Reliability and Validity of an Instrument of COVID-19 Patient-Reported Symptoms in Outpatients

Diana Rofail, PhD; Pip Griffiths, PhD; Giulio Flore, MSc; Mohamed Hussein, PhD; Sumathi Sivapalasingam, MD; Anna J. Podolanczuk, MD; Ana Maria Rodriguez, PhD; Vera Mastey, MS; Chad Gwaltney, PhD

Abstract

IMPORTANCE Patient-reported outcome instruments are key in assessing COVID-19–related symptoms and associated burden. However, a valid and reliable instrument to assess symptom severity and progression among outpatients with COVID-19 is not yet available.

OBJECTIVES To assess the extent to which the Symptoms Evolution of COVID-19 (SE-C19) instrument is valid, reliable, and able to detect symptom changes in outpatients with COVID-19, as well as to establish a definition of symptom resolution.

DESIGN, SETTING, AND PARTICIPANTS In this diagnostic/prognostic study, psychometric properties of SE-C19 were assessed in participants recruited into an ongoing, adaptive, phase 1/2/3, randomized, double-blind, placebo-controlled clinical trial, during 2020 to 2022. Adult outpatients with symptomatic COVID-19 were randomized 1:1:1 to receive 2.4 g or 8.0 g intravenous casirivimab and imdevimab or placebo, in outpatient centers at 114 sites, from 2 countries (US and Mexico).

MAIN OUTCOMES AND MEASURES Reliability, validity, and sensitivity to change of the SE-C19 were assessed. SE-C19 and Patient Global Impression of Severity (PGIS) were administered daily from predose at day 1 to day 29.

RESULTS Analysis was conducted on 657 adult outpatients (342 female patients [52.1%], 562 White patients [85.5%]), and 337 non-Hispanic patients [51.3%]. At baseline, patients reported a mean (SD) of 6.6 (3.9) symptoms (ie, rated as at least mild) with a mean (SD) of 3.8 (3.3) of these symptoms being rated as moderate or severe. Stable patients according to PGIS showed scores with intraclass correlation values indicating moderate-to-good test-retest reliability (ie, 0.50-0.90). At baseline, 20 item scores (87%) varied significantly across PGIS-defined groups, supporting the validity of the SE-C19. A symptom-resolution end point was defined after excluding the item sneezing due to its low ability to discriminate severity levels, and excluding confusion, rash, and vomiting, due to their low prevalence in this population. Symptom resolution required complete absence of all remaining items, except cough, fatigue, and headache, which could be mild or moderate in severity. A total of 19 of 23 items from the SE-C19 instrument were identified as valid and reliable to measure disease-related symptoms in outpatients with COVID-19.

CONCLUSIONS AND RELEVANCE This study identified 19 items that are valid and reliable to measure disease-related symptoms in outpatients with COVID-19, and proposed a definition of symptom resolution for potential use in future clinical trials.

**Introduction**

Clinical presentation of COVID-19 is heterogeneous, ranging from asymptomatic to severe disease and life-threatening complications requiring hospitalization and intensive care. Although the range of symptoms reported by outpatients differ across studies, the most commonly reported include chills, fever, headache, stuffy or runny nose, sore throat, shortness of breath, cough, fatigue, muscle or body aches, nausea, vomiting, loss of smell, and loss of taste. When present, symptoms can last days, weeks, or even months, with symptoms not initially present often manifesting later during the disease. In most patients, symptoms resolve spontaneously; however, the resolution kinetics differ across symptoms and some may remain long after viral clearance (ie, postacute sequelae SARS-CoV-2 infection, also known as long COVID-19).

Most measures to monitor symptoms evolution have been developed for hospitalized patients and are based on objective measures (eg, monitoring oxygen saturation or hemodynamics), but these are rarely applicable to outpatients. Evaluating the longitudinal trajectory of symptom occurrence and severity requires novel reliable and valid measures tailored to this population. Patient-reported outcome (PRO) measures are ideal, because they directly reflect the patient’s perspective. Despite guidance from the US Food and Drug Administration to evaluate the burden of COVID-19 in outpatients participating in clinical trials by using PRO instruments, no valid and reliable instrument to assess outpatient-reported symptom progression in clinical trials or clinical practice is yet available.

Regeneron Pharmaceuticals, Inc. recently developed the Symptoms Evolution of COVID-19 (SE-C19) instrument to address the symptoms experienced by outpatients with COVID-19. The aim of this diagnostic/prognostic study was to assess the extent to which the SE-C19 instrument is valid, reliable, and able to detect symptom changes in outpatients with laboratory-confirmed COVID-19, as well as to establish a definition of symptom resolution.

**Methods**

**Data Source and Study Population**

Psychometric properties of the SE-C19 were studied in symptomatic patients recruited in the Safety, Tolerability, and Efficacy of Anti-Spike (S) SARS-CoV-2 Monoclonal Antibodies for the Treatment of Ambulatory Adult and Pediatric Patients With COVID-19 (NCT04425629) phase 1/2/3 adaptive multicenter trial assessing efficacy and safety of the combination of casirivimab and imdevimab monoclonal antibodies in adult outpatients with COVID-19. In phase 3, there were 114 participating centers from 2 countries (US and Mexico). Patients could be symptomatic or asymptomatic at baseline, but all patients needed a positive reverse transcription-quantitative polymerase chain reaction in nasopharyngeal swab samples at randomization. In addition, patients could not have been previously hospitalized or be currently hospitalized for COVID-19. The study lasted 29 days from randomization at day 1.

The trial was conducted in accordance with the principles of the Declaration of Helsinki, International Council for Harmonization Good Clinical Practice guidelines, and applicable regulatory requirements. The study followed relevant parts of the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guideline. The local institutional review board or ethics committee at each study center oversaw trial conduct and documentation. All the patients provided written informed consent before participating in the trial. Patients were compensated for participating in the main clinical trial.

**Patient-Reported Outcomes**

In addition to clinical data and self-reported demographic data such as race and ethnicity, several PROs were administered in the study. These were the SE-C19, the Patient Global Impression of Severity (PGIS), and the Patient Global Impression of Change (PGIC). The 2 last questionnaires were
used as anchors (benchmarks) in the assessment of SE-C19 psychometrics. The SE-C19 consists of 23 items.\textsuperscript{11} It was developed by Regeneron per the Food and Drug Administration Patient Focused Drug Development Guidance Series.\textsuperscript{14} The SE-C19 was linguistically validated in 7 different languages per the International Society for Pharmacoeconomics and Outcomes Research Taskforce for Translation and Cultural Adaptation.\textsuperscript{15}

The SE-C19 consists of 23 symptoms associated with COVID-19, including respiratory, gastrointestinal, and febrile symptoms. It is administered electronically with patients using their own devices to complete it (ie, a bring your own device [BYOD] approach). If BYOD is not feasible, a web link is provided to the patient.\textsuperscript{16} Patients are first asked to select all experienced symptoms and are then moved to the next screen where they need to confirm these symptoms and then rate the severity of each experienced symptom as mild, moderate, or severe at its worst moment within the last 24 hours. Symptoms not checked in the initial screen are automatically rated as none for that time point.\textsuperscript{11} Before completing the questionnaire, patients are given the opportunity to review the selected list of symptoms and associated severity. The SE-C19 was completed daily for a period of 29 days.

At each time point after completing the SE-C19, patients were also asked to complete the PGIS, a single question that rates the overall symptom severity at their worst moment in the last 24 hours from 0 (no symptoms) up to 3 (severe symptoms).\textsuperscript{16} At day 29, patients completed the PGIC to assess their perception of the degree to which their symptoms had changed since baseline on a 7-point scale from −3 (very much worse) to 3 (very much better).\textsuperscript{17}

**Statistical Analysis**

All psychometric analyses were performed on a modified full analysis set of the pooled treatment groups, defined as all randomized patients with a positive reverse transcription-quantitative polymerase chain reaction in the nasopharyngeal swab samples at randomization in the trial. The analyses were conducted at the individual as well as the combined item level.

Statistical inferential (Mann-Whitney, analysis of variance, and analysis of covariance) tests were performed to assess convergent validity and sensitivity to change, using 2-sided tests at the α = .05 significance level unless stated otherwise. For point estimates, 95% CIs were used when stated.

The distribution of SE-C19 items and PGIS responses was summarized and reported for baseline and up to day 29. All summary statistics were reported for categorical variables in terms of frequency and percentage and in terms of mean (SD) and median (IQR) for all continuous variables. Additionally, correlations between each individual item were assessed using Spearman correlation coefficients, with correlations above 0.4 indicating items that may form a subscale or domain.

Test-retest reliability was assessed using intraclass correlation coefficients (ICCs).\textsuperscript{18} The expectation was that for patients identified as stable according to the PGIS score at 2 points in time, item scores would show similar highly correlated values. ICCs were calculated for each symptom between pairs of consecutive days from days 16 to 22 in selected stable patients (measured by patients presenting the same PGIS score at each pair of days). An ICC value of 0.50 to 0.75 suggests moderate reliability, 0.75 to 0.90, good reliability, and greater than 0.90, excellent reliability.\textsuperscript{19}

The ability of the SE-C19 to discriminate among groups of patients (known-groups validity) was assessed by comparing the difference in SE-C19 score between patients grouped by their self-reported severity at baseline (baseline PGIS score, grouped none-mild or moderate-severe), using a Mann-Whitney test. The expectation was that the item scores would show a high degree of variance between the PGIS defined groups in demonstration of the PRO ability to track severity status.

The ability of the SE-C19 to detect change in disease severity was also assessed. Changes in individual SE-C19 symptom scores between baseline and day 15, and baseline and day 29, were assessed for patients who had reported a change in severity at both periods on the PGIS, using a Mann-Whitney test.\textsuperscript{20} Days 15 and 29 were selected because a change either as a result of the study treatment or due to the normal evolution of the condition could be expected at these 2 time points.
The expectation for the PRO item scores was to show a strong change corresponding to the degree of change defined by the PGIS.

Structural validity was assessed using the longitudinal item response theory (IRT), which measures the ability of each item to discriminate across symptom severities across time points. The main assumption was that the proposed PRO represented a unidimensional latent construct determining the severity of COVID-19 in outpatients, and each item properly represented a specific level of severity in the target population. Single-time point Rasch Measurement Theory (RMT) models were also built according to the results arising from the longitudinal IRT. Iterations continued until model properties were acceptable as defined by the following. First, the response options for each item were appropriate (i.e., the ordering of response options was discriminated consistently by patients across the 4 options [none, mild, moderate, and severe]). Second, the range of expected item scores matched the range of expected patient severities on the same measurement continuum, as shown by Wright maps that compare the distribution of item latent score estimates with those of person latent score estimates. Third, no additional association between items existed beyond those expected by the model (once underlying patient severity had been accounted for). The G^2 χ^2 statistic was then used to assess the degree of association between observed responses after controlling for the underlying severity, and tested through the residual interitem correlations using the Yen Q3 statistics. Finally, the overall item fit was assessed using an iterative procedure where fit to the model was assessed both at the item and patient level, using infit and outfit statistics. A rating scale model was used and adjustments were made where necessary to improve the fit of the SE-C19 to the model.

All analyses performed were prespecified a priori in an analysis plan before database lock of the original trial. Analyses were performed in SAS version 9.4 (SAS Institute) and R Studio version 1.3 (R Project for Statistical Computing).

**Results**

**Baseline Characteristics**

The modified full analysis set criteria were met by 657 patients. Most patients were aged 50 years or younger (463 patients [70.5%]), 342 were female (52.1%), 562 were White (85.5%), and 337 were not Hispanic or Latino (51.3%) (Table 1). Overall, 405 patients (61.6%) had 1 or more factors associated with risk for severe COVID-19, with cardiovascular disease (137 patients [20.8%]) and chronic metabolic disease (89 patients [13.5%]) being the most prevalent.

**Patient-Reported Symptoms at Baseline**

Completion rates of SE-C19 and PGIS were high at baseline (SE-C19, 543 of 649 patients [83.7%]; PGIS, 451 of 649 patients [69.5%]) and remained higher than 68% up to day 22 (eTable 1 in the Supplement). Thereafter, completion rates fell below 50% and reached 48.7% (320 patients) by day 29. At baseline, most patients (532 of the 543 patients completing SE-C19 [98.0%]) reported 1 or more symptoms. On average, patients reported a mean (SD) of 6.6 (3.9; median [IQR] 6.0 [4.0-9.0]) different symptoms with at least mild severity (Table 2). Of these, a mean (SD) of 3.8 symptoms (3.3; median [IQR] 3.0 [1.0-6.0]) were rated as moderate or severe, and a mean (SD) of 0.8 (1.5; median [IQR] 0.0 [0.0-1.0]) as severe. By day 29, a total of 238 of 320 patients (74.4%) who had a valid response on the SE-C19 reported no symptoms (eTable 2 in the Supplement).

For most symptoms, the proportion of patients who reported that they did not experience it (i.e., rated as none) at baseline (ranging between 168 [30.9%] to 529 [97.4%]) outnumbered those who did (Figure 1). The 5 most prevalent symptoms at baseline were cough (reported by 415 patients [69.1%]; severe, 24 patients [4.4%]), fatigue (332 patients [61.1%]), severe, 65 patients [12.0%]), headache (329 patients [60.6%]; severe, 51 patients [9.4%]), body aches (282 patients [52%]; severe, 59 patients [10.9%]), and loss of taste and smell (255 patients [47%]; severe, 109 patients [20.1%]). In the PGIS, most patients scored their COVID-19 symptoms as mild (192 patients [42.6%]).
or moderate (216 patients [47.9%]) at baseline, with 16 [3.5%] and 27 [6.0%] patients reporting none and severe, respectively (data not shown).

Symptoms such as feverish, vomiting, confusion, and rash reached a symptom resolution rate above 95% within the first week from randomization (eTable 3 in the Supplement). For many other symptoms, such as chills, nausea, diarrhea, red or watery eyes, dizziness, pressure or tightness in chest, and stomachache, a similar resolution rate was not reached until the end of the second week.

Table 1. Demographic and Baseline Clinical Characteristics of the Modified Full Analysis Set

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants, No. (%) (N = 657)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>342 (52.1)</td>
</tr>
<tr>
<td>Male</td>
<td>315 (47.9)</td>
</tr>
<tr>
<td>Age group ≤50 y</td>
<td>463 (70.5)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>318 (48.4)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>337 (51.3)</td>
</tr>
<tr>
<td>Not reporteda</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>4 (0.6)</td>
</tr>
<tr>
<td>Asian</td>
<td>9 (1.4)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>61 (9.3)</td>
</tr>
<tr>
<td>White</td>
<td>562 (85.5)</td>
</tr>
<tr>
<td>Unknown</td>
<td>5 (0.8)</td>
</tr>
<tr>
<td>Not reportedb</td>
<td>16 (2.4)</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>137 (20.8)</td>
</tr>
<tr>
<td>Chronic lung disease</td>
<td>65 (9.9)</td>
</tr>
<tr>
<td>Chronic metabolic disease</td>
<td>89 (13.5)</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>10 (1.5)</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>3 (0.5)</td>
</tr>
<tr>
<td>Immunosuppressed</td>
<td>11 (1.7)</td>
</tr>
<tr>
<td>Taking immunosuppressants</td>
<td>6 (0.9)</td>
</tr>
<tr>
<td>Not reportedc</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>Risk factor, ≥1 risk factor for hospitalization due to COVID-19b</td>
<td>405 (61.6)</td>
</tr>
</tbody>
</table>

a Not reported means that the patient chose not to self-identify their race.

Unknown means that the question was not answered.
b Having a comorbidity is a risk factor for hospitalization.

Table 2. Number of Symptoms at Different Rating Levels at Baseline, Day 7, Day 15, Day 22, and Day 29 After Randomization

<table>
<thead>
<tr>
<th>Symptom severity</th>
<th>Symptom rating, mean (SD) or median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline (n = 543)</td>
</tr>
<tr>
<td>At least mild (rating &gt;0)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>6.6 (3.9)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>6.0 (4.0-9.0)</td>
</tr>
<tr>
<td>At least moderate (rating ≥1)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>3.8 (3.3)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>3.0 (1.0-6.0)</td>
</tr>
<tr>
<td>Severe (rating = 3)</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>0.8 (1.5)</td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>0.0 (0.0-1.0)</td>
</tr>
</tbody>
</table>
from randomization. By day 22, 90% or more of patients reported having no symptoms (ie, none) for many of the items, except for cough, headache, loss of taste and smell, and fatigue, for which symptom resolution rates ranged between 85% and 90%.

The correlation between pairs of items were weak for most cases ($r < 0.3$); 14 of 253 item pairs had moderate correlations (between $r = 0.3$ and $r = 0.5$) ([eFigure 1 in the Supplement](/sites/default/files/2022-10/figure1.png)). The 2 symptom pairs that most strongly correlated were headache with body aches ($r = 0.40$), and feverish with chills ($r = 0.48$) ([eFigure 1 in the Supplement](/sites/default/files/2022-10/figure1.png)). Despite some item pairs showing higher correlations, the observed low-to-moderate correlations did not support the need for specific symptom clusters, or overall domains or total scores. Instead, the items are to be considered independently, but with a responder definition according to symptom resolution across the collection of items (presented and discussed in the following section).

### Reliability, Known-Groups Validity, and Ability of the SE-C19 to Detect Change

The test-retest reliability (ie, the ability of the SE-C19 items to provide consistent scores when the symptoms were stable over time) was moderate to good across most items, with ICC estimates ranging from 0.50 to 0.90 ([eTable 4 in the Supplement](/sites/default/files/2022-10/etable4.png)). However, there were some exceptions. Test-retest reliability could not be tested for rash, chest pain, confusion, or vomiting because of low prevalence at the assessed periods. Red or watery eyes, stomachache, and dizziness also had varying profiles of reliability, where acceptable results were seen for some periods but not for others.

Patients with a higher self-reported COVID-19 symptom severity, as measured by the PGIS at baseline, reported more severe mean SE-C19 item scores ([eTable 5 in the Supplement](/sites/default/files/2022-10/etable5.png)). Of the 23 items, 20 (87%) had a significant difference in SE-C19 scores between the PGIS severity groups. The magnitude of differences varied across items, with fatigue showing the largest difference between groups (mean [SD] difference, 0.72 [0.09] of a maximum possible difference of 4), and rash showing the smallest difference (mean [SD] difference, 0.04 [0.03] of a maximum difference of 4).

Most SE-C19 items were able to detect change, with scores decreasing over time among patients who reported improvement on the PGIS ([eTable 6 in the Supplement](/sites/default/files/2022-10/etable6.png)) or PGIC ([eTable 7 in the Supplement](/sites/default/files/2022-10/etable7.png)). The exceptions were confusion, rash, and red or watery eyes, for which the direction of the score change on the SE-C19 was misaligned with the reported improvement on the PGIS or PGIC. In the longitudinal IRT, sneezing was associated with poor ability to discriminate between symptom severities and was, therefore, excluded from the RMT analyses.

---

**Figure 1. Distribution of Responses for Symptoms Evolution of COVID-19 (SE-C19) Items at Baseline**

![Figure 1](/sites/default/files/2022-10/figure1.png)
Examination of the probability curves of selection of severity categories using the RMT analyses suggested that patients were unable to reliably distinguish between the options mild and moderate (Figure 2). As a result, these 2 response options were merged into a single response score mild-moderate, and the RMT model was rerun. In the new model, vomiting, confusion, and rash fell outside the expected severity range for this study population, with mild-moderate rating for these items being associated with a latent severity 3 logits above the average item severity. This suggests that very few patients would report these symptoms (Figure 3). When assessing whether high correlations between items existed beyond those expected by the model (ie, once the underlying patient severity had been accounted for), most items had a low correlation level ($r < 0.20$). Chills and feverish had the highest residual correlation of $r = 0.31$ (eFigure 2 in the Supplement).

The RMT analysis suggested an overall good fit for the items, except for vomiting, confusion, and rash. Loss of taste and smell also showed slightly poorer model fit than expected, but it has a theoretical association with COVID-19 symptoms in this study population (eTables 8 and 9 in the Supplement). The good model fit, and low residual correlations help to reinforce the structural validity of the measure.

**Figure 2. Example of Item Response Curves When Using All 4 Response Options Compared With Merging Mild and Moderate**

**A** All 4 response options

---

**B** Mild and moderate responses merged

$\theta$ represents the underlying severity of COVID-19 symptoms.

**Figure 3. Wright Map for Symptoms Evolution of COVID-19 Items at Baseline, 19 Items, 3-Rating Scale**
Development of a Symptom-Resolution End Point

A symptom-resolution end point was developed. According to the RMT model findings, vomiting, confusion, and rash, together with sneezing, were excluded from the responder definition. A responder can be defined as a patient with resolution (ie, rating of none) of the 19 remaining items, except cough, fatigue, and headache, for which a maximum of mild or moderate could be rated for each of these items.

Discussion

This diagnostic/prognostic study demonstrates that the SE-C19 is a valid and reliable measure able to capture longitudinal changes in the occurrence and severity of COVID-19–related symptoms in outpatients. To our knowledge, this is the first study to establish the reliability and validity of a patient-reported COVID-19 symptom measure for outpatients.

In our study and in line with the literature, patients had heterogeneous symptom presentations at baseline; the most prevalent being cough, fatigue, headache, body aches, and loss of taste and smell. In the current study, however, patients were more likely to report a moderate or severe loss of taste and smell as compared with other symptoms. This was not the case in a US study with COVID-19–positive outpatients in which 4% of loss of taste and smell cases were classified as severe (vs 20.1% in our study). This difference may be due to the primary respondents being physicians rather than patients in the US study, and its retrospective nature.

In line with published data from other longitudinal studies with outpatients, symptoms progression was also heterogeneous. Although some symptoms resolved within the first week, others persisted generally until the end of the second. Over 3 weeks, most symptoms resolved except for cough, headache, and fatigue. Other longitudinal clinical and observational studies reported similar findings. Longitudinal changes may also differ across patients and COVID-19 variants. However, at the time of the study, no major COVID-19 variants had been yet detected and therefore the analysis did not focus on this possibility.

At baseline, most patients (98.0%) reported at least 1 of the 23 symptoms. Despite the heterogeneity of symptom progression, the percentage of symptomatic patients gradually decreased, and by the end of the study most (74.4%) patients were asymptomatic regardless of treatment assignment with different symptoms resolving at different time points. In the context of clinical trials, it is important to use a responder definition that allows assessment of treatment benefit. The responder end point defined using the SE-C19 is based on symptoms resolution, a commonly used end point in clinical trials of infectious diseases.

Limitations

Further studies are needed to assess the sensitivity and specificity of the SE-C19 as a screening tool for outpatients with COVID-19. One limitation is the observed increasing missing data of the SE-C19, particularly toward the end of the trial. This could have possibly biased the estimate of the change in severity between baseline and day 29 if patients with more favorable or unfavorable changes in severity had dropped out. However, the consistency of the finding between day 15 and day 29 suggests that there might be very little bias introduced by the observed missing pattern at day 29. Additional studies using clinical markers or measures may further help understanding of the responder definition developed for clinical trials, as well as the relevance of the other 4 items in monitoring disease. Although the SE-C19 was initially designed for the outpatient setting, its potential use in hospitalized patients should be assessed. With further work evaluating this tool in specific nonclinical trial populations, the SE-C19 could be used either as an additional criterion for discharge decisions, or for follow-up for potential readmission which may occur in up to 20% of patients.

Additionally, while the SE-C19 has been shown to be a valid tool, the current results might not be generalizable directly to a population with a different composition of race and ethnicity, who
might have a different level of literacy. As a result, future studies should evaluate the psychometric validation of the SE-C19 in these different subpopulations.

Conclusions

In this diagnostic/prognostic study, we developed a symptom-resolution definition based on 19 SE-C19 items that could be used as a standard in future clinical trials. The SE-C19 is easy to complete and could also be leveraged for daily symptom monitoring and decision-making in clinical practice. This study provides evidence that the SE-C19 instrument is a valid, reliable, and sensitive measure that can be used to assess COVID-19 symptoms and their progression among outpatients.
Additional Contributions: We thank the study participants, their families, and the clinicians involved in this trial. Medical writing assistance was provided by Hervé Besson, PhD, and Montse Casamayor, MD, PhD, from IQVIA; these writers did not receive any compensation beyond their normal salary. Editorial support was provided by Prime, Knutsford, UK, funded by Regeneron Pharmaceuticals, Inc.

REFERENCES


SUPPLEMENT.
eTable 1. Completion Rates of the SE-C19, PGIS, and PGIC

eTable 2. Percent of Patients Reporting at Least 1 Symptom at Different Rating Levels at Baseline, Day 7, Day 15, Day 22, and Day 29 After Randomization

eTable 3. Percentages of Patients With Reporting No Symptoms (None) at Baseline and Days 7, 15, 22, and 29

eTable 4. Intraclass Correlation Coefficients

eTable 5. Mean Difference in SE-C19 Item Scores at Baseline Between Known Groups Defined by PGIS Value at Baseline (Convergence Validity)

eTable 6. Mean Change in SE-C19 Item Scores From Baseline to Days 15 and 29 by PGIS Scores at Days 15 and 29, Respectively

eTable 7. Mean Change in SE-C19 Item Scores From Baseline to Day 29 by PGIC Scores at Day 29

eTable 8. Goodness-of-Fit Measures for Items for Each Model

eTable 9. RMT Model Person Goodness-of-Fit Measures

eFigure 1. Interitem Correlations

eFigure 2. Heat Map of the Residual Correlation Matrix