

# Epidemiological Data Source Assessment

This tool helps practitioners rate different data sources on things like quality, timeliness and usefulness for modelling SARS-CoV-2. By pooling expert opinions, we can create visual comparisons showing the trade-offs between different data types. This makes it easier to decide which data to use when estimating transmissibility, and helps identify when different sources might conflict with each other.

Please think about a candidate dataset that you might consider for modelling SARS-CoV-2 and answer the following questions which are grouped into six main categories: basic meta-data, scope, resolution, data quality, data utility, and practical considerations with each then having further subcategories. For each of these subcategories you will assign a value between 0 and 5 (for numeric entries), assign a category, or enter free text.

---

\* Indicates required question

## 1. Email \*

---

### Basic meta-data

What kind of data is it and why was it collected? The essentials about the data source.

## 2. Source type \*

The category/classification of data

*Mark only one oval.*

☐ Confirmed cases time series

☐ Hospitalisation time series

☐ Wastewater

☐ Contact tracing

☐ Other: 

---

## 3. Study design \*

Type of study

*Mark only one oval.*☐ Sentinel surveillance☐ Cohort study☐ Routine surveillance☐ Contact survey☐ Other: \_\_\_\_\_

## 4. Description \*

Brief explanation of what the data contains

---

---

---

---

---

## 5. Primary purpose \*

Original intent of data collection

*Mark only one oval.*☐ Clinical management☐ Other: \_\_\_\_\_

## Scope

Who's included in the data and who isn't? Covers the population represented, any subgroups, and how close it is to actual infection timing.

## 6. The source population \*

Overall population from which the data are drawn, eg hospital catchment areas

---

## 7. Target population \*

Who does the data aim to cover

*Tick all that apply.*

- ☐ Specific age groups
- ☐ Risk-specific
- ☐ Convenience
- ☐ Geographic structure
- ☐ Healthcare workers
- ☐ Whole population
- ☐ Other: \_\_\_\_\_

## 8. Stratification/covariates (except spatial-temporal see Resolution section) \*

*Tick all that apply.*

- ☐ Demographic
- ☐ Clinical
- ☐ None
- ☐ Other: \_\_\_\_\_

## 9. Collection type \*

*Mark only one oval.*

- ☐ Routine
- ☐ Triggered
- ☐ One-off

## 10. If triggered, potential triggers

*Mark only one oval.*

- ☐ Greater than a threshold
- ☐ New variant/pathogen detected
- ☐ Other: \_\_\_\_\_

## Resolution

How detailed is the data? Looks at whether it's individual or aggregated, and how fine-grained the time and location information is.

### 11. Data aggregation \*

*Mark only one oval.*

- ☐ Individual
- ☐ Aggregated

### 12. Temporal data \*

Does your data have a temporal component

*Mark only one oval.*

- ☐ Yes      *Skip to question 13*
- ☐ No      *Skip to question 16*

## Resolution - temporal data

Questions related to temporal data

### 13. Collection frequency

How often data are collected

*Mark only one oval.*

- ☐ Continuously
- ☐ Daily
- ☐ Multiple times a week
- ☐ Weekly
- ☐ Fortnightly
- ☐ Monthly
- ☐ Less frequently than monthly

## 14. Reporting frequency

How often releases/updates occur

*Mark only one oval.*

- ☐ Continuously
- ☐ Daily
- ☐ Multiple times a week
- ☐ Weekly
- ☐ Fortnightly
- ☐ Monthly
- ☐ Less frequently than monthly

## 15. Time period covered

*Tick all that apply.*

- ☐ Early outbreak
- ☐ Peak
- ☐ Endemic
- ☐ Continuous
- ☐ Other: \_\_\_\_\_

## Resolution - spatial data

Questions related to spatial data

## 16. Spatial data \*

*Mark only one oval.*

- ☐ Yes
- ☐ No      *Skip to section 7 (Data quality)*

## 17. Spatial resolution

Lowest spatial resolution across geographical areas

*Mark only one oval.*

- ☐ International
- ☐ National
- ☐ Regional
- ☐ Local
- ☐ Hyper-local

## 18. Geographic coverage

Completeness of target geographical areas

*Mark only one oval.*

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high coverage

## Data quality

How trustworthy is it? Rates things like measurement accuracy, potential biases, and reporting delays.

## Data quality - measurement quality

## 19. Quality of case definitions

How standardised is the case definition? For example does it vary across time or space or both. Leave blank if not applicable.

*Mark only one oval.*

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very variable

## 20. Test sensitivity

Sensitivity of any test used to obtain information on infection. Leave blank if not applicable.

Mark only one oval.

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 21. Test specificity

Specificity of any test used to obtain information on infection. Leave blank if not applicable.

Mark only one oval.

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 22. Potential for unexplained variability \*

Mark only one oval.

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 23. Reporting delay \*

Lag between event occurrence and reporting relative to granularity of time considered

Mark only one oval.

	1	2	3	4	5	
Inst:	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very long

Data quality - sources of bias

## 24. Outages

Potential for missing reports of information

Mark only one oval.

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 25. Censoring \*

Potential for censored data (events are known to have happened but the precise time of events is unknown i.e. to within an interval).

Mark only one oval.

	1	2	3	4	5	
Non	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 26. Truncation \*

Potential for truncated data (events are not known to have occurred but may be reported due to lack of observation such as conditioning on case report for onset data).

Mark only one oval.

	1	2	3	4	5	
Non	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 27. Selection (ie unrepresentative relative to target population), including ascertainment bias/underreporting

\*

Potential for these

Mark only one oval.

	1	2	3	4	5	
Non	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high



## 28. If there is the potential for selection bias what kind

*Tick all that apply.*

- ☐ Ascertainment/underreporting
- ☐ Only more severe events observed
- ☐ Response is age (or other factor) dependent
- ☐ Location specific observation
- ☐ Socio-economic factors
- ☐ Confounding from external factors (i.e. environmental factors)
- ☐ Other: \_\_\_\_\_

## Data quality - bias characteristics

If bias of any kind of present what characteristics does it have?

## 29. Time-varying

Potential for biases to vary over time

*Mark only one oval.*

	1	2	3	4	5	
Non	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 30. Direction of bias

Put 3 if unclear

*Mark only one oval.*

	1	2	3	4	5	
Larg	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Large overestimation

## Data utility

What can you actually use it for? Identifies which transmission metrics it can inform and how directly it provides that information.

31. **Quantities informed \***

What target quantities does the data source inform. Does it inform it directly or indirectly (e.g. wastewater indirectly informs prevalence and community prevalence surveys with random sampling directly inform prevalence estimates).

*Tick all that apply.*

	Direct	Indirect	Not informed
<b>Basic reproduction number</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Time-varying reproduction number</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Time-varying / basic reproduction number ratio</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Incidence</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Prevalence</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Heterogeneity in transmission (e.g. superspreading)</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Drivers of transmission</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

32. **Relationship of target population to general population \***

Can information on the target population be generalised to the general population

*Mark only one oval.*

	1	2	3	4	5	
It ca	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very easily

33. Why can't the information on the target population be generalised to the general population
- 

### Practical considerations

How feasible is it to use? Covers accessibility, sustainability, cost, and whether it plays well with other data sources.

34. Scalability \*

How does the data collection scale with outbreak size

⌵ Dropdown

*Mark only one oval.*

- ☐ independent
- ☐ Sub-linearly
- ☐ Linearly
- ☐ Sub-exponentially
- ☐ Exponentially
- ☐ More than exponentially

35. Sustainability \*

Likelihood of continued collection/effort to collect

*Mark only one oval.*

1   2   3   4   5

---

Very ☐ ☐ ☐ ☐ ☐ Very unlikely

---

## 36. Cost \*

How resource-intensive is the collection?

Mark only one oval.

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very high

## 37. Accessibility \*

How easily data can be obtained. For example due to privacy concerns.

Mark only one oval.

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very easily

## 38. Linkage potential \*

Ability to connect with other datasets

Mark only one oval.

	1	2	3	4	5	
Very	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very easily

## 39. What data sources if any have been linked to?

The category/classification of data

Tick all that apply.

- ☐ Confirmed cases time series
- ☐ Hospitalisation time series
- ☐ Wastewater
- ☐ Contact tracing
- ☐ Other: \_\_\_\_\_

## 40. Data format \*

The level of structure in the data source

Mark only one oval.

1   2   3   4   5

Unstructured ☐ ☐ ☐ ☐ ☐ Very structured

## 41. How generalisable are these findings for other pathogens? \*

Mark only one oval.

1   2   3   4   5

Not generalisable ☐ ☐ ☐ ☐ ☐ Very generalisable

## 42. Which pathogens if any could your answers be generalised to?

---

---

This content is neither created nor endorsed by Google.

Google Forms

