WHO Global Clinical Platform
for COVID-19

Data for public health response

WHO Global Clinical Platform for the Clinical Characterization of COVID-19

Statistical Analysis Plan

7 July 2021

World Health Organization

© World Health Organization 2021.

Some rights reserved. This work is available under the CC BY-NC-SA 3.0 IGO licence.

WHO reference number: WHO/2019-nCoV/Clinical/Analytic_plan/2021.1

Contents

Acknowledgements

Data contributors

Individual facilities contributing clinical data are acknowledged in the following WHO web page: https://www.who.int/docs/default-source/coronaviruse/acknowledgements_covid19clindataplat_web1716bf7d-bb9a-410b-9ba8-56a48ee67ca0.pdf?sfvrsn=7edb0993_8

WHO Secretariat

HQ: Dr Silvia Bertagnolio (Department of Global HIV, STI, Hepatitis Programmes), Dr Janet Diaz (Department of Country Readiness Strengthening, Health Emergencies Programme), Dr Soe Soe Thwin (Department of Sexual and Reproductive Health and Research), Dr Ronaldo Silva (Department of Sexual and Reproductive Health and Research), Madeleine Crowe (Department of Country Readiness Strengthening), Firdavs Kurbonov (Department of Sexual and Reproductive Health and Research), Flaminia Sabbatucci (Department of Country Readiness Strengthening), Dr Ndema Habib (Department of Sexual and Reproductive Health and Research), Dr Teresa Kortz (WHO Health Emergencies Programme)

WHO Regional Offices: Dr John Appiah (WHO Regional Office for Africa), Dr Ludovic Reveiz (WHO Regional Office for the Americas), Dr Chiori Kodama (WHO Regional Office for the Eastern Mediterranean), Dr Dina Pfeifer (WHO Regional Office for Europe), Dr Wijesinghe Pushpa (WHO Regional Office for South-East Asia)

Clinical Advisory Group

Rashan Haniffa, *University College Hospital, United Kingdom*
Robert Fowler, *Sunnybrook Health Sciences Centre, Canada*

Bin Cao, *China-Japan Friendship Hospital, China*
Flavia Machado, *Federal University of São Paulo, Brazil*
Gail Carson, *Nuffield Department of Medicine, United Kingdom*
John Amuasi, *Kwame Nkrumah University of Science and Technology, Ghana*
Lee Wallis, *University of Cape Town, South Africa*
Lindsey Baden, *Harvard Medical School, United States of America*
Lucille Blumberg, *National Institute for Communicable Diseases, South Africa*
Michael Hughes, *Harvard TH Chan School of Public Health, United States of America*
Michael Jacobs, *Royal Free London NHS Foundation Trust, United Kingdom*
Natalia Pshyenshya, *Rostov State Medical University (RSMU), Russian Federation*
Paolo Bonfanti, *Hospital San Gerardo, Monza, Italy*
Pisake Lumbiganon, *Khon Kaen University, Thailand*
Richard Kojan, *ALIMA & University of Kinshasa, Democratic Republic of the Congo*
Roger Paredes, *Departament de Salut, Generalitat de Catalunya, Spain*
Sabue Mulangu, *Institut National de Recherche Biomedical, Democratic Republic of the Congo*
Shabina Ariff, *Department of Pediatrics & Child Health, Ministry of Health, Pakistan*
Tim Uyeki, *Centres for Disease Control and Prevention, United States of America*
Yaseen Arabi, *King Saud University, Saudi Arabia*
Yee Sin Leo, *National Centres of Infectious Diseases, Ministry of Health, Singapore*
Yinzhong Shen, *Fudan University, China*

Abbreviations

ACE	angiotensin-converting enzyme
ALT	alanine aminotransferase
APTR	activated Partial Thromboplastin Time Ratio
aPTT	activated partial thromboplastin time
ARB	angiotensin receptor blocker
ARDS	acute respiratory distress syndrome
AST	aspartate aminotransferase
AVPU	alert, verbal, pain, unresponsive
BiPAP	bi-level positive airway pressure
BMI	body mass index
BP	blood pressure
BUN	blood urea nitrogen
CI	confidence interval
CPAP	continuous positive airway pressure
CRP	c-reactive protein
CT	computed tomography
ECMO	extracorporeal membrane oxygenation
ESR	erythrocyte sedimentation rate
FiO2	fraction of inspired oxygen
GCS	Glasgow Coma Scale
HF	high-flow
HR	heart rate
ICU	intensive care unit
IL-6	interleukin-6
IQR	interquartile range
LDH	lactate dehydrogenase
NCDs	non-communicable diseases
NEWS2	national early warning score 2
NSAID	non-steroidal anti-inflammatory drug
PaCO2	partial pressure of carbon dioxide
PaO2	partial pressure of oxygen
PEEP	positive end-expiratory pressure
P/F ratio	ratio of the partial pressure of oxygen to the fraction of inspired oxygen
RR	respiratory rate
RRT	renal replacement therapy
SBP	systolic blood pressure
SGOT	serum glutamic-oxaloacetic transaminase
SGPT	serum glutamic pyruvic transaminase
SpO2	peripheral oxygenation saturation
WBC	white blood cell
WHO	World Health Organization

Chapter 1. Background and Objectives

Introduction: Concerted epidemiological surveillance strategies are needed to better characterize the clinical presentations of COVID-19 in different demographic groups and in the context of varying management approaches worldwide. At this juncture, it is also critical to gain a more comprehensive understanding of the risk factors portending severe COVID-19 so that appropriate preventative or mitigating strategies may be put into place.

Intended purpose: To gather this information, the World Health Organization (WHO) has devised data collection tools for its Member States and a global COVID-19 clinical platform to enable harmonized data collection system submissions. The purpose of this document is to present a succinct description of the proposed analytic plan to generate statistics at global, regional and national levels including among subpopulation on the different clinical characteristics associated with COVID-19 and risk factors associated with poor clinical outcomes. The reports generated and published from these proposed analyses will help clinicians and national programs prepare appropriate management and response strategies.

Rationale: The World Health Organization has launched a Global COVID-19 Clinical Platform, which is intended to provide Member States with a standardized approach and platform to collect clinical data to better characterize the natural history of the disease, identify risk factors for severe disease and describe treatment interventions. The use of a single standardized clinical data tool enables clinical data from around the world to be aggregated and analyzed to gain a better understanding of the disease, inform the public health response and prepare for large-scale clinical trials. See https://www.who.int/teams/health-care-readiness-clinical-unit/covid-19/data-platform for more information.

Objectives of the analysis

1. Description of clinical characteristics
To describe the demographic features, clinical features, underlying conditions, medications, therapeutic interventions, supportive care, laboratory markers, and clinical outcomes (hereafter, collectively called clinical characteristics) among:
- the general population hospitalized with COVID-19
- specific subpopulations such as children, pregnant women, people living with HIV, people infected with tuberculosis (TB) or malaria, individuals with non-communicable diseases or other underlying conditions, severe/critically ill patients, and those from different geographic settings (hereafter, subgroups).

2. Variations in clinical characteristics
To assess variations in clinical characteristics among and between subgroups, as described above.

3. Association of clinical characteristics with outcomes
To identify clinical characteristics associated with disease severity at admission, ICU admission and in-hospital mortality globally, regionally, and nationally using regression models in both:
- the general population hospitalized with COVID-19
- in subgroups, as described above.

Time-to-event analyses for ventilation, ICU admission, and death will also be conducted.

4. Temporal trends
To describe temporal trends in clinical characteristics.

Please refer to **Table 1** entitled "**Overview of Analytic Schema**" below for more details.

Registry design: The WHO Global Clinical Platform is an open platform where Members States and individual facilities are invited to contribute anonymized patient data. Data contributors include a convenience sample of facilities willing to contribute data to the WHO Global Clinical Platform for COVID-19, research networks including multiple facilities and national registries. Increasing attempts will be made to include a representative sample of clinics.

Study population: Anonymized patients hospitalized with clinically suspected or laboratory-confirmed COVID-19.

Participant inclusion criteria: All patients, regardless of age, admitted to a hospital or health facility with laboratory-confirmed or clinically suspected COVID-19, will be included in the analytic sample. Patients with a negative laboratory result for SARS-CoV2 will be excluded.

Contribution to the Data Platform: All Member States are invited by WHO to participate and contribute clinical data through official communication from WHO Regional Offices. The extent of data contribution and data representativeness was expected to vary greatly among countries. In countries where funding is available to support data collection and data entry, a large network of clinics (up to 50 clinics per country) was trained to contribute data to the WHO Global Clinical Platform. In countries where a national registry of clinical data from hospitalized cases was established, the data shared with WHO is a population of hospitalized patient census in the country. In other countries, data contribution is limited to conveniently selected hospitals. Through regular review of the literature, authors of studies of clinical characterization of COVID-19 are invited to contribute data.

Main parameters/endpoints: Primary descriptive parameters include demographics (age, gender), presence of underlying conditions, use of chronic medications, clinical features on admission and during the hospitalization, laboratory findings on admission and during hospitalization, clinical interventions on admission and during hospitalization (oxygen use, ventilator use, use of therapeutics) and patient outcomes (dead, discharged, referral).

The patient outcomes described above will be used for secondary analysis to determine associations between baseline characteristics and severity of disease and outcomes.

Standardized data collection tool: Case report forms can be found at https://www.who.int/teams/health-care-readiness-clinical-unit/covid-19/data-platform

Table 1. Overview of Analytic Schema

Objective 1: Description of Clinical Characteristics	**Objective 2:** Variations of Clinical characteristics	**Objective 3:** Association of clinical characteristics with outcomes	**Objective 4:** Temporal trends
Descriptive Analysis of variables 1. Demographics 2. Clinical features 3. Underlying conditions and co-infections 4. Medications received 5. Therapeutic interventions 6. Supportive care 7. Laboratory markers 8. Clinical outcomes Stratified by: 1. Age 2. Gender 3. Severity of illness at admission To be applied to subpopulations (see sample size section).	**Bivariate analysis** to assess differences among and between subpopulations 1. Pregnant women compared to non-pregnant women 2. Children compared to adults 3. HIV+ compared to HIV - 4. TB co-infected patients compared to non-TB patients 5. Malaria co-infected pts compared to non-malaria patients 6. Other subpopulations, including people with NCDs or other underlying conditions, or co-infections will be considered 7. Populations from different geographic areas	**Regression analysis** to estimate odds ratio/relative risk or hazards ratio for clinical outcomes Specified outcomes include: 1. Disease severity at hospital admission 2. Mortality 3. ICU admission To be applied to subpopulations (see sample size section).	**Time-series analysis** to assess temporal trends in clinical characterization and management

Chapter 2. Sample Size

Study design

This study design may be described as passive clinical surveillance. It was pre-determined that the minimal sample size to conduct descriptive analysis per country was 300 patients. For regional reports, the minimal sample size required was at least 1200 patients from four countries (300 patients in each country). For global reports, the minimal sample size was at least 7200 patients derived from at least four countries in each of the 6 WHO regions

Sample size to assess associations between baseline characteristics and outcomes

The sample sizes to assess associations between baseline characteristics and severity of disease and outcomes will be calculated as detailed in **Annex 1**. For example, for three different baseline event rates, **Figure 1**. displays the total sample needed (on the y-axis) to detect an odds ratio (on the x-axis) at p=0.05 significance level with 90% statistical power using a two-sided, two-sample (independent) test of proportions with equal samples in each group. Detailed sample size tables are provided following the graph. The sample size needed to estimate risk of mortality will be calculated as detailed in **Annex 1**. See **Figure 2** and **Table 4** for an example.

Chapter 3. Definitions

Pre-specified subgroups:

- Age in completed years
    - o <18 years old
        - ◦ 0-4 years old
        - ◦ 5-13 years old
        - ◦ 14-17 years old
    - o 18 - 45 years old
    - o 46 - 65 years old
    - o 66 -75 years old
    - o >75 years old
- Sex
- Subpopulations include children, pregnant women, people living with HIV, patients with TB and individuals with other coinfections
- Other subpopulations (e.g., individuals with NCDs or other underlying conditions or infections; populations from different geographic areas) may be considered.
- Classification of COVID-19 severity (mild/moderate vs severe/critical)

Cases were defined as severe or critical if they met one or more of the following conditions at hospital admission:

- SpO2: <90%
- Respiratory rate >30 breaths/minute in adults and children over 5 years old, ≥ 60 breaths/minute in children under 2 months old, ≥ 50 in children 2-11 months old, and ≥ 40 in children 1-5 years old
- Received extracorporeal membrane oxygenation (ECMO)
- Admitted to an Intensive Care Unit (ICU)
- Received an inotrope or vasopressor
- Received oxygen therapy, either invasive or non-invasive ventilation

Cases not meeting all the conditions described above, and those meeting the conditions below were considered as mild or moderate:

- SpO2: ≥90% without supplemental oxygen
- Respiratory rate: ≤30 breaths/minute in adults and children over 5 years old, < 60 breaths/minute in children under 2 months old, < 50 in children 2-11 months old, and < 40 in children 1-5 years old
- Did not receive oxygen therapy, either invasive or non-invasive ventilation

See **Table 2** for descriptions of demographic and clinical characteristics.

Chapter 4. Statistical Considerations

General considerations: All analyses will be conducted in SAS version 9.4 (Copyright (c) 2016 by SAS Institute Inc., Cary, North Carolina, United States of America) or R version 3.6.3 (R: A Language and Environment for Statistical Computing, R Core Team, R Foundation for Statistical Computing, Vienna, Austria 2020, *https://www.R-project.org*). Maps will be generated as deemed necessary using ArcGIS Pro Release 2.5.0. (Environmental Systems Research Institute (ESRI), 2020. Redlands, CA.)

Missing data: For each analysis, the denominator will represent data that is available. Hospital admission date will be considered essential and critical to ascertain the hospitalized population to be analyzed. Records with missing admit dates will be excluded from the analysis unless there is confirmation from data contributors that data refer to a cohort of hospitalized patients. It is understood that some data variables could be missing in entire data sets (if collected retrospectively) or in some individual cases. Due to this potential heterogeneity, imputation will not be done but may be considered on a case-by-case basis.

Descriptive statistics: According to the variable type, descriptive statistics will be:

- Quantitative criteria: number of observations (N), mean, standard deviation (SD), median and inter-quartile range (IQR), as applicable.
- Qualitative criteria: number of observations (N), absolute frequency (n), and relative frequency (%).
- Percentages will be calculated on the number of participants with documented data.

Comparison: Where appropriate and if data are available, country-specific distributions may be compared to WHO global surveillance data.

Limitations: If more than 80% of the contributions come from one country, then limitations (namely, the lack of generalizability of the findings to external populations) of such a sample will be clearly described in the report.

Chapter 5. Description of Analytic Objectives

First objective: description of clinical characteristics

Primary descriptions of the population and prespecified subgroups will be conducted for demographic data, signs/symptoms and laboratory features on admission, presence of chronic diseases, treatments received, and outcomes. As the case report form (CRF) is updated, the list of variables below will be updated accordingly.

Table 2. Description of Clinical Characteristics

Variable	Statistical methods
Age in completed years <18 years old • 0-4 years old • 5-13 years old • 14-17 years old 18 - 45 years old 46 - 65 years old 66 -75 years old >75 years old	*Median (IQR) or mean (SD), categorical
Gender Male, female If female, pregnancy status	n/N (%)
Occupational exposures Healthcare worker, laboratory worker	n/N (%)
Admission vital signs and Anthropometrics	
Temperature (continuous and categorical >37.5C)	*Median (IQR), mean (SD), n/N (%)
Heart rate (continuous and categorical >100 p/ min in adults and adolescent; > 120 in children up to 12 years)	*Median (IQR), mean (SD), n/N (%)
Respiratory rate (continuous and categorical >20 b/min in adolescent and adults, and > 40 in children up to 12 years)	*Median (IQR), mean (SD), n/N (%)
BP (continuous and categorical systolic >140 mmHg in adults)	*Median (IQR), mean (SD), n/N (%)
Oxygen saturation (on room air or oxygen) (continuous and categorical: <90%, <94%)	*Median (IQR), mean (SD), n/N (%)
Severe dehydration	n/N (%)
Capillary refill (sternal)	n/N (%)
Glasgow Coma Scale (GCS)	*Median (IQR), mean (SD)
Mental status (AVPU)	n/N (%)
Mid-upper arm circumference	*Median (IQR), mean (SD)
Height (cm), Weight (kg)	*Median (IQR), mean (SD)
BMI (continuous and categorical >30 and > 40)	*Median (IQR), mean (SD), n/N (%)
Severity classification	n/N (%)

Chronic conditions None, chronic cardiac disease, hypertension, chronic pulmonary disease, asthma, chronic liver disease, chronic kidney disease, chronic neurological disease, HIV (on ART, not on ART), diabetes, current smoking, tuberculosis (active and previous), asplenia, malignant neoplasm, obesity, autoimmune diseases, current smoking, other (free text field)	n/N (%)
Pre-admission or chronic medications (within 14 days) ACE inhibitors, ARBs, NSAIDs, chloroquine/ hydroxychloroquine, azithromycin, lopinavir-ritonavir, other	n/N (%)
Signs and symptoms on admission and during hospitalization History of fever (>38ºC), cough, shortness of breath OR tachypnea, sore throat, running nose, wheezing, chest pain, loss of taste, loss of smell, headache, stroke (ischemic stroke vs. intracerebral haemorrhage), vomiting/nausea, diarrhea, fatigue/malaise, altered consciousness, seizures, abdominal pain, conjunctivitis, skin rash, skin ulcers, lymphadenopathy, inability to walk, bleeding (specify site)	n/N (%)
Laboratory tests on admission and during hospitalization Blood indices: Haemoglobin, WBC count, Haematocrit, Platelets, APTT/APTR, PT, INR Metabolic indices: Sodium, Potassium, Creatinine, Urea (BUN),Total bilirubin, LDH, ALT/ SGPT, AST/SGOT, Creatine kinase Inflammatory markers: Procalcitonin, CRP, Troponin, Lactate, ESR, D-dimer, Ferritin, IL-6	*Median (IQR), mean (SD), n/N (%)
Diagnostic testing Chest x-ray/CT if yes, presence of infiltrates	n/N (%)
Treatments on admission and during hospitalization Antiviral (specify), corticosteroid (route, agent, max dose), antibiotic (specify), antimalarial, (specify), antifungal (specify), systemic anticoagulation, experimental agent (specify), oral/orogastric fluids, IV fluids, NSAID, ACE inhibitor, ARBs	n/N (%), trend test will be conducted in keeping with the dashboard output

Pathogen testing at any time during hospitalization Influenza (if positive, type), coronavirus (if positive, type), other respiratory pathogens, viral hemorrhagic fever (specify), other pathogens of public health interest (specify), falciparum malaria, non-falciparum malaria, HIV, Variants of SARS-CoV-2, bacterial and fungal infections (specify)	n/N (%)
Vaccination status	n/N (%)
Reinfection	n/N (%)
ICU admission	n/N (%)
Respiratory and critical care interventions Renal replacement therapy/hemodialysis Inotropes/Vasopressors Extracorporeal membrane support (ECMO) Prone position Oxygen therapy Flow (1-5, 6-10, 11-15 and >15 L/m) Source of oxygen: Piped, Cylinder or Concentrator Interface: Mask, Nasal prongs; Mask with reservoir, high flow nasal cannula, NIV or CPAP mask Non-invasive ventilation: BiPAP or CPAP Invasive ventilation ECMO	n/N (%)
Hospital outcomes Discharge status (alive, still hospitalized, transferred, death, palliative discharge, unknown), ability to care for self at discharge (same, better, unknown)	n/N (%)
Complications any time during hospitalization Shock, seizure, meningitis/encephalitis, anemia, cardiac arrhythmia, cardiac arrest, pneumonia, bronchiolitis, ARDS, bacteremia, bleeding, endocarditis, myocarditis/pericarditis, acute renal injury, pancreatitis, liver dysfunction, cardiomyopathy, stroke (ischemic, hemorrhagic), other	n/N (%)
Time to event outcomes Symptom onset (day of disease at the time of admission): days between first/earliest symptom and admission Length of hospital stay Length of ICU stay Days from hospital admission to ICU/HDU admission Days from hospital admission until transfer or death (reported as total, for survivors, and non-survivors)	*Median (IQR), mean (sd)

* Note that the tests used for descriptive statistics (parametric vs. non-parametric) will be determined by the normality of distribution for continuous variables.

Second objective: variations on clinical characteristics

Comparison of the distribution of clinical characteristics (patient demographics, clinical features, therapeutic interventions, lab markers, underlying conditions, co-infections (based on the pathogen list), medications, and clinical outcomes) within and across pre-specified subgroups.

- The statistical tests of comparison across subgroups will include t-test and ANOVA parametric tests for normally distributed continuous variables, and the Wilcoxon rank-sum and Kruskal-Wallis non-parametric tests in case of skewed (non-normally) distributed continuous variables. For categorical variables, the Chi-square test will be used, with Fisher's Exact test applied in case of variables with small cell count.
- A p-value <0.05 will be statistically significant and 95% confidence intervals (95% CI) will be reported.

Third objective: clinical characteristics associated with outcomes

Bivariate analysis: will be conducted to determine predictor variables that are reported in >80% of anonymized cases and their association with the primary outcome of death, need for invasive mechanical ventilation, or disease severity. Predictors considered a priori of clinical importance (age, gender, disease severity at admission) will be considered irrespective of bivariate analysis. For all others, only predictor variables with a p-value <0.10 on bivariate analyses and not highly correlated with other variables, using a correlation matrix threshold of a >0.8, will be considered for addition to the model as covariates/predictors.

Multivariate regression: We will then conduct a log-binomial generalized estimating equation (GEE)/ generalized linear Model (GLM) multivariable regression (primary outcome of death and need for invasive mechanical ventilation) to determine independent risk factors for mortality[1,2]. The comparison of the risk of dying between different levels of predictors will be conducted by calculating risk ratios (RRs) and 95% confidence intervals. Both unadjusted and adjusted RRs will be reported. Predictors will be considered as potential confounders and effect modifiers and tested for interaction when appropriate. The GEE/GLM model will account for clustering due to the correlation of data belonging to different countries and sites. Sensitivity analysis can then be done by excluding high contributing countries to assess the stability of estimates.

Time-to-event analysis: We will perform a series of time-to-event analyses using proportional hazards regression modeling[3]. The primary outcomes include death, mechanical ventilation, and ICU admission. Included predictors must be measured in >80% of subjects; statistically significant in this cohort at a p-value <0.10 on bivariate analyses or considered to be a priori of clinical importance; and not highly correlated with other variables using a correlation matrix threshold of >0.8. If <80% reported, then restricted set sensitivity analysis may be considered to minimize selection bias. Only predictors measured before the outcome will be included in the respective model. We will compare models with likelihood ratio testing, and test for interaction between predictors when appropriate as detailed below. We will censor unknown outcomes after the last observed time and then conduct a Wald test to determine if there was a trend in survival across the severity of illness. Finally, we will assess the proportional hazard assumption when comparing subgroups of interest, using visual inspection of a smoothed hazard ratio (smoothed scaled Schoenfeld residual plots) and a plot of the log cumulative hazard vs. the log time plot, as well as the Schoenfeld test for non-proportional hazards.

[1] Liang, K.-Y.; and Zeger, S. L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models." Biometrika 73:13–22.
[2] Fitzmaurice, G.M.; Laird, N.M.; Ware, J.H. (2011). "Applied Longitudinal Analysis, 2nd Edition". New York: John Wiley & Sons.
[3] Lee, E.T. (1992). "Statistical Methods for Survival Data Analysis, 2nd Edition". New York: John Wiley & Sons.

Methods for stepwise PH regression modelling:[3]

1. The values of -2Log Likelihood (-2LogL) between the univariate model and the null (baseline hazard) model will be compared to determine which variables significantly reduce the value of -2LogL.
2. All factors in (1) which are important by bivariate analysis (p-value of <0.10) will be fitted together in a full model.
3. Only the factors that when omitted from the model in (2) lead to a significant increase in -2LogL ($p<0.10$) are retained in the model.
4. Finally, variables not important on their own (in (1) vs (2) will be added into the model to determine whether they can be important in the presence of the other. This is done until when no factor added or omitted from the model does not significantly change the value of -2LogL.
5. Interaction effects between levels of the important variables will also be assessed.
6. Adjusted and unadjusted hazard ratios (from univariate and the final multivariate models) and 95% CI will be reported.

Fourth objective: temporal trends

Longitudinal data analysis: This will be conducted on time-series panel data to explore the change in event rates over time by modeling the relationship between the proportion of characteristics and observation time.

Moment-based method of generalized estimation equations, a statistical methodology developed for the analysis of repeated observations, will be used to estimate population-averaged effects where observation time is considered as a discreet variable (in months) and included in the model as a covariate. To adjust for correlation from non-independent repeated observations from the same facility (or country or region), generalized linear regression models will be fitted specifying Poisson distribution with log-link and considering as correlation structure both exchangeable (if this correlation is not expected to vary by time) and auto-regressive (for correlations that vary over time).

Furthermore, indicators for partitioned time intervals (before or post-vaccination for example) could also be included in the model to explore the association of change in rates with the partitioning event. Segmented regression time series would be the methodology for this exploratory analysis.

For all models fitted in the proposed longitudinal framework, there is potential for bias from missing observations that may or may not be at random, and from measurements observed over unequal time intervals if the time interval is associated with both the explanatory variable and outcome. Methodologies including different correlation structures will be considered to minimize potential bias.

[3] Lee, E.T. (1992). “Statistical Methods for Survival Data Analysis, 2nd Edition”. New York: John Wiley & Sons.

Chapter 6: Publication Plan

WHO reports will be made available on the WHO website on approval from the contributors, WHO country offices, and respective ministries of health, following standard WHO practice. All sites/ data contributors have full access to their dataset and are given instructions and training on setting up R-Studio and R-code to generate descriptive reports.

WHO Country report: the contribution of new datasets from countries will trigger the update of the country report, which will be shared with ministries of health.

WHO Regional report: to be conducted when there is sufficient country representation based on sample size calculations, as described in Chapter 2. The contribution of new datasets from countries will trigger the update of the regional report.

WHO Global report: to be conducted when there is sufficient regional representation based on sample size calculations, as described in Chapter 2. The contribution of new datasets from countries/regions will trigger the update of the global report.

WHO Global report by subpopulations: to be conducted when there is sufficient representation. The contribution of new datasets from countries will trigger the update of the global report.

Dashboard: dynamic metric showing COVID-19 clinical characterization and use of therapeutics and other interventions will be created.

Peer-reviewed publications: reports generated from the above-mentioned analysis in certain instances may be published in peer-reviewed journals using standard authorship criteria. The SAP for these publications may be reviewed in discussion with data contributors.

WHO guidelines: the reports will also serve as an important information source to inform COVID-19 guideline development at WHO.

Annex: Sample size calculation

For Odds Ratios

Fig. 1. displays the total sample needed for three different baseline event rates (on the y-axis) to detect an Odds Ratio (on the x-axis) at p=0.05 significance level with 90% statistical power using a two-sided, two-sample (independent) test of proportions with equal samples in each group. Detailed sample size tables are provided following the graph. See **Table 3** for sample size estimation for a given Odds Ratio.

Hypothetical Example (indicated by red arrow in **Figure 1** and in red text in **Table 3**), if ICU admission was 10% in the <60-year age group, a sample of 2000 (1000 in each age group) would be needed to detect an Odds Ratio of 1.60 at p=0.5 significance level, with 90% statistical power.

Fig. 1. Examples of sample size calculation for Odds Ratios

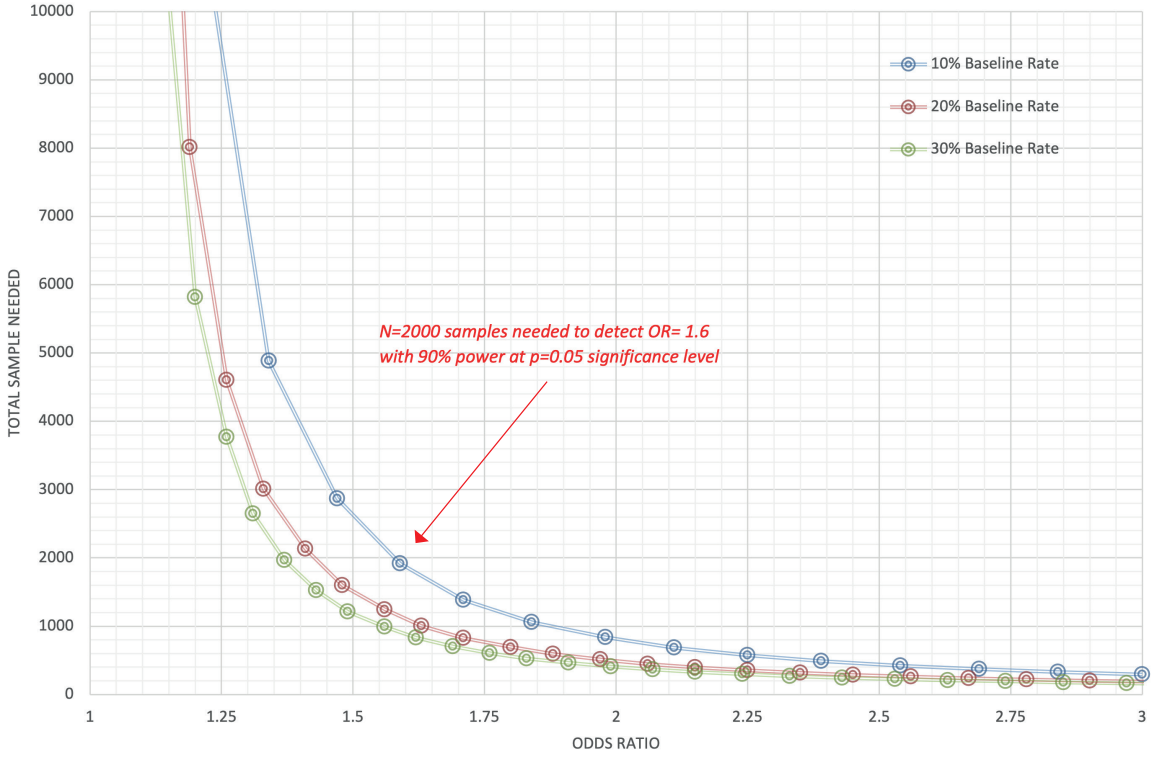

Table 3. Total sample needed for given Odds Ratios

% in Base Group	% in Test Group	Alpha	Power	Odds Ratio	Total Sample Needed
0.1	0.11	0.05	0.9	1.11	39894
0.1	0.12	0.05	0.9	1.23	10484
0.1	0.13	0.05	0.9	1.34	4884
0.1	0.14	0.05	0.9	1.47	2870
0.1	0.15	0.05	0.9	1.59	1916
0.1	0.16	0.05	0.9	1.71	1384
0.1	0.17	0.05	0.9	1.84	1056
0.1	0.18	0.05	0.9	1.98	838
0.1	0.19	0.05	0.9	2.11	684
0.1	0.2	0.05	0.9	2.25	574
0.1	0.21	0.05	0.9	2.39	488
0.1	0.22	0.05	0.9	2.54	422
0.1	0.23	0.05	0.9	2.69	370
0.1	0.24	0.05	0.9	2.84	328
0.1	0.25	0.05	0.9	3	294
0.2	0.21	0.05	0.9	1.06	68896
0.2	0.22	0.05	0.9	1.13	17630
0.2	0.23	0.05	0.9	1.19	8012
0.2	0.24	0.05	0.9	1.26	4604
0.2	0.25	0.05	0.9	1.33	3008
0.2	0.26	0.05	0.9	1.41	2132
0.2	0.27	0.05	0.9	1.48	1596
0.2	0.28	0.05	0.9	1.56	1246
0.2	0.29	0.05	0.9	1.63	1002
0.2	0.3	0.05	0.9	1.71	826
0.2	0.31	0.05	0.9	1.8	694
0.2	0.32	0.05	0.9	1.88	592
0.2	0.33	0.05	0.9	1.97	512
0.2	0.34	0.05	0.9	2.06	448
0.2	0.35	0.05	0.9	2.15	396
0.2	0.36	0.05	0.9	2.25	354
0.2	0.37	0.05	0.9	2.35	318
0.2	0.38	0.05	0.9	2.45	286
0.2	0.39	0.05	0.9	2.56	260
0.2	0.4	0.05	0.9	2.67	238
0.2	0.41	0.05	0.9	2.78	218
0.2	0.42	0.05	0.9	2.9	202
0.2	0.43	0.05	0.9	3.02	186
0.3	0.31	0.05	0.9	1.05	89490
0.3	0.32	0.05	0.9	1.1	22672
0.3	0.33	0.05	0.9	1.15	10208
0.3	0.34	0.05	0.9	1.2	5814
0.3	0.35	0.05	0.9	1.26	3766
0.3	0.36	0.05	0.9	1.31	2646
0.3	0.37	0.05	0.9	1.37	1966
0.3	0.38	0.05	0.9	1.43	1522
0.3	0.39	0.05	0.9	1.49	1214
0.3	0.4	0.05	0.9	1.56	994
0.3	0.41	0.05	0.9	1.62	830
0.3	0.42	0.05	0.9	1.69	704
0.3	0.43	0.05	0.9	1.76	604
0.3	0.44	0.05	0.9	1.83	526
0.3	0.45	0.05	0.9	1.91	462
0.3	0.46	0.05	0.9	1.99	410
0.3	0.47	0.05	0.9	2.07	366
0.3	0.48	0.05	0.9	2.15	328
0.3	0.49	0.05	0.9	2.24	296
0.3	0.5	0.05	0.9	2.33	270
0.3	0.51	0.05	0.9	2.43	246
0.3	0.52	0.05	0.9	2.53	226
0.3	0.53	0.05	0.9	2.63	208
0.3	0.54	0.05	0.9	2.74	192
0.3	0.55	0.05	0.9	2.85	178
0.3	0.56	0.05	0.9	2.97	166
0.3	0.57	0.05	0.9	3.09	154

For Hazard Ratios

Fig. 2. below displays on the y-axis the statistical power for the log-rank test to compare two groups of equal size and detect a significant risk ratio on the x-axis, (for example the ratio of hazards at a fixed time point), at p=0.05 level, for three different events in the test group. See **Table 4** for sample size estimation for a given Hazard Ratio.

Hypothetical Example (indicated by the red arrow in Figure 2 and in red text in Table 4)): To estimate the increased risk of mortality related to severity of disease at admission, a total of 159 deaths would need to be observed (100 in the Severe group and 59 in Mild group) to detect with 90% power a Hazard Ratio of 1.70 for mortality that is significant at p=0.5 level.

Fig. 2. Examples of sample size calculation for Hazard Ratios

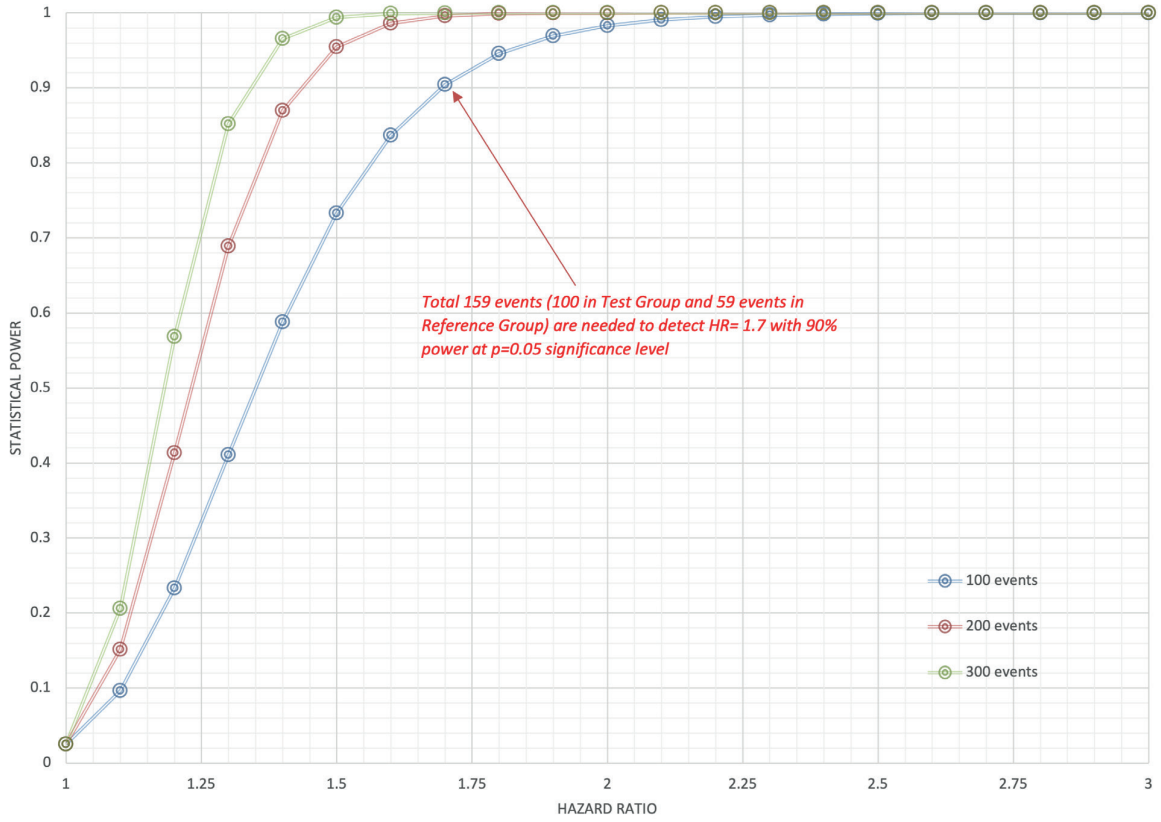

Table 4. Total number of events needed for given Hazard Ratios

Hazard Ratio	Events in Test Group	Events in Reference Group	Total Events	Alpha	Power
1	100	100	200	0.05	0.025
1.1	100	91	191	0.05	0.09646
1.2	100	83	183	0.05	0.23299
1.3	100	77	177	0.05	0.41098
1.4	100	71	171	0.05	0.58793
1.5	100	67	167	0.05	0.73304
1.6	100	63	163	0.05	0.8369
1.7	100	59	159	0.05	0.90445
1.8	100	56	156	0.05	0.94559
1.9	100	53	153	0.05	0.96955
2	100	50	150	0.05	0.9831
2.1	100	48	148	0.05	0.99065
2.2	100	45	145	0.05	0.99481
2.3	100	43	143	0.05	0.9971
2.4	100	42	142	0.05	0.99836
2.5	100	40	140	0.05	0.99907
2.6	100	38	138	0.05	0.99946
2.7	100	37	137	0.05	0.99969
2.8	100	36	136	0.05	0.99981
2.9	100	34	134	0.05	0.99989
3	100	33	133	0.05	0.99993
1	200	200	400	0.05	0.025
1.1	200	182	382	0.05	0.15163
1.2	200	167	367	0.05	0.41325
1.3	200	154	354	0.05	0.68921
1.4	200	143	343	0.05	0.86994
1.5	200	133	333	0.05	0.95463
1.6	200	125	325	0.05	0.98611
1.7	200	118	318	0.05	0.9961
1.8	200	111	311	0.05	0.99896
1.9	200	105	305	0.05	0.99973
2	200	100	300	0.05	0.99993
2.1	200	95	295	0.05	0.99998
2.2	200	91	291	0.05	1
2.3	200	87	287	0.05	1
2.4	200	83	283	0.05	1
2.5	200	80	280	0.05	1
2.6	200	77	277	0.05	1
2.7	200	74	274	0.05	1
2.8	200	71	271	0.05	1
2.9	200	69	269	0.05	1
3	200	67	267	0.05	1
1	300	300	600	0.05	0.025
1.1	300	273	573	0.05	0.20601
1.2	300	250	550	0.05	0.5683
1.3	300	231	531	0.05	0.852
1.4	300	214	514	0.05	0.9656
1.5	300	200	500	0.05	0.994
1.6	300	188	488	0.05	0.99914
1.7	300	176	476	0.05	0.99989
1.8	300	167	467	0.05	0.99999
1.9	300	158	458	0.05	1
2	300	150	450	0.05	1
2.1	300	143	443	0.05	1
2.2	300	136	436	0.05	1
2.3	300	130	430	0.05	1

2.4	300	125	425	0.05	1
2.5	300	120	420	0.05	1
2.6	300	115	415	0.05	1
2.7	300	111	411	0.05	1
2.8	300	107	407	0.05	1
2.9	300	103	403	0.05	1
3	300	100	400	0.05	1
1	400	400	800	0.05	0.025
1.1	400	364	764	0.05	0.25977
1.2	400	333	733	0.05	0.69212
1.3	400	308	708	0.05	0.93447
1.4	400	286	686	0.05	0.9919
1.5	400	267	667	0.05	0.99932
1.6	400	250	650	0.05	0.99996
1.7	400	235	635	0.05	1
1.8	400	222	622	0.05	1
1.9	400	211	611	0.05	1
2	400	200	600	0.05	1
2.1	400	190	590	0.05	1
2.2	400	182	582	0.05	1
2.3	400	174	574	0.05	1
2.4	400	167	567	0.05	1
2.5	400	160	560	0.05	1
2.6	400	154	554	0.05	1
2.7	400	148	548	0.05	1
2.8	400	143	543	0.05	1
2.9	400	138	538	0.05	1
3	400	133	533	0.05	1
1	500	500	1000	0.05	0.025
1.1	500	455	955	0.05	0.31251
1.2	500	417	917	0.05	0.78595
1.3	500	385	885	0.05	0.97254
1.4	500	357	857	0.05	0.99825
1.5	500	333	833	0.05	0.99993
1.6	500	313	813	0.05	1
1.7	500	294	794	0.05	1
1.8	500	278	778	0.05	1
1.9	500	263	763	0.05	1
2	500	250	750	0.05	1
2.1	500	238	738	0.05	1
2.2	500	227	727	0.05	1
2.3	500	217	717	0.05	1
2.4	500	208	708	0.05	1
2.5	500	200	700	0.05	1
2.6	500	192	692	0.05	1
2.7	500	185	685	0.05	1
2.8	500	179	679	0.05	1
2.9	500	172	672	0.05	1
3	500	167	667	0.05	1

For suggestions or comments on the SAP please email:

covid_clinplatform@who.int

World Health Organization