RESEARCH ARTICLE

Distribution–free hyperrectangular tolerance regions for setting multivariate reference regions in laboratory medicine

Wei Liu[1] | Frank Bretz[2] | Mario Cortina–Borja[3]

[1]Southampton Statistical Sciences Research Institute and School of Maths, University of Southampton, Southampton, SO17 1BJ, UK

[2]Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

[3]Population, Policy and Practice Research and Teaching Department, Great Ormond Street Institute of Child Health, University College London, London, WC1N 1EH, UK

Correspondence
Mario Cortina–Borja, Population, Policy and Practice Research and Teaching Department, Great Ormond Street Institute of Child Health, University College London, London, WC1N 1EH, UK.
Email: m.cortina@ucl.ac.uk

Reference regions are important in laboratory medicine to interpret the test results of patients, and usually given by tolerance regions. Tolerance regions of $p\ (\geq 2)$ dimensions are highly desirable when the test results contains p outcome measures. Nonparametric hyperrectangular tolerance regions are attractive in real problems due to their robustness with respect to the underlying distribution of the measurements and ease of intepretation, and methods to construct them have been recently provided by Young and Mathew [Stat Methods Med Res. 2020;29:3569-3585]. However, their validity is supported by a simulation study only. In this paper, nonparametric hyperrectangular tolerance regions are constructed by using Tukey's [Ann Math Stat. 1947;18:529-539; Ann Math Stat. 1948;19:30-39] elegant results of equivalence blocks. The validity of these new tolerance regions is proven mathematically in [Ann Math Stat. 1947;18:529-539; Ann Math Stat. 1948;19:30-39] under the only assumption that the underlying distribution of the measurements is continuous. The methodology is applied to analyze the kidney function problem considered in Young and Mathew [Stat Methods Med Res. 2020;29:3569-3585].

KEYWORDS
nonparametric tolerance interval, nonparametric tolerance region, reference range, reference region, tolerance interval, tolerance region

1 | INTRODUCTION

Reference intervals, or regions if multiple outcome measures are involved, are powerful tools in laboratory medicine to aid decision making and their use has become increasingly prevalent in clinical practice.[1-6] The comparator limits assist the clinician in determining a context for an individual value, in order to answer the natural question from the patient "are my test results typical with respect to a healthy population?"

Incorrectly estimating the reference interval of a sensitive clinical marker of physiological function has enormous public health implications. For example, underestimating the upper limit of a reference interval would mean classifying a large number of people as diseased thus affecting the doses of medication prescribed.[7,8] Construction of appropriate reference intervals is therefore crucial in laboratory medicine practice. Well–known general references are available.[1-3,5,9,10]

Reference intervals are constructed from observed sample data and purport to include a pre-specified $100P\%$ (commonly set to 95%) of the values in the population of interest. Their main point is to classify any future observations that fall outside these intervals as atypical and thus may warrant further investigation.

As the reference intervals depend on the random sample, the proportion of the population contained in the intervals is also random. Hence all one can hope for is that the reference intervals contain $100P\%$ of the population with a large confidence with respect to the randomness in the sample. Since tolerance intervals are constructed to contain $100P\%$ of the population values with a pre-specified large confidence γ, they are eminently suitable as reference intervals.[9,11-16] Since the first publication on tolerance intervals by Wilks,[17] various parametric and nonparametric procedures are readily available for use as reference intervals.[16,18-24]

On the other hand, it is argued strongly in References [13-16] that P-confidence prediction intervals (which are also called P-expectation tolerance intervals) should not be used as reference intervals since prediction intervals contain $100P\%$ of the population values with a probability as low as half especially when the sample size is large.[13] So prediction intervals or regions are not considered in this paper.

While tolerance intervals for univariate measurements are well-established, multivariate tolerance regions are limited,[9,24] either specifically for multivariate normal distributions [16,25-29] or nonparametric.[18,30-33] Young and Mathew [18] and Lucagbo and Mathew [16] provide excellent and timely reviews of the topic.

It is argued eloquently in Wellek [33] that hyperrectangular reference regions for multivariate measurements are particularly useful due to their ease of interpretation. Specifically, a hyperrectangular reference region of any dimension $p(\geq 2)$ can be used to assess straightforwardly whether a future observation is outside the region and so atypical and, if it is, in which of the p measurements. It is also known [34] that parametric tolerance intervals could have true confidence level dramatically smaller than the nominal level γ when the underlying distribution departs only very slightly from the assumed form of the parametric distribution, which highlights the importance of nonparametric tolerance intervals. It is therefore timely that Young and Mathew [18] have constructed nonparametric hyperrectangular tolerance regions as reference regions. However the validity of their method is supported only by a simulation study for the specific underlying distributions considered in their paper.

In this article, nonparametric hyperrectangular tolerance regions of $100P\%$ content and γ confidence are constructed by using Tukey's [30,31] equivalence blocks (EBs). They are distribution–free in that the "$100P\%$ content and γ confidence" requirement is guaranteed so long as the underlying multivariate distribution of the measurements is continuous. The gist of our construction method is that the true confidence level can be computed exactly depending on the total number of EBs without specifying the multivariate probability model assumed to generate the observed data. Besides, as noted by Di Bucchianico et al [35] and Amerise,[36] there is no canonical ordering in higher dimensions thus auxiliary ordering functions are implied in Tukey's method. This means that one cannot look at the data and select the best ordering that results in the smallest hyperrectangle.

Distribution–free hyperrectangular tolerance regions are considered in Section 2. The methods described there are then applied in Section 3 to the real problem of defining reference regions for kidney function in normal adolescents considered in Young and Mathew.[18] Finally, concluding remarks are given in Section 4. The data and R code for reproducing all the results in this paper are available at http://www.personal.soton.ac.uk/wl/RefHypeRecta/.

2 | DISTRIBUTION–FREE HYPERRECTANGULAR TOLERANCE REGIONS

In this section, distribution–free hyperrectangular tolerance regions of $100P\%$ content and γ confidence are constructed using Tukey's [30,31] EBs.

Let $\mathbf{X} = (X_1, \ldots, X_p)'$ denote a (generic) random vector of p measurements of interest from an individual of the population, which has a continuous p-dimensional distribution with its cumulative distribution function (cdf) denoted as F. For example, for the kidney function problem considered in Section 3, $p = 3$ measurements are of interest, given by $X_1 =$ urine albumin–to–creatinine ratio (UACR), $X_2 =$ uric acid (UA), and $X_3 =$ serum creatinine (SC). A random sample of n observations from the population is available and denoted by

$$\mathbf{X}_1 = \begin{pmatrix} X_{11} \\ \vdots \\ X_{1p} \end{pmatrix}, \ldots, \mathbf{X}_n = \begin{pmatrix} X_{n1} \\ \vdots \\ X_{np} \end{pmatrix} \overset{i.i.d.}{\sim} F.$$

Let $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ denote the random sample, and we further assume that $\mathbf{X}$ is independent of $\mathcal{X}$ (and also has cdf F). A tolerance region of $100P\%$ content and γ confidence is constructed using the sample $\mathcal{X}$, is denoted as $T(\mathcal{X})$, and satisfies

$$\Pr_{\mathcal{X}} \left\{ \Pr_{\mathbf{X}|\mathcal{X}} \{\mathbf{X} \in T(\mathcal{X})\} \geq P \right\} \geq \gamma, \tag{1}$$

where the (inner) probability $\Pr_{\mathbf{X}|\mathcal{X}} \{\mathbf{X} \in T(\mathcal{X})\}$ is computed taking $\mathbf{X}$ as random while $\mathcal{X}$, and so the region $T(\mathcal{X})$, is fixed, and the (outer) probability $\Pr_{\mathcal{X}} \{\}$ is computed taking $\mathcal{X}$ as random. The (inner) probability $\Pr_{\mathbf{X}|\mathcal{X}} \{\mathbf{X} \in T(\mathcal{X})\}$ is called the coverage probability or content of $T(\mathcal{X})$, which clearly depends on $\mathcal{X}$ and so is random. The inequality in (1) guarantees that the content of $T(\mathcal{X})$ is at least P with a confidence of at least γ with respect to the randomness of the sample $\mathcal{X}$.

The elegant results of Tukey [30,31] state that if $T(\mathcal{X})$ is the union of $k\,(1 \leq k \leq n)$ EBs (whose construction will be considered below) then

$$\Pr_{\mathbf{X}|\mathcal{X}} \{\mathbf{X} \in T(\mathcal{X})\} \sim Beta(k, n - k + 1), \tag{2}$$

where $Beta(k, n - k + 1)$ denotes the beta distribution with parameters k and $n - k + 1$. The cdf of $Beta(k, n - k + 1)$ is denoted as $\mathcal{B}_{k,n-k+1}(\cdot)$. Hence if $T(\mathcal{X})$ is constructed as the union of k EBs then the requirement in (1) becomes

$$1 - \mathcal{B}_{k,n-k+1}(P) \geq \gamma. \tag{3}$$

In order to use the fewest EBs to form $T(\mathcal{X})$ as such union, the smallest $k\,(1 \leq k \leq n)$ that satisfies this constraint, denoted as k_0, will be used. It is straightforward to find k_0 from (3) numerically, and it is clear that k_0 depends on P, γ and n only. Due to the discreteness of the number of EBs k_0 used and the sample size n, the true confidence level of the tolerance region $T(\mathcal{X})$ is given by $1 - \mathcal{B}_{k_0,n-k_0+1}(P)$ from (3), and most likely to be larger than γ. This is of course well–known in the case of distribution–free tolerance intervals for one single measurement.[22]

It remains to show how the $n + 1$ EBs of Tukey [30,31] can be constructed based on the sample $\mathcal{X}$, from which a tolerance region $T(\mathcal{X})$ of hyperrectangular shape can be formed as the union of k_0 EBs for $p \geq 2$.

For the case of $p = 1$, let the order statistics of the sample of n observations on the only measurement be denoted as $X_{[1]1} < \cdots < X_{[n]1}$. Then the $n + 1$ EBs are given by

$$(-\infty, X_{[1]1}], \ (X_{[1]1}, X_{[2]1}], \ \ldots, \ (X_{[n-1]1}, X_{[n]1}], \ (X_{[n]1}, \infty).$$

The union of the first k_0 EBs is given by $T(\mathcal{X}) = (-\infty, X_{[k_0]1}]$, which is a $100P\%$ content and γ confidence upper tolerance interval for the population. The union of the last k_0 EBs is given by $T(\mathcal{X}) = (X_{[n-k_0+1]1}, \infty)$, which is a $100P\%$ content and γ confidence lower tolerance interval for the population. The union of the middle k_0 EBs is given by $T(\mathcal{X}) = (X_{[m]1}, X_{[m+k_0]1}]$ which is a $100P\%$ content and γ confidence two–sided tolerance interval for the population; here $m = (n - k_0 + 1)/2$ if $n - k_0 + 1$ is an even number, and $m = (n - k_0)/2$ or $(n - k_0)/2 + 1$ if $n - k_0 + 1$ is an odd number. Due to the assumption that the underlying distribution is continuous, there is no difference to the requirement (1) whether or not a boundary point $X_{[i]1}$ is included in the EB. The same is true for a general $p\,(\geq 1)$.[30,31]

2.1 | The case of $p = 2$

For $p = 2$, the $n + 1$ EBs can be constructed in the following way using Tukey's method [30,31] in order to form two–sided tolerance rectangles of the form $T(\mathcal{X}) = \left\{ (x_1, x_2)' : \ L_1(\mathcal{X}) < x_1 \leq U_1(\mathcal{X}), L_2(\mathcal{X}) < x_2 \leq U_2(\mathcal{X}) \right\}$.

Step 1 Order the n observations on the first measurement $X_{11}, \ldots, X_{n1}$ as $X_{[1]1} < \cdots < X_{[n]1}$. Construct the first EB as

$$B_1 = \left\{ (x_1, x_2)' : \ x_1 > X_{[n]1} \right\} \subset R^2$$

where R^2 denotes the whole (x_1, x_2)-plane.

Step 2 Delete the sample observation having the largest first measurement $X_{[n]1}$ and denote the remaining $n-1$ sample observations as $\mathbf{X}_{1,1}, \ldots, \mathbf{X}_{n-1,1}$ with $\mathbf{X}_{i,1} = (X_{i1,1}, X_{i2,1})'$ for $i = 1, \ldots, n-1$. Order the $n-1$ observations on the second measurement $X_{12,1}, \ldots, X_{(n-1)2,1}$ as $X_{[1]2,1} < \cdots < X_{[n-1]2,1}$. Construct the second EB as

$$B_2 = \left\{ (x_1, x_2)' : (x_1, x_2)' \notin B_1 \text{ and } x_2 > X_{[n-1]2,1} \right\} \subset R^2.$$

Step 3 Delete the sample observation having the largest second measurement $X_{[n-1]2,1}$ and denote the remaining $n-2$ sample observations as $\mathbf{X}_{1,2}, \ldots, \mathbf{X}_{n-2,2}$ with $\mathbf{X}_{i,2} = (X_{i1,2}, X_{i2,2})'$ for $i = 1, \ldots, n-2$. Order the $n-2$ observations on the first measurement $X_{11,2}, \ldots, X_{(n-2)1,2}$ as $X_{[1]1,2} < \cdots < X_{[n-2]1,2}$. Construct the third EB as

$$B_3 = \left\{ (x_1, x_2)' : (x_1, x_2)' \notin B_1 \cup B_2 \text{ and } x_1 \leq X_{[1]1,2} \right\} \subset R^2.$$

Step 4 Delete the sample observation having the smallest first measurement $X_{[1]1,2}$ and denote the remaining $n-3$ sample observations as $\mathbf{X}_{1,3}, \ldots, \mathbf{X}_{n-3,3}$ with $\mathbf{X}_{i,3} = (X_{i1,3}, X_{i2,3})'$ for $i = 1, \ldots, n-3$. Order the $n-3$ observations on the second measurement $X_{12,3}, \ldots, X_{(n-3)2,3}$ as $X_{[1]2,3} < \cdots < X_{[n-3]2,3}$. Construct the fourth EB as

$$B_4 = \left\{ (x_1, x_2)' : (x_1, x_2)' \notin B_1 \cup B_2 \cup B_3 \text{ and } x_2 \leq X_{[1]2,3} \right\} \subset R^2.$$

Step 5 Delete the sample observation having the smallest second measurement $X_{[1]2,3}$.
Starting now with the remaining $n-4$ sample observations, repeat the steps above in the given order until the n–th EB B_n is constructed. Note that it is possible that not the full cycle of all the four steps above may have been used by the time that B_n is constructed.
The last EB B_{n+1} is simply given by the complement in R^2 of the set $B_1 \cup \cdots \cup B_n$. Hence $B_1 \cup \cdots \cup B_n \cup B_{n+1} = R^2$.

Figure 1 gives an illustration on the construction of the $n+1 = 8$ EBs from a given sample of $n = 7$ observations. The $n = 7$ observations are plotted in the figure as crosses.
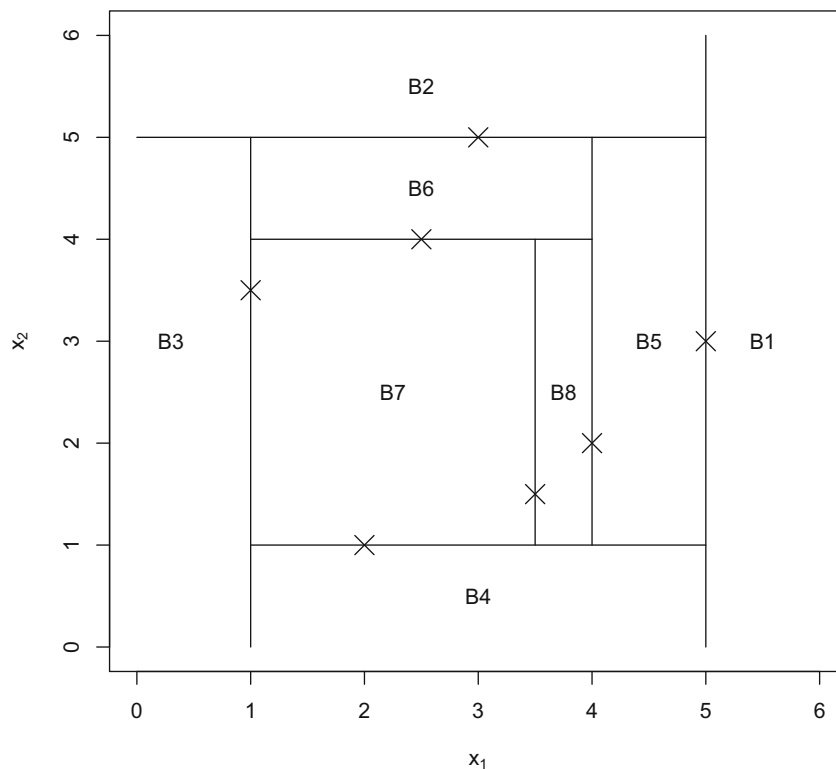

FIGURE 1 The 8 EBs forming two–sided tolerance rectangles constructed from the 7 observations denoted by crosses.

Now the union of the last k_0 EBs, given by $T(\mathcal{X}) = B_{n-k_0+2} \cup \cdots \cup B_{n+1}$, is a 100$P$% content and γ confidence tolerance region for the population. For the example in Figure 1, if $k_0 \leq 4$ then $T(\mathcal{X})$ is a rectangle; if $k_0 > 4$ then $T(\mathcal{X})$ is not a rectangle. In general, k_0 must be no larger than $n - 3$ in order that $T(\mathcal{X})$ is a rectangle. The existence of such a k_0 is guaranteed only if the sample size n is sufficiently large such that the inequality in (3) is satisfied with $k = n - 3$, that is,

$$1 - \mathcal{B}_{n-3,4}(P) \geq \gamma. \tag{4}$$

This minimum sample size can easily be computed numerically (eg, using our R code provided), and it depends on the given P and γ only. For example, for $P = 95\%$ and $\gamma = 0.95$, the sample size n must be at least 153 in order that $T(\mathcal{X})$ is a rectangle. If the available sample size is smaller than the minimum sample size then P or γ has to be reduced in order that $T(\mathcal{X})$ is a rectangle. For example, for $P = 45\%$ and $\gamma = 0.60$, the minimum sample size n is reduced to 7 in order that $T(\mathcal{X})$ is a rectangle. This consideration of minimum sample size requirement is similar to that developed for the one single measurement case discussed in Reference 22 in relation to constructing distribution–free tolerance intervals.

For given P, γ, and sample size n that is no smaller than the minimum sample size, we therefore are able to find k_0 from (3) and construct a distribution–free tolerance rectangle $T(\mathcal{X})$ of 100P% content and γ confidence as given above. As already pointed out, the true confidence level of the tolerance rectangle is given by $1 - \mathcal{B}_{k_0,n-k_0+1}(P)$, which is likely to be larger than γ. For example, when $P = 95\%$, $\gamma = 0.95$ and $n = 200$, k_0 is computed from (3) to be 196 and the true confidence level is $1 - \mathcal{B}_{k_0,n-k_0+1}(P) = 0.974$ which is larger than the nominal $\gamma = 0.95$. As another example, when $P = 45\%$, $\gamma = 0.60$ and $n = 7$, k_0 is computed to be 4 and the true confidence level is 0.608 which is larger than the nominal $\gamma = 0.60$. Hence for the example in Figure 1 which has sample size $n = 7$, the rectangle formed by the union of the last $k_0 = 4$ EBs, given by $T(\mathcal{X}) = B_5 \cup B_6 \cup B_7 \cup B_8$, is a $P = 45\%$ content and $\gamma = 0.60$ confidence tolerance region for the population.

It is noteworthy that there are numerous alternative ways to construct the EBs. For example, the order of the four steps above can be changed to any one of the 4! permutations of the four steps in one full cycle. Or, the orders of the four steps in different cycles could be altered. All these alternative constructions will result in valid EBs and tolerance rectangles so long as the orders are chosen independently of the sample $\mathcal{X}$.[30,31] That is, one cannot look at the data and select the best ordering in order to have the smallest rectangle $T(\mathcal{X})$. When the sample size n is sufficiently large, one can envisage that the difference between the resultant tolerance rectangles from these alternative constructions of EBs should be small. So this will not be pursued further in this paper.

Next we consider the construction of EBs in order to form one–sided tolerance "rectangles" of the form $T(\mathcal{X}) = \left\{ (x_1, x_2)' : x_1 \leq U_1(\mathcal{X}), x_2 \leq U_2(\mathcal{X}) \right\}$ (or upper semi-space in [1]).

Step 1 Order the n observations on the first measurement $X_{11}, \ldots, X_{n1}$ as $X_{[1]1} < \cdots < X_{[n]1}$. Construct the first EB as

$$B_1 = \left\{ (x_1, x_2)' : x_1 > X_{[n]1} \right\} \subset R^2.$$

Step 2 Delete the sample observation having the largest first measurement $X_{[n]1}$ and denote the remaining $n - 1$ sample observations as $\mathbf{X}_{1,1}, \ldots, \mathbf{X}_{n-1,1}$ with $\mathbf{X}_{i,1} = (X_{i1,1}, X_{i2,1})'$ for $i = 1, \ldots, n - 1$. Order the $n - 1$ observations on the second measurement $X_{12,1}, \ldots, X_{(n-1)2,1}$ as $X_{[1]2,1} < \cdots < X_{[n-1]2,1}$. Construct the second EB as

$$B_2 = \left\{ (x_1, x_2)' : (x_1, x_2)' \notin B_1 \text{ and } x_2 > X_{[n-1]2,1} \right\} \subset R^2.$$

Step 3 Delete the sample observation having the largest second measurement $X_{[n-1]2,1}$.
Starting now with the remaining $n - 2$ sample observations, repeat the steps 1 and 2 above in the given order until the n–th EB B_n is constructed. Note that it is possible that not the full cycle of all the two steps above are used when B_n is constructed.
The last EB B_{n+1} is simply given by the complement in R^2 of the set $B_1 \cup \cdots \cup B_n$.

This construction basically deletes the steps 3 and 4 in the construction of the two–sided tolerance rectangle in order not to have lower bounds on x_1 and x_2. Figure 2 gives an illustration on the construction of the $n + 1 = 8$ EBs from the same sample of $n = 7$ observations in Figure 1.
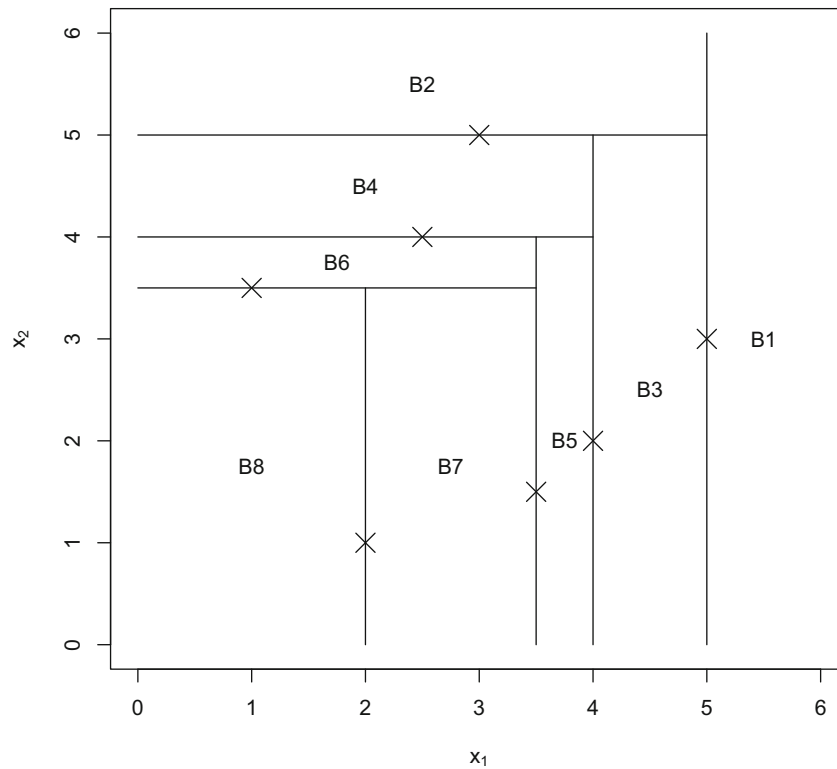
FIGURE 2   The 8 EBs forming one–sided tolerance rectangles from the same sample of $n = 7$ observations in Figure 1.

Now the union of the last k_0 EBs, given by $T(\mathcal{X}) = B_{n-k_0+2} \cup \cdots \cup B_{n+1}$, is a $100P\%$ content and γ confidence tolerance region for the population as before. For the example in Figure 2, if $k_0 \leq 6$ then $T(\mathcal{X})$ is a one–sided "rectangle" required; if $k_0 > 6$ then $T(\mathcal{X})$ is not a one–sided "rectangle". It is clear that, in general, k_0 must be no larger than $n - 1$ in order that $T(\mathcal{X})$ is a one–sided "rectangle". Similar to the two–sided case above, the existence of such a k_0 is guaranteed only if the sample size n satisfies (3) with $k = n - 1$, that is, $1 - \mathcal{B}_{n-1,2}(P) \geq \gamma$. For example, for $P = 95\%$ and $\gamma = 0.95$, the sample size n must be at least 93 in order that $T(\mathcal{X})$ is a one–sided "rectangle". For given P, γ, and n that is not smaller than the minimum sample size required, k_0 is computed from (3) as before. For the example in Figure 2 with sample size $n = 7$, $P = 45\%$ and $\gamma = 0.60$, k_0 is computed to be 4 as before. The one–sided "rectangle" formed by the union of the last $k_0 = 4$ EBs, given by $T(\mathcal{X}) = B_5 \cup B_6 \cup B_7 \cup B_8$, is therefore a $P = 45\%$ content and $\gamma = 0.60$ confidence tolerance region for the population. As before, its true confidence level is 0.608 which is larger than the nominal confidence level $\gamma = 0.60$.

The constructions for two–sided or one–sided tolerance "rectangles" given above can clearly be adapted to construct mix-sided tolerance "rectangles" of the form, say, $T(\mathcal{X}) = \left\{(x_1, x_2)' : L_1(\mathcal{X}) < x_1 \leq U_1(\mathcal{X}), x_2 \leq U_2(\mathcal{X})\right\}$. Details are omitted to save space.

2.2 | The case of $p \geq 2$

For a general $p \geq 2$, the $n + 1$ EBs can be constructed in the following way in order to form two–sided p–dimensional tolerance hyperrectangles (ie, boxes) $T(\mathcal{X}) = \left\{(x_1, \ldots, x_p)' : L_i(\mathcal{X}) < x_i \leq U_i(\mathcal{X}), i = 1, \ldots, p\right\}$.

Step 1   Order the n observations on the first measurement $X_{11}, \ldots, X_{n1}$ as $X_{[1]1} < \cdots < X_{[n]1}$. Construct the first EB as

$$B_1 = \left\{(x_1, \ldots, x_p)' : x_1 > X_{[n]1}\right\} \subset R^p$$

where R^p denotes the whole $(x_1, \ldots, x_p)$-space.

Step 2 Delete the sample observation having the largest first measurement $X_{[n]1}$ and denote the remaining $n - 1$ sample observations as $\mathbf{X}_{1,1}, \ldots, \mathbf{X}_{n-1,1}$ with $\mathbf{X}_{i,1} = (X_{i1,1}, X_{i2,1}, \ldots, X_{ip,1})'$ for $i = 1, \ldots, n-1$. Order the $n - 1$ observations on the second measurement $X_{12,1}, \ldots, X_{(n-1)2,1}$ as $X_{[1]2,1} < \cdots < X_{[n-1]2,1}$. Construct the second EB as

$$B_2 = \left\{ (x_1, \ldots, x_p)' : (x_1, \ldots, x_p)' \notin B_1 \text{ and } x_2 > X_{[n-1]2,1} \right\}.$$

$\ldots$

Step p Delete the sample observation having the largest $(p-1)$–th measurement $X_{[n-(p-2)](p-1),p-2}$ and denote the remaining $n - (p-1)$ sample observations as $\mathbf{X}_{1,(p-1)}, \ldots, \mathbf{X}_{n-(p-1),(p-1)}$ with $\mathbf{X}_{i,1} = (X_{i1,(p-1)}, X_{i2,(p-1)}, \ldots, X_{ip,(p-1)})'$ for $i = 1, \ldots, n-(p-1)$. Order the $n - (p-1)$ observations on the p–th measurement $X_{1p,(p-1)}, \ldots, X_{(n-(p-1))p,(p-1)}$ as $X_{[1]p,(p-1)} < \cdots < X_{[n-(p-1)]p,(p-1)}$. Construct the p–th EB as

$$B_p = \left\{ (x_1, \ldots, x_p)' : (x_1, \ldots, x_p)' \notin \cup_{i=1}^{p-1} B_i \text{ and } x_p > X_{[n-(p-1)]p,(p-1)} \right\}.$$

Step p + 1 Delete the sample observation having the largest p–th measurement $X_{[n-(p-1)]p,(p-1)}$ and denote the remaining $n - p$ sample observations as $\mathbf{X}_{1,p}, \ldots, \mathbf{X}_{n-p,p}$ with $\mathbf{X}_{i,p} = (X_{i1,p}, X_{i2,p}, \ldots, X_{ip,p})'$ for $i = 1, \ldots, n-p$. Order the $n - p$ observations on the first measurement $X_{11,p}, \ldots, X_{(n-p)1,p}$ as $X_{[1]1,p} < \cdots < X_{[n-p]1,p}$. Construct the $(p+1)$–th EB as

$$B_{p+1} = \left\{ (x_1, \ldots, x_p)' : (x_1, \ldots, x_p)' \notin \cup_{i=1}^{p} B_i \text{ and } x_1 \leq X_{[1]1,p} \right\}.$$

Step p + 2 Delete the sample observation having the smallest first measurement $X_{[1]1,p}$ and denote the remaining $n - (p+1)$ sample observations as $\mathbf{X}_{1,(p+1)}, \ldots, \mathbf{X}_{n-(p+1),(p+1)}$ with $\mathbf{X}_{i,(p+1)} = (X_{i1,(p+1)}, X_{i2,(p+1)}, \ldots, X_{ip,(p+1)})'$ for $i = 1, \ldots, n-(p+1)$. Order the $n - (p+1)$ observations on the second measurement $X_{12,(p+1)}, \ldots, X_{(n-(p+1))2,(p+1)}$ as $X_{[1]2,(p+1)} < \cdots < X_{[n-(p+1)]2,(p+1)}$. Construct the $(p+2)$–th EB as

$$B_{p+2} = \left\{ (x_1, \ldots, x_p)' : (x_1, \ldots, x_p)' \notin \cup_{i=1}^{p+1} B_i \text{ and } x_2 \leq X_{[1]2,(p+1)} \right\}.$$

$\ldots$

Step 2p Delete the sample observation having the smallest $(p-1)$–th measurement $X_{[1](p-1),2p-2}$ and denote the remaining $n - (2p-1)$ sample observations as $\mathbf{X}_{1,(2p-1)}, \ldots, \mathbf{X}_{n-(2p-1),(2p-1)}$ with $\mathbf{X}_{i,(2p-1)} = (X_{i1,(2p-1)}, X_{i2,(2p-1)}, \ldots, X_{ip,(2p-1)})'$ for $i = 1, \ldots, n-(2p-1)$. Order these $n - (2p-1)$ observations on the p–th measurement $X_{1p,(2p-1)}, \ldots, X_{(n-(2p-1))p,(2p-1)}$ as $X_{[1]p,(2p-1)} < \cdots < X_{[n-(2p-1)]p,(2p-1)}$. Construct the $(2p)$–th EB as

$$B_{2p} = \left\{ (x_1, \ldots, x_p)' : (x_1, \ldots, x_p)' \notin \cup_{i=1}^{2p-1} B_i \text{ and } x_p \leq X_{[1]\, p,(2p-1)} \right\}.$$

Step 2p + 1 Delete the sample observation having the smallest p–th measurement $X_{[1]p,(2p-1)}$.

Starting now with the remaining $n - 2p$ sample observations, repeat the steps above in the given order until the n–th EB B_n is constructed. Note that it is possible that not the full cycle of all the $2p$ steps above are used when B_n is constructed.

The last EB B_{n+1} is simply given by the complement in R^p of the set $B_1 \cup \cdots \cup B_n$. Hence $B_1 \cup \cdots \cup B_n \cup B_{n+1} = R^p$.

As before the union of the last k_0 EBs, given by $T(\mathcal{X}) = B_{n-k_0+2} \cup \cdots \cup B_{n+1}$, is a $100P\%$ content and γ confidence tolerance region for the population. If $k_0 \leq n - (2p-1)$ then $T(\mathcal{X})$ is a two–sided p–dimensional tolerance hyperrectangle (ie, box) as required. Similarly as before, the existence of such a k_0 is guaranteed only if the sample size n is sufficiently larger such that $1 - \mathcal{B}_{n-(2p-1),2p}(P) \geq \gamma$. For example, for $P = 95\%$, $\gamma = 0.95$ and $p = 3$, the sample size n must be at least 208 in order that $T(\mathcal{X})$ is a three dimensional box.

For the real problem considered in Section 3 below, it is required to construct a three-dimensional rectangular tolerance region of the form

$$T(\mathcal{X}) = \left\{ (x_1, x_2, x_3)' \; : \; x_1 \leq U_1(\mathcal{X}), \text{ and } L_i(\mathcal{X}) < x_i \leq U_i(\mathcal{X}) \text{ for } i = 2, 3 \right\}. \tag{5}$$

For this, the EBs can be constructed in a similar way as above but with $p = 3$, and *Step $p + 1 = 4$* is deleted in order not to have a lower bound on the measurement X_1.

It is clear that the construction given in this section can also be extended to construct distribution–free $100P\%$ content and γ confidence hyperrectangular tolerance regions of one–sided or two–sided or mix-sided for a general $p(\geq 2)$.

It is worth to emphasize that the hyperrectangular tolerance regions constructed above using k_0 EBs have a confidence level of at least γ from the requirement (3), with the exact confidence level given by $1 - \mathcal{B}_{k_0, n-k_0+1}(P)$, under the only assumption that the underlying distribution of the p measurements is continuous. This is mathematically proven in References 30,31. On the other hand, the hyperrectangular tolerance regions of Young and Mathew [18] are constructed not using EBs. The two key ingredients are (1) the number of observations k_0' to be retained based on which the tolerance regions are constructed, and (2) the identification of the $n - k_0'$ observations to be deleted (or trimmed). It is noteworthy that k_0' (involving some ad hoc adjustment) is an asymptotic approximation to k_0 as $n \to \infty$ while ingredient (2) utilizes data depth (cf. Reference 33). But the true confidence levels of these tolerance regions are shown to be reasonably close to, and more likely on the conservative side of, the nominal level γ only by simulation studies under the very limited underlying joint distributions considered in Reference 18. It should be noted that neither liberal nor conservative confidence level (in comparison with the nominal level γ) is desirable. A too conservative confidence level indicates that the tolerance region could be bigger than necessary and so has reduced sensitivity in detecting out of range individuals or observations in applications. The simulation study in Reference 37 on the true confidence levels of the hyperrectangular and upper-semi tolerance regions of Young and Mathew [18] under various bivariate mixed normal and mixed t distributions shows that the true confidence levels could be considerably larger than γ sometimes. For example, when $P = 90\%$, $\gamma = 0.90$ and $n = 1000$, the true confidence level of the hyperrectangular tolerance region (using simplicial depth) of Young and Mathew [18] is estimated to be 0.944 from simulation (Reference 36, p. 26, tab. 4.3). But the true confidence level of the hyperrectangular tolerance region constructed in this paper is $1 - \mathcal{B}_{k_0, n-k_0+1}(P) = 1 - \mathcal{B}_{913, 1000-913+1}(0.9) = 0.908$, which is much closer to $\gamma = 0.90$.

We also point out that nonparametric hyperrectangular tolerance regions have also been proposed and studied in Reference 38 (Sects. 5.6 and 5.7), even though their true confidence levels could be well short of the nominal level γ especially when p is large and the sample size n is not in thousands based on the simulation results for a multivariate lognormal distribution.[38]

3 | REFERENCE REGIONS FOR KIDNEY FUNCTION IN NORMAL ADOLESCENTS

This problem is addressed in Young and Mathew [18] which provides detailed background information. In essence, accurate reference regions about kidney function are important for detecting the presence and cause of, and for treating, kidney disease. The reference population of healthy adolescents is defined to be "US adolescents between 12 and 17 years old, not pregnant, having a blood pressure <120/80 mmHg, without diabetes, not using prescription medications within the previous 30 days, and a Z–score for weight-to-height ratio ≤ 2, which is a healthy level according to the World Health Organization".[18] The three important measurements of kidney function are: $X_1 = $ UACR, $X_2 = $ UA and $X_3 = $ SC. A sample of 5255 observations is extracted from the survey NHANES 1999–2014. This sample is further broken down by gender, resulting in $n_m = 2529$ males and $n_f = 2726$ females.

The aim is to establish $P = 95\%$ content and $\gamma = 0.95$ confidence reference regions of the form in (5) separately for males and females using the available reference samples of sizes n_m and n_f.

Using the method given in Section 2.2, the required reference regions are easily computed and given in the column "Mix–sided region (new)" of Table 1. In particular, for $n_m = 2529$, we have $k_{0m} = 2421$ and so the reference region for males is the union of 2421 EBs. The exact confidence level is $1 - \mathcal{B}_{k_{0m}, n_m-k_{0m}+1}(P) = 0.9518$ irrespective the distribution of the three measurements. The only assumption required is that the underlying distribution of the three measurements is continuous, which is entirely plausible. It is also noteworthy that the marginal distributions of UACR and SC are clearly not normal. The corresponding reference region of Young and Mathew [18] is given in the column "Mix–sided region (YM)" of Table 1, and taken from tab. 3 of their paper.

For the reference region for females we have $n_f = 2726$ and $k_{0f} = 2609$ so the exact confidence level is $1 - \mathcal{B}_{k_{0f}, n_f-k_{0f}+1}(P) = 0.9532$

TABLE 1 Three-dimensional rectangular reference regions for UACR, UA and SC.

	Mix-sided region (YM)	Mix-sided region (new)	Two-sided region (new)	Upper one-sided region (new)
Male				
UACR	≤3.2 mg/g	≤4.417	0.012–5.417	≤ 2.732
UA	2.8–8.2 mg/dL	2.800–8.200	2.700–8.400	≤ 8.000
SC	0.3–1.2 mg/dL	0.300–1.200	0.300–1.200	≤ 1.120
Female				
UACR	≤4.3 mg/g	≤5.241	0.017–5.622	≤ 3.964
UA	2.4–6.5 mg/dL	2.400–6.700	2.400–6.800	≤ 6.400
SC	0.3–1.0 mg/dL	0.300–0.970	0.300–1.000	≤ 0.900

The two–sided and upper one–sided reference regions (boxes) are also computed and given in Table 1. From Table 1, it can be seen that, for this particular application, the reference limits for UA and SC are hardly different between the two methods. But the upper limits on UACR of the new method are considerably larger than those of Young and Mathew's [18] method. Due to the (mathematical) validity of the new method (while the validity of the tolerance region of Young and Mathew [18] is only supported by simulation results), the new tolerance regions are certainly reliable to use. As expected, the two–sided regions have larger upper limits and smaller lower limits, and the upper one–sided regions have the smallest upper limits.

4 | CONCLUSIONS

Reference regions of $p(\geq 2)$ dimensions are important in laboratory medicine when several measurements on a patient are available. Nonparametric reference regions are attractive due to their robustness with respect to the underlying distribution. Nonparametric hyperrectangular reference regions have the additional advantage of being simple to interpret in real problems. Hyperrectangular reference regions have been recently given by Young and Mathew.[18] However, the validity of these reference regions is supported only by simulation results for limited underlying joint distributions and limited configurations of p, P, γ, and n.

The nonparametric hyperrectangular reference regions provided in this paper are distribution–free in the sense that their validity is proved mathematically with the only assumption that the underlying distribution is continuous. These reference regions are constructed by using the beautiful results on EBs developed by Tukey.[30,31] R code is available for easy computation of the new reference regions, which is demonstrated with the real problem of constructing reference regions for kidney functions of adolescences.

A non–rectangular two dimensional tolerance region is constructed under the bivariate normality assumption in Liu et al [39] as a reference region for detecting growth hormone misuse by elite athletes. We note that distribution–free tolerance regions of that particular shape can also be constructed by using the ideas in this paper.

Finally, we note that nonparametric hyperrectangular prediction regions can also be constructed by using Tukey's [30,31] EBs. However, following the view in Reference 13 that prediction regions are not suitable as reference regions, the construction of prediction regions falls beyond the scope of this paper.

DATA AVAILABILITY STATEMENT

The data and R code for reproducing all the results in this paper are available at http://www.personal.soton.ac.uk/wl/RefHypeRecta/.

ORCID

Wei Liu https://orcid.org/0000-0002-4719-0345
Frank Bretz https://orcid.org/0000-0002-2008-8340
Mario Cortina–Borja https://orcid.org/0000-0003-0627-2624

REFERENCES

1. Horn PS, Pesce AJ. *Reference Intervals: A User's Guide*. Washington, DC: AACC Press; 2005.
2. Clinical and Laboratory Standard Institute. *Defining, Establishing, and Verifying Reference Intervals in Clinical Laboratory: Approved Guideline*. 3rd ed. Wayne, PA: PLSI; 2008.
3. Henny J, Hyltoft PP. Reference values: from philosophy to a tool for laboratory medicine. *Clin Chem Lab Med*. 2004;42:686-691. doi:10.1515/CCLM.2004.117
4. Whyte MB, Kelly P. The normal range: it is not normal and it is not a range. *Postgrad Med J*. 2018;94:613-616. doi:10.1136/postgradmedj-2018-135983
5. Geffré A, Friedrichs K, Harr K, et al. Reference values: a review. *Vet Clin Pathol*. 2009;38:288-298. doi:10.1111/j.1939-165X.2009.00179.x
6. Häggström M. Establishment and clinical use of reference ranges. *WikiJ Med*. 2014;1:1. doi:10.15347/wjm/2014.003
7. Wise J. Choosing wisely: how the UK intends to reduce harmful medical overuse. *Bmj*. 2017;356:j370. doi:10.1136/bmj.j370
8. Samuels MH, Kolobova I, Smeraglio A, et al. Effect of thyroid function variations within the laboratory reference range on health status, mood, and cognition in levothyroxine-treated subjects. *Thyroid*. 2016;26:1173-1184. doi:10.1089/thy.2016.0141
9. Albert A, Harris EK. *Multivariate Interpretation of Clinical Laboratory Data*. New York, NY: Marcel–Dekker; 1987.
10. Harris EK, Boyd JC. *Statistical Basis of Reference Values in Laboratory Medicine*. New York, NY: Marcel Dekker; 1995.
11. Katki HA, Engels EA, Rosenberg PS. Assessing uncertainty in reference intervals via tolerance intervals: application to a mixed model describing HIV infection. *Stat Med*. 2005;24:3185-3198. doi:10.1002/sim.2171
12. Dong X, Mathew T. Central tolerance regions and reference regions for multivariate normal population. *J Multivar Anal*. 2015;134:50-60. doi:10.1016/j.jmva.2014.10.009
13. Liu W, Bretz F, Cortina-Borja M. Reference range: which statistical intervals to use? *Stat Methods Med Res*. 2021;30:523-534. doi:10.1177/0962280220961793
14. Wellek S, Jennen-Steinmetz C. Reference ranges: why tolerance intervals should not be used. Comment on Liu, Bretz and Cortina–Borja, reference range: which statistical intervals to use? *Stat Methods Med Res*. 2022;31:2255-2256. doi:10.1177/09622802221114538
15. Liu W, Bretz F, Cortina-Borja M. Authors' response. (response to the comment by Wellek and Jennen–Steinmetz.). *Stat Methods Med Res*. 2022;31:2257. doi:10.1177/09622802221111376
16. Lucagbo M, Mathew T. Rectangular tolerance regions and multivariate normal reference regions in laboratory medicine. *Biom J*. 2023;65:2100180. doi:10.1002/bimj.202100180
17. Wilks SS. Determination of sample sizes for setting tolerance limits. *Ann Math Stat*. 1941;12:91-96. doi:10.1214/aoms/1177731788
18. Young DS, Mathew T. Nonparametric hyperrectangular tolerance and prediction regions for setting multivariate reference regions in laboratory medicine. *Stat Methods Med Res*. 2020;29:3569-3585. doi:10.1177/0962280220933910
19. Guttman I. *Statistical Tolerance Regions: Classical and Bayesian*. London: Griffin; 1970.
20. Hahn G, Meeker WQ. *Statistical Intervals: A Guide to Practitioners*. 2nd ed. New York, NY: Wiley; 1991.
21. Guttman I. Tolerance regions. In: Kotz S, ed. *Encyclopaedia of Statistical Sciences*. 2nd ed. New York, NY: Wiley; 2006:8644-8659.
22. Krishnamoorthy K, Mathew T. *Statistical Tolerance Regions: Theory, Applications, and Computation*. New York, NY: Wiley; 2009.
23. Meeker WQ, Hahn GJ, Escobar LA. *Statistical Intervals: a Guide for Practitioners and Researchers*. 2nd ed. New York: Wiley; 2017.
24. Young DS. Tolerance: an R package for estimating tolerance intervals. *J Stat Softw*. 2010;36:1-39. doi:10.18637/jss.v036.i05
25. Trost DC. Multivariate probability–based detection of drug–induced hepatic signals. *Toxicol Rev*. 2006;25:37-45. doi:10.2165/00139709-200625010-00003
26. Krishnamoorthy K, Mathew T. Comparison of approximation methods for computing tolerance factors for a multivariate normal population. *Dent Tech*. 1999;41:234-249. doi:10.1080/00401706.1999.10485672
27. Krishnamoorthy K, Mondal S. Improved tolerance factors for multivariate normal distributions. *Commun Stat Simul Comput*. 2006;35:461-478. doi:10.1080/03610910600591883
28. Mbodj M, Mathew T. Approximate ellipsoidal tolerance regions for multivariate normal populations. *Stat Probab Lett*. 2015;97:41-45. doi:10.1016/j.spl.2014.11.001
29. Liu W, Bretz F, Hayter AJ, Kiatsupaibul S. Computation of tolerance ellipses for bivariate and trivariate normal populations. *J Stat Comput Simul*. 2022;92:3630-3638. doi:10.1080/00949655.2022.2076091
30. Tukey JW. Nonparametric estimation II: statistical equivalence blocks and tolerance regions–the continuous case. *Ann Math Stat*. 1947;18:529-539. doi:10.1214/aoms/1177730343

31. Tukey JW. Non-parametric estimation III. Statistically equivalent blocks and multivariate tolerance regions-the discontinuous case. *Ann Math Stat.* 1948;19:30-39. doi:10.1214/aoms/1177730287

32. Li J, Liu RY. Multivariate spacings based on data depth: construction of nonparametric multivariate toleranceregions. *Ann Stat.* 2008;36:1299-1323. doi:10.1214/07-aos505

33. Wellek S. On easily interpretable multivariate reference regions of rectangular shape. *Biom J*. 2011;53:491-511. doi:10.1002/bimj.201000147

34. Liu W, Bretz F, Böhning D, et al. Comparison of normal distribution based and nonparametric decision limits on the GH–2000 score for detecting growth hormone misuse (doping) in sport. *Biom J*. 2020;63:187-200. doi:10.1002/bimj.202000019

35. Di Bucchianico A, Einmahl JHJ, Mushkudiani NA. Smallest nonparametric tolerance regions. *Ann Stat*. 2001;29:1320-1343. doi:10.1214/aos/1013203456

36. Amerise IL. A direct method for constructing distribution–free tolerance regions. *Qual Quant*. 2023;57:3941-3954. doi:10.1007/s11135-022-01549-9

37. Jiang F. *Nonparametric Hyper Rectangular and Upper Semi–Spacetolerance Regions under Bivariate Mixed–Normal and Bivariatemixed–t Distributions: A Simulation Study*. MSc dissertation. UK: University of Southampton; 2022.

38. Lucagbo MD. *Rectangular Statistical Regions with Applications in Laboratory Medicine and Calibration*. PhD Thesis. Baltimore County, USA: University of Maryland; 2021, https://www.proquest.com/dissertations-theses/rectangular-statistical-regions-with-applications/docview/2540710922/se-2

39. Liu W, Bretz F, Böhning D, et al. Combined statistical decision limits based on two GH-2000 scores for the detection of growth hormone misuse. *Stat Methods Med Res*. 2022;31:1439-1448. doi:10.1177/09622802221093730

How to cite this article: Liu W, Bretz F, Cortina–Borja M. Distribution–free hyperrectangular tolerance regions for setting multivariate reference regions in laboratory medicine. *Statistics in Medicine*. 2024;1-11. doi: 10.1002/sim.10019