Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

Analysis of a marker for cancer of the thyroid with a limit of detection

Nicholas T. Longford, José Rafael Tovar Cuevas & Carlos Alvear

To cite this article: Nicholas T. Longford, José Rafael Tovar Cuevas & Carlos Alvear (2016): Analysis of a marker for cancer of the thyroid with a limit of detection, Journal of Applied Statistics, DOI: 10.1080/02664763.2016.1247792

To link to this article: http://dx.doi.org/10.1080/02664763.2016.1247792

Published online: 24 Oct 2016.

JOURNAL OF APPLIED STATISTICS, 2016
http://dx.doi.org/10.1080/02664763.2016.1247792

Taylor & Francis
Taylor & Francis Group

Analysis of a marker for cancer of the thyroid with a limit of detection

Nicholas T. Longford [a], José Rafael Tovar Cuevas[b] and Carlos Alvear[b]

[a]Department of Medicine, Imperial College, London SW10 9NH, UK; [b]Escuela de Estadística, Universidad del Valle, Ciudad Universitaria Meléndez, Edificio 357, 25360 Cali, Colombia

ABSTRACT

Limit of detection (LoD) is a common problem in the analysis of data generated by instruments that cannot detect very small concentrations or other quantities, resulting in left-censored measurements. Methods intended for data that are not subject to this problem are often difficult to modify for censoring. We adapt the simulation-extrapolation method, devised originally for fitting models with measurement error, to dealing with LoD in conjunction with a mixture analysis. The application relates the levels of thyroglobulin in individuals with cancer of the thyroid before and after treatment with radioactive iodine I–131. We conclude that the fitted mixture components correspond to levels of effectiveness of the treatment.

ARTICLE HISTORY
Received 10 August 2015
Accepted 9 October 2016

KEYWORDS
Cancer; extrapolation; limit of detection; mixture model; thyroid

AMS CLASSIFICATION
62P10; 62N99; 62J99

1. Introduction

Cancer of the thyroid gland is a rare type of cancer and the differentiated carcinoma is a common form of malignancy associated with the endocrine system [12,30]. Cancer of the thyroid affects more frequently women than men and between 85% and 90% of the malignancies correspond to differentiated thyroid carcinoma [9].

Outright removal of the gland, *thyroidectomy*, is a radical treatment for thyroid cancer. It is followed by an ablation therapy with iodine I–131 and suppression of the thyrotropin hormone. High levels of thyroglobulin (Tg) are a reliable proxy for the recurrence of cancer. Its concentration is established by analyses of the patient's blood sample prior to ablation therapy (Tg-pre) and one year later (Tg-post). The units of measurement are nanograms per milliliter of blood (ng/mL). The threshold of 2.0 ng/mL was recently proposed by Mejía *et al.* [20]. Values above this threshold after the therapy, Tg-post > 2.0, indicate the presence of cancerous cells, to be interpreted as a recurrence or a case of persisting cancer. The laboratory analysis follows a standard procedure using instruments that can detect only concentrations greater than a known lower limit, referred to as the *limit of detection* (LoD). Similar limits are commonly encountered in medical research and practice. We analyse a data set of 91 pairs of such measurements made before and after iodine therapy. Sex and age of the patients are also recorded; there are 74 women and 17 men, with ages ranging

CONTACT Nicholas T. Longford sntlnick@sntl.co.uk Department of Medicine, Imperial College, 369 Fulham Road, 4th floor, London SW10 9NH, UK

from 12 to 86 years, with mean 47.6 and median 50. Men are younger on average (mean 42.2 and median 39). The values of Tg-pre and Tg-post are extremely skewed, with a few very large values. Most values of Tg-post are lower than Tg-pre and many of them (46, that is, 50%) are below LoD.

Prior to therapy, the smallest recorded value is 0.1, in four instances. At the other extreme, the largest three values of Tg-pre are 672, 2474 and 3000, and 142, 220 and 586 for Tg-post. The values of Tg-pre and Tg-post were obtained in different laboratories, by personnel who used different conventions for rounding and recording values below LoD. The smallest values are rounded to one or two decimal places and greater values are rounded to one decimal place or to integers. Values below LoD were recorded as zero in 34 instances and as 0.1, the LoD, in 12 instances.

Figure 1 presents the data. Both axes are on the multiplicative scale, plotting the values $x = \log(1 + \text{Tg-pre})$ and $y = \log(1 + \text{Tg-post})$, although the axes are labelled on the original scale. A similar plot on the original scale would have very poor resolution. All values subject to LoD are recoded to 0.1. A small amount of random noise is added to the data points in both horizontal and vertical directions to better indicate observations with similar values. Grey colour is used for men and the age is indicated by the size of the symbol (greater symbol for older patients, but with different scales in the two panels). The points for the youngest and oldest patients, one man and seven women, are marked by white centers.

With a few exceptions, the concentrations after the therapy are lower than before, in many cases well below the conventionally adopted value of 2.0 ng/mL for the threshold between recurrent and non-recurrent subjects (horizontal dots in the diagram). Some patients had values below this threshold even before the I–131 therapy.

Figure 1 suggests that there are two kinds of patients. For some, the concentrations have been reduced by the treatment to very low levels, in many cases below LoD. For others, the levels have been reduced to values related to the original concentrations. This motivates

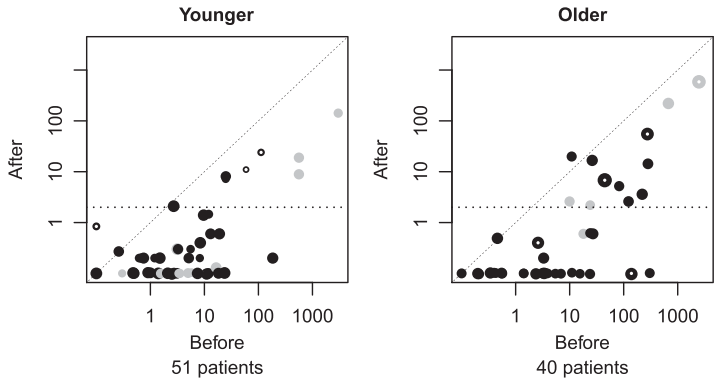

Figure 1. Concentration of thyroglobulin before and one year after therapy (Tg-pre and Tg-post). Patients aged 50 years and below are classified as younger, and above 50 years as older. The patients whose discs are marked by small white centres are the youngest (aged 12, 18 and 21 years) and the oldest (aged 71, 72, 75, 81 and 86 years). The horizontal dots mark the value of 2.0 ng/mL, adopted as the threshold between recurrent and non-recurrent cases.

the analysis described in the next section, where we combine a mixture regression model with the simulation-extrapolation method (SimEx), Carroll *et al.* [4], adapted to dealing with values below LoD. In the original version of SimEx, data are simulated according to the assumed measurement error model, but with inflated error variance $\sigma^2 + \Delta$, for several levels of inflation $\Delta > 0$. Estimates that refer to the absence of measurement error are then obtained by extrapolation of the Δ-specific results to $\Delta = -\sigma^2$. We replace the simulation step by forming the data set that would be obtained if the LoD threshold were increased. This step entails no uncertainty, and so it requires no simulations.

For background to the treatment of thyroid cancer, we refer to Asare and Wang [1] and references therein. Patient outcomes several years after ablation are studied by Ballal *et al.* [2]. Quality of life of patients after thyroidectomy is assessed by Rubic *et al.* [26]. Dysfunction of the salivary gland, a side effect of the I–131 treatment, is studied by Lee *et al.* [14]. Aspects of radiation safety and performance of the Greek hospitals in which I–131 treatment is provided are addressed by Vogiatzi *et al.* [31].

1.1. Limit of detection and mixture models

The role of LoD in the diagnosis of thyroid cancer is illustrated by Hänscheid et al. [10]. See also Wartofsky and Van Nostrand [32] for background to the use of iodine–131 in the diagnosis and treatment of thyroid cancer. The relevance of LoD and of related concepts to the management of cancer (diagnosis, treatment and post-treatment assessment of success) is discussed by Stamey [29] on the example of prostate-specific antigen. Methods for determining the value of LoD in molecular detection assays are described in Browne and Whitcomb [3], Rajakovic *et al.* [24] and Milbury *et al.* [22].

Methods for analysis of data with LoD and some generalisations of LoD are discussed by Lambert *et al.* [13]. They point out the errors and pitfalls of some common ways of treating values that are subject to LoD, and relate the problem to left-censoring. Biomedical applications (to biomarkers and longitudinal analysis, respectively) of data subject to LoD are presented by Vexler *et al.* [33,34]. The methods they and others use are constructed for specific applications and are difficult to adapt for other settings. We propose an approach that is generic, easy to apply in all settings and problems that would have a tractable solution if there were no LoD.

Mixture models are a prominent example in Dempster *et al.* [7], the seminal paper on the expectation–maximisation (EM) algorithm. The application is elaborated and its numerous extensions developed by McLachlan and Peel [19] and others. A set of applications in medical sciences is presented by Schlattmann [28]. For applications to large-scale surveys of household income, see Longford and Pittau [18]. The issue of model selection (how many mixture components to declare) is addressed by Hennig and Liao [11]. Some aspects of fitting the Bayesian versions of these models by Markov chain Monte Carlo methods are discussed by Richardson and Green [25]. A wealth of material relevant to mixtures, including many applications, are collected in Mengersen *et al.* [27].

2. Methods

In this section, we apply mixture modelling to data introduced in Figure 1, which is subject to LoD. We consider a mixture model with two components. One component involves

ordinary regression of Tg-post on the three recorded background variables, Tg-pre, sex and age, and the other is a normally distributed random sample, involving no covariates. We cannot take for granted that the two fitted components will have the anticipated interpretation as qualified successes (regression) and total successes (random sample); see Longford [17] for a related discussion. However, a single regression would seem not to be sufficient and the data set is not extensive enough for a mixture model with more than two components.

Denote the value of LoD by L'. We want to estimate a parameter or, more generally, a *target* θ. If all the values were recorded precisely, not subject to LoD, we would evaluate an estimator $\hat{\theta}$ of θ. In our case, the smallest possible concentration is zero, so the absence of LoD is equivalent to LoD set to zero.

We address the problem of concentrations below LoD by the method developed by Longford [16]. It is based on the following idea. We can easily generate the hypothetical data set that would be recorded if LoD were equal to any value $L > L'$. We specify how to substitute values for the entries that are subject to LoD. Examples of such substitution processes are the (constant) zero, L, $\frac{1}{2}L$, and a random draw from the uniform distribution on $(0, L)$. For the actual limit L', we define $\hat{\theta}_{L'}$ as the estimator $\hat{\theta}$ evaluated on the data set generated by applying a substitution process. This estimator depends on both L' and the substitution process.

We can evaluate $\hat{\theta}_L$, the version of $\hat{\theta}_{L'}$ with L' replaced by L, for any value $L > L^{'}$, because we can easily construct the data set that would have been obtained if LoD were higher than L'. Of course, we cannot do this for any $L < L'$. We refer to $\hat{\theta}_L$, treated as a function of L, as the *estimation function* for $\hat{\theta}$. The analysis concludes by extrapolating the realised values of this function to $L = 0$.

This extrapolation step is difficult when $\hat{\theta}_L$ is not a smooth function or has substantial curvature. In our example, as well as in Longford [16], $\hat{\theta}_L$ is much smoother when L is substituted for all the values that are below LoD L. The main device to relieve the problem of extrapolation when $\hat{\theta}_L$ is distinctly nonlinear is to transform the scale of L to $M = g^{-1}(L)$, where g is an increasing function such that $\hat{\theta}_{g(M)}$, a function of M, would have very little curvature. This can be done by trial and error, although [16] makes some more clinical suggestions. Setting the values of L might appear as another issue, but it presents no problems when the original analysis is relatively simple. That is the case with our data, even though the analysis involves iterations.

We thus proceed by the following steps. For a selected set of values of L greater than L', we fit the mixture model with the two components described earlier (an ordinary regression and a random sample). In the analysis with a particular value of L, we reset to L all the values of the outcome (Tg-post) and of the principal covariate (Tg-pre) that would be below this LoD. The two variables are transformed as $\log(x + 1 - L)$, so that the values recorded as below LoD are transformed to zero and ordinary regression (including homoscedasticity) is palatable. Then for each target θ (e.g. a regression parameter) we plot the estimation function $\hat{\theta}_L$ and extrapolate it to $L = 0$. If this task is non-trivial, we experiment with transformations of L that convert the problem closer to one of linear extrapolation.

The values of age (in years) are truncated (left-censored) at 40, because age is meant to represent a handicap (the 'age' factor) in the surgery, applicable only to older patients. This truncation affects 23 patients (25%). Details of fitting the mixture model with two

components by the EM algorithm are given in the Appendix. The algorithm is implemented as a user-defined function in R [23]. We assume that the variances of the two components are identical. This constraint is implemented by pooling the variance estimates in the M-step of each iteration. The weights in the pooling are set to the current estimates of the marginal probabilities of the components. Without this constraint, the estimates of the mean and variance of the random sample converge to zero.

As an aside, we note that LoD can be regarded as a problem of missing data, and addressed also by the EM algorithm. In the terminology established by Dempster *et al.* [7], the recorded values form the *incomplete data set*. The *complete data set* coincides with it for values above LoD, and is known only to be in the interval $(0, L')$ for the rest of the values. The E-step of the EM algorithm entails estimation of the summaries of the complete data that appear linearly in the complete-data loglikelihood. The M-step applies the complete-data method, that is, evaluation of $\hat{\theta}$, using the summaries obtained in the preceding E-step. The E- and M-steps are iterated till convergence. Combining two kinds of EM algorithm, one for fitting a mixture and the other for LoD, raises some computational difficulties, including slow convergence. That is why we prefer to deal with LoD by extrapolation.

3. Results

For orientation, we first give details of the model fit for $L = 0.5$, selected arbitrarily. This limit applies to 13 values before and 63 values (69%) after the therapy. These values are reset to 0.5, and the transformation $\log(1 - L + x)$ moves them to zero. The other transformed values, those with $x > L$, are positive. The iterations of the EM algorithm require an initial solution. We use the ordinary regression fit to all the values above the limit L for component 1 and 0.0 for the expectation in component 2. The marginal probability $\hat{p}$ is set initially to 0.5 for both components. The iterations are stopped when the precision to eight decimal places is reached. This takes 36 iterations. However, the estimates and the value of the deviance (-2 loglikelihood) do not change in the first five decimal places in the last 12 iterations. With other choices of initial values, for $\hat{p}$ in particular, we obtain the same solution after a similar number of iterations. Setting the convergence criterion to eight decimal places may appear excessive, but it is useful for checking that the algorithm is numerically stable and the convergence is genuine. The additional computing is of no consequence. We have checked on several examples that the same result is obtained with a wide range of initial solutions. As anticipated by Dempster *et al.* [7], the first few iterations move the estimates most of the way toward the solution, and then the convergence slows down. Methods for accelerating the convergence, such as [21], are not necessary in our case.

The estimates and the fitted deviance are displayed in Table 1. The first row gives the initial solution (for completeness), and the second row the fit obtained at the convergence of the EM algorithm. In component 1, the regression on Tg-pre is important.

Table 1. Fit of the two-component mixture with $L = 0.5$.

	Component 1						Component 2	
Iteration	Intercept	Tg-pre	Age	Sex	Res. var.	p	Mean	Deviance
0	−1.455	0.632	0.016	0.170	1.078	0.500	0.000	265.06
36	−1.431	0.659	0.024	−0.215	0.197	0.318	0.084	167.23

For example, the predicted difference between two subjects who differ only in the values of Tg-pre by 2.0 points on the log scale (e.g. one value $e^{2.0} = 7.4$ times greater than the other) is 1.32 on the log scale, which corresponds to $e^{1.32} = 3.74$ times greater value. A more appropriate comparison is $\exp(1.32 + 0.197/2) = 4.12$, using the identity $\mathrm{E}\{\exp(X)\} = \exp\{\mathrm{E}(X) + \frac{1}{2}\,\mathrm{var}(X)\}$ for a normally distributed random variable X. Further adjustment should be made for the uncertainty about the moments of X, see Longford [15]. This is not necessary for an informal comparison of the importance of the regression parameter estimates. Since we ignore the shift in the log-transformation, by $1-L$, these statements hold only approximately, for values much greater than L.

Two subjects who differ in their backgrounds only in age, by 30 years (e.g. one 40 years or younger and the other 70 years old), differ in their predicted outcomes on the log scale by $30 \times 0.024 = 0.71$, that is, one expected outcome is about twice as large as the other. The difference between the two sexes (estimate -0.215) is much smaller, corresponding to about nine years of difference in age. These statements ignore the sampling variation of the estimators. The comparison of the deviances suggests that the mixture model fits much better than the regression on its own, which we used as the initial solution. We note, however, that the distribution of the related likelihood ratio test statistic is not χ^2_2 because the related null-hypothesis is at the boundary of the parameter space.

Figure 2 presents the fit of the mixture model graphically. The subjects are represented in the diagram by points with colour indicating sex (grey for men) and size linearly related to the conditional probability of belonging to the first component (regression). These probabilities are evaluated in each E-step of the EM algorithm; see the Appendix for details. We use the probabilities obtained in the concluding iteration. The assignment to the components is as anticipated. The subjects with outcomes below LoD are assigned to the second (random-sample) component with high probabilities. The parallel straight lines in the diagram are for predicting log(1 + Tg-post) for 50- and 60-year-old patients. The lines are $-1.431 + 0.659x + 0.024a - 0.215\,s$, where $x = \log(1 + \text{Tg-pre})$, a is the age in years (set to 50 or 60), and $s = 1$ for women and $s = 2$ for men. The fitted lines for 60-year-old men and 50-year-old women nearly coincide.

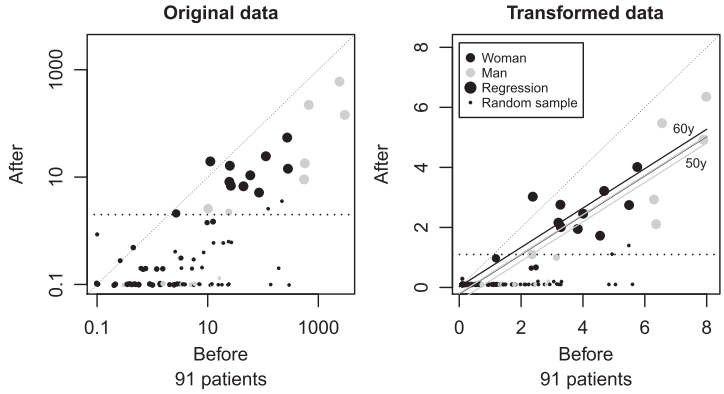

Figure 2. Mixture model fit with $L = 0.5$. The size of the symbols is linearly related to the conditional probability of belonging to component 1.

The mixture model parameter estimates with LoD (re-)set to values of $L = 0.1, 0.2, \ldots, 0.9$ are plotted in Figure 3. There is no point in evaluating them on a finer grid because most values of Tg-post are rounded and would not be altered by a small change of L. The extrapolation task is simple for the intercept (-1.57), the slope on Tg-pre (0.69), the residual variance (0.22) and the marginal probability of component 1 (0.36). The results of extrapolation are given in parentheses.

Extrapolation for age entails some uncertainty that is difficult to resolve because the apparent curvature is the result of a single evaluation, at $L' = 0.1$. Without it, the extrapolation task would be simple and would yield the estimate 0.0242. If we ignore the estimates for $L > 0.2$, we would obtain the estimate 0.0235 at $L = 0$. The range of uncertainty, 0.0007, corresponds to the difference of 0.021 on the log scale for two subjects who differ by 30 years of age. That corresponds to about 2.1% greater or smaller contrast of predictions for two such subjects on the original scale. This we regard as insubstantial, especially in view of the sampling variation of the slope on age, which we study in the next section.

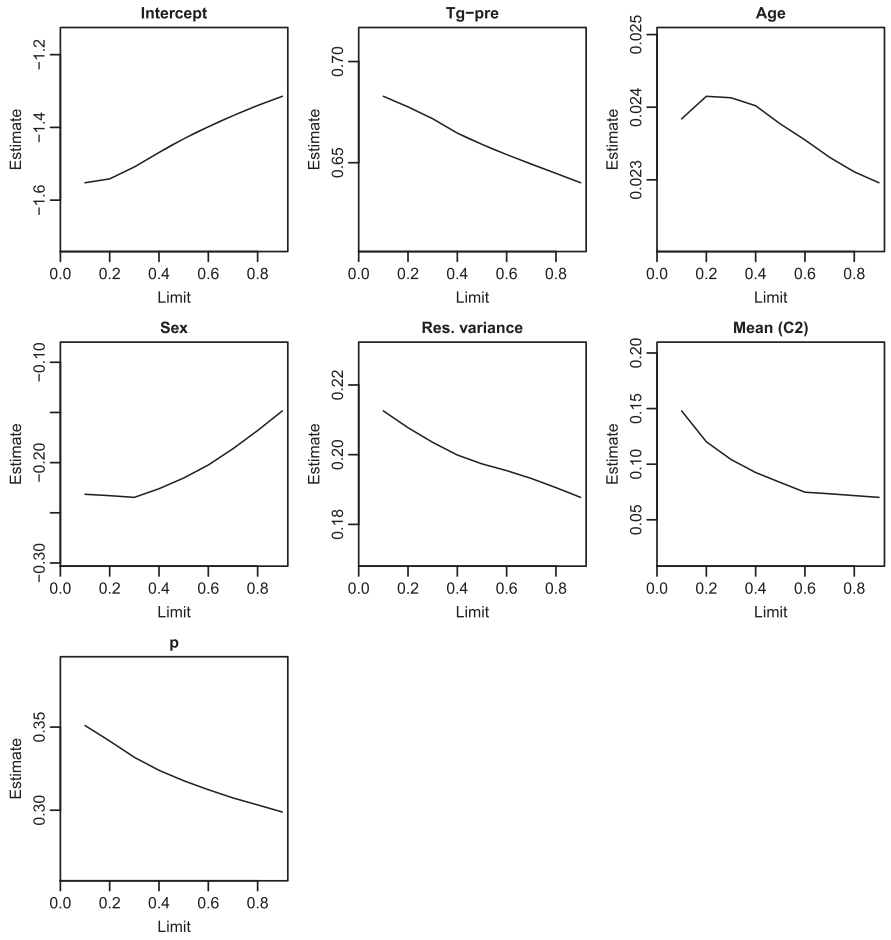

Figure 3. Extrapolation of the mixture-model parameter estimates from $L \geq L' = 0.1$ to $L = 0$.

Estimation of the sex difference entails a similar problem, although less acute, because the curvature of the function of estimates is only mild, and the estimation function is very close to linearity for $L \in (0.1, 0.3)$. The extrapolation, using no sophisticated device, yields a value between -0.25 and -0.23. The range of uncertainty, 0.02, is similar to the uncertainty about predictions for two subjects 30 years apart in age. This uncertainty is much smaller than the sampling variation of the estimator of the sex difference estimated in the next section. (Recall that there are only 17 men in the sample.)

We illustrate linearisation of the extrapolation task on estimation of the mean of component 2, denoted by μ. The left-hand panel of Figure 4 reproduces the function $\hat{\mu}_L$ from Figure 3. Powers, or the Box–Cox family of transformations, are always the first candidates for transforming L, but in this instance they are not useful. We therefore resort to other well-known families of functions. We consider transformations $\lambda = \log(CL + 1)$, where $C > 0$ is a parameter to be set. For small values of C, $\hat{\mu}_L$ is a convex function of λ, but its curvature diminishes with increasing C, and for $C \doteq 20$ it deviates from linearity only slightly. The transformation $g(L) = \log(20L + 1)$ is applied in the right-hand panel of Figure 4. Two alternative linear extrapolations are indicated by dashes. They result in 0.200 and 0.205. We regard the associated uncertainty, additional to sampling variation, as acceptable.

The assignment of the subjects to the two components may be of interest. Estimation of the 91 conditional probabilities requires extrapolation of their values from $L \geq 0.1$ to $L = 0$. The task is illustrated in Figure 5, where these conditional probabilities are plotted as functions of L. For most subjects, the estimated conditional probability depends on L very weakly and extrapolation is very simple. In a few cases, the estimation function changes its direction abruptly, and the extrapolation to $L = 0$ is subject to considerable uncertainty. However, the assignment is uncertain whenever the probability of assignment (to component 1), denoted by $\hat{r}$, is distant from both zero and unity. Thus, some subjects almost certainly belong to component 2: 12 have conditional probabilities $\hat{r}$ smaller than 0.01 for all values of L and further 8 have $\hat{r}$ smaller than 0.05. Fourteen subjects almost certainly belong to component 1, since their values of $\hat{r}$ are greater than 0.99 for all $L > 0.1$; their estimation functions are overprinted in the diagram. One further subject has $\hat{r} > 0.95$ throughout.

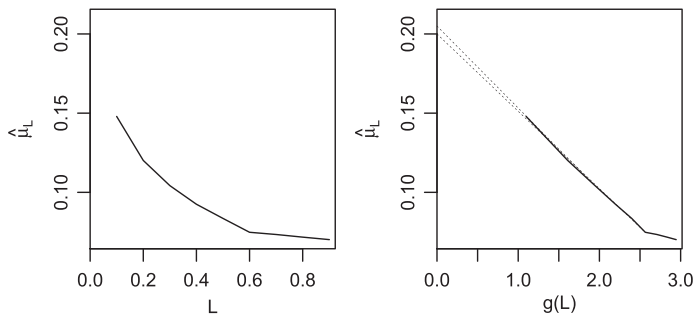

Figure 4. Linearising the extrapolation task for the mean of component 2, using transformation $g(L) = \log(20L + 1)$.

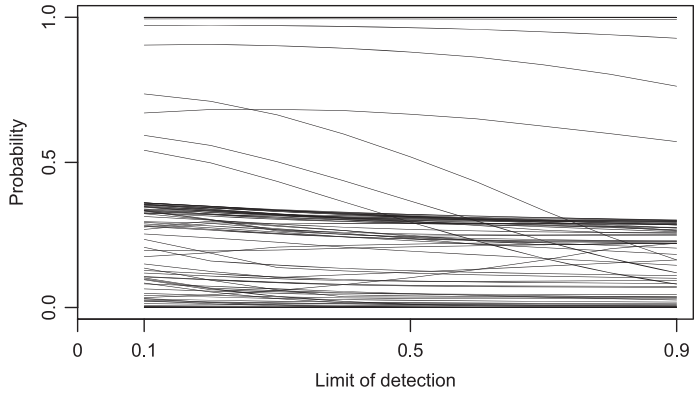

Figure 5. Estimation functions for the conditional probabilities of assignment to component 1.

The four subjects with probabilities $\hat{r}$ in the range $(0.5, 0.8)$ for $L = 0.1$ have moderate values of Tg-pre, in the range $(2.7, 24.0)$, and values of the outcome Tg-post in the range $(1.4, 2.2)$, close to the standard of 2.0, so it is indeed difficult to judge whether they are qualified or unqualified successes. The cluster of subjects with values of $\hat{r}$ around 0.35, close to the fitted marginal probability $\hat{p} = 0.36$, have very low values of both Tg-pre and Tg-post. They are at the intersection of the regression fitted for component 1 and the mean fitted for component 2.

3.1. Sampling variation

The (estimated) standard errors of the parameter estimates cannot be obtained from the EM algorithm directly. The standard errors that the M-step analysis yields at the concluding iteration underestimate the relevant targets because they ignore the uncertainty about the components to which the observations belong. The LoD presents additional uncertainty that compounds the problem of estimating the standard errors.

We estimate the standard errors by bootstrap, Davison and Hinkley [6] and Efron and Tibshirani [8]. We replicate the analysis 200 times on samples of units drawn from the realised data set at random and with replacement. Each sample has size 91, the same as the realised data set. Figure 6 illustrates the extrapolation task for the 200 bootstrap replicate model fits. It shows that the task is simple for all but a few replicates for each parameter, but the replicate extrapolations to $L = 0$ have substantial sampling variation. It confirms that the uncertainty about the extrapolation is small in relation to sampling variation.

The correct application of bootstrap entails extrapolation for each replication and parameter, that is, 200×7 extrapolation tasks, several of them requiring some improvisation because the estimation functions are neither smooth nor linear. We avoid this by extrapolating the averages over the bootstrap replicates (solid grey line in each panel). Confidence intervals for the parameters can be established similarly, by extrapolating the 2.5th and 97.5th percentiles of the bootstrap replicates at the values of L. As an alternative, the (symmetric) confidence intervals can be based on the standard errors estimated by bootstrap. In the diagram they are drawn by grey dashes.

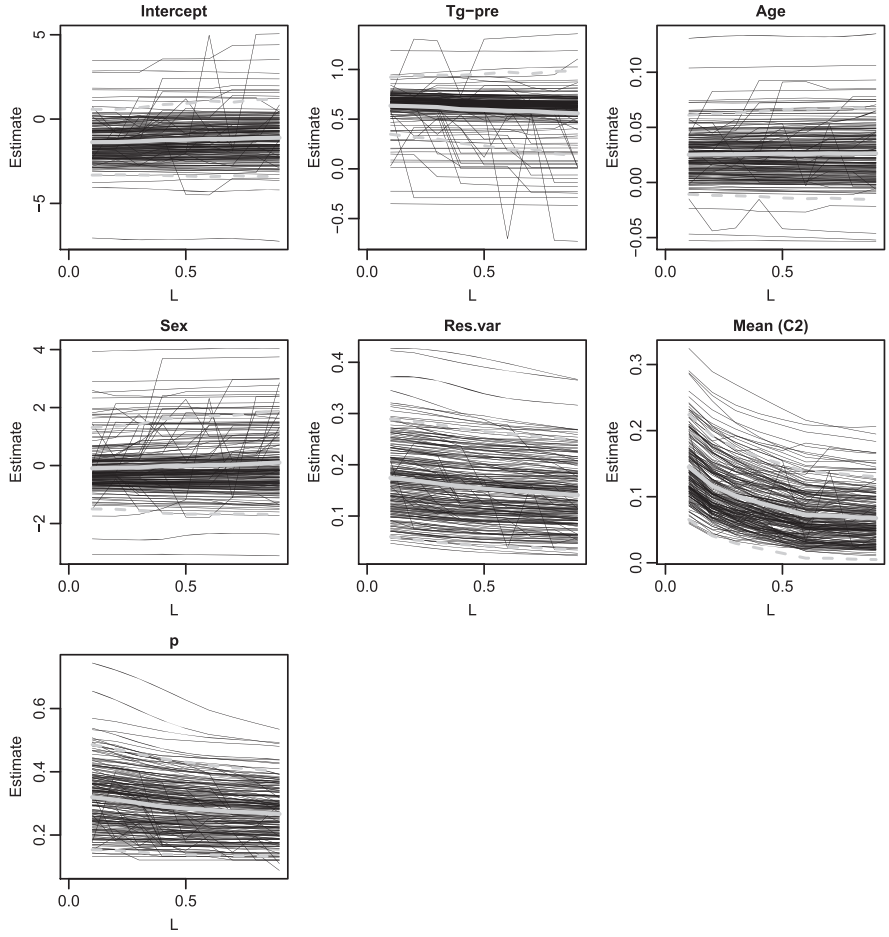

Figure 6. Bootstrap replicate analyses with increased LoD. The averages of the replicate estimates are drawn by thick grey lines and the 95% pointwise bootstrap confidence limits by grey dashes.

The estimates and estimated standard errors obtained by extrapolation of the bootstrap means and standard deviations are displayed in Table 2. They reflect the substantial sampling variation observed in Figure 6 and confirm that the uncertainty stemming from LoD is a small part of the overall uncertainty about the estimated parameters. Note that the sampling distribution of the mean of component 2 is skewed to the right. The cause of this is the lower bound of zero for the outcomes on the log scale.

In the bootstrap procedure, we drew samples with no stratification. If we stratified on sex, so that every sample would have exactly 17 men, the standard error for sex would be smaller, and other standard errors would probably also be reduced. This conclusion is based on the assumption that the sampling variance of an estimator has the form $v = D\sigma^2(1/n_{\mathrm{M}} + 1/n_{\mathrm{F}})$, where n_{F} and n_{M} are the numbers of women and men in the sample and D is a positive scalar. The smaller of the sample sizes, n_{M}, has a stronger influence

Table 2. Estimates and estimated standard errors of the mixture model parameters.

	Component 1						Component 2
	Intercept	Tg-pre	Age	Sex	Res. var.	p	Mean (C2)
Estimate	−1.57	0.70	0.024	−0.23	0.22	0.20	0.36
Bootstrap							
Mean	−1.38	0.64	0.025	−0.12	0.18	0.20	0.33
St. dev.	1.20	0.16	0.021	0.82	0.07	0.05	0.11

on ν. With stratification, $1/n_{\mathrm{M}}$ is constant across replications; without it, it varies, and its expectation is smaller than the realised value $1/n_{\mathrm{M}}$, because $1/x$ is a concave function. Of course, the scalar D may also vary across replications; that is why our conclusion is tenuous. However, we believe that the original sample was not stratified, and therefore bootstrap with stratification would represent the sampling variation with a bias.

4. Conclusion

The analysis presented in this paper illustrates a modular approach in which methods for important features of the data (LoD and mixture) are seamlessly combined. The combination is computationally much more demanding than each method in isolation, but the programming effort additional to the two methods is only modest.

This approach is applicable generally to analysing data that has a nuisance feature characterised by a parameter with a known value, in our case LoD with $L = 0.1$ and, in the original application, Cook and Stefanski [5], measurement error with a known or estimated variance. Inference is sought for the setting in which the feature is absent ($L = 0$). The method devised for this setting is applied to the data set with the feature inflated (LoD increased) to a range of levels ($L > 0.1$), and the results are extrapolated to $L = 0$. This is not a mechanical fail-safe process, but when it works we need no methodological development that is specific to a narrow class of problems. The only diagnostics additional to that required for the original method is to assess the quality of the extrapolation, and that is straightforward when the extrapolation is linear.

The results of the analysis support the original conjecture that the mixture components represent two groups of patients: those for whom the surgery and the subsequent therapy were an unqualified success, and those left with some residue of thyroid cancer. This is a step toward the goal of setting the dosage of I–131 to near the patient-specific minimum for which a low value of Tg-post is obtained after the treatment. This has to be achieved without any experimentation. The optimal dosage is likely to depend on several factors related to the stage of the disease and the initial compromise of the patient's immune system, which would have to be recorded.

The code developed specifically for this application can be obtained from the first author (NTL) on request.

Acknowledgments

The authors are grateful to doctors Sánchez de Guzmán and Mejía for providing the data analysed in this paper. Research described in this manuscript was initiated when the first author was a visitor in the School of Statistics, Universidad del Valle, Cali, Colombia.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Nicholas T. Longford http://orcid.org/0000-0003-4129-9726

References

[1] E.A. Asare and T.S. Wang, *Comparative effectiveness in thyroid cancer: Key questions and how to answer them*, Cancer Treat. Res. 164 (2015), pp. 67–87.
[2] S. Ballal, R. Soundararajan, A. Garg, S. Chopra, and C. Bal, *Intermediate-risk differentiated thyroid carcinoma patients who were surgically ablated do not need adjuvant radioiodine therapy: Long-term outcome study*, Clin. Endocrinol. 84 (2016), pp. 408–416.
[3] R.W. Browne and B.W. Whitcomb, *Procedures for determination of detection limits: Application to high-performance liquid chromatography analysis of fat soluble vitamins in human serum*, Epidemiology 21 (2010), pp. 4–9.
[4] R.J. Carroll, D. Ruppert, and L.A. Stefanski, *Measurement Error in Nonlinear Models*, Chapman & Hall, London, 1995.
[5] J. Cook and L.A. Stefanski, *A simulation-extrapolation method for parametric measurement error models*, J. Amer. Statist. Assoc. 89 (1994), pp. 1314–1328.
[6] A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge, 1997.
[7] A.P. Dempster, N.M. Laird, and D.B. Rubin, *Maximum likelihood estimation from incomplete data via the EM algorithm*, J. Roy. Statist. Soc. Ser. B 39 (1977), pp. 1–38.
[8] B. Efron and R. Tibshirani, *An Introduction to Bootstrap*, Chapman & Hall, London, 1993.
[9] G.E. Guzmán, G. Sánchez de Guzmán and J.R. Tovar, *Relación entre la presencia de tiroiditis linfocítica con la gravedad y recurrencia/persistencia del carcinoma diferenciado de tiroides (In Spanish: Relationship between the presence of lymphocytic thyroidis and the severity and recurrence/persistence of differentiation thyroid carcinoma)*, Rev. Argent. Endocrinol. Metab. 51 (2014), pp. 169–177.
[10] H. Hänscheid, M. Lassmann, A.K. Buck, C. Reiners, and F.A. Verburg, *The limit of detection in scintigraphic imaging with I–131 in patients with differentiated thyroid carcinoma*, Phys. Med. Biol. 59 (2014), pp. 2353–2368.
[11] C. Hennig and T.F. Liao, *How to find an appropriate clustering for mixed type variables with application to socio-economic stratification*, Appl. Statist. 62 (2013), pp. 309–369.
[12] R.T. Kloos, *Papillary thyroid cancer: medical management and follow-up*, Curr. Treat. Options Oncol. 6 (2005), pp. 323–338.
[13] D. Lambert, B. Peterson, and I. Terpenning, *Nondetects, detection limits, and the probability of detection*, J. Amer. Statist. Assoc. 86 (1991), pp. 266–277.
[14] H.N. Lee, J.Y. An, K.M. Lee, E.J. Kim, W.S. Choi, and D.Y. Kim, *Salivary gland dysfunction after radioactive iodine (I–131) therapy in patients following total thyroidectomy: emphasis on radioactive iodine therapy dose*, Clin. Imaging 39 (2015), pp. 397–400.
[15] N.T. Longford, *Inference with the lognormal distribution*, J. Statist. Plann. Inference 139 (2009), pp. 2329–2340.
[16] N.T. Longford, *Handling the limit of detection by extrapolation*, Stat. Med. 31 (2012), pp. 3133–3146.
[17] N.T. Longford, *Discussion of M.S. Handcock, A.E. Raftery, and J.M. Tantrum Model-based clustering for social networks*, J. Roy. Statist. Soc. Ser. A 170 (2007), pp. 301–354 (the discussion on p. 329).
[18] N.T. Longford and M.G. Pittau, *Stability of the household income in European countries in the 1990's*, Comput. Statist. Data Anal. 51 (2006), pp. 1364–1383.
[19] G. McLachlan and D. Peel, *Finite Mixture Models*, Wiley, New York, 2000.

[20] A. Mejía, J.R. Tovar, and C.T. Gutiérrez, *Níveis de tiroglobulina prévia à ablaç ao e persistência/recorrência precoce do câncer diferenciado da tireoide (In Portuguese: Pre-ablation levels of thyroglobulin and persistence/recurrence associated with early differentiated thyroid cancer)*, Rev. Cienc. Salud 12 (2014), pp. 9–21.
[21] X.-L. Meng and D. van Dyk, *The EM algorithm – an old folk-song sung to a fast new tune*, J.Roy. Statist. Soc. Ser. B 59 (1997), pp. 511–567.
[22] C.A. Milbury, Q. Zhong, J. Lin, M. Williams, J. Olson, D.R. Link, and B. Hutchison, *Determining lower limits of detection of digital PCR assays for cancer-related gene mutations*, Biomol. Detect. Quant. 1 (2014), pp. 8–22.
[23] R Development Core Team, *R: A Language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
[24] L.V. Rajakovic, D.D. Markovic, V.N. Rajakovic-Ognjanovic, and D.Z. Antanasijevic, *The approaches for estimation of limit of detection for IPC–MS trace analysis of arsenic*, Talanta 102 (2012), pp. 79–87.
[25] S. Richardson and P. Green, *On Bayesian analysis of mixtures with an unknown number of components*, J. Roy. Statist. Soc. Ser. B 59 (1997), pp. 731–792.
[26] M. Rubic, S.K. Kuna, V. Tesic, T. Samardzic, M. Despot, and D. Huic, *The most common factors influencing quality of life of thyroid cancer patients*, Psychiatr. Danub. 26 (2014), pp. 520–527.
[27] K. Mengersen, C. Robert, and M. Titterington (eds), *Mixtures: Estimation and Applications*, Wiley, Chichester.
[28] P. Schlattmann, *Medical Applications of Finite Mixture Models*, Springer, Heidelberg, 2009.
[29] T. Stamey, *Lower limits of detection, biological detection limits, functional sensitivity, or residual cancer detection limit? Sensitivity reports on prostate-specific antigen assays mislead clinicians*, Clin. Chem. 46 (1996), pp. 849–852.
[30] R.M. Tuttle, R. Leboeuf, and A.J. Martorella, *Papillary thyroid cancer: monitoring and therapy*, Endocrinol. Metab. Clin. North America 36 (2007), pp. 753–778.
[31] S. Vogiatzi, A. Liossis, and M. Lamprinakou, *Thyroid cancer radioiodine therapy: Health service performance and radiation safety*, Radiat. Prot. Dosimetry 165 (2015), pp. 434–438.
[32] L. Wartofsky and D. Van Nostrand (eds), *Thyroid Cancer: A Comprehensive Guide to Clinical Management*, 3rd ed., Springer, New York, 2016.
[33] A. Vexler, A. Liu, E. Eliseeva, and E.F. Schisterman, *Maximum likelihood ratio tests for comparing the discriminatory ability of biomarkers subject to limit of detection*, Biometrics 64 (2008), pp. 895–903.
[34] A. Vexler, J. Yu, and A.D. Hutson, *Likelihood testing populations modeled by autoregressive process subject to the limit of detection in applications to longitudinal biomedical data*, J. Appl. Statist. 38 (2011), pp. 1333–1346.

Appendix. EM algorithm for mixtures

We assume that one component of the mixture follows an ordinary regression, with its vector of observations $\mathbf{y}_1 = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}_{n_1}, \sigma^2\mathbf{I}_{n_1})$, and the other is a normal random sample, $\mathbf{y}_2 \sim \mathcal{N}(\mathbf{0}_{n_2}, \sigma^2\mathbf{I}_{n_2})$; n_1 and n_2 are the (unknown) sample sizes of the components; $\mathbf{0}$ and $\mathbf{I}$ are the vector of zeros and the identity matrix, with its sizes given in the subscripts. The first iteration of the EM algorithm requires an initial solution. Its setting is not critical. The obvious choice is the ordinary regression fit to all the values above LoD for component 1, and zero mean for component 2.

An iteration of the EM algorithm comprises an E- and an M-step. In the E-step, the conditional probability of belonging to each component is estimated for every observation. For component 1, this is equal to

$$\hat{r}_{i1} = \frac{\hat{p}_1 f_1(y_i)}{\hat{p}_1 f_1(y_i) + (1-\hat{p}_1) f_2(y_i)},$$

where $f_1(y_i) = \phi(y_i; \mathbf{x}_i\hat{\boldsymbol{\beta}}, \hat{\sigma})$ is the density of component 1 and $f_2(y_i) = \phi(y_i, \hat{\mu}, \hat{\tau})$ the density of component 2 defined for observation i; ϕ denotes the standard normal density. The densities are

evaluated at the current (provisional) fit $(\hat{\boldsymbol{\beta}}, \hat{\sigma})$. For component 2, $\hat{r}_{i2} = 1 - \hat{r}_{i1}$. Note that $f_1, f_2, \hat{p}_1$ and $\hat{r}$ are iteration specific, but we do not indicate this by an (additional) index t, to have a less cluttered notation.

In the M-step, the models for the two components are fitted separately, using the conditional probabilities $\hat{r}_{kh}$ as weights. The regression parameter vector is updated in iteration t as

$$\hat{\boldsymbol{\beta}}_t = (\mathbf{X}^\top \hat{\mathbf{R}}_t \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{R}}_t \mathbf{y},$$

where $\hat{\mathbf{R}}_t$ is the diagonal matrix of the probabilities $\hat{\mathbf{r}} = (\hat{r}_{11}, \hat{r}_{21}, \ldots, \hat{r}_{n1})^\top$, $n = 91$. The mean of component 2 is estimated by $\hat{\mu}_t = \mathbf{y}^\top(\mathbf{1} - \hat{\mathbf{r}})/(n - \mathbf{1}^\top \hat{\mathbf{r}})$; $\mathbf{1}$ is the vector of ones. The residual variance is estimated by pooling the estimates

$$\frac{1}{\hat{r}_+}(\mathbf{y}^\top \hat{\mathbf{R}}_t \mathbf{y} - \hat{\boldsymbol{\beta}}^\top \mathbf{X}^\top \hat{\mathbf{R}}_t \mathbf{X} \hat{\boldsymbol{\beta}}),$$

where $\hat{r}_+ = \hat{\mathbf{r}}^\top \mathbf{1}$ and $(\mathbf{y} - \hat{\mu})^\top (\mathbf{y} - \hat{\mu})/(n - \hat{r}_+)$, with weights $\hat{r}_+$ and $n - \hat{r}_+$.

The iterations are stopped when the values of the parameter estimates (and of the fitted incomplete-data loglikelihood) are changed by less than a prescribed value, such as 10^{-8}. We evaluate the negative log-norm $Q_t = -\log_{10}(\|\boldsymbol{\xi}_t - \boldsymbol{\xi}_{t-1}\|)$ for the vector of all estimates $\boldsymbol{\xi}$ (supplemented by other quantities, such as the loglikelihood), and the iterations are stopped when $Q_t > 8$; Q_t can be loosely interpreted as the number of decimals in the achieved precision.