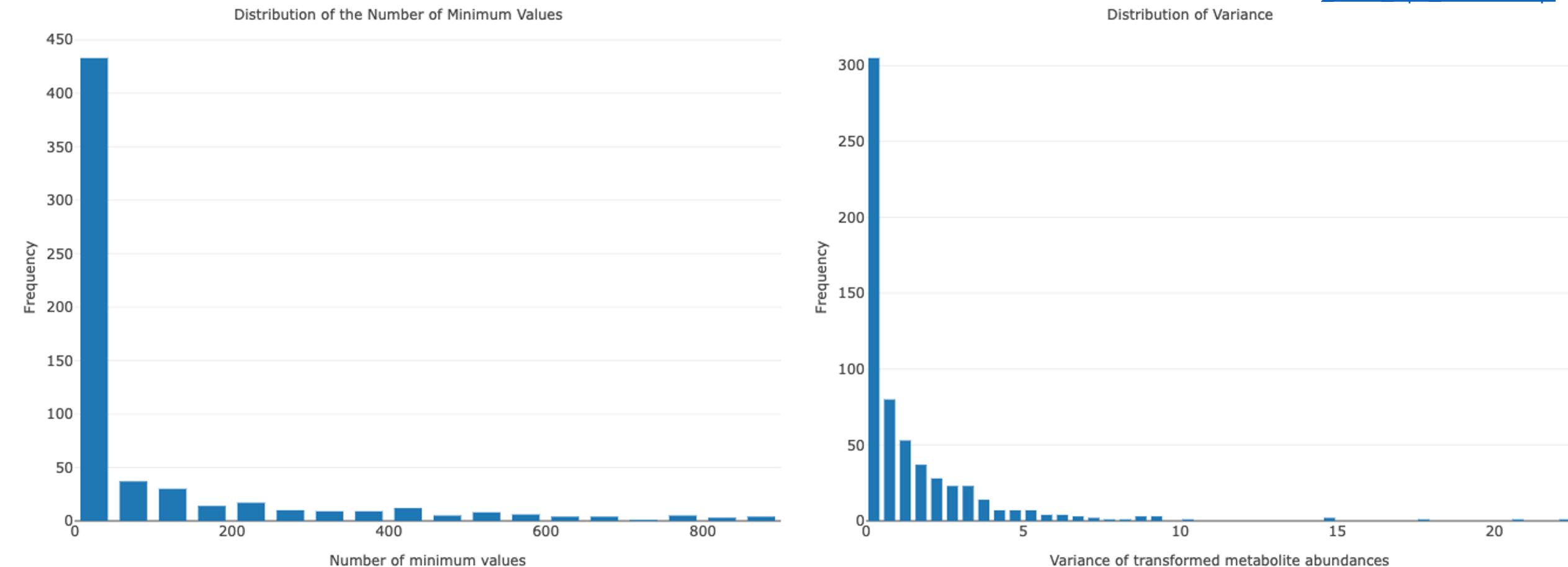
#### 2 Warnings

- The variable(s) female, fasted have been removed from adjvars because of: too few unique non-missing values
- The variable(s) multivitamin.2 have been removed from adjvars because of: correlated with another predictor



# Step 1: readCOMETSinput()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)



# Single Cohort Analysis using R Package

## *RcometsAnalytics Workflow*

### Step 1:

`readCOMETSinput()`

Read in input sheet  
with all data

### Step 2:

`getModelData()`

Prepares models to  
be run

### Step 3:

`runModel()`

Runs model as  
specified by  
`getModelData()`



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

**Prespecified Models:** Already defined in input sheet

**Interactive Models:** useful for finalizing which models should be run

- Model is not predefined in the input sheet
- Model type and parameters specified in R

**Super-batch Mode:** quickest option for cohort analysts

- Very useful for running many models at once
- Results are standardized (e.g., one Excel file per model)



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

**Prespecified Models:** Already defined in input sheet

**Interactive Models:** useful for finalizing which models should be run

- Model is not predefined in the input sheet
- Model type and parameters specified in R

**Super-batch Mode:** quickest option for cohort analysts

- Very useful for running many models at once
- Results are standardized (e.g., one Excel file per model)



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Set up pre-specified models based on input sheet

```
# set up "1 Age" model as specified in the input sheet
exmodeldata <- RcometsAnalytics::getModelData(exmetadata, modlabel="1 Age")
```

## Set up interactive mode to specify in R

```
# set up metabolites ~ age model using interactive mode
exmodeldata2 <- RcometsAnalytics::getModelData(exmetadata, modelspec="Interactive",
                                               exposures=c("age"))
```



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Set up pre-specified models based on input sheet

```
# set up "1 Age" model as specified in the input sheet
exmodeldata <- RcometsAnalytics::getModelData(exmetadata, modlabel="1 Age")
```

## Set up interactive mode to specify in R

```
# set up metabolites ~ age model using interactive mode
exmodeldata2 <- RcometsAnalytics::getModelData(exmetadata, modelspec="Interactive",
                                               exposures=c("age"))
```



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Set up pre-specified models based on input sheet

```
# set up "1 Age" model as specified in the input sheet
exmodeldata <- RcometsAnalytics::getModelData(exmetadata, modlabel="1 Age")
```

## Set up interactive mode to specify in R

```
# set up metabolites ~ age model using interactive mode
exmodeldata2 <- RcometsAnalytics::getModelData(exmetadata,
                                               exposures=c("age"))
```

From Input Sheet →

MODEL	OUTCOMES	EXPOSURE	EXCLUDED
1 Age	All metabolites	age	
2 Age Spearman Min 10 Subjects	All metabolites	age	
3 Age Multivariable adjusted	All metabolites	age	
4 Age Multivariable adjusted stratified	All metabolites	age	
5 Age Multivariable adjusted stratified subset	All metabolites	age	



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

**Prespecified Models:** Already defined in input sheet

**Interactive Models:** useful for finalizing which models should be run

- Model is not predefined in the input sheet
- Model type and parameters specified in R

**Super-batch Mode:** quickest option for cohort analysts

- Very useful for running many models at once
- Results are standardized (e.g., one Excel file per model)



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Set up pre-specified models based on input sheet

```
# set up "1 Age" model as specified in the input sheet
exmodeldata <- RcometsAnalytics::getModelData(exmetadata, modlabel="1 Age")
```

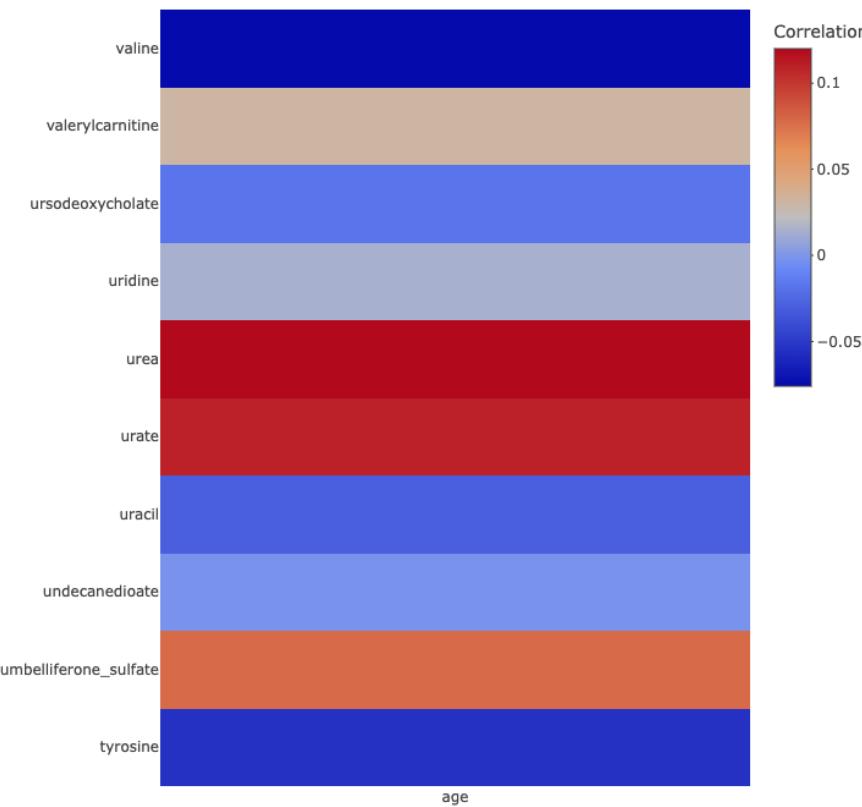
## Set up interactive mode to specify in R

```
# set up metabolites ~ age model using interactive mode
exmodeldata2 <- RcometsAnalytics::getModelData(exmetadata, modelspec="Interactive",
exposures=c("age"))
```

# Simple Plots with RcometsAnalytics



[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)



- **RcometsAnalytics can produce simple plots**
- **Allows easy comparison between models**



# Step 2: getModelData()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Set up and run stratified correlation analysis

```
exmodeldata2 <- RcometsAnalytics::getModelData(exmetadata, modelspec="Interactive",  
                                               outcomes=c("lactose", "lactate"),  
                                               exposures=c("age", "bmi_grp"), strvars="race_grp")
```

```
excorrdata2 <- RcometsAnalytics::runModel(exmodeldata2, exmetadata, "test_cohort")
```

```
RcometsAnalytics::showModel(excorrdata2, nlines=3)
```

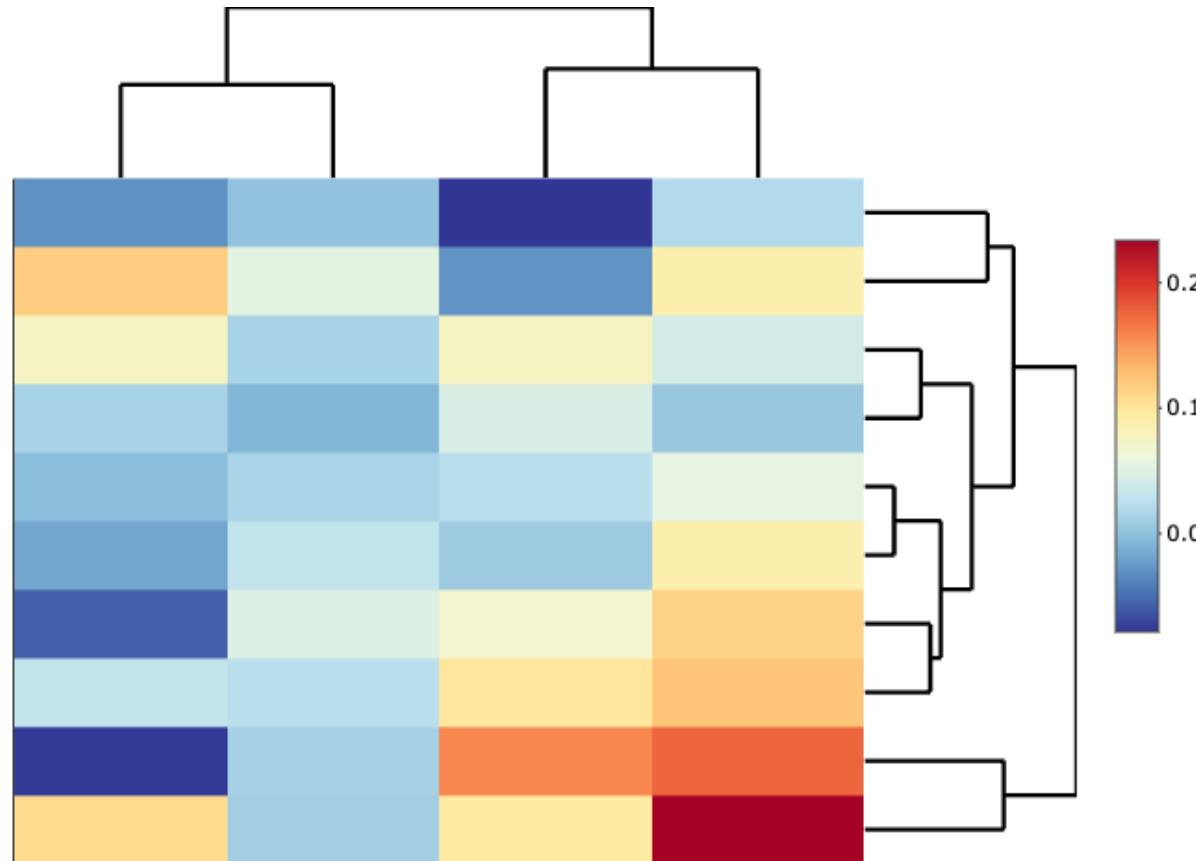


Allows specification of metabolites,  
outcomes, stratification, etc.

# Simple Plots with RcometsAnalytics



[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)



- With 2 or more outcome and exposure specifications, can produce heatplot with hierarchical clustering

# Single Cohort Analysis using R Package

## *RcometsAnalytics Workflow*

### Step 1:

`readCOMETSinput()`

Read in input sheet  
with all data

### Step 2:

`getModelData()`

Prepares models to  
be run

### Step 3:

`runModel()`

**Runs model as  
specified by  
`getModelData()`**



# Step 3: runModel()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Run pre-specified models based on input sheet

```
excorrdata <- RcometsAnalytics::runModel(exmodeldata, exmetadata, "test_cohort")
RcometsAnalytics::OutputListToExcel(filename="test_cohort_corr1.xlsx", excorrdata)

## Output saved to file: test_cohort_corr1.xlsx
```



# Step 3: runModel()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

Run pre-specified models based on input sheet

```
excorrdata <- RcometsAnalytics::runModel(exmodeldata, exmetadata, "test_cohort")
RcometsAnalytics::OutputListToExcel(filename = "test_cohort_corr1.xlsx", excorrdata)
```

```
## Output saved to file: test_cohort_corr1.xlsx
```



**Call model data  
specified with  
Step 2:  
getModelData()**



# Step 3: runModel()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Run pre-specified models based on input sheet

```
excorrdata <- RcometsAnalytics::runModel(exmodeldata, exmetadata, test_cohort)  
RcometsAnalytics::OutputListToExcel(filename="test_cohort_corr1.xlsx", excorrdata)  
  
## Output saved to file: test_cohort_corr1.xlsx
```



**Call metabolite  
data from input  
sheet in Step 1:  
readCometsInput()**



# Step 3: runModel()

[https://bit.ly/NIH\\_NCATS  
Met Epi workshop](https://bit.ly/NIH_NCATS_Met_Epi_workshop)

## Run pre-specified models based on input sheet

```
excorrdata <- RcometsAnalytics::runModel(exmodeldata, exmetadata, "test_cohort")
RcometsAnalytics::OutputListToExcel(filename="test_cohort_corr1.xlsx", excorrdata)

## Output saved to file: test_cohort_corr1.xlsx
```



**Name cohort for  
easy file saving**

# Single Cohort Analysis using R Package

**Prespecified Models:** Already defined in input sheet

**Interactive Models:** useful for finalizing which models should be run

- Model is not predefined in the input sheet
- Model type and parameters specified in R

**Super-batch Mode:** quickest option for cohort analysts

- Very useful for running many models at once
- Results are standardized (e.g., one Excel file per model)

# Super-batch mode in RcometsAnalytics

```
exallmodels <- RcometsAnalytics::runAllModels(exmetadata, cohortLabel = "test_cohort") # whatever cohortLabel you choose will be used in autosaved files of results
```

```
## Running 1 Age  
## Output saved to file: 1_Age_test_cohort_2024-05-08.xlsx  
## Running 2 Age Spearman Min 10 Subjects  
## Output saved to file: 2_Age_Spearman_Min_10_Subjects_test_cohort_2024-05-08.xlsx  
## Running 3 Age Multivariable adjusted  
## Output saved to file: 3_Age_Multivariable_adjusted_test_cohort_2024-05-08.xlsx  
## Running 4 Age Multivariable adjusted stratified
```

```
## Warning in runModel.addRemVars(rem.obj, tmp[rem], varSet, "linearly dependent",  
## : The variable(s) race_grp.2 have been removed from adjvars because of:  
## linearly dependent  
## Warning in runModel.addRemVars(rem.obj, tmp[rem], varSet, "linearly dependent",  
## : The variable(s) race_grp.2 have been removed from adjvars because of:  
## linearly dependent
```

# Super-batch mode in RcometsAnalytics

```
exallmodels <- RcometsAnalytics::runAllModels(exmetadata, cohortLabel = "test_cohort") # whatever cohortLabel you choose will be used in autosaved files of results
```

```
## Running 1 Age  
## Output saved to file: 1_Age_test_cohort_2024-05-08.xlsx  
## Running 2 Age Spearman Min 10 Subjects  
## Output saved to file: 2_Age_Spearman_Min_10_Subjects_test_cohort_2024-05-08.xlsx  
## Running 3 Age Multivariable adjusted  
## Output saved to file: 3_Age_Multivariable_adjusted_test_cohort_2024-05-08.xlsx  
## Running 4 Age Multivariable adjusted stratified
```

```
## Warning in runModel.addRemVars(rem.obj, tmp[rem], varSet, "linearly dependent",  
## : The variable(s) race_grp.2 have been removed from adjvars because of:  
## linearly dependent  
## Warning in runModel.addRemVars(rem.obj, tmp[rem], varSet, "linearly dependent",  
## : The variable(s) race_grp.2 have been removed from adjvars because of:  
## linearly dependent
```

# Single Cohort Analysis using R Package

## *RcometsAnalytics Workflow*

Step 1:

readCOM

Read in in  
with al

Step 2:

**Completed!**

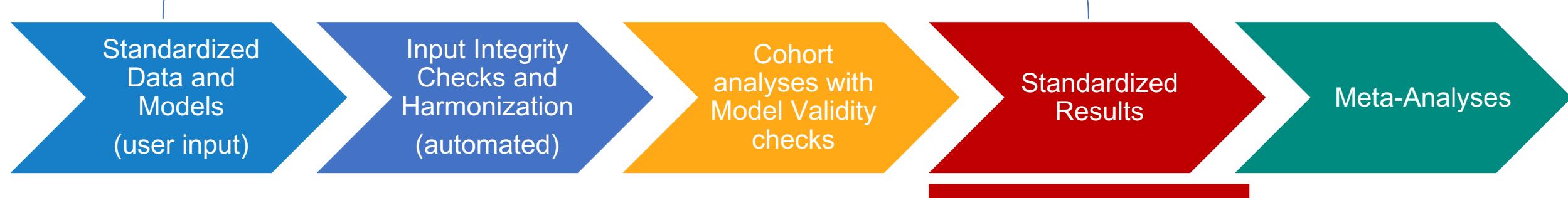
Step 3:

odel()

odel as  
ied by

getModelData()

Same process and input template for each cohort



## American Journal of Epidemiology



Society for  
Epidemiologic  
Research



JOHNS HOPKINS  
BLOOMBERG SCHOOL  
of PUBLIC HEALTH

### COMETS Analytics: An online tool for analyzing and meta-analyzing metabolomics data in large research consortia

Marinella Temprosa, Steven C Moore, Krista A Zanetti, Nathan Appel, David Ruggieri, Kaitlyn M Mazzilli, Kai-ling Chen, Rachel S Kelly, Jessica A Lasky-Su, Erikka Loftfield ... Show more

Author Notes

American Journal of Epidemiology, kwab120, <https://doi.org/10.1093/aje/kwab120>

Published: 22 April 2021 Article history ▾

# COMETS Analytics Results Output

A	B	C	D	E	F	G	H	I	J	
1	run	outcomespec	exposurespec	term	nobs	message	adjvars	adjvars.removed	adjspec	outcome_uid
2	1	_1_2_3_benzenetriol_sulfate_2	age		1000					CHEM100006374
3	2	_1_2_dipalmitoylglycerol	age		1000					HMDB07098
4	3	_1_2_propanediol	age		1000					HMDB01881
5	4	_1_3_7_trimethylurate	age		1000					HMDB02123
6	5	_1_3_dimethylurate	age		1000					HMDB01857
7	6	_1_3_dipalmitoylglycerol	age		1000					HMDB31011
8	7	_1_5_anhydroglucitol_1_5_ag_	age		1000					HMDB02712
9	8	_1_6_anhydroglucose	age		1000					HMDB00640
10	9	_1_7_dimethylurate	age		1000					HMDB11103
11	10	_1_arachidonoylglycerophoscho	age		1000					HMDB10395
12	11	_1_arachidonoylglycerophoseth	age		1000					HMDB11517
13	12	_1_arachidonoylglycerophosino	age		1000					HMDB61690
14	13	_1_arachidonoylglyercophospha	age		1000					CHEM100004442
15	14	_1_arachidonylglycerol	age		1000					HMDB11549_HMDB11572
16	15	_1_dihomo_linolenylglycerol_	age		1000					CHEM100006121
17	16	_1_docosahexaenoylglycerol	age		1000					HMDB11587
18	17	_1_linolenoylglycerol	age		1000					HMDB11569
19	18	_1_linolenoylglycerophosphoch	age		1000					HMDB10388
20	19	_1_linoleoylglycero	age		1000					CHEM100001040
21	20	_1_linoleoylglycerophosphocho	age		1000					HMDB10386
22	21	_1_linoleoylglycerophosphoeth	age		1000					HMDB11507
23	22	_1_linoleoylglycerophosphoino	age		1000					CHEM100001778

ModelSummary

Effects

Errors\_Warnings

Table1

Info

+

# COMETS Analytics Results Output

	A	B	C	D	E	F	G	H
1	run	outcomespec	exposurespec	term	estimate	pvalue	pvalue.adj	metabolite_name
2	1	_1_2_3_benzenetriol_sulfate_2	age	age	0.0196317	0.53519711	0.674238009	1,2,3-benzenetriol sulfate (2)
3	2	_1_2_dipalmitoylglycerol	age	age	-0.002059914	0.94812718	0.975262133	1,2-dipalmitoylglycerol
4	3	_1_2_propanediol	age	age	-0.018211498	0.565138774	0.70325823	1,2-propanediol
5	4	_1_3_7_trimethylurate	age	age	-0.005041788	0.87348221	0.918432465	1,3,7-trimethylurate
6	5	_1_3_dimethylurate	age	age	-0.033353691	0.292015562	0.454988621	1,3-dimethylurate
7	6	_1_3_dipalmitoylglycerol	age	age	-0.037405635	0.237283406	0.387647489	1,3-dipalmitoylglycerol
8	7	_1_5_anhydroglucitol_1_5_ag_	age	age	-0.062754328	0.047262559	0.1173879	1,5-anhydroglucitol (1,5-AG)
9	8	_1_6_anhydroglucose	age	age	0.014837138	0.639333801	0.751217217	1,6-anhydroglucose
10	9	_1_7_dimethylurate	age	age	-0.000926926	0.976645018	0.992339856	1,7-dimethylurate
11	10	_1_arachidonoylglycerophoscho	age	age	-0.043015271	0.174086348	0.305651604	1-arachidonoylglycerophosphocholine (20:4n6)*
12	11	_1_arachidonoylglycerophoseth	age	age	0.023423856	0.459359187	0.611478133	1-arachidonoylglycerophosphoethanolamine*
13	12	_1_arachidonoylglycerophosino	age	age	0.007963309	0.801419088	0.872766999	1-arachidonoylglycerophosphoinositol*
14	13	_1_arachidonoylglyercophospha	age	age	-0.080660255	0.010720813	0.03574946	1-arachidonoylglyercophosphate
15	14	_1_arachidonylglycerol	age	age	0.017039721	0.590433173	0.719782498	1-arachidonylglycerol
16	15	_1_dihomo_linolenylglycerol_	age	age	-0.001828356	0.953951406	0.97796025	1-dihomo-linolenylglycerol (alpha, gamma)
17	16	_1_docosahexaenoylglycerol	age	age	0.029746118	0.347378508	0.515165699	1-docosahexaenoylglycerol
18	17	_1_linolenoylglycerol	age	age	0.043133719	0.172902584	0.304448066	1-linolenoylglycerol
19	18	_1_linolenoylglycerophosphoch	age	age	-0.026908831	0.395313085	0.564337138	1-linolenoylglycerophosphocholine (18:3n3)*
20	19	_1_linoleoylglycero	age	age	0.052351224	0.098014768	0.199623411	1-linoleoylglycerol (1-monolinolein)
21	20	_1_linoleoylglycerophosphocho	age	age	-0.038000112	0.2299065	0.378633078	1-linoleoylglycerophosphocholine (18:2n6)
22	21	_1_linoleoylglycerophosphoeth	age	age	0.00073162	0.981564981	0.994587402	1-linoleoylglycerophosphoethanolamine*
23	22	_1_linoleoylglycerophosphoino	age	age	-0.024161788	0.445333023	0.601987781	1-linoleoylglycerophosphoinositol*

ModelSummary Effects Errors\_Warnings Table1 Info +

# COMETS Analytics Results Output

	A	B	C	D	E	F	G	H
1	run	outcomespec	exposurespec	term	estimate	pvalue	pvalue.adj	metabolite_name
2	1	_1_2_3_benzenetriol_sulfate_2	age	age	0.0196317	0.53519711	0.674238009	1,2,3-benzenetriol sulfate (2)
3	2	_1_2_dipalmitoylglycerol	age	age	-0.002059914	0.94812718	0.975262133	1,2-dipalmitoylglycerol
4	3	_1_2_propanediol	age	age	-0.018211498	0.565138774	0.70325823	1,2-propanediol
5	4	_1_3_7_trimethylurate	age	age	-0.005041788	0.87348221	0.918432465	1,3,7-trimethylurate
6	5	_1_3_dimethylurate	age	age	-0.033353691	0.292015562	0.454988621	1,3-dimethylurate
7	6	_1_3_dipalmitoylglycerol	age	age	-0.037405635	0.237283406	0.387647489	1,3-dipalmitoylglycerol
8	7	_1_5_anhydroglucitol_1_5_ag_	age	age	-0.062754328	0.047262559	0.1173879	1,5-anhydroglucitol (1,5-AG)
9	8	_1_6_anhydroglucose	age	age	0.014837138	0.639333801	0.751217217	1,6-anhydroglucose
10	9	_1_7_dimethylurate	age	age	-0.000926926	0.976645018	0.992339856	1,7-dimethylurate
11	10	_1_arachidonoylglycerophoscho	age	age	-0.043015271	0.174086348	0.305651604	1-arachidonoylglycerophosphocholine (20:4n6)*
12	11	_1_arachidonoylglycerophoseth	age	age	0.023423856	0.459359187	0.611478133	1-arachidonoylglycerophosphoethanolamine*
13	12	_1_arachidonoylglycerophosino	age	age	0.007963309	0.801419088	0.872766999	1-arachidonoylglycerophosphoinositol*
14	13	_1_arachidonoylglyercophospha	age	age	-0.080660255	0.010720813	0.03574946	1-arachidonoylglyercophosphate
15	14	_1_arachidonylglycerol	age	age	0.017039721	0.590433173	0.719782498	1-arachidonylglycerol
16	15	_1_dihomo_linolenylglycerol_	age	age	-0.001828356	0.953951406	0.97796025	1-dihomo-linolenylglycerol (alpha, gamma)
17	16	_1_docosahexaenoylglycerol	age	age	0.029746118	0.347378508	0.515165699	1-docosahexaenoylglycerol
18	17	_1_linolenoylglycerol	age	age	0.043133719	0.172902584	0.304448066	1-linolenoylglycerol
19	18	_1_linolenoylglycerophosphoch	age	age	-0.026908831	0.395313085	0.564337138	1-linolenoylglycerophosphocholine (18:3n3)*
20	19	_1_linoleoylglycero	age	age	0.052351224	0.098014768	0.199623411	1-linoleoylglycerol (1-monolinolein)
21	20	_1_linoleoylglycerophosphocho	age	age	-0.038000112	0.2299065	0.378633078	1-linoleoylglycerophosphocholine (18:2n6)
22	21	_1_linoleoylglycerophosphoeth	age	age	0.00073162	0.981564981	0.994587402	1-linoleoylglycerophosphoethanolamine*
23	22	_1_linoleoylglycerophospho..._1	age	age	0.024161788	0.445333023	0.601987781	1-linoleoylglycerophosphoinositol*

ModelSummary

Effects

Errors\_Warnings

Table1

Info

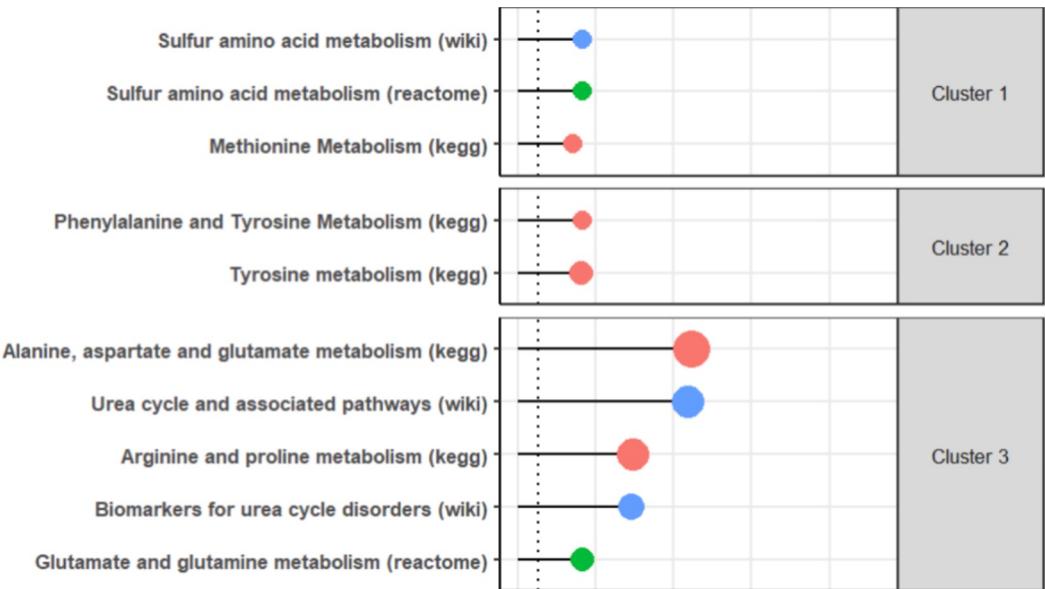
# *COMETS Analytics to Aid Interpretation in Metabolomic Epidemiology*

# RaMP in COMETS Analytics

- (*Relational database of Metabolomics Pathways*)  
RaMP can be used to visualize results and aid interpretation
  - Integrated into the RcometsAnalytics package
- Easy way to:
  - Integrate additional ‘omics
  - Perform pathway enrichments
  - And more!

# RaMP in COMETS Analytics

- Easily visualize COMETS Analytics results using RaMP plotting tools
- Identify important pathways to aid biological interpretation



RaMP Vignette  
<https://bit.ly/3UA65EA>

# *Final Thoughts*

# **COMETS Analytics is continuously being updated and improved!**

1. Do you have suggested features or changes you would like to see incorporated?
2. Could COMETS Analytics be applied to your own analyses?

# Acknowledgements

## COMETS Data Infrastructure Group

Steven Moore (NCI)  
Ella Temprosa (GWU)  
Eoin Fahy (UCSD)  
Vicky Stevens (DLH Corp.)

And others in COMETS Data Infrastructure Group!

## Developers and Bioinformaticians

Bill Wheeler (IMS)  
Kailing Chen (CBIIT)  
Brian Park (CBIIT)  
Madu Kanigicherla (CBIIT)  
Phyllip Cho (CBIIT)  
Andy Patt (NCATS/NIH)  
John Braisted (NCATS/NIH)

## COMETS Early Career Working Group

Haley Chatelaine (NCATS)  
Waylon Hastings (Tulane)  
Kate Shutta (HSPH)  
Ivy Tran (BWH)

## Funding

NCATS IRP  
NCI DCCPS  
NCI DCEG  
NIH STRIDES

***Users and other COMETS Interest Groups!  
Thank you for your feedback!***

Contact us at [comets.analytics@gmail.com](mailto:comets.analytics@gmail.com)