# **RESEARCH ARTICLE**

# Modeling observations with a detection limit using a truncated normal distribution with censoring

Justin R. Williams<sup>1\*</sup> <sup>1</sup>, Hyung-Woo Kim<sup>2</sup> and Catherine M. Crespi<sup>1</sup>

## Abstract

Background: When data are collected subject to a detection limit, observations below the detection limit may be considered censored. In addition, the domain of such observations may be restricted; for example, values may be required to be non-negative.

**Methods:** We propose a method for estimating population mean and variance from censored observations that accounts for known domain restriction. The method finds maximum likelihood estimates assuming an underlying truncated normal distribution.

**Results:** We show that our method, tcensReg, has lower bias, Type I error rates, and mean squared error than other methods commonly used for data with detection limits such as Tobit regression and single imputation under a range of simulation settings from mild to heavy censoring and truncation. We further demonstrate the consistency of the maximum likelihood estimators. We apply our method to analyze vision quality data collected from ophthalmology clinical trials comparing different types of intraocular lenses implanted during cataract surgery. All of the methods yield similar conclusions regarding non-inferiority, but estimates from the tcensReg method suggest that there may be greater mean differences and overall variability.

**Conclusions:** In the presence of detection limits, our new method tcensReg provides a way to incorporate known domain restrictions in dependent variables that substantially improves inferences.

Keywords: Contrast sensitivity, limited dependent variables, Limited domain, Visual acuity

## Background

Censoring, in which the value of an observation is not known exactly but rather is only known to be above or below a specific value, is prevalent in many data settings. Censoring occurs with time-to-event data, but can also occur when measurements are subject to a detection limit (DL). A detection limit is defined as the lowest quantity or concentration of a compound that can be reliably detected with a given analytical method [1]. Quantities below the DL can be considered censored. Detection limits and the

\*Correspondence: williazo@g.ucla.edu

<sup>1</sup>Department of Biostatistics, University of California Los Angeles, Charles E. Young Dr. South, 90095 Los Angeles, USA

## tered in epidemiology [2, 3], hydrology [4], chemistry [5], toxicology [6], and economics [7, 8]. Estimation of the parameters of a normal distribution

censored observations associated with them are encoun-

based on a sample with censored observations has a long history of investigation. [9] was one of the first authors to develop maximum likelihood estimation methods for this data setting. Much of the early work focused on single mean models [9-12] or estimation using order statistics [13, 14]. Regression with a dependent variable subject to known censoring gained prominence with the work of [15], who developed the Tobit model, also called a censored regression or Tobit regression model. Linear regres-

© The Author(s). 2020 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.







Full list of author information is available at the end of the article

sion methods with an unspecified censored distribution were later developed by [16].

In settings in which censored data arise due to a DL, estimation is sometimes performed by singly imputing the DL or 1/2 DL for censored observations. While these methods are known to yield biased estimates of the mean and standard deviation, they are still routinely applied due to convenience. Single imputation of the DL yields upwardly biased estimates of the mean [4]. There can also be substantial bias using 1/2 DL imputed values, with the direction of the bias depending on the underlying data mechanism [1, 2, 4]. [2] noted that the bias of parameter estimates when using 1/2 DL is substantial unless the proportion of censored observations is small, defined as less than 10%.

Related to but distinct from censoring is the concept of truncation. A truncated distribution is a conditional distribution that results from restricting the domain of some other probability distribution. For example, the zero-truncated Poisson distribution is the distribution of a Poisson random variable conditional on the value of the random variable being non-zero. Truncation is a strict restriction of the domain of the random variable; observations outside the domain cannot occur. In contrast, when an observation is censored, its true value is known to lie beyond the censoring threshold, and such true values are permitted to occur. Truncation is a property of the population, while censoring is a result of the sampling mechanism.

Methods have been developed for estimating the parameters of a truncated normal distribution. [9] developed methods for the single mean model, while later work by [17] provided solutions for settings involving normal distributions with known or unknown truncation points and single or double truncation. [18] extended the theoretical framework to settings in which the distribution satisfies regularity conditions but is not necessarily normal. Work by [19] expanded the truncated regression framework by demonstrating the consistency and asymptotic normality properties of the maximum likelihood estimator and identifying consistent initial estimators.

Both censored and truncated data are often referred to as limited dependent variables in the economics literature [8, 15]. There are currently methods of parameter estimation applicable to censored observations from normally distributed random variables and uncensored observations from truncated normal random variables. However, estimation for the case in which observations are censored from a truncated normal random variable, are lacking. In this manuscript we propose maximum likelihood estimation techniques for such data. The methods can be seen as analogous to those of [15] with the latent normal distribution replaced with a latent truncated normal distribution. The paper is organized as follows. In "Methods" we introduce a motivating example from an analysis of visual quality data from clinical trials of intraocular lenses implanted during cataract surgery and the maximum likelihood estimation procedure. The "Results" compares the performance of the proposed new methodology to other common methods using both simulations and our motivating example. In the "Discussion" we reflect on the significance of the results, discusses limitations and note possible areas for extending the work. Finally the "Conclusion" sections highlights the importance of the method when modeling observations subject to a detection limit.

## Methods

## **Problem motivation**

Our application concerns left censored non-negative observations arising from contrast sensitivity testing in clinical trials for intraocular lenses implanted during cataract surgeries.

Contrast sensitivity measures the visual quality experienced by a subject by testing his or her ability to distinguish increasingly finer increments of light versus dark. Being unable to distinguish objects when contrast is low, i.e., when there is little difference between light and dark, can make everyday tasks such as night driving, navigating new settings, or perceiving distances difficult [20]. Cataract patients tend to have especially poor contrast sensitivity due to the clouding of the natural lens. The primary treatment for cataracts is surgery, during which the natural clouded lens is removed and replaced with a new synthetical intraocular lens (IOL). The patient should see improvement in visual acuity and visual quality following IOL implantation. As such, contrast sensitivity is an important clinical outcome for patients who receive IOLs during cataract surgery.

Contrast sensitivity testing is performed using standardized charts with alternating light and dark bars, referred to as gratings. To determine the contrast sensitivity of a patient, the intensity of the contrast between the bars as well as the spacing is reduced until the patient is no longer able to perceive separate bars. The contrast is defined as the relative difference in luminance of the bars from the background and may be calculated using the Weber contrast,  $\frac{L_{max}-L_{min}}{L_{background}}$ , Michelson contrast,  $\frac{L_{max}-L_{min}}{L_{max}+L_{min}}$ , or RMS contrast,  $\frac{L_{\sigma}}{L_{\mu}}$ , where  $L_{max}$ ,  $L_{min}$ ,  $L_{background}$ ,  $L_{\mu}$ , and  $L_{\sigma}$ are luminance maximum, minimum, background, mean and standard deviation respectively [21]. Typically when using gratings to test contrast, the Michelson contrast is preferred. Contrast sensitivity is defined as the reciprocal of the threshold contrast, which is the lowest contrast that the patient can identify the grating. The spacing between the bars is measured in cycles per degree (CPD)

with higher values of CPD indicating less space between bars. The testing is performed across a variety of different CPD levels, under either bright or dim lighting conditions, and with or without glare. In general, contrast sensitivity scores are lower when testing is performed under dim lighting with glare.

At each CPD level, the subject is presented with a sample grating followed by 8 gratings that progressively decrease the intensity of the image contrast. Figure 1 shows the testing setup used for 12.0 CPD, which is one of the most common visual quality outcomes analyzed in ophthalmic clinical trials. The subject is first asked if they can identify the striped pattern in the sample image, and is subsequently shown each column starting from column 1 and asked to identify whether the striped pattern is in the top, bottom, or neither grating. A contrast sensitivity score is recorded as the lowest level of identifiable contrast, i.e., the lowest intensity contrast for which the patient is able to correctly identify the striped pattern. The contrast sensitivity score ranges from 0 to 8, with 0 representing the ability to identify only the sample grating and 8 corresponding to the last column with the lowest image contrast. If a subject is unable to identify the sample grating, the value is recorded as -1. The scores of 0-8 are converted to continuous log contrast sensitivity values based on the manufacturers' recommendations as shown in Table 1. Note that by definition, the contrast sensitivity values using the Michelson formula range from  $[1, \infty)$ , meaning that the log contrast sensitivity scores must be non-negative, i.e.,  $[0, \infty)$ .

Problems arise using this scoring approach when patients are unable to identify the sample grating, i.e., have contrast sensitivity threshold score of -1. The true log contrast sensitivity values for these patients are only known to be below the sample threshold. This is equivalent to having left censored observations. However, we also know that log contrast sensitivity values must be non-negative, implying that the distribution of log contrast sensitivity scores is left truncated at zero. Therefore, to accurately estimate the mean and standard deviation of log contrast sensitivity for a particular IOL, both the censoring and truncation need to be accounted for.

Our data come from two prospective clinical trials for IOLs implanted during cataract surgery. The first study, ClinicalTrials.gov Identifier NCT01510717, compared a monofocal lens and a multifocal lens in a double blind randomized parallel group study with bilateral IOL implantation, and the second study, ClinicalTrials.gov Identifier NCT01424189, compared two different multifocal lenses in a nonrandomized parallel assignment multicenter study again with bilateral IOL implantation. Our analysis will be restricted to data reported for binocular (both eyes open) testing, under dim lighting with and without glare. All observations were taken 6 months after surgery.

Figure 2 shows the marginal histograms with kernel density smoothers of each IOL by glare condition. Visual inspection of the data suggest that it may be reasonable to assume an underlying normal distribution for the log contrast sensitivity scores, with scores being censored at the lower-bound detection limit.

The goal of the analysis was to estimate the difference in mean log contrast sensitivity between the monofocal and each multifocal lenses. Historically, monofocal lenses have provided patients with better contrast sensitivity, but multifocal lenses are often preferred by patients for visual acuity as they provide both distance and near vision with increased spectacle independence. A common clinical trial hypothesis is to test whether the multifocal lens is non-inferior to the monofocal lens with respect to contrast sensitivity. Based on regulatory guidelines, losses of 0.3 log units are considered to be clinically significant when they occur at two or more spatial frequencies [22]. The non-inferiority margin for log contrast sensitivity is set at half of this clinically significant magnitude, i.e., loss of 0.15 log units. Our goal was thus to estimate pairwise differences in mean log contrast sensitivity between the monofocal lens and each multifocal lens to test for non-inferiority of visual quality.



		Contrast sensitivity threshold scores								
		0	1	2	3	4	5	6	7	8
	1.5	0.60	0.90	1.07	1.22	1.37	1.52	1.67	1.82	1.97
	3.0	0.70	1.00	1.17	1.34	1.49	1.63	1.78	1.93	2.08
CPD <sup>2</sup>	6.0	0.91	1.21	1.38	1.55	1.70	1.84	1.99	2.14	2.29
	12.0	0.61	0.91	1.08	1.25	1.40	1.54	1.69	1.84	1.99
	18.0	0.17	0.47	0.64	0.81	0.96	1.10	1.25	1.40	1.55

## Table 1 Manufacturer Log Contrast Sensitivity Value<sup>1</sup>

<sup>1</sup> Based on scoring instructions from http://www.vectorvision.com/csv1000-norms/ accessed on 08NOV2018 <sup>2</sup> CPD = Cycles per Degree

Data analyzed in "Application results" section is from 12.0 CPD test highlighted in grey

Detection limit (DL) for 12.0 CPD is 0.61 highlighted in the red box and therefore 1/2 DL is 0.31

## Statistical methods

This section first develops maximum likelihood estimation for a sample from a truncated normal distribution with censored observations, i.e., a single mean model. The methods are then extended to include a linear predictor for the mean. We then propose an extension to handle heteroskedastic variances. We focus on the setting in which the truncation and censoring occur only on the lower, or left, end of the distribution, i.e., left truncation and left censoring. The results can be generalized to right truncated and right censoring. Finally, the process of finding the maximum likelihood estimates is described using different optimization techniques.

### Single mean model

As a first step, we define the likelihood function for censored observations from a normal distribution. Assume a



contrast sensitivity threshold scores via Table 1. Detection limit for 12 CPD occurs at  $\nu = 0.61$ . Histogram is shown in the background with Gaussian kernel density estimate in the foreground with bandwidth set to 0.2

latent normal distribution with mean  $\mu$  and variance  $\sigma^2$  for the random variable  $X^*$ . Following [15], assume that the values of  $X^*$  are left-censored at  $\nu$ ,  $\nu \in \mathbb{R}$ , to produce the random variable *X* defined as

$$X_{i} = \begin{cases} \nu & \text{if } X_{i}^{*} \leq \nu, \\ X_{i}^{*} & \text{if } X_{i}^{*} > \nu. \end{cases}$$
(1)

Here, censored observations are reported as the DL  $\nu$ . The values of *X* represent the observed values in the sample, which are a partial representation of the values of  $X^*$ . Assume that a total of *n* observations are independently drawn with  $n_0$  observations censored, i.e.,  $n_0 = \sum_{i=1}^{n} 1_{\{x_i = \nu\}}$ , and  $n_1$  observations uncensored, i.e.,  $n_1 = \sum_{i=1}^{n} 1_{\{x_i > \nu\}}$ . The likelihood function for such data is

$$L(\mu,\sigma) = \left[\Phi\left(\frac{\nu-\mu}{\sigma}\right)\right]^{n_0} \left[\frac{1}{\sigma}\right]^{n_1} \prod_{i \in S_1} \phi\left(\frac{x_i-\mu}{\sigma}\right),$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  denote the standard normal pdf and cdf and  $S_1$  is the set of all uncensored observations. Therefore the log-likelihood is

$$l(\mu,\sigma) = n_0 \ln\left[\Phi\left(\frac{\nu-\mu}{\sigma}\right)\right] - n_1 \ln(\sigma) + \sum_{i \in S_1} \ln\left[\phi\left(\frac{x_i-\mu}{\sigma}\right)\right].$$

Maximum likelihood estimates  $\hat{\mu}$  and  $\hat{\sigma}$  can be found using iterative optimization techniques, such as the Newton-Raphson algorithm, as discussed in [15].

Now we replace the latent normal distribution above with a latent truncated normal distribution, truncated from the left at the value *a*. Call this random variable  $Y^*$ , which has the following pdf and cdf from [17]:

$$\begin{split} f_{Y_i^*}\left(y_i^*\right) &= \frac{1}{1 - \Phi\left(\frac{a - \mu}{\sigma}\right)} \left[\frac{1}{\sigma} \phi\left(\frac{y_i^* - \mu}{\sigma}\right)\right],\\ F_{Y_i^*}\left(y_i^*\right) &= \frac{\Phi\left(\frac{y_i^* - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{a - \mu}{\sigma}\right)}. \end{split}$$

The distribution of  $Y^*$  is a scaled version of a normally distributed random variable, obtained by dividing the pdf by the constant  $1 - \Phi\left(\frac{a-\mu}{\sigma}\right)$  to obtain a proper probability density function that integrates to one. We will denote a latent truncated normal random variable with left truncation at a constant  $a \in \mathbb{R}$  as

$$Y^* \sim TN\left(\mu, \sigma^2, a\right)$$

where we assume that the truncation value a is known and therefore fixed. For non-negative variables, a = 0.

Note that the parameter  $\mu$  denotes the mean of the underlying normal distribution prior to truncation, rather than the mean of the truncated normal distribution, which is  $\mu_{TN} = \mu + \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}\sigma$  [8]. Throughout the paper we focus on the estimation of this underlying central tendency parameter  $\mu$  rather than  $\mu_{TN}$ .

The log-likelihood for the truncated normal distribution with n independent observations drawn from this distribution is

$$l(\mu,\sigma) = -n\ln\left[1 - \Phi\left(\frac{a-\mu}{\sigma}\right)\right] - n\ln(\sigma) + \sum_{i=1}^{n}\ln\left[\phi\left(\frac{y_i^* - \mu}{\sigma}\right)\right].$$

There is no closed form for the maximum likelihood estimates for  $\mu$  and  $\sigma$  for the truncated normal distribution. However, [17] described methods for iteratively solving for the maximum likelihood estimates and [19] showed that the estimates were consistent and asymptotically normal.

Further suppose that the values of  $Y^*$  are censored at  $\nu > a$ ,  $\nu \in \mathbb{R}$ , to produce the random variable *Y*, defined as

$$Y_{i} = \begin{cases} \nu & \text{if } Y_{i}^{*} \leq \nu, \\ Y_{i}^{*} & \text{if } Y_{i}^{*} > \nu, \end{cases}$$
(2)

where  $\nu$  is a known constant. For example in the contrast sensitivity problem, the testing procedure has an inherent detection limit and cannot detect values below 0.61 log units. The log contrast sensitivity scores must also be non-negative as discussed earlier in "Problem motivation" section, signifying an implicit truncation value of a = 0.

The pdf for the truncated random variable with censoring can be expressed as

$$f_{Y_i}(y_i) = 1_{\{y_i = \nu\}} \left[ \frac{\Phi\left(\frac{\nu - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{a - \mu}{\sigma}\right)} \right] + 1_{\{y_i > \nu\}} \left[ \frac{1}{\sigma\left(1 - \Phi\left(\frac{a - \mu}{\sigma}\right)\right)} \phi\left(\frac{y_i - \mu}{\sigma}\right) \right].$$
(3)

The first term of this equation captures censored observations, in which case the data are reported as the detection limit value  $\nu$  and the cdf of the truncated normal is used to provide information for the likelihood. In the second term, the observation is not censored and we simply use the pdf for the truncated normal. Figure 3 shows an example of the probability density function for such a random variable.

Out of a total of *n* observations, again let  $n_0$  be censored such that  $n_0 = \sum_{i=1}^n 1_{\{y_i = \nu\}}$ , and  $n_1$  be uncensored,  $n_1 = \sum_i^n 1_{\{y_i > \nu\}}$ . Let *S* define the set of all observations,  $S_1$  be the set of uncensored observations, and  $S_0$  be the set of censored observations, i.e.,  $S_0 \cup S_1 = S$ . Assuming that the observations are drawn independently, the likelihood is



$$L(\mu,\sigma) = \left(\frac{1}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^n \left[\Phi\left(\frac{\nu-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right]^{n_0} \\ \left(\frac{1}{\sigma}\right)^{n_1} \prod_{i \in S_1} \phi\left(\frac{y_i - \mu}{\sigma}\right).$$
(4)

Taking the log of the likelihood, we have

$$l(\mu,\sigma) = -n\ln\left[1 - \Phi\left(\frac{a-\mu}{\sigma}\right)\right] + n_0\ln\left[\Phi\left(\frac{\nu-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right] - n_1\ln(\sigma) \quad (5)$$
$$+ \sum_{i\in\mathcal{S}_1}\ln\left[\phi\left(\frac{y_i-\mu}{\sigma}\right)\right].$$

Similar to the censored only and truncated only loglikelihoods, the maximum likelihood estimators for  $\mu$  and  $\sigma$  for the truncated normal distribution with censoring do not have a closed form but can be estimated using the iterative process discussed in "Obtaining maximum likelihood estimates" section.

## Linear predictor for the mean

The goal of many applications is to understand how certain predictors influence the mean or to compare the means of different populations. This can be accomplished by using the linear predictor  $\mathbf{X}_i^T \boldsymbol{\beta}$ , i.e.,

$$Y_i^* \sim TN\left(\mathbf{X}_i^T \boldsymbol{\beta}, \sigma^2, a\right),$$

where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p-1})^T$  is a  $(p-1) \times 1$  vector of parameters with  $p \geq 2$ . Again, note that  $\mathbf{X}_i^T \boldsymbol{\beta}$  is the mean of the underlying normal distribution rather than the mean of the truncated normal distribution, i.e.,  $\mathbb{E}[X_i^*]$  rather than  $\mathbb{E}[Y_i^*]$ . All of the unknown parameters can be collected into the vector  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)^T$  which has length p. The corresponding pdf and cdf are

$$f_{Y_{i}^{*}}\left(y_{i}^{*}\right) = \frac{1}{1 - \Phi\left(\frac{a - \mathbf{x}_{i}^{T} \boldsymbol{\beta}}{\sigma}\right)} \left[\frac{1}{\sigma} \phi\left(\frac{y_{i}^{*} - \mathbf{x}_{i}^{T} \boldsymbol{\beta}}{\sigma}\right)\right]$$
$$F_{Y_{i}^{*}}\left(y_{i}^{*}\right) = \frac{\Phi\left(\frac{y_{i}^{*} - \mathbf{x}_{i}^{T} \boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{a - \mathbf{x}_{i}^{T} \boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(\frac{a - \mathbf{x}_{i}^{T} \boldsymbol{\beta}}{\sigma}\right)}.$$

Suppose that this truncated normal distribution is then censored at the value  $\nu$ , where  $\nu > a$ . Let the notation  $a_i^* = \frac{a - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}$  denote a standardized version of the constant

Page 7 of 25

*a*. The likelihood of the truncated normal distribution with censoring can be expressed as

$$l(\boldsymbol{\beta}, \sigma^{2}) = -\sum_{i=1}^{n} \ln\left[1 - \Phi\left(a_{i}^{*}\right)\right] + \sum_{i \in S_{0}} \ln\left[\Phi\left(\nu_{i}^{*}\right) - \Phi\left(a_{i}^{*}\right)\right] - n_{1}\ln(\sigma) + \sum_{i \in S_{1}} \ln\left[\phi\left(\frac{y_{i} - \mathbf{x}_{i}^{T}\boldsymbol{\beta}}{\sigma}\right)\right].$$
(6)

## Heteroskedastic variances

We now relax the assumption of homogeneous variance. We consider the case of independent groups with different variances. Assume we have samples drawn independently from *J* truncated normals, with each population having a common truncation value but possibly different variance, according to the model

$$Y_{ij}^* \sim TN\left(\mathbf{X}_{ij}^T \boldsymbol{\beta}, \sigma_j^2, a\right), \quad i = 1, \dots, n_j \text{ and } j = 1, \dots, J,$$

where  $n_j$  is the number of observations in group j and  $Y_{ij}^* \perp Y_{i'j'}^*$  for all  $i \neq i'$  and  $j \neq j'$ . Assume that observations are censored at the value  $\nu$  to create a sample of independent random variables with pdf

$$f_{Y_{ij}}(y_{ij}) = 1_{\{y_{ij}=\nu\}} \left[ \frac{\Phi\left(\frac{\nu - \mathbf{x}_{ij}^T \boldsymbol{\beta}}{\sigma_j}\right) - \Phi\left(\frac{a - \mathbf{x}_{ij}^T \boldsymbol{\beta}}{\sigma_j}\right)}{1 - \Phi\left(\frac{a - \mathbf{x}_{ij}^T \boldsymbol{\beta}}{\sigma_j}\right)} \right] \\ + 1_{\{y_{ij}>\nu\}} \left[ \frac{1}{\sigma_j \left(1 - \Phi\left(\frac{a - \mathbf{x}_{ij}^T \boldsymbol{\beta}}{\sigma_j}\right)\right)} \phi\left(\frac{y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}}{\sigma_j}\right) \right].$$

Because groups are independent, the log likelihood becomes

$$l\left(\boldsymbol{\beta}, \sigma_{1}^{2}, \dots, \sigma_{j}^{2}\right) = \sum_{j=1}^{J} \sum_{i=1}^{n_{j}} -\ln\left[1 - \Phi\left(a_{ij}^{*}\right)\right] + \sum_{i \in S_{0j}} \ln\left[\Phi\left(v_{ij}^{*}\right) - \Phi\left(a_{ij}^{*}\right)\right] - n_{1j}\ln(\sigma_{j}) \quad (7) + \sum_{i \in S_{1j}} \ln\left[\phi\left(\frac{y_{ij} - \mathbf{x}_{ij}^{T}\boldsymbol{\beta}}{\sigma_{j}}\right)\right],$$

where  $a_{ij}^* = \frac{a - \mathbf{x}_{ij}^T \boldsymbol{\beta}}{\sigma_j}$ ,  $S_{0j}$  and  $S_{1j}$  are the sets of censored observations and uncensored observations respectively in the  $j^{th}$  group, and  $n_{1j}$  is the number of uncensored observations in the  $j^{th}$  group.

## Obtaining maximum likelihood estimates

Our goal is to find the values  $\theta$  that maximize the log-likelihoods of Eqs. 5, 6, and 7, where  $\theta$  corresponds to the appropriate mean and standard deviation parameters, i.e.,  $\theta = (\beta, \sigma)$ . However, closed form solutions do not exist. To obtain maximum likelihood estimates, an iterative procedure is required. One approach is to use the Newton-Raphson algorithm using Taylor series expansion, as discussed in Chapter 14 in [23]. While the Newton-Raphson method has attractive local convergence guarantees and reliable performance [24], each step requires evaluation of the Hessian matrix, which can become computationally expensive for a large set of predictors. Alternative optimization routines, such as the quasi-Newton BFGS [25-28] or the conjugate gradient [29], which require only evaluation of the likelihood function and corresponding gradient, often require additional evaluations but have reduced memory and faster computing time.

Within R, other optimization packages such as the max-Lik package from [30] can also be used to find maximum likelihood estimates. This package is called in the censored only and truncated only maximum likelihood estimation packages in R such as censReg by [31] and truncreg by [32].

We developed the standalone R package tcensReg available in CRAN to solve the novel likelihood equation of the truncated normal distribution with censoring. This software package uses analytic results of the gradient and Hessian via Newton-Raphson optimization for the corresponding model of interest in either Eq. 6 or Eq. 7 derived in Appendix A. Several other optimization routines are available within the software including conjugate gradient, maxLik, and BFGS. This package uses familiar model syntax and has additional functionality to estimate parameters for the censReg and truncreg packages, respectively.

## Results

## Simulation study

We conducted a simulation study to compare the performance of our method to that of five methods of estimating the mean and standard deviation from a truncated normal distribution with censored observations. Method 1 is the gold standard method which uses the true uncensored observations, and accounts for the truncation in the estimation procedure by using the appropriate truncated log-likelihood. Method 2 uses the same uncensored truncated data but does not adjust for truncation in the normal distribution likelihood function. Methods 3-6 use the censored truncated observations but differ in how they treat censoring and truncation. Method 3 imputes all censored values with the detection limit and uses maximum likelihood estimation with a normal distribution likelihood, while Method 4 imputes all values as half of the detection limit and also uses normal maximum likelihood estimation. Both the DL and 1/2 DL imputation methods have been shown to perform poorly in cases with censoring [1, 2, 4], and are expected to show even worse performance in this setting since they do not account for censoring or truncation. However, they are still sometimes used in practice. Method 5 uses Tobit regression, which incorporates censoring into the likelihood. This method is often recommended when the assumption of normality seems reasonable. Finally, Method 6 is our proposed maximum likelihood estimation procedure described in the "Methods" section which takes into account not only the censoring but also the truncation of the underlying distribution. These six methods will be referred to as: 1) Gold Standard (GS), 2) Uncensored with no truncation adjustment (Uncens NT), 3) Detection Limit (DL), 4) 1/2 Detection Limit (1/2 DL), 5) Tobit regression (Tobit), and 6) censored regression with truncation adjustment (tcensReg), our proposed method.

Each of these methods makes different assumptions about the underlying data generating process. Both the GS and Uncens NT methods observe  $Y_i^*$ . The GS method uses the true generating distribution, i.e.,  $Y_i^* \sim$  $TN(\mu, \sigma^2, a)$ . With Uncens NT, the distribution is incorrectly assumed to be not truncated, i.e.,  $Y_i^* \sim N(\mu, \sigma^2)$ . The remaining methods observe only the censored data,  $Y_i$ , along with the detection limit v. For DL, 1/2 DL, and Tobit methods, the observed data are assumed to be generated from a normal distribution  $Y_i \sim N(\mu, \sigma^2)$ , but handle the censored observations differently. In DL and 1/2 DL, point masses are placed at either v or v/2 before estimating the parameters  $\mu$  and  $\sigma^2$ . The Tobit method instead assumes that the censored observations fall within the tail region below the detection limit, and an appropriate term is incorporated into the likelihood function. The tcensReg method correctly assumes that the censored observations are generated from an underlying truncated normal distribution,  $Y_i \sim TN(\mu, \sigma^2, a)$ , with information in the censored observations incorporated using the truncated normal tail probabilities. In simulation studies, we can test each of these methods against the gold standard, since the complete data generating process is known. However, in practice the true underlying distribution is unknown, meaning only the DL, 1/2 DL, Tobit, and tcensReg methods are appropriate to implement.

#### Set-up

Four different sets of simulation studies were conducted to compare the six methods. In the first simulation study, values from a single mean model were drawn to compare performance of the methods in terms of bias and mean squared error (MSE) for estimating the mean and standard deviation of the underlying latent distribution. The second simulation study focused on estimating the difference of the means of two independent populations and their common variance. In the third simulation study, the performance of the methods in a non-inferiority test setting similar to that of our motivating example was assessed. Finally, the last simulation study examined the consistency of the maximum likelihood estimation procedures as a function of sample size in a single mean model. The simulations were conducted in R version 3.6.2[33].

For the first simulation study, values were simulated from a truncated normal distribution,

$$Y_i^* \sim TN\left(\mu, \sigma^2, a\right)$$
,

and then censored. A constant value of a = 0 was used to represent zero-truncation. The values of  $\mu$  and  $\sigma^2$  were chosen to approximate the marginal distributions of the log contrast sensitivity scores from the application introduced in "Problem motivation" section. Typical values of mean log contrast sensitivity ranged from 0.7 to 1.1, and thus we used  $\mu \in \{0.7, 0.8, 0.9, 1.0, 1.1\}$ . Based on standard deviations observed in the contrast sensitivity data, for the simulation we used  $\sigma \in \{0.40, 0.45, 0.50\}$ . This created a total of 15 (5 × 3) parameter combinations for  $\mu$  and  $\sigma^2$ .

Observations from a truncated normal distribution,  $Y_i^* \sim TN(\mu, \sigma^2, a)$ , can be simulated by transforming samples from a uniform distribution on the interval [0, 1] using the inverse probability transformation:

$$Y_i^* = \Phi^{-1}\left\{p \times \left[1 - \Phi\left(\frac{a - \mu}{\sigma}\right)\right] + \Phi\left(\frac{a - \mu}{\sigma}\right)\right\} \times \sigma + \mu,$$

where *p* represents the sample from the uniform distribution [34]. This method of inverse transformation sampling is implemented with the tcensReg software package. After transforming to the appropriate truncated normal distribution, another dataset was generated by censoring the observations at v = 0.61. Values that fell below v were either replaced with the DL (0.61), 1/2 DL (0.305), or marked as censored for Tobit and tcensReg estimation.

Each of the six estimation methods was used to estimate  $\mu$  and  $\sigma$ . The data simulation and estimation procedures were repeated for B = 10,000 replications with a sample size of n = 100. For each of the six methods, two performance metrics were calculated for  $\theta \in {\{\mu, \sigma\}}$ : average bias  $(\bar{\theta} - \theta)$  and mean squared error (MSE;  $1/B \sum_{k=1}^{B} [\hat{\theta}_k - \theta]^2$ ) where  $\bar{\theta} = 1/B \sum_{k=1}^{B} \hat{\theta}_k$ . When reporting results for simulations throughout this manuscript, log MSE rather than MSE was used as the MSE for the various procedures sometimes differed by orders of magnitude. Applying this monotonic transformation allowed for easier comparison among methods.

In the second simulation study, we simulated values from two truncated normal populations with different means but common variance and truncation value, i.e.,

$$Y_1^* \sim TN\left(\mu_1, \sigma^2, a\right) \qquad Y_2^* \sim TN\left(\mu_2, \sigma^2, a\right)$$

A constant value of a = 0 was used to represent zerotruncation and a range of values of  $\mu_1$ ,  $\mu_2$ , and  $\sigma^2$  were selected to produce data similar to the application. Population 1 approximated a monofocal intraocular lens, while Population 2 approximated a multifocal intraocular lens. Means for monofocal lenses took values  $\mu_1 \in \{1.0, 1.1\}$ , and the difference between multifocal and monofocal lenses was set to range from no difference to a clinically significant difference of 0.3 log units [22], i.e.,  $\delta \in$  $\{-0.3, -0.2, -0.1, 0\}$ . Therefore the mean of Population 2 was set as  $\mu_2 = \mu_1 + \delta$ . The common standard deviation was assumed to take values  $\sigma \in \{0.40, 0.45, 0.50\}$ . The two-population simulation had a total of 24 (2 × 4 × 3) parameter combinations for  $\mu_1$ ,  $\mu_2$ , and  $\sigma^2$ .

A total of 100 observations from each population were sampled. Again, for each population a separate censored dataset was generated using the censoring threshold v = 0.61, and values that fell below v were replaced with the DL (0.61), 1/2 DL (0.305) or marked as censored for Tobit and tcensReg estimation. Each of the six estimation methods was used to estimate the mean difference between Population 1 and 2,  $\delta$ , and the common standard deviation,  $\sigma$ . Data simulation and estimation were repeated for B = 10,000 replications. For each of the six methods, average bias and MSE were calculated.

The third simulation study assessed the performance of the methods in the context of a non-inferiority test. Here we focused on the Type I error rates of the various methods. Type I error rates are particularly important for non-inferiority tests because, if non-inferiority is falsely accepted, patients may decide between products based on non-efficacy factors such as price and side effects assuming the products to be similar when in fact one is truly superior. For the non-inferiority test, data were simulated with a true difference of  $\delta = -0.15$  in the two population model, while varying  $\mu_1 \in \{1.0, 1.1\}$  and  $\sigma \in$ {0.40, 0.45, 0.50} for a total of 6 different non-inferiority scenarios. Then each of the six methods was used to construct  $1 - \alpha$  confidence intervals for  $\delta$ . The test is specified as one-sided because it is known that multifocal contrast sensitivity is less than monofocal contrast sensitivity. Constructing  $1 - \alpha$  confidence intervals and comparing the lower bound to the non-inferiority margin will result in a one-sided hypothesis test at the  $\alpha$  level. If the lower bound of the confidence interval does not cover the true value of  $\delta$ , then the non-inferiority hypothesis would be falsely accepted. A total of 100 observations were drawn from each population to construct the confidence intervals. The hypothesis test was repeated for B = 10,000 replications and the Type I error rate was calculated as the percent of replications where the lower bound of the  $1-\alpha$  confidence interval was greater than -0.15.

The final simulation study compared the performance of the methods as a function of the number of observations (*n*). The procedure from the first simulation was repeated fixing  $\mu = 0.9$  and  $\sigma = 0.45$ , but sample size varied from  $n \in \{100, 200, 400, 800, 1600\}$ . Average bias and MSE were calculated for estimation of  $\mu$  and  $\sigma$ . To compare the rate of Type I error as a function of sample size, the procedure for the third simulation was repeated with  $\mu_1 = 1.0$  and  $\sigma = 0.45$  while allowing the sample size per population to vary as above, i.e.,  $n = n_1 = n_2 \in \{100, 200, 400, 800, 1600\}$ .

The choice of initial starting values  $\theta^{(0)}$  for the tcensReg method is important since optimization algorithms can provide local rather than global convergence. To ensure that the starting values are reasonable, we recommend using initial estimates from a censored regression, i.e., Tobit, model. These estimates showed excellent rates of convergence for our simulation settings.

## Simulation results

Table 2 shows expected censoring and truncation percentages and the ratio of censoring to truncation for each of the 15 parameter value scenarios in the single mean model. The expected percent of censoring ranges from 10.8%-38.6% and truncation was typically  $\leq 5\%$  but was

**Table 2** Expected Percentages of Truncation and Censoring inSingle Mean Simulation Study

μ	σ	Expected Censoring %	Expected Truncated %	Censoring:Truncation Ratio*
	0.50	15.17%	1.39%	10.91
1.1	0.45	13.18%	0.73%	18.05
	0.40	10.76%	0.30%	35.87
	0.50	19.95%	2.28%	8.75
1.0	0.45	18.23%	1.31%	13.92
	0.40	15.96%	0.62%	25.74
	0.50	25.42%	3.59%	7.08
0.9	0.45	24.24%	2.28%	10.63
	0.40	22.47%	1.22%	18.42
	0.50	31.44%	5.48%	5.74
0.8	0.45	31.04%	3.77%	8.23
	0.40	30.15%	2.28%	13.22
	0.50	37.84%	8.08%	4.68
0.7	0.45	38.38%	5.99%	6.41
	0.40	38.64%	4.01%	9.64

Note that censoring threshold set at v = 0.61 and truncation value at a = 0. \*This column compares the expected percent of censoring to the expected percent of truncation, e.g., when  $\mu = 1.1$  and  $\sigma = 0.50$  we expect 10.91 times more censoring than truncation. slightly higher when the mean was closer to the truncation value of 0, i.e.,  $\mu \in \{0.7, 0.8\}$ . The ratio of censoring to truncation varies from 4.68 to 35.87 where values greater than 1 indicate more censoring than truncation. In general, for a fixed  $\mu$ , as  $\sigma$  increases the expected censoring and truncation increase while the ratio of censoring to truncation decreases.

Performance metrics of the six methods for  $\mu$  are shown in Fig. 4. The left panel shows results with respect to average bias. The figure shows that the gold standard was essentially unbiased with only slight negative bias for low values of  $\mu$ . The average bias of tcensReg was slightly negative for all scenarios, meaning that the estimated mean was smaller than the true value. The slight negative bias increased as the amount of censoring increased, i.e., as  $\mu$ decreased for a fixed  $\sigma$  and as  $\sigma$  increased. However, the absolute bias remained small and was always below 2% of the true  $\mu$  value. The Tobit, Uncens NT and DL methods consistently had positive bias, with the estimated mean consistently larger than the true value. The amount of bias also increased more rapidly for these other methods as censoring increased, compared to tcensReg. The 1/2 DL method had variable trends in bias with positive bias in some settings and negative bias in others. In particular, when  $\sigma=$  0.40, the 1/2 DL method overestimated  $\mu$  when  $\mu \leq 0.8$  and underestimated it when  $\mu > 0.8$ .

The tcensReg method was consistently closer to the gold standard compared to the Tobit method, especially in situations with high censoring, i.e.,  $\sigma = 0.50$ . There were

high censoring settings in which the tcensReg method and Tobit method performed similarly; for example, absolute bias was similar when  $\mu = 0.9$  and  $\sigma = 0.40$ , which had 22.47% expected censoring. For this scenario, the ratio of censoring to truncation is 18.42 meaning that there is more than 18 times more expected censoring to truncation. For cases where the ratio of censoring to truncation is low, the tcensReg method tends to outperform the Tobit method. When the ratio of censoring to truncation is greater, the Tobit method is comparable in terms of absolute bias to the tcensReg method, but with different directions of bias.

The precision of the estimates of the mean as reflected by their log MSE is shown in the right panel of Fig. 4. MSE was transformed to the log scale as there was a great disparity between the values for the DL point imputation method and the other five methods. Here, the average log MSE of each method should be compared to that of the gold standard. The DL method consistently had the highest log MSE of all of the methods. The tcensReg method had higher log MSE than the Tobit, 1/2 DL, and Uncens NT methods, suggesting that these methods were outperforming tcensReg. However, the average log MSE for these three methods often fell below that of the gold standard, which reflects an underestimation of the true sampling variance. In contrast, the tcensReg method always maintained a log MSE greater than or about equal to the gold standard. Overall, the tcensReg method avoided false precision.



The left panel of Fig. 5 provides evidence that other methods such as Tobit, 1/2 DL, and Uncens NT also systematically underestimated the variance of the underlying distribution. The Tobit and Uncens NT always have negative bias, meaning that these methods underestimated  $\sigma$ , with this bias being especially pronounced when  $\sigma$  is large and  $\mu$  is small. Similar to the bias for  $\mu$ , the bias for the 1/2 DL method when estimating  $\sigma$  is variable with positive bias in low censoring scenarios and negative bias in higher censoring scenarios. Overall the average bias for the tcensReg method was generally closest to the gold standard with only slight overestimates of  $\sigma$ . However, we note that again the ratio of censoring to truncation appears to play an important role when comparing the Tobit and tcensReg methods. The Tobit method slightly outperforms the tcensReg method for scenarios with  $\sigma = 0.4$ . In these settings the censoring to truncation ratio is higher as shown in Table 2, meaning that censoring is expected to occur much more frequently than truncation.

On the right panel of Fig. 5 we see results for the log MSE for each method. Again, the tcensReg method protected against false precision, with log MSE slightly greater than that of the gold standard. The 1/2 DL method had artificially low MSE for almost all parameter scenarios, while the Uncens NT alternated between over- and underestimation. The Tobit method was often the closest method to the gold standard.

In the two population simulation study, the parameter of primary interest was the difference of the means,  $\delta$ . Figure 6 shows the performance of each method with respect to average bias for  $\delta$ . For all values of  $\delta$ , the tcensReg and gold standard methods were essentially unbiased. The other four methods all had positive bias, corresponding to underestimation of the true difference in means. Similar to the single mean model, the greatest difference between the tcensReg method and the other methods occurred when the amount of censoring was greater, i.e., when  $\mu_1 = 1.0$  and  $|\delta|$  is large for a fixed  $\sigma$ .

For all six methods, the average bias of  $\hat{\delta}$  when  $\delta = 0$  was approximately 0. As our simulation study for the single mean model showed, bias generally increases as the mean decreases for fixed  $\sigma$ ; see Fig. 4. When the two means were equal, i.e.,  $\delta = 0$ , the biases for estimating each mean were also equal and thus there was no bias for estimating  $\delta$ .

Additional performance results for log MSE of  $\delta$  and common standard deviation  $\sigma$  in the two-population simulation are available in Appendix B, Figs. 11 and 12 respectively. These results were similar to those from the single mean model with tcensReg avoiding errors of false precision and having optimal average bias compared to the gold standard.

Results for testing the non-inferiority hypothesis with  $\alpha = 0.05$  are shown in Table 3. For all scenarios, the tcensReg method had slightly higher Type I error rates than the gold standard and the Tobit method had slightly higher Type I error rates than the tcensReg method. The





DL method had Type I error rates 2.5-4 times higher than the nominal value. The other single imputation method, 1/2 DL, had inflated Type I errors in the medium and high variance settings. Especially when censoring was higher, i.e.,  $\mu_1 = 1.0$  and  $\mu_2 = 0.85$ , the other methods had significantly higher Type I error rates compared to tcensReg.

Performance results for the numerical consistency of the maximum likelihood estimates with respect to  $\mu$  in a single mean model are displayed in Fig. 7. In the left

panel average bias for the tcensReg and gold standard methods decreased as sample size increased, and is near zero for sample sizes above 200. All other methods show no noticeable differences in average bias as the sample size increased. As shown in the right panel of Fig. 7, the average log MSE remains approximately constant for the tcensReg method as sample size increases, while Tobit, Uncens NT, and DL begin to perform noticeably worse. The 1/2 DL method still shows evidence of false precision as the average log MSE is below the gold standard for all

Table 3 Type	I Error Rates for	Non-inferiority	y Test in Simu	lation Study
--------------	-------------------	-----------------	----------------	--------------

$\mu_1$	$\mu_2$	δ	σ	GS <sup>1</sup>	Uncens NT <sup>2</sup>	DL <sup>3</sup>	1/2 DL	Tobit <sup>4</sup>	tcensReg <sup>5</sup>
1.1	0.95	-0.15	0.40	0.0552	0.0609	0.1240	0.0477	0.0576	0.0555
1.0	0.85	-0.15	0.40	0.0563	0.0709	0.1874	0.0577	0.0652	0.0594
1.1	0.95	-0.15	0.45	0.0606	0.0736	0.1424	0.0648	0.0712	0.0642
1.0	0.85	-0.15	0.45	0.0642	0.0863	0.2046	0.0766	0.0776	0.0682
1.1	0.95	-0.15	0.50	0.0582	0.0755	0.1456	0.0687	0.0670	0.0599
1.0	0.85	-0.15	0.50	0.0603	0.0887	0.1996	0.0831	0.0754	0.0641

**Note**: nominal  $\alpha$  set to 0.05

<sup>1</sup>GS = Gold Standard, i.e., uncensored observations with truncation adjustment

<sup>2</sup>Uncens NT = Uncensored data with no truncation adjustment

<sup>3</sup>DL = detection limit

<sup>4</sup>Tobit = Tobit censored regression with no truncation adjustment

<sup>5</sup>tcensReg = Censored regression with truncation adjustment



values of *n*. Similar results for the estimates of  $\sigma$ , shown in Appendix B Fig. 13, demonstrate that the bias of the tcensReg method decreases as a function of sample size with consistent MSE performance. Finally, Table 4 shows that Type I error rates for GS and tcensReg generally decrease as a function of sample size. Other methods have Type I error rates that escalate with sample size, with Uncens NT, DL, 1/2 DL, and Tobit all having more than 2.9 times the nominal value with a sample size of n = 1600.

**Table 4** Type 1 Error Rates as Sample Size Increases for

 Simulation Study

n	GS <sup>1</sup>	Uncens NT <sup>2</sup>	DL <sup>3</sup>	1/2 DL	Tobit <sup>4</sup>	tcensReg <sup>5</sup>
100	.0569	.0762	.1835	.0645	.0662	.0578
200	.0633	.1020	.2974	.0831	.0813	.0661
400	.0610	.1286	.4698	.1023	.0945	.0640
800	.0580	.1794	.7238	.1303	.1102	.0606
1600	.0576	.2707	.9410	.1861	.1452	.0616

Note: nominal  $\alpha$  set to 0.05

 $\label{eq:GS} ^1\text{GS} = \text{Gold Standard, i.e., uncensored observations with truncation adjustment} \\ ^2\text{Uncens NT} = \text{Uncensored data with no truncation adjustment}$ 

<sup>3</sup>DL = detection limit

<sup>4</sup>Tobit = Tobit censored regression with no truncation adjustment

<sup>5</sup>tcensReg = Censored regression with truncation adjustment

## Application results

We now apply the methods to our contrast sensitivity application introduced in "Problem motivation" section. The goal was to compare visual quality, measured as contrast sensitivity, for monofocal vs multifocal lenses implanted following cataract surgery, using data collected from two clinical trials. Table 5 shows the number of participants who received each type of IOL and the percent of patients whose observation was censored under each glare condition. Each group has over 150 patients; Multifocal 2 is larger than the others due to a 2:1 allocation ratio in that trial. As mentioned previously, the amount of

<b>Table 5</b> Observed Censoring Percentage fe	or Intraocul	ar Lenses
(IOLs) under Dim Lighting at 12.0 CPD		

	Observed Censo		ring %	
IOL Type	n	With Glare	Without Glare	
Monofocal*	159	17.6%	12.6%	
Multifocal 1*	153	29.4%	20.9%	
Multifocal 2**	377	24.7%	11.4%	
Multifocal 3**	185	28.1%	18.4%	

\*Data from ClinicalTrials.gov Identifier NCT01510717

\*\*Data from ClinicalTrials.gov Identifier NCT01424189

censoring is greater with glare than without glare, with an average difference of 9% censoring. However, all groups have at least moderate levels of censoring, ranging from 11%-29%.

Separate models were fit for each pairwise comparison between the monofocal lens and each multifocal lens with and without glare. Initially, each pairwise comparison was fit assuming the standard deviation for lens type was heteroskedastic. Figure 8 displays result for the 95% confidence interval of each groups standard deviation,  $\hat{\sigma}_{j}$ . The standard deviations for the multifocal lenses are sufficiently close to the monofocal lens for both glare conditions with a high level of overlap in the confidence intervals. This suggested that a common standard deviation for lens type is appropriate. In the subsequent analysis, each model was re-run assuming a common standard deviation in the two groups. The parameters of interest included the mean of each population (monofocal or multifocal), the mean difference between populations, and the common standard deviation. Unlike the simulation design, we observe only the censored observations meaning we are restricted to four methods to estimate the parameters. DL imputation for censored observations, 1/2 DL imputation, Tobit regression and our censored truncated method, tcensReg.

We also tested whether the multifocal lens is noninferior to the monofocal lenses with respect to contrast sensitivity. Based on regulatory guidelines, the noninferiority margin for contrast sensitivity is set at -0.15 log units [22]. The non-inferiority test was one-sided test with  $\alpha = 0.05$ , conducted by constructing a two-sided 90% confidence interval (CI) for  $\delta$  and comparing the lower bound of the CI to the non-inferiority margin. To establish non-inferiority, the lower bound for the CI must be above the non-inferiority margin.

Figure 9 shows 90% CIs for  $\delta$  using the four methods. For all of the methods, the lower bound of the CIs are below the non-inferiority margin, for all three lens comparisons and both glare conditions. Thus using any of these methods, we would be unable to conclude that a particular multifocal lens is non-inferior to the monofocal lens in terms of contrast sensitivity. For each comparison, the tcensReg method had the largest estimate of  $|\delta|$  and the DL method had the smallest estimate. The point estimates of  $\delta$  from the Tobit and 1/2 DL methods were intermediate and similar.

The tcensReg method yielded the longest CIs (mean length = 0.164 log units). Mean CI lengths for the Tobit, 1/2 DL and DL methods were 0.158, 0.156 and 0.133 log units, respectively. The differences in CI length are a





reflection of differences in estimates of  $\sigma$ , shown in Fig. 10. The DL method had consistently low estimates of  $\sigma$ , while the tcensReg method had the highest estimated values of  $\sigma$ . The 1/2 DL and Tobit methods gave similar estimates intermediate between the other two methods.

Interpreting these results in light of our simulation studies, if the latent normality assumption holds for these data, we would expect the estimates from the tcensReg method to have the least bias. The other methods typically underestimated both the difference in means and the population standard deviation, which would tend to lead to confidence intervals that are more likely to falsely exclude the non-inferiority margin and thus to higher Type I error rates, especially when there is a high rate of censoring. The wider confidence intervals observed in Fig. 9 from the tcensReg may better protect against Type I errors.

As an additional way to compare the proposed tcensReg model against potential competitors, we compared goodness of fit. As suggested by [35], we focused on measures of fit based on the Pseudo  $R^2$ . For limited dependent variables, such as those arising from a censored truncated normal distribution, the authors found that the Pseudo  $R^2$  metric from [36] performed best for model comparison. Table 6 shows the results for this Pseudo  $R^2$  metric for each pairwise comparison with glare. The tcensReg method consistently had the highest value of the Pseudo  $R^2$ , providing evidence that this model was the most appropriate for the data.

## Discussion

In this manuscript, we developed a maximum likelihood method for estimating parameters from a truncated normal distribution when observations are subject to censoring. We have also developed the R package tcensReg to implement the method. We showed in simulations that our method has substantially less bias for estimating the mean and standard deviation than other commonly used methods for a range of simulation settings. Our method also had close to the nominal Type I error rate for noninferiority testing, while single imputation with either the detection limit or half the detection limit and Tobit regression often had inflated Type I error rates when the censoring rate was high.

The single mean simulation study showed that as the levels of censoring and truncation increased, the bias for all other methods generally increased, often dramatically. As expected, point imputation of the detection limit was consistently the worst performing method, but even



the Tobit method had large bias under certain conditions. An important factor when comparing the Tobit and tcensReg method in the single mean model was not only the raw levels of censoring and truncation but the ratio of censoring to truncation. For low censoring to truncation ratios, generally below 20, the tcensReg method outperformed Tobit, while high ratios tended to have similar performance in mean estimation and worse performance in standard deviation estimation. This trend may point to the scenarios where the censoring dominates truncation in terms of estimation and thus using the Tobit method can lead to more precise estimates, particularly of the standard deviation. It is important to note that the tcensReg method was the only method to have positive bias in estimation of the standard deviation while other methods were typically negatively biased. Positive bias in this sense leads to conservative hypothesis testing.

In the two population simulation study, we observed trends similar to those for the single mean model. The tcensReg method uniformly outperformed all other methods in estimating the average difference and standard deviation between the populations with respect to average bias. The greatest difference between the methods occurred when the censoring rate was high and the difference between the population means was large. Unlike the single mean model, the ratio of censoring to truncation did not appear to play a significant role in the accuracy of parameter estimation for the twopopulation method. Similar to the single mean model, the

**Table 6** Goodness of Fit Metrics for Intraocular Lenses (IOLs)

 under Dim Lighting at 12.0 CPD with Glare

IOL Comparison	Method	Psuedo R <sup>2</sup> *
	DL	0.0100
Multi 1 vs. Mono	1/2 DL	0.0137
Marti 1 V3. Morio	Tobit	0.0150
	tcensReg	0.0180
	DL	0.0112
Multi 2 vs. Mono	1/2 DL	0.0124
Marti 2 V3. Morio	Tobit	0.0132
	tcensReg	0.0159
	DL	0.0182
Multi 3 vs. Mono	1/2 DL	0.0216
Marti 9 VS. Morio	Tobit	0.0234
	tcensReg	0.0292

\*Pseudo  $R^2$  is calculated based on definition from [36]

to outperform the gold standard. Across all parameter scenarios, the tcensReg method had significantly lower average bias and did not fall victim to the false precision fallacy.

Results from the non-inferiority simulation study confirmed the trends observed in the previous simulation studies. Methods other than tcensReg tended to underestimate the true variability of the underlying data generating mechanism, which led to precise but biased results. This bias tended to increase as the censoring increased, and this combined with underestimated values of the true standard deviation led to substantially inflated Type I error rates. The tcensReg method consistently had estimates closest to the gold standard.

In the simulation study, the maximum likelihood estimates were shown to be numerically consistent for the tcensReg method. The average bias for population mean and standard deviation tended towards 0 as the sample size grew, and Type I error rates approached the nominal level.

In the application, the main difference among the methods was a greater estimate of the difference between means when using the tcensReg method. The tcensReg method also had the widest confidence interval length due to higher estimates of the variance. The DL method results suggested that it would be the most likely method to lead to a finding of non-inferiority, but as shown in the simulation study, this method can have highly inflated Type I error rates. The 1/2 DL, Tobit, and tcensReg methods are more likely to capture the true difference and variability in the estimate, with the tcensReg method showing evidence that the true difference may be even greater than previously thought using either of the other two methods. In our particular application, none of these methods differed with respect to the ultimate conclusion of the noninferiority hypothesis test. However, in cases in which non-inferiority is more marginal, the choice of method could make a difference. Goodness of fit metrics based on Pseudo  $R^2$  provided additional evidence that the tcensReg method was appropriate to model the observed contrast sensitivity scores. These goodness of fit metrics can be used in conjunction with subject matter knowledge about the data generating process when selecting a model to be used.

The goal of the estimation methods presented here is to estimate the mean and standard deviation of the latent normal distribution rather than the mean and standard deviation of the truncated normal distribution. This affects the interpretation of the results. When censoring is less than 50 percent, the mean of the latent normal corresponds to the mode of the truncated normal. In many settings the mode of the truncated normal may be a more clinically meaningful measure of central tendency than the mean of the truncated normal. For example, in contrast sensitivity testing, it may be clinically relevant to understand what factors affect the mode rather than the mean in order to target where the greatest proportion of patients lie.

The methodology as presented makes a strong parametric assumption of normality. For studies with small to moderate sample sizes, checking the reasonableness of the normality assumption may be difficult and the assumption may be increasingly tenuous as the amount of censoring and truncation increase. While the unobserved data may not be exactly normal, a normality assumption at least reflects an assumption that the unobserved part of the distribution has a monotone decreasing shape, which is frequently a reasonable assumption. In general, whenever an analysis involves unverifiable assumptions, conducting a sensitivity analysis is prudent. The tcensReg method can be viewed as an option in the toolkit of the statistician that can be used as part of a sensitivity analysis, which might also include Tobit regression.

Another potential limitation is that the detection threshold and truncation value are assumed to be known. While assuming that the detection threshold is known is often reasonable, the truncation value is not observable in the data due to the censoring and so must be incorporated in the analysis based on subject matter expertise.

In this manuscript, we have focused on the setting of left censoring and left truncation. The methods can be easily extended to handle right censoring with right truncation. Future work could extend the methods to handle censoring in both tails of the distribution. In our application, we noted some suggestion of censoring due to an upper detection threshold; see Fig. 2. Conducting the analysis adjusting for left and right censoring could potentially improve the accuracy of the estimates.

Other possible extensions include to other non-normal parametric distributions, or to linear mixed effect models with repeated measurements using the extended Newton-Raphson technique proposed by [37]. An additional area for future work is criteria for model and variable selection. Work by [38] introduced the concept of using Focused Information Criteria when performing variable selection for Tobit models. Based on our work, we suggest using Pseudo  $R^2$  from [36] when comparing models and AIC or BIC for variable selection in tcensReg models.

## Conclusion

In the presence of detection limits, the values of observations below the detection limit are censored. In addition, in many settings, the response domain is known to be restricted, often with zero truncation. We demonstrated a new estimation method that accounts for this restriction and thereby substantially improves inferences.

## **Appendix A: Analytic derivations**

## Formula for the gradient

In the model discussed in "Linear predictor for the mean" section, the parameters to be estimated are  $\theta$  =  $(\boldsymbol{\beta}, \sigma)^T$  where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p-1})^T$  from Eq. 6. The single mean model is a special case of the linear equation where  $\mathbf{X}_{i}^{T}\boldsymbol{\beta}=\mu.$ 

Investigating the log likelihood from Eq. 6, it is evident that calculating the partial derivatives with respect to  $\ln(\sigma)$  rather than  $\sigma$  is optimal. Then to obtain estimates and standard errors for  $\sigma$ , one can apply the inverse transformation and delta method.

Note that another possible parametrization using one-to-one functions is  $\delta = \frac{\beta}{\sigma}$  and  $\eta = \frac{1}{\sigma}$  used by [15]. However, we present results for the derivation using  $\beta$  and  $\ln(\sigma)$  analogous to methods used by [31] in the R package censReg.

Throughout this derivation, let  $c_i^* = \frac{c - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma}$  for some constant c. The following properties will be used to calculate the appropriate derivatives:

- $\frac{\partial}{\partial x} \Phi(x) = \phi(x)$   $\frac{\partial}{\partial x} \phi(x) = -x\phi(x)$   $\frac{\partial}{\partial \ln(\sigma)}\sigma = \frac{\partial}{\partial \ln(\sigma)} \exp \left[\ln(\sigma)\right] = \sigma$   $\frac{\partial}{\partial \ln(\sigma)}c_i^* = -c_i^*$

We define the gradient as

$$\nabla l(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial l}{\partial \beta_1} \\ \vdots \\ \frac{\partial l}{\partial \beta_{p-1}} \\ \frac{\partial l}{\partial \ln(\sigma)} \end{pmatrix}, \text{ for } p \ge 2.$$

Taking the derivative of the log-likelihood with respect to  $\beta_k$ ,

$$\frac{\partial l}{\partial \beta_k} = -\sum_{i=1}^n \frac{x_{ik}\phi\left(a_i^*\right)}{\sigma\left[1 - \Phi\left(a_i^*\right)\right]} - \sum_{i \in S_0} \frac{x_{ik}\left[\phi\left(v_i^*\right) - \phi\left(a_i^*\right)\right]}{\sigma\left[\Phi\left(v_i^*\right) - \Phi\left(a_i^*\right)\right]} + \frac{1}{\sigma^2}\sum_{i \in S_1} x_{ik}\left(y_i - \mathbf{x}_i^T\boldsymbol{\beta}\right) \text{ for } k = 1, \dots, p-1.$$
(8)

Then taking the derivative with respect to  $\ln(\sigma)$ ,

$$\frac{\partial l}{\partial \ln(\sigma)} = -\sum_{i=1}^{n} \frac{a_{i}^{*}\phi(a_{i}^{*})}{1 - \Phi(a_{i}^{*})} - \sum_{i \in S_{0}} \frac{\left[v_{i}^{*}\phi(v_{i}^{*}) - a_{i}^{*}\phi(a_{i}^{*})\right]}{\Phi(v_{i}^{*}) - \Phi(a_{i}^{*})} - n_{1} + \frac{1}{\sigma^{2}} \sum_{i \in S_{1}} \left(y_{i} - \mathbf{x}_{i}^{T}\boldsymbol{\beta}\right)^{2}$$
(9)

Therefore, the gradient vector is

$$\nabla l(\boldsymbol{\theta}) = \begin{pmatrix} -\sum_{i=1}^{n} \frac{x_{i1}\phi(a_{i}^{*})}{\sigma[1-\Phi(a_{i}^{*})]} - \sum_{i\in S_{0}} \frac{x_{i1}[\phi(v_{i}^{*})-\phi(a_{i}^{*})]}{\sigma[\Phi(v_{i}^{*})-\Phi(a_{i}^{*})]} \\ + \frac{1}{\sigma^{2}} \sum_{i\in S_{1}} x_{i1} \left( y_{i} - \mathbf{x}_{i}^{T} \boldsymbol{\beta} \right) \\ \vdots \\ -\sum_{i=1}^{n} \frac{x_{i(p-1)}\phi(a_{i}^{*})}{\sigma[1-\Phi(a_{i}^{*})]} - \sum_{i\in S_{0}} \frac{x_{i(p-1)}[\phi(v_{i}^{*})-\phi(a_{i}^{*})]}{\sigma[\Phi(v_{i}^{*})-\Phi(a_{i}^{*})]} \\ + \frac{1}{\sigma^{2}} \sum_{i\in S_{1}} x_{i(p-1)} \left( y_{i} - \mathbf{x}_{i}^{T} \boldsymbol{\beta} \right) \\ -\sum_{i=1}^{n} \frac{a_{i}^{*}\phi(a_{i}^{*})}{1-\Phi(a_{i}^{*})} - \sum_{i\in S_{0}} \frac{[v_{i}^{*}\phi(v_{i}^{*})-a_{i}^{*}\phi(a_{i}^{*})]}{\Phi(v_{i}^{*})-\Phi(a_{i}^{*})} - n_{1} \\ + \frac{1}{\sigma^{2}} \sum_{i\in S_{1}} \left( y_{i} - \mathbf{x}_{i}^{T} \boldsymbol{\beta} \right)^{2} \end{pmatrix}$$
(10)

## Formula for the hessian

The Hessian is the matrix of second derivatives,

$$\nabla^2 l(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_1^2} & \cdots & \frac{\partial^2 l}{\partial \beta_1 \partial \beta_{p-1}} & \frac{\partial^2 l}{\partial \beta_1 \partial \ln(\sigma)} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 l}{\partial \beta_{p-1} \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \beta_{p-1}^2} & \frac{\partial^2 l}{\partial \beta_{p-1} \partial \ln(\sigma)} \\ \frac{\partial^2 l}{\partial \ln(\sigma) \partial \beta_1} & \cdots & \frac{\partial^2 l}{\partial \ln(\sigma) \partial \beta_{p-1}} & \frac{\partial^2 l}{\partial \ln^2(\sigma)} \end{bmatrix}.$$

Note that this Hessian matrix is symmetric so that  $\nabla^2 l(\boldsymbol{\theta})_{ij} = \nabla^2 l(\boldsymbol{\theta})_{ji}$  for  $i \neq j$ .

The individual components of this matrix are calculated as

$$\frac{\partial^{2}l}{\partial\beta_{k}\partial\beta_{l}} = -\sum_{i=1}^{n} \frac{x_{ik}x_{il} \left\{a_{i}^{*} \left[1 - \Phi\left(a_{i}^{*}\right)\right] \phi\left(a_{i}^{*}\right) - \phi^{2}\left(a_{i}^{*}\right)\right\}}{\sigma^{2} \left[1 - \Phi\left(a_{i}^{*}\right)\right]^{2}} \\ -\sum_{i \in S_{0}} \frac{x_{ik}x_{il} \left\{\begin{bmatrix}v_{i}^{*} \phi\left(v_{i}^{*}\right) - a_{i}^{*} \phi\left(a_{i}^{*}\right)\right] \left[\Phi\left(v_{i}^{*}\right) - \Phi\left(a_{i}^{*}\right)\right]\right\}}{+\left[\phi\left(v_{i}^{*}\right) - \phi\left(a_{i}^{*}\right)\right]^{2}} \\ -\frac{1}{\sigma^{2}} \sum_{i \in S_{1}} x_{ik}x_{il} \text{ for } k = 1, \dots, p-1 \text{ and } l = k, \dots, p-1,$$
(11)

$$\frac{\partial^{2} l}{\partial \beta_{k} \partial \ln(\sigma)} = \sum_{i=1}^{n} \frac{x_{ik} \left\{ \left[ 1 - \Phi\left(a_{i}^{*}\right) \right] \phi\left(a_{i}^{*}\right) \left[ 1 - \left(a_{i}^{*}\right)^{2} \right] + a_{i}^{*} \phi^{2}\left(a_{i}^{*}\right) \right\} \right\}}{\sigma \left[ 1 - \Phi\left(a_{i}^{*}\right) \right]^{2}} - \left\{ \frac{x_{ik} \left\{ \left( \Phi\left(v_{i}^{*}\right) - \Phi\left(a_{i}^{*}\right)\right) \left( \phi\left(v_{i}^{*}\right) \left[ 1 - \left(v_{i}^{*}\right)^{2} \right] - \phi\left(a_{i}^{*}\right) \left[ 1 - \left(a_{i}^{*}\right)^{2} \right] \right) \right\}}{- \left[ \phi\left(v_{i}^{*}\right) - \phi\left(a_{i}^{*}\right) \right] \left[ \phi\left(v_{i}^{*}\right) v_{i}^{*} - \phi\left(a_{i}^{*}\right) a_{i}^{*} \right]}{\sigma \left[ \Phi\left(v_{i}^{*}\right) - \Phi\left(a_{i}^{*}\right) \right]^{2}} - \frac{2}{\sigma^{2}} \sum_{i \in S_{1}} x_{ik} \left( y_{i} - \mathbf{x}_{i}^{T} \boldsymbol{\beta} \right) \text{ for } k = 1, \dots, p - 1,$$

$$(12)$$

$$\frac{\partial^{2} l}{\partial \ln^{2}(\sigma)} = \sum_{i=1}^{n} \frac{a_{i}^{*} \left\{ \left[ 1 - \Phi\left(a_{i}^{*}\right) \right] \phi\left(a_{i}^{*}\right) \left[ 1 - \left(a_{i}^{*}\right)^{2} \right] + a_{i}^{*} \phi^{2}\left(a_{i}^{*}\right) \right\}}{\left[ 1 - \Phi\left(a_{i}^{*}\right) \right]^{2}} - \sum_{i \in S_{0}} \frac{\left\{ \left( \Phi\left(v_{i}^{*}\right) - \Phi\left(a_{i}^{*}\right)\right) \left( \phi\left(v_{i}^{*}\right) \left[\left(v_{i}^{*}\right)^{3} - v_{i}^{*}\right] - \phi\left(a_{i}^{*}\right) \left[\left(a_{i}^{*}\right)^{3} - a_{i}^{*}\right] \right) \right\}}{\left[ \Phi\left(v_{i}^{*}\right) - \Phi\left(a_{i}^{*}\right) \right]^{2}} - \frac{2}{\sigma^{2}} \sum_{i \in S_{1}} \left( y_{i} - \mathbf{x}_{i}^{T} \boldsymbol{\beta} \right)^{2}.$$
(13)

## Gradient for heteroskedastic model

Assuming that there are samples from *J* independent populations as discussed in "Heteroskedastic variances" section, the parameters to be estimated are  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_1, \dots, \sigma_J)^T$  where  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p-1})^T$ . Again, the form of log likelihood from Eq. 7 suggests calculating the partial derivatives with respect to  $\ln(\sigma_j)$  rather than  $\sigma_j$  is optimal. Let  $c_{ij}^* = \frac{c - \mathbf{x}_{ij}^T \boldsymbol{\beta}}{\sigma_j}$  for some constant *c*.

We define the gradient as

$$\nabla l(\boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial l}{\partial \beta_{1}} \\ \vdots \\ \frac{\partial l}{\partial \beta_{p-1}} \\ \frac{\partial l}{\partial \ln(\sigma_{1})} \\ \vdots \\ \frac{\partial l}{\partial \ln(\sigma_{j})} \end{pmatrix}, \text{ for } p \ge 2 \text{ and } J \ge 1.$$

Note that the case where J = 1 is equivalent to the case in "Formula for the gradient" section.

Taking the derivative of the log-likelihood with respect to  $\beta_k$ ,

$$\frac{\partial l}{\partial \beta_{k}} = \sum_{j=1}^{J} \left\{ \sum_{i=1}^{n_{j}} -\frac{x_{ijk}\phi\left(a_{ij}^{*}\right)}{\sigma_{j}\left[1-\Phi\left(a_{ij}^{*}\right)\right]} - \sum_{i\in S_{0j}} \frac{x_{ijk}\left[\phi\left(v_{ij}^{*}\right)-\phi\left(a_{ij}^{*}\right)\right]}{\sigma_{j}\left[\Phi\left(v_{ij}^{*}\right)-\Phi\left(a_{ij}^{*}\right)\right]} + \frac{1}{\sigma_{j}^{2}} \sum_{i\in S_{1j}} x_{ijk}\left(y_{ij}-\mathbf{x}_{ij}^{T}\boldsymbol{\beta}\right) \right\} \quad for \ k = 1, \dots, p-1,$$
(14)

and with respect to  $\ln(\sigma_j)$ ,

$$\frac{\partial l}{\partial \ln(\sigma_j)} = \sum_{i=1}^{n_j} -\frac{a_{ij}^* \phi\left(a_{ij}^*\right)}{1 - \Phi\left(a_{ij}^*\right)} - \sum_{i \in S_{0j}} \frac{\left[\nu_{ij}^* \phi\left(\nu_{ij}^*\right) - a_{ij}^* \phi\left(a_{ij}^*\right)\right]}{\Phi\left(\nu_{ij}^*\right) - \Phi\left(a_{ij}^*\right)} - n_{1j} + \frac{1}{\sigma_j^2} \sum_{i \in S_{1j}} \left(y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta}\right)^2 \text{ for } j = 1, \dots, J.$$
(15)

## Therefore, the gradient vector is

$$\nabla l(\boldsymbol{\theta}) = \begin{pmatrix} \sum_{j=1}^{J} \sum_{i=1}^{n_j} -\frac{x_{ij1}\phi(a_{ij}^*)}{\sigma_j [1-\Phi(a_{ij}^*)]} - \sum_{i \in S_{0j}} \frac{x_{ij1}[\phi(v_{ij}^*)-\phi(a_{ij}^*)]}{\sigma_j [\Phi(v_{ij}^*)-\Phi(a_{ij}^*)]} \\ + \frac{1}{\sigma_j^2} \sum_{i \in S_{1j}} x_{ij1} \left( y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta} \right) \\ \vdots \\ \sum_{j=1}^{J} \sum_{i=1}^{n_j} -\frac{x_{ij(\rho-1)}\phi(a_{ij}^*)}{\sigma_j [1-\Phi(a_{ij}^*)]} - \sum_{i \in S_{0j}} \frac{x_{ij(\rho-1)}[\phi(v_{ij}^*)-\phi(a_{ij}^*)]}{\sigma_j [\Phi(v_{ij}^*)-\Phi(a_{ij}^*)]} \\ + \frac{1}{\sigma_j^2} \sum_{i \in S_{1j}} x_{ij(\rho-1)} \left( y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta} \right) \\ \sum_{i=1}^{n_1} -\frac{a_{i1}^*\phi(a_{i1}^*)}{1-\Phi(a_{i1}^*)} - \sum_{i \in S_{0i}} \frac{[v_{i1}^*\phi(v_{i1}^*)-a_{i1}^*\phi(a_{i1}^*)]}{\Phi(v_{i1}^*)-\Phi(a_{i1}^*)} - n_{11} \\ + \frac{1}{\sigma_i^2} \sum_{i \in S_{1i}} (y_{i1} - \mathbf{x}_{i1}^T \boldsymbol{\beta})^2 \\ \vdots \\ \sum_{i=1}^{n_{i1}} -\frac{a_{ij}^*\phi(a_{ij}^*)}{1-\Phi(a_{ij}^*)} - \sum_{i \in S_{0j}} \frac{[v_{ij}^*\phi(v_{ij}^*)-a_{ij}^*\phi(a_{ij}^*)]}{\Phi(v_{ij}^*)-\Phi(a_{ij}^*)} - n_{1j} \\ + \frac{1}{\sigma_j^2} \sum_{i \in S_{1j}} (y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\beta})^2 \end{pmatrix}$$

$$(16)$$

## Hessian for heteroskedastic model

The Hessian matrix for parameters  $\theta$  in "Gradient for heteroskedastic model" section is derived by taking further partial derivatives. This matrix takes the form

$$\nabla^{2}l(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^{2}l}{\partial\beta_{1}^{2}} & \cdots & \frac{\partial^{2}l}{\partial\beta_{1}\partial\beta_{p-1}} & \frac{\partial^{2}l}{\partial\beta_{1}\partial\ln(\sigma_{1})} & \cdots & \frac{\partial^{2}l}{\partial\beta_{1}\partial\ln(\sigma_{f})} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^{2}l}{\partial\beta_{p-1}\partial\beta_{1}} & \cdots & \frac{\partial^{2}l}{\partial\beta_{p-1}^{2}} & \frac{\partial^{2}l}{\partial\beta_{p-1}\partial\ln(\sigma_{1})} & \cdots & \frac{\partial^{2}l}{\partial\beta_{p-1}\partial\ln(\sigma_{f})} \\ \frac{\partial^{2}l}{\partial\ln(\sigma_{1})\partial\beta_{1}} & \cdots & \frac{\partial^{2}l}{\partial\ln(\sigma_{f})\partial\beta_{p-1}} & \frac{\partial^{2}l}{\partial\ln^{2}(\sigma_{1})} & \cdots & \frac{\partial^{2}l}{\partial\ln(\sigma_{f})\partial\ln(\sigma_{f})} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^{2}l}{\partial\ln(\sigma_{f})\partial\beta_{1}} & \cdots & \frac{\partial^{2}l}{\partial\ln(\sigma_{f})\partial\beta_{p-1}} & \frac{\partial^{2}l}{\partial\ln(\sigma_{f})\partial\ln(\sigma_{1})} & \cdots & \frac{\partial^{2}l}{\partial\ln^{2}(\sigma_{f})} \end{bmatrix}.$$

Note that since the groups are assumed to be independent,  $\frac{\partial^2 l}{\partial \ln(\sigma_j) \partial \ln(\sigma_k)} = 0$  for all  $j \neq k$ , which reduces the Hessian matrix to

.

$$\nabla^{2}l(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial^{2}l}{\partial\beta_{1}^{2}} & \cdots & \frac{\partial^{2}l}{\partial\beta_{1}\partial\beta_{p-1}} & \frac{\partial^{2}l}{\partial\beta_{1}\partial\ln(\sigma_{1})} & \cdots & \frac{\partial^{2}l}{\partial\beta_{1}\partial\ln(\sigma_{l})} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^{2}l}{\partial\beta_{p-1}\partial\beta_{1}} & \cdots & \frac{\partial^{2}l}{\partial\beta_{p-1}^{2}} & \frac{\partial^{2}l}{\partial\beta_{p-1}\partial\ln(\sigma_{1})} & \cdots & \frac{\partial^{2}l}{\partial\beta_{p-1}\partial\ln(\sigma_{l})} \\ \frac{\partial^{2}l}{\partial\ln(\sigma_{1})\partial\beta_{1}} & \cdots & \frac{\partial^{2}l}{\partial\ln(\sigma_{1})\partial\beta_{p-1}} & \frac{\partial^{2}l}{\partial\ln^{2}(\sigma_{1})} & 0 \\ \vdots & \ddots & \vdots & \ddots \\ \frac{\partial^{2}l}{\partial\ln(\sigma_{l})\partial\beta_{1}} & \cdots & \frac{\partial^{2}l}{\partial\ln(\sigma_{l})\partial\beta_{p-1}} & 0 & & \frac{\partial^{2}l}{\partial\ln^{2}(\sigma_{l})} \end{bmatrix}$$

The individual components of this matrix are calculated as

$$\begin{aligned} \frac{\partial^{2} l}{\partial \beta_{k} \partial \beta_{l}} &= \sum_{j=1}^{l} \left\{ \sum_{i=1}^{n_{j}} - \frac{x_{ijk} x_{ijl} \left[ a_{ij}^{*} \left[ 1 - \Phi \left( a_{ij}^{*} \right) \right] \phi \left( a_{ij}^{*} \right) - \phi^{2} \left( a_{ij}^{*} \right) \right]}{\sigma_{k} \sigma_{l} \left[ 1 - \Phi \left( a_{ij}^{*} \right) \right]^{2}} \right] \\ &- \sum_{i \in S_{ij}} \frac{x_{ijk} x_{ijl}}{\sigma_{k} \sigma_{l} \left[ \Phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right]^{2}}{\sigma_{k} \sigma_{l} \left[ \Phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right]^{2}} \right] \\ &- \frac{1}{\sigma_{j}^{2}} \sum_{i \in S_{ij}} x_{ijk} x_{ijl} \right\} for k = 1, \dots, p - 1 and l = k, \dots, p - 1, \\ \frac{\partial^{2} l}{\partial \beta_{k} \partial \ln(\sigma_{j})} &= \sum_{i=1}^{n_{i}} \frac{x_{ijk} \left\{ \left[ 1 - \Phi \left( a_{ij}^{*} \right) \right] \phi \left( a_{ij}^{*} \right) \left[ 1 - \left( a_{ij}^{*} \right)^{2} \right] + a_{ij}^{*} \phi^{2} \left( a_{ij}^{*} \right) \right] \right\} \\ &- \frac{\partial^{2} l}{\partial \beta_{k} \partial \ln(\sigma_{j})} &= \sum_{i=1}^{n_{i}} \frac{x_{ijk} \left\{ \left[ 1 - \Phi \left( a_{ij}^{*} \right) \right] \phi \left( a_{ij}^{*} \right) \left[ 1 - \left( a_{ij}^{*} \right)^{2} \right] + a_{ij}^{*} \phi^{2} \left( a_{ij}^{*} \right) \right] \right\} \\ &- \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right) \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right]^{2} \right] - \phi \left( a_{ij}^{*} \right) \left[ 1 - \left( a_{ij}^{*} \right)^{2} \right] \right\} \\ &- \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right] \left[ \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right]^{2} \right] - \phi \left( a_{ij}^{*} \right) \left[ 1 - \left( a_{ij}^{*} \right)^{2} \right] \right\} \\ &- \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right] \left[ \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right]^{2} \right] \\ &- \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right] \left[ \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right]^{2} \right] \right] \\ &- \left[ \left( \phi \left( v_{ij}^{*} \right) - x_{ij} \right] for k = 1, \dots, p - 1 and j = 1, \dots, J, \\ \\ &\frac{\partial^{2} l}{\partial \ln^{2}(\sigma_{j})} = \sum_{i=1}^{n_{j}} \frac{a_{ij}^{*} \left[ \left( 1 - \Phi \left( a_{ij}^{*} \right) \right] \phi \left( a_{ij}^{*} \right) \left[ \left( 1 - \phi \left( a_{ij}^{*} \right) \right]^{2} \right] + a_{ij}^{*} \phi^{2} \left( a_{ij}^{*} \right) \right] \\ &- \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right] \left( \phi \left( v_{ij}^{*} \right)^{3} - v_{ij}^{*} \right] - \phi \left( a_{ij}^{*} \right) \left[ \left( a_{ij}^{*} \right)^{3} - a_{ij}^{*} \right] \right) \right\} \\ &- \sum_{i \in S_{ij}} \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right) \left( \phi \left( v_{ij}^{*} \right) \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^{*} \right) \right] \right]^{2} \\ &- \sum_{i \in S_{ij}} \left[ \left( \phi \left( v_{ij}^{*} \right) - \phi \left( a_{ij}^$$

## **Appendix B: Additional figures**









Fig. 13 Performance of Maximum Likelihood Estimate for  $\sigma$  as Function of Sample Size. The vertical dashed black line on the left figure corresponds to zero bias. GS = Gold Standard, i.e., uncensored observations with truncation adjustment; Uncens NT = Uncensored data with no truncation adjustment; DL = detection limit; Tobit = Tobit censored regression with no truncation adjustment; tcensReg = Censored regression with truncation adjustment

#### Abbreviations

CI: confidence interval; CPD: cycles per degree; DL: Detection Limit; GS: gold standard; IOL: intraocular lens; ISO: international standards organization; MSE: mean squared error; tcensReg: censored regression with truncation adjustment; TN: truncated normal; Uncens NT: uncensored with no truncation adjustment

#### Acknowledgements

Not applicable.

#### Authors' contributions

JRW performed analyses, implemented software, and wrote first draft of the article. HWK contributed to analyses, oversaw initial methodology design, and reviewed manuscript. CMC provided additional methodological extensions, edited manuscript, and oversaw analyses. The author(s) read and approved the final manuscript.

#### Funding

Not applicable.

#### Availability of data and materials

R code for generating data used in the simulations discussed in "Simulation study" is available at https://github.com/williazo/tcensReg/tree/master/inst/ script. The R package tcensReg is available as a stable release on CRAN at https://CRAN.R-project.org/package=tcensReg with developmental version of the software hosted on the corresponding GitHub page, https://github.com/ williazo/tcensReg. Data used for the application in "Application results" section is proprietary to Alcon Laboratories, Inc.

#### Ethics approval and consent to participate

Data analyzed for the motivating example were collected as part of clinical trials conducted by Alcon Laboratories, Inc. Details on these trials can be found on ClinicalTrials.gov using the identifiers NCT01510717 and NCT01424189. All participants or subjects' legally authorized representatives were provided and signed informed consent to participate in these trials.

#### Consent for publication

Not applicable

#### **Competing interests**

Hyung-Woo Kim is a former employee of Alcon Laboratories, Inc.

#### Author details

<sup>1</sup> Department of Biostatistics, University of California Los Angeles, Charles E. Young Dr. South, 90095 Los Angeles, USA. <sup>2</sup> Apellis Pharmaceuticals, Inc., 100 5th Avenue, 02451 Waltham, USA.

#### Received: 12 December 2019 Accepted: 26 May 2020 Published online: 29 June 2020

#### References

- Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. Appl Occup Environ Hyg. 1990;5(1): 46–51.
- Lubin JH, Colt JS, Camann D, Davis S, Cerhan JR, Severson RK, Bernstein L, Hartge P. Epidemiologic evaluation of measurement data in the presence of detection limits. Environ Health Perspect. 2004;112(17): 1691–6.
- Schisterman EF, Vexler A, Whitcomb BW, Liu A. The limitations due to exposure detection limits for regression models. Am J Epidemiol. 2006;163(4):374–83.
- 4. Helsel DR. Less than obvious-statistical treatment of data below the detection limit. Environ Sci Technol. 1990;24(12):1766–74.
- 5. Analytical Methods Committee. Recommendations for the definition, estimation and use of the detection limit. Analyst. 1987;112(2):199–204.
- Zaugg SD, Sandstrom MW, Smith SG, Fehlberg KM. Methods of Analysis by the U.S. Geological Survey National Water Quality Laboratory-Determination of Pesticides in Water by C-18 Solid Phase Extraction and Capillary-Column Gas Chromatography/Mass Spectrometry with

Selected-Ion Monitoring. Open-File Report 95-181, U.S. Geological Survey, Denver, Colorado. 1995. http://pubs.er.usgs.gov/publication/ofr95181.

- McDonald JF, Moffitt RA. The uses of tobit analysis. Rev Econ Stat. 1980;62(2):318–21.
- Greene WH. Econometric Analysis. In: Chap. 19. "Limited Dependent Variables, Truncation, Censoring and Sample Selection". 8th edn. United Kingdom: Pearson; 2018.
- Hald A. Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. Skandinavisk Aktuarietidskrift. 1949;32:119–34.
- 10. Gupta AK. Estimation of the mean and standard deviation of a normal population from a censored sample. Biometrika. 1952;39(3):260–73.
- Harter HL, Moore AH. Iterative maximum-likelihood estimation of the parameters of normal population from singly and doubly censored samples. Biometrika. 1966;53(1-2):205–13.
- 12. Tiku M. Estimating the mean and standard deviation from a censored normal sample. Biometrika. 1967;54(1):155–65.
- Sarhan AE, Greenberg BG. Estimation of location and scale parameters by order statistics from singly and doubly censored samples. Ann Math Stat. 1956;27(2):427–51.
- 14. Dixon WJ. Simplified estimation from censored normal samples. Ann Math Stat. 1960;31(2):385–91.
- 15. Tobin J. Estimation of relationships for limited dependent variables. Econometrica. 1958;26(1):24–36.
- Buckley J, James I. Linear regression with censored data. Biometrika. 1979;66(3):429–36.
- 17. Cohen Jr AC. Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. Ann Math Stat. 1950;21(4):557–69.
- Halperin M. Maximum likelihood estimation in truncated samples. Ann Math Stat. 1952;23(2):226–38.
- 19. Amemiya T. Regression analysis when the dependent variable is truncated normal. Econometrica. 1973;41(6):997–1016.
- 20. Owsley C, Sloane ME. Contrast sensitivity, acuity, and the perception of 'real-world' targets. Br J Ophthalmol. 1987;71(10):791–6.
- 21. Pelli DG, Bex P. Measuring contrast sensitivity. Vis Res. 2013;90:10-14.
- ISO 11979-7. Ophthalmic implants Intraocular lenses Part 7: Clinical investigations of intraocular lenses for the correction of aphakia. Vernier, Geneva, CH: Standard, International Organization for Standardization; 2018.
- 23. Lange K. Numerical Analysis for Statisticians. 2nd Edn. New York: Springer; 2010.
- 24. Ypma TJ. Historical development of the newton-raphson method. Soc Ind Appl Math. 1995;37(4):531–51.
- Broyden CG. The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations. IMA J Appl Math. 1970;6(1):76–90.
- 26. Fletcher R. A new approach to variable metric algorithms. Comput J. 1970;13(3):317–22.
- 27. Goldfarb D. A family of variable-metric methods derived by variational means. Math Comput. 1970;24(109):23–26.
- 28. Shanno DF. Conditioning of quasi-Newton methods for function minimization. Math Comput. 1970;24(111):647–56.
- 29. Fletcher R, Reeves CM. Function minimization by conjugate gradients. Comput J. 1964;7(2):149–54.
- Henningsen A, Toomet O. maxLik: A package for maximum likelihood estimation in R. Comput Stat. 2011;26(3):443–58.
- 31. Henningsen A. Estimating Censored Regression Models in R using the censReg Package. R Package vignettes. 2010;5(2):12.
- Croissant Y, Zeileis A. Truncreg: Truncated Gaussian Regression Models. R package version 0.2-5. 2018. https://CRAN.R-project.org/package= truncreg.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019. https:// www.R-project.org/, R Foundation for Statistical Computing.
- Burkardt J. The Truncated Normal Distribution [PDF File]. https://people. sc.fsu.edu/~jburkardt/presentations/truncated\_normal.pdf. Accessed 9 June 2020.
- Veall MR, Zimmermann KF. Pseudo-r2 measures for some common limited dependent variable models. J Econ Surv. 1996;10(3):241–59.

- 36. McKelvey RD, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. J Math Sociol. 1975;4(1):103–20.
- Lindstrom MJ, Bates DM. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. J Am Stat Assoc. 1988;83(404):1014–22.
- Zhang X, Wan AT, Zhou SZ. Focused information criteria, model selection, and model averaging in a Tobit model with a nonzero threshold. J Bus Econ Stat. 2012;30(1):132–42.

## **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

#### At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

