##########
Data dictionary
##########

Description of datasets accompanying publication Hay & Kennedy-Shaffer et al. 2021, PMID: 34083451.

NOTE: all of these data files are already publicly available at J. A. Hay, L. Kennedy-Shaffer, jameshay218/virosolver_paper: Publication release, version v1.0.1, Zenodo (2021); http://doi.org/10.5281/zenodo.4776834.

Open to the hyperlink (https://zenodo.org/record/4776834#.YbOsWvHMJqs). All data are stored in the “data” folder.

##########
Dataset 1
##########
Filename: Data S1.csv
Description: SARS-CoV-2 RT-qPCR nasopharyngeal testing results collected from nursing homes in Massachusetts, USA during early 2020 (April to May).
Variables:

‘location’ — one of four anonymized locations.

`week` — integer indicator of sample collection week relative to study period.

`collection_date` — full date of sample collection as M/D/Y.

`result` — binary indicator if sample was deemed positive (POS) or negative (NEG) for SARS-CoV-2.

`sampleid` — sample indicator, simply used for indexing data.

`RP` — cycle threshold value of the control target (human RP). NA indicates not detected or invalid result.

`N2` — cycle threshold value against N2 gene. NA indicates not detected (Ct < 40) or invalid result.

`N1` — cycle threshold value against N1 gene. NA indicates not detected (Ct < 40) or invalid result.

##########
Dataset 2
##########
Filename: Data S2.csv
Description: SARS-CoV-2 RT-qPCR cycle threshold values from routine inpatient nasopharyngeal swab testing on the Hologic Panther Fusion instrument at the Brigham and Women’s hospital, Boston, Massachusetts, USA. Note that patients tested on this instrument were those tested in hospital for reasons other than suspect COVID-19 diagnosis (ie. Routine screening). Only positive samples (Ct < 40) are included.
Variables:

‘coll_date’ — full date of sample collection as M/D/Y.

‘ORF1ab_Ct` — cycle threshold value against the ORF1ab gene.

##################
Dataset 3
##################
Filename: biobot_data_melted.csv
Description: Publicly available viral load measurements from wastewater RT-qPCR testing in Boston, Massachusetts, USA. NOTE: these data were obtained from https://www.mwra.com/biobot/biobotdata.htm.
Variables:

'date' — full date when measurement was taken as M/D/Y.

'copies` — raw viral load measurement in copies/mL

'avg' — 7-day average of the 'copies' variable.

'lower' — lower confidence interval on the viral load measurement (confidence interval width not specified in original data, assumed 95% CI).

'upper' — upper confidence interval on the viral load measurement (confidence interval width not specified in original data, assumed 95% CI).

'system' — factor whether measurement corresponds to the "Northern" or "Southern" measurement system.

##################
Dataset 4
##################
Filename: borremans_urt_pos.csv
Description: Data from the publication Borremans et al. 2020 eLife 2020;9:e60122 DOI: 10.7554/eLife.60122. These data give the result of a meta-analysis calculating the percentage of individuals who are expected to test RT-qPCR positive for SARS-CoV-2 as a function of days since symptom onset.
Variables:

'time_since_onset' — number of days since symptom onset.

'pos' — percentage of samples testing positive.

'n' — total number of samples collected on this 'time_since_onset'.

'lower_95' — lower 95% confidence interval on 'pos'.

'upper_95' — upper 95% confidence interval on 'pos'.

###################
Dataset 5/6
###################
Filename: us-counties.csv and us-states.csv
Description: COVID-19 case counts over time stratified by US state or county. These data are publicly available at: https://github.com/nytimes/covid-19-data
Variables:

'date' — date of report as M/D/Y.

'county' — name of US county.

'state' — name of US state.

‘fips’ — Federal Information Processing Standards key.
‘cases’ — number of COVID-19 attributed deaths on this date.
‘deaths’ — number of COVID-19 cases recorded on this date.

##########################
Other datasets
##########################
Note that there are a number of simulated datasets included with the publication which are used to generate figures and illustrate simulation analyses. These are not included here as they are fully simulated and reproducible from the code provided alongside the publication, available at:

J. A. Hay, L. Kennedy-Shaffer, jameshay218/virosolver_paper: Publication release, version v1.0.1, Zenodo (2021); http://doi.org/10.5281/zenodo.4776834.