Simultaneous Confidence Intervals with more Power to Determine Signs

Yoav Benjamini

Department of Statistics and Operations Research

The Sackler School of Mathematical Sciences

Tel-Aviv University, Israel

Vered Madar

Department of Biostatistics

University of North Carolina, Chapel Hill

Philip B. Stark

Department of Statistics

University of California, Berkeley

10 March 2011

Abstract

We develop new simultaneous confidence intervals for the components of a multivariate mean. The intervals determine the signs of the parameters more frequently than standard intervals do: the set of data values for which each interval includes parameter values with only one sign is larger. When one or more estimated means are small, the new intervals sacrifice some length to avoid crossing zero. But when all the estimated means are large, the new intervals coincide with standard simultaneous confidence intervals, so there is no sacrifice of precision. The improved ability to determine signs is remarkable. For example, if two means are to be estimated and the intervals are allowed to be at most 80% longer than standard intervals, when only one mean is small its sign is determined almost as well as by a one-sided test that ignores multiplicity and has a pre-specified direction. When both are small the sign is determined better than by two-sided tests that ignore multiplicity. The intervals are constructed by inverting level-α tests to form a $1-\alpha$ confidence set, then projecting that set onto the coordinate axes to get confidence intervals. The tests have hyperrectangular acceptance regions that minimize the maximum amount by which the acceptance region protrudes from the orthant that contains the hypothesized parameter value, subject to a constraint on the maximum side length of the hyperrectangle. R and SAS scripts are available online.

Key Words: Non-equivariant hypothesis test, hyperrectangular acceptance region

1 Introduction

Standard simultaneous confidence intervals for the components of a multivariate mean serve at least two purposes: They express the joint uncertainty in estimates of the components and they classify the sign of each component as positive, negative, or indeterminate. For the first purpose, shorter intervals are preferable and standard intervals perform well. For the second, length is less important than whether the intervals include values of only one sign. One-sided confidence intervals classify signs well, but have infinite length—and the direction of the interval (upper or lower) must be pre-specified. Here we construct confidence intervals that adaptively trade length for the ability to classify the sign as nonnegative or nonpositive more frequently than standard intervals do, without pre-specifying a direction, while maintaining simultaneous coverage probability. Where the data make it easy to draw conclusions about the signs of the parameters, the new intervals are identical to conventional intervals. But where conventional intervals cannot determine the sign of one or more parameters, the new intervals sometimes can, at the cost of some length.

The new intervals extend work by Benjamini, Hochberg and Stark [2], Benjamini and Stark [3] and Madar [11]. Benjamini et al. [2] construct a $1-\alpha$ two-sided univariate confidence interval with nearly the same power to determine the sign of the parameter as $1-\alpha$ one-sided confidence intervals, without pre-specifying whether to use an upper or a lower one-sided interval.

Benjamini and Stark [3] develop a simultaneous confidence procedure with more power than conventional intervals to determine the signs of the components of an n-dimensional location parameter. Their intervals result from inverting a family of hypothesis tests whose acceptance regions are hyperrectangles of a fixed size and shape, centered at the hypothesized parameter value, but the orientation of the hyperrectangle depends on the relative sizes of the components of the hypothesized parameter value.

Here, we introduce a family $\{A_\theta\}_{\theta \in \Re^n}$ of acceptance regions that leads to simultaneous confidence intervals more directly analogous to the individual confidence intervals of Benjamini et al. [2]: These acceptance regions, called *quasi-conventional* (QC), protrude as little as possible from the orthant that contains the hypothesized parameter value, subject to a constraint on the level of the test and on the side lengths of the hyperrectangle. The QC confidence intervals that result from inverting the tests are not centered at the unbiased estimate when one or more components of that estimate is small. Allowing asymmetry—which biases the tests—increases the power to determine the signs of the components of the mean.

QC acceptance regions are equivariant under permutations and reflections of the coordinates but not under translation. The same is true of the hyperrectangular acceptance regions considered by [3], but those hyperrectangles have fixed aspect ratios and are centered at the hypothesized parameter value; only the orientation of the hyperrectangle varies with the parameter. Those regions yield unbiased tests. Allowing bias, as we do here, increases the power to determine the signs of the components.

Section 2 reviews the duality between confidence intervals and tests. The QC family of acceptance regions is presented in Section 3. QC confidence intervals are presented in Section 4. Section 5 presents some bivariate illustrations and a trivariate example from the Women's Health Initiative study of Hormone Replacement Therapy. Section 6 discusses further properties and possible generalizations of QC intervals. Appendix A contains technical details and proofs, including an explicit characterization of the extreme points of the QC confidence set, which determine the endpoints of the confidence intervals.

2 Tests and Confidence Sets

We seek simultaneous confidence intervals for the components of $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$ from the n-dimensional datum $\mathbf{X} = (X_j)_{j=1}^n$, where $\{X_j - \mu_j\}_{j=1}^n$ are iid with cdf F, and F has a symmetric, continuous, unimodal density $f(x)$ that is strictly decreasing for $x \geq 0$ in the support of f. Each X_j might be an unbiased estimator of μ_j computed from more than one raw observation. Estimating $\boldsymbol{\mu}$ from independent Gaussian observations $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2 I)$ is an example. Section 6 discusses joint confidence intervals when the components of $\mathbf{X}$ are correlated and gives some simulation results for correlated Gaussian estimators of $\{\mu_j\}$.

We want the confidence intervals with simultaneous coverage probability $1 - \alpha > 1/2$; i.e., the chance that all n intervals cover their parameters should be at least $1 - \alpha$. We want the intervals to determine the signs of $\{\mu_j\}$; that is, for the confidence interval for μ_j to contain values of only one sign. And we want the intervals to be short.

Suppose that for each $\theta \in \Re^n$, A_θ is the acceptance region for a level–α test of the hypothesis that $\boldsymbol{\mu} = \theta$ using the datum $\mathbf{X} = (X_j)_{j=1}^n$. Then

$$S_A(\mathbf{X}) \equiv \{\theta \in \Re^n : \mathbf{X} \in A_\theta\} \tag{1}$$

is a $1-\alpha$ simultaneous confidence set for $\boldsymbol{\mu}$ [10, pp. 89–90]. Simultaneous confidence intervals for the components of $\boldsymbol{\mu}$ can be constructed by projecting $S_A(\mathbf{X})$ onto the coordinate axes: For $j = 1, \ldots, n$, define

$$\mathcal{I}_j(\mathbf{X}) \equiv \left[\inf\{\theta_j : \theta \in S_A(\mathbf{X})\},\ \sup\{\theta_j : \theta \in S_A(\mathbf{X})\}\right]. \tag{2}$$

Then

$$\Pr_{\boldsymbol{\mu}}\left\{\bigcap_{j=1}^{n} \{\mathcal{I}_j(\mathbf{X}) \ni \mu_j\}\right\} \geq 1 - \alpha. \tag{3}$$

Hence, the intervals $\{\mathcal{I}_j\}$ are simultaneous $1 - \alpha$ confidence intervals for $\{\mu_j\}$. Below, we tailor $\{A_\theta\}$ so that $\mathcal{I}_j$ determines the sign of μ_j more often than conventional simultaneous intervals do.

3 Acceptance Regions

The conventional choice of α-level acceptance regions is a set of hypercubes centered at the hypothesized parameter values:

$$B_\theta \equiv \mathop{\times}_{j=1}^{n} [\theta_j - c_\alpha, \theta_j + c_\alpha], \tag{4}$$

where $c_\alpha \equiv F_{(1+(1-\alpha)^{1/n})/2}$ is the pth quantile of F. This family of acceptance regions is equivariant under permutations of the coordinates, reflections around the coordinate axes, and translations. The corresponding conventional confidence intervals are

$$\mathcal{I}_j^B(X) \equiv [X_j - c_\alpha, X_j + c_\alpha]. \tag{5}$$

We shall see that inverting a family of tests that is not equivariant under translations produces simultaneous confidence intervals that determine the signs of the components of $\boldsymbol{\mu}$ more frequently than conventional intervals do.

Suppose θ_0 and θ_1 differ in the sign of their jth component. The confidence set $S_A(\mathbf{X})$ does not determine the sign of the jth component of $\boldsymbol{\mu}$ if $\mathbf{X} \in A_{\theta_0} \cap A_{\theta_1}$. Hence, if we wish to determine the signs of the components as frequently as possible, the acceptance region A_θ should be confined as nearly as possible to the orthant in which θ lies. We consider only hyperrectangular acceptance regions, which correspond to conventional confidence sets when the regions are hypercubes centered at the parameter. Sidak [14] discusses of the merits of hyperrectangular acceptance regions.

Let $\mathcal{A}(\theta)$ denote the set of all hyperrectangles $H = \times_{j=1}^{n} [\theta_j - \ell_j(\theta), \theta_j + u_j(\theta)]$ that satisfy the significance-level constraint

$$\Pr_\theta\{\mathbf{X} \notin H\} \leq \alpha \tag{6}$$

and a side-length constraint

$$\ell_j(\theta) + u_j(\theta) \leq C, \; j = 1, \ldots, n. \tag{7}$$

We will drop the argument θ when that does not introduce ambiguity. Limiting the maximum side length to C limits the length of the confidence intervals that result from inverting the family of tests to less than $2C$.

Let $\mathcal{Z}(\theta) \equiv \{j : \theta_j = 0\}$ and $\mathcal{N}(\theta) \equiv \{j : \theta_j \neq 0\}$. (The mnemonic is that $\mathcal{Z}$ stands for the zero components and $\mathcal{N}$ for the non-zero components.) We define the QC acceptance region A_θ for $\theta \geq \mathbf{0}$ as follows:

1. If there exist hyperrectangles $H \in \mathcal{A}(\theta)$ for which $\ell_j = u_j = c_\alpha$, $j \in \mathcal{Z}(\theta)$, and $\theta_j - \ell_j \geq 0$, $j \in \mathcal{N}(\theta)$, then A_θ is the one with the smallest maximum side length.

2. Otherwise, A_θ is the hyperrectangle $H \in \mathcal{A}(\theta)$ with $\ell_j = u_j = c_\alpha$, $j \in \mathcal{Z}(\theta)$, for which $\min_{j \in \mathcal{N}(\theta)}(\theta_j - \ell_j)$ is largest.

Thus, A_θ is the conventional hypercube centered at θ whenever the smallest nonzero component of θ is c_α or larger. When some component of θ is less than c_α, A_θ is a hyperrectangle that contains only positive values of components $j \in \mathcal{N}(\theta)$ and has maximum side length not exceeding C, if such a hyperrectangle can satisfy the significance-level constraint. When that is impossible, A_θ is the hyperrectangle with side length not exceeding C that protrudes as little as possible into other orthants for $j \in \mathcal{N}(\theta)$. The protrusion of the acceptance region into orthants other than the one θ belongs to can be reduced or eliminated by lengthening the sides of the acceptance region for large components of θ and by allowing A_θ to be centered at a point other than θ. This is the key to the new method.

When θ is not in the positive orthant, the QC acceptance region A_θ is defined by reflecting the negative components about their coordinate axes. So, for example,

$$\ell_j((\theta_1, \ldots, -\theta_j, \ldots, \theta_n)) = u_j((\theta_1, \ldots, \theta_j, \ldots, \theta_n)). \tag{8}$$

The QC acceptance regions are equivariant under reflections about the axes and permutations of the coordinates: If π is a permutation of $(1, \ldots, n)$, then

$$\ell_j((\theta_{\pi(i)})_{i=1}^n) = \ell_{\pi(j)}(\theta) \tag{9}$$

and

$$u_j((\theta_{\pi(i)})_{i=1}^n) = u_{\pi(j)}(\theta). \tag{10}$$

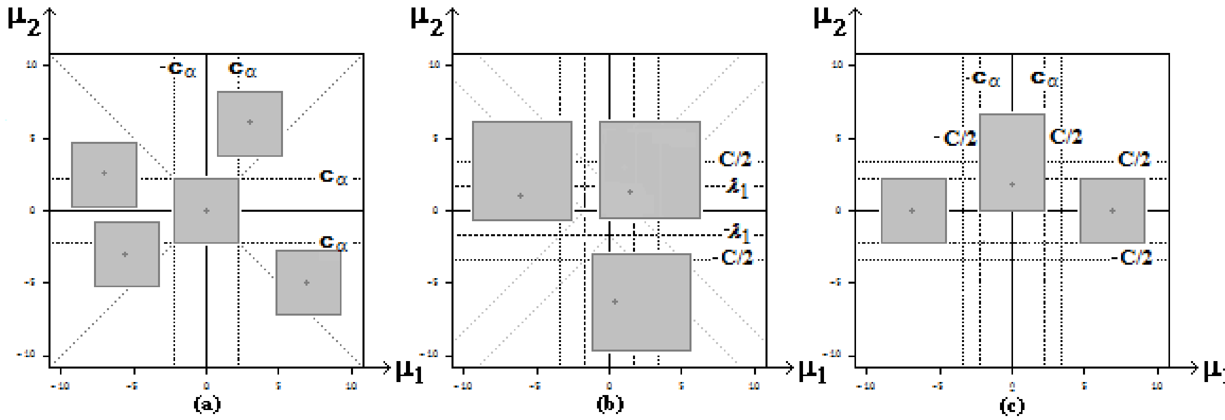

Figure 1: Bivariate QC acceptance regions

(a) Squares with side length c_α centered at θ when $\min(|\theta_1|, |\theta_2|) \geq c_\alpha$ or $\theta = \mathbf{0}$; (b) Squares with side length C that are centered at θ in one coordinate when $\min\{|\theta_1|, |\theta_2|\} < c_\alpha$ and $\Big||\theta_2| - |\theta_1|\Big| \geq C/2 - \lambda_1$ (top left and bottom); Squares with side length C that are not centered at θ in either coordinates when $\min\{|\theta_1|, |\theta_2|\} < c_\alpha$ and $\Big||\theta_2| - |\theta_1|\Big| < C/2 - \lambda(\theta) \leq C/2 - \lambda_1$ (top right). (c) Rectangles when one component of θ is zero.

Appendix A characterizes these acceptance regions precisely. Figure 1 shows exemplar QC bivariate acceptance regions, which can be squares centered at θ, squares centered at a point other than θ, or rectangles, depending on the magnitudes of the components of θ.

4 Confidence sets

The confidence set for $\boldsymbol{\mu}$ is $S(\mathbf{X}) = \{\theta \in \Re^n : \mathbf{X} \in A_\theta\}$. The simultaneous confidence intervals for $\{\mu_j\}_{j=1}^n$ are, for each j,

$$\begin{aligned} \mathcal{I}_j(X) &\equiv [\inf\{\theta_j : \theta \in S(\mathbf{X})\},\ \sup\{\theta_j : \theta \in S(\mathbf{X})\}] \\ &= [\inf\{\theta_j : \mathbf{X} \in A_\theta\},\ \sup\{\theta_j : \mathbf{X} \in A_\theta\}]\,. \end{aligned} \tag{11}$$

This amounts to projecting the convex hull of $S(\mathbf{X})$ onto the coordinate axes. The endpoints of the intervals for different components might be attained by different parameter vectors, so the intervals can be jointly conservative. The set $S(\mathbf{X})$ is hard to use directly or interpret. The n confidence intervals, one for each component of the parameter, are more useful in practice for simultaneous inference.

Since the acceptance regions are equivariant under reflection, the confidence intervals are too. We therefore may focus on the case $\mathbf{X} \geq 0$; other cases are constructed by reflecting the confidence set about the coordinate axes of those components of $\mathbf{X}$ that are negative. Treating the vector $\mathbf{X}$ as fixed, we denote the confidence interval for μ_j by (L_j, U_j), $j = 1, \ldots, n$. The confidence intervals depend on the datum $\mathbf{X}$ in a surprisingly simple way, described below.

Define

$$\lambda_k \equiv \min\{x : (2F(C/2) - 1)^{n-k} \times (F(x) + F(C - x) - 1)^k \geq 1 - \alpha\}, \tag{12}$$

$$\mathcal{C} \equiv \{j : X_j \leq C\}, \tag{13}$$

$$\mathcal{C}(j) \equiv \{i \neq j : C - X_i \geq X_j\}, \tag{14}$$

and

$$\kappa(j) \equiv \#\{i \neq j : C - X_i \geq X_j\} = \#\mathcal{C}(j). \tag{15}$$

These functions allow us to bracket the endpoints of the interval narrowly *a priori*. In the most complex case, the lower endpoint can be found exactly by solving an optimization problem with one variable:

$$h_k(x) \equiv x - \max_y \{y : [2F(C/2)-1]^{n-k-1} \times [F(C-x)-F(-x)]^k \times [F(C-y)-F(-y)] \geq 1-\alpha\}. \tag{16}$$

The upper confidence bound U_j for θ_j is never larger than $X_j + C/2$, since no acceptance region extends below θ_j by more than $C/2$.

Theorem 1 *Upper Confidence Bounds*

1. *If* $\#\mathcal{C} = 0$, $U_j = X_j + c_\alpha$ *for all* j.
2. *If* $\#\mathcal{C} = 1$, $U_j = X_j + c_\alpha$ *for* $j \in \mathcal{C}$ *and* $U_j = X_j + C/2$ *for* $j \notin \mathcal{C}$.
3. *If* $\#\mathcal{C} > 1$, $U_j = X_j + C/2$ *for all* j.

The lower confidence bound L_j for θ_j is never below $X_j - (C - \lambda_1)$, since no acceptance region contains values of a component that are larger than from the corresponding component of θ by more than $C - \lambda_1$.

Theorem 2 *Lower Confidence Bounds*

1. *If* $X_j > C$ *and* $\#\mathcal{C} = 0$, $L_j = X_j - c_\alpha$.
2. *If* $X_j > C$ *and* $\#\mathcal{C} > 0$, $L_j = X_j - C/2$.
3. *If* $\lambda_{\kappa(j)+1} < X_j \leq C$, $L_j = (X_j - (C - \lambda_1))_+$.
4. *If* $\lambda_{\kappa(j)} < X_j \leq \lambda_{\kappa(j)+1}$, $L_j = h_{\kappa(j)}(X_j)$.
5. *If* $0 < X_j \leq \lambda_{\kappa(j)}$ *then* $L_j = X_j - C/2$.

6. If $X_j = 0$ and $\#\mathcal{C} = 1$, $L_j = 0 - c_\alpha$.

Proofs of both theorems are in the appendix. R and SAS code for computing QC intervals is available at the URL `www.math.tau.ac.il/~ybenja`.

5 Examples and Illustrations

5.1 Bivariate Confidence Regions and Intervals

Figure 2 shows QC bivariate confidence sets and simultaneous confidence intervals for representative values of $\mathbf{X}$. The intervals are sometimes of the form $X_j \pm c_\alpha$, but not when any component of $\mathbf{X}$ is close to zero.

Figure 3 contrasts the values of $\mathbf{X}$ for which conventional simultaneous intervals determine the signs of the components of $\boldsymbol{\mu}$ with the set for which QC intervals determine those signs. The set of data values for which QC confidence intervals determine the sign of at least one component of $\boldsymbol{\mu}$ strictly includes the set for which conventional intervals do, so the QC intervals indeed determine the sign more frequently. The values themselves are surprising: almost too good to be true. For instance, suppose that the QC intervals are allowed to be 1.8 times as long as the conventional intervals in the worst case. Then if one component of $\mathbf{X}$ is large, the sign of both parameters is determined when the smaller component of $\mathbf{X}$ is larger in magnitude than $\lambda_1 = 1.65$. This is comparable to 1.645, the threshold to determine sign of a component using a one-sided regular interval—with a pre-determined direction. The signs of both components of the parameter are determined when both components of the datum are larger than $\lambda_2 = 1.95$. This is smaller than 1.965, the threshold to infer the signs of the components separately, not simultaneously. QC intervals have remarkable power to determine signs.

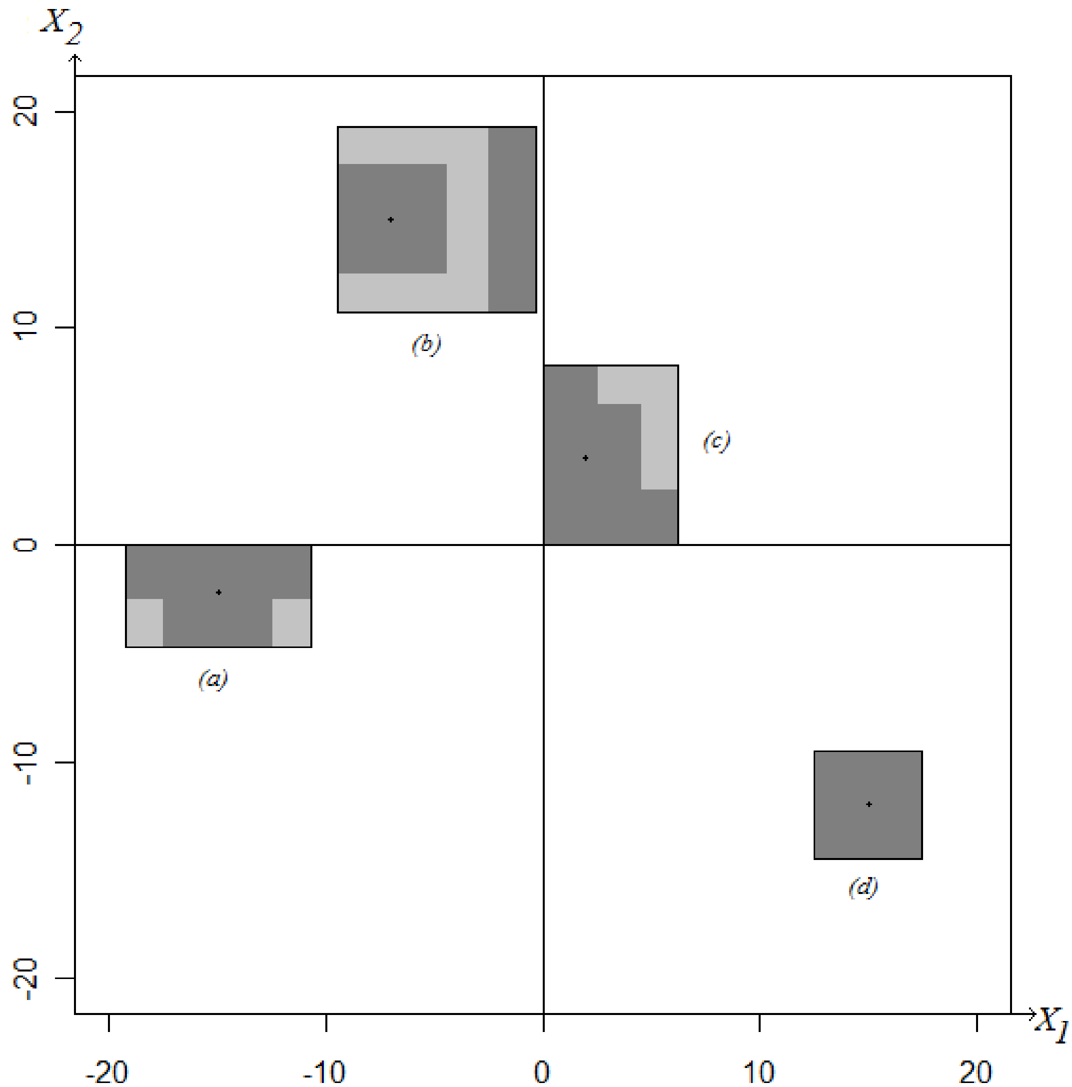

Figure 2: Bivariate QC confidence sets and confidence intervals

Let ${}_L\mathbf{X}_U$ denote the 95% confidence interval around the estimator X. (a) $\mathcal{I}_1 = {}_{-19.02}{-15}_{-10.98}$ and $\mathcal{I}_2 = {}_{-4.43}{-2.2}_{0.00}$ (b) $\mathcal{I}_1 = {}_{-9.23}{-7}_{-0.60}$ and $\mathcal{I}_2 = {}_{10.97}\mathbf{1}5_{19.03}$ (c) $\mathcal{I}_1 = {}_{0.00}\mathbf{1}.98_{6.01}$ and $\mathcal{I}_2 = {}_{0.00}\mathbf{4}_{8.03}$ (d) $\mathcal{I}_1 = {}_{12.76}\mathbf{1}5_{17.24}$ and $\mathcal{I}_2 = {}_{-14.23}{-12}_{-9.77}$. $C/2 = 1.8c_\alpha$ in these examples.

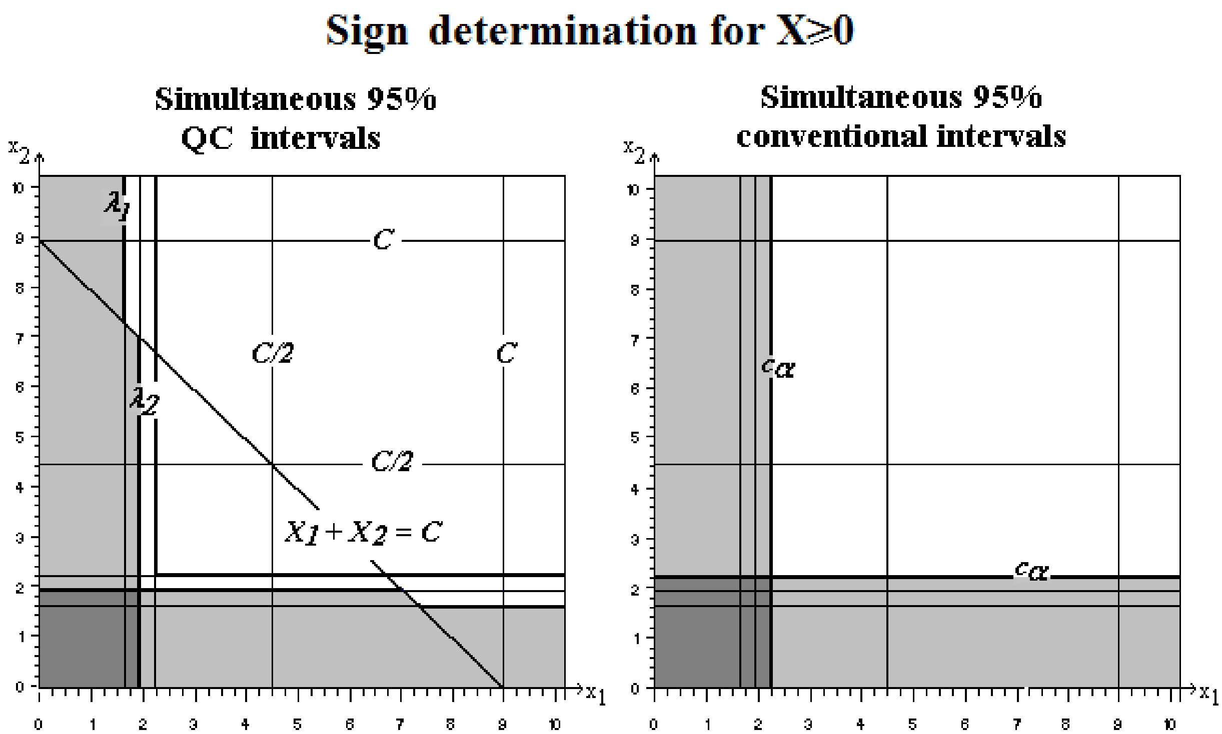

Figure 3: Sign determinations by QC and conventional simultaneous confidence intervals.

(Left) Data values for which 95% QC intervals determine the sign of one or both components of $\boldsymbol{\mu}$, for $C/2 = 1.8c_\alpha$. ($\lambda_1 = 1.65 < \lambda_2 = 1.95 < c_\alpha = 2.24$) (Right): Data values for which 95% conventional intervals determine the sign of one or both components of $\boldsymbol{\mu}$. The white regions are data values for which both components of $\boldsymbol{\mu}$ are determined to be nonnegative, the light gray regions are data values for which one component is determined to be nonnegative, and the dark gray regions are data values for which neither component is determined to be nonnegative.

5.2 Women's Health Initiative Trial of Hormone Replacement Therapy

The results of the Women's Health Initiative (WHI) randomized controlled clinical trial of Estrogen plus Progestin hormone therapy for postmenopausal women are reported in [13]. The primary endpoint for success of the therapy was a decrease in Coronary Heart Disease (CHD); the primary adverse endpoint was Invasive Breast Cancer (IBC); and there was a combined endpoint called "Global Health Index" (GHI), which combined risks and benefits. Larger values of the three parameters indicate worse health. The trial was stopped early because treatment unexpectedly increased CHD and increased IBC beyond a predetermined threshold. The GHI indicated that, overall, risk outweighed benefit.

The study reported simultaneous confidence intervals and intervals that were not adjusted for multiplicity. Conclusions from the two sets of intervals differed: The unadjusted intervals showed increases in GHI and the risk of IBC and CHD, as mentioned above. The simultaneous intervals were consistent with no increase in risk for any of the endpoints. The clinical recommendations of the study were based on the unadjusted confidence intervals.

Table 1 shows the estimated hazard ratio (HR) for the three endpoints, unadjusted confidence intervals, conventional simultaneous confidence intervals, and QC simultaneous confidence intervals for two choices of C. (All are based on the normal approximation to the log odds ratio.) Computing the QC intervals is described in appendix B. The QC 95% simultaneous confidence intervals support the clinical recommendations of the study while maintaining simultaneous confidence.

Endpoint	HR	Unadjusted	Conventional	QC ($C/2 = 1.2c_\alpha$)	QC ($C/2 = 1.8c_\alpha$)
IBC	1.26	$[1.00, 1.59]$	$[0.95, 1.67]$	$[0.90, 1.77]$	$[0.76, 2.1]$
CHD	1.29	$[1.02, 1.63]$	$[0.97, 1.72]$	$[1.00, 1.82]$	$[1.00, 2.16]$
GHI	1.15	$[1.03, 1.28]$	$[1.01, 1.31]$	$(1.00, 1.35]$	$(1.00, 1.45]$

Table 1: Estimated hazard rates, unadjusted (non-simultaneous) 95% confidence intervals, conventional simultaneous, and QC simultaneous 95% confidence intervals for the three endpoints in the Estrogen + Progestin Women's Health Initiative study of hormone-replacement therapy. All the intervals are based on the normal approximation to the log odds ratio. See appendix B.

6 Discussion

Quasi-conventional (QC) simultaneous confidence intervals determine the signs of the components of a multidimensional location parameter $\boldsymbol{\mu}$ more often than conventional simultaneous confidence intervals do. QC intervals are based on a family of hypothesis tests with non-equivariant hyperrectangular acceptance regions that exploit asymmetry (which entails bias) to reduce the the amount by which the acceptance region for θ protrudes from the orthant that contains θ. Inverting these tests and projecting the convex hull of the resulting confidence set onto the coordinate axes yields QC simultaneous confidence intervals.

When all components of the datum $\mathbf{X}$ are all large, QC intervals are identical to conventional simultaneous confidence intervals. But when any component of $\mathbf{X}$ is small, the QC intervals determine the signs of components of $\boldsymbol{\mu}$ more often, power purchased by an increase in length compared with conventional intervals. The increase in length is controlled by a parameter C.

The QC intervals include parameter values of only one sign for some values of $|X_i| < c_\alpha$.

When C is not much larger than $2c_\alpha$ (the length of conventional simultaneous intervals), QC intervals determine signs better than conventional two-sided intervals that ignore multiplicity. They do not exclude 0 until $|X_i| \geq c_\alpha$. Madar [12] defines QC acceptance regions differently for components of $\boldsymbol{\mu}$ that are equal to zero, resulting in intervals that are open at 0 for some data for which the QC intervals presented here are closed. Since it is implausible that the point null hypothesis $\mu = 0$ is *exactly* true, whether the intervals are open or closed at zero has little effect on their utility, so in the present paper we simplified the definition for clarity of exposition. The software for computing QC intervals available online uses the more complicated definition.

QC confidence intervals have simultaneous confidence level $1 - \alpha$ if the estimators of the components of $\boldsymbol{\mu}$ are independent. If the estimators are dependent but the acceptance regions have probability at least $1 - \alpha$ under that dependence, QC confidence intervals still attain their nominal level. Some QC hyperrectangles calibrated for independent, jointly Gaussian estimators can have probability less than $1-\alpha$ under dependence, but simulations show that the resulting intervals remain nearly conservative [12]. For example, for bivariate Gaussian estimators with $C/2 = 1.8c_\alpha$ and $\alpha = 0.05$, the simultaneous coverage probability of QC confidence intervals designed to have 95% confidence when the components of the data are independent have estimated coverage above 94.94%($\pm$0.005%) for all values of the correlation coefficient. Probability inequalities for hyperrectangular regions for dependent Gaussian and other elliptically contoured densities explain this empirical finding [14, 15, 7, 12].

Joint confidence sets can be tailored for inferences about scale rather than location, following the strategy outlined in [3]. Constructing confidence sets to attain other goals can be useful too. For instance [5, 6, 9] address confidence sets for bioequivalence, and [8] and [?] address inference conditional on the event that the estimator exceeds a threshold. We see the present work as a contribution in the larger context of optimizing confidence sets for specific scientific applications.

QC methods guarantee simultaneous coverage, but not all inference problems with multiple parameters require simultaneity: It is often enough to adjust for selection effects by controlling the False Coverage Statement Rate (FCR) [4]. Combining FCR with the univariate confidence intervals of [2], yields more powerful selection-adjusted sign determinations.

A Derivations and Proofs

This section characterizes QC acceptance regions in a way that helps find the extreme points of the confidence sets and shows how to project the confidence sets to find simultaneous confidence intervals.

A.1 Characterizing A_θ

Assume without loss of generality that $\theta \geq 0$. As noted above, acceptance regions for θ in other orthants are obtained by reflection.

The significance-level constraint, together with symmetry and unimodality of f, requires $C \geq 2c_\alpha$. Setting $C = 2c_\alpha$ reproduces the conventional confidence intervals, so the interesting case is $C > 2c_\alpha$. For technical reasons, we require the support of f to contain the interval $[-C, C]$; otherwise, we might as well decrease C, because an acceptance region satisfying the side-length constraint could have significance level $\alpha = 0$.

It follows from properties 1 and 2 and inequality 7 (see section 3) that for $\theta \geq 0$,

$$\ell_j \leq u_j, \tag{17}$$

$$\ell_j + u_j \leq C, \tag{18}$$

and hence

$$\ell_j \leq C/2. \tag{19}$$

Define

$$
\begin{aligned}
p(c) &\equiv F(c) - F(-c), \\
t &= t(\theta) \equiv \min_{j \in \mathcal{N}(\theta)} \theta_j, \\
z &= z(\theta) \equiv \#\mathcal{Z}(\theta).
\end{aligned}
$$

The acceptance region A_θ can be characterized using two functions. The first is $C(\theta)$, the smallest possible maximum side length of a hyperrectangular acceptance region that gives a test with the significance level α, has sides $[-c_\alpha, c_\alpha]$ for $j \in \mathcal{Z}(\theta)$, and contains only nonnegative values for the components $j \in \mathcal{N}(\theta)$:

$$
C(\theta) \equiv \inf \left\{ x : [p(c_\alpha)]^z \times \prod_{j \in \mathcal{N}(\theta)} [F(\min(\theta_j, x/2)) + F(x - (\min(\theta_j, x/2))) - 1] \geq 1 - \alpha \right\}. \tag{20}
$$

Note that $C(\theta) \geq 2c_\alpha$. (It can be infinite—we define the infimum over the empty set to be infinity.) If $C(\theta) \leq C$, there is a hyperrectangular acceptance region for a level α test of the hypothesis $\boldsymbol{\mu} = \theta$ that has side lengths no larger than C and is entirely confined to the positive orthant. If $C(\theta) > C$, A_θ crosses at least one axis.

The second function is $\lambda(\theta)$, the value of ℓ_j for the smallest nonzero θ_j; the acceptance region protrudes from the positive orthant by $(\lambda(\theta) - t(\theta))_+$:

$$
\begin{aligned}
\lambda(\theta) \equiv \inf \{ \quad x : \quad & [p(c_\alpha)]^z \times [p(C/2)]^{\#\{j \in \mathcal{N}(\theta) : \theta_j \geq C/2 + t(\theta) - x\}} \times \\
& \times \prod_{j \in \mathcal{N}(\theta) : \theta_j < C/2 + t(\theta) - x} [F(\theta_j + x - t(\theta)) + F(C - (\theta_j + x - t(\theta))) - 1] \\
& \geq 1 - \alpha \\
\} . &
\end{aligned} \tag{21}
$$

If $C(\theta) > C$, then A_θ contains $\mathbf{x} \in \Re^n$ with $x_j = t(\theta) - \lambda(\theta) < 0$ for some $j \in \mathcal{N}(\theta)$. If $C(\theta) \leq C$, then $\lambda(\theta) - t(\theta) \leq 0$.

Recall that $\ell_j = u_j = c_\alpha$ for $j \in \mathcal{Z}(\theta)$. The values of ℓ_j and u_j for $j \in \mathcal{N}(\theta)$ can be characterized using $C(\theta)$:

- If $C(\theta) = 2c_\alpha$, then $\ell_j = u_j = c_\alpha$, $j \in \mathcal{N}(\theta)$.
- If $2c_\alpha < C(\theta) \le C$, then $\ell_j(\theta) = \min(\theta_j, C(\theta)/2)$ and $u_j = C(\theta) - \ell_j(\theta)$, $j \in \mathcal{N}(\theta)$.
- If $C(\theta) > C$, then for $j \in \mathcal{N}(\theta)$, $u_j = C - \ell_j$ and

$$\ell_j = \begin{cases} C/2, & \theta_j \ge C/2 - (\lambda(\theta) - t(\theta)) \\ \theta_j + (\lambda(\theta) - t(\theta)), & \text{otherwise.} \end{cases} \tag{22}$$

In the first case, A_θ is the conventional hypercube acceptance region. In the second case, the sides of A_θ have equal length for $j \in \mathcal{N}(\theta)$, A_θ contains only positive values for the components $j \in \mathcal{N}(\theta)$, and A_θ is not centered at θ. In the third case, the sides of A_θ have equal length C for $j \in \mathcal{N}(\theta)$, A_θ contains negative values for some components $j \in \mathcal{N}(\theta)$, and A_θ is not centered at θ.

Any particular hyperrectangle H with side lengths no less than $2c_\alpha$ and no greater than C is the acceptance region for at most one θ unless H crosses two or more coordinate axes equally. On the other hand, if (i) H crosses two or more coordinate axes equally, (ii) the side lengths of H are equal to C for $j \in \mathcal{N}$, and (iii) H does not protrude too far from the positive orthant, then there can be a manifold of values of θ that have H as their acceptance region. For instance, in dimension $n = 2$, the hyperrectangle $H = [0, C] \times [0, C]$ is the acceptance region for $\theta = (\lambda_1, C/2)$, $\theta = (C/2, \lambda_1)$, $\theta = (\lambda_2, \lambda_2)$, and infinitely many other values of θ. (Note that H gives a biased test for all these parameters: The chance of rejecting the null is larger than it is for $\theta = (C/2, C/2)$, which has a different acceptance region, $[C/2 - c_\alpha, C/2 + c_\alpha] \times [C/2 - c_\alpha, C/2 + c_\alpha]$.) The manifold $\Theta(H)$ of values of θ that have a given acceptance region H plays an important role in inverting the tests to form confidence intervals.

A.2 Inverting and Projecting A_θ

Proof of theorem 1 *Upper Confidence Bounds.*

Recall that $\mathbf{X} \geq 0$ is fixed. Note that $\mathbf{X}$ is always in the acceptance region for $\theta = \mathbf{X}$. The proof follows the numbered assertions in the theorem.

1. Any parameter θ with one or more components close enough to zero to cause $C(\theta)$ to be larger than c_α is so close to zero that A_θ cannot include $\mathbf{X}$.

2. Observe that θ cannot be close enough to the axes in components $k \notin \mathcal{C}$ to cause u_j to be larger than c_α. Now consider $j \notin \mathcal{C}$. Starting with $\theta = \mathbf{X}$, decrease the component θ_k, $k \in \mathcal{C}$, until $C(\theta) = C$, which is obviously possible. Then the component θ_j can be increased to $X_j + C/2$; the resulting A_θ includes $\mathbf{X}$. Since $\ell_j \leq C/2$, this construction is extremal.

3. Starting with $\theta = \mathbf{X}$, decrease any component θ_k, $k \in \mathcal{C}$, $k \neq j$, until $C(\theta) = C$, which again is possible. Then increase θ_j to $X_j + C/2$; the resulting $A_\theta \ni \mathbf{X}$. Since $\ell_j \leq C/2$, this construction is extremal.

□

Proof of theorem 2 *Lower Confidence Bounds.*

The proof of (1) is immediate. To show (2), note that $\eta_j = X_j - C/2$ is feasible since there is another $i \in \mathcal{C}$, $i \neq j$, for which η_i can be reduced towards 0 until all other sides of the acceptance region have length C and are centered. If $\eta_j < X_j - C/2$, the acceptance region for $\boldsymbol{\eta}$ cannot cross 0 while having jth sidelength no larger than C. Therefore, what matters for the lower confidence bound is the upper extent of the acceptance region, ℓ_j, but $\ell_j \leq C/2$.

The other parts of Theorem 2 follow from a series of lemmas, starting with two utility lemmas.

Lemma A.1 *Suppose* $\mathbf{X} \in A_\theta$ *and* $C(\theta) \in (c_\alpha, C)$. *Then there exists* $\boldsymbol{\eta} \in \Re^n$ *such that* $\mathbf{X} \in A_{\boldsymbol{\eta}}$, $C(\boldsymbol{\eta}) = C$ *and* $|\eta_i| \leq |\theta_i|$, *and* $sgn(\eta_i) = sgn(\theta_i)$, $i = 1, \ldots, n$.

Proof. Suppose $C(\theta) \in (c_\alpha, C)$ with $\theta \geq 0$. We have $X_i \in [\theta_i - \ell_i, \theta_i + u_i]$, $i = 1, \ldots, n$, with $\ell_i = u_i = c_\alpha$, $i \in \mathcal{Z}$ and $\ell_i + u_i = C(\theta)$, $i \in \mathcal{N}$. For some k, $0 < |\theta_k| < c_\alpha \leq C$ (or else $C(\theta) = 2c_\alpha$). For that k, $|X_k| < C$ and $\mathrm{sgn}(X_k) = \mathrm{sgn}(\theta_k)$, or else $\mathbf{X} \notin A_\theta$. Define

$$\gamma_k(\theta) \equiv \arg\inf\{a\theta_k : a \in [0,1],\ \ \boldsymbol{\beta} \in \Re^n,\ \ \beta_i = \theta_i,\ i \neq k,\ \ \text{and } \beta_k = a\theta_k \text{ and } C(\boldsymbol{\beta}) \leq C\}. \tag{23}$$

Let $\eta_i = \theta_i$, $i \neq k$, and let $\eta_k = \gamma_k(\theta)$. Then

1. $C(\boldsymbol{\eta}) = C$
2. $|\eta_i| \leq |\theta_i|$ and $\mathrm{sgn}(\eta_i) = \mathrm{sgn}(\theta_i)$, $i = 1, \ldots, n$
3. For $i \neq k$, $\ell_i(\boldsymbol{\eta}) \geq \ell_i(\theta)$ and $u_i(\boldsymbol{\eta}) \geq u_i(\theta)$, so $[\eta_i - \ell_i(\boldsymbol{\eta}), \eta_i + u_i(\boldsymbol{\eta})] \supset [\theta_i - \ell_i(\theta), \theta_i + u_i(\theta)] \ni X_i$
4. If $\theta_k \geq 0$, $[\eta_k - \ell_k(\boldsymbol{\eta}), \eta_k + u_k(\boldsymbol{\eta})] = [0, C] \ni X_k$, and if $\theta_k < 0$, $[\eta_k - \ell_k(\boldsymbol{\eta}), \eta_k + u_k(\boldsymbol{\eta})] = [-C, 0] \ni X_k$.

□

Lemma A.2 *Suppose* $\mathbf{X} \in A_{\boldsymbol{\eta}}$ *where* $\eta_j < 0$ *and* $\eta_i > 0$, $\forall i \neq j$. *Suppose* $\eta_k > |\eta_j|$ *for some* $k \in \mathcal{C}(j)$. *Define* $\boldsymbol{\eta}'$ *such that* $\eta'_i = \eta_i$, $i \neq k$, *and* $\eta'_k = -\eta_j$. *Then:*

1. $\mathbf{X} \in A_{\boldsymbol{\eta}'}$

2. $\ell_j(\boldsymbol{\eta}') \geq \ell_j(\boldsymbol{\eta})$.

Proof. Since $A_{\boldsymbol{\eta}}$ crosses orthants, $\ell_i(\theta) + u_i(\theta) = C$, $\forall i \in \mathcal{N}$. Suppose $X_i \leq C - X_j$. If we set $\eta'_i = -\eta_j < \eta_i$, then $\ell_j(\boldsymbol{\eta}') \geq \ell_j(\boldsymbol{\eta})$ and $\ell_j(\boldsymbol{\eta}') = \ell_i(\boldsymbol{\eta}')$. Since $A_{\boldsymbol{\eta}} \ni \mathbf{X}$,

$$\eta_j + \ell_j(\boldsymbol{\eta}) \geq X_j, \tag{24}$$

and since $\ell_j(\boldsymbol{\eta}') \geq \ell_j(\boldsymbol{\eta})$,

$$\eta_j + \ell_j(\boldsymbol{\eta}') \geq X_j. \tag{25}$$

□

Theorem 2 follows from the previous two lemmas and a few more specific results:

Lemma A.3 *If* $\lambda_{\kappa(j)+1} < X_j$, $L_j \geq 0$.

Proof. The acceptance regions are equivariant under reflections around the axes, so it suffices to consider $\theta \geq 0$ and imagine varying the signs of some components of the datum.

Suppose that the maximum protrusion of $A_{\boldsymbol{\eta}}$ from the positive orthant is at least X_j, so that there is a datum with jth component $-X_j$ that is in the acceptance region for $\boldsymbol{\eta}$. That is, suppose that $(\lambda(\boldsymbol{\eta}) - t(\boldsymbol{\eta}))_+ \geq X_j$. The acceptance region $A_{\boldsymbol{\eta}}$ is always centered in every coordinate in which it does not cross an axis by at least X_j: The ith side is $[\eta_i - C/2, \eta_i + C/2]$ unless $C/2 - \eta_i \geq X_j$. (Otherwise, the maximum protrusion could be reduced by making $A_{\boldsymbol{\eta}}$ more nearly symmetric in the ith direction.) It follows that if $A_{\boldsymbol{\eta}}$ crosses any axis, it is symmetric in every direction that does not cross maximally.

By the definition of λ_k, for $A_{\boldsymbol{\eta}}$ to protrude from the positive orthant by $x > \lambda_{k-1}$, it must protrude by x from the positive orthant in at least k components. (If it protruded in fewer components than that, it would be symmetric in enough components to allow greater asymmetry in those components that cross axes, and hence would protrude less than x.)

The acceptance region for $\boldsymbol{\eta}$ cannot protrude across an axis by more than x and also include a value on the same side of that axis that is above $C - x$, because the side lengths do not exceed C.

Combining these three facts shows that if $X_j > \lambda_{k-1}$ and there are not at least $k - 1$ other components i for which $C - X_i \geq X_j$, there is no $\boldsymbol{\eta} \geq 0$ with an acceptance region that includes the value $-X_j$ in the jth coordinate.

$\square$

Lemma A.4 *If* $\lambda_{\kappa(j)} < X_j \le \lambda_{\kappa(j)+1}$*, then* $L_j = h_{\kappa(j)}(X_j)$*.*

Proof. Define $\eta_i = X_i + C/2$ for $i \notin \mathcal{C}(j)$, $i \neq j$. Define $\eta_i = \lambda_{\kappa(j)+1} - X_j$ otherwise. Since $\lambda_{\kappa(j)} < X_j \le \lambda_{\kappa(j)+1}$, $\boldsymbol{\eta} \ge 0$. Then $\boldsymbol{\eta}$ has exactly $\kappa(j)+1$ equal coordinates, so $A\boldsymbol{\eta}$ protrudes from the positive orthant at most by $\lambda_{\kappa(j)+1}$. Hence for $i = j$ and for $i \in \mathcal{C}(j)$, $\eta_i - \lambda_{\kappa(j)+1} = \lambda_{\kappa(j)+1} - X_j - \lambda_{\kappa(j)+1} = -X_j$.

For $i \in \mathcal{C}(j)$, $-X_j + C \ge X_i$, and for $i = j, X_j < C/2$ implies that $-X_j + C \ge X_j$. Construct $\boldsymbol{\eta}'$ so that $\eta'_j = -\eta_j$ and $\eta'_i = \eta_i$ for $i \neq j$. Then $\mathbf{X} \in A_{\boldsymbol{\eta}'}$. There is a manifold of parameter values sharing this acceptance region: $\Theta \equiv \{\theta : A_\theta = A_{\boldsymbol{\eta}'}\}$. The lower confidence bound for θ_j is no larger than

$$- \max_{\theta \in \Theta} \{\theta_j : \theta_i \le X_j - \lambda_{\kappa(j)+1} \text{ for } i \in \mathcal{C}(j)\}, \tag{26}$$

which is the maximization problem solved by $h_{\kappa(j)}(X_j)$ if θ_i is set to 0 for all $i \in \mathcal{C}(j)$.

$\square$

Lemma A.5 *If* $X_j \le \lambda_{\kappa(j)}$*,* $L_j = X_j - C/2$*.*

Proof. Define $\eta_j = C/2 - X_j$, and for $i \neq j$ define $\eta_i = \lambda_{\kappa(j)} - X_j$ if $i \in \mathcal{C}(j)$ and $\eta_i = X_i + C/2$ for $i \notin \mathcal{C}(j)$. Since $t(\eta) = \lambda_{\kappa(j)} - X_j$, and $\lambda(\eta) = \lambda_{\kappa(j)}$,

$$\eta_j = C/2 - X_j = C/2 - (\lambda_{\kappa(j)} - (\lambda_{\kappa(j)} - X_j)) = C/2 - (\lambda(\eta) - t(\eta)), \tag{27}$$

The condition in equation (22) (case 3) in the definition of the acceptance region is satisfied, and hence $\ell_j = C/2$. Therefore, $\eta_j - C/2 = (C/2 - X_j) - C/2 = -X_j$ is the lower edge of the acceptance region in direction j: The region includes $-X_j$. By reflection, $\mathbf{X}$ is in the acceptance region for $\boldsymbol{\eta}'$, where $\eta'_j = X_j - C/2$, and $\eta'_i = \eta_i$ for $i \neq j$.

$\square$

Lemma A.6 *If $X_j = 0$ and $\#\mathcal{C} = 1$, then $L_j = -c_\alpha$.*

Proof. Consider $\eta_i = X_i - c_\alpha$ for all i, so that $\eta_j = -c_\alpha$. This acceptance region for this parameter value contains $\mathbf{X}$, as shown above. For an acceptance region for a parameter with jth component less than $-c_\alpha$ to include $\mathbf{X}$ would require that for some $i \neq j$, $u_i + \ell_i > 2c_\alpha$. But this is impossible because $|X_i| \leq C$, $\forall i \neq j$.

□

These completes the proof of the six cases of theorem 2.

B Calculating New Confidence Intervals for WHI

We rely on the fact that the hazard ratio estimates, transformed to log-odds ratios, are approximately Gaussian distributed. We infer standard errors from the widths of the unadjusted 95% confidence intervals reported in the study.

The transformed, studentized datum is $\mathbf{X} = (1.947, 2.134, 2.558)$. For $\alpha = 0.05$ and $C/2 = 1.2c_\alpha$ we compute: $\lambda_1 = 1.728, \lambda_2 = 1.992, \lambda_3 = 2.125$, and $C/2 = 2.865$ (taking $C/2 = 1.8c_\alpha$ yields $\lambda_1 = 1.645, \lambda_2 = 1.955, \lambda_3 = 2.121$, and $C/2 = 4.298$).

To apply the results from section A.2, first note that $X_2 + X_3 \leq C$, so $\kappa(j) = 3$, $\forall j$. From $X_1 < \lambda_2$ it follows that the confidence interval for IBC is $\mathcal{I}_1(\mathbf{X}) = [X_1 - C/2, X_1 + C/2]$.

Next, $\mathcal{I}_2(\mathbf{X}) = [0, X_2 + C/2]$, since $c_\alpha > X_2 > \lambda_3$, and $\mathcal{I}_3(\mathbf{X}) = (0, X_3 + C/2]$ because $C > X_3 > c_\alpha$.

Transforming back into confidence intervals for the hazard ratio on the original scale produces the simultaneous 95% intervals in Table 1.

References

[1] I.M. Anonymous. Blinded. *JASA*, submitted, 2010.

[2] Y. Benjamini, Y. Hochberg, and P.B. Stark. Confidence intervals with more power to determine the sign: Two ends constrain the means. *J. Am. Stat. Assoc.*, 93:309–317, 1998.

[3] Y. Benjamini and P.B. Stark. Non-equivariant simultaneous confidence intervals less likely to contain zero. *J. Am. Stat. Assoc.*, 91:329–337, 1996.

[4] Y. Benjamini and Y. Yekutieli. False discovery rate controlling confidence intervals for selected parameters. *J. Am. Stat. Assoc.*, 100(469):71–80, 2005.

[5] R.L. Berger and J.C. Hsu. Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statist. Sci.*, 11:283–319, 1996.

[6] L.D Brown, G. Casella, and J.T.G. Hwang. Optimal confidence sets, bioequivalence, and the limacon of Pascal. *J. Am. Stat. Assoc.*, 90:880–889, 1995.

[7] S. Das Gupta, M.L. Eaton, I. Olkin, M. Perlman, L.J. Savage, and M. Sobel. Inequalities on the probability context of convex regions for elliptically contoured distributions. In *Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability*, volume 2, pages 241–265. Univ. California Press, 1972.

[8] Zhong H. and R.L. Prentice. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics*, 9:621–634, 2008.

[9] J.C. Hsu, J.T.G. Hwang, H. Liu, and S.J. Ruberg. Confidence intervals associated with tests for bioequivalence. *Biometrika*, 81:103–114, 1994.

[10] E.L. Lehmann. *Testing Statistical Hypotheses.* John Wiley and Sons, New York, 2nd edition, 1986.

[11] V. Madar. Non equivariant confidence intervals less likely to contain zero. Master's thesis, Tel-Aviv Univ., 2001.

[12] V. Madar. *Simultaneous confidence intervals for sign determination.* PhD thesis, Tel-Aviv Univ., 2008.

[13] J.E. Rossouw, G.L. Anderson, R.L. Prentice, and A.Z LaCroix. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. *JAMA.*, 288:321–333, 2002.

[14] Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *jasa*, 62:626–633, 1967.

[15] Z. Šidák. On probabilities of rectangles in multivariate student distributions: their dependence on correlations confidence regions for the means of multivariate normal distributions. *Ann. Math. Statist.*, 42:169–175, 1971.