Biometrika (2011), **xx**, x, *pp*. 1–18

Simultaneous Confidence Intervals with more Power to Determine Signs

BY 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

SUMMARY

Many medical studies draw inferences about multiple endpoints, but ignore the statistical implications of multiplicity. Effects inferred to be positive when there is no adjustment for multiplicity can lose their statistical significance when multiplicity is taken into account—perhaps explaining why such adjustments are so often omitted. We develop new simultaneous confidence intervals that mitigate this problem. They determine the signs of parameters more frequently than standard simultaneous confidence intervals: The set of data values for which each interval includes parameter values with only one sign is larger. When one or more parameter estimates are small, the new intervals sacrifice some length to avoid crossing zero. But when all the parameter estimates 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 four Normal means are to be estimated at 95% confidence and the intervals are allowed to be about 45% longer standard simultaneous intervals, when only one estimate is small, the new procedure determines the sign of the corresponding parameter essentially as well as a one-sided test that ignores multiplicity and has a pre-specified direction. 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 implementations are available online.

Key Words: Multiplicity, non-equivariant hypothesis test, one-sided test

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 procedures 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.

Effects inferred to be positive when there is no adjustment for multiplicity can lose their statistical significance when multiplicity is taken into account. The Women's Health Initiative (WHI) randomized controlled clinical trial of Estrogen plus Progestin hormone therapy for post-menopausal women (Rossouw et al., 2002) illustrates this problem. The study reported simultaneous confidence intervals and individual confidence intervals that were not adjusted for multiplicity, for three primary endpoints. Implications differed: The unadjusted intervals showed harm for two of three endpoints. The simultaneous intervals showed increase only for a combined endpoint measuring global health. The clinical recommendations of the study were based on the unadjusted confidence intervals.

Ignoring the statistical implications of multiplicity in medical research (research that is not conducted for regulatory purposes) seems to be the rule rather than the exception. In an editorial discussing WHI, Fletcher & Colditz (2001) write: "The authors present both nominal and rarely used adjusted CIs to take into account multiple testing, thus widening the CIs. Whether such adjustment should be used has been questioned ..." Benjamini (2010) finds that failure to adjust for multiplicity is endemic in medical research: He estimates that less than 25% of the studies reported in the New England Journal of Medicine that involve multiple testing took any account of multiplicity.

Multiplicity should be taken into account, as is required in regulatory studies. But the tendency of standard techniques for dealing with multiplicity to weaken clinical conclusions can be mitigated. For instance, in some cases the goal of testing can be relaxed to control the false discovery rate rather than the familywise error rate (Benjamini & Yekutieli, 2005). Here we take a different approach that still controls the familywise error rate: We develop simultaneous 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 work surprisingly well: Suppose that four Normal means are to be estimated at confidence level 95% and the intervals are allowed to be 45% longer than conventional simultaneous intervals. When only one observation is small, the new intervals determine the sign of the corresponding mean essentially as well as a one-sided test that ignores multiplicity and has a pre-specified direction. When all four estimates are small, the new intervals still determine the sign almost as well as two-sided tests that ignore multiplicity. Since many medical studies—for ethical and cost reasons—operate near the boundary of statistical significance, the new intervals help exactly where it matters.

The new intervals extend work by Benjamini et al. (1998), Benjamini & Stark (1996), and Madar (2001). Benjamini et al. (1998) 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 & Stark (1996) develop a simultaneous confidence procedure with more power than conventional intervals to determine the signs of the components of an n-dimensional location parameter.

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. (1998): These acceptance regions, called quasi-conventional, 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 quasi-conventional 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.

Quasi-conventional 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 Benjamini & Stark (1996), 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 quasi-conventional family of acceptance regions is presented in Section 3. Quasi-conventional confidence intervals are presented in Section 4. Section 5 presents some bivariate illustrations; a comparison with one-sided, unadjusted, and conventional intervals in four dimensions; a trivariate example from the Women's Health Initiative study of Hormone Replacement Therapy; and a 10-dimensional example from a study on coffee and mortality. Section 6 discusses further properties and possible generalizations of quasi-conventional intervals. Appendix 7 contains technical details and proofs, including an explicit characterization of the extreme points of the quasi-conventional confidence set, which determine the endpoints of the confidence intervals.

2. Tests and Confidence Sets

We seek simultaneous confidence intervals for the components of $\mu = (\mu_j)_{j=1}^n$ 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 μ from independent Gaussian observations $\mathbf{X} \sim \mathscr{N}(\mu, \sigma^2 I)$ 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 $\mu = \theta$ 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 μ (Lehmann, 1986, pp. 89–90). Simultaneous confidence intervals for the components of μ can be constructed by projecting $S_A(\mathbf{X})$ onto the coordinate axes: For $j = 1, \ldots, n$, define

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

Then

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

Hence, the intervals $\{\mathscr{I}_j\}$ are simultaneous $1-\alpha$ confidence intervals for $\{\mu_j\}$. Below, we tailor the family $\{A_\theta\}$ so that $\mathscr{I}_j$ 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. The conventional acceptance region for the hypothesis $\mu = \theta$ is

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

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

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

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 μ 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 μ 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. Šidák (1967) discusses of the merits of hyperrectangular acceptance regions.

Let $\mathscr{A}(\theta)$ denote the set of all hyperrectangles $H = \mathop{\times}_{j=1}^{n} [\theta_j - \ell_j(\theta), \theta_j + u_j(\theta)]$ 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 at most $3C/2 - \lambda_1$, where $\lambda_1 > 0$, defined in equation [12], depends on the distribution F and the dimension n.

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

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

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

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

Appendix 7·1 characterizes these acceptance regions precisely. Figure 1 shows exemplar bivariate quasi-conventional 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 μ 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,

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

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 report, to interpret, and to use directly. The n confidence intervals, one for each component of the parameter, are more useful.

Since the acceptance regions are equivariant under reflection, the confidence intervals are too. We therefore 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 $\mathbf{X}$ in a surprisingly simple way, described below.

Define

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

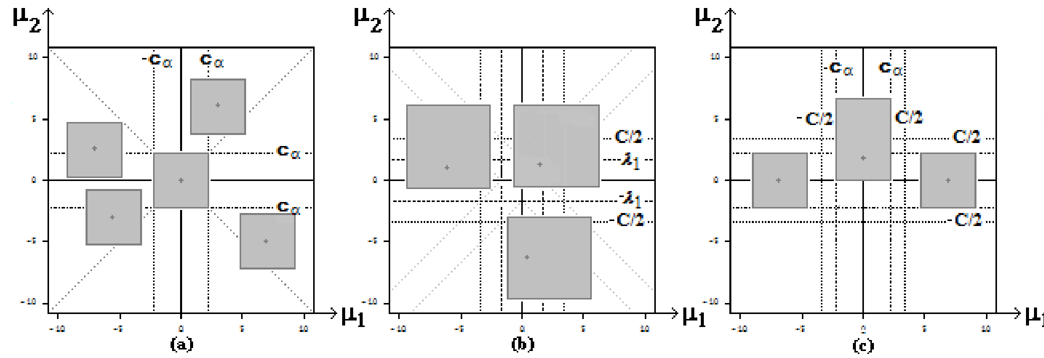

Fig. 1. Bivariate quasi-conventional acceptance regions

(a) Squares with side length c_α centered at θ when $\min(|\theta_1|,|\theta_2|) \geq c_\alpha$ or $\theta = 0$; (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.

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

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

and

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

These functions allow us to bracket the endpoints of the interval well *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 $C/2$. 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 $C - \lambda_1$. Hence, the length of any confidence interval is at most $3C/2 - \lambda_1$.

Theorem. Upper Confidence Bounds

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

Theorem. Lower Confidence Bounds

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

Proofs of both theorems are in the appendix 7. R and SAS code for computing quasi-conventional intervals is available at the URL `www.math.tau.ac.il/~ybenja`.

5. Examples and Illustrations

5·1. *Bivariate Normal Confidence Regions and Intervals*

Figure 2 shows quasi-conventional bivariate 95% confidence sets and simultaneous confidence intervals for a Normal mean, for representative values of **X**. The intervals are sometimes of the form $X_j \pm c_\alpha$, but not when any component of **X** is close to zero.

Figure 3 contrasts the values of **X** for which conventional simultaneous intervals determine the signs of the components of μ with the set for which quasi-conventional intervals determine those signs. The set of data values for which quasi-conventional confidence intervals determine the sign of at least one component of μ strictly includes the set for which conventional intervals do, so the quasi-conventional intervals indeed determine the sign more frequently. The values themselves seem almost too good to be true. For instance, suppose that the quasi-conventional acceptance regions have $C/2 = 1.8c_\alpha$. The conventional simultaneous intervals have length $2c_\alpha = 4.78$, while the quasi-conventional intervals have maximum length $3C/2 - \lambda_1 = 10.445$, no more than $21/3$ times as long as the conventional intervals in the worst case. Then if one component of **X** is large, the sign of both parameters is determined when the smaller component of **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. Quasi-conventional intervals have remarkable power to determine signs.

5·2. *Example: Four-dimensional sign determinations*

This section compares quasi-conventional sign determinations with those of one- and two-sided unadjusted confidence intervals and conventional simultaneous confidence intervals in 4-dimensional Normal examples, at 95% confidence level. We vary X_1 from 1 to 3 in increments of 0.05 and find the lower endpoints of 95% confidence intervals for μ_1.

To examine the effect of the number of small and large observations, we set some of the values of X_2, X_3, and X_4 to 10 and the rest to 0.5. Figure 4 plots the lower endpoints for $C/2 = 1.2c_\alpha$ (the upper row of subplots) and $C/2 = 1.8c_\alpha$ (the lower row). In the first panel in each row, all three of $\{X_2, X_3, X_4\}$ are equal to 0.5 and none is equal to 10. In the subsequent panels in each row, the number observations equal to 10 increases from 0 to 3.

The result is striking: Even when all three other components of **X** are small (0.5), the quasi-conventional intervals determine the sign of μ_1 for values of X_1 roughly midway between those for which unadjusted two-sided intervals and conventional simultaneous intervals do. When only two are small, the sign of μ_1 is determined for values of X_1 nearly as small as unadjusted two-sided intervals require. When the other three components of **X** are large, the quasi-conventional

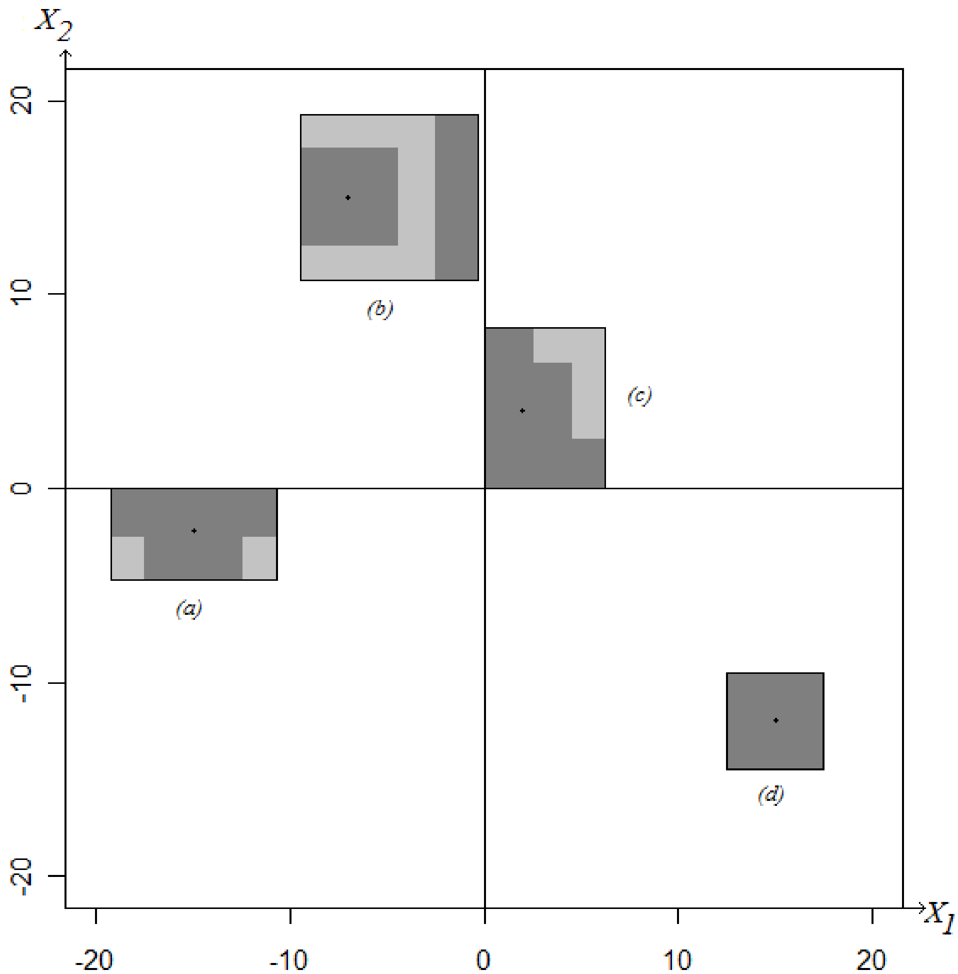

Fig. 2. Bivariate quasi-conventional confidence sets and confidence intervals for a bivariate Normal mean, $C/2 = 1.8c_\alpha$

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

intervals determine the sign of μ_1 essentially as well as *unadjusted one-sided intervals* with pre-specified direction. That is, the quasi-conventional interval then allows μ_1 to be inferred to be nonnegative for values of X_1 very close to $z_{1-\alpha}$.

Most of the benefit of quasi-conventional is evident even when $C/2 = 1.2c_\alpha$, for which the maximum length of the quasi-conventional confidence intervals is $3C/2 - \lambda_1 = 7.251$, about 45% longer than the the standard simultaneous confidence interval, which has length 4.989. The incremental improvement in sign determinations by allowing the acceptance regions to be 80%

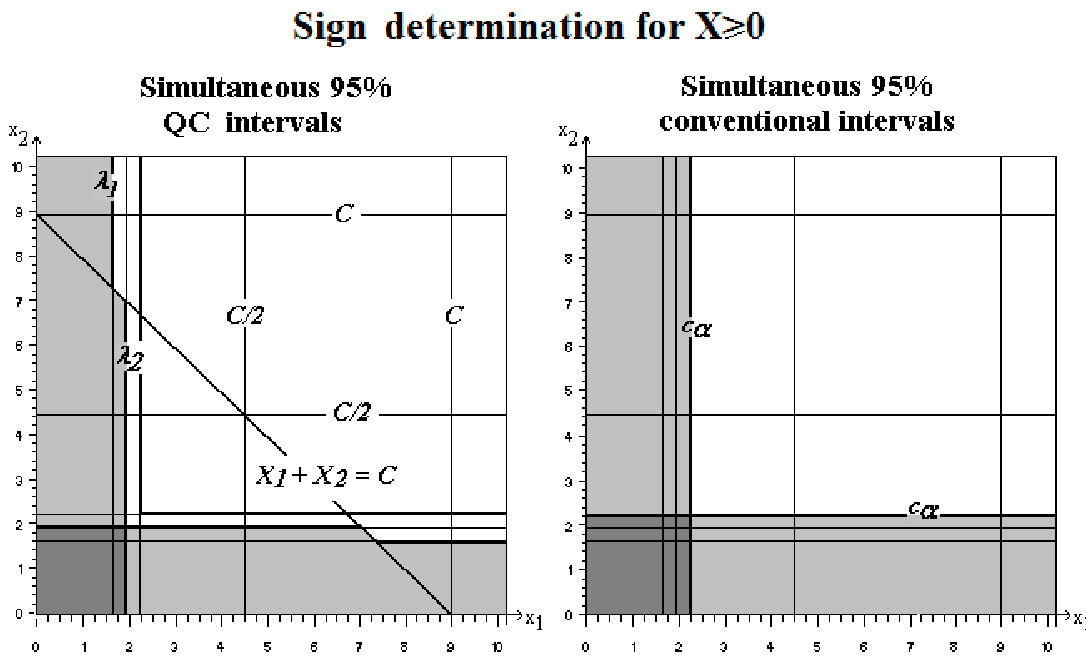

Fig. 3. Sign determinations by quasi-conventional and conventional simultaneous confidence intervals, $C/2 = 1.8c_\alpha$.

Left panel: Data values for which 95% quasi-conventional intervals determine the sign of one or both components of μ. Right panel: Data values for which 95% conventional intervals determine the sign of one or both components of μ. The white regions are data values for which both components of μ 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.

longer than standard hypercube acceptance regions, leading to a maximum confidence interval length about 137% longer than the standard simultaneous confidence interval, is small. Of course, this depends on the dimension of the problem and on that fact that the estimators are Gaussian in this example.

The quasi-conventional lower endpoint as a function of X_1 is step-like: When X_1 is too small to allow the sign of μ_1 to be determined, the lower endpoint is below even the lower endpoint of the conventional simultaneous interval. But as X_1 becomes large enough to allow μ_1 to be determined to be nonnegative, the lower endpoint abruptly rises to zero. It is equal to zero for a range of X_1, rather than crossing zero at a single point. (It eventually rises, but for larger values of X_1 than these plots show.) A cost of the improved ability to determine signs is that the lower endpoint of the confidence interval for μ_1 is equal to zero for values of X_1 for which the lower endpoint of a conventional simultaneous interval is strictly positive. And quasi-conventional confidence intervals are generally longer than conventional simultaneous intervals when one or more components of $\mathbf{X}$ are small. Quasi-conventional intervals are not a free lunch: They just let you start dinner early.

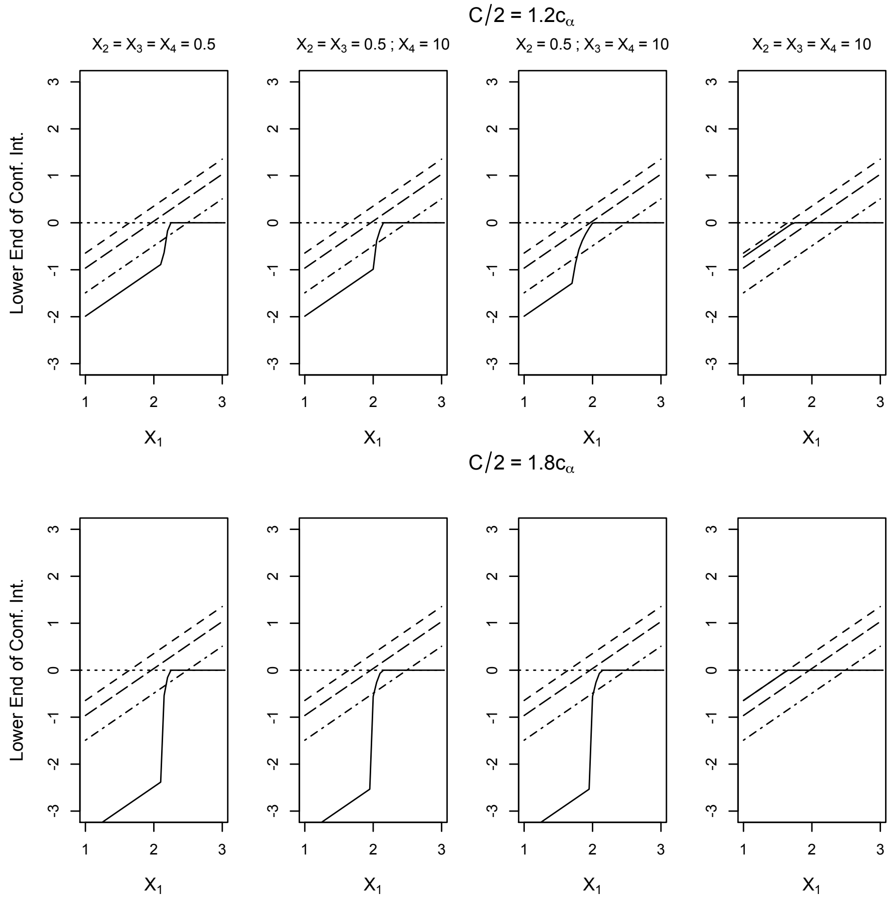

Fig. 4. Lower endpoints of 95% confidence intervals for the first component μ_1 of a 4-dimensional Normal mean μ as a function of X_1, the estimated value of μ_1.

Unadjusted one-sided lower confidence bounds (small dashes), unadjusted two-sided (large dashes), conventional two-sided simultaneous (alternating dashes and dots), and quasi-conventional simultaneous (solid lines). In the top row, $C/2 = 1.2c_\alpha$, which constrains the length of the quasi-conventional confidence interval to be at most 45% longer than standard simultaneous intervals. In the bottom row, $C/2 = 1.8c_\alpha$, which constrains the quasi-conventional intervals to be at most 137% longer than standard simultaneous intervals. In each row, the number of other estimates (X_2, X_3, X_4) that are "large" increases from 0 to 3. Estimates that are "large" are set to 10; the rest are set to 0.5.

5·3. *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 Rossouw et al. (2002). 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

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 quasi-conventional (QC) simultaneous 95% confidence intervals for the three endpoints in the Estrogen + Progestin Women's Health Initiative study of hormone-replacement therapy. The intervals are based on the normal approximation to the log odds ratio.*

treatment unexpectedly increased CHD and increased IBC beyond a predetermined threshold. The GHI indicated that, overall, risk outweighed benefit.

As mentioned in the introduction, the study reported simultaneous confidence intervals and intervals that were not adjusted for multiplicity, and the two sets of intervals supported different conclusions: The unadjusted intervals showed increases in GHI and the risk of IBC and CHD, while the simultaneous intervals were consistent with no increase in risk of IBC and CHD.

Table 1 shows the estimated hazard ratio (HR) for the three endpoints, unadjusted confidence intervals, conventional simultaneous confidence intervals, and quasi-conventional simultaneous confidence intervals for two choices of C. (All are based on the normal approximation to the log odds ratio.) Computing the quasi-conventional intervals is described in appendix 8. The quasi-conventional 95% simultaneous confidence intervals showed increase in CHD, the primary endpoint, as well as for GHI, and hence support the clinical recommendations of the study while maintaining simultaneous confidence. Note that in this case, CHD is an efficacy endpoint that became an adverse endpoint: The fact that quasi-conventional intervals have essentially the same power as a one-sided test—but without the need to pre-specify the direction—is crucial.

5·4. *Coffee and Mortality*

Lopez-Garcia et al. (2008) report an observational study of the association between coffee consumption and mortality from cardiovascular disease, cancer, and all causes. The study included 18 years follow-up in 41,736 men and 24 years in 41,736 men and 86,214 women. The raw results show a positive association of coffee intake and mortality from all causes. However, after adjustments for age, smoking status, alcohol consumption, and BMI, a Cox proportional hazard model shows weak negative association of relative risk of mortality from all causes with increasing coffee consumption—but only for women. The study reports confidence intervals for relative risk; those intervals do not appear to take multiplicity into account, even though there were inferences for three endpoints, two genders, and five consumption groups within each gender.

Table 2 shows the estimated relative risks, the reported (unadjusted) 95% confidence intervals, conventional simultaneous 95% confidence intervals, and quasi-conventional simultaneous 95% confidence intervals using $C/2 = 1.8c_\alpha$. The simultaneous intervals are for the single endpoint of mortality from all causes, but are simultaneous for the 10 gender-by-consumption groups. The unadjusted intervals are consistent with an increased risk of mortality for men for all consumption groups, but the relative risk of mortality is inferred to be less than one for women who drink 5 or more cups per week, four of the five consumption groups.

The conventional simultaneous intervals are consistent with an increase in mortality for women who drink 5–7 cups per week or more than 6 cups per day, two of the groups for which unadjusted confidence intervals show a decrease in mortality. The quasi-conventional intervals

Consumption	1c/mo–4c/week	5–7c/week	2–3c/day	4–5c/day	$\geq$ 6c/day
			Men		
Estimated RR	1.07	1.02	0.97	0.93	0.80
Unadjusted	[0.99, 1.16]	[0.95, 1.11]	[0.89, 1.05]	[0.81, 1.07]	[0.62, 1.04]
Conventional	[0.95, 1.21]	[0.90, 1.16]	[0.86, 1.09]	[0.76, 1.14]	[0.54, 1.17]
QC	[0.86, 1.32]	[0.82, 1.27]	[0.79, 1.19]	[0.64, 1.34]	[0.40, 1.58]
			Women		
Estimated RR	0.98	0.93	0.82	0.74	0.83
Unadjusted	[0.91, 1.05]	[0.87, 0.98]	[0.77, 0.87]	[0.68, 0.81]	[0.73, 0.95]
Conventional	[0.88, 1.09]	[0.86, 1.01]	[0.75, 0.90]	[0.65, 0.85]	[0.64, 1.01]
QC	[0.82, 1.18]	[0.81, 1.00]	[0.70, 1.00]	[0.58, 1.00]	[0.58, 1.00]

Table 2. *Estimated relative risk, unadjusted confidence intervals, conventional simultaneous confidence intervals, and quasi-conventional (QC) simultaneous confidence intervals for risk of mortality from all causes. Reference group:* <1 *c/month. 95% Confidence intervals. The quasi-conventional intervals use* $C/2 = 1.8c_\alpha$.

support (essentially) the same conclusions as the unadjusted intervals: Women in the four groups who drink more than 5 cups per week do not have an elevated risk of mortality relative to the control group, which consumed less than 1 cup per month. Notice that the quasi-conventional intervals are rather longer for men and for the group of women who consumed 1 cup per month to 4 cups per week, groups for which even the unadjusted intervals did not permit an inference about whether the relative risk exceeds 1.

6. Discussion

Quasi-conventional simultaneous confidence intervals determine the signs of the components of a multidimensional location parameter μ more often than conventional simultaneous confidence intervals do. Quasi-conventional 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 quasi-conventional simultaneous confidence intervals.

When all components of the datum $\mathbf{X}$ are all large, quasi-conventional intervals are identical to conventional simultaneous confidence intervals. But when any component of $\mathbf{X}$ is small, the quasi-conventional intervals determine the signs of components of μ more often, power purchased by an increase in length compared with conventional intervals. The increase in length is controlled by a parameter C: The maximum length is $3C/2 - \lambda_1$, where $c_\alpha \geq \lambda_1 > 0$.

The quasi-conventional 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), quasi-conventional intervals determine signs better than conventional two-sided intervals that ignore multiplicity. They do not exclude 0 until $|X_i| \geq c_\alpha$. Madar (2008) defines quasi-conventional acceptance regions differently for components of μ that are equal to zero, resulting in intervals that are open at 0 for some data for which the quasi-conventional 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 available at `www.math.tau.ac.il/~ybenja`. for computing quasi-conventional intervals uses the more complicated definition.

Quasi-conventional confidence intervals have simultaneous confidence level $1-\alpha$ if the estimators of the components of μ are independent. If the estimators are dependent but the acceptance regions have probability at least $1-\alpha$ under that dependence, quasi-conventional confidence intervals still attain their nominal level. Some quasi-conventional 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 (Madar, 2008). For example, for bivariate Gaussian estimators with $C/2 = 1.8c_\alpha$ and $\alpha = 0.05$, quasi-conventional confidence intervals designed to have 95% confidence when the components of the data are independent have estimated simultaneous coverage above 94.94% (s.e. 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 (Šidák, 1967, 1971; Das Gupta et al., 1972; Madar, 2008).

Joint confidence sets can be tailored for inferences about scale rather than location, following the strategy outlined in Benjamini & Stark (1996). Constructing confidence sets to attain other goals can be useful too. For instance Berger & Hsu (1996), Brown et al. (1995), **?**, and Hsu et al. (1994) address confidence sets for bioequivalence, and Zhong & Prentice (2008) and Benjamini & Weinstein (2010) 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.

Quasi-conventional 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) (Benjamini & Yekutieli, 2005). Combining FCR with the univariate confidence intervals of Benjamini et al. (1998), yields more powerful selection-adjusted sign determinations.

In the two examples in section 5·3 and 5·4, quasi-conventional simultaneous confidence intervals allow the same clinical conclusions as unadjusted confidence intervals, while conventional simultaneous confidence intervals do not. Of course, quasi-conventional intervals will not always make the same sign determinations as unadjusted intervals. The cases where they differ are where adjusting for simultaneity protects against selection bias. The cost of the improved ability to determine the signs of parameters compared with conventional simultaneous confidence intervals is that the intervals are wider: Estimating effect sign comes at the expense of estimating effect size.

7. Derivations and Proofs

This section characterizes quasi-conventional 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.

7·1. *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 $\theta \geq 0$,

$$\ell_j \leq u_j, \tag{A1}$$

$$\ell_j + u_j \leq C, \tag{A2}$$

and hence

$$\ell_j \leq C/2. \tag{A3}$$

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 $j \in \mathcal{N}(\theta)$:

$$C(\theta) \equiv \inf \left\{ x : [p(c_\alpha)]^z \times \prod_{j \in \mathcal{N}(\theta)} [F(\min(\theta_j, x/2)) + F(x - (\min(\theta_j, x/2))) - 1] \geq 1 - \alpha \right\}. \tag{A4}$$

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 $\mu = \theta$ 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 $(\lambda(\theta) - t(\theta))_+$:

$$\begin{aligned} \lambda(\theta) \equiv \inf\{ \; x : \; & [p(c_\alpha)]^z \times [p(C/2)]^{\#\{j \in \mathcal{N}(\theta) : \theta_j \geq C/2 + t(\theta) - x\}} \times \\ & \times \prod_{j \in \mathcal{N}(\theta) : \theta_j < C/2 + t(\theta) - x} [F(\theta_j + x - t(\theta)) + F(C - (\theta_j + x - t(\theta))) - 1] \\ & \geq 1 - \alpha \\ \}. & \end{aligned} \tag{A5}$$

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 $2c_\alpha < C(\theta) \leq C$, 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

$$\ell_j = \begin{cases} C/2, & \theta_j \geq C/2 - (\lambda(\theta) - t(\theta)) \\ \theta_j + (\lambda(\theta) - t(\theta)), & \text{otherwise.} \end{cases} \tag{A6}$$

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

$j \in \mathscr{N}$, 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.

7·2. *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 below 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 $k \notin \mathscr{C}$ to cause u_j to be larger than c_α. Now consider $j \notin \mathscr{C}$. Starting with $\theta = \mathbf{X}$, decrease the component θ_k, $k \in \mathscr{C}$, 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, $k \in \mathscr{C}$, $k \neq j$, 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 $i \in \mathscr{C}$, $i \neq j$, 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 η 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. Suppose $\mathbf{X} \in A_\theta$ and $C(\theta) \in (c_\alpha, C)$. Then there exists $\eta \in \Re^n$ such that $\mathbf{X} \in A_\eta$, $C(\eta) = C$ and $|\eta_i| \leq |\theta_i|$, and $\text{sgn}(\eta_i) = \text{sgn}(\theta_i)$, $i = 1, \ldots, n$. □

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 $\ell_i = u_i = c_\alpha$, $i \in \mathscr{Z}$ and $\ell_i + u_i = C(\theta)$, $i \in \mathscr{N}$. For some k, $0 < |\theta_k| < c_\alpha \leq C$ (or else $C(\theta) = 2c_\alpha$). For that k, $|X_k| < C$ and $\text{sgn}(X_k) = \text{sgn}(\theta_k)$, or else $\mathbf{X} \notin A_\theta$. Define

$$\gamma_k(\theta) \equiv \arg\inf\{a\theta_k : a \in [0,1],\ \beta \in \Re^n,\ \beta_i = \theta_i,\ i \neq k,\ \text{ and } \beta_k = a\theta_k \text{ and } C(\beta) \leq C\}. \quad \text{(A7)}$$

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

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

□

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

1. $\mathbf{X} \in A_{\eta'}$
2. $\ell_j(\eta') \geq \ell_j(\eta)$. □

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

$$\eta_j + \ell_j(\eta) \geq X_j, \tag{A8}$$

and since $\ell_j(\eta') \geq \ell_j(\eta)$,

$$\eta_j + \ell_j(\eta') \geq X_j. \tag{A9}$$

□

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

Lemma. 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 $A\eta$ 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 η. That is, suppose that $(\lambda(\eta) - t(\eta))_+ \geq X_j$. The acceptance region $A\eta$ 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 $A\eta$ more nearly symmetric in the ith direction.) It follows that if $A\eta$ crosses any axis, it is symmetric in every direction that does not cross maximally.

By the definition of λ_k, for $A\eta$ 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 η 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 of the acceptance region cannot 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 $\eta \geq 0$ with an acceptance region that includes the value $-X_j$ in the jth coordinate.

□

Lemma. If $\lambda_{\kappa(j)} < X_j \leq \lambda_{\kappa(j)+1}$, then $L_j = h_{\kappa(j)}(X_j)$. □

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

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

$$-\max_{\theta \in \Theta}\{\theta_j : \theta_i \leq X_j - \lambda_{\kappa(j)+1} \text{ for } i \in \mathscr{C}(j)\}, \tag{A10}$$

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

□

Lemma. If $X_j \leq \lambda_{\kappa(j)}$, $L_j = X_j - C/2$. □

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

$$\eta_j = C/2 - X_j = C/2 - (\lambda_{\kappa(j)} - (\lambda_{\kappa(j)} - X_j)) = C/2 - (\lambda(\eta) - t(\eta)), \tag{A11}$$

The condition in equation (A6) (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 η', where $\eta'_j = X_j - C/2$, and $\eta'_i = \eta_i$ for $i \neq j$.
□

Lemma. If $X_j = 0$ and $\#\mathscr{C} = 1$, 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.

8. 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 7·2, 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 $\mathscr{I}_1(\mathbf{X}) = [X_1 - C/2, X_1 + C/2]$.

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

BENJAMINI, Y. (2010). Simultaneous and selective inference: Current successes and future challenges. *Biometrical Journal* **52**, 708721. DOI: 10.1002/bimj.200900299.

BENJAMINI, Y., HOCHBERG, Y. & STARK, P. (1998). Confidence intervals with more power to determine the sign: Two ends constrain the means. *J. Am. Stat. Assoc.* **93**, 309–317.

BENJAMINI, Y. & STARK, P. (1996). Non-equivariant simultaneous confidence intervals less likely to contain zero. *J. Am. Stat. Assoc.* **91**, 329–337.

BENJAMINI, Y. & WEINSTEIN, A. (2010). Selection adjusted confidence intervals with more power to determine the sign. *J. Am. Stat. Assoc.* **submitted**.

BENJAMINI, Y. & YEKUTIELI, Y. (2005). False discovery rate controlling confidence intervals for selected parameters. *J. Am. Stat. Assoc.* **100**, 71–80.

BERGER, R. & HSU, J. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statist. Sci.* **11**, 283–319.

BROWN, L., CASELLA, G. & HWANG, J. (1995). Optimal confidence sets, bioequivalence, and the limacon of Pascal. *J. Am. Stat. Assoc.* **90**, 880–889.

DAS GUPTA, S., EATON, M., OLKIN, I., PERLMAN, M., SAVAGE, L. & SOBEL, M. (1972). Inequalities on the probability context of convex regions for elliptically contoured distributions. In *Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2. Univ. California Press.

FLETCHER, S. & COLDITZ, G. (2001). Failure of estrogen plus progestin therapy for prevention. *J. Amer. Med. Assoc.* **288**, 366369.

HSU, J., HWANG, J., LIU, H. & RUBERG, S. (1994). Confidence intervals associated with tests for bioequivalence. *Biometrika* **81**, 103–114.

LEHMANN, E. (1986). *Testing Statistical Hypotheses*. New York: John Wiley and Sons, 2nd ed.

LOPEZ-GARCIA, E., VAN DAM, R., LI, T., RODRIGUEZ-ARTALEJO, F. & HU, F. (2008). The relationship of coffee consumption with mortality. *Ann. Internal Medicine* **148**, 904–914.

MADAR, V. (2001). *Non equivariant confidence intervals less likely to contain zero*. Master's thesis, Tel-Aviv Univ.

MADAR, V. (2008). *Simultaneous confidence intervals for sign determination*. Ph.D. thesis, Tel-Aviv Univ.

ROSSOUW, J., ANDERSON, G., PRENTICE, R. & LACROIX, A. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. *J. Amer. Med. Assoc.* **288**, 321–333.

ŠIDÁK, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc.* **62**, 626–633.

ŠIDÁK, Z. (1971). 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.

ZHONG, H. & PRENTICE, R. (2008). Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics* **9**, 621–634.