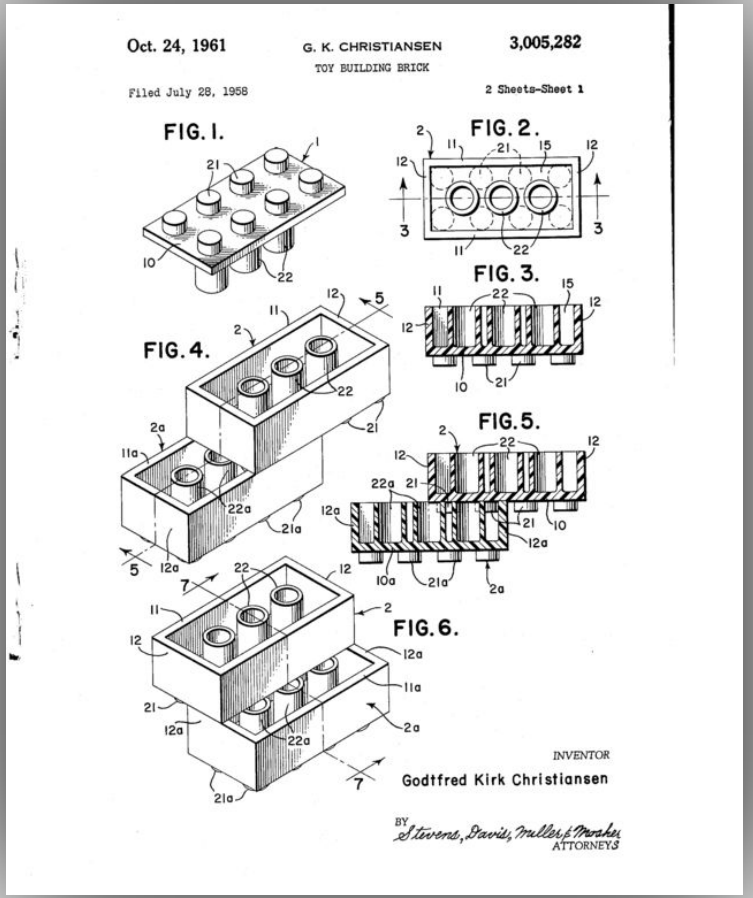
Interoperability between {epireview} & Epiverse-TRACE

Joshua W. Lambert
July 2024

Interoperable design



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

Epidemiological parameters as an *end in themselves**

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**

Case fatality rate	Reproduction number
What was the severity in the last outbreak?	Is an outbreak of this pathogen usually subcritical (i.e. $R < 1$)?

* Direct interpretation without calculation

Epidemiological parameters as an *end in themselves**

Case fatality rate	Reproduction number
What was the severity in the last outbreak?	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?

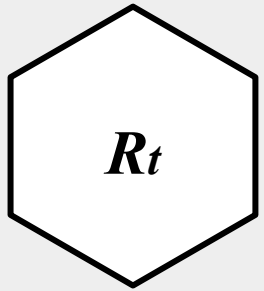
Calculate R_t

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

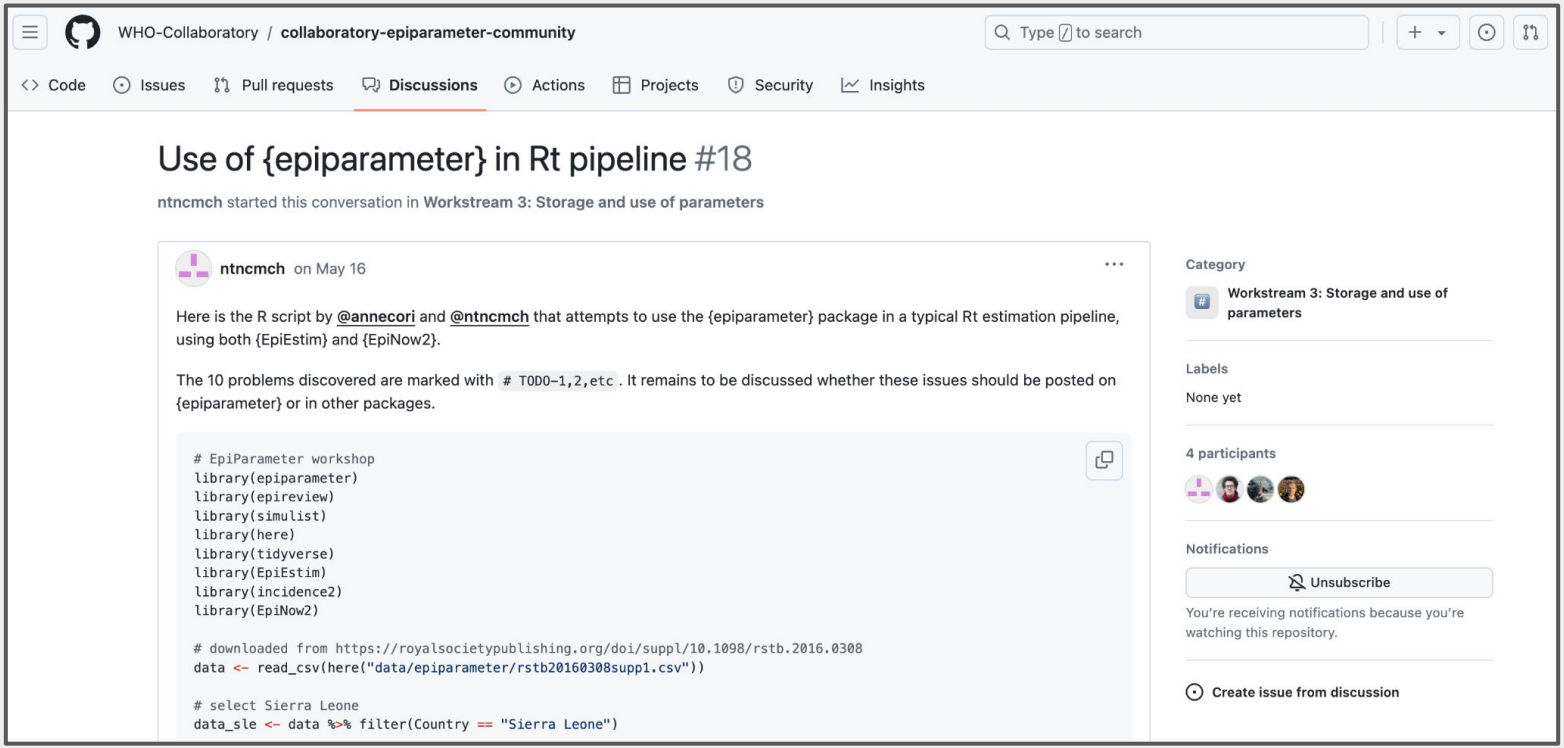
+

+

+



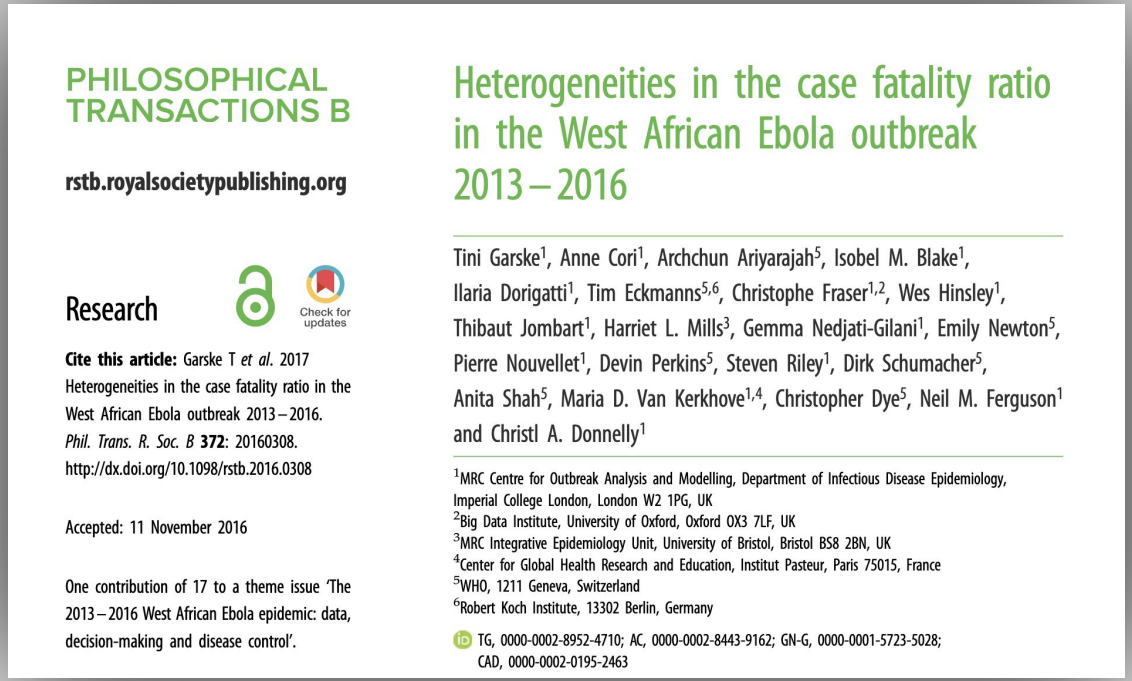
Epidemiological pipelines



Credit Anne Cori & Anton Camacho

Estimate R_t pipeline

1. Read in data



PHILOSOPHICAL TRANSACTIONS B

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

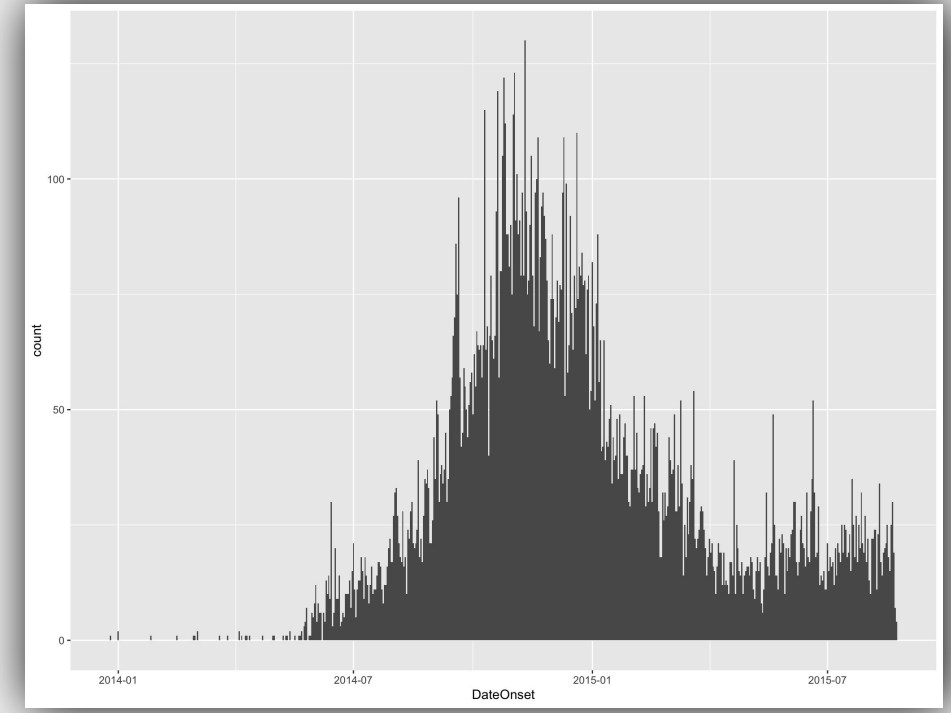
[1]MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London W2 1PG, UK
[2]Big Data Institute, University of Oxford, Oxford OX3 7LF, UK
[3]MRC Integrative Epidemiology Unit, University of Bristol, Bristol BS8 2BN, UK
[4]Center for Global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

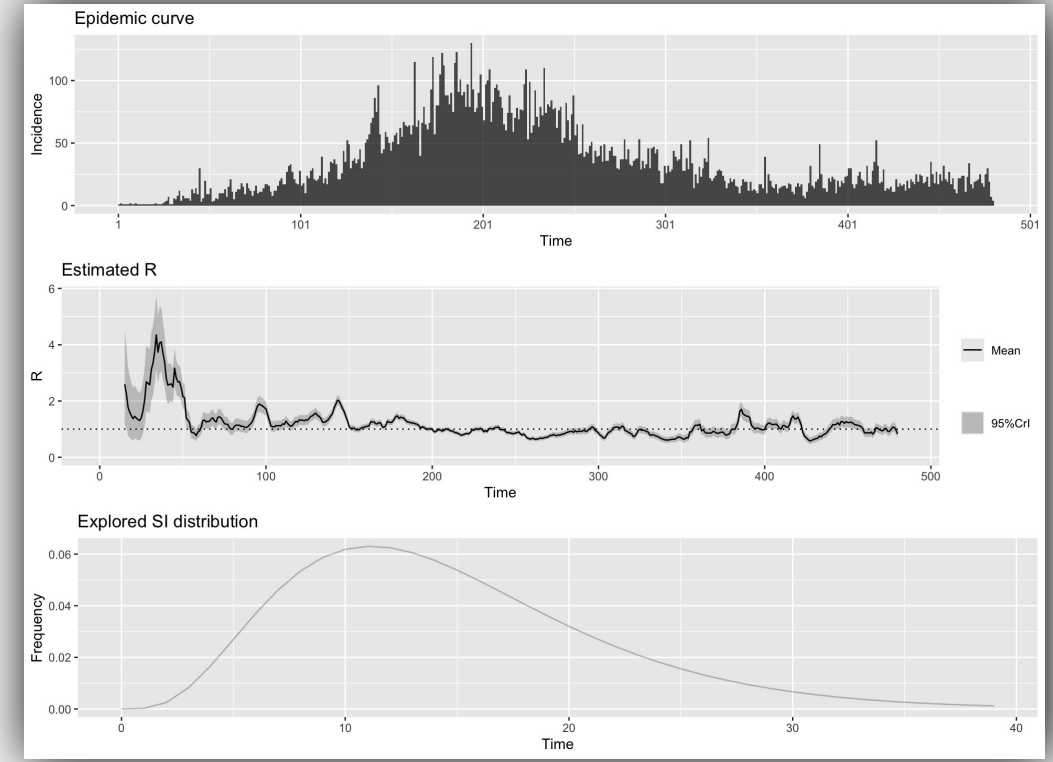
7. Convert epireview `data.frame` to `<epidist>` object

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

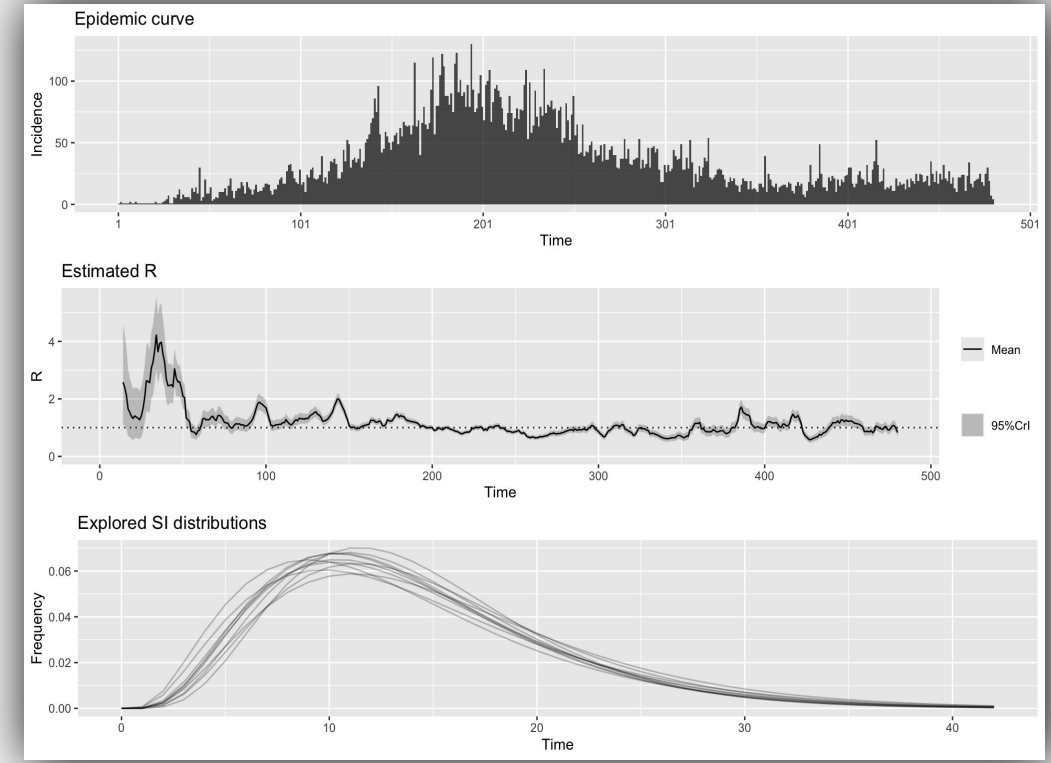
10. Calculate R_t with {EpiNow2}



Estimate R_t pipeline

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

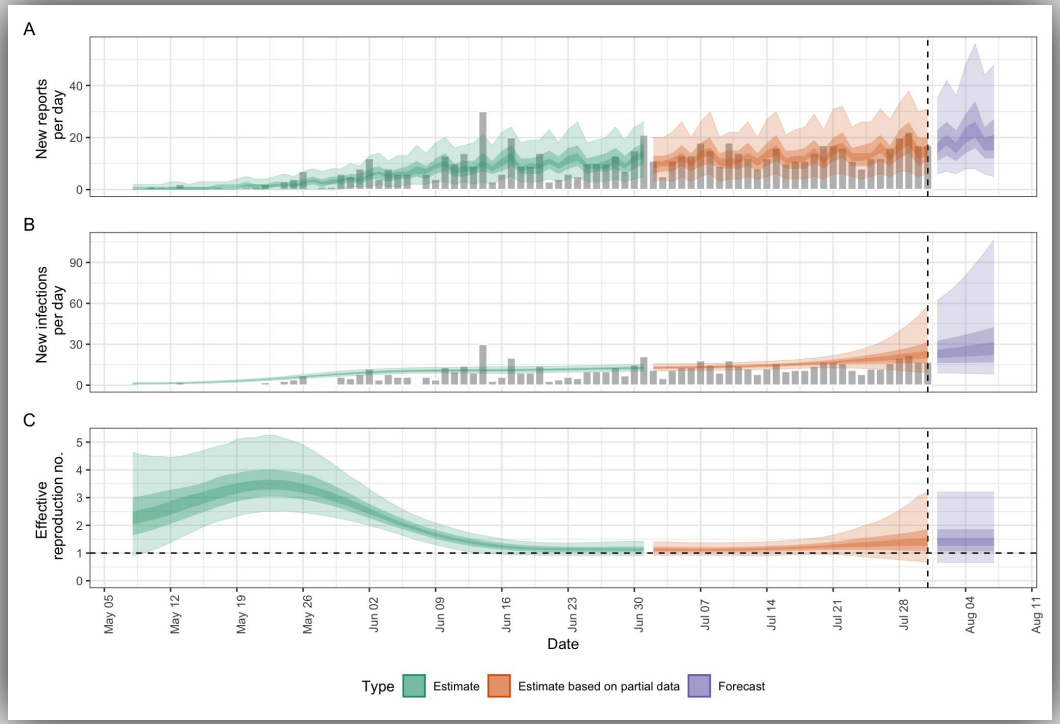
10. Calculate R_t with {EpiNow2}



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

sbfnk on May 21

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

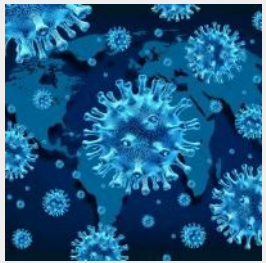
```
# TODO-9: epinow accept incidence2 objects
```

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

↑ 1 0 replies



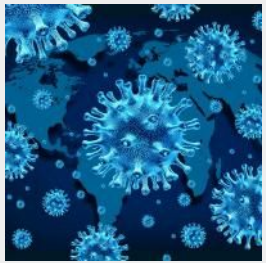
Epiverse TRACE
powered by data.org

IMPERIAL

Collaboratory
Pandemic and Epidemic Intelligence

EpiParameter Community

Tangent: Collaborative epi analytics and package ecosystem



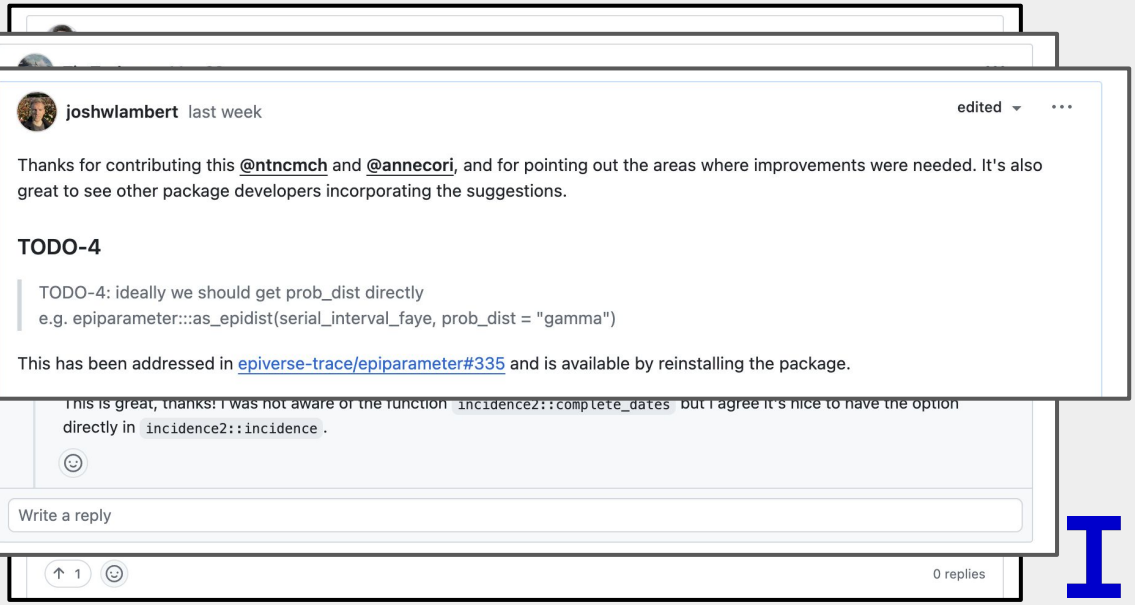
Epiverse TRACE powered by data.org

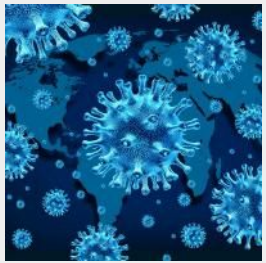
IMPERIAL

Collaboratory
Pandemic and Epidemic Intelligence

EpiParameter Community

Tangent: Collaborative epi analytics and package ecosystem





Epiverse TRACE
powered by data.org

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}

```r
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}

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

```r
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}

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

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

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

Idealised estimate R_t pipeline

From

```r
# 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(discretize(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

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

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

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

Estimate the final size of an epidemic

```r
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

```r
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

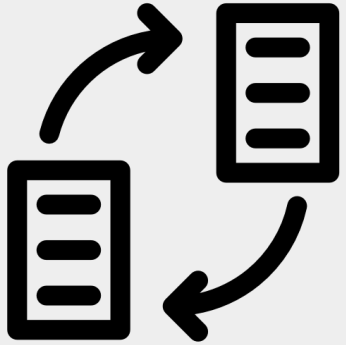
- Ensure data complies with definitions (data dictionary)

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:

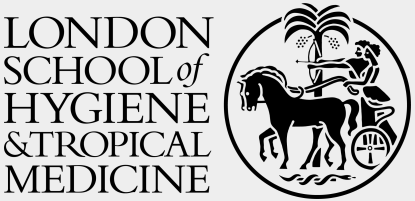
- Entirely reproducible in a single environment (no manual copy and pasting or moving between R and Excel)
- Leverage well tested packages and functions to make robust analyses
- Predictable interoperable pipelines can be packaged into automated reports (e.g. {episoap})

Cons:

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

Thanks for listening

Any Questions?



data.org