A minimal common outcome measure set for COVID-19 clinical research

WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection

Clinical research is necessary for an effective response to an emerging infectious disease outbreak. However, research efforts are often hastily organised and done using various research tools, with the result that pooling data across studies is challenging. In response to the needs of the rapidly evolving COVID-19 outbreak, the Clinical Characterisation and Management Working Group of the WHO Research and Development Blueprint programme, the International Forum for Acute Care Trialists, and the International Severe Acute Respiratory and Emerging Infections Consortium have developed a minimum set of common outcome measures for studies of COVID-19. This set includes three elements: a measure of viral burden (quantitative PCR or cycle threshold), a measure of patient survival (mortality at hospital discharge or at 60 days), and a measure of patient progression through the health-care system by use of the WHO Clinical Progression Scale, which reflects patient trajectory and resource use over the course of clinical illness. We urge investigators to include these key data elements in ongoing and future studies to expedite the pooling of data during this immediate threat, and to hone a tool for future needs.

Introduction
Clinical research is essential for an effective public health response during an emerging infectious outbreak.¹⁻³ Research enables early description of the nature, extent, epidemiology, and prognosis of the outbreak, and guides the selection of management strategies that benefit the largest number of patients. However, pandemic research is challenging because a new outbreak represents an unknown threat. Data must be accumulated rapidly to guide a response for which priorities are uncertain and the geographical reach is unknown. This information informs patient management but is also crucial for resource planning to ensure a benefit for the greatest number of people, and for public health measures to restrict the spread of disease and protect those who are directly involved in the response.

Reliable management conclusions require reproducible and widely accepted metrics to describe the emerging threat—to define the natural history (including infectivity and clinical course), to understand the spread and consequences for health-care systems, and to evaluate the effect of interventions that could modify the clinical course. Because these metrics are chosen rapidly and defined and measured differently from one study to the next, data sharing across studies could be facilitated if investigators agree to collect data for a set of common outcome measures.

The concept of a core outcome set has been championed by the Core Outcome Measures in Effectiveness Trials initiative.⁴ A core outcome set is defined as “an agreed standardised collection of outcomes that should be measured and reported for a specific area of health”.⁵ This strategy comprises a minimal set of outcomes that can be routinely recorded, independent of whether or not the study includes primary or secondary outcomes, so that the results of clinical trials in a particular disease can be reliably synthesised and compared. Collecting data for a core outcome set does not restrict the selection of primary or secondary outcome measures for a study. Rather, it ensures that specific data elements that are essential for the study of the disease are routinely collected and available. The development of a core outcome set presupposes previous experience with the disease, and so although the rationale is relevant to studies of a new disease, the method differs. In the evolving research response to an emerging pandemic, in which data are collected quickly and coordination of activities is difficult, a common minimal outcome set could be invaluable in understanding the epidemiology, evaluating therapies, and guiding a public health response.

As part of a WHO-led international collaborative response to the COVID-19 outbreak, a working group on clinical characterisation and management developed a minimum set of outcome measures for studies of the emerging outbreak. We describe a rapid consensus process to create this core outcome set, drawing on input from researchers, clinicians, patients, funders, and policy makers.

Development of the common minimal outcome set
The initiative was led by the Clinical Characterisation and Management Working Group established by WHO as a component of the research and development roadmap process in response to COVID-19. Members of this group comprised an international panel with expertise in clinical trials, epidemiology, virology, infectious diseases, critical care, and public health, as well as funders and policy makers. The working group met by videoconference and at a face-to-face meeting in Geneva, Switzerland, from Feb 11 to Feb 12, 2020, to discuss issues relevant to research into the clinical management of patients during the evolving outbreak. We agreed that a minimal but comprehensively collected outcome set could facilitate study design and data sharing, and that this set should include information on viral burden, clinical course, and survival measured at a more distant timepoint from randomisation (eg, 60 days rather than 7 days). Our goal
Endpoints used in clinical studies planned or done during the COVID-19 outbreak

| Table: Endpoints used in clinical studies planned or done during the COVID-19 outbreak |
|------------------|----------------------------------|
| Registered studies (n=1135)* | Specific metrics |
| Visceral burden 149 | Quantitative PCR |
| Mortality 118 | - |
| Duration of hospital or intensive care unit stay 32 | - |
| Symptoms 45 | Fever, vital signs, cough |
| Progression and resolution 175 | Multiple measure, scales |
| Lung injury and function 101 | Oxygen saturation, Murray score, oxygenation index |
| Other clinical measures 117 | - |
| Imaging findings 76 | CT scan, chest x-ray, echocardiogram |
| Biomarkers 73 | C-reactive protein, cardiac enzymes, cytokines |
| Depression, anxiety, long-term quality of life 63 | - |
| Co-infection, acute kidney injury, myocardial injury 15 | - |

*Not all studies listed their specific outcomes, or were studies whose outcomes did not reflect patient outcomes.

Review of clinical research databases and response to the questionnaire

As of April 21, 2020, there were 1135 planned or ongoing observational studies or clinical trials of patients with COVID-19 in the WHO International Clinical Trials Registry Platform. Although 41 different countries had registered clinical studies, 792 (69·8%) were based in China. The interventions varied and included antiviral drugs, mesenchymal stem cells, various immunomodulatory drugs, corticosteroids, convalescent plasma, and traditional Chinese medicines. Most of these studies used viral burden, mortality or length of hospitalisation, and progression or resolution of clinical symptoms as trial endpoints (table); measures of lung function were the primary endpoint in 101 (8·9%) studies.

We received input from 67 individuals in response to the first mailing of the outcomes questionnaire. These people represented 43 different research or professional networks and 25 different countries. 63 respondents indicated an ability to recruit patients to clinical trials of COVID-19. Our review of the published literature, combined with input from the questionnaires, identified three core domains to be included in a minimal common outcome set: mortality, viral load, and clinical course (progression and recovery). Additional outcome measures that might be considered for a core outcomes set are shown in panel 1, reflecting the spectrum of reported variables, rather than recommending their incorporation or the cutoff values that could be used for clinical assessment.

Mortality

Estimates from data up to March 1, 2020, place the mortality of SARS-CoV-2 infection at 1·4–5·7%. All respondents to our survey agreed that mortality was important to include in a set of minimal outcome measures and 50 (75%) of them agreed that hospital discharge was the appropriate timepoint to evaluate mortality status. 23 (34%) replies indicated a preference for one or more landmark timepoints to ascertain the mortality status, ranging from 28 days to 1 year. Potential limitations of the use of hospital discharge as a mortality endpoint included that patients in low-income and middle-income countries might leave hospital against medical advice when the prognosis is poor to avoid the costs of hospitalisation, that in a pandemic the need for care might exceed hospital resources with the result that patients would be managed at home, and that such an endpoint might miss hospital readmission and subsequent death. Moreover, mortality is dependent on the availability of resources and so might vary from one geographical area to another, particularly when need overwhelms available capacity.
Mortality is an intuitively sensible outcome for any disease that has a considerable attributable mortality risk. The extent of this risk for SARS-CoV-2 infection is unknown, but appears to be about 5% for patients. Emerging clinical data suggest that acute sudden death from pulmonary embolism, rather than a failure to resolve organ dysfunction, might be responsible for death in some cases. For this reason, we recommend that survival status be routinely collected in all studies and that the time for mortality ascertainment be sufficiently long to capture delayed deaths, ideally at hospital discharge or at 60 days. Such a timepoint could miss patients who are discharged in anticipation that they will die at home and patients who are discharged only to return with progressive illness, although the latter cohort can be evaluated by recording mortality at the last hospital discharge for SARS-CoV-2 infection.

Viral burden
Most respondents (49 [73%]) agreed that a measure of viral burden was an appropriate core outcome. Quantitative PCR (qPCR) to quantify viral copies was considered to be the best measure, with threshold cycle values during PCR as an alternative. Nasopharyngeal swabs are associated with the highest viral load.

qPCR is the most reliable method to detect the coronavirus responsible for COVID-19, although there are reports that radiographic signs might be present in patients with a negative qPCR test. qPCR quantifies viral transcripts in the selected sample relative to a standard control RNA using the ratio of amplification cycles needed to detect the virus. Thus, an alternate measure of viral load is the threshold cycle at which viral transcripts can be detected. Since SARS-CoV-2 is predominantly (at least initially) a respiratory pathogen, we recommend detection in specimens obtained from the upper or lower respiratory tract, but recognise that the virus might also be present in the faeces of patients infected with the virus. Quantification of viral burden provides no insight into the clinical status of the patient but does provide strong evidence of the presence of the pathogen, and it can be used to measure pathogen burden in response to treatment.

Non-mortal clinical outcomes
Adapted from a previously used measure that is an ordinal scale based on patient symptoms and location within the hospital and a scale applied in a Chinese trial measuring the efficacy of lopinavir and ritonavir in combination with interferon-alfa-2b, we modified an ordinal scale to measure clinical progression and recovery on the basis of location and supportive measures used within the health-care system. Our modifications sought to capture the entire spectrum of clinical illness from an asymptomatic carrier to death, and to provide greater resolution at the more severe end of the disease spectrum. The scale ranged from 0 (virus-free) to 10 (dead), with increasing numbers reflecting the severity of symptoms in ambulatory patients, patients treated in hospital, and patients admitted to an intensive care unit or high-dependency unit.

We sought input from the 67 survey respondents into the structure of the planned scale and used these comments to revise the measure. In particular, respondents noted that separating patients at the lower, less severe end of the scale might be difficult and that greater granularity might be provided at the upper, more severe end of the scale. Particular note was made of the limitations of the construct: the scale is largely untested and it is unknown how gradations of the scale correlate with mortality risk. The scale is ordinal, rather than numeric, and should probably be analysed with appropriate ordinal approaches, such as non-parametric tests or enumeration of transitions between classes on the scale, although this issue is controversial.

Data for the clinical response score would ideally be collected daily while the patient is being studied in the context of an observational study or randomised controlled trial. Since the variables measure symptoms or location and support within the health-care system,
Panel 2: A proposed core outcome measure set for clinical studies of COVID-19

Viral burden
Semiquantitative viral RNA of severe acute respiratory syndrome coronavirus 2 as measured by quantitative PCR or cycle threshold; nasopharyngeal swabs are associated with the highest viral load

Survival
All-cause mortality at hospital discharge or at 60 days

Clinical progression
WHO Clinical Progression Scale measured daily over the course of the study

<table>
<thead>
<tr>
<th>Patient State</th>
<th>Descriptor</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uninfected</td>
<td>Uninfected; no viral RNA detected</td>
<td>0</td>
</tr>
<tr>
<td>Ambulatory mild disease</td>
<td>Asymptomatic; viral RNA detected</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Symptomatic; independent</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Symptomatic; assistance needed</td>
<td>3</td>
</tr>
<tr>
<td>Hospitalised: moderate disease</td>
<td>Hospitalised; no oxygen therapy*</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Hospitalised; oxygen by mask or nasal prongs</td>
<td>5</td>
</tr>
<tr>
<td>Hospitalised: severe diseases</td>
<td>Hospitalised; oxygen by NIV or high flow</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Intubation and mechanical ventilation; pO2/FiO2 ≥150 or SpO2/FiO2 ≥200</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mechanical ventilation pO2/FiO2 &lt;150 (SpO2/FiO2 &lt;200) or vasopressors</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Mechanical ventilation pO2/FiO2 &lt;150 and vasopressors, dialysis, or ECMO</td>
<td>9</td>
</tr>
<tr>
<td>Dead</td>
<td>Dead</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure: WHO clinical progression scale

Emerging infectious disease epidemic. The scale provides a measure of illness severity across a range from 0 (not infected) to 10 (dead) with data elements that are rapidly obtainable from clinical records. Modelling in other disease states has shown that distinction is greater when seven or more classes are used, particularly at the lower range of disease severity.24 This spectrum, from the absence of infection to death, enables the scale to be used across a broad range of studies. Clinical and virological absence of infection is suggestive of a cure for patients who are initially infected or suggestive of a misdiagnosis for those individuals included in a trial. The WHO Clinical Progression Scale can also function as the entry criterion for patients in a vaccine trial. At the other end of the severity spectrum, the scale recognises that mechanical ventilation provides support that is survivable, although that probability is affected by both the severity of respiratory failure and the development of additional physiological organ dysfunction.

Tracking progression through the health-care system is potentially confounded by variability in the structure and capacity of those systems. Despite this variability, the health-care system is where patients who are infected receive their care and the burden of an emerging pandemic is felt both by the patient as acute illness and by the health-care system as strained resources. Systems with abundant or even excess capacity might care for patients in hospital or within the intensive care unit, whereas systems in resource-limited settings must rely on improvisation with available services. This issue creates a potential bias for studies that report locale in the health-care system as an outcome. We have tried to minimise this bias. First, we have done this by recognising that patients might be hospitalised for isolation and accommodate for this factor in the outcome scale. Second, the scale does not measure admission to an intensive care unit but instead focuses on the support that is typically provided there, and so a patient who is ventilated outside the hospital would have a high score. Therefore, intensive care is a process rather than a geographical location.

The scale has challenges. At the lower end of the scale the measures are subjective; differentiation between hospitalisation for quarantine versus hospitalisation for clinical support might be difficult. Quantification of subjective symptoms is similarly challenging. At the upper end of the scale, the use of life support measures is variable, reflecting not only on the patient's baseline comorbidities but also on the regional practice preferences. Although the scale has inherent face validity on the basis of its elements, this strategy must be tested and validated in independent data sets. A need for validation as a trial outcome does not preclude its use as a measure of treatment intensity within clinical trials of COVID-19.

The scale is intentionally presented as a simple minimal data set, focusing on variables relevant to most or all patients included in cohort studies or clinical trials. Special populations, such as pregnant women, are not included,

For more on the International Severe Acute Respiratory and emerging Infection Consortium see https://isaric.tghn.org/
For more on REMAP-CAP see https://www.remapcap.org/
but pregnancy outcomes would be important to monitor in women of child-bearing age.

There are a number of ways that the WHO Clinical Progression Scale might be used to identify a population for study and to track the progress of patients with COVID-19 within clinical trials. At the time of trial randomisation, the scale can serve to identify an appropriate cohort for study. Vaccine trials could recruit patients with a score of 0 and use any progression across the scale as endpoints. Large studies of patients with mild disease could recruit patients with a score of 3 or less and use progression to the need for hospitalisation or admission to intensive care units as a study endpoint. Similarly, studies of patients with severe disease could restrict recruitment to patients with a score of 5 or more and measure efficacy as either survival time or successful recovery to a lower score, for example, a value of less than 4 indicating a discharge from hospital. The scale can be modelled in a number of different ways, including median values at a fixed timepoint, time to a defined state, aggregate values over time, or change from baseline.

Integration into clinical research
COVID-19 research is rapidly changing, is globally collaborative, and is crucially dependent on new and unproven models of data aggregation. We urge medical professionals who care for patients with COVID-19 and those researchers who study the clinical characteristics of the illness to contribute data and to recruit patients to trials across a spectrum of platforms (panel 3).

In summary, we present a novel model of a minimal set of common outcome measures for ongoing and future studies responding to this outbreak. Further testing and validation of the measure are needed and this process might result in further modifications to its structure. The WHO Clinical Progression Scale has been developed to facilitate data pooling across cohort studies and clinical trials, with the objective of expediting the exchange of knowledge to benefit patients infected with SARS-CoV-2 and to inform optimal resource planning. To this end, and independent of the design and reporting of individual studies, we urge researchers to record these data elements and to share these results with the international community. Platforms and agreements for doing so are under development.

Contributors
JCM, SM, and JD, the Writing Committee, designed the project. JCM wrote the first draft and reviewed data from online databases. JCM, SM, and JD edited subsequent drafts of the paper and approved the final manuscript.

WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection

Declaration of interests
JCM is co-chair of the WHO Working Group on Clinical Characterisation and Management of COVID-19 infection and chair of the International Forum for Acute Care Trialists. He reports travel support from the Bill & Melinda Gates Foundation, and personal fees from Asahi Kasei Pharma America. SM is co-chair of the WHO Working Group on Clinical Characterisation and Management of COVID-19 infection. JD is an employee of WHO.

Acknowledgments
Supported in part by a grant from the Canadian Institutes of Health Research.

References