

Tests of BDL Bias and Variance - v.1a

James R. Hunter, D.Sc.
Laboratório de Retrovirologia, UNIFESP

2026-01-04

Purpose

This file will test the determination of bias and variance when calculating viral loads and other variables that are subject to detection limits. For HIV-1 patients, the objective of antiretroviral treatment is to make their plasma RNA viral load zero.

However, there is a practical limit below which the assays cannot distinguish between the presence of virus and having no virus (the zero objective). This is called the “detection limit” and patients who achieve this level are said to have a viral load “BDL”, below detection limit. One of the most important consequences of exactitude in measuring true lack of the presence of virus is the phrase “undetectable = untransmittable” [G. C. Chang et al. [1]][2]. This frees, at least emotionally, the individual from concern about transmitting the virus to a partner.

There have been various strategies to deal with calculations of viral load in a panel of patients. One has been to eliminate the cases from the calculation, another has been to treat the viral level of those patients who are BDL as 0, assuming they are really without any virus. Still another strategy has been to assign a value equal to the detection limit. Others have used a value of half the detection limit, i.e., half of the difference between the detection limit and zero.

Name	Strategy	Value Assigned to “BDL”
censor	eliminate case	no value
zero	treat as zero	0
dl	treat as detection limit	detection limit
halfdl	treat as half of detection limit	detection limit/2

This study was inspired by a blog post by Nikolas Siccha, “Lower Limits of Detection or Quantification”. [3] He develops a Bayesian model to explore the pharmacodynamics of drugs that cause the measurement of interest to fall below detection limits. As it deals with the dynamics over time of a drug’s presence and efficacy in a patient, his question is more complex than the question I would like to address here.

! Important

What is the relationship between the true value of viral load below detection limit and the strategies used to estimate it and how does that effect analyses of the viral load of a sample or population?

Method

I will create a panel of 1,000 PLWH artificially with estimates of their viral load at a baseline and after 48 simulated weeks of antiretroviral treatment (ART) when some of them will have viral loads reduced to BDL, which I will define here as 50 copies per mL.¹ According to a 2009 clinical trial (STARTMRK) of Raltegravir, a 1st generation integrase inhibitor, 86.1% of the RTG arm of naive patients had viral loads BDL by week 48. [4] Likewise, a clinical trial (SINGLE) of a second generation integrase inhibitor, Dolutagravir, showed 88% of the DTG arm with viral loads below detection limit by week 48. [5] For my panel, I will use a proportion of 85% of patients achieving true viral loads between 0 and 50 by the 48 week mark. 85% is a bit more conservative than either of the real trials, but consistent with their results.

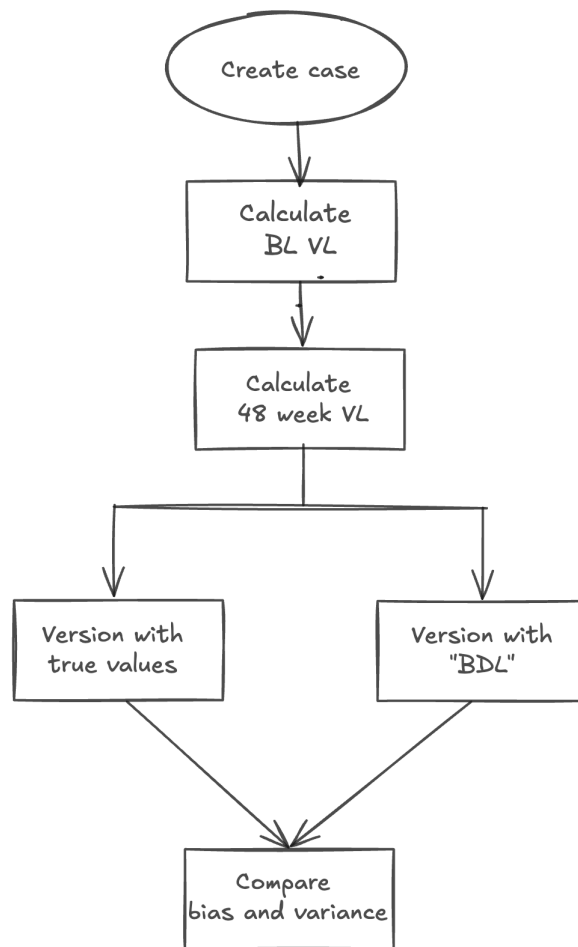
The baseline viral loads will be determined by calculating 1,000 normal variates from a truncated normal distribution with a mean of 900 copies per mL and a standard deviation of 1,000 copies. Since values below detection limit make no sense for this simulation, these values will be removed by using the truncated normal distribution.[6]

The 48 week data will be reduced dramatically (and unrealistically)² by 97.5% and the number of values below 50 copies per mL will be calculated. I will then create a second version substituting the true values below the detection limit with BDL. Using the four strategies outlined in Table 1, I will then compare the bias and variance of each of these strategies for substituting the true value of the viral load. Finally, I will propose a method for imputing values that is commonly applied in studies using the R statistical language.

The flow chart below shows a schematic of the workflow of this study.

¹While 50 copies was a common level for detection limit until 2024, since then it has been reduced to 30 or 20 copies depending on the measurement technology used. I am using 50 here as the purpose is to explore the bias and variance of samples rather than to test the detection limit itself.

²While unrealistic, this large reduction is useful to give us a number of values BDL that can be tested against the 4 strategies.



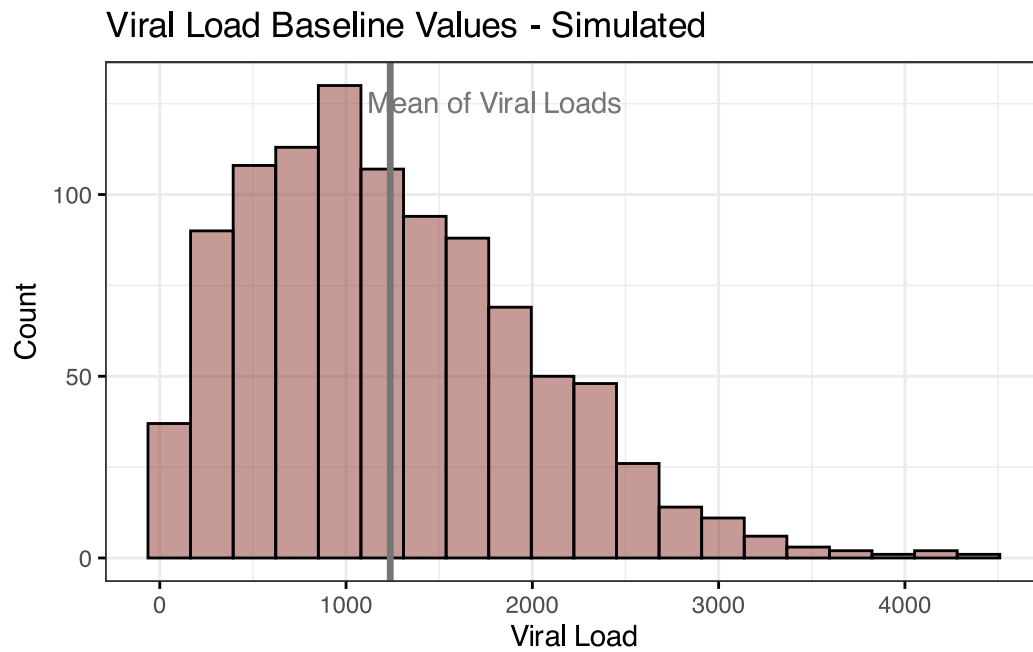
Baseline

Descriptive Statistics

bl

N: 1000

	bl
-----	-----
Mean	1237.22
Std.Dev	753.15
Min	51.00
Median	1116.50
Max	4395.00
IQR	1058.25
CV	0.61



Week 48 Viral Load

We will calculate the viral load at 48 weeks to be 97.5% less than the original viral load.

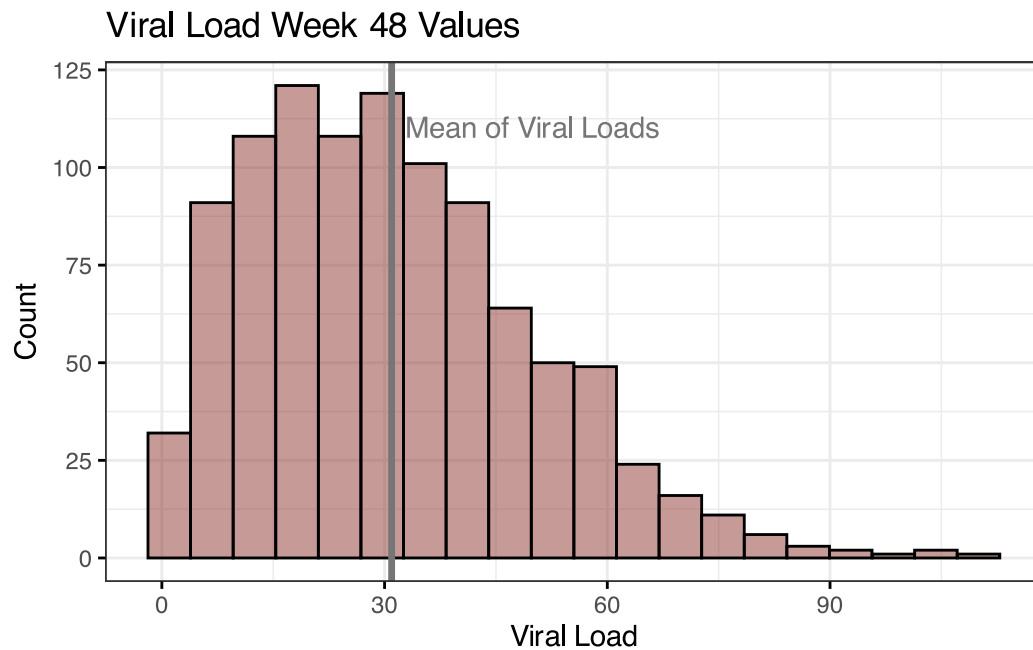
Descriptive Statistics

vldf\$w48

N: 1000

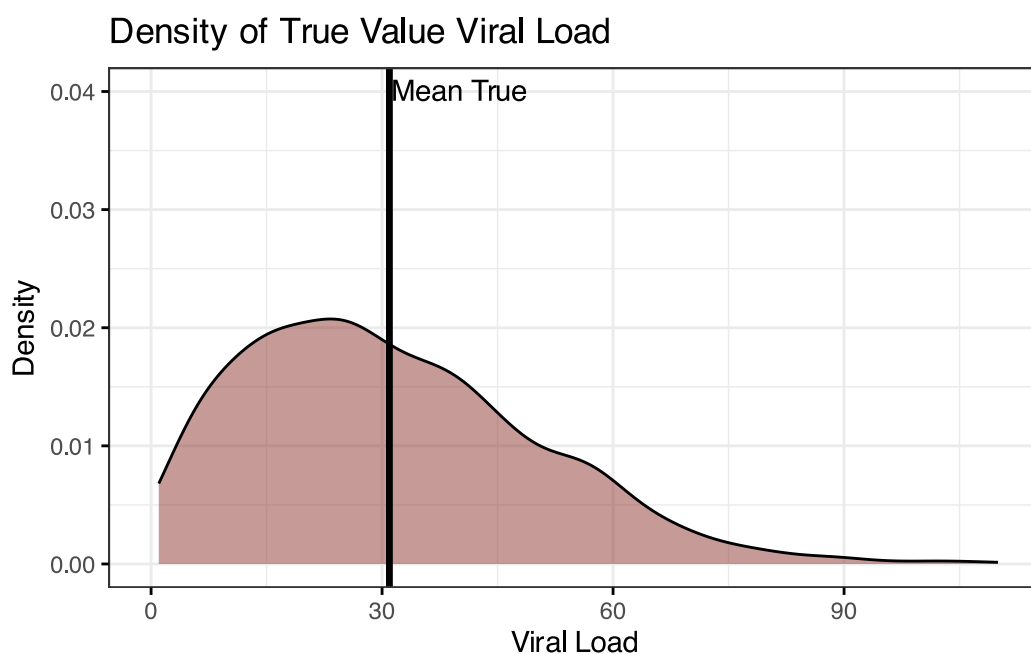
	w48

Mean	30.95
Std.Dev	18.84
Min	1.00
Median	28.00
Max	110.00
IQR	27.00
CV	0.61



To give you an idea of what we have now, here are the first 10 cases in our dataset. While we know the true values of the viral load, the assay itself will only give us a BDL signal instead of the value. This occurs 835 times out of the 1,000 cases. The question is what can we do about it and what does that do to the distribution of viral loads that we are trying to use to make an inference about HIV-1 treatment strategies.

	Week 48	
	True Value	BDL
Baseline		
2,271	57	57
335	8	bdl
1,263	32	bdl
1,533	38	bdl
1,304	33	bdl
794	20	bdl
2,412	60	60
805	20	bdl
2,918	73	73
837	21	bdl



Bias and Variance

The measures I am going to use to test the strategies listed at the beginning of this analysis are bias and variance. Bias measures how much the strategy distorts the distribution of true values. I will show this by comparing the mean, median, and standard deviation of the results of the strategies to those of the true values and through a density visualization of the distributions. The variance is rather more straightforward as we can find the statistical variance as determined by the classical formula for this quantity.

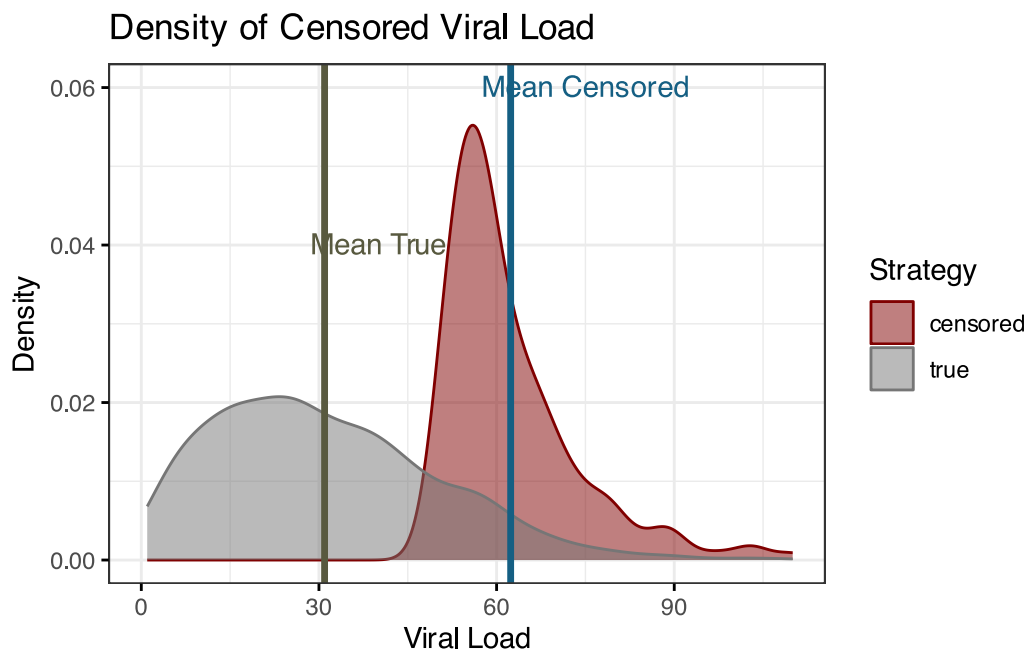
The Four Strategies

Censorship

In this strategy, which is very commonly applied when laboratories confront results of assays that do not indicate numerical values, simply removes those cases from the study.

When we do that here, we throw away 835 cases, meaning 83.5% of the cases, which likely contain many other nuggets of useful information. Before we even arrive at judging bias and variance, this reckless disposal of worthwhile information is the principal defect of the strategy. This dataset is not arbitrarily skewed toward having cases with BDL viral load after 48 weeks. As pointed out at the beginning, many studies show results in the range of 85%.

Week 48 Viral Load							
Strat- egy	Mean	Diff True Mean	Std. Dev.	Vari- ance	Median	# of Cases	Co- efficient Var.
true	30.95	0.00	18.84	355.02	28	1000	0.61
cen- sored	62.38	31.42	11.23	126.21	59	165	0.18



As you can see from the table and the density graph, the mean of the censored viral loads at 48 weeks is more than twice the true value as is the median value. That is because all values less than 50 were removed from the dataset because we did not know their true value; they were “BDL”. The censorship strategy leaves a highly biased dataset. The variance is greatly reduced, approximately a third the variance of the true values. Again, this is due to censoring—eliminating values within a certain range. Here the eliminated range is from the lowest value among the true values (1) up to 49.

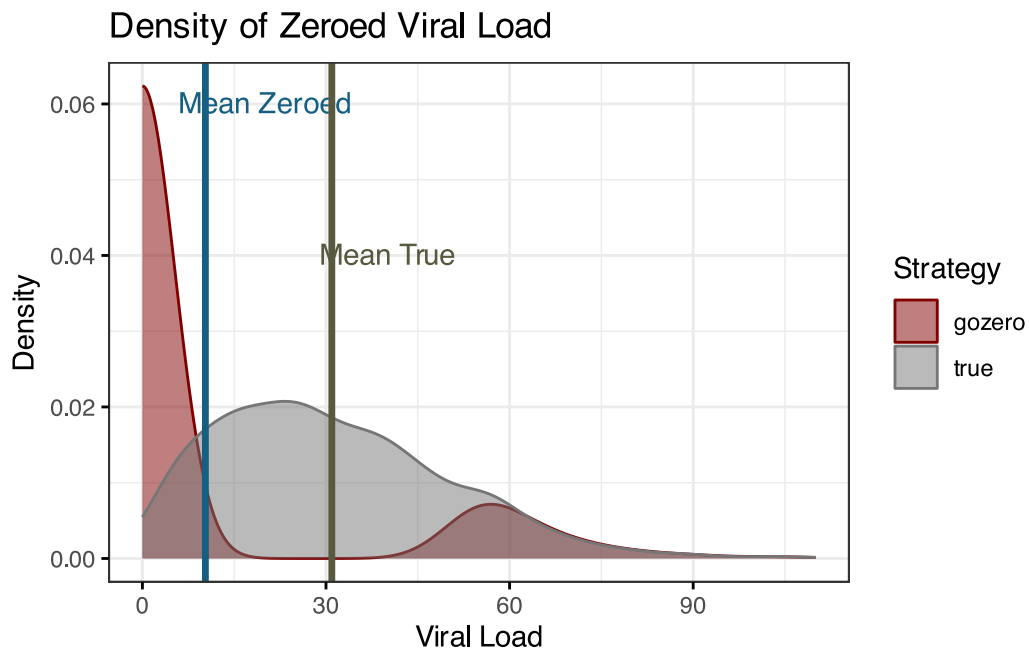
Treat as 0

This is a very attractive and tempting option for any research team. The objective of HIV treatment and cure projects is to eliminate the virus, drive it to zero. It would be very tempting to simply treat any viral load value that cannot be specified for being below the detection limit is to treat it as indistinguishable from zero. This has support in many cases by assuming that someone whose viral load is BDL is no longer communicable. The

tendency is to assume that if the viral load cannot be counted, it is because it is not there. However, we know that is not the case. Below Detection Limit does not substantively translate to a value of zero. So, this strategy as well comes with conceptual as well as calculation issues that argue against implementing it.

However, we need to test the implications of this strategy in terms of bias and variance as well. In this strategy, we will take all those values that are “bdl” in the dataset and substitute 0 for them.

Week 48 Viral Load							
Strat- egy	Mean	Diff True Mean	Std. Dev.	Vari- ance	Median	# of Cases	Co- efficient Var.
true	30.95	0.00	18.84	355.02	28	1000	0.61
cen- sored	62.38	31.42	11.23	126.21	59	165	0.18
Zero	10.29	-20.66	23.61	557.30	0	1000	2.29



Here, the mean of the distribution now reflects the 835 cases that have viral loads that cannot be accurately measured. Although we know that if we had been able to measure these values, they would have produced a mean of 30.95 and a variance of 355.02. However, because of the huge numbers of false zeros that have been inserted, the mean

has now fallen to 10.29. Because there are now the mass of zero values and a lesser number of high values, with few in the middle (the trough of the graph on the density graph), the variance has increased to 557.3 . Since the zero values represent such a large portion of the dataset, the median has fallen to 0.

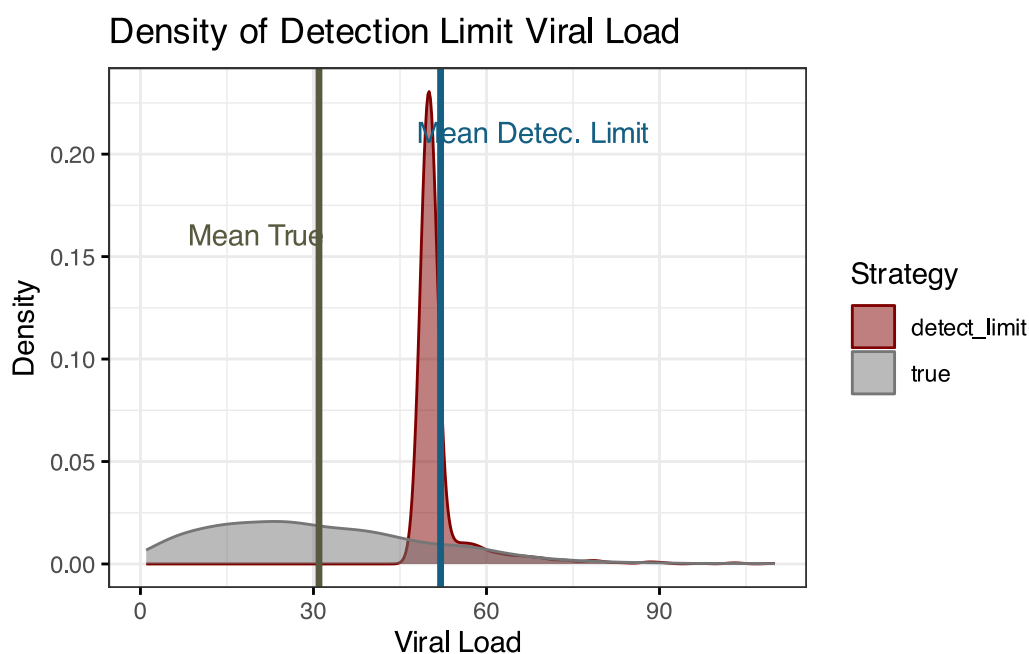
Thus, while tempting, this strategy also distorts dramatically the true nature of the data and bases medical and research decision-making on a false premise.

Treat as Detection Limit

Just as the prior strategy was optimistic, this strategy is pessimistic. In this strategy, we set all the BDL values to the detection limit itself, which we are setting at 50 copies for this analysis. However, it is similar in that it chooses a single, arbitrary value for all the BDL cases.

This too distorts the underlying true distribution of viral loads.

Week 48 Viral Load							
Strat- egy	Mean	Diff True Mean	Std. Dev.	Vari- ance	Median	# of Cases	Co- efficient Var.
true	30.95	0.00	18.84	355.02	28	1000	0.61
cen- sored	62.38	31.42	11.23	126.21	59	165	0.18
Zero	10.29	-20.66	23.61	557.30	0	1000	2.29
Detec- tion Limit	52.04	21.09	6.47	41.84	50	1000	0.12



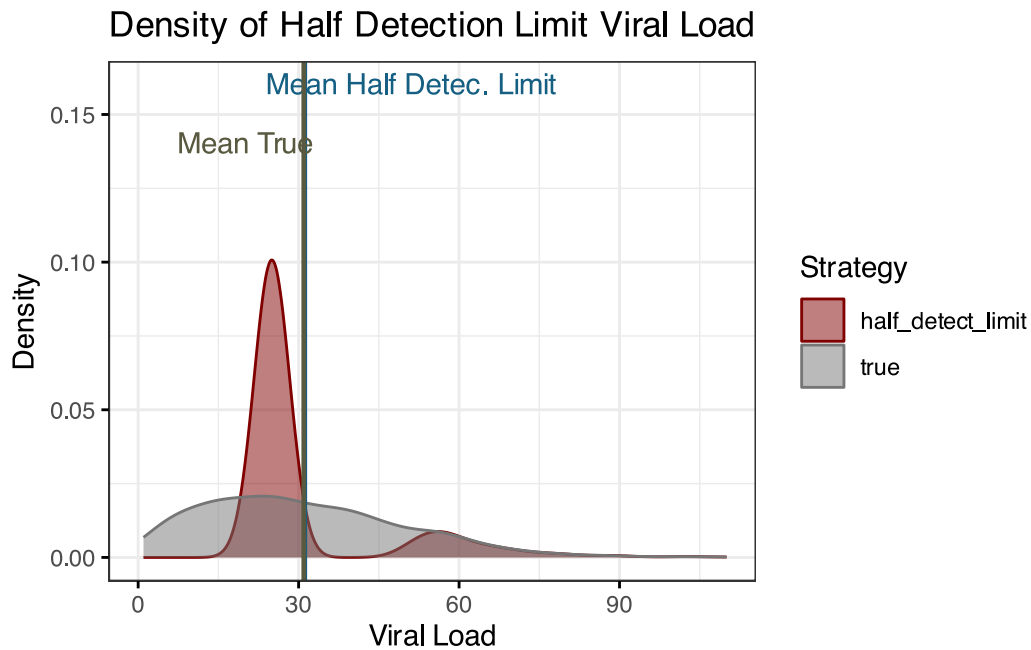
As with the censoring strategy, substituting the detection limit for BDL pushes the distribution of viral loads into a narrow range around that detection limit. As can be seen in the table above, the mean and the median are very close together and the variance now is less than half that of censored strategy, the lowest previous value. While the mean of this strategy is closer to the mean of the true values than the censoring strategy (a difference of 21.09 copies versus 31.42 copies), the concentration of values around 50 copies limits the accuracy and utility of this strategy.

Treat as Half of the Detection Limit

The final of these strategies that I have come across is that of treating the BDL values as one-half of the detection limit, in this case 25. I'm not sure what the justification for this is beyond being a specific value somewhere between zero and the detection limit. As with the other strategies, this arbitrariness leads to considerable bias in calculations and then inferences related to viral load.

Week 48 Viral Load							
Strat- egy	Mean	Diff True Mean	Std. Dev.	Vari- ance	Median	# of Cases	Co- efficient Var.
true	30.95	0.00	18.84	355.02	28	1000	0.61
cen- sored	62.38	31.42	11.23	126.21	59	165	0.18

Week 48 Viral Load							
Strat- egy	Mean	Diff True Mean	Std. Dev.	Vari- ance	Median	# of Cases	Co- efficient Var.
Zero	10.29	-20.66	23.61	557.30	0	1000	2.29
Detect- ion Limit	52.04	21.09	6.47	41.84	50	1000	0.12
Half De- tection Limit	31.17	0.21	14.61	213.38	25	1000	0.47



Empirically, this strategy does a relatively good job of estimating the distribution of the true values of the viral load distribution. The difference between the mean under this strategy and the true mean is very small (0). As with all the strategies that focus on a single value as a substitute for BDL, the density of the distribution is focused very narrowly around the half mean value with the only distant cluster being those viral loads that were greater than 50 initially. These appear as the concentration whose center is just below 60. This works well as the 835 values that are classed as BDL actually have a mean of 24.74 , a value also extremely close to the assigned value of 25. As the detection

limit declines, this approximation will become even closer. So, in terms of distortion of the data, this strategy offers an acceptable practical solution.

Imputation of Values Below Detection Limit

All of these four strategies have the defect of substituting a range of values by a single value, losing the variance inherent in what I have called here the true values. We are considering a single variable, isolated from covariates that describe the phenotype of the person whose viral load we are estimating. If we had a set of variables, we could estimate values for the censored BDL values using one of a number of systems of multiple imputation, such as multiple regression imputation, or multiple imputation by chained equations (MICE)³. [7]

However, in the case that we simply have viral load to consider, the only way we can “fill in the blanks” of viral loads below detection limit is to determine the distribution that describes the data and estimate the parameters of that distribution. Knowing these quantities, we can estimate a series of random values to stand in for the missing BDL values. The process of estimating which of the many known statistical distributions the empirical data is closest to is well known. Various distribution estimating algorithms are available in R. In this discussion, I will focus on `fitdistrplus` which uses maximum likelihood (mle) optimization to settle on a set of parameters that best fit the empirical data provided⁴. [8]

! Important

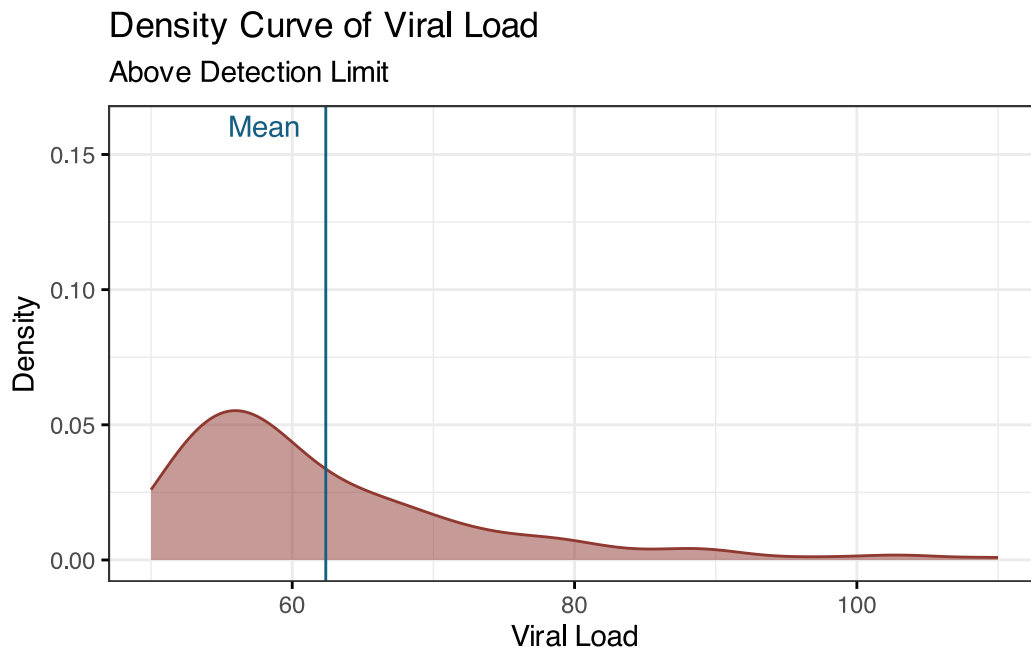
Remember that although we know the true values of our artificial dataset that are below the detection limit of 50, with real data, we will only have NA, BDL, or other indication that the machine could not produce a reliable value. So, we would be limited to working with only those values above the detection limit.

Determining the Distribution of the Original Values

The process of choosing a “best fitting” distribution involves testing the fit of various known distributions (such as normal, uniform, gamma, etc.) against the known values of the dataset. To prepare for this, it would be useful to plot the values above 50 to see if it suggests any of the typical curves of distributions.

³In a later post, I will address some of these techniques applied to the problem of the machine limits in determining values for viral load and other assays of HIV-1.

⁴mle is not the only optimization algorithm available in `fitdistrplus`. However, it is the one I will focus on here.



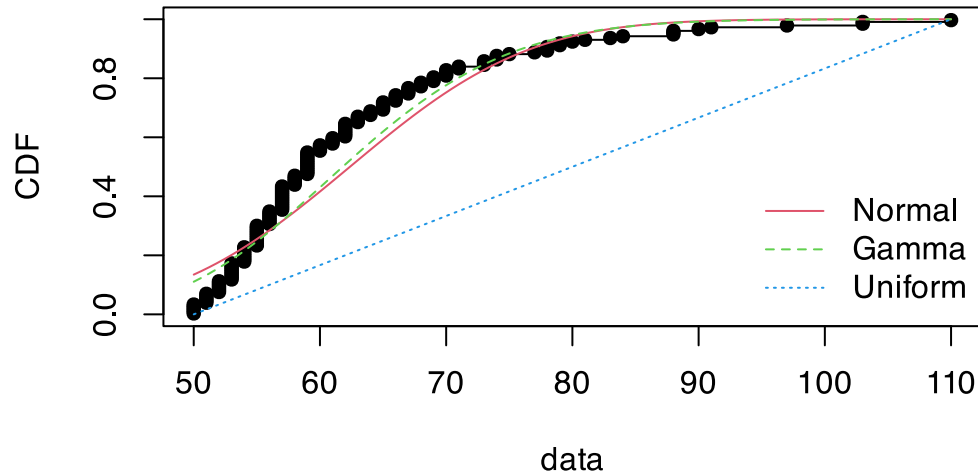
While this form shows a single peak that resembles a normal distribution, it is skewed to the right (skewness = 1.6968377). This makes it not directly a normal distribution. The Gamma distribution is a family of distributions whose shape is determined by two parameters. Among the possible curves of the distribution are a number that appear similar to our viral load data. For our analysis, we will test the Normal, Gamma, and Uniform distributions. I am adding the Uniform distribution to our inquiry list so we can see what a distribution that definitely doesn't fit the data looks like⁵.

Goodness of Fit of Distributions

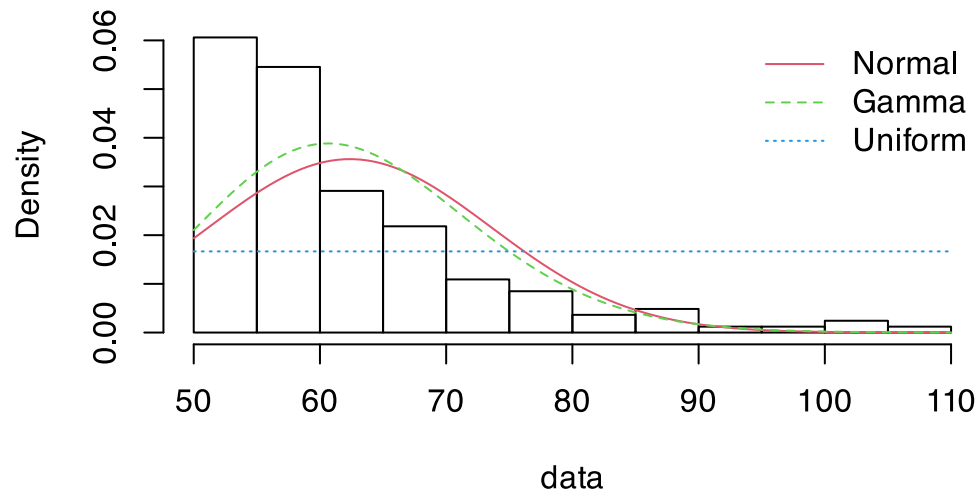
Distribution	logLikelihood	Cramer-von MisesStatistic	Akaike's InformationCriterion
Normal	-632.7550	1.3122234	1269.510
Gamma	-618.8550	0.9623501	1241.710
Uniform	-675.5669	18.1146970	1355.134

⁵The Uniform distribution doesn't fit the data as it assumes that the probability that the variable will assume any individual value is equal to the probability of any other value. Therefore, no peaks possible. This is clearly not the case for our data.

Empirical and theoretical CDFs



Histogram and theoretical densities



The table shows three evaluation criteria for the three distributions: Loglikelihood of the distribution fitting the empirical data, the result of the Cramer-von Mises one-sample test[9], and the Akaike's Information Criterion (AIC).[10] For the Loglikelihood test, the maximum value is what we are seeking. In the other two tests, the minimum value would suggest the best distribution to use.

💡 Tip

Remember that the Loglikelihood values are all negative, so the maximum value is the one closest to 0.

The graphs very clearly show how far away from the shape of the distribution is the Uniform model. The other two distributions are reasonably close to each other, but the Gamma distribution is slightly better both in the table and on the graphs.

i Note

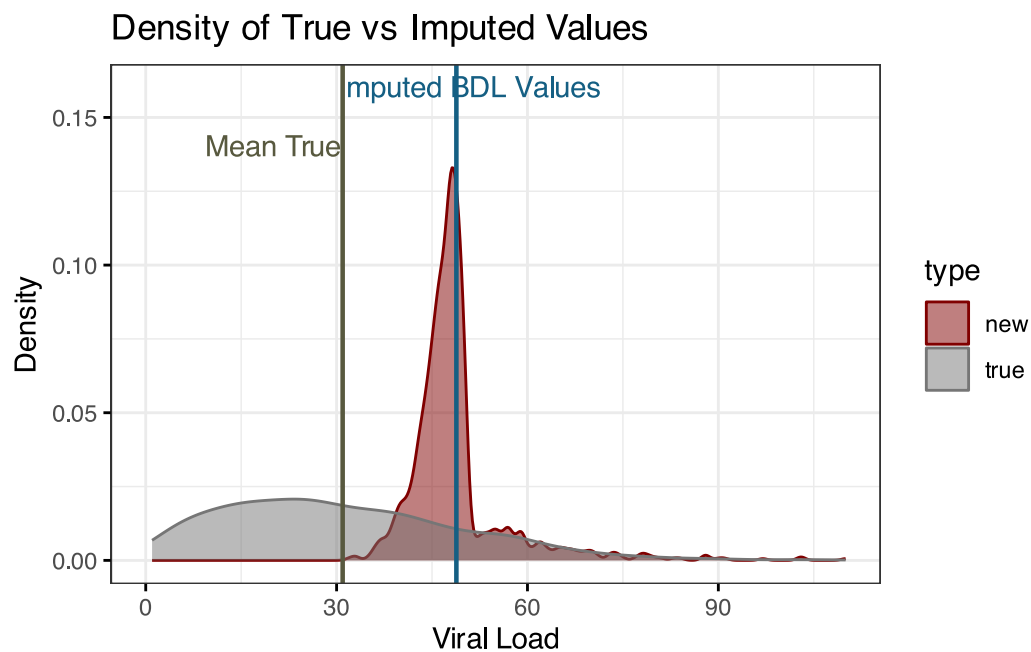
As a practical matter, I usually focus on the AIC to help me choose among models. Because of the simplicity in its calculation, it is a criterion that makes choosing relative strengths easy. The theory behind it is very much more complex than the calculation of it, but in terms of practical implementation, it makes sense to use.

Calculating New Values for BDL

So, we will use the Gamma distribution with the parameters specified by distribution fitting function (`fitdistrplus::fitdist()`) to calculate new values to replace those 835 values that are BDL. We will then test that distribution against the true values to see if this imputation of values approximates the true distribution better than the Half Detection Limit strategy.

Week 48 Viral Load							
Strat- egy	Mean	Diff True Mean	Std. Dev.	Vari- ance	Median	# of Cases	Co- efficient Var.
true	30.95	0.00	18.84	355.02	28	1000	0.61
cen- sored	62.38	31.42	11.23	126.21	59	165	0.18
Zero	10.29	-20.66	23.61	557.30	0	1000	2.29
Detec- tion Limit	52.04	21.09	6.47	41.84	50	1000	0.12
Half De- tection Limit	31.17	0.21	14.61	213.38	25	1000	0.47

Week 48 Viral Load							
Strat- egy	Mean	Diff True Mean	Std. Dev.	Vari- ance	Median	# of Cases	Co- efficient Var.
Im- puted BDL Values	48.83	17.87	8.09	65.50	48	1000	0.17



As both the graph and the last line of the table indicate, univariate imputation based on goodness-of-fit of the observed data (i.e, the data above detection limit) is not a good approximation of the true values originally shown. It is considerably worse than the Half Detection Limit strategy and involves much more work to achieve any result. As stated above, if the viral load or other variable subject to detection limits is part of a larger model with other variates that describe the clinical and virological state of the individuals, then MICE or another multiple imputation strategy may be desirable. This I will address in a later post.

Bibliography

- [1] G. C. Chang et al., “Brief Report: Undetectable HIV-1 Viral Load Among Virally Suppressed People Living With HIV: Implications for Undetectable Equals Untransmittable (U=U) in Resource-Limited Settings Using Dried Spot Testing,” *Journal of*

Acquired Immune Deficiency Syndromes (1999), vol. 100, no. 4, pp. 289–293, Dec. 2025, doi: 10.1097/QAI.00000000000003745.

- [2] C. Combescure et al., “How reliable is an undetectable viral load?,” *HIV medicine*, vol. 10, no. 8, pp. 470–476, Sep. 2009, doi: 10.1111/j.1468-1293.2009.00714.x.
- [3] N. Siccha, “Lower limits of detection or quantification.” [Online]. Available: https://www.generable.com/post/lloqs?utm_source=substack&utm_medium=email
- [4] J. L. Lennox et al., “Safety and efficacy of raltegravir-based versus efavirenz-based combination therapy in treatment-naïve patients with HIV-1 infection: a multicentre, double-blind randomised controlled trial,” *Lancet (London, England)*, vol. 374, no. 9692, pp. 796–806, Sep. 2009, doi: 10.1016/S0140-6736(09)60918-1.
- [5] S. L. Walmsley et al., “Dolutegravir plus abacavir-lamivudine for the treatment of HIV-1 infection,” *The New England Journal of Medicine*, vol. 369, no. 19, pp. 1807–1818, Nov. 2013, doi: 10.1056/NEJMoa1215541.
- [6] O. Mersmann, H. Trautmann, D. Steuer, and B. Bornkamp, *Truncated Normal Distribution*. 2023. [Online]. Available: <https://github.com/olafmersmann/truncnorm>
- [7] S. v. Buuren and K. Groothuis-Oudshoorn, “mice : Multivariate Imputation by Chained Equations in R,” *Journal of Statistical Software*, vol. 45, no. 3, 2011, doi: 10.18637/jss.v045.i03.
- [8] M. L. Delignette-Muller and C. Dutang, “fitdistrplus : An R Package for Fitting Distributions,” *Journal of Statistical Software*, vol. 64, no. 4, 2015, doi: 10.18637/jss.v064.i04.
- [9] W. contributors, “Cramér–von Mises criterion.” [Online]. Available: https://en.wikipedia.org/w/index.php?title=Cram%C3%A9r%E2%80%93von_Mises_criterion&oldid=1323176763
- [10] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. Oxford University Press, 2004. [Online]. Available: <https://books.google.com.br/books?id=tlshj56pl0IC>