Welcome: Consultation on Defining a Standardized Data Model for Epidemiological Parameters Extraction

Tuesday 20th June 2023, 14:00 -16:00 CET

World Health Organization

Meeting objective and agenda

Objective:

To seek consensus on the data model for the extraction and representation of essential epidemiological parameters

Timetable for TWG Meeting: 20th June 2023 14:00 – 16:00 CET	
14:00 – 14:05	**Welcome and introduction** – Julia Fitzner and Patricia Ndumbi, WHO
14:05 – 14:10	**Summary of previous workshop** – Megan Evans, WHO
14:10 – 14:15	**Extracting data from the literature** - Lisa Waddell, PHAC
14:15 – 14:20	**Menti ice breaker** - Carmen Tamayo Cuartero, Epiverse/LSHTM
14:20– 14:50	**Interactive exercise on Public Health Use Cases and Data Requirements for Epidemiological Parameters** - Joshua Lambert, Epiverse/LSHTM
14:50 – 15:50	**Review and discussion of data model survey results** – Finlay Campbell, WHO
15:50 – 16:00	**Closing remarks, next steps, and final survey** – Patricia Ndumbi, WHO

Global repository of epidemiological parameters

Problem Statement

There is currently no global repository of epidemiological parameters that the global modelling community can access, use, and contribute to.

Future State

A global repository of epidemiological parameters that is publicly accessible by modellers, epidemiologist, subject matter experts and decision makers to inform public health response.

Proposed approach (5 workstreams)

World Health Organization

Feedback from the first TWG

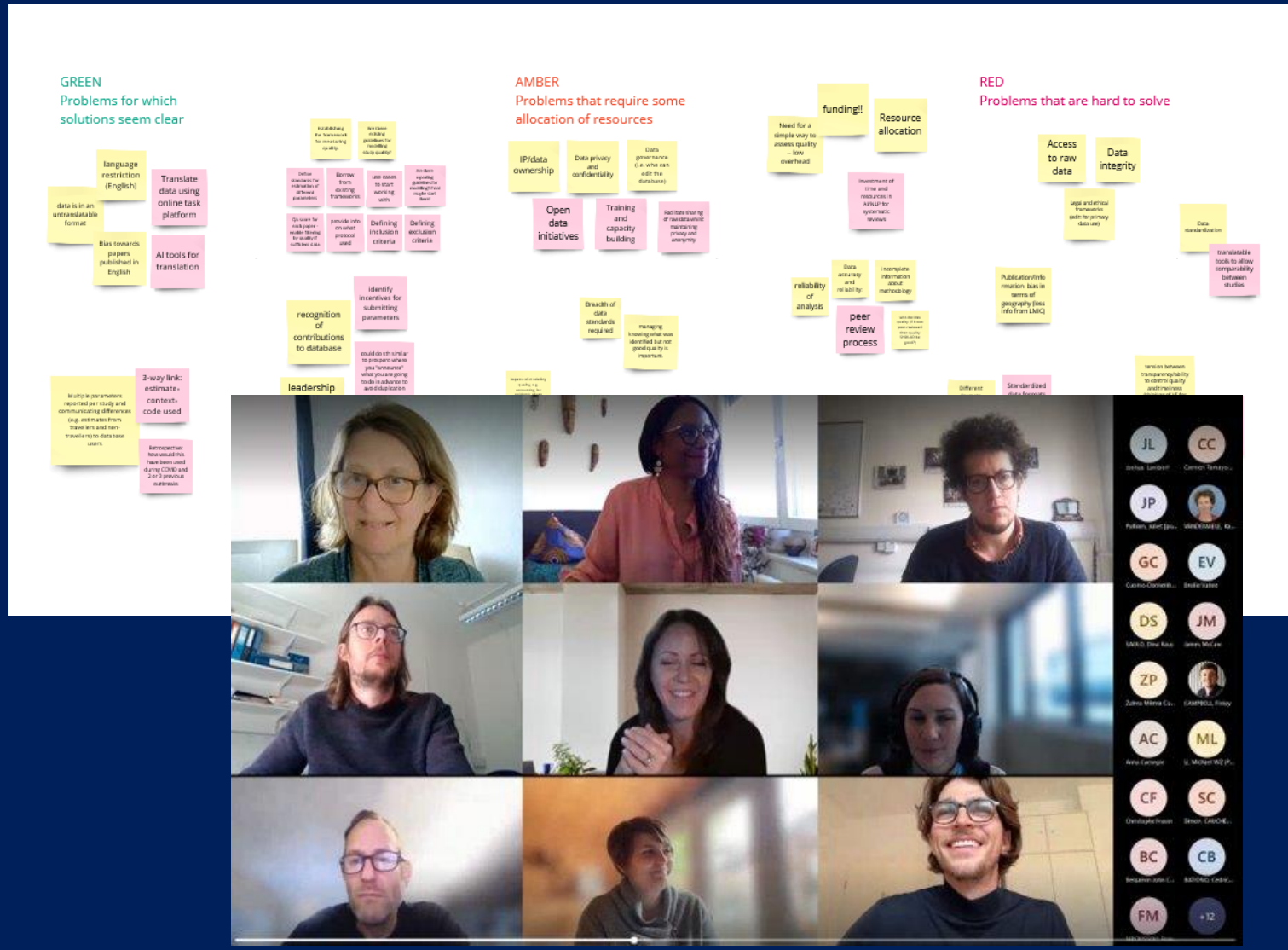
The Workshop

Attendees: Academia and Public Health

- Epidemiology / Surveillance
- Modelers
- Knowledge Synthesis Experts

Positive Feedback from Workshop

Interest in Workstreams 1-5



Miro Board Activity and Attendees

World Health Organization

Key Discussion Points

Menti Board 1: Ideal characteristics of the repository

Menti Board 2: Potential benefits of the repository

World Health Organization

Extracting data from the literature

Estimating the incubation period of monkeypox virus during the 2022 multi-national outbreak

Kelly Charniga, Nina B. Masters, Rachel B. Slayton, Lucas Gosdin, Faisal S. Minhaj, David Philpott, Dallas Smith, Shannon Gearhart, Francisco Alvarado-Ramy, Clive Brown, Michelle A. Waltenburg, Christine M. Hughes, Yoshinori Nakazawa

doi: https://doi.org/10.1101/2022.06.22.22276713

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

Abstract

Monkeypox is a zoonotic disease endemic in Central and West Africa. In May 2022, an outbreak of monkeypox characterized by human-to-human transmission was detected in multiple non-endemic countries. We estimated the incubation period for monkeypox using information from 22 probable (N = 1) and confirmed (N = 21) monkeypox cases in patients reported in the United States through June 6, 2022. We pooled U.S. patient data with the data from 18 confirmed cases in patients reported from the Netherlands through May 31, 2022. The mean incubation period from exposure to first symptom onset was 7.6 days (95% credible interval: 6.2 – 9.7), and the 95th percentile was 17.1 days (95% CrI: 12.7–24.3). These findings align with current CDC recommendations for monitoring close contacts of people with monkeypox for 21 days after their last exposure.

EMERGING INFECTIOUS DISEASES®

ISSN: 1080-60

EID Journal > Volume 29 > Number 4—April 2023 > Main Article

Volume 29, Number 4—April 2023

Dispatch

Serial Interval and Incubation Period Estimates of Monkeypox Virus Infection in 12 Jurisdictions, United States, May–August 2022

Zachary J. Madewell[1], Kelly Charniga[1], Nina B. Masters, Jason Asher, Lily Fahrenwald, William Still, Judy Chen, Naama Kipperman, David

On This Page

EMERGING INFECTIOUS DISEASES®

EID Journal > Volume 28 > Number 12—December 2022 > Main Article

Volume 28, Number 12—December 2022

Research

National Monkeypox Surveillance, Central African Republic, 2001–2021

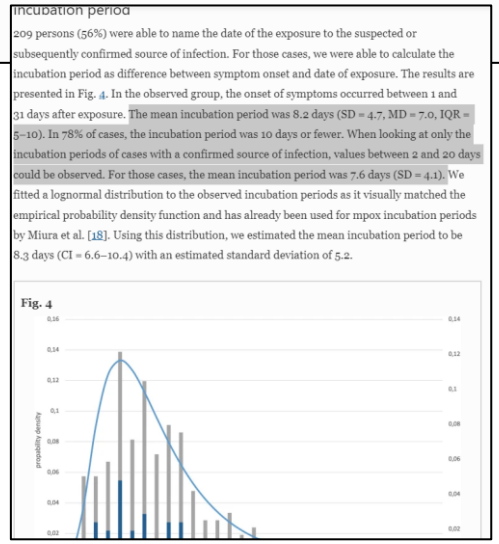
Camille Besombes, Festus Mbrenga, Laura Schaeffer, Christian Malaka, Ella Gonofio, Jordi Landier, Ulrich Vickos, Xavier Konamna,

Research | Published: 14 February 2023

Mpox outbreak 2022: an overview of all cases reported to the Cologne Health Department

Sophia Toya Kröger, Max Christian Lehmann, Melanie Treutlein, Achim Fiethe, Annelene Kossow, Annika Küfer-Weiß, Johannes Nießen & Barbara Grüne

Infection (2023) | Cite this article



Incubation period

209 persons (56%) were able to name the date of the exposure to the suspected or subsequently confirmed source of infection. For those cases, we were able to calculate the incubation period as difference between symptom onset and date of exposure. The results are presented in Fig. 4. In the observed group, the onset of symptoms occurred between 1 and 31 days after exposure. The mean incubation period was 8.2 days (SD = 4.7, MD = 7.0, IQR = 5–10). In 78% of cases, the incubation period was 10 days or fewer. When looking at only the incubation periods of cases with a confirmed source of infection, values between 2 and 20 days could be observed. For those cases, the mean incubation period was 7.6 days (SD = 4.1). We fitted a lognormal distribution to the observed incubation periods as it visually matched the empirical probability density function and has already been used for mpox incubation periods by Miura et al. [18]. Using this distribution, we estimated the mean incubation period to be 8.3 days (CI = 6.6–10.4) with an estimated standard deviation of 5.2.

Fig. 4

Mpox	Source study	Incubation period	Method highlights
Study 1	Case report	Malaise 3 days/Rash 5 days after exposure	Descriptive information on 1 observation from UK early in 2022 outbreak.
Study 2	Summary of surveillance data collected at country level during early outbreak	**exposure to first symptom onset**: • mean 7.6 days (95% CrI: 6.2–9.7) • median 6.4 (95% CrI: 5.1 – 7.9) and a standard deviation of 1.8 days (95% CrI: 1.6–2.2). • 95th percentile 17.1 days (95% CrI: 12.7–24.3) *Same data provided for **exposure to rash onset**	US data up to June 6, n=22 +18 cases from the Netherlands. Analysis included constructing a doubly censored dataset with assumed log-normal distribution of the incubation period and use of the Metropolis-Hastings Markov chain Monte Carlo algorithm for calibration.
Study 3	Summary of an outbreak from single city.	• mean 8.2 days (SD = 4.7). • **Only confirmed cases:** mean 7.6 days (SD = 4.1) and range 2 to 20 days	368 cases Cologne Germany up to Sept 17 2022.
Study 4	Retrospective study of 16 historical outbreaks in Congo	exposure and symptom onset: • median 7 (range 0–17; IQR 1–13) days.	16 outbreaks, 2001-2021 with 327 persons investigated. Only 29 people had incubation period data.

Case reports – descriptive information on a parameter.
Does this meet the minimum requirements for inclusion?

M			Method highlights
Study 1	Case report	Malaise 3 days/Rash 5 days after exposure	Descriptive information on 1 observation from UK early in 2022 outbreak.
Study 2	Summary of surveillance data collected at country level during early outbreak	**exposure to first symptom onset**: • mean 7.6 days (95% CrI: 6.2–9.7) • median 6.4 (95% CrI: 5.1 – 7.9) and a standard deviation of 1.8 days (95% CrI: 1.6–2.2). • 95th percentile 17.1 days (95% CrI: 12.7–24.3) *Same data provided for **exposure to rash onset**	US data up to June 6, n=22 +18 cases from the Netherlands. Analysis included constructing a doubly censored dataset with assumed log-normal distribution of the incubation period and use of the Metropolis-Hastings Markov chain Monte Carlo algorithm for calibration.
Study 3	Summary of an outbreak from single city.	• mean 8.2 days (SD = 4.7). • **Only confirmed cases:** mean 7.6 days (SD = 4.1) and range 2 to 20 days	368 cases Cologne Germany up to Sept 17 2022.
Study 4	Retrospective study of 16 historical outbreaks in Congo	exposure and symptom onset: • median 7 (range 0–17; IQR 1–13) days.	16 outbreaks, 2001-2021 with 327 persons investigated. Only 29 people had incubation period data.

Mpox	Source study	Incubation period	
Study 1	Case report	Malaise 3 days/Ra	
Study 2	Summary of surveillance data collected at country level during early outbreak	**exposure to first symptom onset**: • mean 7.6 days (95% CrI: 6.2–9.7) • median 6.4 (95% CrI: 5.1 – 7.9) and a standard deviation of 1.8 days (95% CrI: 1.6–2.2). • 95th percentile 17.1 days (95% CrI: 12.7–24.3) *Same data provided for **exposure to rash onset**	US data up to June 6, n=22 +18 cases from the Netherlands. Analysis included constructing a doubly censored dataset with assumed log-normal distribution of the incubation period and use of the Metropolis-Hastings Markov chain Monte Carlo algorithm for calibration.
Study 3	Summary of an outbreak from single city.	• mean 8.2 days (SD = 4.7). • **Only confirmed cases:** mean 7.6 days (SD = 4.1) and range 2 to 20 days	368 cases Cologne Germany up to Sept 17 2022.
Study 4	Retrospective study of 16 historical outbreaks in Congo	exposure and symptom onset: • median 7 (range 0–17; IQR 1–13) days.	16 outbreaks, 2001-2021 with 327 persons investigated. Only 29 people had incubation period data.

Summary estimates can be provided in many ways. e.g. mean, median with different measures of variability.
Some parameters may need additional information to specify the definition of the parameter.

Mpox	Source study	Incubation period	Method highlights
Study 1	Case report	Malaise 3 days/Rash 5 days after eposure	Descriptive information on 1 observation from UK early in 2022 outbreak.
Study 2	Summary of surveillance data collected at country level during early outbreak	**exposure to first symptom onset**: • mean 7.6 days (95% CrI: 6.2–9.7) • median 6.4 (95% CrI: 5.1 – 7.9) and a standard deviation of 1.8 days (95% CrI: 1.6–2.2). • 95th percentile 17.1 days (95% CrI: 12.7–24.3) *Same data provided for **exposure to rash onset**	US data up to June 6, n=22 +18 cases from the Netherlands. Analysis included constructing a doubly censored dataset with assumed log-normal distribution of the incubation period and use of the Metropolis-Hastings Markov chain Monte Carlo algorithm for calibration.
Study 3	Summary of an outbreak from single city.	• mean 8.2 days (SD = 4.7). • **Only confirmed cases:** mean 4.1) and range 2 to 20 days	368 cases Cologne Germany up to
Study 4	Retrospective study of 16 historical outbreaks in Congo	exposure and symptom onset: • median 7 (range 0–17; IQR 1–13) days.	16 outbreaks, 2001-2021 with 327 persons investigated. Only 29 people had incubation period data.

What should the minimum data criteria for inclusion be? e.g. are median – range or IQR data useful/useable?

The data model

Provides the structure to capture what parameter estimates exist for a disease parameter.

- This will allow rapid assessment of:
  - How many studies estimate the parameter?
  - How much agreement is there across studies?
  - How good or generalizable are the estimates?

- Considerations for today's discussion:
  - The data from the literature is reported in variable ways --> does that data model accommodate what is necessary to capture a range of reporting styles?
  - Not all parameter estimates will be useful --> considerations for minimum criteria for parameter information to be included from a study?
  - What kind of contextual data do we need for each entry so we can use the data without having to open each paper?

Mentimer

Link: https://www.menti.com/al9kw7oa68d2
Or visit **Menti.com** and use code: **3614 8289**
Or scan the QR code

Epidemiological Parameter Public Health Use Case Examples

Modeling disease transmission dynamics and estimating reproduction numbers

- Transmission speed and epidemic potential (e.g., R_0 & R_t)
- Disease natural history (e.g., incubation period)
- Regional transmission
- Accounting for delays (e.g., reporting delay)

Evaluating the impact of interventions and control measures

- Transmission dynamics (e.g., R_0 and R_t)
- Hospital capacity
- Serological analysis & population immunity
- Superspreading

Assessing the effectiveness of vaccination campaigns

- Transmission in vaccinated & unvaccinated groups
- Vaccine effectiveness
- Transmission dynamics for variants/subtypes
- Hospital capacity
- Vaccination strategy (e.g., ring versus mass vaccination)

Exercise example

Example use-case

Understand transmission in the first two weeks of an outbreak in region *x*

What parameters are required?

- Mean and range for a generation time or serial interval
- Distribution parameters

What other contextual information is required/desired?

- When and where was the data collected (used to estimate parameters)

What are the potential limitations?

- Varying reporting styles
- Lack of contextual information

Miro exercise

Objective: To encourage participants to think about the epidemiological parameters needed for specific use cases, the contextual information required, and the potential limitations related to parameter collection and database implementation.

Quantifying the spread of a disease in a specific population or geographic area and assessing the impact of a public health interventions.

Required Parameters	Contextual Information	Limitations
• *Basic and effective reproduction number*	• *Region* • *Demography* • *Time window of data collection* • *Importation versus local transmission*	• *Inconsistent reporting* • *Missing information in report/article* • *Parameters from past outbreaks may not be representative on current outbreak*

Key message

- The database model/structure should accommodate a diverse range of epidemiological parameters relevant to various public health use cases.
- Contextual information is crucial for accurate parameter representation in the database.
- Recognizing the limitations and challenges associated with parameter collection and database implementation is essential for developing an effective and robust database model/structure.

Review and discussion of survey results

Aim:

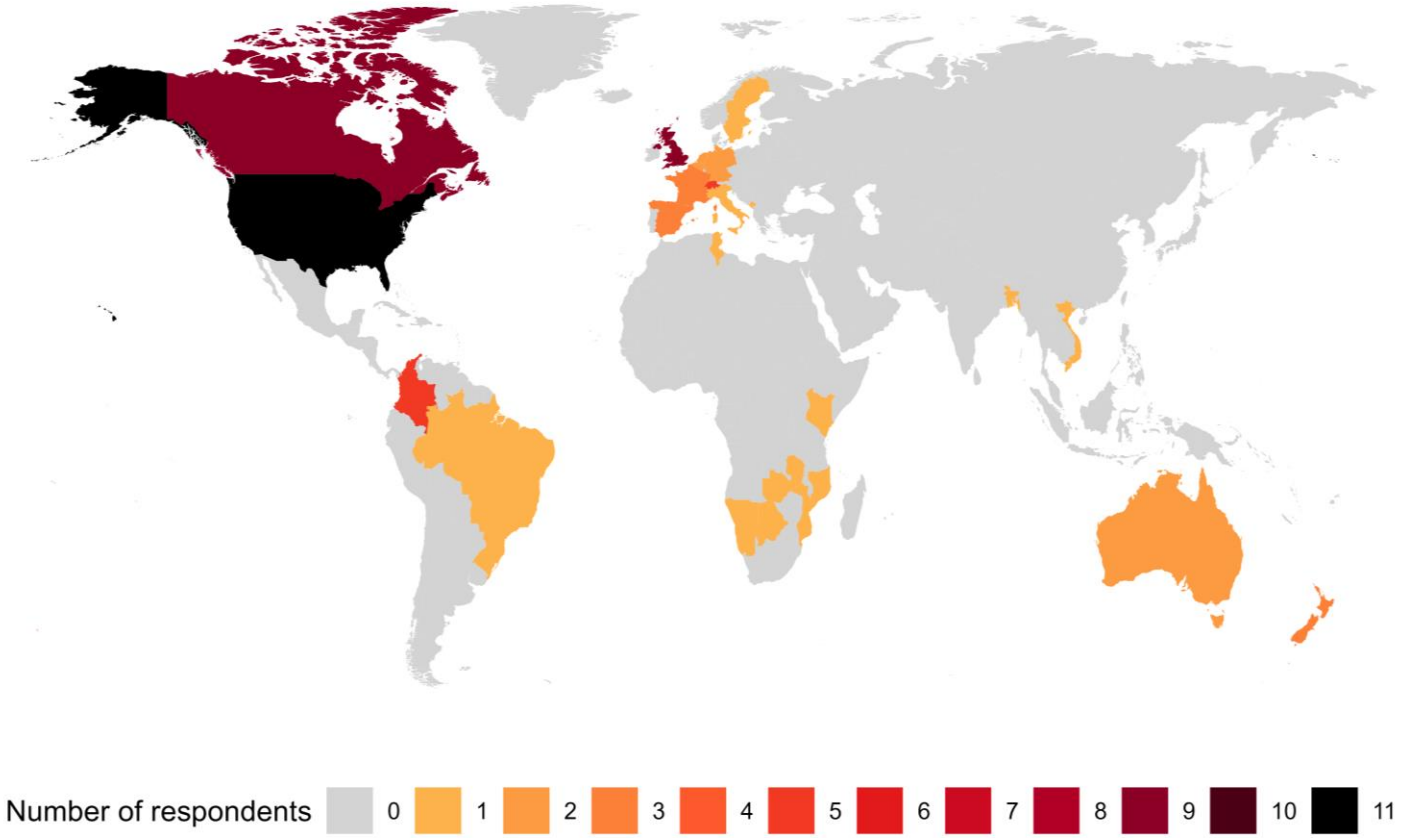
Consolidate database structure in accordance with community perspectives and needs

Structure:

1. Overview of survey participants - who is the “community”?
2. Survey results, section by section
   a) Summary of results and feedback
   b) Brief discussion of key questions and issues
3. Concluding discussion

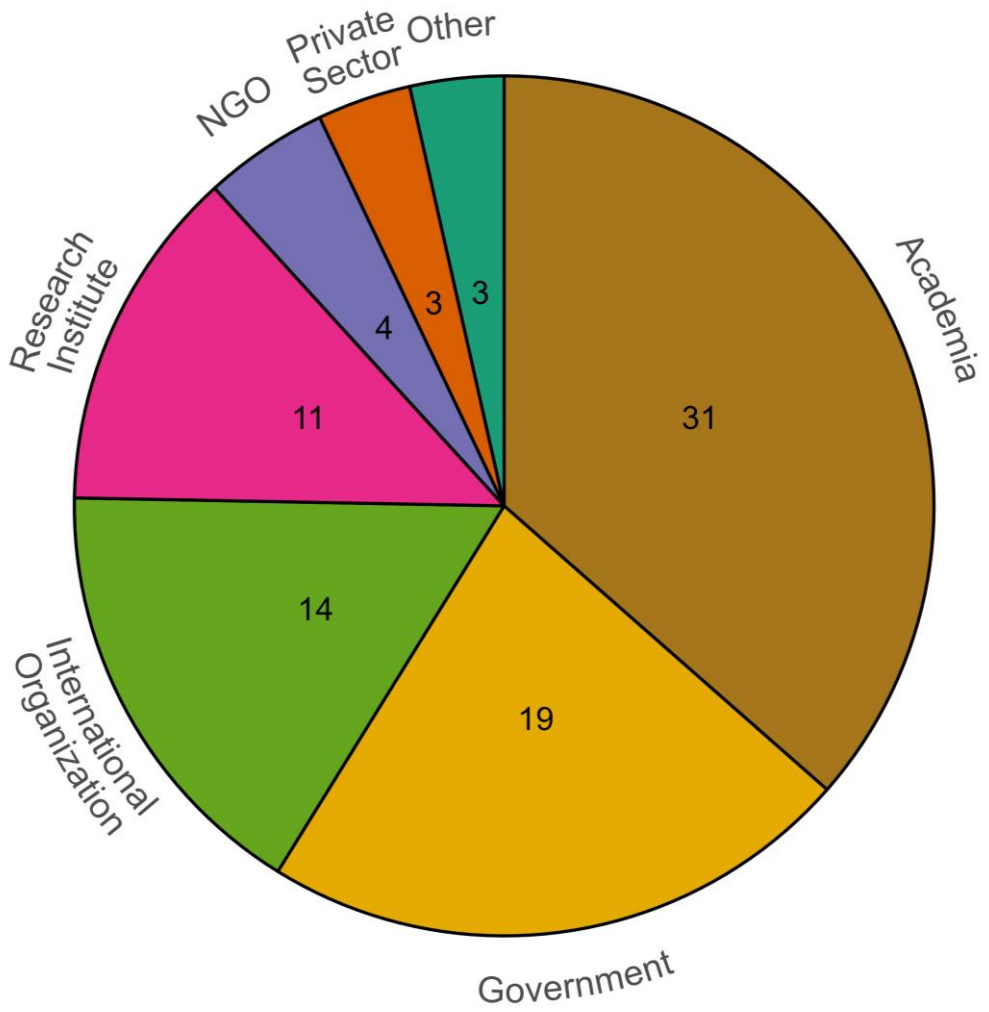
Section 1: Survey Participants

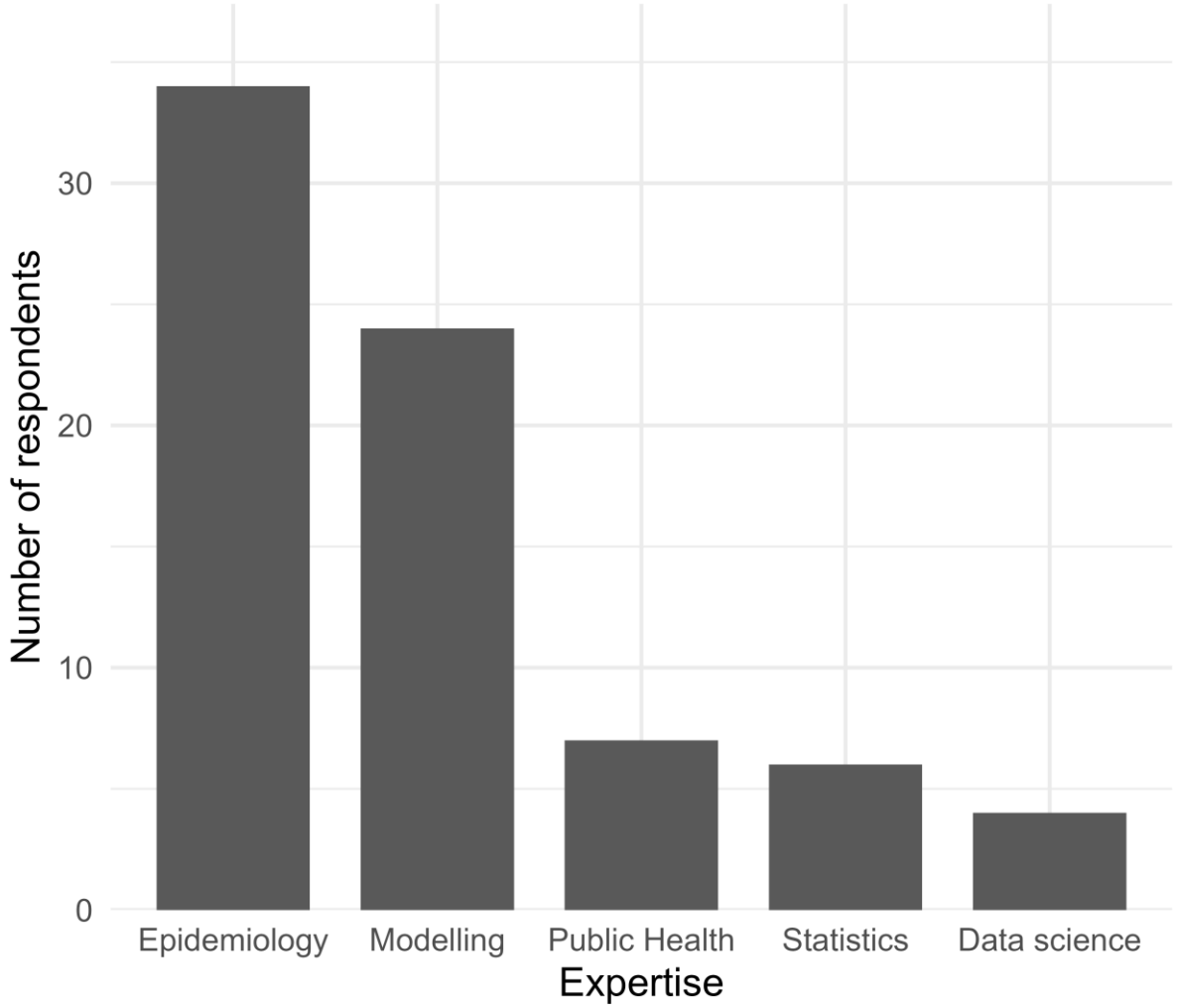
86 respondents from 25 countries and 53 Institutions



University of Melbourne
University of Bern
Public Health Agency of Canada
National Institute for Public Health and the Environment (RIVM)
University of Canterbury
Ministry of Health
McMaster University
Javeriana University
University of Liverpool
Imperial College London
Washington State University
Health Science Institute in Aragon (IACS)
HCMC Center for Disease Control
University of Warwick
PATH
Eli Lilly and Company
Unicef
Coordinating centre for Health Alerts and Emergencies
WHO
Child Health Research Foundation
CDC
IETS
LSHTM
WCO-Nepal
INSERM
Ausvet
INRAE
Swiss TPH
CDPH
Emory
Centers for Disease Control and Prevention
Departamento de Salud, Colombia
University of Georgia
First Nations Health Authority
SENASICA
University of Namibia
Hasselt University
MSF
UNICEF
Pasteur Institute of Tunis
IPH
ISI Foundation
Institute for Health Metrics and Evaluation
Wageningen University
Institute of Tropical Medicine
Lancaster University
University of the Basque Country
Massey University
Oswaldo Cruz Foundation
Public Health Agency of Sweden
University College London
University of California, San Francisco
University of Glasgow

A variety of sectors and expertise





Section 2: Survey Results

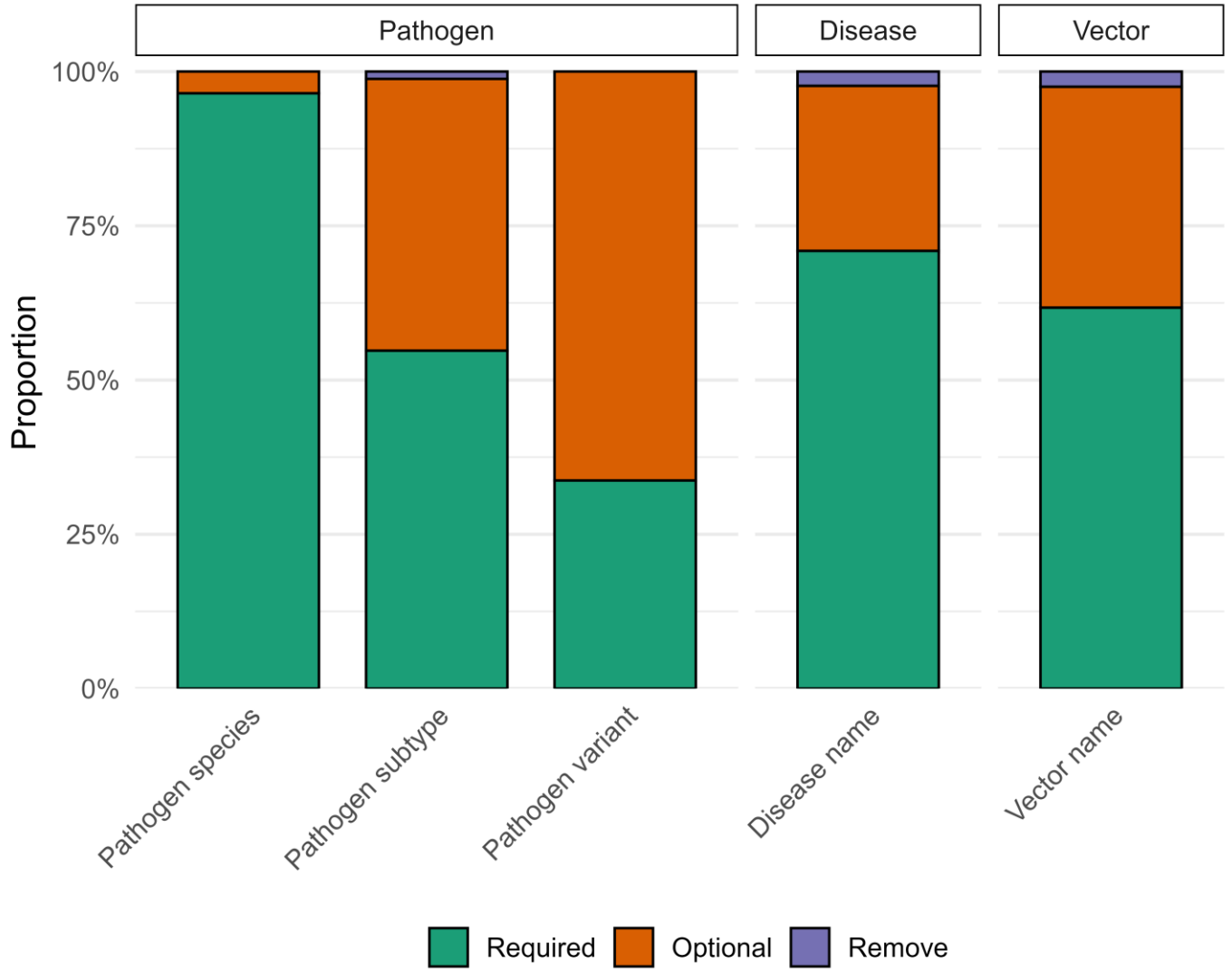
Pathogen, disease and vector designations

Results & Feedback

- Importance of subtypes, variants depends on pathogen (e.g. influenza vs measles)
- Higher-level classifications like pathogen family could be useful (e.g. flavivirus, alphavirus)
- Automated field-completion if pathogen species is provided?

Discussion

- Should pathogen species be selected from a list or free-text?



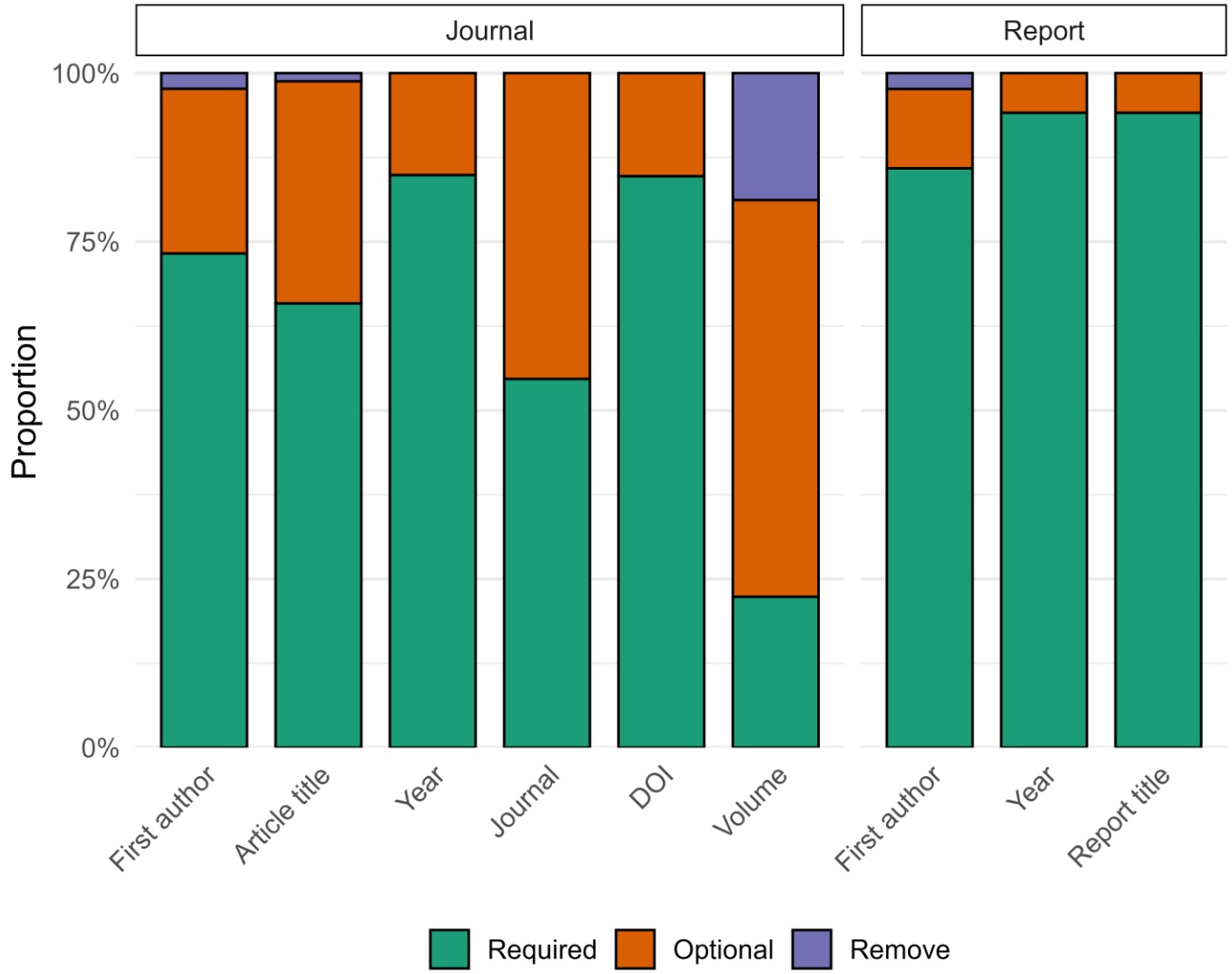
Source

Results & Feedback

- Unique identifier (UID) needed: DOI for article and URL for unpublished report
- Automatically generate citations from UID?

Discussion

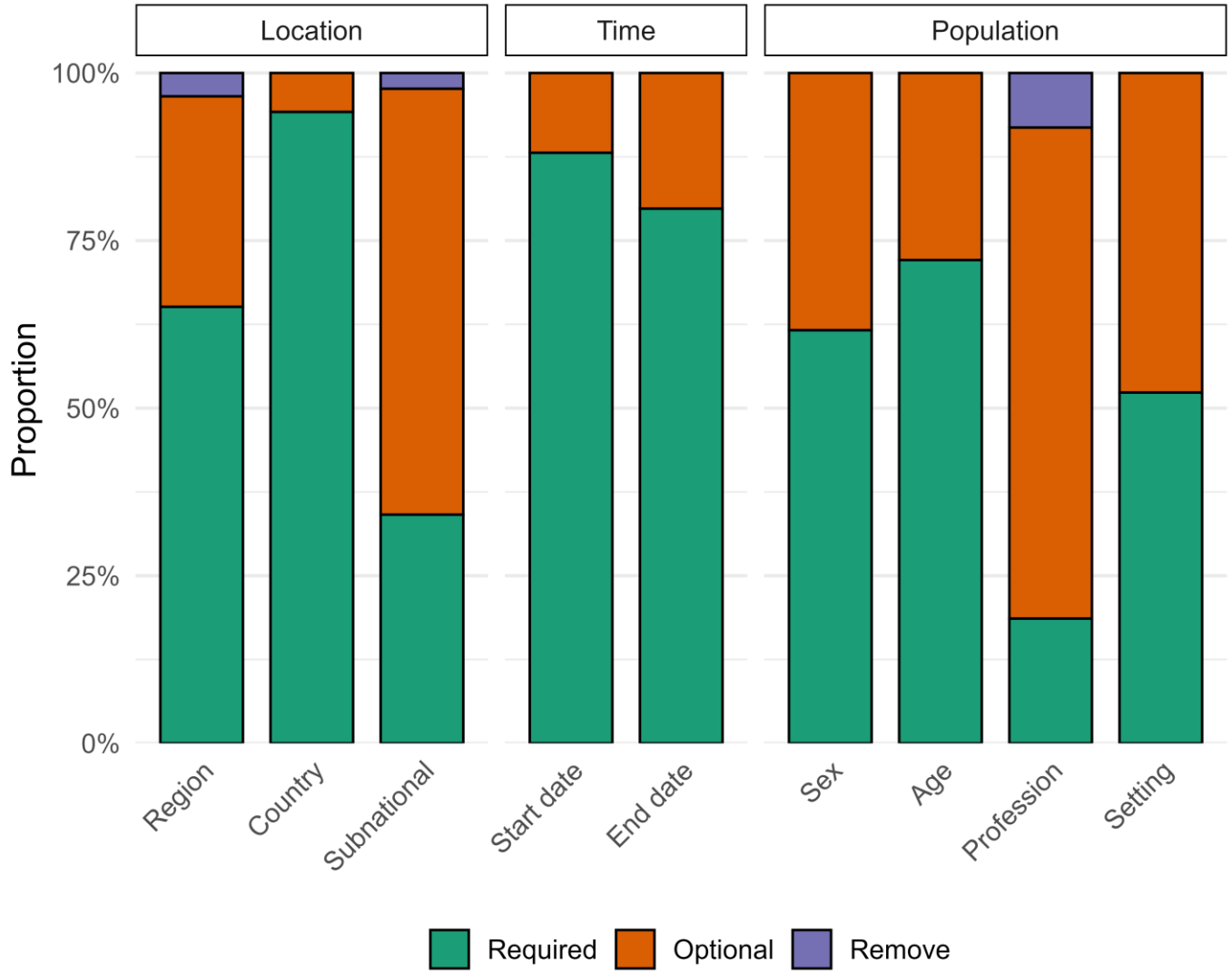
- Should everything besides UID be optional?



Context

Results & Feedback

- Basic location and time indicators considered necessary by most respondents
- Context is important for some parameter types (e.g. transmission rates highly context specific)
- Population data needed for disaggregated parameters
- Large number of optional population covariates that **vary between parameters and studies** and cannot be specified clearly ahead of time



Context

Results & Feedback

- Basic location and time indicators considered necessary by most respondents
- Context is important for some parameter types (e.g. transmission rates highly context specific)
- Population data needed for disaggregated parameters
- Large number of optional population covariates that **vary between parameters and studies** and cannot be specified clearly ahead of time

Discussion

- What does the start date refer to?
- Should population be included? As a free-text field?
- Should “population” be used for characterising representativeness of the study population or for disaggregating parameter estimates?

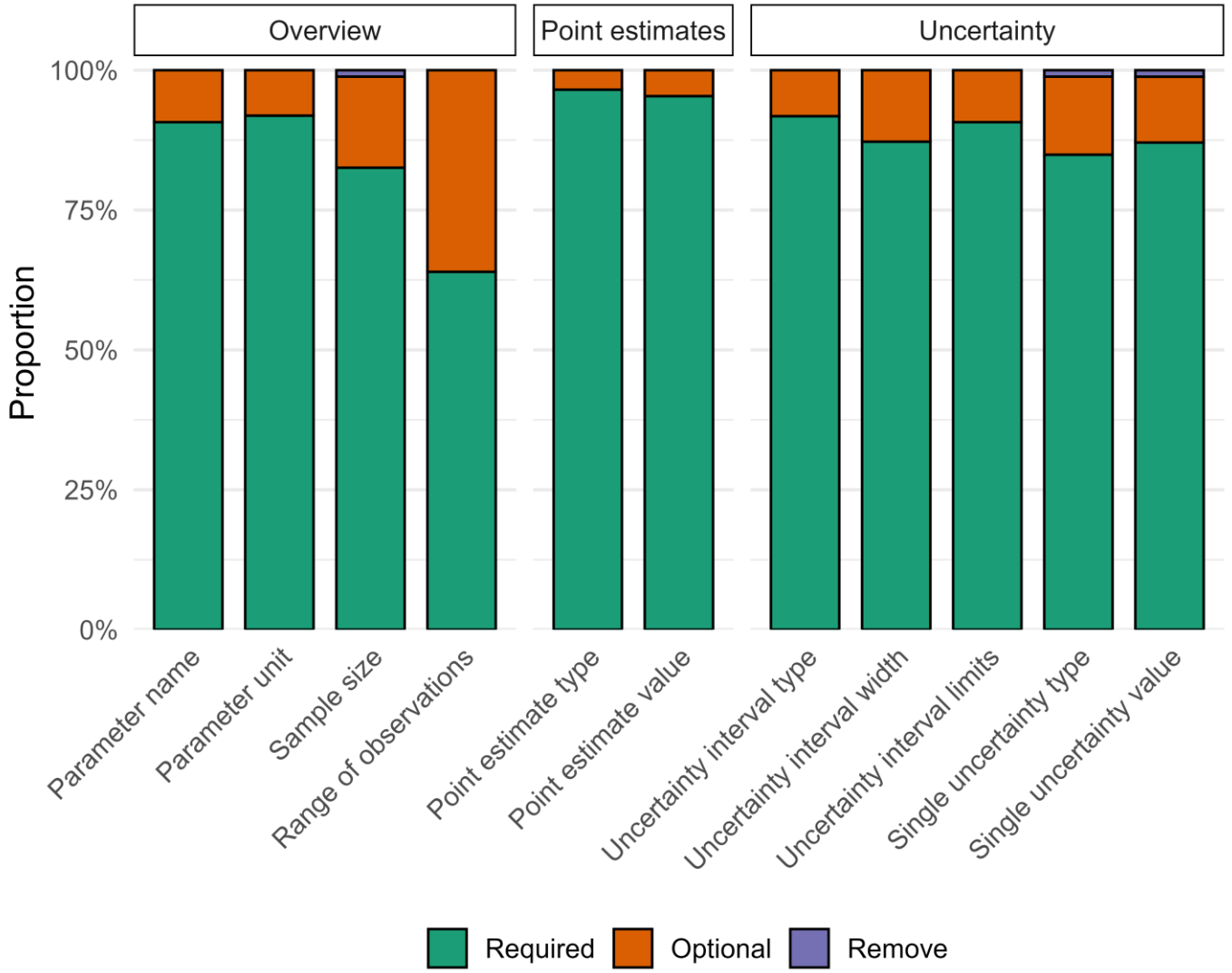
Parameter estimates

Results & Feedback

- Clear parameter definitions needed
- Uncertainty listed as a “required” field by >90% of participants

Discussion

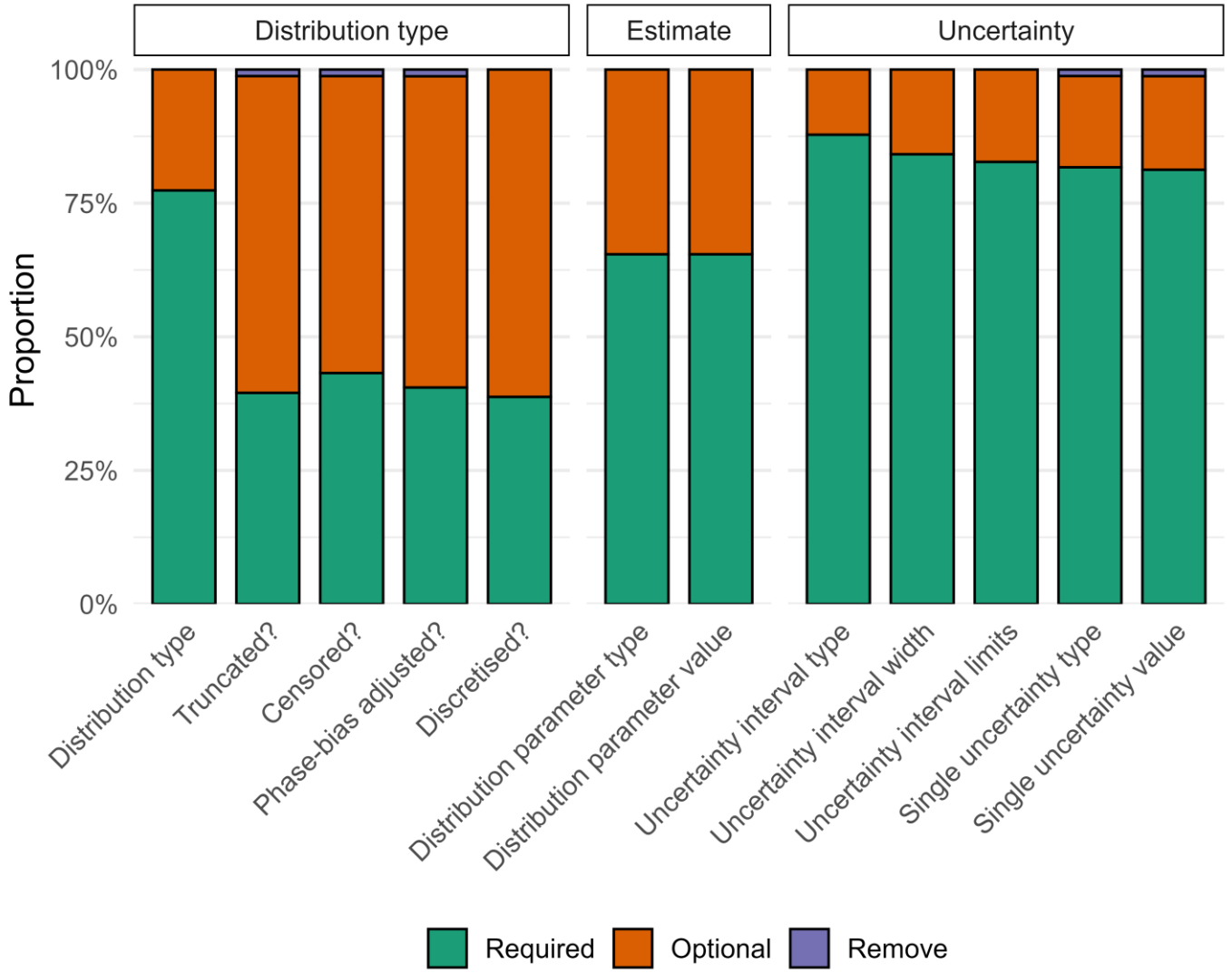
- Should a measure of uncertainty be necessary for submission?
- Can/should parameter names be selected from pre-specified list?



Distribution estimates

Results & Feedback

- At a minimum: **central estimate, variance and distribution type**. Parameters can be calculated from these if necessary.
- How to deal with non-parametric distributions? Provide the empirical distribution and calculate summary statistics?
- Parameterizations of distributions differ a lot. Include a reference list of distributions and parameterizations?



Distribution estimates

Results & Feedback

- At a minimum: **central estimate, variance and distribution type**. Parameters can be calculated from these if necessary.
- How to deal with non-parametric distributions? Provide the empirical distribution and calculate summary statistics?
- Parameterizations of distributions differ a lot. Include a reference list of distributions and parameterizations?

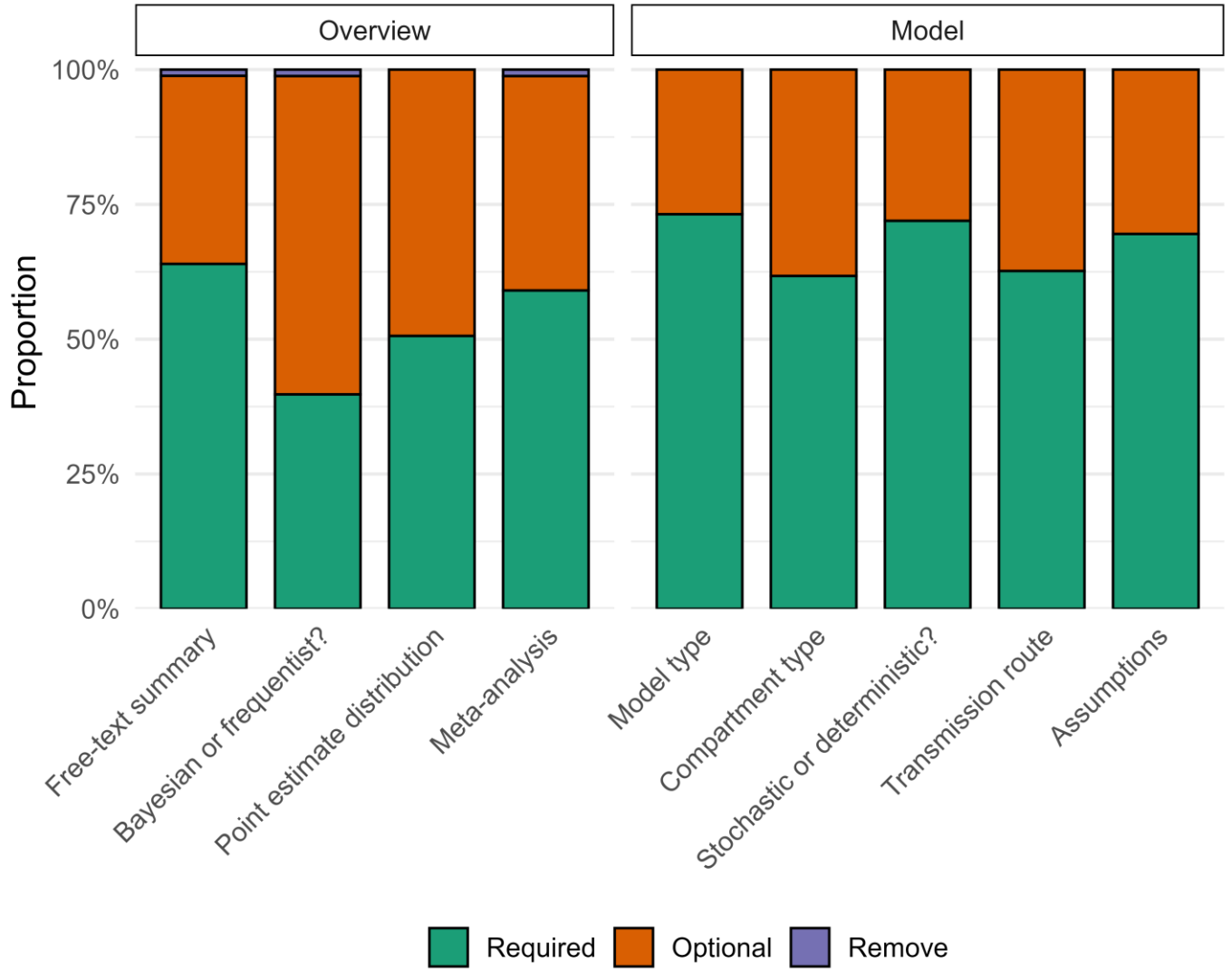
Discussion

- Are central estimate, variance and distribution type a valid minimum submission?
- Is uncertainty around the distribution parameters necessary for submission?
- Can empirical distributions be uploaded to the database or only summary statistics?

Methods

Results & Feedback

- Specify study type (e.g. descriptive, observational study)
- Models have unique features that cannot be specified according to predefined categories: simple categorization may be misleading.
- The ultimate reference must be the original paper. Unclear how useful it would be to collect these details in free-text fields, unless you can decide upon a short list of valid values.
- Meta-analysis requires a lot more information to be extracted. I am not sure meta-analyses have a place in this database.



Methods

Results & Feedback

- Specify study type (e.g. descriptive, observational study)
- Models have unique features that cannot be specified according to predefined categories: simple categorization may be misleading.
- The ultimate reference must be the original paper. Unclear how useful it would be to collect these details in free-text fields, unless you can decide upon a short list of valid values.
- Meta-analysis requires a lot more information to be extracted. I am not sure meta-analyses have a place in this database.

Discussion

- Methods: free-text field, free-text with sections, drop-down list or reference to original paper?
- Should meta-analyses be included in the database?

Section 3: Concluding discussion

Concluding discussion

Results & Feedback

- “Not everything needs to be in the database”
- “Database structure should drive best practice, it’s better to have a small database of high-quality information than a large database that is flooded with submissions that are hard or risky to interpret”
- “What about sensitive data? We should get thinking about people's locations, privacy and equity.”

Discussion

- Emphasis on structured fields, free text or minimal submissions with links to source?
- Emphasis on quality or quantity?

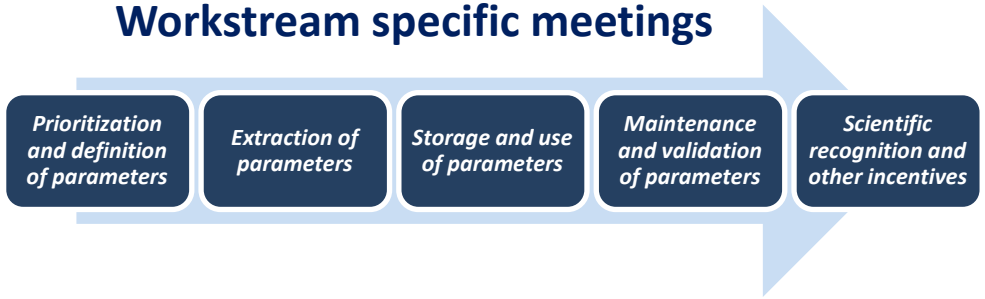
Global repository of epidemiological parameters – Way Forwards

Key Outputs

- GitHub repository and Community Forum
- Standardized Protocol
- Parameters database
- Analytical tools
- ShinyApp for parameter visualization
- Documented lessons learned

Immediate next steps

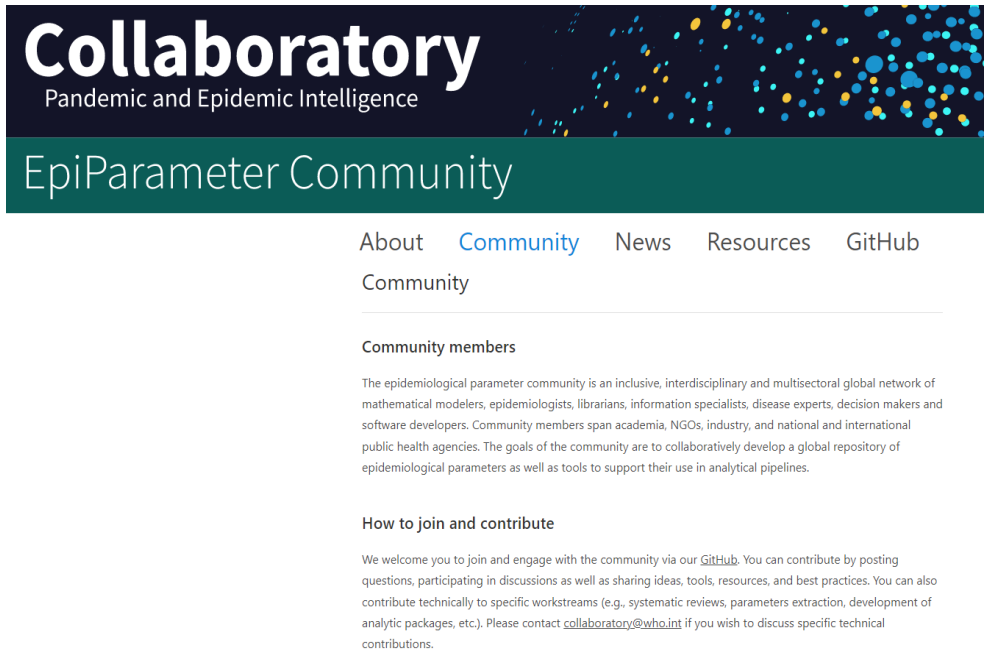
Workstream specific meetings



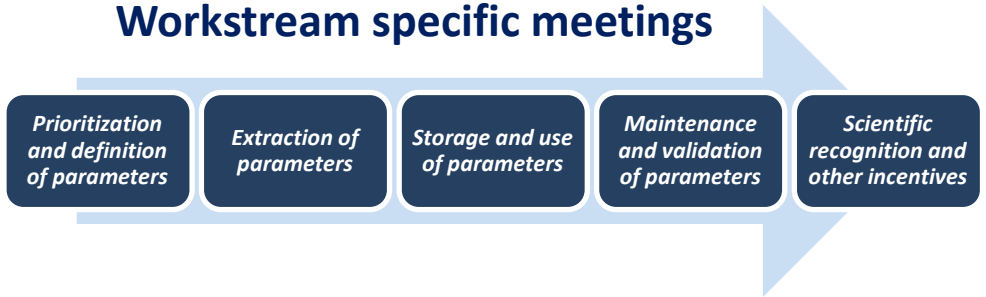
- Data model curation
- Use case exploration
- Quality assurance
- Reporting guidance
- Beta testing
- Etc.

How to join and Contribute

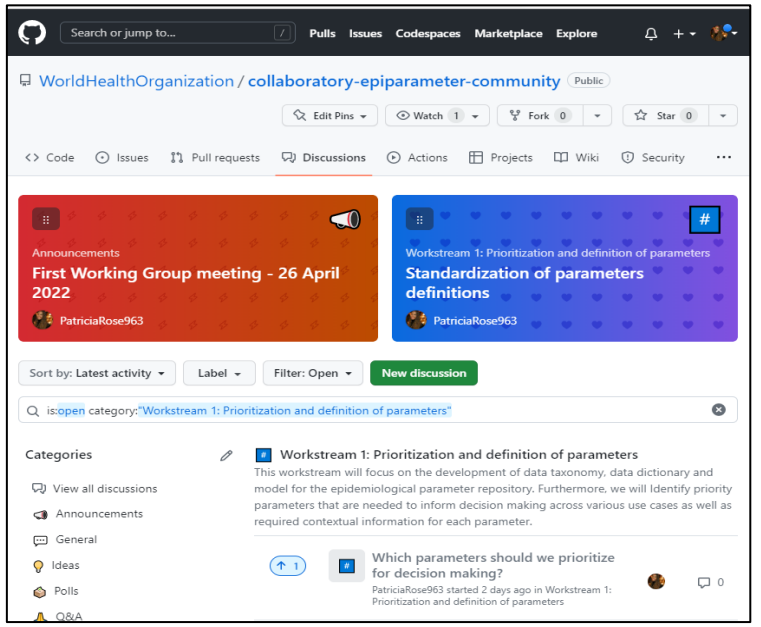
Website



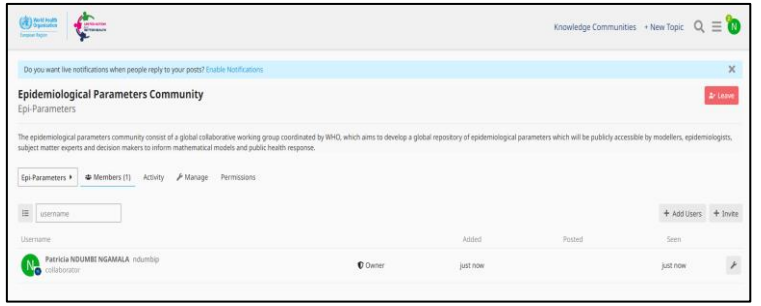
Workstream specific meetings



GitHub



Discourse *(Exploration stage)*



Participant survey

Link in the chat:
https://docs.google.com/forms/d/e/1FAIpQLSeAJQY64PrQV9zpFet9t5Y9oK_MkgMeh6k7x3IoGA4oUtrY6g/viewform?vc=0&c=0&w=1&flr=0