Open Forum Infectious Diseases

MAJOR ARTICLE

Development of a Standardized Data Collection Tool for Evaluation and Management of Coronavirus Disease 2019

Stephen R. Morris,[1,a] Yoichiro Natori,[2,a] Douglas Salguero,[10] Alejandro Mantero,[11] Ruixuan Ma,[11] Daniela F. de Lima Corvino,[1] Anmary Fernandez,[1] Alex Lazo,[1] Christine A. Vu,[3] Lauren Bjork,[6] David Serota,[10] Jennifer Quevedo,[7] Ana Vega,[3] Meshell Maxam,[3] Kailynn DeRonde,[3] Pablo Barreiro,[4] Patricia Raccamarich,[10] Maria Romero Alvarez,[1] Dimitra Skiada,[1] Shuba Balan,[1] Maya Ramanathan,[1] Gregory Holt,[5] Jose Gonzales-Zamora,[10] Gio J. Baracco,[8,9] Susanne Doblecki-Lewis,[10] Lilian M. Abbo,[10] Paola N. Lichtenberger,[8,9] and Maria L. Alcaide[10]

[1]Jackson Memorial Hospital/University of Miami, Miami, Florida, USA, [2]Miami Transplant Institute, Jackson Health System, Division of infectious Diseases, Department of Medicine, University of Miami, Miami, Florida, USA, [3]Jackson Memorial Hospital, Department of Pharmacy, Miami, Florida, USA, [4]Hospital Carlos III—La Paz, Unit of Infectious Diseases, European University, Madrid, Spain, [5]University of Miami, Department of Medicine, Division of Pulmonary/Critical Care Medicine, Miami, Florida, USA, [6]Miami Veterans Affairs Medical Center, Department of Pharmacy, Miami, Florida, USA, [7]University of Miami Hospital, Department of Pharmacy, Miami, Florida, USA, [8]University of Miami, Department of Medicine, Division of Infectious Diseases, Miami, Florida, USA, [9]Miami Veterans Affairs Medical Center, Infectious Disease Section, Miami, Florida, USA, [10]University of Miami, Department of Medicine, Division of Infectious Diseases, Miami, Florida, USA, and [11]University of Miami, Department of Public Health Sciences, Division of Biostatistics, Miami, Florida, USA

Background. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for coronavirus disease 2019 (COVID-19), a disease that had not been previously described and for which clinicians need to rapidly adapt their daily practice. The novelty of SARS-CoV-2 produced significant gaps in harmonization of definitions, data collection, and outcome reporting to identify patients who would benefit from potential interventions.

Methods. We describe a multicenter collaboration to develop a comprehensive data collection tool for the evaluation and management of COVID-19 in hospitalized patients. The proposed tool was developed by a multidisciplinary working group of infectious disease physicians, intensivists, and infectious diseases/antimicrobial stewardship pharmacists. The working group regularly reviewed literature to select important patient characteristics, diagnostics, and outcomes for inclusion. The data collection tool consisted of spreadsheets developed to collect data from the electronic medical record and track the clinical course after treatments.

Results. Data collection focused on demographics and exposure epidemiology, prior medical history and medications, signs and symptoms, diagnostic test results, interventions, clinical outcomes, and complications. During the pilot validation phase, there was <10% missing data for most domains and components. Team members noted improved efficiency and decision making by using the tool during interdisciplinary rounds.

Conclusions. We present the development of a COVID-19 data collection tool and propose its use to effectively assemble harmonized data of hospitalized individuals with COVID-19. This tool can be used by clinicians, researchers, and quality improvement healthcare teams. It has the potential to facilitate interdisciplinary rounds, provide comparisons across different hospitalized populations, and adapt to emerging challenges posed by the pandemic.

Keywords. coronavirus-19; data collection tool; hospitalized.

In December 2019, severe pneumonia cases of unknown cause emerged in Wuhan, Hubei, China, with clinical presentations greatly resembling viral pneumonia [1]. A novel enveloped ribonucleic acid beta coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was deemed responsible for coronavirus disease 2019 (COVID-19) and rapidly became a pandemic virus. As of June 10, 2020, there have been over 7 million cases confirmed of COVID-19 worldwide and approximately 2 million cases in the United States [2]. The World Health Organization (WHO)/China Joint Mission estimated approximately 6.1% of all cases develop critical disease, with an overall case fatality rate of 3.8% [3].

Several patient characteristics have emerged as risk factors for severe illness and mortality. These include older age, male gender, smoking, black race, underlying comorbidities (ie, cardiovascular disease, diabetes, hypertension, chronic lung disease, cancer, chronic kidney disease, and obesity), symptomatology, and imaging findings at presentation to the hospital [4–14]. In addition, certain laboratory findings have also been associated with poor outcomes (ie, lymphopenia, high neutrophil-to-lymphocyte ratio, elevated transaminases, elevated lactate dehydrogenase [LDH], inflammatory biomarkers [C-reactive protein {CRP}, ferritin, and D-dimer]), red blood cell distribution width (RDW), troponin, as well as indicators

Received 2 July 2020; editorial decision 23 July 2020; accepted 24 July 2020.

[a]S. R. M. and Y. N. contributed equally to this manuscript.

Correspondence: Maria L. Alcaide, MD, Professor of Clinical Medicine, Department of Medicine Division of Infectious Diseases, University of Miami, 1120 NW 14th St., Suite 864, Miami, FL 33136 (malcaide@med.miami.edu).

Open Forum Infectious Diseases®

DOI: 10.1093/ofid/ofaa320

of acute kidney injury [5, 15–23]. However, these biomarkers require further validation to guide clinical practice in a novel pandemic [3, 15].

Several agents have been used to treat patients during the current pandemic, based on in vitro antiviral/anti-inflammatory activity or experience in other illnesses [24–26]. Some are Food and Drug Administration (FDA)-approved products being used off-label, and others are investigational drugs available from the manufacturer via compassionate use programs or as part of ongoing clinical trials. As of June 10, 2020, no drugs have been licensed by the FDA for COVID-19, although remdesivir, an antiviral with potentially broad activity, has received FDA emergency use authorization for use in COVID-19 [25, 27]. However, results from the supporting clinical trials have only indicated efficacy in reducing illness duration and have not shown a statistically significant survival benefit [28, 29]. Limited literature [30] and anecdotal experience [31] suggest that timing of interventions and appropriate patient selection are important considerations. On April 21, 2020, the National Institutes of Health published national guidelines recommending that the use of therapeutic agents for the management of COVID-19 should be done in the context of a clinical trial [25]. Because not all clinicians have access to clinical trials and knowledge evolves very rapidly, it is critical to collect systematic data on presentation, management, and clinical outcomes to provide feedback to frontline clinicians and to improve patient safety and quality of care [26].

The large disease burden attributed to an emerging pathogen has led to a strong focus by the media, public health organizations, and the research community leading to an explosion of rapidly published scientific manuscripts of limited long-term analysis. Despite this, knowledge gaps remain, and there is a need for harmonized definitions, data collection, and outcome reporting across many centers to identify those most likely to benefit from certain therapies and those who are most likely to have poor outcomes [26].

In the setting of the considerable unknowns faced daily by our clinical teams during the pandemic, we developed a system to study the presentation and outcomes of patients seen across our diverse practice sites in Miami, Florida. In this study, we describe the development and pilot validation of a comprehensive data collection tool that was rapidly adopted in our hospital settings. The use of such a tool can harmonize systematic data collection within and between clinical sites to help guide clinical practice and research during this pandemic.

METHODS

Study Setting

This study was conducted in 3 large academic centers in Miami, Florida. As of June 10, 2020, the counties served by the participating centers (Broward and Miami-Dade) reported 27 625 cases of COVID-19 [32]. Multiple factors could potentially contribute to poor outcomes in our setting: older average age, significant volume of visiting travelers, and high proportion of those who are economically disadvantaged or lack health insurance [33, 34]. The University of Miami Hospital, Miami Veterans Affairs Medical Center (VAMC), and Jackson Health System hospitals serve a large and diverse patient population. This includes a high volume of specific patient populations for which outcomes are not well described in COVID-19 literature: veterans, persons with human immunodeficiency virus (HIV), malignancy, or other immunocompromising conditions.

Development of the Data Collection Tool

The proposed tool was a set of spreadsheets developed by a multidisciplinary team of infectious disease physicians, intensivists, and antimicrobial stewardship pharmacists (ASP). The development phase was conducted at the Miami VAMC and Jackson Memorial Hospital, and the pilot validation phase was conducted at the University of Miami Hospital and all Jackson Memorial Hospital, under an IRB-approved protocol (protocol number 20200424). The team members formed a working group that met via telephone or video conferences at least biweekly during the early phase of the COVID-19 pandemic arriving in Miami, Florida in early March 2020. Based on available literature and Centers for Disease Control and Prevention (CDC) guidance [35], we determined a set of sociodemographic and comorbidity risk factors, clinical signs and symptoms, and diagnostic tests that could be obtained from the electronic medical records (EMRs) (Table 1) or patient interview. Laboratory tests were chosen based on literature correlating elevation of infection or inflammatory biomarkers (eg, CRP, ferritin, LDH, D-Dimer, troponin, neutrophil-lymphocyte ratio, RDW, procalcitonin) with greater likelihood of severe/critical COVID-19 [5, 16–24], or to calculate intensive care unit (ICU) physiologic severity scores [36]. Antiviral, anti-inflammatory, and supportive therapies included in the data collection tool were selected based on available literature and upon review of hospital protocols at our institution and other settings [20, 24, 26].

Based on the characteristics of the study (retrospective review of data obtained for purposes of clinical care and quality improvement), a waiver of informed consent was obtained from the Institutional Review Board (IRB). Health Insurance Portability and Accountability Act (HIPAA)-compliant spreadsheets were developed to track data trends and facilitate biweekly discussions. All elements included in the final data collection tool were discussed during the working group meetings, during international forums with clinicians from countries who encountered the pandemic early (Spain, Italy, China), and during interdisciplinary rounds with clinical providers—because this strategy has been shown to improve communication and foster agreement on the plan of care [37]. Factors were updated based on frequent forums and literature review until all members of the working group came to a mutual agreement. In

Table 1. Baseline Characteristics of the Pilot Validation Cohort

Characteristics	Total Cohort (n = 200)
Hospital	
Jackson Memorial Hospital	113 (56.5%)
University of Miami Hospital	87 (43.5%)
Age (median, IQR)	63 (49–73)
Gender	
Male	118 (59.0%)
Female	82 (41.0%)
Ethnicity	
Hispanic	99 (49.5%)
Black	57 (28.5%)
White/Caucasian	25 (12.5%)
Asian/Pacific Islander	9 (4.5%)
Middle East/North Africa/Central Asia	2 (1.0%)
Other/not specified	6 (3.0%
Missing	2 (1.0%)
Comorbidities	
Chronic ventilator dependence	4 (2.0%)
Asthma or COPD	33 (16.5%)
Obstructive Sleep Apnea	5 (2.5%)
Congestive Heart Failure[i]	18 (9.0%)
Hypertension	115 (57.5%)
Diabetes Mellitus	64 (32.0%)
Coronary artery disease (occlusive)	19 (9.5%)
End Stage Renal Disease	13 (6.5%)
Chronic Kidney Disease	24 (12.0%)
Cirrhosis	1 (0.5%)
Solid Organ Transplant	4 (2.0%)
HIV	10 (5.0%)
Malignancy[ii]	19 (9.5%)
Smoking (recent)	14 (7.0%)
Alcohol use (recent)	19 (9.5%)
Obesity (BMI > 30)	88 (44.0%)
Pregnancy	1 (0.5%)
Exposure Epidemiology	
Community[c]	145 (72.5%)
Facility[d]	29 (14.5)
Ship	17 (8.5%)
Foreign Travel	14 (7.0%)
Health Care Worker[e]	7 (3.5%)
Symptom Duration (median, IQR)	4 (2–7)
Symptoms	
Fever	145 (72.5%)
Cough	148 (74.0%)
Dyspnea	139 (69.5%)
Signs[f]	
Fever	134 (67.0%)
Tachypnea	73 (36.5%)
Hypoxia	101 (50.5%)
Tachycardia	103 (51.5%)
Hypotension	11 (5.5%)
Disease Severity[g]	
Mild	10 (5.0%)
Moderate	88 (44.0%)
Severe	83 (41.5%)
Critical	11 (5.5%)
Critical with MODS	8 (4.0%)
WHO Ordinal Severity Scale	
3: Hospitalized, no oxygen	85 (42.5%)
4: Oxygen by nasal prongs or face mask	86 (43.0%)
5: Noninvasive ventilation/high-flow oxygen	11 (5.5%)
6: Intubation and Mechanical Ventilation	11 (5.5%)
7: Ventilation and additional support (RRT/ECMO/vasopressors)	8 (4.0%)

Abbreviations: BMI, body mass index; COPD, chronic obstructive pulmonary disease; ECMO, extracorporeal membrane oxygenation; HIV, human immunodeficiency virus; IQR, interquartile range; MODS, multiple organ dysfunction syndrome; RRT, renal replacement therapy; WHO, World Health Organization.

[a] Five patients had congestive heart failure with reduced ejection fraction (EF ≤ 40%).

[b]Malignancy: 11 solid, 9 hematologic, 1 patient with both; 4 on active chemotherapy.

[c]If none of the other exposure locations was noted, the patient was labeled community exposure, by default.

[d]Facility: residence in nursing home, assisted living, long term care facility, prison, homeless shelter. Note: visiting one of the previously mentioned places did not qualify and there was a separate category for ship.

[e]Healthcare worker was someone performing their job duties in patient care areas of the hospital. patients who had visited a hospital did not qualify.

[f]Definition of abnormal vital signs: fever - temperature ≥37.5°C; tachypnea - respiration rate ≥30/minute; hypoxia - SaO_2 on room air ≤93% or PaO_2; FiO_2 ratio - <300 (mechanical or noninvasive ventilation); tachycardia - heart rate >90 beats per minute; hypotension - shock with vasopressor use.

[g]Definition of disease severity: mild - no signs, symptoms, or imaging consistent with pneumonia; moderate - signs and/or symptoms consistent with pneumonia and compatible imaging; severe - tachypnea or hypoxia (as defined above); critical - respiratory failure requiring mechanical ventilation but no other organ dysfunction requiring support; critical with MODS - respiratory failure requiring mechanical ventilation and other organ failure requiring support (renal replacement therapy, vasopressors, ECMO). Liver injury and need for transfusion not included.

[i,ii]These comorbidities have been associated with poor outcomes from COVID-19.

addition, the working group decided to track important dates: illness onset, hospital admission, as well as beginning and end dates of ICU care, mechanical ventilation and other life-support treatments, and antiviral/anti-inflammatory therapies. Working group members created a note template to evaluate those with confirmed or suspected COVID-19 using the above elements, and this was distributed to the infectious disease teams for use at their discretion.

The working group decided on standardized definitions of terms and format of data reporting to facilitate data analysis. Binary (ie, 1 = yes, 0 = no) format was used unless additional complexity existed. In these cases, either a limited set of result categories was established, or free-text format was used. Care was taken to distinguish data that could not be obtained (left blank) versus normal/negative (recorded as "0"). If date of illness onset could not be ascertained, it was considered missing data (left blank). The data tool noted patients with chronic respiratory failure who were dependent on mechanical ventilation (yes/no) or supplemental oxygen at rest (yes/no).

Each vital sign was noted as normal ("0") or abnormal ("1") using the most severe value on day of admission: fever - temperature ≥37.5°C; tachypnea - respiration rate ≥30/minute; hypoxia - SaO_2 on room air ≤93% or PaO_2/FiO_2 ratio <300 (mechanical or noninvasive ventilation); tachycardia - heart rate >90 beats per minute; hypotension - shock with vasopressor use.

The definitions for abnormal vital signs and COVID severity stage were based on early published natural history studies [3, 6, 16, 20], which were recommended for disease staging in later guidelines [26]. However, we distinguished 2 different stages within the previously defined "critical" stage: those with only respiratory failure were defined as "critical", and those with respiratory failure and failure of other organ systems requiring support were defined as "critical with MODS" (multiple organ dysfunction syndrome)—mild - no signs, symptoms, or imaging consistent with pneumonia; moderate - signs and/or symptoms consistent with pneumonia and compatible imaging; severe - tachypnea or hypoxia (as defined above); critical - respiratory failure requiring mechanical ventilation, but no other organ dysfunction requiring support; critical with MODS - respiratory failure requiring mechanical ventilation and other organ failure requiring support (renal replacement therapy, vasopressors, extracorporeal membrane oxygenation [ECMO]). Liver injury and need for transfusion were not included.

This modification was made based on our treatment protocol and our experience that MODS confers different prognosis and responsiveness to interventions. The WHO ordinal scale for clinical improvement [38] was used to track clinical course and need for respiratory support.

In addition, in categories in which normal cutoffs were different than typical clinical practice (fever - ≥37.5°C vs ≥38.2°C typically used for nonneutropenic patients), we included a free text field to record the actual value, to facilitate later categorization or quantitative analysis. Time-updated vital signs were reported as daily ranges. Time-updated laboratory tests used first morning arterial blood gas (ABG) results and the results from other laboratory tests drawn closest in time to the ABG results. For radiology results, abnormal features were new findings compared with baseline imaging, and if no prior comparison was available, the findings were presumed to be new.

The order of specific variables and groups were modified based on feedback from regular meetings of the working group. Final groupings and order of variables represented the most user-friendly format balancing the ability to extract data within our EMR systems during real-time care, to allow easy quality control and to further analyze data. Specifically, organization of the data collection tool aimed to group data that would be extracted at specific time points during hospitalization.

Pilot Validation of the Data Collection Tool

During the pilot validation phase, the data collection tool was used to track data obtained from patients being evaluated for confirmed or suspected COVID-19 during 31 consecutive days (March 23–April 23, 2020) at the participating sites. During the development phase, CDC criteria for persons under investigation included either (1) compatible clinical syndrome (fever and either cough or dyspnea) and epidemiologic risk factor (travel to a high-transmission country or close contact with confirmed case within 14 days of symptom onset) or (2) severe febrile lower respiratory illness requiring hospitalization without an alternative diagnosis [39]. However, guidance was updated on March 4, 2020 to suggest testing be considered at the discretion of the treating clinician. Thus, patients studied during the pilot validation phase presented with one of the following: viral syndrome, fever, acute dyspnea, cough, hypoxia, or abnormal chest imaging. The tool was applied regardless of epidemiologic risk factors.

The pilot validation phase was performed in the context of normal duties by ASP and the infectious disease team (fellows and/or staff physicians), using a convenience sample size. Data were entered and results were shared with stakeholders (infectious disease team, intensivists, and ASP team) during interdisciplinary rounds. The clinical data collected during the pilot validation phase was input into secure clinical data management system at each site: SharePoint or RedCap. The percentage of missing data was calculated by the number of patients in whom the data was missing by the total participants included in the pilot validation process. Missing data for diagnostic testing was calculated based on any data obtained before beginning any antiviral or anti-inflammatory therapy. Missing data for interventions and clinical outcomes were calculated based on data available at date of discharge or death. Missing data for other outcomes were obtained after evaluation by the infectious disease team.

RESULTS

The data collection tool was pilot tested with 200 patients admitted to 1 of the 2 participating centers over a 31-day period, and patient characteristics are described in Table 1. The final data collection tool contained 2 main components with worksheets for infectious diseases and ASP teams (Supplemental Appendix; summarized in Table 2) and additional worksheets for definitions of abbreviations and terms, laboratory reference ranges, and data dictionary. The infectious diseases fellows and ASP found that the required data could be easily and efficiently collected, and all working group members were satisfied with its functionality. Stakeholders indicated that use of the tool improved decision making during interdisciplinary rounds by allowing rapid review of patient's clinical presentation, treatment, and clinical course. The major data domains, specific data, and results from data collection during the pilot validation phase is specified in Table 2.

In both the development and pilot validation phase, scenarios arose that challenged our current framework of data reporting. When this occurred, a consensus decision was made within the

Table 2. Data Collection Tool Domains, Timing, Variables, and Percentage of Missing Data From the Electronic Medical Record (EMR)

Domain	Timing and Source	Variables	Missing (%) N = 200
Demographics and Exposure Epidemiology	Timing: Baseline Primary Source: EMR • Provider notes of patient-reported data Other source: • Interdisciplinary rounds report	Age, Gender, Race/Ethnicity Exposure Epidemiology • Community • Foreign Travel • Health Care Worker • Facility: nursing home/long-term care, assisted living, jail/prison • Cruise Ship	1.0% Race/ethnicity 7.0% Foreign Travel No missing data for age, gender, or other epidemiologic exposure groups
Prior Medical History and Medications	Timing: Baseline Primary Source: EMR • Provider notes and medication reconciliation Other source: • Interdisciplinary rounds report	Selected Home Medications • Angiotensin Converting Enzyme inhibitors • Angiotensin receptor blockers • Nonsteroidal anti-inflammatory drugs • Corticosteroids Comorbidities/Substance Use • Smoking (current) • Alcohol (recent) Special Populations (eg) • Pregnancy[a] • HIV infection[b] • Malignancy[c]	0%–2.0% in all comorbidities
Signs and Symptoms of Illness	Timing: • Symptoms: Baseline • Vital Signs: Baseline and time-updated (daily range, hospital day 1, 2, 3, etc) Primary Source: EMR • Symptoms: Provider notes documenting illness onset date per patient report • Vital Signs: EMR flowsheets Other source: • Interdisciplinary rounds report	Vital signs at admission Date of onset of first symptom All symptoms up to admission	6.0% Date of illness onset No missing data for • Most common symptoms (fever, cough, dyspnea) • Vital signs at admission <20% missing data on other nonspecific viral symptoms at admission (eg, sore throat, headache, nausea, diarrhea) >20% missing data for uncommon symptoms[d]
Diagnostic Test Results	Timing: • Baseline • Time-updated: Hospital day 1,2,3 etc.[e] • End of follow up[f] Source: EMR • Laboratory • Microbiology • Radiology Reports	Arterial Blood Gas Blood Count Chemistry Indices C-Reactive Protein Lactate Dehydrogenase Ferritin D-Dimer Interleukin-6 Triglycerides Albumin Procalcitonin Troponin Respiratory Virus PCR Panel[g] SARS-CoV-2 PCR Chest x-ray Computerized Tomography Cultures (Blood, respiratory, other sterile sites)	97.5% Triglycerides 56.6% Troponin 54.0% Procalcitonin 46.0% Respiratory viral PCR panel 87.5% CT scan All other tests: <10% missing data[h]

Table 2. Continued

Domain	Timing and Source	Variables	Missing (%) N = 200
Interventions Specific Treatments Supportive Care Modalities	Timing •Baseline • Time-updated: Hospital day 1, 2, 3 etc • End of follow up Source: EMR • Provider Notes • Medication Administration Record •Flowsheets	Antiviral Treatments • Hydroxychloroquine • Lopinavir-Ritonavir • Azithromycin • Oseltamivir • Remdesivir • Tacrolimus (continuation of home medicine) Anti-inflammatory/other adjunctive treatments • Tocilizumab • Corticosteroids • Intravenous Immunoglobulin • Anticoagulation • Other Anti-inflammatory agents Other therapies: eg, • Pulmonary Vasodilators • Inhaled Nitric Oxide • Stem Cell Therapy • Convalescent Plasma Supportive care [i] • Supplemental Oxygen Delivery • Noninvasive Ventilation • Mechanical Ventilation • Extra-Corporal Membrane Oxygenation • Renal Replacement Therapy • Vasopressor Use WHO ordinal scale for clinical improvement [j]	No missing data
Clinical outcomes and complications	Timing •Baseline • Time-updated: Hospital day 1, 2, 3, etc • End of follow-up Source: EMR • Flowsheets: dates of intubation/ extubation, pressor/ECMO/RRT • Provider Notes: Complications, dates of ICU admission/ discharge, dates of hospital admission/ discharge	Days requiring various life support modalities • Intensive Care Unit admission • Intubation and Mechanical ventilation • Intensive Care Unit length of stay • Hospital length of stay Mortality (in-hospital) • 14 days • 30 days Adverse events	No missing data

Abbreviations: CT, computed tomography; EMR, electronic medical records; ICU, intensive care unit; HIV, human immunodeficiency virus; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2 virus; WHO, World Health Organization.

[a]Pregnancy status and gestational age. All females of child-bearing age had a point-of-care pregnancy test unless they reported menopause.

[b]Duration of infection, viral load nadir CD4, recent CD4 (before COVID-19 illness) and adherence to antiretroviral therapy.

[c]Specific type (organ/primary cell type), treatment (yes/no), specific treatment (free text therapy name).

[d]Many of which were described during or after the development phase of our study (eg, dysguesia, anosmia).

[e]Used first morning arterial blood gas (ABG) results and laboratory tests drawn closest in time to the ABG results.

[f]End of follow-up for diagnostic data is defined as last data available before death or discharge from the hospital.

[g]Assay: BioFire for Jackson Memorial Hospital, GenMark for University of Miami Hospital.

[h]Significant missing data for interleukin-6 and hepatitis B serologies since these were only drawn in patients considered for Tocilizumab.

[i]Including respiratory support parameters: oxygen liters per minute, PEEP, FiO_2, respiratory rate, tidal volume, prone positioning, or paralytic use.

[j]Incorporates functional capacity and level of oxygenation and ventilation support needed to stabilize patient in the emergency department or initial level of care.

working group to minimize reporting bias and maximize harmonization. After this, relevant prior data underwent review and reclassification, if necessary, to ensure consistency.

After data extraction, we assessed the frequency of missing data (Table 2). Among demographic data, we had no missing data for gender or age and only 1.0% without race/ethnicity

recorded. We had no missing data for epidemiologic exposures except 7.0% for history of foreign travel. All females of childbearing age had a point-of-care pregnancy test. Based on documentation available in the EMR of the participating institutions, there was <10% missing data on comorbidities of interest. Date of illness onset had 6.0% missing data, but presence of symptoms at admission was variably reported. There was no missing data for the most common symptoms (fever, cough, dyspnea), <20% missing data for other symptoms commonly associated with viral infections (eg, sore throat, headache, nausea, diarrhea), and a greater percentage of missing data for uncommon symptoms (eg, dysgeusia, anosmia)—which were often described during or after the development phase of our study. Data availability on most objective measures was excellent, and there was <10% missing data except for a few specific tests: respiratory viral polymerase chain reaction panel (46.0%: limited supply), computed tomography scan (87.5%: limited use based on infection control concerns), and certain laboratory tests such as triglycerides (97.5%), procalcitonin (54.0%), and troponin (56.6%). These laboratory tests were often not sent unless severe disease and potential association with severe COVID-19 was described during or after our development phase [19, 22, 40].There was also significant missing data for interleukin-6 level and hepatitis B serologies, but these were obtained only in patients who were candidates for tocilizumab. There was no missing data regarding use of specific interventions, complications, or clinical outcomes.

DISCUSSION

We present the development of a COVID-19 data collection tool, which efficiently tracked comprehensive clinical data and provided up-to-date information to guide decision making in an era of rapidly evolving data. We propose its use to effectively collect harmonized data of individuals admitted to hospitals with confirmed or suspected SARS-CoV-2 infection [39]. In the context of a pandemic, we recommend selecting patients for evaluation patients based on clinical presentation, as described above, without requiring specific epidemiologic risk factors. Given the gaps and rapidly evolving knowledge in understanding risk factors, clinical presentation, treatment options, and clinical outcomes, coordinated data collection tools are needed to maximize utility of clinical data and to improve patient safety and quality of care.

Timely risk stratification in patients infected with SARS-CoV-2 and patient selection for interventions has proved challenging. Some groups have reported favorable performance of risk prediction tools using early clinical laboratory tests and patient characteristics [17, 18]. However, to date, there is no widely validated diagnostic that can reliably identify those at high risk for respiratory failure requiring mechanical ventilation or death. Additional challenges of expanding the COVID-19 knowledge base include differences in biomarker assays, protocols specifying the roles of certain interventions that vary by institution and over time, and variable definitions for abnormal vital signs and disease staging. Infectious Diseases Society of America (IDSA) guidelines [26] specifically emphasized the need to report relevant objective clinical outcomes and use standardized disease staging definitions that use readily obtainable clinical criteria, like the WHO/China Joint Mission [3]. Data collection tools have been reported for emerging infectious diseases and other conditions, and they have had a positive impact on patient safety, quality improvement, research, and clinical care [41, 42].

The COVID-19 data collection tool can be used by clinicians, researchers, and quality control staff, and adapted to their own setting during this pandemic. It has the potential to enable comparisons across different hospitalized populations in the future and to be rapidly adapted to the emerging challenges posed by the pandemic. In addition, it can be used easily and safely in settings with limited technology by using spreadsheets within secure data collection systems (RedCap, SharePoint).

A few general observations in our study are worthy of discussion. The performance of the data tool during collection of data in all the major domains listed was enhanced by using the EMR, but use of the tool was not dependent on review of EMR documentation. Daily interdisciplinary rounds featured presentation of new patients by the primary team who personally evaluated and interviewed the patient. In addition, availability of data on diagnostic testing was dependent on the practice pattern of the primary medical team and infectious disease team. For example, the likelihood of missing data for diagnostic testing obtained before antiviral/anti-inflammatory therapy was influenced by the adherence to recommendations for baseline laboratory testing by the primary medical team. In other cases, detailed information regarding uncommon or rare symptoms was not gathered or documented. Often, there was low suspicion for COVID-19, and this information could not be obtained at the time of infectious disease evaluation due to patient condition. All of these examples highlight the need for systematic data collection, and we recommend creating a site-specific note template to prompt clinicians to obtain data relevant for care of those with confirmed or suspected COVID-19.

There are a few opportunities for improvement that should be noted. First, the 2 centers in the pilot validation phase had different EMR systems, but we did not study differences in missing data or time needed to extract data between the 2 sites. Such comparison would have been confounded by differences in data collectors and is inconsistent with our proposed recommendation for others to adapt and optimize this tool locally. Thus, we cannot determine which EMR system had organization and functionality best suited to our tool. However, some EMR have downloadable data functionality, and data tool organization should be optimized to receive output data in this manner, if available. Second, the scientific

community should work toward consensus on how to report dynamic changes of certain biomarkers. This is important when data from multiple sites are combined for reporting because reference ranges and assays for laboratory tests may be different and may require normalizing procedures during analysis (eg, converting to percentage or fold change from baseline or pretreatment value). Third, we encountered few transplanted patients in the study. We believe a separate data collection tool should be developed and validated for recipients of solid organ and hematopoietic stem cell transplants. Likewise, there are other patient populations not well represented in our study that require further validation with this tool: chronic respiratory failure, obstructive sleep apnea, cirrhosis, HIV, and pregnancy. Fourth, there were no patients who had clinical documentation from recent care at other medical centers, but outside data would need to be accounted for in further improvements to the tool. In addition, insurance status and type are important variables not built into our data tool that should be included on further iterations. Uninsured patients who are hospitalized receive fewer services and are more likely to experience in-hospital mortality than insured patients [43]. Finally, this data tool only captured basic baseline data on patients with chronic respiratory failure; need for mechanical ventilation and need for supplemental oxygenation. More detailed data should be collected on such patients including support settings, level of exertion at assessment, and other factors—based on input from critical care specialists.

We believe these limitations can be overcome. Each center should include variables relevant to management of COVID-19 patients within local protocols and based on the monitoring and diagnostic capabilities of the medical center. This tool should then be adapted to the sophistication and organization of the EMR platform, validated locally, and improved based on feedback from members of interdisciplinary care teams.

At this time, no intervention has been shown to be effective in a randomized controlled trial for the most relevant clinical endpoints: mortality, rate of progression to *acute* respiratory distress syndrome, and need for mechanical ventilation [25, 26, 28, 29]. Thus, current guidelines do not recommend any specific therapies and recommend use of available products only within the context of a clinical trial [25, 26]. Until more trial results are published, clinicians and hospital systems are faced with treating patients who are currently ill, leading to difficult decisions on the role of multiple unproven therapies and on which clinical trials to pursue.

CONCLUSIONS

We developed this data collection tool to track patient data relevant to the initial evaluation and ongoing management of hospitalized patients with COVID-19 infection. It can be adapted and applied to facilitate research involving cohort studies investigating patient characteristics associated with poor outcomes, specific therapies associated with clinical improvement, and appropriate timing or patient selection for therapy [44]. Thus, data collected through this tool can inform future clinical practice, improve patient safety and quality of care, and provide feedback for the design and conduct of trials evaluating the efficacy of the existing and novel therapies for COVID-19.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We thank Dr. Paul Mariani (Lt. Col., US Air Force) and Dr. Jonas Gonseth-Garcia (Pan-American Health Organization/World Health Organization) for graciously offering to review the manuscript.

Author contributions. S. R. M. designed the data tool, coordinated Institutional Review Board (IRB) submission, conducted the development and pilot validation phase, analyzed the data, and wrote the manuscript. Y. N. designed the data tool and conducted the pilot validation phase D. S. and P. R. assisted in preparation of manuscript and creation of tables. A. M. and R. M. assisted in developing the RedCap site. D. D. C., A. F., A. L., C. A. V., L. B., J. Q., D. S., A. V., M. M., K. D., M. R. A., D. S., and S. B., G. H., and J. G.-Z. were involved in designing the data tool and its use in the pilot validation phase. P. N. L. and G. J. B. were involved in the development phase. P. B. consulted with the team on clinical experience of coronavirus disease 2019 (COVID-19) management in Spain S. D.-L. provided oversight for study design and IRB submission. L. A. was involved in the design of the data tool and provided study oversight during pilot validation phase. M. L. A. conceived the initiative and provided oversight and assisted in preparation of the manuscript and tables. All authors reviewed and edited the final version of the manuscript.

Disclaimer. The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Financial support. This study was funded by the Miami Center for AIDS Research (CFAR) (P30AI073961) and the University of Miami Clinical and Translational Science Institute, Grant Number 1UL1TR000460, from the National Center for Advancing Translational Sciences and the National Institute on Minority Health and Health Disparities.

Potential conflicts of interest. All authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

References

1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet **2020**; 395:497–506.
2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis **2020**; 20:533–4.
3. Aylward B, Liang W. Report of the WHO-China joint mission on Coronavirus disease 2019 (COVID-19). WHO-China Jt Mission Coronavirus Dis **2020**.
4. Zheng Z, Peng F, Xu B, et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. J Infect **2020**; 81:e16–25.
5. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet **2020**; 395:1054–62.
6. Wu Z, McGoogan JM. Characteristics of and important lessons from the Coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report

of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA **2020**; doi: 10.1001/jama.2020.2648.
7. Liang W, Guan W, Chen R, et al. Cancer patients in SARS-CoV-2 infection: a nationwide analysis in China. Lancet Oncol **2020**; 21:335–7.
8. CDC COVID-19 Response Team. Preliminary estimates of the prevalence of selected underlying health conditions among patients with Coronavirus disease 2019–United States, February 12-March 28, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:382–6.
9. Lighter J, Phillips M, Hochman S, et al. Obesity in patients younger than 60 years is a risk factor for Covid-19 hospital admission. Clin Infect Dis **2020**; doi: 10.1093/cid/ciaa415.
10. Kalligeros M, Shehadeh F, Mylona EK, et al. Association of obesity with disease severity among patients with Coronavirus disease 2019. Obesity (Silver Spring) **2020**; 28:1200–4.
11. Vardavas CI, Nikitara K. COVID-19 and smoking: a systematic review of the evidence. Tob Induc Dis **2020**; 18:20.
12. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA **2020**; doi: 10.1001/jama.2020.6775.
13. Onder G, Rezza G, Brusaferro S. Case-fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. JAMA **2020**; doi: 10.1001/jama.2020.4683.
14. Garg S, Kim L, Whitaker M, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed Coronavirus disease 2019–COVID-NET, 14 states, March 1-30, 2020. MMWR Morb Mortal Wkly Rep **2020**; 69:458–64.
15. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel Coronavirus-infected pneumonia in Wuhan, China. JAMA **2020**; 323:1061–9.
16. Wu C, Chen X, Cai Y, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with Coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern Med **2020**; doi: 10.1001/jamainternmed.2020.0994.
17. Liang W, Liang H, Ou L, et al.; China Medical Treatment Expert Group for COVID-19. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Intern Med **2020**; doi: 10.1001/jamainternmed.2020.2033.
18. Gong J, Ou J, Qiu X, et al. A tool to early predict severe Corona Virus disease 2019 (COVID-19): a multicenter study using the risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis **2020**; doi: 10.1093/cid/ciaa443.
19. Lippi G, Plebani M. Procalcitonin in patients with severe coronavirus disease 2019 (COVID-19): a meta-analysis. Clin Chim Acta **2020**; 505:190–1.
20. Guan WJ, Ni ZY, Hu Y, et al.; China Medical Treatment Expert Group for Covid-19. Clinical characteristics of Coronavirus disease 2019 in China. N Engl J Med **2020**; 382:1708–20.
21. Ruan Q, Yang K, Wang W, et al. Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. Intensive Care Med **2020**; 46:846–8.
22. Shi S, Qin M, Shen B, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. JAMA Cardiol **2020**; doi: 10.1001/jamacardio.2020.0950.
23. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet **2020**; 395:507–13.
24. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic treatments for Coronavirus disease 2019 (COVID-19): a review. JAMA **2020**; doi: 10.1001/jama.2020.6019.
25. COVID-19 Treatment Guidelines Panel. Coronavirus disease 2019 (COVID-19) treatment guidelines. National Institutes of Health. Available at https://www.covid19treatmentguidelines.nih.gov/. Accessed 28 May 2020.
26. Bhimraj A, Morgan RL, Shumaker AH, et al. Infectious Diseases Society of America guidelines on the treatment and management of patients with COVID-19. Clin Infect Dis **2020**; doi: 10.1093/cid/ciaa478.
27. US Food and Drug Administration. Emergency Use Authorization (EUA) for emergency use of remdesivir for the treatment of hospitalized 2019 coronavirus disease (COVID-19) patients. Available at: https://www.fda.gov/media/137564/download. Accessed 1 May 2020.
28. Beigel JH, Tomashek KM, Dodd LE, et al.; ACTT-1 Study Group Members. Remdesivir for the treatment of Covid-19–preliminary report. N Engl J Med **2020**; doi: 10.1056/NEJMoa2007764.
29. Wang Y, Zhang D, Du G, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. Lancet **2020**; 395:1569–78.
30. Cao B, Wang Y, Wen D, et al. A trial of Lopinavir-Ritonavir in adults hospitalized with severe Covid-19. N Engl J Med **2020**; 382:1787–99.
31. Siddiqi HK, Mehra MR. COVID-19 illness in native and immunosuppressed states: a clinical-therapeutic staging proposal. J Heart Lung Transplant **2020**; 39:405–7.
32. Florida Department of Health. Florida's COVID-19 Data and Surveillance Dashboard. Available at: https://fdoh.maps.arcgis.com/apps/opsdashboard/index.html#/8d0de33f260d444c852a615dc7837c86. Accessed 10 June 2020.
33. United States Census Bureau. Quick Facts: Miami-Dade County. Available at: https://www.census.gov/quickfacts/fact/table/miamidadecountyflorida,FL/PST045219. Accessed 5 May 2020.
34. World Population Review. Miami-Dade county, Florida population 2020. Available at: https://worldpopulationreview.com/us-counties/fl/miami-dade-county-population/. Accessed 5 May 2020.
35. Human Infection with 2019 Novel Coronavirus CDC. Available at: https://www.cdc.gov/coronavirus/2019-ncov/downloads/pui-form.pdf. Accessed 4 March 2020.
36. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med **1996**; 22:707–10.
37. Terra SM. Interdisciplinary rounds. Prof Case Manag **2015**; 20:299–307.
38. World Health Organization. "WHO R&D Blueprint novel Coronavirus: COVID-19 Therapeutic Trial Synopsis" (2020). **2020**. Available at: https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf. Accessed 4 March 2020.
39. National Notifiable Diseases Surveillance System: COVID-19 Interim Case Definition. Available at: https://wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/. Accessed 27 February 2020.
40. Mehta P, McAuley DF, Brown M, et al.; HLH Across Speciality Collaboration, UK. COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet **2020**; 395:1033–4.
41. Li Y, Hu Y, Yu J, Ma T. Retrospective analysis of laboratory testing in 54 patients with severe- or critical-type 2019 novel coronavirus pneumonia. Lab Invest **2020**:1–7. doi: 10.1038/s41374-020-0431-6.
42. Soucie JM. Public health surveillance and data collection: general principles and impact on hemophilia care. Hematology **2012**; 17:S144–6.
43. Institute of Medicine (US) Committee on the Consequences of Uninsurance. Care Without Coverage: Too Little, Too Late. Washington DC: National Academies Press; **2002**.
44. Swedish Research Study. Medical center is conducting the REACT COVID-19: real world evidence for anti-cytokine therapy in COVID-19, an international, multi-center, observational study. Available at: https://www.swedish.org/for-health-professionals/research/react-covid19. Accessed 4 May 2020.