

# Epidemiology is a data science – if only epidemiologists knew

Success stories of modernizing data science practice and education in epidemiology

Konrad H. Stopsack, BIPS

Travis A. Gerke, cStructure

Emily Riederer, Capital One

Malcolm Barrett, Stanford University

Wednesday, June 11, 2025

# Data Science

Medicine

Epidemiology

Informatics

Biostatistics

...

Health Science

Population Science

Computer Science

Data Science

... Science



# Data Science

Medicine

Health Science

Epidemiology

Population Science

Informatics

Computer Science

Biostatistics

Data Science

...

... Science

## Epidemiologic data science

Caring for our data before we put it into a model or into a paper.

# Epidemiologists care about their data



# Epidemiologists care about their data



Eur J Epidemiol (2017) 32:863–865  
DOI 10.1007/s10654-017-0314-3

ESSAY

## The growing rift between epidemiologists and their data

Kenneth J. Rothman<sup>1,2</sup>

Cutter Lecture, Harvard School of Public Health, December 2014

# Welcome to this SER session!



Konrad Stopsack (BIPS)  
*Epidemiologic research*

**Table creation**

# Welcome to this SER session!



Konrad Stopsack (BIPS)  
*Epidemiologic research*  
**Table creation**



Travis Gerke (PCCTC/cStructure)  
*Clinical trials/Startup*  
**DAG/LLM-based causal workflows**

# Welcome to this SER session!



Konrad Stopsack (BIPS)  
*Epidemiologic research*

**Table creation**



Travis Gerke (PCCTC/cStructure)  
*Clinical trials/Startup*  
**DAG/LLM-based causal workflows**



Emily Riederer (Capital One)  
*Finance*  
**Project organization and workflows**

# Welcome to this SER session!



Konrad Stopsack (BIPS)  
*Epidemiologic research*

**Table creation**

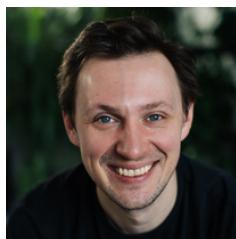


Travis Gerke (PCCTC/cStructure)  
*Clinical trials/Startup*

**DAG/LLM-based causal workflows**

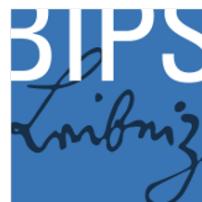


Emily Riederer (Capital One)  
*Finance*  
**Project organization and workflows**



Malcolm Barrett (Stanford)  
*Research/programming*  
**Code review**

# Questions?



Please submit your questions using the Whova App as we go along.

A screenshot of the Whova mobile application interface. At the top, there's a blue header bar with the time '10:41', signal strength, '5G', battery level '100%', a back arrow labeled 'Home', a 'Details' button, and a share icon. Below the header, a circular badge indicates 'Symposia'. The main content area displays a summary of a symposium titled 'Epidemiology is a data science – if only epidemiologists knew: Success stories of modernizing data science practice and education in epidemiology'. It shows the date 'Wed, Jun 11', time '10:15 AM - 11:45 AM', and location 'Commonwealth'. Below this, there are sections for '122 Attending', '0 Session Polls', '3 Likes', '0 Comments', and a red-bordered section for '0 Questions Asked'. A 'See more' link is visible at the bottom of the summary. At the very bottom, there's a 'Speaker' section with a profile picture and name 'Konrad Stopsack'.

# Questions?

Please submit your questions using the Whova App as we go along.

10:41 5G 100%

< Home Details ⌂

Symposia

**Epidemiology is a data science – if only epidemiologists knew: Success stories of modernizing data science practice and education in epidemiology**

Wed, Jun 11  
10:15 AM - 11:45 AM

Location: Commonwealth

122 Attending

0 Session Polls

3 Likes

0 Comments

0 Questions Asked

Overview

Chair(s): Konrad Stopsack Presenters: · Konrad Stopsack - Use Case 1: No more hand-typed numbers – Epidemiologic results tables in 20...

See more

Speaker

Konrad Stopsack

10:42 5G 100%

< Questions for Epidemiology is a d... ⌂

Sort: Votes Filter: All

Ask a Question

View as Attendee

# Use Case 1

## No more hand-typed numbers – Epidemiologic results tables in 2025

Konrad H. Stopsack

Professor and Chair, Department of Epidemiologic Methods and Etiologic Research  
Leibniz Institute for Prevention Research and Epidemiology – BIPS  
Bremen, Germany

I have no relevant financial relationships to disclose.

# How do our analyses produce results?

## 1. Write code in a text editor

```
1 data work.cribriiform;
2   merge work.cases cribri.cribriiform_final;
3   by id;
4   if missing(cribriiform) then delete;
5   label cribriiform ='Morphology';
6   run;
7
8 proc phreg data=work.cribriiform;
9   class gleason (ref='2') cribriiform(ref='0');
10  model lethalfu*eventlethal(0)=cribriiform gleason / ties=efron;
11  hazardratio cribriiform / at(gleason=ref) diff=ref;
12  output out=phres wtresschoen=crib_res;
13 run;
```

# How do our analyses produce results?

1. Write code in a text editor

2. Submit code to a server that houses the data

```
1 data work.cribriiform;
2   merge work.cases cribri.cribriiform_final;
3   by id;
4   if missing(cribriiform) then delete;
5   label cribriiform ='Morphology';
6   run;
7
8 proc phreg data=work.cribriiform;
9   class gleason (ref='2') cribriiform(ref='0');
10  model lethalfu*eventlethal(0)=cribriiform gleason / ties=efron;
11  hazardratio cribriiform / at(gleason=ref) diff=ref;
12  output out=phres wtresschoen=crib_res;
13 run;
```

# How do our analyses produce results?

1. Write code in a text editor
2. Submit code to a server that houses the data
3. Receive a log file containing the results

```

1  data work.cribriform;
2    merge work.cases cribri.cribriform_final;
3    by id;
4    if missing(cribriform) then delete;
5    label cribriform ='Morphology';
6    run;
7
8  proc phreg data=work.cribriform;
9    class gleason (ref='2') cribriform(ref='0');
10   model lethalfu*eventlethal(0)=cribriform gleason / ties=efron;
11   hazardratio cribriform / at(gleason=ref) diff=ref;
12   output out=phres wtresschoen=crib_res;
13  run;

```

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
cribriform	1	0.97671	0.21231	21.1646	<.0001	2.656 Morphology

Hazard Ratios for Morphology				
Description	Point Estimate	Confidence Limits	95% Wald	
cribriform Unit=1	2.656	1.752 4.026		

# How do our analyses produce results?

1. Write code in a text editor
2. Submit code to a server that houses the data
3. Receive a log file containing the results
4. Type the results into Word

```

1 data work.cribriiform;
2 merge work.cases cribri.cribriiform_final;
3 by id;
4 if missing(cribriiform) then delete;
5 label cribriiform ='Morphology';
6 run;

7
8 proc phreg data=work.cribriiform;
9   class gleason (ref='2') cribriiform(ref='0');
10  model lethalfu*eventlethal(0)=cribriiform gleason / ties=efron;
11  hazardratio cribriiform / at(gleason=ref) diff=ref;
12  output out=phres wtresschoen=crib_res;
13 run;

```



**Table 3.** Lethal disease associated with invasive cribriform 4 (NC4) in the prospective prostate cancer cohorts with and the Physicians' Health Study.

	Total	Lethal	HR <sup>a</sup>	95% CI <sup>a</sup>
Overall	218	43		
Unadjusted			2.66	1.75 to 4.03
Model 1: + Gleason			1.62	1.05 to 2.49
Model 2: + Age, BMI			1.67	1.08 to 2.59
Model 3: + cTNM			1.45	0.92 to 2.27

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Hazard Ratio
cribriiform	1	0.97671	0.21231	21.1646	<.0001 2.656 Morphology

Hazard Ratios for Morphology			
Description	Point Estimate	95% Wald Confidence Limits	
cribriiform Unit=1	2.656	1.752 4.026	

This process is such a pain

# This process is such a pain

1. It is unclear which data were read in and which code produced which result

# This process is such a pain

1. It is unclear which data were read in and which code produced which result
2. Numbers are manually transcribed and rounded

# This process is such a pain

1. It is unclear which data were read in and which code produced which result
2. Numbers are manually transcribed and rounded
3. Even the tiniest change to the analysis requires redoing many manual steps

# This process is such a pain

1. It is unclear which data were read in and which code produced which result
2. Numbers are manually transcribed and rounded
3. Even the tiniest change to the analysis requires redoing many manual steps

# Wish list for epidemiologic results tables, 2025

## This process is such a pain

1. It is unclear which data were read in and which code produced which result
2. Numbers are manually transcribed and rounded
3. Even the tiniest change to the analysis requires redoing many manual steps

## Wish list for epidemiologic results tables, 2025

1. Descriptive and inferential results should go hand-in-hand

## This process is such a pain

1. It is unclear which data were read in and which code produced which result
2. Numbers are manually transcribed and rounded
3. Even the tiniest change to the analysis requires redoing many manual steps

## Wish list for epidemiologic results tables, 2025

1. Descriptive and inferential results should go hand-in-hand
2. Code should produce a publication-ready table

## This process is such a pain

1. It is unclear which data were read in and which code produced which result
2. Numbers are manually transcribed and rounded
3. Even the tiniest change to the analysis requires redoing many manual steps

## Wish list for epidemiologic results tables, 2025

1. Descriptive and inferential results should go hand-in-hand
2. Code should produce a publication-ready table
3. Sensitivity analyses should be easy and quick

## This process is such a pain

1. It is unclear which data were read in and which code produced which result
2. Numbers are manually transcribed and rounded
3. Even the tiniest change to the analysis requires redoing many manual steps

## Wish list for epidemiologic results tables, 2025

1. Descriptive and inferential results should go hand-in-hand
2. Code should produce a publication-ready table
3. Sensitivity analyses should be easy and quick

*Let us make creating epidemiologic results tables easy, reproducible, and fun!*

# Example data built into R

```
data(flchain, package = "survival")
```

Monoclonal gammopathy of unclear significance (MGUS) and mortality, Olmsted County, MN, 1995–2003

age	sex	sample_yr	creatinine	MGUS status	futime	death
97	F	1997	1.7	No MGUS	0.23	1
92	F	2000	0.9	No MGUS	3.51	1
94	F	1997	1.4	No MGUS	0.19	1
92	F	1996	1.0	No MGUS	0.31	1
93	F	1996	1.1	No MGUS	2.84	1
90	F	1997	1.0	No MGUS	3.71	1
90	F	1996	0.8	No MGUS	7.81	1
90	F	1999	1.2	No MGUS	1.02	1
93	F	1996	1.2	No MGUS	9.06	1
91	F	1996	0.8	No MGUS	3.63	1

... plus 7864 additional observations.

Kyle R *et al.* Prevalence of monoclonal gammopathy of undetermined significance. *N Engl J Med.* 2006.

Dispenzieri A *et al.* Use of nonclonal serum immunoglobulin free light chains to predict overall survival in the general population. *Mayo Clin Proc.* 2012.

# Creating epidemiologic results tables, 2025

```
1 library(rifftable)
2
3 tribble(
4   ~type,           ~label,
5   "total",        "Participants",
6   "events/time",  "Deaths/person-years",
7   "cuminc",       "10-year risk",
8   "cumincratio", "10-year risk ratio (95% CI)",
9   "hr",           "Hazard ratio (95% CI)"
10 ) |>
11   mutate(
12     exposure = "mgus_factor",
13     event = "death",
14     time = "futime",
15     arguments = list(list(timepoint = 10)))
16 ) |>
17   rifftable(data = flchain) |>
18   rt_gt()
```

# Creating epidemiologic results tables, 2025

```
1 library(rifftable)
2
3 tribble(
4   ~type,           ~label,
5   "total",        "Participants",
6   "events/time", "Deaths/person-years",
7   "cuminc",       "10-year risk",
8   "cumincratio", "10-year risk ratio (95% CI)",
9   "hr",           "Hazard ratio (95% CI)"
10 ) |>
11   mutate(
12     exposure = "mgus_factor",
13     event = "death",
14     time = "futime",
15     arguments = list(list(timepoint = 10)))
16 ) |>
17   rifftable(data = flchain) |>
18   rt_gt()
```

MGUS status	No MGUS	MGUS
Participants	7759	115
Deaths/person-years	2153/77632	16/1292
10-year risk	0.24	0.11
10-year risk ratio (95% CI)	1 (reference)	0.45 (0.30, 1.01)
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)

# Sensitivity analyses made easy

```
1 tribble(
2   ~type,           ~label,
3   "events/total", "Deaths/participants",
4   "hr",           "Hazard ratio (95% CI)"
5 ) |>
6   mutate(
7     exposure = "mgus_factor",
8     event = "death",
9     time = "futime"
10 ) |>
11   riftable(data = flchain) |>
12   rt_gt()
```

# Sensitivity analyses made easy

```

1 tribble(
2   ~type,      ~label,
3   "events/total", "Deaths/participants",
4   "hr",        "Hazard ratio (95% CI)"
5 ) |>
6   mutate(
7     exposure = "mgus_factor",
8     event = "death",
9     time = "futime"
10 ) |>
11   rifftable(data = flchain) |>
12   rt_gt()

```

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)

# Sensitivity analyses made easy

```

1 tribble(
2   ~type,      ~label,
3   "events/total", "Deaths/participants",
4   "hr",       "Hazard ratio (95% CI)"
5 ) |>
6   mutate(
7     exposure = "mgus_factor",
8     event = "death",
9     time = "futime"
10 ) |>
11   riftable(data = flchain) |>
12   rt_gt()

```

```

1 tribble(
2   ~type,      ~label,          ~confounders,
3   "events/total", "Deaths/participants",    """",
4   "hr",        "Hazard ratio (95% CI)",      """",
5   "hr",        " Creatinine-adjusted",        "+ creatinine",
6   "hr",        " Creatinine/age-adjusted",    "+ creatinine + age",
7 ) |>
8   mutate(
9     exposure = "mgus_factor",
10    event = "death",
11    time = "futime"
12 ) |>
13   riftable(data = flchain) |>
14   rt_gt()

```

	<b>MGUS status</b>	<b>No MGUS</b>	<b>MGUS</b>
Deaths/participants	2153/7759	16/115	
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)	

# Sensitivity analyses made easy

```

1 tribble(
2   ~type,      ~label,
3   "events/total", "Deaths/participants",
4   "hr",        "Hazard ratio (95% CI)"
5 ) |>
6   mutate(
7     exposure = "mgus_factor",
8     event = "death",
9     time = "futime"
10 ) |>
11   riftable(data = flchain) |>
12   rt_gt()

```

```

1 tribble(
2   ~type,      ~label,          ~confounders,
3   "events/total", "Deaths/participants",    """",
4   "hr",        "Hazard ratio (95% CI)",      """",
5   "hr",        " Creatinine-adjusted",       "+ creatinine",
6   "hr",        " Creatinine/age-adjusted",    "+ creatinine + age",
7 ) |>
8   mutate(
9     exposure = "mgus_factor",
10    event = "death",
11    time = "futime"
12 ) |>
13   riftable(data = flchain) |>
14   rt_gt()

```

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)
Creatinine-adjusted	1 (reference)	0.52 (0.32, 0.84)
Creatinine/age-adjusted	1 (reference)	0.89 (0.54, 1.45)

# Sensitivity analyses made easy

```

1 tribble(
2   ~type,      ~label,
3   "events/total", "Deaths/participants",
4   "hr",        "Hazard ratio (95% CI)"
5 ) |>
6   mutate(
7     exposure = "mgus_factor",
8     event = "death",
9     time = "futime"
10 ) |>
11  riftable(data = flchain) |>
12  rt_gt()

```

```

1 tribble(
2   ~type,      ~label,           ~confounders,
3   "events/total", "Deaths/participants",    """",
4   "hr",        "Hazard ratio (95% CI)",      """",
5   "hr",        "  Creatinine-adjusted",      "+ creatinine",
6   "hr",        "  Creatinine/age-adjusted", "+ creatinine + age",
7 ) |>
8   mutate(
9     exposure = "mgus_factor",
10    event = "death",
11    time = "futime"
12 ) |>
13  riftable(data = flchain) |>
14  rt_gt()

```

```

1 tribble(
2   ~type,      ~label,          ~ci,
3   "events/total", "Deaths/participants", NA,
4   "",          "Hazard ratio",      NA,
5   "hr",        "  with 95% CI",  NA,
6   "hr",        "  with 80% CI",  0.8,
7   "hr",        "  with 99.5% CI", 0.995
8 ) |>
9   mutate(
10  exposure = "mgus_factor",
11  event = "death",
12  time = "futime",
13  confounders = "+ creatinine"
14 ) |>
15 riftable(data = flchain) |>
16 rt_gt()

```

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)
Creatinine-adjusted	1 (reference)	0.52 (0.32, 0.84)
Creatinine/age-adjusted	1 (reference)	0.89 (0.54, 1.45)

# Sensitivity analyses made easy

```

1 tribble(
2   ~type,      ~label,
3   "events/total", "Deaths/participants",
4   "hr",        "Hazard ratio (95% CI)"
5 ) |>
6   mutate(
7     exposure = "mgus_factor",
8     event = "death",
9     time = "futime"
10) |>
11  riftable(data = flchain) |>
12  rt_gt()

```

```

1 tribble(
2   ~type,      ~label,          ~confounders,
3   "events/total", "Deaths/participants",    """",
4   "hr",        "Hazard ratio (95% CI)",    """",
5   "hr",        "  Creatinine-adjusted",    "+ creatinine",
6   "hr",        "  Creatinine/age-adjusted", "+ creatinine + age",
7 ) |>
8   mutate(
9     exposure = "mgus_factor",
10    event = "death",
11    time = "futime"
12) |>
13  riftable(data = flchain) |>
14  rt_gt()

```

```

1 tribble(
2   ~type,      ~label,      ~ci,
3   "events/total", "Deaths/participants", NA,
4   "",           "Hazard ratio",  NA,
5   "hr",        "  with 95% CI", NA,
6   "hr",        "  with 80% CI",  0.8,
7   "hr",        "  with 99.5% CI", 0.995
8 ) |>
9   mutate(
10  exposure = "mgus_factor",
11  event = "death",
12  time = "futime",
13  confounders = "+ creatinine"
14) |>
15  riftable(data = flchain) |>
16  rt_gt()

```

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)
Creatinine-adjusted	1 (reference)	0.52 (0.32, 0.84)
Creatinine/age-adjusted	1 (reference)	0.89 (0.54, 1.45)

MGUS status	No MGUS	MGUS
Deaths/participants	2153/7759	16/115
Hazard ratio		
with 95% CI	1 (reference)	0.52 (0.32, 0.84)
with 80% CI	1 (reference)	0.52 (0.37, 0.71)
with 99.5% CI	1 (reference)	0.52 (0.26, 1.04)

# Stratified and regression analyses at once

```
1 tribble(
2   ~type,           ~label,           ~stratum,
3   "events/total", "Deaths/participants",  "",
4   "hr",            "Hazard ratio (95% CI)", c("F", "M"),
5   "",              "*Female*",          "F",
6   "events/total", " Deaths/participants", "F",
7   "hr",            " Hazard ratio (95% CI)", "F",
8   "",              "*Male*",            "M",
9   "events/total", " Deaths/participants", "M",
10  "hr",            " Hazard ratio (95% CI)", "M",
11 ) |>
12 mutate(
13   exposure = "mgus_factor",
14   event = "death",
15   time = "futime",
16   effect_modifier = "sex"
17 ) |>
18 riftable(data = flchain) |>
19 rt_gt()
```

# Stratified and regression analyses at once

```

1 tribble(
2   ~type,      ~label,      ~stratum,
3   "events/total", "Deaths/participants",    "",
4   "hr",        "Hazard ratio (95% CI)",    c("F", "M"),
5   "",          "*Female*",           "F",
6   "events/total", " Deaths/participants",  "F",
7   "hr",        " Hazard ratio (95% CI)",  "F",
8   "",          "*Male*",            "M",
9   "events/total", " Deaths/participants", "M",
10  "hr",        " Hazard ratio (95% CI)", "M",
11 ) |>
12 mutate(
13   exposure = "mgus_factor",
14   event = "death",
15   time = "futime",
16   effect_modifier = "sex"
17 ) |>
18 rifftable(data = flchain) |>
19 rt_gt()

```

	<b>MGUS status</b>	<b>No MGUS</b>	<b>MGUS</b>
Deaths/participants	2153/7759	16/115	
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)	
<i>Female</i>			
Deaths/participants	1154/4282	11/68	
Hazard ratio (95% CI)	1 (reference)	0.55 (0.30, 0.99)	
<i>Male</i>			
Deaths/participants	999/3477	5/47	
Hazard ratio (95% CI)	1 (reference)	0.32 (0.13, 0.76)	

# Different exposure? No problem

```

1 tribble(
2   ~type,      ~label,          ~stratum,
3   "events/total", "Deaths/participants",  "", 
4   "hr",       "Hazard ratio (95% CI)",  c("F", "M"),
5   "",         "*Female*",        "F",
6   "events/total", "Deaths/participants",  "F",
7   "hr",       "Hazard ratio (95% CI)",  "F",
8   "",         "*Male*",          "M",
9   "events/total", "Deaths/participants",  "M",
10  "hr",      "Hazard ratio (95% CI)",  "M",
11 ) |>
12 mutate(
13   exposure = "mgus_factor",
14   event = "death",
15   time = "futime",
16   effect_modifier = "sex"
17 ) |>
18 riftable(data = flchain) |>
19 rt_gt()

```

```

1 tribble(
2   ~type,      ~label,          ~stratum,
3   "events/total", "Deaths/participants",  "", 
4   "hr",       "Hazard ratio (95% CI)",  c("F", "M"),
5   "",         "*Female*",        "F",
6   "events/total", "Deaths/participants",  "F",
7   "hr",       "Hazard ratio (95% CI)",  "F",
8   "",         "*Male*",          "F",
9   "events/total", "Deaths/participants",  "M",
10  "hr",      "Hazard ratio (95% CI)",  "M",
11 ) |>
12 mutate(
13   exposure = "age_group",
14   event = "death",
15   time = "futime",
16   effect_modifier = "sex"
17 ) |>
18 riftable(data = flchain) |>
19 rt_gt()

```

	<b>MGUS status</b>	<b>No MGUS</b>	<b>MGUS</b>
Deaths/participants	2153/7759	16/115	
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)	
<i>Female</i>			
Deaths/participants	1154/4282	11/68	
Hazard ratio (95% CI)	1 (reference)	0.55 (0.30, 0.99)	
<i>Male</i>			
Deaths/participants	999/3477	5/47	
Hazard ratio (95% CI)	1 (reference)	0.32 (0.13, 0.76)	

	<b>Age at inclusion, years</b>	<b>50-&lt;65</b>	<b>65-80</b>	<b>80+</b>
Deaths/participants	443/4373	1088/2736	638/765	
Hazard ratio (95% CI)	1 (reference)	4.4 (4.0, 5.0)	17 (15, 19)	
<i>Female</i>				
Deaths/participants	193/2272	528/1538	444/540	
Hazard ratio (95% CI)	1 (reference)	4.5 (3.8, 5.3)	20 (17, 24)	
<i>Male</i>				
Deaths/participants	250/2101	560/1198	194/225	
Hazard ratio (95% CI)	1 (reference)	4.6 (4.0, 5.3)	16 (13, 19)	

# Different exposure? No problem

```

1 tribble(
2   ~type,      ~label,          ~stratum,
3   "events/total", "Deaths/participants",  "", 
4   "hr",       "Hazard ratio (95% CI)",  c("F", "M"),
5   "",         "*Female*",        "F",
6   "events/total", "Deaths/participants",  "F",
7   "hr",       "Hazard ratio (95% CI)",  "F",
8   "",         "*Male*",          "M",
9   "events/total", "Deaths/participants",  "M",
10  "hr",      "Hazard ratio (95% CI)",  "M",
11 ) |>
12 mutate(
13   exposure = "mgus_factor",
14   event = "death",
15   time = "futime",
16   effect_modifier = "sex"
17 ) |>
18 riftable(data = flchain) |>
19 rt_gt()

```

```

1 tribble(
2   ~type,      ~label,          ~stratum,
3   "events/total", "Deaths/participants",  "", 
4   "hr",       "Hazard ratio (95% CI)",  c("F", "M"),
5   "",         "*Female*",        "F",
6   "events/total", "Deaths/participants",  "F",
7   "hr",       "Hazard ratio (95% CI)",  "F",
8   "",         "*Male*",          "F",
9   "events/total", "Deaths/participants",  "M",
10  "hr",      "Hazard ratio (95% CI)",  "M",
11 ) |>
12 mutate(
13   exposure = "age_group",
14   event = "death",
15   time = "futime",
16   effect_modifier = "sex"
17 ) |>
18 riftable(data = flchain) |>
19 rt_gt()

```

	<b>MGUS status</b>	<b>No MGUS</b>	<b>MGUS</b>
Deaths/participants	2153/7759	16/115	
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)	
<i>Female</i>			
Deaths/participants	1154/4282	11/68	
Hazard ratio (95% CI)	1 (reference)	0.55 (0.30, 0.99)	
<i>Male</i>			
Deaths/participants	999/3477	5/47	
Hazard ratio (95% CI)	1 (reference)	0.32 (0.13, 0.76)	

	<b>Age at inclusion, years</b>	<b>50-&lt;65</b>	<b>65-80</b>	<b>80+</b>
Deaths/participants	443/4373	1088/2736	638/765	
Hazard ratio (95% CI)	1 (reference)	4.4 (4.0, 5.0)	17 (15, 19)	
<i>Female</i>				
Deaths/participants	193/2272	528/1538	444/540	
Hazard ratio (95% CI)	1 (reference)	4.5 (3.8, 5.3)	20 (17, 24)	
<i>Male</i>				
Deaths/participants	250/2101	560/1198	194/225	
Hazard ratio (95% CI)	1 (reference)	4.6 (4.0, 5.3)	16 (13, 19)	

# Different exposure? No problem

```

1 tribble(
2   ~type,      ~label,           ~stratum,
3   "events/total", "Deaths/participants",  "", 
4   "hr",        "Hazard ratio (95% CI)",  c("F", "M"),
5   "",          "*Female*",           "F",
6   "events/total", "Deaths/participants", "F",
7   "hr",        "Hazard ratio (95% CI)", "F",
8   "",          "*Male*",            "M",
9   "events/total", "Deaths/participants", "M",
10  "hr",       "Hazard ratio (95% CI)", "M",
11 ) |>
12   mutate(
13     exposure = "mgus_factor",
14     event = "death",
15     time = "futime",
16     effect_modifier = "sex"
17   ) |>
18   rifftable(data = flchain) |>
19   rt_gt()

```

```

1 tribble(
2   ~type,      ~label,
3   "events/total", "Deaths/participant",
4   "hr",        "Hazard ratio (95% CI)",
5   "",          "*Female*",           "F",
6   "events/total", "Deaths/participant",
7   "hr",        "Hazard ratio (95% CI)",
8   "",          "*Male*",            "M",
9   "events/total", "Deaths/participant",
10  "hr",       "Hazard ratio (95% CI)",
11 ) |>
12   mutate(
13     exposure = "age_group",
14     event = "death",
15     time = "futime",
16     effect_modifier = "sex"
17   ) |>
18   rifftable(data = flchain) |>
19   rt_gt()

```



	<b>MGUS status</b>	<b>No MGUS</b>	<b>MGUS</b>
Deaths/participants	2153/7759	16/115	
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)	
<i>Female</i>			
Deaths/participants	1154/4282	11/68	
Hazard ratio (95% CI)	1 (reference)	0.55 (0.30, 0.99)	
<i>Male</i>			
Deaths/participants	999/3477	5/47	
Hazard ratio (95% CI)	1 (reference)	0.32 (0.13, 0.76)	

	<b>Age at inclusion, years</b>	<b>50-&lt;65</b>	<b>65-80</b>	<b>80+</b>
Deaths/participants	443/4373	1088/2736	638/765	
Hazard ratio (95% CI)	1 (reference)	4.4 (4.0, 5.0)	17 (15, 19)	
<i>Female</i>				
Deaths/participants	193/2272	528/1538	444/540	
Hazard ratio (95% CI)	1 (reference)	4.5 (3.8, 5.3)	20 (17, 24)	
<i>Male</i>				
Deaths/participants	250/2101	560/1198	194/225	
Hazard ratio (95% CI)	1 (reference)	4.6 (4.0, 5.3)	16 (13, 19)	

# riftable's data model

## The design

```

1 design <- tribble(
2   ~type,           ~label,
3   "total",        "Participants",
4   "events/time", "Deaths/person-years",
5   "cuminc",       "10-year risk",
6   "cumincratio", "10-year risk ratio (95% CI)",
7   "hr",           "Hazard ratio (95% CI)"
8 ) |>
9   mutate(
10   exposure = "mgus_factor",
11   event = "death",
12   time = "futime",
13   arguments = list(list(timepoint = 10))
14 )

```

## The data

```

# A tibble: 7,874 × 7
  age sex sample_yr creatinine mgus_factor futime death
  <dbl> <fct>    <dbl>      <dbl> <fct>      <dbl> <dbl>
1 97 F     1997      1.7 No MGUS  0.233     1
2 92 F     2000      0.9 No MGUS  3.51      1
3 94 F     1997      1.4 No MGUS  0.189     1
4 92 F     1996      1   No MGUS  0.315     1
5 93 F     1996      1.1 No MGUS  2.84      1
6 90 F     1997      1   No MGUS  3.71      1
7 90 F     1996      0.8 No MGUS  7.81      1
8 90 F     1999      1.2 No MGUS  1.02      1
9 93 F     1996      1.2 No MGUS  9.06      1
10 91 F    1996      0.8 No MGUS  3.63      1
# i 7,864 more rows

```

# riftable's data model

## The design

```
# A tibble: 5 × 6
  type      label            exposure event time  arguments
  <chr>     <chr>           <chr>    <chr> <chr> <list>
1 total     Participants    mgus_fa... death futi... <named list>
2 events/time Deaths/person-years mgus_fa... death futi... <named list>
3 cuminc    10-year risk    mgus_fa... death futi... <named list>
4 cumincratio 10-year risk ratio (9... mgus_fa... death futi... <named list>
5 hr        Hazard ratio (95% CI) mgus_fa... death futi... <named list>
```

## The data

```
# A tibble: 7,874 × 7
  age sex   sample_yr creatinine mgus_factor futime death
  <dbl> <fct>    <dbl>       <dbl> <fct>      <dbl> <dbl>
1 97   F        1997       1.7  No MGUS    0.233   1
2 92   F        2000       0.9  No MGUS    3.51    1
3 94   F        1997       1.4  No MGUS    0.189   1
4 92   F        1996       1    No MGUS    0.315   1
5 93   F        1996       1.1  No MGUS    2.84    1
6 90   F        1997       1    No MGUS    3.71    1
7 90   F        1996       0.8  No MGUS    7.81    1
8 90   F        1999       1.2  No MGUS    1.02    1
9 93   F        1996       1.2  No MGUS    9.06    1
10 91  F        1996       0.8  No MGUS    3.63    1
# i 7,864 more rows
```

# rifftable's data model

## The design

```
# A tibble: 5 × 6
  type      label            exposure event time  arguments
  <chr>     <chr>          <chr>    <chr> <chr> <list>
1 total     Participants    mgus_fa... death futi... <named list>
2 events/time Deaths/person-years mgus_fa... death futi... <named list>
3 cuminc    10-year risk    mgus_fa... death futi... <named list>
4 cumincrato 10-year risk ratio (9... mgus_fa... death futi... <named list>
5 hr        Hazard ratio (95% CI) mgus_fa... death futi... <named list>
```

## The data

```
# A tibble: 7,874 × 7
  age sex   sample_yr creatinine mgus_factor futime death
  <dbl> <fct>    <dbl>       <dbl> <fct>      <dbl> <dbl>
1 97 F      1997       1.7 No MGUS  0.233  1
2 92 F      2000       0.9 No MGUS  3.51   1
3 94 F      1997       1.4 No MGUS  0.189   1
4 92 F      1996       1  No MGUS  0.315   1
5 93 F      1996       1.1 No MGUS  2.84   1
6 90 F      1997       1  No MGUS  3.71   1
7 90 F      1996       0.8 No MGUS  7.81   1
8 90 F      1999       1.2 No MGUS  1.02   1
9 93 F      1996       1.2 No MGUS  9.06   1
10 91 F     1996       0.8 No MGUS  3.63   1
# i 7,864 more rows
```

## The table produced by `rifftable()`

```
1 rifftable(
2   design = design,
3   data = flchain
4 )
```

```
# A tibble: 5 × 3
`MGUS status` `No MGUS` MGUS
<chr>          <chr>    <chr>
1 Participants  7759    115
2 Deaths/person-years 2153/77632 16/1292
3 10-year risk  0.24    0.11
4 10-year risk ratio (95% CI) 1 (reference) 0.45 (0.30, 1.01)
5 Hazard ratio (95% CI) 1 (reference) 0.44 (0.27, 0.73)
```

# rifftable's data model

## The design

```
# A tibble: 5 × 6
  type      label            exposure event time  arguments
  <chr>     <chr>           <chr>    <chr> <chr> <list>
1 total     Participants    mgus_fa... death futi... <named list>
2 events/time Deaths/person-years mgus_fa... death futi... <named list>
3 cuminc    10-year risk    mgus_fa... death futi... <named list>
4 cumincrato 10-year risk ratio (9... mgus_fa... death futi... <named list>
5 hr        Hazard ratio (95% CI) mgus_fa... death futi... <named list>
```

## The data

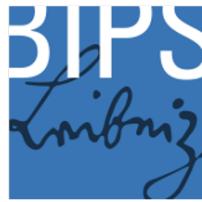
```
# A tibble: 7,874 × 7
  age sex   sample_yr creatinine mgus_factor futime death
  <dbl> <fct>    <dbl>       <dbl> <fct>      <dbl> <dbl>
1 97 F      1997       1.7 No MGUS  0.233  1
2 92 F      2000       0.9 No MGUS  3.51   1
3 94 F      1997       1.4 No MGUS  0.189   1
4 92 F      1996       1  No MGUS  0.315   1
5 93 F      1996       1.1 No MGUS  2.84   1
6 90 F      1997       1  No MGUS  3.71   1
7 90 F      1996       0.8 No MGUS  7.81   1
8 90 F      1999       1.2 No MGUS  1.02   1
9 93 F      1996       1.2 No MGUS  9.06   1
10 91 F     1996       0.8 No MGUS  3.63   1
# i 7,864 more rows
```

## The table produced by `rifftable()`

```
1 rifftable(
2   design = design,
3   data = flchain
4 ) |>
5   rt_gt()
```

MGUS status	No MGUS	MGUS
Participants	7759	115
Deaths/person-years	2153/77632	16/1292
10-year risk	0.24	0.11
10-year risk ratio (95% CI)	1 (reference)	0.45 (0.30, 1.01)
Hazard ratio (95% CI)	1 (reference)	0.44 (0.27, 0.73)

# Great software for analyses and tables exists

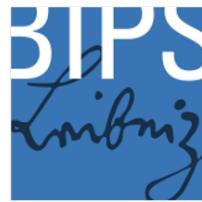


- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

# Great software for analyses and tables exists

## Why rfttable

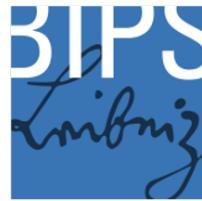
- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality



# Great software for analyses and tables exists

- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

## Why rifttable

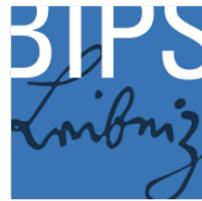


- Integrates into real data analysis

# Great software for analyses and tables exists

- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

## Why rifttable

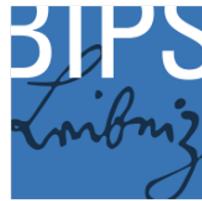


- Integrates into real data analysis
- Shows descriptive and inferential statistics side-by-side

# Great software for analyses and tables exists

- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

## Why rifttable



- Integrates into real data analysis
- Shows descriptive and inferential statistics side-by-side
- Makes stratified analyses easy  
(and table 2 fallacies hard)

# Great software for analyses and tables exists

- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

## Why rifttable



- Integrates into real data analysis
- Shows descriptive and inferential statistics side-by-side
- Makes stratified analyses easy  
(and table 2 fallacies hard)
- Rounds estimates well

# Great software for analyses and tables exists

- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

## Why rifttable



- Integrates into real data analysis
- Shows descriptive and inferential statistics side-by-side
- Makes stratified analyses easy  
(and table 2 fallacies hard)
- Rounds estimates well
- Facilitates sensitivity analyses

# Great software for analyses and tables exists

- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

## Why rifttable



- Integrates into real data analysis
- Shows descriptive and inferential statistics side-by-side
- Makes stratified analyses easy  
(and table 2 fallacies hard)
- Rounds estimates well
- Facilitates sensitivity analyses
- Extends with custom estimators

# Great software for analyses and tables exists

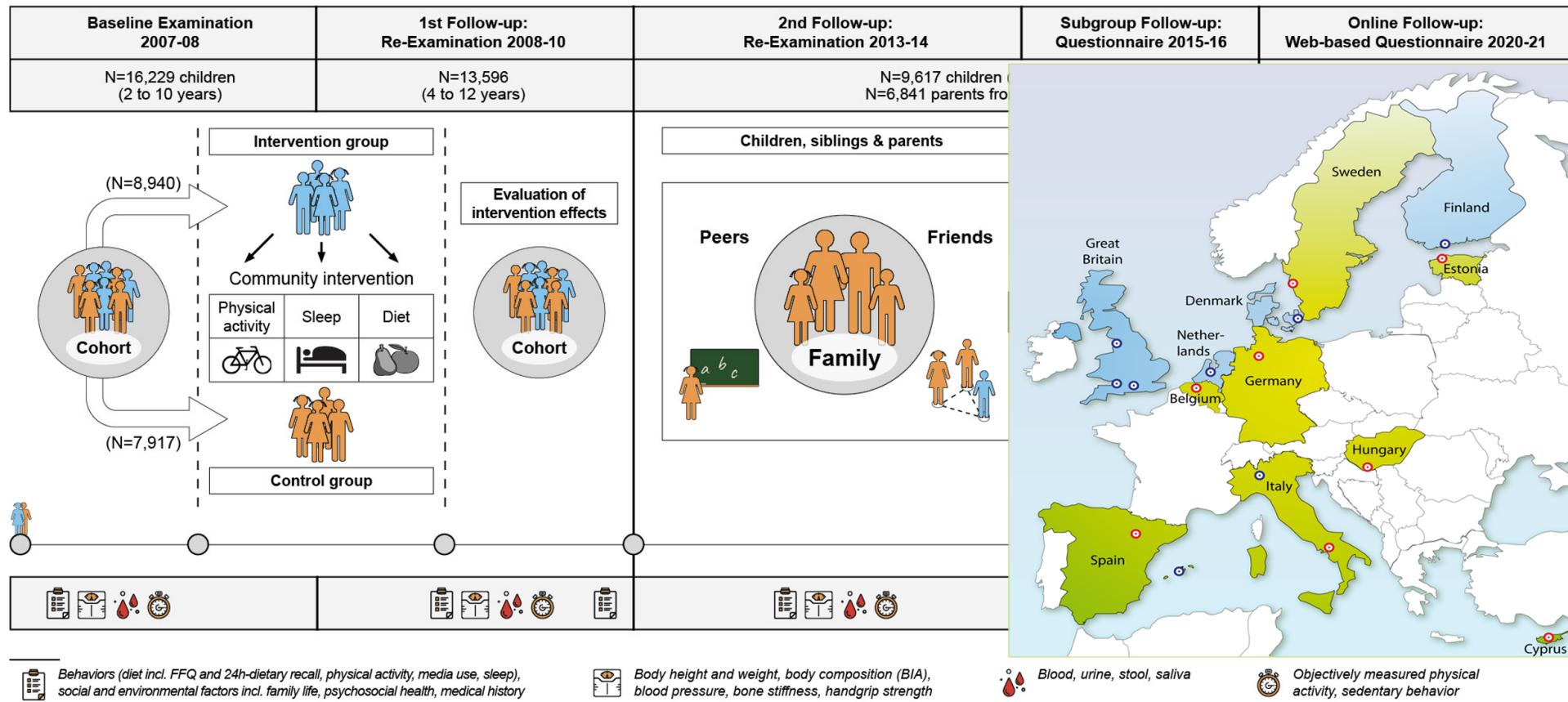
- Excel calculator **Episheet**  
by Kenneth Rothman
- R package **gtsummary**  
by Daniel Sjoberg
- Many others, of varying scope and quality

## Why rifttable



- Integrates into real data analysis
- Shows descriptive and inferential statistics side-by-side
- Makes stratified analyses easy  
(and table 2 fallacies hard)
- Rounds estimates well
- Facilitates sensitivity analyses
- Extends with custom estimators
- Has a website with copy-paste examples

# The IDEFICS/I.Family cohort



Ahrens W et al. The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes.* 2011;35:S3-S15.

Ahrens W et al. Cohort profile: The transition from childhood to adolescence in European children – How I.Family extends the IDEFICS cohort. *Int J Epidemiol.* 2017;46:1394-1395.

# The IDEFICS/I.Family cohort: Table 1

```
idefics |>
  filter(survey == 0) |> # Baseline only
  table1_design(
    age, country, migrant,
    bmi_score_cole_12, bmi_m, bmi_f, one_parent,
    birth_w,
    energy,
    by = gender
  ) |>
  mutate(
    digits = if_else(
      outcome %in% c("energy", "birth_w"),
      true = 0,
      false = NA
    )
  ) |>
  rifftable(diff_digits = 1) |>
  rt_gt() |>
  gt::tab_footnote(footnote = "Statistics shown in
```

Ahrens W *et al.* The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes.* 2011;35:S3-S15.

Ahrens W *et al.* Cohort profile: The transition from childhood to adolescence in European children – How I.Family extends the IDEFICS cohort. *Int J Epidemiol.* 2017;46:1394-1395.

# The IDEFICS/I.Family cohort: Table 1

```
idefics |>
  filter(survey == 0) |> # Baseline only
  table1_design(
    age, country, migrant,
    bmi_score_cole_12, bmi_m, bmi_f, one_parent,
    birth_w,
    energy,
    by = gender
  ) |>
  mutate(
    digits = if_else(
      outcome %in% c("energy", "birth_w"),
      true = 0,
      false = NA
    )
  ) |>
  rifftable(diff_digits = 1) |>
  rt_gt() |>
  gt::tab_footnote(footnote = "Statistics shown are count (percent) or median (interquartile range).")
```

Characteristic	Boys	Girls
N	8251	7979
Age at enrollment, years	6.2 (4.4, 7.5)	6.3 (4.5, 7.6)
Country		
Italy	1167 (14%)	1083 (14%)
Estonia	849 (10%)	870 (11%)
Cyprus	1222 (15%)	1161 (15%)
Belgium	980 (12%)	945 (12%)
Sweden	927 (11%)	882 (11%)
Germany	1053 (13%)	1013 (13%)
Hungary	1284 (16%)	1284 (16%)
Spain	769 (9%)	741 (9%)
Parent(s) with migration history	1251 (16%)	1253 (16%)
Unknown	378	353
z-score of BMI by Cole (2012)	0.2 (-0.5, 1.0)	0.3 (-0.4, 1.1)
Maternal BMI, kg/m <sup>2</sup>	23.0 (20.9, 25.9)	22.9 (20.9, 25.8)
Unknown	536	519
Paternal BMI, kg/m <sup>2</sup>	26.0 (24.1, 28.4)	26.0 (24.0, 28.4)
Unknown	1294	1291
Single-parent family	1148 (15%)	1061 (14%)
Unknown	633	594
Birth weight, g	3415 (3060, 3760)	3280 (2970, 3600)
Unknown	490	467
Usual energy intake, kcal/day	1563 (1388, 1728)	1431 (1304, 1546)
Unknown	3101	2935

Statistics shown are count (percent) or median (interquartile range).

Ahrens W *et al.* The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes.* 2011;35:S3-S15.

Ahrens W *et al.* Cohort profile: The transition from childhood to adolescence in European children – How I.Family extends the IDEFICS cohort. *Int J Epidemiol.* 2017;46:1394-1395.

# riftable, v0.7.1

https://stopsack.github.io/riftable/

80% Search

**riftable** 0.7.1 Get started FAQ Example Tables More ▾

## Automated, Reproducible Generation of Results Tables: Bridging the Rift Between Epidemiologists and Their Data

Tables are the key format in which epidemiologists present their results. Many results tables in applied studies merely show point estimates and confidence intervals, or even p-values, from regression models: a “growing rift between epidemiologists and their data” (Rothman 2017). “Actual,” descriptive data, such as counts stratified by exposure and a main confounder or effect modifier, are often lacking.

**riftable** creates presentation-ready results tables for epidemiologists in an automated, reproducible fashion. The user provides the final analytical dataset and specifies the design of the table, with rows and/or columns defined by exposure(s), effect modifier(s), and estimands as desired, allowing to show descriptors and inferential estimates in one table – bridging the rift between epidemiologists and their data, one table at a time.

### Installation

The riftable package can be installed from CRAN:

```
install.packages("riftable")
```

Development versions can be installed from [GitHub](#) using:

```
remotes::install_github("stopsack/riftable")
```

The latter installation procedure requires the remotes package, obtainable via `install.packages("remotes")`.

### Example

```
library(riftable)

example_design <- tibble::tribble(
  ~label, ~type, ~stratum,
  "Overall", "", "",
  "- Deaths/N", "outcomes/total", c("Low", "High"),
  ...)
```

<https://stopsack.github.io/riftable>

# riftable, v0.7.1



https://stopsack.github.io/riftable/

riftable 0.7.1 Get started FAQ Example Tables More ▾

## Automated, Reproducible Generation of Results Tables: Bridging the Rift Between Epidemiologists and Their Data

Tables are the key format in which epidemiologists present their results. Many results tables in applied studies merely show point estimates and confidence intervals, or even p-values, from regression models: a “growing rift between epidemiologists and their data” (Rothman 2017). “Actual,” descriptive data, such as counts stratified by exposure and a main confounder or effect modifier, are often lacking.

**riftable** creates presentation-ready results tables for epidemiologists in an automated, reproducible fashion. The user provides the final analytical dataset and specifies the design of the table, with rows and/or columns defined by exposure(s), effect modifier(s), and estimands as desired, allowing to show descriptors and inferential estimates in one table – bridging the rift between epidemiologists and their data, one table at a time.

### Installation

The riftable package can be installed from CRAN:

```
install.packages("riftable")
```

Development versions can be installed from GitHub using:

```
remotes::install_github("stopsack/riftable")
```

The latter installation procedure requires the remotes package, obtainable via `install.packages("remotes")`.

### Example

```
library(riftable)

example_design <- tibble::tribble(
  ~label, ~type, ~stratum,
  "Overall", "", ""),
  "- Deaths/N", "outcomes/total", c("Low", "High"),
  ...)
```

Also...

- Quantiles and quantile regression
- Ratios of continuous outcomes
- Competing events
- Clustering
- Trends (slopes)
- Stratified and joint models
- Inverse probability-weighted estimates
- Regression-based risk ratios and risk differences for binary outcomes

<https://stopsack.github.io/riftable>

# riftable, v0.7.1



https://stopsack.github.io/riftable/

riftable 0.7.1 Get started FAQ Example Tables More ▾

## Automated, Reproducible Generation of Results Tables: Bridging the Rift Between Epidemiologists and Their Data

Tables are the key format in which epidemiologists present their results. Many results tables in applied studies merely show point estimates and confidence intervals, or even p-values, from regression models: a “growing rift between epidemiologists and their data” (Rothman 2017). “Actual,” descriptive data, such as counts stratified by exposure and a main confounder or effect modifier, are often lacking.

**riftable** creates presentation-ready results tables for epidemiologists in an automated, reproducible fashion. The user provides the final analytical dataset and specifies the design of the table, with rows and/or columns defined by exposure(s), effect modifier(s), and estimands as desired, allowing to show descriptors and inferential estimates in one table – bridging the rift between epidemiologists and their data, one table at a time.

### Installation

The riftable package can be installed from CRAN:

```
install.packages("riftable")
```

Development versions can be installed from GitHub using:

```
remotes::install_github("stopsack/riftable")
```

The latter installation procedure requires the remotes package, obtainable via `install.packages("remotes")`.

### Example

```
library(riftable)

example_design <- tibble::tribble(
  ~label, ~type, ~stratum,
  "Overall", "", ""),
  "- Deaths/N", "outcomes/total", c("Low", "High"),
  ...)
```

Also...

- Quantiles and quantile regression
- Ratios of continuous outcomes
- Competing events
- Clustering
- Trends (slopes)
- Stratified and joint models
- Inverse probability-weighted estimates
- Regression-based risk ratios and risk differences for binary outcomes

<https://stopsack.github.io/riftable>

# Regression models for risk ratios and risk differences

In cohort studies [...], there is little reason to consider an odds ratio. While odds ratios are not the primary measure of interest, they are frequently reported because of the statistical model used by the investigators to analyze their data.

# Regression models for risk ratios and risk differences

In cohort studies [...], there is little reason to consider an odds ratio. While odds ratios are not the primary measure of interest, they are frequently reported because of the statistical model used by the investigators to analyze their data.

Exposure	Unexposed	Exposed
Total	1000	1000
Cases/non-cases	20/980	40/960
Risk	2%	4%
Risk ratio	1 (reference)	2.00 (1.18, 3.4)
Odds ratio	1 (reference)	2.04 (1.20, 3.6)

# Regression models for risk ratios and risk differences

In cohort studies [...], there is little reason to consider an odds ratio. While odds ratios are not the primary measure of interest, they are frequently reported because of the statistical model used by the investigators to analyze their data.

Exposure	Unexposed	Exposed
Total	1000	1000
Cases/non-cases	20/980	40/960
Risk	2%	4%
Risk ratio	1 (reference)	2.00 (1.18, 3.4)
Odds ratio	1 (reference)	2.04 (1.20, 3.6)

Exposure	Unexposed	Exposed
Total	1000	1000
Cases/non-cases	960/40	980/20
Risk	96%	98%
Risk ratio	1 (reference)	1.02 (1.01, 1.04)
Odds ratio	1 (reference)	2.04 (1.20, 3.6)

# Why are we still reporting odds ratios in cross-sectional and cohort studies?

## 1. Log-binomial models do not converge

```
1 data(breastcancer, package = "risks")
2 glm(
3   formula = death ~ stage + receptor,
4   data = breastcancer,
5   family = binomial(link = "log")
6 )
```

Error: no valid set of coefficients has been found: please supply starting values

# Why are we still reporting odds ratios in cross-sectional and cohort studies?

## 1. Log-binomial models do not converge

```
1 data(breastcancer, package = "risks")
2 glm(
3   formula = death ~ stage + receptor,
4   data = breastcancer,
5   family = binomial(link = "log")
6 )
```

Error: no valid set of coefficients has been found: please supply starting values

# Why are we still reporting odds ratios in cross-sectional and cohort studies?

## 1. Log-binomial models do not converge

```
1 data(breastcancer, package = "risks")
2 glm(
3   formula = death ~ stage + receptor,
4   data = breastcancer,
5   family = binomial(link = "log")
6 )
```

Error: no valid set of coefficients has been found: please supply starting values

## 2. “Poisson” models with robust variance often do not converge either

# Why are we still reporting odds ratios in cross-sectional and cohort studies?

## 1. Log-binomial models do not converge

```
1 data(breastcancer, package = "risks")
2 glm(
3   formula = death ~ stage + receptor,
4   data = breastcancer,
5   family = binomial(link = "log")
6 )
```

Error: no valid set of coefficients has been found: please supply starting values

## 2. “Poisson” models with robust variance often do not converge either

# Why are we still reporting odds ratios in cross-sectional and cohort studies?

## 1. Log-binomial models do not converge

```
1 data(breastcancer, package = "risks")
2 glm(
3   formula = death ~ stage + receptor,
4   data = breastcancer,
5   family = binomial(link = "log")
6 )
```

Error: no valid set of coefficients has been found: please supply starting values

## 2. “Poisson” models with robust variance often do not converge either

## 3. Marginal standardization/g-computation/g-formula with time-fixed exposure works whenever a logistic model can be fit

# Why are we still reporting odds ratios in cross-sectional and cohort studies?

## 1. Log-binomial models do not converge

```
1 data(breastcancer, package = "risks")
2 glm(
3   formula = death ~ stage + receptor,
4   data = breastcancer,
5   family = binomial(link = "log")
6 )
```

Error: no valid set of coefficients has been found: please supply starting values

## 2. “Poisson” models with robust variance often do not converge either

## 3. Marginal standardization/g-computation/g-formula with time-fixed exposure works whenever a logistic model can be fit

# Why are we still reporting odds ratios in cross-sectional and cohort studies?

## 1. Log-binomial models do not converge

```
1 data(breastcancer, package = "risks")
2 glm(
3   formula = death ~ stage + receptor,
4   data = breastcancer,
5   family = binomial(link = "log")
6 )
```

Error: no valid set of coefficients has been found: please supply starting values

## 2. “Poisson” models with robust variance often do not converge either

## 3. Marginal standardization/g-computation/g-formula with time-fixed exposure works whenever a logistic model can be fit

## 4. And there is Miettinen’s case-duplication approach

# We no longer need to report odds ratios in cross-sectional and cohort studies

```
1 library(risks)
2 summary(
3   riskratio(
4     formula = death ~ stage + receptor,
5     data = breastcancer
6   )
7 )
```

# We no longer need to report odds ratios in cross-sectional and cohort studies

```
1 library(risks)
2 summary(
3   riskratio(
4     formula = death ~ stage + receptor,
5     data = breastcancer
6   )
7 )
```

Risk ratio model, fitted via marginal standardization of a logistic model with delta method (margstd\_delta).

Call:

```
stats::glm(formula = death ~ stage + receptor, family = binomial(link =
"logit"),
           start = "(no starting values)")
```

Coefficients: (3 not defined because of singularities)

	Estimate	Std. Error	z value	Pr(> z )
stageStage I	0.0000	0.0000	NaN	NaN
stageStage II	0.8989	0.3875	2.320	0.0203 *
stageStage III	1.8087	0.3783	4.781	1.75e-06 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 228.15 on 191 degrees of freedom

Residual deviance: 185.88 on 188 degrees of freedom

AIC: 193.88

Number of Fisher Scoring iterations: 4

Confidence intervals for coefficients: (delta method)

	2.5 %	97.5 %
stageStage I	0.0000000	0.000000

# Estimating and reporting adjusted risk ratios and risk differences made easy

```
1 tribble(
2   ~type,           ~label,
3   "outcomes/total", "Deaths/total",
4   "risk",          "Risk",
5   "rr",            "Risk ratio (95% CI)",
6   "rd",            "Risk difference (95% CI)",
7 ) |>
8   mutate(
9     outcome = "death",
10    exposure = "stage",
11    confounders = "+ receptor"
12  ) |>
13  rifftable(data = breastcancer) |>
14  rt_gt() |>
15  gt::tab_footnote(footnote = "Adjusted for hormone receptor status.")
```

# Estimating and reporting adjusted risk ratios and risk differences made easy

```
1 tribble(
2   ~type,           ~label,
3   "outcomes/total", "Deaths/total",
4   "risk",          "Risk",
5   "rr",            "Risk ratio (95% CI)",
6   "rd",            "Risk difference (95% CI)",
7 ) |>
8   mutate(
9     outcome = "death",
10    exposure = "stage",
11    confounders = "+ receptor"
12  ) |>
13  rifftable(data = breastcancer) |>
14  rt_gt() |>
15  gt::tab_footnote(footnote = "Adjusted for hormone receptor status.")
```

Stage	Stage I	Stage II	Stage III
Deaths/total	7/67	26/96	21/29
Risk	0.10	0.27	0.72
Risk ratio (95% CI)	1 (reference)	2.46 (1.15, 5.3)	6.1 (2.91, 13)
Risk difference (95% CI)	0 (reference)	0.16 (0.05, 0.28)	0.57 (0.38, 0.77)
Adjusted for hormone receptor status.			

# Compare methods if you like. In practice, simply use *g*-computation for RRs and RDs

```

1 tribble(
2   ~type,           ~label,           ~arguments,
3   "outcomes/total", "Deaths/total", NA,
4   "risk",          "Risk",          NA,
5   "",              "Risk ratio (95% CI)", NA,
6   "rr",            "  *g*-computation, delta method CI", list(approach = "margstd_delta"),
7   "rr",            "  *g*-computation, bootstrap CI", list(approach = "margstd_boot"),
8   "rr",            " Poisson model, robust CI", list(approach = "robpoisson"),
9   "rr",            " Case-duplication approach", list(approach = "duplicate"),
10  "rd",            "Risk difference (95% CI)", NA,
11 ) |>
12 mutate(
13   outcome = "death",
14   exposure = "stage",
15   confounders = "+ receptor"
16 ) |>
17 riftable(data = breastcancer) |>
18 rt_gt() |>
19 gt::tab_footnote(footnote = "Adjusted for hormone receptor status.")

```

Stage	Stage I	Stage II	Stage III
Deaths/total	7/67	26/96	21/29
Risk	0.10	0.27	0.72
Risk ratio (95% CI)			
g-computation, delta method CI	1 (reference)	2.46 (1.15, 5.3)	6.1 (2.91, 13)
g-computation, bootstrap CI	1 (reference)	2.46 (1.09, 5.9)	6.1 (3.0, 15)
Poisson model, robust CI	1 (reference)	2.52 (1.17, 5.4)	5.9 (2.78, 13)
Case-duplication approach	1 (reference)	2.52 (1.16, 5.5)	5.9 (2.79, 13)
Risk difference (95% CI)	0 (reference)	0.16 (0.05, 0.28)	0.57 (0.38, 0.77)
Adjusted for hormone receptor status.			

# risks, v0.4.3



A screenshot of a web browser showing the risks package homepage at <https://stopsack.github.io/risks/>. The page title is "risks 0.4.3". Below the title are navigation links: "Get started", "Reference", "Articles", and "Changelog".

## risks: Estimating risk ratios and risk differences using regression



### Installation

The **risks** package can be installed from CRAN:

```
install.packages("risks")
```

Development versions can be installed from [GitHub](#) using:

```
remotes::install_github("stopsack/risks")
```

### Summary

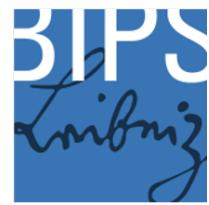
The **risks** package fits regression models for risk ratios (RR) and risk differences (RD). The package refers to "risk," but "prevalence" can be substituted throughout.

What is the association between an exposure (smoker/nonsmoker, age in years, or underweight/lean/overweight/obese) and the risk of a binary outcome (dead/alive, disease/healthy), perhaps adjusting for confounders (men/women, years of education)? For such questions, many studies default to reporting odds ratios, which may exaggerate associations when the outcome is common. Odds ratios are often used because they are easily obtained from logistic regression models. Obtaining risk ratios or risk differences, especially adjusting for confounders, has typically required more advanced biostatistics and programming skills, including in R.

The **risks** package makes estimating adjusted risk ratios and risk differences as simple as fitting a logistic regression model. No advanced programming or biostatistics skills are required. Risk ratios or risk differences are returned whenever the data would allow for fitting a logistic model.

<https://stopsack.github.io/risks>

# Summary: Epidemiologic results tables in 2025



# Summary: Epidemiologic results tables in 2025

- The tables in our manuscripts can and should be made directly by software

# Summary: Epidemiologic results tables in 2025

- The tables in our manuscripts can and should be made directly by software
- Several R packages create various types of tables

# Summary: Epidemiologic results tables in 2025

- The tables in our manuscripts can and should be made directly by software
- Several R packages create various types of tables
  - The **riftable** package is designed specifically for epidemiologists and with best-practice result reporting in mind

# Summary: Epidemiologic results tables in 2025

- The tables in our manuscripts can and should be made directly by software
- Several R packages create various types of tables
  - The **riftable** package is designed specifically for epidemiologists and with best-practice result reporting in mind
  - One perk are regression models for adjusted risk ratios and risk differences for binary outcomes from the **risks** package

# Summary: Epidemiologic results tables in 2025

- The tables in our manuscripts can and should be made directly by software
- Several R packages create various types of tables
  - The **riftable** package is designed specifically for epidemiologists and with best-practice result reporting in mind
  - One perk are regression models for adjusted risk ratios and risk differences for binary outcomes from the **risks** package
- This automation makes table creation easier, faster, and reproducible

# Summary: Epidemiologic results tables in 2025

- The tables in our manuscripts can and should be made directly by software
- Several R packages create various types of tables
  - The **riftable** package is designed specifically for epidemiologists and with best-practice result reporting in mind
  - One perk are regression models for adjusted risk ratios and risk differences for binary outcomes from the **risks** package
- This automation makes table creation easier, faster, and reproducible
  - We can focus on thinking about our results instead of transcribing numbers

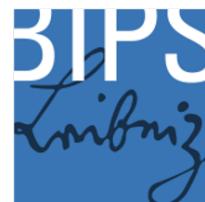
# Summary: Epidemiologic results tables in 2025

- The tables in our manuscripts can and should be made directly by software
- Several R packages create various types of tables
  - The **riftable** package is designed specifically for epidemiologists and with best-practice result reporting in mind
  - One perk are regression models for adjusted risk ratios and risk differences for binary outcomes from the **risks** package
- This automation makes table creation easier, faster, and reproducible
  - We can focus on thinking about our results instead of transcribing numbers

**Thank you for your attention – Please reach out**

[stopsack@leibniz-bips.de](mailto:stopsack@leibniz-bips.de)

# We are hiring!



The [Leibniz Institute for Prevention Research and Epidemiology – BIPS](#) is one of Germany's largest and oldest research institutes focused on epidemiology.

Applications are open now for our PhD Program in Epidemiology, Statistics, and Prevention and Implementation Science, with an October start date.



[https://tinyurl.com/  
bips-phd-program](https://tinyurl.com/bips-phd-program)

Reach out about this opportunity and others  
[stopsack@leibniz-bips.de](mailto:stopsack@leibniz-bips.de)

