

# Interoperability between {epireview} & Epiverse-TRACE

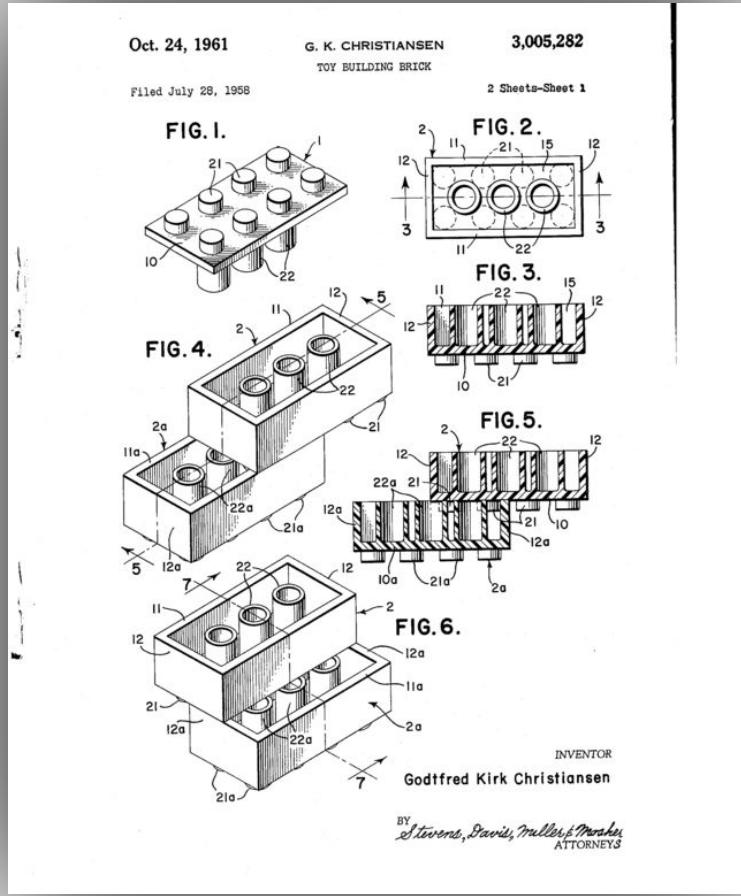


---

Joshua W. Lambert  
July 2024

# Interoperable design

Lego Patent (1961), Wikimedia commons, Christiansen, Godtfred Kirk.



# Epidemiological parameters as an *end in themselves*<sup>\*</sup>

Case fatality rate

What was the severity in the last outbreak?

Reproduction number

Is an outbreak of this pathogen usually subcritical (i.e.  $R < 1$ )?

\* Direct interpretation without calculation

# Epidemiological parameters as an *end in themselves*\*<sup>\*</sup>

Case fatality rate

What was the severity in the last outbreak?

Reproduction number

Is an outbreak of this pathogen usually subcritical (i.e.  $R < 1$ )?



\* Direct interpretation without calculation

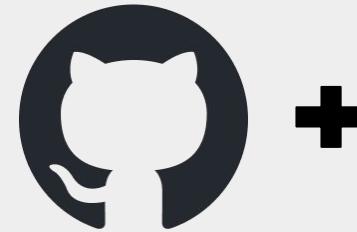
# Epidemiological parameters as an *end in themselves*\*<sup>\*</sup>

Case fatality rate

What was the severity in the last outbreak?

Reproduction number

Is an outbreak of this pathogen usually subcritical (i.e.  $R < 1$ )?



\* Direct interpretation without calculation

# Epidemiological parameters as a *means to an end*

Declare end of the outbreak

What is the maximum or 95th percentiles of the incubation period to estimate if transmission has ceased?

Calculate  $R_t$

What are the best fitting parameters for a generation time or serial interval for this outbreak?

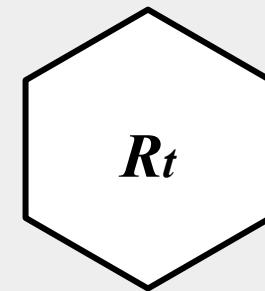
# Epidemiological parameters as a *means to an end*

Declare end of the outbreak

What is the maximum or 95th percentiles of the incubation period to estimate if transmission has ceased?

Calculate  $R_t$

What are the best fitting parameters for a generation time or serial interval for this outbreak?





# Epidemiological pipelines

WHO-Collaboratory / collaboratory-epiparameter-community

Type ⌘ to search

Code Issues Pull requests Discussions Actions Projects Security Insights

## Use of {epiparameter} in Rt pipeline #18

ntncmch started this conversation in Workstream 3: Storage and use of parameters

ntncmch on May 16

Here is the R script by [@annecori](#) and [@ntncmch](#) that attempts to use the {epiparameter} package in a typical Rt estimation pipeline, using both {EpiEstim} and {EpiNow2}.

The 10 problems discovered are marked with `# TODO-1,2,etc`. It remains to be discussed whether these issues should be posted on {epiparameter} or in other packages.

```
# EpiParameter workshop
library(epiparameter)
library(epireview)
library(simulist)
library(here)
library(tidyverse)
library(EpiEstim)
library(incidence2)
library(EpiNow2)

# downloaded from https://royalsocietypublishing.org/doi/suppl/10.1098/rstb.2016.0308
data <- read_csv(here("data/epiparameter/rstb20160308supp1.csv"))

# select Sierra Leone
data_sle <- data %>% filter(Country == "Sierra Leone")
```

Category: Workstream 3: Storage and use of parameters

Labels: None yet

4 participants:

Notifications: Unsubscribe

You're receiving notifications because you're watching this repository.

Create issue from discussion

Credit Anne Cori & Anton Camacho

# Estimate $R_t$ pipeline

## I. Read in data

PHILOSOPHICAL  
TRANSACTIONS B

[rstb.royalsocietypublishing.org](http://rstb.royalsocietypublishing.org)

Research  

Cite this article: Garske T *et al.* 2017  
Heterogeneities in the case fatality ratio in the  
West African Ebola outbreak 2013–2016.  
*Phil. Trans. R. Soc. B* **372:** 20160308.  
<http://dx.doi.org/10.1098/rstb.2016.0308>

Accepted: 11 November 2016

One contribution of 17 to a theme issue 'The  
2013–2016 West African Ebola epidemic: data,  
decision-making and disease control'.

Heterogeneities in the case fatality ratio  
in the West African Ebola outbreak  
2013–2016

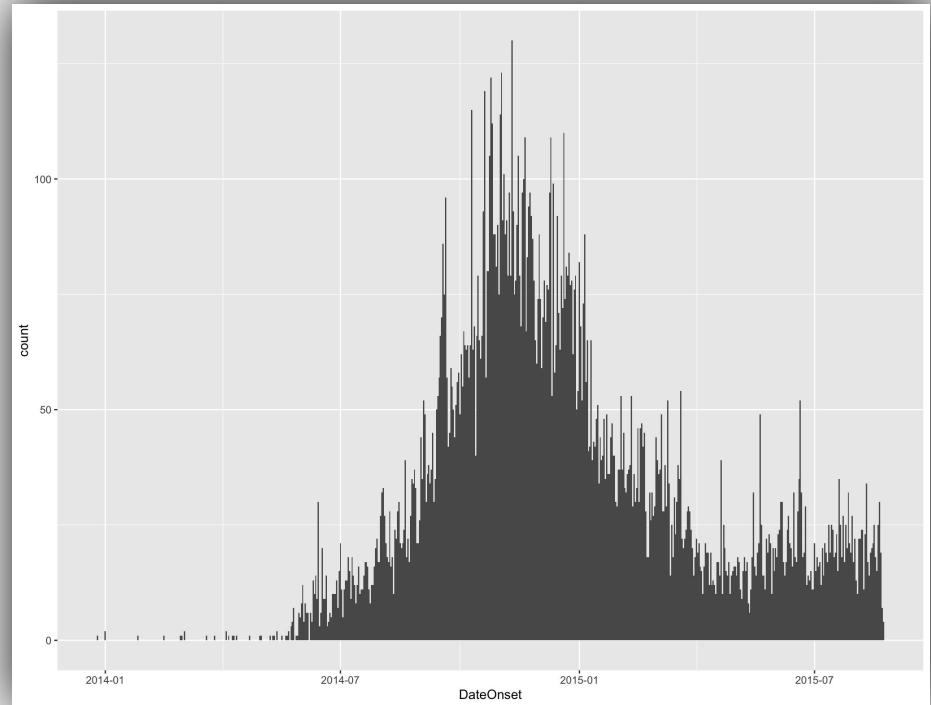
Tini Garske<sup>1</sup>, Anne Cori<sup>1</sup>, Archchun Ariyarajah<sup>5</sup>, Isobel M. Blake<sup>1</sup>,  
Ilaria Dorigatti<sup>1</sup>, Tim Eckmanns<sup>5,6</sup>, Christophe Fraser<sup>1,2</sup>, Wes Hinsley<sup>1</sup>,  
Thibaut Jombart<sup>1</sup>, Harriet L. Mills<sup>3</sup>, Gemma Nedjati-Gilani<sup>1</sup>, Emily Newton<sup>5</sup>,  
Pierre Nouvellet<sup>1</sup>, Devin Perkins<sup>5</sup>, Steven Riley<sup>1</sup>, Dirk Schumacher<sup>5</sup>,  
Anita Shah<sup>5</sup>, Maria D. Van Kerkhove<sup>1,4</sup>, Christopher Dye<sup>5</sup>, Neil M. Ferguson<sup>1</sup>  
and Christl A. Donnelly<sup>1</sup>

<sup>1</sup>MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London W2 1PG, UK  
<sup>2</sup>Big Data Institute, University of Oxford, Oxford OX3 7LF, UK  
<sup>3</sup>MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK  
<sup>4</sup>Center for Global Health Research and Education, Institut Pasteur, Paris 75015, France  
<sup>5</sup>WHO, 1211 Geneva, Switzerland  
<sup>6</sup>Robert Koch Institute, 13302 Berlin, Germany

 TG, 0000-0002-8952-4710; AC, 0000-0002-8443-9162; GN-G, 0000-0001-5723-5028;  
CAD, 0000-0002-0195-2463

# Estimate $R_t$ pipeline

2. Subset to Sierra Leone
3. Plot daily incidence curve  
(epi curve)



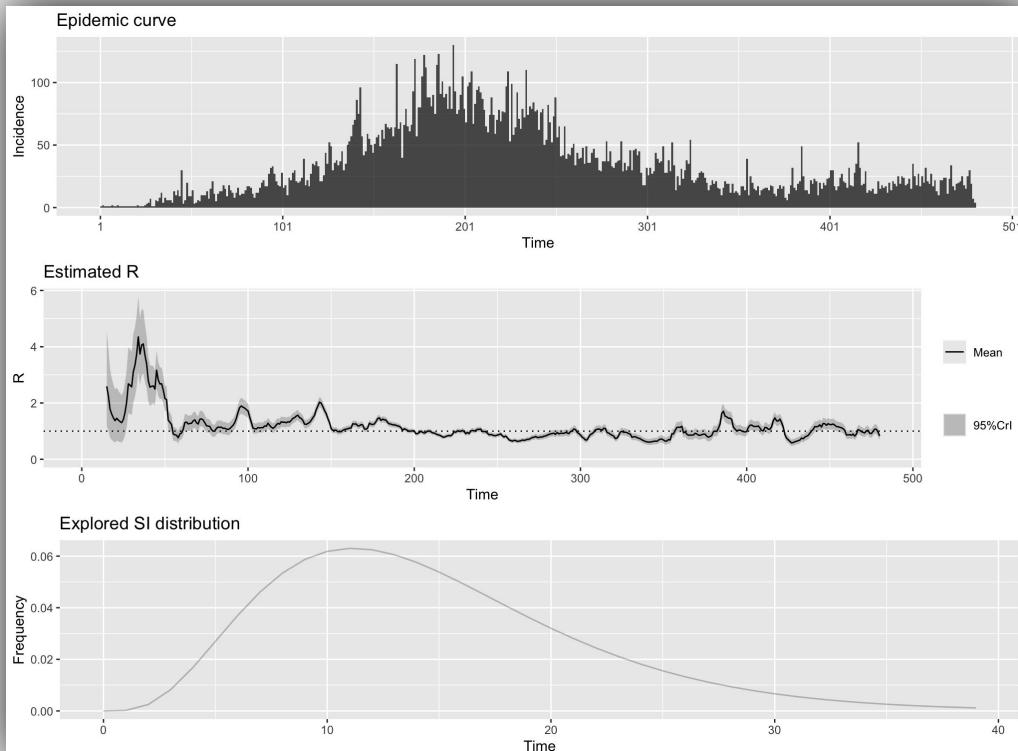
# Estimate $R_t$ pipeline

4. Convert line list data.frame to <incidence> object
5. Load Ebola epiparameters from {epireview}
6. Subset to serial interval from Faye et al. (2015)
7. Convert epireview data.frame to <epidist> object
8. Calculate PMF of discretised serial interval (<epidist> methods)



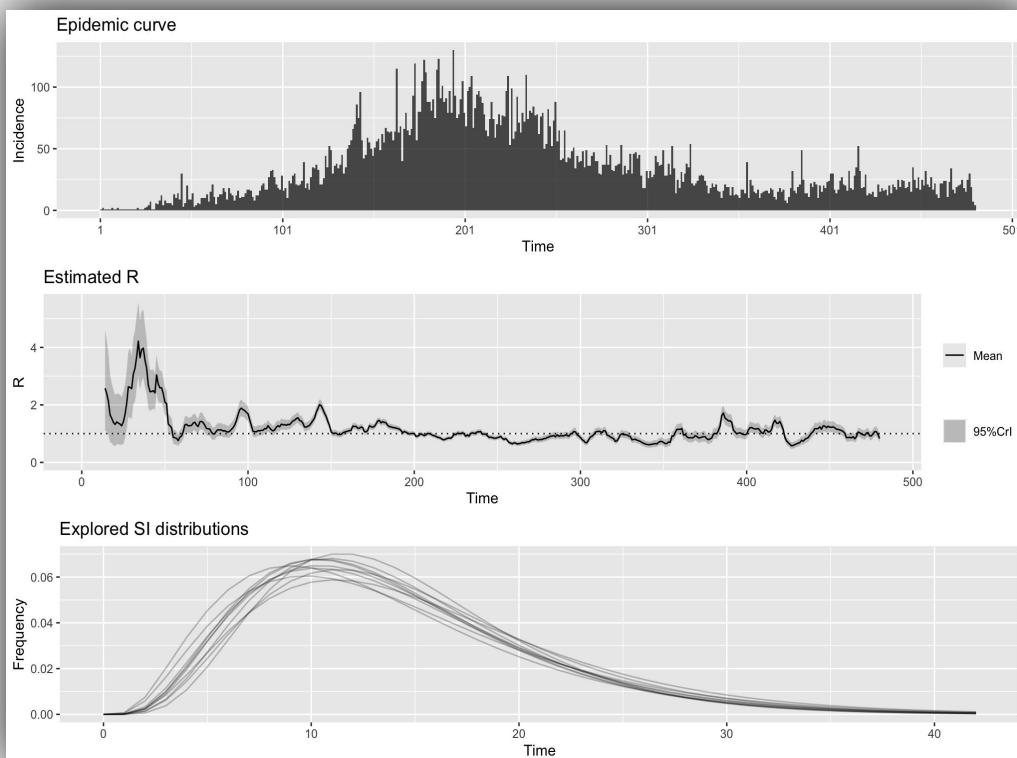
# Estimate $R_t$ pipeline

9. Calculate  $R_t$  with {EpiEstim}:
  - a. SI PMF
  - b. SI mean and SD
  - c. Uncertain SI
10. Calculate  $R_t$  with {EpiNow2}



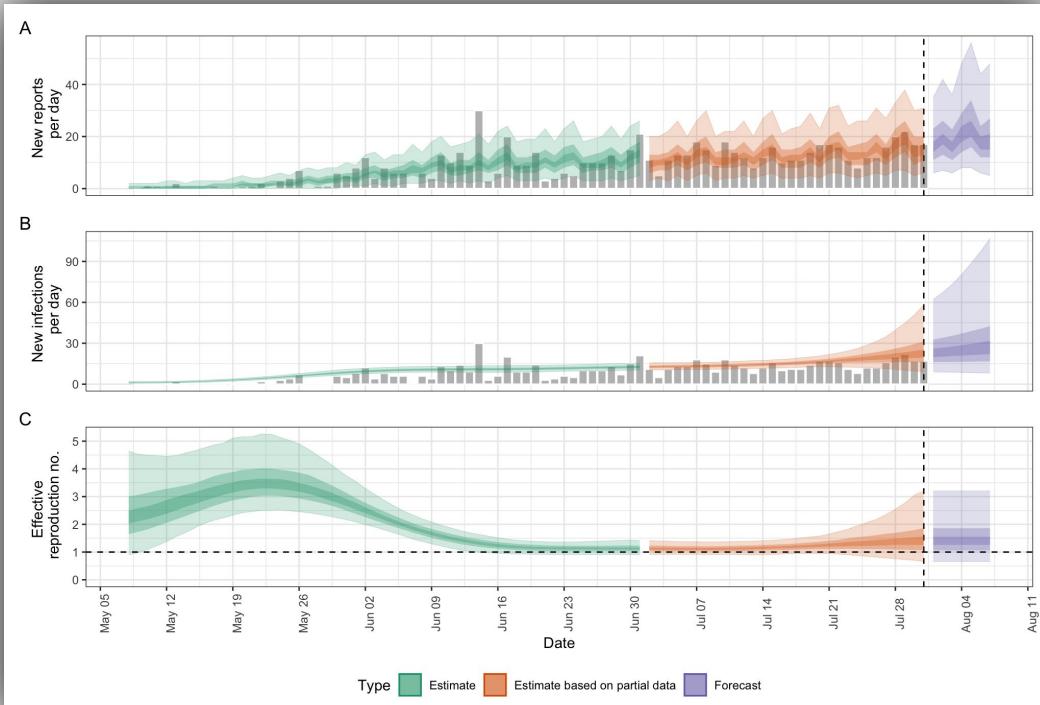
# Estimate $R_t$ pipeline

9. Calculate  $R_t$  with {EpiEstim}:
  - a. SI PMF
  - b. SI mean and SD
  - c. Uncertain SI
10. Calculate  $R_t$  with {EpiNow2}



# Estimate $R_t$ pipeline

9. Calculate  $R_t$  with {EpiEstim}:
  - a. SI PMF
  - b. SI mean and SD
  - c. Uncertain SI
10. Calculate  $R_t$  with {EpiNow2}



# *Tangent*: Collaborative epi analytics and package ecosystem

This is great, thanks for sharing! Just a couple of comments relating to EpiNow2:

```
# TODO-8: write function that compute a max based on a  
# given percentile (e.g. max such that 99% of the density)
```

See [epiforecasts/EpiNow2#533](#)

```
# TODO-9: epinow accept incidence2 objects
```

We're thinking of making the column names more flexible in EpiNow2, which would address this to some degree. We could also add an S3 method to accept incidence2 objects but I'm not sure it's worth the overhead if all it does is rename some columns (happy to convinced otherwise).

[epiforecasts/EpiNow2#505](#)

```
## NOTE this is slow
```

This is largely a consequence of the settings used. We're hoping to clarify this in the future:  
[epiforecasts/EpiNow2#629](#)

0 replies



# IMPERIAL

## Collaboratory

Pandemic and Epidemic Intelligence

EpiParameter Community

# *Tangent*: Collaborative epi analytics and package ecosystem

TimTaylor on May 23  
@annecori, @ntncmch - I've just release incidence 2.3.0 to CRAN (binaries waiting to be built but source there) and you can now do:

```
inc_sle <-  
  incidence2::incidence(data_sle, date_index = "DateOnset", complete_dates = TRUE)
```

hopefully that crosses off TODO-1.

1 reply

ntncmch on May 28 Author  
This is great, thanks! I was not aware of the function `incidence2::complete_dates` but I agree it's nice to have the option directly in `incidence2::incidence`.

0 replies



# IMPERIAL

**Collaboratory**  
Pandemic and Epidemic Intelligence

EpiParameter Community

# *Tangent*: Collaborative epi analytics and package ecosystem

joshwlambert last week edited ...

Thanks for contributing this [@ntncmch](#) and [@anneccori](#), and for pointing out the areas where improvements were needed. It's also great to see other package developers incorporating the suggestions.

**TODO-4**

TODO-4: ideally we should get prob\_dist directly  
e.g. `epiparameter:::as_epidist(serial_interval_faye, prob_dist = "gamma")`

This has been addressed in [epiverse-trace/epiparameter#335](#) and is available by reinstalling the package.

This is great, thanks! I was not aware of the function `incidence2:::complete_dates` but I agree it's nice to have the option directly in `incidence2:::incidence`.

Write a reply ↑ 1

0 replies



# IMPERIAL

## Collaboratory

Pandemic and Epidemic Intelligence

EpiParameter Community

# Idealised estimate $R_t$ pipeline

1. Data Import and cleaning
2. Convert to <incidence>
3. Load {epireview} parameters
4. Convert to discrete <epidist> with  
subsetting within `as_epidist()`
5. Calculate  $R_t$  with {EpiEstim}

# Idealised estimate $R_t$ pipeline

1. Data Import and cleaning
2. Convert to `<incidence>`
3. Load `{epireview}` parameters
4. Convert to discrete `<epidist>` with  
subsetting within `as_epidist()`
5. Calculate  $R_t$  with `{EpiEstim}`

```
inc_sle <- incidence2::incidence(  
  data_sle,  
  date_index = "DateOnset",  
  complete_dates = TRUE  
)
```

# Idealised estimate $R_t$ pipeline

1. Data Import and cleaning
2. Convert to `<incidence>`
3. Load `{epireview}` parameters
4. Convert to discrete `<epidist>` with  
subsetting within `as_epidist()`
5. Calculate  $R_t$  with `{EpiEstim}`

```
inc_sle <- incidence2::incidence(  
  data_sle,  
  date_index = "DateOnset",  
  complete_dates = TRUE  
)
```

```
ebola <- epireview::load_epidata("ebola")  
evd_SI_ep <- epiparameter::as_epidist(  
  ebola,  
  epi_dist = "serial interval",  
  author = "Faye",  
  prob_dist = "gamma",  
  discretise = TRUE  
)
```

# Idealised estimate $R_t$ pipeline

1. Data Import and cleaning
2. Convert to `<incidence>`
3. Load `{epireview}` parameters
4. Convert to discrete `<epidist>` with  
subsetting within `as_epidist()`
5. Calculate  $R_t$  with `{EpiEstim}`

```
inc_sle <- incidence2::incidence(  
  data_sle,  
  date_index = "DateOnset",  
  complete_dates = TRUE  
)
```

```
ebola <- epireview::load_epidata("ebola")  
evd_SI_ep <- epiparameter::as_epidist(  
  ebola,  
  epi_dist = "serial interval",  
  author = "Faye",  
  prob_dist = "gamma",  
  discretise = TRUE  
)
```

```
res <- estimate_R(  
  incid = inc_sle,  
  method = "non_parametric_si",  
  config = make_config(evd_SI_ep)  
)
```

# Idealised estimate $R_t$ pipeline

From

```
# transform LL into incidence TS
inc_sle <- incidence2::incidence(data_sle, date_index = "DateOnset") %>%
  complete()
  date_index = seq(min(date_index), max(date_index), 1),
  fill = list(count = 0, count_variable = "DateOnset")
)

# get Ebola serial interval from EpiReview via Epiparameter
ebola <- epireview::load_epidata("ebola")
params <- ebola[["params"]]
param_evd_faye <- params %>% filter(covidence_id == 4966)
serial_interval_faye <-
  param_evd_faye %>% filter(parameter_type == "Human delay - serial interval")

serial_interval_faye %>% select(
  pathogen,
  article_label,
  parameter_type,
  parameter_value_type,
  parameter_value,
  parameter_unit
)

# create epidist object from epireview
evd_SI_ep <- epiparameter:::as_epidist(serial_interval_faye)
evd_SI_ep_param <-
  convert_summary_stats_to_params("gamma",
    mean = evd_SI_ep$summary_stats$mean,
    sd = evd_SI_ep$summary_stats$sd)
evd_SI_ep$prob_dist <-
  distributional::dist_gamma(shape = evd_SI_ep_param$shape,
    rate = 1 / evd_SI_ep_param$scale)

is_parameterised(evd_SI_ep)
plot(evd_SI_ep, day_range = 0:50)
evd_SI_pmf <- density(discretise(evd_SI_ep), at = 0:50)

# EpiEstim example -----
res1 <- estimate_R(incid = inc_sle$count,
  method = "non_parametric_si",
  config = make_config(list(si_distr = evd_SI_pmf)))
plot(res1)

# epiestim v2
res2 <- estimate_R(incid = inc_sle$count,
  method = "parametric",
  config = make_config(
    list(
      mean_si = evd_SI_ep$summary_stats$mean,
      std_si = evd_SI_ep$summary_stats$sd
    )
  ))
plot(res2)
```

To

```
inc_sle <- incidence2::incidence(
  data_sle,
  date_index = "DateOnset",
  complete_dates = TRUE
)
```

```
ebola <- epireview::load_epidata("ebola")
evd_SI_ep <- epiparameter:::as_epidist(
  ebola,
  epi_dist = "serial interval",
  author = "Faye",
  prob_dist = "gamma",
  discretise = TRUE
)
```

```
res <- estimate_R(
  incid = inc_sle,
  method = "non_parametric_si",
  config = make_config(evd_SI_ep)
)
```



# Estimate the final size of an epidemic

```
ebola <- epireview::load_epidata(pathogen = "ebola")
ebola_params <- ebola$params
ebola_r <- ebola_params[grep(
  pattern = "Reproduct", x = ebola_params$parameter_type
), ]
ebola_r_ep <- epiparameter::as_epidist(ebola_r[10, ])
finalsize::final_size(r0 =
ebola_r_ep$summary_stats$mean)
#>      demo_grp    susc_grp susceptibility p_infected
#> 1  demo_grp_1  susc_grp_1                 1  0.7968124
```



# Estimate the case fatality risk

```
data(ebola1976, package = "cfr")
ebola <- epireview::load_epidata(pathogen = "ebola")
ebola_params <- ebola$params
ebola_onset_to_death <- ebola_params[
  grep(pattern = "Death", x = ebola_params$parameter_type),
]
onset_to_death <- epiparameter::as_epidist(ebola_onset_to_death[27, ])
cfr::cfr_static(data = ebola1976, delay_density = as.function(onset_to_death))

#> Total cases = 245 and p = 0.959: using Normal approximation to binomial likelihood.
#>   severity_estimate severity_low severity_high
#> 1           0.9742      0.8356      0.9877
```

# Design patterns for interoperability

## Mapping between fields

- Clear and consistent naming of fields

## Validation on both sides

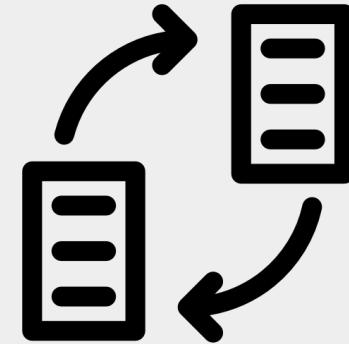
- Ensure data complies with definitions (data dictionary)

## Minimise overlap in package functionality

- Well-defined scope of packages, classes and methods

## Utilise existing infrastructure

- Class dispatch in R (S3)
- Use existing classes (distribution and citation)



[Data icons created by  
Freepik - Flaticon](#)



# Balancing benefits and downsides of interoperability

Pros:

- Entirely reproducible in a single environment (no manual copy and pasting or moving between R and Excel)
- Leverage well tested packages and functions to make robust analyses
- Predictable interoperable pipelines can be packaged into automated reports (e.g. {episoap})

Cons:

- Tightly coupled packages can create fragile code when upstream dependencies have breaking changes



# Thanks for listening

---

## Any Questions?

LONDON  
SCHOOL *of*  
HYGIENE  
& TROPICAL  
MEDICINE



**data.org**