TITLE PAGE

Title

Statistical Challenges when Analyzing SARS-CoV-2 RNA Measurements Below the Assay Limit of Quantification in COVID-19 Clinical Trials

Running Title

Analyzing SARS-CoV-2 RNA in COVID Trials

AUTHORS

Carlee B Moser, PhD, Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Boston, 02115, USA

Kara W Chew, MD, MS, Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, 90024, USA

Mark J Giganti, PhD, Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Boston, 02115, USA

Jonathan Z Li, MD, MMSc, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Cambridge, 02139, USA

Evgenia Aga, MS, Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Boston, 02115, USA

Justin Ritz, MS, Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Boston, 02115, USA

Alexander L Greninger, MD, PhD, Department of Laboratory Medicine & Pathology, University of Washington, Seattle, 98195, USA

Arzhang Cyrus Javan, MD, MPH, DTM&H, National Institutes of Health, Rockville, 20852, USA

Rachel Bender Ignacio, MD, MPH, Department of Medicine, University of Washington, Seattle, 98195, USA

Eric S Daar, MD, Lundquist Institute at Harbor-UCLA Medical Center, Torrance, 90502, USA

David A Wohl, MD, Department of Medicine, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, 27599, USA

Judith S Currier, MD, MSc, Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, 90024, USA

Joseph J Eron, MD, Department of Medicine, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, 27599, USA

Davey M Smith, MD, MAS, Department of Medicine, University of California, San Diego, La Jolla, 92093, USA

Michael D Hughes, PhD, Department of Biostatistics and Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Boston, 02115, USA

For the ACTIV-2/A5401 Study Team

Word Count: 145 (abstract) / 3465 (text)

Corresponding Author

Carlee B. Moser, PhD, Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, 651 Huntington Ave, FXB 513, Boston, MA, 02115, USA, phone: (617) 432-2526, fax: (617) 432-3163, email: cmoser@sdac.harvard.edu

FOOTNOTE PAGE

Disclosures

CBM participated on a Data Safety Monitoring Board for the BONE STAR study. KWC has received research funding to the institution from Merck Sharp & Dohme and is a consultant for Pardes Biosciences. JZL has consulted for Abbvie and received research support from Merck ALG reports contract testing from Abbott, Cepheid, Novavax, Pfizer, Janssen and Hologic and research support from Gilead. RBI received consulting fees from SeaGen and Resverlogix unrelated to this work and research funding through her institution from Novartis and Ascentage Pharma Group. ESD receives consulting fees from Gilead Sciences, Merck, and GSK/ViiV and research support through the institution from Gilead Sciences and GSK/ViiV. DAW has received funding to the institution to support research and honoraria for advisory boards and consulting from Gilead Sciences. JSC has consulted for Merck and Company. JJE is an ad hoc consultant to GSK/VIR, data monitoring committee (DMC) chair for Adagio Phase III studies. DMS has consulted for the following companies: Evidera, Fluxergy, Kiadis, Linear Therapies, Matrix BioMed, Arena Pharmaceuticals, VxBiosciences, Model Medicines, Bayer Pharmaceuticals, Signant Health and Brio Clinical. MJG, EA, JR, MN, ACJ and MDH report no competing interests.

Funding Statement

This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number UM1 AI068634 (MDH), UM1 AI068636 (JSC) and UM1 AI106701 (GA). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Meeting Information

N/A

Corresponding Author

Carlee B. Moser, PhD, Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, 651 Huntington Ave, FXB 513, Boston, MA, 02115, USA, phone: (617) 432-2526, fax: (617) 432-3163, email: cmoser@sdac.harvard.edu

ABSTRACT

Most clinical trials evaluating COVID-19 therapeutics include assessments of antiviral activity. In recently completed outpatient trials, changes in nasal SARS-CoV-2 RNA levels from baseline were commonly assessed using analysis of covariance (ANCOVA) or mixed models for repeated measures (MMRM) with single-imputation for results below assay lower limits of quantification (LLoQ). Analyzing changes in viral RNA levels with singly-imputed values can lead to biased estimates of treatment effects. In this paper, using an illustrative example from the ACTIV-2 trial, we highlight potential pitfalls of imputation when using ANCOVA or MMRM methods, and illustrate how these methods can be used when considering values <LLoQ as censored measurements. Best practices when analyzing quantitative viral RNA data should include details about the assay and its LLoQ, completeness summaries of viral RNA data, and outcomes among participants with baseline viral RNA ≥LLoQ, as well as those with viral RNA <LLoQ.

Key Words: SARS-CoV-2 RNA, COVID-19, linear regression for censored data, randomized trial

Trial Registration: ClinicalTrials.gov Identifier: NCT04518410

BACKGROUND

Clinical trials designed to evaluate COVID-19 therapeutics should have clinically meaningful endpoints. FDA guidance states that clinical outcomes, such as the proportion of participants hospitalized or time to symptom recovery, are recommended as primary outcomes in phase III outpatient COVID-19 trials [1]. However, it also states that viral shedding should be measured to assess antiviral activity, primary virology outcomes are acceptable in phase II, and quantitative and qualitative virological assessments are encouraged.

In typical COVID-19 randomized trials, samples such as nasopharyngeal swabs, anterior or mid-turbinate nasal swabs, oropharyngeal swabs, saliva, or plasma, are collected longitudinally for SARS-CoV-2 RNA testing before and after intervention. Repeat sampling from early timepoints is common and in phase III typically includes one to four timepoints (Supplemental Table 1).

To evaluate virologic efficacy, SARS-CoV-2 RNA, henceforth called viral RNA (vRNA), is measured with quantitative reverse transcription polymerase chain reaction (RT-qPCR) assays. Like other nucleic acid assays, SARS-CoV-2 RNA assays have limits between which vRNA is accurately quantified, called the lower limit of quantification (LLoQ) and upper limit of quantification (ULoQ). For results >ULoQ, samples can be rerun with dilution to obtain quantifiable values. Assays may also indicate whether results <LLoQ are detectable or not.

Recent outpatient COVID-19 therapeutic trials considered various vRNA outcome measures and statistical methods. Most commonly, vRNA changes from baseline were analyzed using

analysis of covariance (ANCOVA) at each timepoint or mixed models for repeated measures (MMRM). With these methods, single-imputation was used to assign values for vRNA results <LLoQ (Supplemental Table 1) [2–18]. However, such imputation can introduce bias in estimating the magnitudes of treatment effects, as uncertainty for values <LLoQ isn't captured [19].

Using an illustrative example from the ACTIV-2 COVID-19 outpatient treatment trial, we describe bias that may arise when estimating treatment effects using single-imputation with ANCOVA and MMRM. Drawing on the HIV literature [19], we describe and discuss alternative approaches for analyzing vRNA changes, that may be more appropriate by considering vRNA values <LLoQ as censored measurements. Finally, we provide recommendations for the analysis and presentation of results concerning vRNA changes in future trials.

METHODS

ACTIV-2 (NCT04518410) is an adaptive platform trial designed to evaluate potential outpatient therapeutics for COVID-19[20]. Our illustrative example includes 114 participants randomized to receive tixagevimab/cilgavimab intravenously or placebo; the primary results previously reported [21]. Nasopharyngeal swabs were collected before treatment at Day 0 (baseline) and Days 3, 7 and 14 for SARS-CoV-2 RNA quantitative testing using a RT-qPCR assay with LLoQ of 2 $\log_{10}$ copies/ml [22]. All results >ULoQ were rerun with dilution to obtain quantifiable results. ACTIV-2 was approved by a central institutional review board (IRB), Advarra (Pro00045266), with additional local IRB review and approval as required by participating sites. All participants provided written informed consent.

As this manuscript aims to illustrate and discuss different approaches to analyze vRNA changes, we provide an overview in Table 1, but integrate descriptions of each method in the Results. For methods that use imputed values for results <LLoQ, two commonly-used single-imputation strategies (Supplemental Table 1) were assessed:

(1) "*LLoQ-imputation*": impute values <LLoQ as the LLoQ,

(2) "*½LLoQ-imputation*": impute values <LLoQ as ½ the LLoQ.

See Supplemental Methods for additional details on model specifications and sample SAS software code.

RESULTS

Descriptive summaries of vRNA across timepoints for the 114 participants are shown in Table 2A and Figure 1A and 1B. At baseline, 15 participants (13%) had missing vRNA (Supplemental Figure 1). There was a chance imbalance in vRNA between the randomized arms, with median vRNA in the active arm 1.0 $\log_{10}$ copies/ml higher than the placebo arm, and a higher proportion of participants with vRNA ≥LLoQ (72% versus 62%).

Following the recommendation of Marschner et al. [19], we separately considered data for participants with vRNA <LLoQ from those ≥LLoQ at baseline. For those with vRNA <LLoQ at baseline (N=33), vRNA remained <LLoQ at all follow-up timepoints in both arms, suggesting peak vRNA may have been achieved before enrollment. For the remaining analyses, we focus on the 66 participants with vRNA ≥LLoQ at baseline. The proportion with vRNA <LLoQ increased over time: 27% and 28% at Day 3, 62% and 54% at Day 7, and 93% and 89% at Day 14 for the active and placebo arms, respectively (Table 2B and Figure 1C and 1D).

159

160 Analyzing vRNA at a Single Timepoint

161 *1. Using imputed values leads to biased estimates*

162 Fifty-five (83%) of the 66 participants had vRNA results at Day 3 (Supplemental Figure 1). For
163 these 55 participants, at baseline there was a modest difference (0.33 $\log_{10}$ copies/ml) in mean
164 vRNA: 5.61 and 5.28 $\log_{10}$ copies/ml for the active and placebo arms, respectively.

165

166 Using *LLoQ-imputation*, the mean vRNA at Day 3 was 3.43 and 3.97 $\log_{10}$ copies/ml for the
167 active and placebo arms, respectively, with estimated mean changes from baseline of -2.18 and
168 -1.30 $\log_{10}$ copies/ml. Within each arm, the estimated mean changes are conservative and
169 biased because for participants with vRNA <LLoQ at Day 3, the true changes are at least as
170 large in magnitude as the imputed changes. Using *½LLoQ-imputation* gives mean changes that
171 are larger (more negative) compared to *LLoQ-imputation*: -2.45 and -1.58 $\log_{10}$ copies/ml for the
172 active and placebo arms, respectively. This imputation still results in biased estimates, but with
173 an unknown direction (estimated changes may be larger or smaller than the truth). For both
174 approaches, the larger mean change in the active arm could reflect higher average baseline
175 values, and thus larger changes are observable. Since the estimated mean changes within each
176 arm are biased, the estimated difference between arms will be biased, and further bias may be
177 introduced with the baseline imbalances.

178

179 The estimated difference in mean change for the active versus placebo arms at Day 3 was -0.87
180 $\log_{10}$ copies/ml using *LLoQ-imputation* and -0.86 $\log_{10}$ copies/ml using *½LLoQ-imputation* (Table
181 3A). Although these estimates are similar, this may not be the case in other datasets when

using the two approaches. By Day 14, when ~90% of participants had vRNA <LLoQ (and hence had imputed changes), the estimated difference in mean change between arms was approximately equal to the baseline mean difference for both imputation approaches. If all participants had vRNA <LLoQ at Day 14, the difference in mean change would equal the difference in mean vRNA at baseline, despite the choice of imputed value and underlying true difference. With larger proportions <LLoQ, differences between arms can reflect chance imbalances at baseline rather than true differences.

2. Adjusting for baseline can help address baseline imbalances

Although adjusting for baseline doesn't remove the bias in estimating differences between arms using singly-imputed values, it may help reduce the impact of baseline imbalances in mean vRNA when assessing treatment effects.

The estimated differences in mean changes between arms using standard linear regression are shown in Table 3A-B. In adjusted analyses, differences between arms have some attenuation at each timepoint compared with unadjusted analyses, reflecting the adjustment for higher baseline vRNA levels in the active arm.

3. *Analysis methods considering vRNA <LLoQ as censored*

Statisticians refer to vRNA values <LLoQ as being left-censored because the if the true vRNA could be measure it would be a value between zero and LLoQ (i.e., a value to the left of LLoQ). This contrasts with right-censoring like in survival analysis where, for example, participants alive at the end of follow-up have time of death greater than (to the right of) the time at the end of

follow-up. Statistical methods used for survival analysis can be used to analyze vRNA data, with the small adaptation that values are left-censored rather than right-censored. Change in vRNA is defined as the difference in vRNA at the follow-up time minus the baseline. However, for follow-up vRNA values that are <LLoQ or left-censored, the change in vRNA is calculated as the LLoQ minus baseline vRNA, and is also left-censored.

Linear regression using software designed to handle censored data (known as tobit regression) is a possible method. Using this approach, adjusting for baseline vRNA, the estimated difference between arms in mean change from baseline to Day 3 was -0.97 $\log_{10}$ copies/ml (95% confidence interval [CI]: -1.81, -0.13) favoring the active arm (Table 3C), and is somewhat larger than the differences in mean change by either imputation approach (Table 3B). At Day 7, the difference in mean change from baseline was -1.36 $\log_{10}$ copies/ml, also favoring the active arm (95% CI: -2.31, -0.41), which is much larger than differences observed by either imputation approach, illustrating the potential bias using those methods when the proportion with vRNA <LLoQ increases. We didn't pursue an analysis of mean changes to Day 14 using tobit regression because of the high level of censoring (~90%) and hence the inability to check model assumptions.

As with standard linear regression, there is an assumption that the errors in the model are normally distributed. These errors are estimated by the residuals calculated as the observed vRNA value minus the predicted model value. The distributional assumption can be evaluated with quantile-quantile (Q-Q) plots, comparing the quantiles of the observed distribution of the residuals (calculated using Kaplan-Meier methods to account for censored residuals) against the corresponding quantiles of a standard normal distribution. If the assumption was satisfied,

the plots would show linear associations. Figure 2 shows Q-Q plots for the distribution of standardized residuals from models for change from baseline, adjusting for baseline. For the models of change from baseline to Days 3 and 7, the Q-Q plots appear reasonably linear, supporting normality assumptions. We note, however, the more restricted range of the Q-Q plot for changes to Day 7, as shown by the lack of standardized residuals below -1. This reflects the higher proportion of censored values at Day 7; thus, the normality assumption cannot be verified for the tail of the distribution, corresponding to large negative changes from baseline.

4. Quantile regression as an alternative distribution-free method

An alternative to tobit regression is quantile regression applied to assay-censored data, for example to model median change in vRNA. With this approach, there are no assumptions concerning the distribution of the errors in the model. However, there is an assumption that the median change has linear associations with continuous covariates in the model, including baseline vRNA.

At Day 3, the adjusted difference between arms in median change from baseline was -1.17 $\log_{10}$ copies/ml (95% CI: -2.42, 0.07) favoring the active arm. This is reasonably similar to the adjusted difference in mean change of -0.97 $\log_{10}$ copies/ml obtained from tobit regression, though estimated without making the assumption of normally distributed errors. There is a somewhat narrower CI for the difference in means, versus difference in medians, reflecting the gain in precision from assuming a normal distribution for the errors. At Day 7, the adjusted difference in median change was -0.96 $\log_{10}$ copies/ml, also favoring the active arm. However, it wasn't possible to obtain a CI from the numerical methods used to fit the model, due to the high

proportion of participants with vRNA <LLoQ at Day 7. At Day 14, the higher proportion with vRNA <LLoQ meant the difference in median change between arms couldn't be estimated.

Analyzing Repeated vRNA Over Time

5. Imputed values can affect estimates from MMRM due to correlation structure

Another strategy in several recent COVID-19 trials has been to use an MMRM with single-imputation for vRNA values <LLoQ [2–14]. These models estimate the difference in mean vRNA change in each arm at each timepoint, in a similar manner as linear regression models fit separately by timepoint. However, MMRMs incorporate a stronger assumption about the distribution of errors across timepoints, specifically that they follow a multivariate normal distribution with a specified correlation structure. Using this assumption, a global test evaluating the null hypothesis of no difference between arms in vRNA change at any timepoint can be undertaken. The stronger assumption may provide improved precision in estimating the differences in mean change at each timepoint by borrowing information between timepoints. However, this assumption may not be appropriate when using singly-imputed values for measurements <LLoQ as the correlation structure is affected by imputation. As an example, participants with vRNA <LLoQ at Days 7 and 14 will have identical imputed changes at both timepoints leading to higher correlations of errors in the model, than if the actual values <LLoQ were observed. To illustrate the impact of this, Table 3E shows results from MMRMs for changes from baseline to Days 3, 7, and 14. Compared with the estimates from models fitted separately at each timepoint (Table 3B), the borrowing of information through the correlation structure leads to smaller estimated differences in mean change between arms, particularly at Day 3 and to a lesser extent at Day 7 for both imputation approaches. This attenuation is driven by including Day 14, where ~90% of participants had vRNA <LLoQ; removal of this timepoint

from the MMRM reduces the magnitude of the attenuation (Table 3F). The estimates remain biased, however, for the same reasons as those obtained from separate regression models at each timepoint.

Extensions to MMRM that account for censored data exist (also known as linear mixed effects models for censored responses[LMEC]), but still require the multivariate normality assumption[23,24]. A caveat with these models is that they can be difficult to implement in standard statistical software, especially as the number of timepoints increases. Estimated differences between arms in mean change from baseline to Days 3 and 7 from LMEC are shown in Table 3G. The estimates are similar to those from the tobit regression models fitted separately at Days 3 and 7 (Table 3C). The stronger multivariate normal assumption leads to small gains in precision at Day 7 as seen by the narrower CI, though the gain at Day 3, where there's less censoring, is negligible. As with the separate regression models, we didn't pursue LMEC over the three days, as the high level of censoring at Day 14 meant that a normality assumption couldn't be reasonably verified.

Analyzing Proportion of Participants with vRNA <LLoQ Over Time

6. *Strategies that don't rely on quantitative values may be preferred with large % <LLoQ*

When there is a high proportion of participants with vRNA <LLoQ at one or more timepoints, it may be more appropriate to focus on how this proportion changes with time. This could be analyzed over time using log-binomial models fit using generalized estimating equations (GEE). However, due to problems with numerical algorithms, in ACTIV-2 we used Poisson regression models modified for binary outcomes [25] fit using GEEs with independence working correlation structure and robust standard errors, adjusting for baseline vRNA. When implementing this

model across the three days, the proportion with vRNA <LLoQ didn't differ between arms (Supplemental Table 2). When excluding the Day 14 measurements, where ~90% of participants had vRNA <LLoQ, the results for Days 3 and 7 were almost identical, confirming this method isn't sensitive to including timepoints with high proportions <LLoQ. This strategy can lead to loss in statistical power compared to analyses of quantitative vRNA, so is best reserved for when high proportions of participants are expected to have vRNA <LLoQ at one or more timepoints. However, there is also no need to restrict the analysis population to participants with vRNA ≥LLoQ, potentially providing more comprehensive analyses of qualitative vRNA in the overall study population.

DISCUSSION

In this paper we summarize methods commonly used in outpatient COVID-19 therapeutic trials for analyzing quantitative changes in SARS-CoV-2 RNA over time, and through an illustrative example from the ACTIV-2 study, highlight potential pitfalls. In ACTIV-2, our primary virology analyses focused on comparing the proportion of participants with vRNA <LLoQ over time, and examined vRNA levels rather than changes. As the pandemic has evolved and we have learned more about viral trajectories and variability, so has our thinking about the best analytic strategy. Since designing ACTIV-2, we have implemented exploratory analyses examining treatment effects on changes in vRNA over time using tobit regression models with adjustment for baseline RNA, restricted to participants with baseline vRNA ≥LLoQ, a method we advocate for in this paper [19,26,27].

In our illustrative example, the primary focus was on the population with quantifiable vRNA at baseline, which has been a focus in recent COVID-19 studies. This was reasonable in our

analysis as none who were <LLoQ at baseline had quantifiable vRNA at later timepoints. Including these individuals in analyses using imputed values would have led to imputed changes of zero and likely attenuation of the estimated mean changes. Regression analyses for censored data are more complex if such individuals are included, requiring strong, unverifiable assumptions about the distribution of vRNA changes over time among those with baseline vRNA <LLoQ. Looking more broadly across the study population in phase II placebo-controlled evaluations in ACTIV-2 (N=1565 enrolled with a median of 6 days from symptom onset), we observed that only 14% (of 287) of those with vRNA <LLoQ at baseline later had quantifiable vRNA. As new studies are developed, potentially with enrollment closer to onset of symptoms, the decision to exclude those <LLoQ at baseline should be carefully scrutinized, as doing so could remove individuals on an upward viral load trajectory and we lack understanding of these trajectories in the setting of vaccination, reinfection, and emergent variants. At a minimum, documenting viral shedding changes among participants with baseline vRNA <LLoQ is important, and analyses stratified by level (<LLoQ and ≥LLoQ at baseline) might be pursued.

The methods considered in this paper aren't exhaustive of imputation or modeling strategies, but were chosen to align with methods from recent publications of COVID-19 trials. We focus on single-imputation, and don't evaluate the performance of multiple-imputation strategies, which are more complicated and rely on distributional assumptions to support the imputation, but may reduce potential biases with imputation highlighted in this paper [28,29]. We also haven't evaluated the statistical performance of these methods through formal simulation studies, which may add further insights to benefits or downsides of the analytic strategies, particularly when high proportions of participants have vRNA <LLoQ during follow-up, where verification of model assumptions becomes more difficult. We also haven't considered potential biases due to missing data, for example, missingness arising due to hospitalization, if hospitalized participants

have higher vRNA levels. In designing studies, the impact on power and precision in estimating treatment effects needs consideration [30]. Finally, analysis of vRNA changes among participants with baseline levels above a threshold (e.g., the LLoQ) leads to estimated mean changes within each arm that are affected by regression to the mean, though estimated differences in mean changes between randomized arms are not. Despite these limitations, our paper highlights key issues and considerations when analyzing SARS-CoV-2 RNA data from outpatient treatment trials. These methods aren't only applicable in the COVID-19 setting, but should be considered when analyzing any biomarker that is measured with an assay with an LLoQ.

Recommendations

The best practices in analyzing SARS-CoV-2 RNA from outpatient trials depend on the number of timepoints and proportion of results <LLoQ. Regardless of the planned analysis, some key details should be reported to facilitate interpretation.

1. Provide sufficient details of the RT-qPCR assay, including the LLoQ.
2. Explain who is included in the analysis, such as via a CONSORT-type diagram (see Supplemental Figure 1), including an accounting of missing data and the reasons for missing (e.g., death, hospitalization, loss to follow-up, sample not obtained, sample processing/shipping issue).
3. If restricting the analysis population to those with quantifiable baseline vRNA, describe outcomes among those with vRNA <LLoQ.
4. Although we don't recommend the use of single-imputation, if used, the choice of imputed values should be provided, and implications for interpretation of results discussed.

5. Include descriptive summaries of vRNA by treatment arm and timepoint. We suggest including two figures (see Figure 1): distributions of quantitative levels (e.g., box and whisker plots) and distribution of vRNA categories (e.g., <LLoQ versus ≥LLoQ).

Analytic strategies to estimate differences between arms we recommend:

1. Methods that address censoring without imputation, such as tobit or median regression, or LMEC [23,24] be prioritized. But with increased censoring:
    a. Normality assumptions underlying regression analysis for censored data cannot be evaluated over the full range of the distribution, and dropping timepoints with high levels of censoring from analysis may be appropriate.
    b. Differences in medians (and their CIs) between arms might not be estimable from quantile regression.
2. Alternatively, consider non-parametric tests to analyze quantitative vRNA, such as the censored version of the Wilcoxon test (Gehan-Wilcoxon) which is implementable in standard software as a stratified test to account for baseline vRNA.
3. Comparing the proportion of participants with vRNA <LLoQ between arms over time may be preferred if there are high amounts of censoring.
4. With early, frequent measurements (e.g., daily), more complex extensions of LMEC that evaluate viral dynamics (e.g., estimating initial increases and subsequent vRNA decay) [20,31–34], or time-to-viral clearance via methods for time-to-event data [4–7,10,35,36] might be used.

ACKNOWLEDGMENTS

We thank the study participants, site staff, site investigators, and the entire ACTIV-2/A5401 study team; the AIDS Clinical Trials Group, including Lara Hosey, Jhoanna Roa, and Nilam Patel; the ACTG Laboratory Center, including Grace Aldrovandi and William Murtaugh; Frontier Science, including Marlene Cooper, Howard Gutzman, and Kevin Knowles; the Harvard Center for Biostatistics in AIDS Research (CBAR) and ACTG Statistical and Data Analysis Center (SDAC) including Ronald Bosch and Linda Harrison; the National Institute of Allergy and Infectious Diseases (NIAID) / Division of AIDS (DAIDS); Bill Erhardt; the Foundation for the National Institutes of Health and the Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV) partnership, including Stacey Adams; and the PPD clinical research business of Thermo Fisher Scientific. We also thank AstraZeneca for partnership in the ACTIV-2 study and for donation of study drug for the cohort used as the illustrative example in this analysis.

REFERENCES

1. COVID-19: Developing Drugs and Biological Products for Treatment or Prevention. **2021**; .

2. Weinreich DM, Sivapalasingam S, Norton T, et al. REGEN-COV Antibody Combination and Outcomes in Outpatients with Covid-19. N Engl J Med. Massachusetts Medical Society; **2021**; 385(23):e81.

3. Weinreich DM, Sivapalasingam S, Norton T, et al. REGN-COV2, a Neutralizing Antibody Cocktail, in Outpatients with Covid-19. New England Journal of Medicine. Massachusetts Medical Society; **2021**; 384(3):238–251.

4. Dougan M, Azizad M, Mocherla B, et al. A randomized, placebo-controlled clinical trial of bamlanivimab and etesevimab together in high-risk ambulatory patients with COVID-19 and validation of the prognostic value of persistently high viral load. Clin Infect Dis. **2021**; :ciab912.

5. Dougan M, Nirula A, Azizad M, et al. Bamlanivimab plus Etesevimab in Mild or Moderate Covid-19. New England Journal of Medicine. Massachusetts Medical Society; **2021**; 385(15):1382–1392.

6. Dougan M, Azizad M, Chen P, et al. Bebtelovimab, alone or together with bamlanivimab and etesevimab, as a broadly neutralizing monoclonal antibody treatment for mild to moderate, ambulatory COVID-19 [Internet]. medRxiv; 2022 [cited 2023 Feb 9]. p. 2022.03.10.22272100. Available from: https://www.medrxiv.org/content/10.1101/2022.03.10.22272100v1

7. Fischer WA, Eron JJ, Holman W, et al. A phase 2a clinical trial of molnupiravir in patients with COVID-19 shows accelerated SARS-CoV-2 RNA clearance and elimination of infectious virus. Science Translational Medicine. American Association for the Advancement of Science; **2021**; 14(628):eabl7430.

8. Montgomery H, Hobbs FDR, Padilla F, et al. Efficacy and safety of intramuscular administration of tixagevimab–cilgavimab for early outpatient treatment of COVID-19 (TACKLE): a phase 3, randomised, double-blind, placebo-controlled trial. The Lancet Respiratory Medicine. Elsevier; **2022**; 10(10):985–996.

9. Gupta A, Gonzalez-Rojas Y, Juarez E, et al. Effect of Sotrovimab on Hospitalization or Death Among High-risk Patients With Mild to Moderate COVID-19: A Randomized Clinical Trial. JAMA. **2022**; 327(13):1236–1246.

10. Jagannathan P, Andrews JR, Bonilla H, et al. Peginterferon Lambda-1a for treatment of outpatients with uncomplicated COVID-19: a randomized placebo-controlled trial. Nat Commun. Nature Publishing Group; **2021**; 12(1):1967.

11. Alemany A, Millat-Martinez P, Corbacho-Monné M, et al. High-titre methylene blue-treated convalescent plasma as an early treatment for outpatients with COVID-19: a randomised, placebo-controlled trial. The Lancet Respiratory Medicine. Elsevier; **2022**; 10(3):278–288.

12. Reis G, Moreira Silva EAS, Medeiros Silva DC, et al. Early Treatment with Pegylated Interferon Lambda for Covid-19. New England Journal of Medicine. Massachusetts Medical Society; **2023**; 388(6):518–528.

13. Ramakrishnan S, Nicolau DV, Langford B, et al. Inhaled budesonide in the treatment of early COVID-19 (STOIC): a phase 2, open-label, randomised controlled trial. The Lancet Respiratory Medicine. Elsevier; **2021**; 9(7):763–772.

14. Mitjà O, Corbacho-Monné M, Ubals M, et al. Hydroxychloroquine for Early Treatment of Adults with Mild Covid-19: A Randomized-Controlled Trial. Clin Infect Dis. **2020**; :ciaa1009.

15. Hammond J, Leister-Tebbe H, Gardner A, et al. Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. New England Journal of Medicine. Massachusetts Medical Society; **2022**; 386(15):1397–1408.

16. Jayk Bernal A, Gomes da Silva MM, Musungaie DB, et al. Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. New England Journal of Medicine. Massachusetts Medical Society; **2022**; 386(6):509–520.

17. Gottlieb RL, Vaca CE, Paredes R, et al. Early Remdesivir to Prevent Progression to Severe Covid-19 in Outpatients. New England Journal of Medicine. Massachusetts Medical Society; **2022**; 386(4):305–315.

18. Mukae H, Yotsuyanagi H, Ohmagari N, et al. Efficacy and Safety of Ensitrelvir in Patients With Mild-to-Moderate Coronavirus Disease 2019 (COVID-19): The Phase 2b Part of a Randomized, Placebo-Controlled, Phase 2/3 Study. Clinical Infectious Diseases. **2022**; :ciac933.

19. Marschner IC, Betensky RA, DeGruttola V, Hammer SM, Kuritzkes DR. Clinical trials using HIV-1 RNA-based primary endpoints: statistical analysis and potential biases. J Acquir Immune Defic Syndr Hum Retrovirol. **1999**; 20(3):220–227.

20. Chew KW, Moser C, Daar ES, et al. Antiviral and clinical activity of bamlanivimab in a randomized trial of non-hospitalized adults with COVID-19. Nat Commun. Nature Publishing Group; **2022**; 13(1):4931.

21. Bender Ignacio RA, Chew KW, Moser C, et al. Tixagevimab/Cilgavimab IM and IV for COVID-19: A Randomized Controlled ACTIV-2 Trial. Seattle, WA: Conference on Retroviruses and Opportunistic Infections; 2023.

22. Berg MG, Zhen W, Lucic D, et al. Development of the RealTime SARS-CoV-2 quantitative Laboratory Developed Test and correlation with viral culture as a measure of infectivity. J Clin Virol. **2021**; 143:104945.

23. Hughes JP. Mixed effects models with censored data with application to HIV RNA levels. Biometrics. **1999**; 55(2):625–629.

24. Vaida F, Liu L. Fast Implementation for Normal Mixed Effects Models With Censored Response. J Comput Graph Stat. **2009**; 18(4):797–817.

25. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol. **2004**; 159(7):702–706.

26. Taiwo BO, Chew KW, Moser C, et al. Phase 2 safety and antiviral activity of SAB-185, a novel polyclonal antibody therapy for non-hospitalized adults with COVID-19. J Infect Dis. **2023**; :jiad013.

27. Moser C, Li JZ, Eron JJ, et al. Predictors of SARS-CoV-2 RNA From Nasopharyngeal Swabs and Concordance With Other Compartments in Nonhospitalized Adults With Mild to Moderate COVID-19. Open Forum Infect Dis. **2022**; 9(11):ofac618.

28. Uh H-W, Hartgers FC, Yazdanbakhsh M, Houwing-Duistermaat JJ. Evaluation of regression methods when immunological measurements are constrained by detection limits. BMC Immunol. **2008**; 9:59.

29. Lubin JH, Colt JS, Camann D, et al. Epidemiologic Evaluation of Measurement Data in the Presence of Detection Limits. Environ Health Perspect. **2004**; 112(17):1691–1696.

30. Hughes MD. Analysis and design issues for studies using censored biomarker measurements with an example of viral load measurements in HIV clinical trials. Stat Med. **2000**; 19(23):3171–3191.

31. Watson JA, Kissler SM, Day NPJ, Grad YH, White NJ. Characterizing SARS-CoV-2 Viral Clearance Kinetics to Improve the Design of Antiviral Pharmacometric Studies. Antimicrob Agents Chemother. 66(7):e00192-22.

32. Perelson AS, Ke R. Mechanistic Modeling of SARS-CoV-2 and Other Infectious Diseases and the Effects of Therapeutics. Clin Pharmacol Ther. **2021**; 109(4):829–840.

33. Wang S, Pan Y, Wang Q, Miao H, Brown AN, Rong L. Modeling the viral dynamics of SARS-CoV-2 infection. Math Biosci. **2020**; 328:108438.

34. Ke R, Zitzmann C, Ho DD, Ribeiro RM, Perelson AS. In vivo kinetics of SARS-CoV-2 infection and its relationship with a person's infectiousness. Proc Natl Acad Sci U S A. **2021**; 118(49):e2111477118.

35. Hoffmann-La Roche. A Phase II Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Antiviral Activity, Safety, Pharmacokinetics, and Efficacy of RO7496998 (AT-527) in Non-Hospitalized Adult Patients With Mild or Moderate COVID-19 [Internet]. clinicaltrials.gov; 2022 Oct. Report No.: results/NCT04709835. Available from: https://clinicaltrials.gov/ct2/show/results/NCT04709835

36. Golan Y, Campos JAS, Woolson R, et al. Favipiravir in Patients With Early Mild-to-moderate Coronavirus Disease 2019 (COVID-19): A Randomized Controlled Trial. Clinical Infectious Diseases. **2023**; 76(3):e10–e17.

Table 1: Summary of Analytic Methods Considered in Our Illustrative Example for the Analysis of Changes from Baseline in SARS-CoV-2 RNA

Methods	No. of Timepoints	Handing Values <LLoQ	Pros	Caveats/Issues
Analysis of covariance (ANCOVA)/linear regression	1	Single Imputation	Easy to implement in standard software. With small proportion <LLoQ, impact of imputation is likely modest.	Using imputation results in biased estimates of differences between randomized arms in mean change. Normality assumption in model may be violated Need to restrict to those ≥LLoQ at baseline to calculate changes.
Linear regression for censored data (tobit regression)	1	Not Required	Easy to implement in standard software. Analyses considering censored measurements avoids bias that may be created by using imputed	Normality assumption in model cannot be confirmed when large proportion of data are censored. Need to restrict to those ≥LLoQ at baseline to calculate changes.

			values.	
Median regression for censored data	1	Not Required	Easy to implement in standard software. Distribution free model removes assumptions about distribution of the errors.	Model cannot be fitted when large proportion of data are censored. Need to restrict to those ≥LLoQ at baseline to calculate changes.
Mixed models for repeated measures (MMRM)	> 1	Single Imputation	Easy to implement in standard software. Global test of, no difference between randomized arms across timepoints, can be easily generated.	Using imputation results in biased estimates of the difference between randomized arms in mean change, with the bias at one time dependent on the proportion <LLoQ at other times (as information is shared among times through an assumed correlation structure). Multivariate normality assumption may be violated. Need to restrict to those ≥LLoQ at baseline to calculate changes.
MMRM for Censored Data (Linear mixed effects models for censored data,	> 1	Not Required	Analyses considering censored measurements avoids bias that may be	Increase complexity in implementing model in standard software as the number of timepoints increases. Multivariate normality assumption difficult to verify,

LMEC)			created by using imputed values. Global test of, no difference between randomized arms across timepoints, can be easily generated. Possible improved precision by sharing information over timepoints through an assumed model.	particularly when large proportion of data are censored at one or more times. Need to restrict to those ≥LLoQ at baseline to calculate changes.
Binary Regression	≥ 1	Not Required	Easy to implement in standard software. Includes all participants, regardless of baseline value. Estimation of treatment effects not influenced by the	Loss of statistical power when dichotomizing outcome from continuous variable to a binary variable.

			proportion <LLoQ.	

LLoQ = Lower Limit of Quantification; ANCOVA = Analysis of Covariance; MMRM = Mixed Model Repeated Measures; LMEC = Linear Mixed Effects Models with Censored Response

Table 2: Distribution of SARS-CoV-2 RNA by Study Visit in each Treatment Arm in overall cohort (A) and among those with vRNA ≥LLoQ at Baseline/Day 0 (B)

A: All participants in cohort (N=114)			
Visit		Active (N=58)	Placebo (N=56)
Baseline	Median (quartiles)	4.0 (<LLoQ, 6.6)	3.0 (<LLoQ, 5.9)
	<LLoQ, n (%)	14 (28)	19 (38)
	Missing, n	9	6
Day 3	Median (quartiles)	<LLoQ (<LLoQ, 3.9)	<LLoQ (<LLoQ, 3.9)
	<LLoQ, n (%)	26 (52)	24 (52)
	Missing, n	8	10
Day 7	Median (quartiles)	<LLoQ (<LLoQ, 2.2)	<LLoQ (<LLoQ, 2.2)
	<LLoQ, n (%)	37 (74)	35 (71)
	Missing, n	8	7
Day 14	Median (quartiles)	<LLoQ (<LLoQ, <LLoQ)	<LLoQ (<LLoQ, <LLoQ)
	<LLoQ, n (%)	49 (98)	45 (97)
	Missing, n	11	7
B: All participants with vRNA ≥LLoQ at baseline (N=66)			
Visit		Active (N=35)	Placebo (N=31)
Baseline	Median (quartiles)	5.5 (3.7, 8.0)	5.0 (3.1, 6.7)
	<LLoQ, n (%)	0 (0)	0 (0)
	Missing, n	0	0
Day 3	Median (quartiles)	3.0 (<LLoQ, 4.5)	3.4 (<LLoQ, 5.9)
	<LLoQ, n (%)	8 (27)	7 (28)
	Missing, n	5	6
Day 7	Median (quartiles)	<LLoQ (<LLoQ, 2.5)	<LLoQ (<LLoQ, 3.3)
	<LLoQ, n (%)	18 (62)	14 (54)

	Missing, n	6	5
Day 14	Median (quartiles) <LLoQ, n (%) Missing, n	<LLoQ (<LLoQ, <LLoQ) 27 (93) 6	<LLoQ (<LLoQ, <LLoQ) 24 (89) 4

LLoQ = Lower Limit of Quantification

Table 3: Differences between Treatment Arms in SARS-CoV-2 RNA (log_{10} copies/ml) change from baseline– Mean/Median[a], 95% CI and p-value among those with quantifiable baseline vRNA

Imputation	Day 3	Day 7	Day 14
A: Linear regression model with imputation, separate model by day – unadjusted			
LLoQ imputation	-0.87 (-1.70, -0.06) p=0.037	-0.82 (-1.79, 0.15) p=0.09	-0.25 (-1.30, 0.81) p=0.64
½LLoQ imputation	-0.86 (-1.69, -0.04) 0.041	-0.90 (-1.82, 0.01) p=0.053	-0.29 (-1.32, 0.74) p=0.58
B: Linear regression model with imputation, separate model by day – adjusted for baseline			
LLoQ imputation	-0.74 (-1.41, -0.06) p=0.034	-0.56 (-1.01, -0.11) p=0.015	-0.06 (-0.18, 0.07) p=0.38
½LLoQ imputation	-0.77 (-1.53, 0.002) 0.050	-0.69 (-1.29, -0.09) p=0.024	-0.11 (-0.37, 0.16) p=0.42
C: Linear regression model for censored data (tobit regression), separate model by day – adjusting for baseline			
N/A	-0.97 (-1.81, -0.13) p=0.023	-1.36 (-2.31, -0.41) p=0.005	Not Obtained[b]
D: Median regression model for censored data, separate model by day – adjusting for baseline			
N/A	-1.17 (-2.42, 0.07) p=0.07	-0.96 (NE, NE) NE	NE
E: MMRM across all three days (Day 3, 7 and 14) with imputation – adjusting for baseline			
LLoQ imputation	-0.39 (-1.23, 0.45) p=0.36	-0.49 (-0.95, -0.04) p=0.032	-0.07 (-0.20, 0.06) p=0.27
½LLoQ imputation	-0.52 (-1.44, 0.40)	-0.60 (-1.21, 0.01)	-0.13 (-0.40, 0.14)

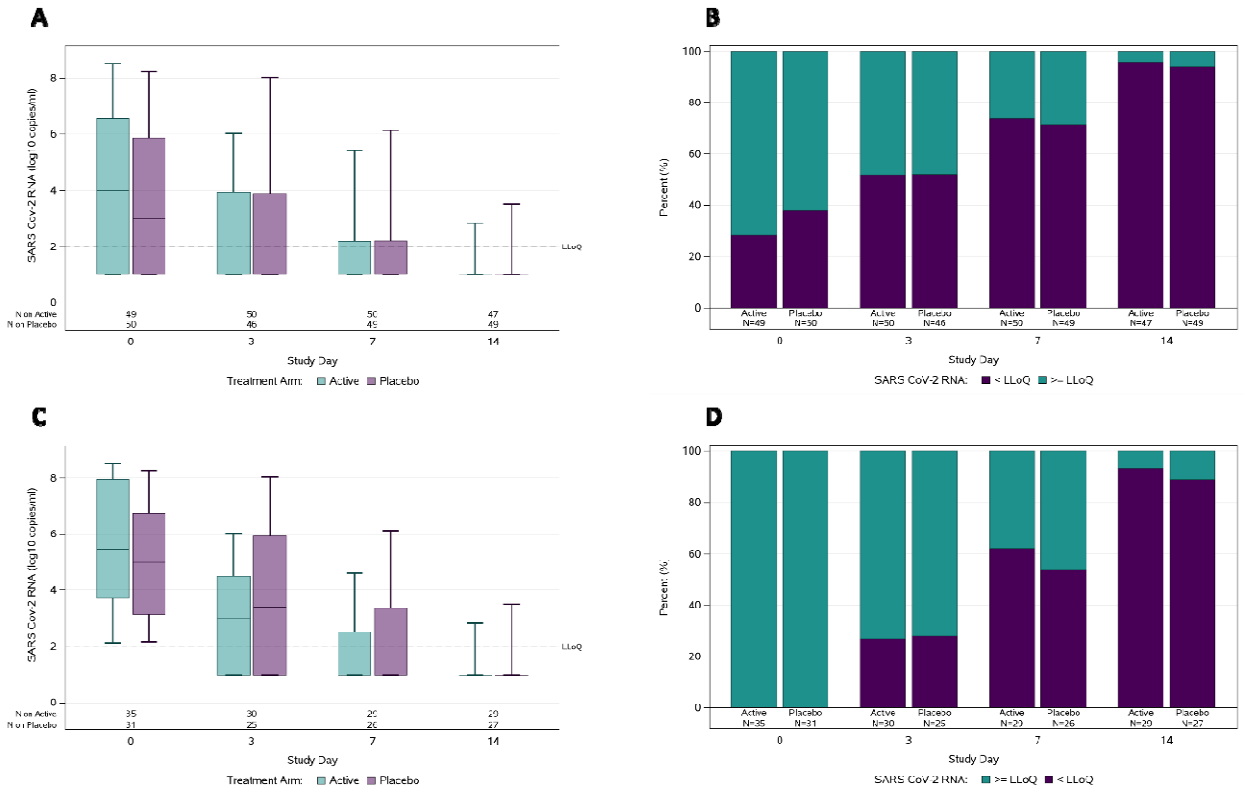
	p=0.26	p=0.052	p=0.33
F: MMRM across Days 3 and 7 with imputation – adjusting for baseline			
LLoQ imputation	-0.65 (-1.36, 0.07) p=0.08	-0.58 (-1.01, -0.15) p=0.009	- -
½LLoQ imputation	-0.72 (-1.50, 0.06) p=0.07	-0.71 (-1.29, -0.13) p=0.018	- -
G: MMRM for censored data across Days 3 and 7– adjusting for baseline			
N/A	-1.10 (-1.94, -0.26) p=0.011	-1.33 (-2.23, -0.43) p=0.004	- -

[a]Differences in mean change provided except for (D), which is difference in median change.

[b]Results are not shown at Day 14 for the linear regression model for censored data because model assumptions cannot be reasonably verified due to the high level of censoring at Day 14.

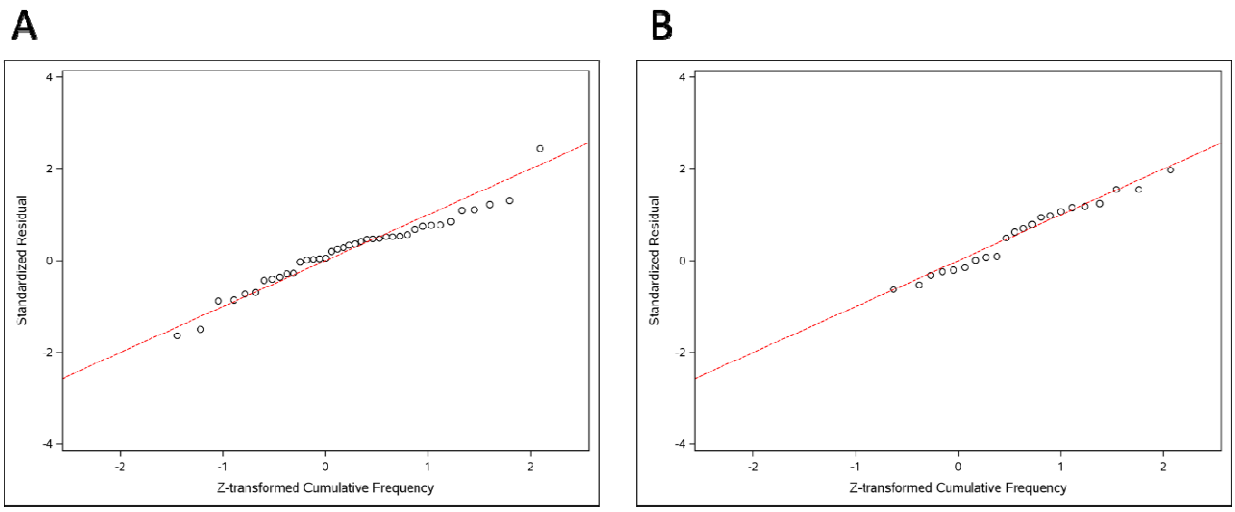
LLoQ = Lower Limit of Quantification; N/A = Not Applicable; NE = Not Estimable; MMRM = Mixed Model for Repeated Measures.

Figure 1: Distribution of SARS-CoV-2 RNA from nasopharyngeal swabs in Active and Placebo arms by study visit in overall cohort (A and B) and among those with vRNA ≥LLoQ at Baseline/Day 0 (C and D).



Levels of SARS-CoV-2 RNA (log_{10} copies/ml) with horizontal line = median, box=interquartile range, whiskers = minimum/maximum (A and C); results below the LLoQ are plotted using an imputed value of 1 log_{10} copies/ml. Proportion with quantifiable SARS-CoV-2 RNA (green) and unquantifiable (purple) (B and D). LLOQ = Lower Limit of Quantification.

Figure 2: Quantile-quantile (Q-Q) plot for linear regression model for censored data for change in vRNA from baseline to Day 3 (A) and to Day 7 (B), both models included an indicator variable for treatment versus placebo and adjusted for baseline vRNA



Standardized residuals (for the non-censored observations) calculated by dividing the residuals by their standard deviation (estimated from the fitted model). Quantiles for a standard normal distribution plotted on the x-axis take account of censored residuals.