Simultaneous Confidence Intervals with more Power to Determine Signs

Philip B. Stark
Department of Statistics
University of California, Berkeley

Joint work with:
Yoav Benjamini, Tel-Aviv University
Vered Madar, University of North Carolina

12 May 2011
Fourth Erich L. Lehmann Symposium
Rice University

Outline

Background

- Simultaneous Confidence Intervals
- Uses of Confidence Intervals
- Tests

Informal Problem Statement

- Determining Signs
- QC Acceptance Regions

QC Confidence Intervals

- Upper Confidence Bounds
- Lower Confidence Bounds

Illustrations

Women's Health Initiative

Coffee and mortality

Multivariate Location Model

- Datum $\mathbf{X} = (X_j)_{j=1}^n$.
- $\{X_j - \mu_j\}_{j=1}^n$ iid with cdf F.
- F has a symmetric, continuous, unimodal density $f(x)$, strictly decreasing for $x \geq 0$ in the support of f.
- Want to learn about μ.

Multivariate Location Model

- Datum $\mathbf{X} = (X_j)_{j=1}^n$.
- $\{X_j - \mu_j\}_{j=1}^n$ iid with cdf F.
- F has a symmetric, continuous, unimodal density $f(x)$, strictly decreasing for $x \geq 0$ in the support of f.
- Want to learn about μ.

Multivariate Location Model

- Datum $\mathbf{X} = (X_j)_{j=1}^n$.
- $\{X_j - \mu_j\}_{j=1}^n$ iid with cdf F.
- F has a symmetric, continuous, unimodal density $f(x)$, strictly decreasing for $x \geq 0$ in the support of f.
- Want to learn about μ.

Multivariate Location Model

- Datum $\mathbf{X} = (X_j)_{j=1}^n$.
- $\{X_j - \mu_j\}_{j=1}^n$ iid with cdf F.
- F has a symmetric, continuous, unimodal density $f(x)$, strictly decreasing for $x \geq 0$ in the support of f.
- Want to learn about $\boldsymbol{\mu}$.

Simultaneous Confidence Intervals

Simultaneous confidence intervals for the components of $\boldsymbol{\mu}$ are random intervals $\{\mathcal{I}_j(\mathbf{X})\}_{j=1}^n$ such that

$$\mathbf{P}_\theta \left\{ \bigcap_{j=1}^n \{\mathcal{I}_j(\mathbf{X}) \ni \theta_j\} \right\} \geq 1-\alpha, \quad \forall \theta \in \Theta.$$

C.f. non-simultaneous intervals:

$$\mathbf{P}_\theta \{\mathcal{I}_j \ni \theta_j\} \geq 1-\alpha, \quad \forall \theta \in \Theta.$$

Generally, simultaneous intervals have to be longer than non-simultaneous intervals.

Simultaneous Confidence Intervals from a Confidence Set

If $S(\mathbf{X})$ is a $1-\alpha$ confidence set for $\boldsymbol{\mu}$, then

$$\mathcal{I}_j \equiv \left[\inf_{\theta \in S(\mathbf{X})} \theta_j, \sup_{\theta \in S(\mathbf{X})} \theta_j \right], \quad j = 1, \ldots, n$$

are simultaneous confidence intervals for the components of $\boldsymbol{\mu}$.

$\mathcal{I}_j$ is the projection of the convex hull of $S(\mathbf{X})$ onto the jth axis.

Compare projection of ellipsoids versus hypercubes.

What are Confidence Intervals Good For?

- Express uncertainty in estimates of parameters
- Also allow inferences about signs of parameters: positive, indeterminate, negative
- Short intervals desirable to minimize uncertainty, but not necessarily for sign determination
- C.f. 1-sided versus 2-sided hypothesis tests

What are Confidence Intervals Good For?

- Express uncertainty in estimates of parameters
- Also allow inferences about signs of parameters: positive, indeterminate, negative
- Short intervals desirable to minimize uncertainty, but not necessarily for sign determination
- C.f. 1-sided versus 2-sided hypothesis tests

What are Confidence Intervals Good For?

- Express uncertainty in estimates of parameters
- Also allow inferences about signs of parameters: positive, indeterminate, negative
- Short intervals desirable to minimize uncertainty, but not necessarily for sign determination
- C.f. 1-sided versus 2-sided hypothesis tests

What are Confidence Intervals Good For?

- Express uncertainty in estimates of parameters
- Also allow inferences about signs of parameters: positive, indeterminate, negative
- Short intervals desirable to minimize uncertainty, but not necessarily for sign determination
- C.f. 1-sided versus 2-sided hypothesis tests

Tests and Acceptance Regions

The set A_θ is the acceptance region for a level-α test of the hypothesis that $\boldsymbol{\mu} = \theta$ if

$$\mathbf{P}_\theta\{\mathbf{X} \in A_\theta\} \geq 1 - \alpha.$$

Reject the hypothesis $\boldsymbol{\mu} = \theta$ if $\mathbf{X} \notin A_\theta$.

The test is *unbiased* if

$$\mathbf{P}_\theta\{\mathbf{X} \in A_\theta\} \geq \mathbf{P}_{\boldsymbol{\nu}}\{\mathbf{X} \in A_\theta\}, \quad \forall \boldsymbol{\nu} \in R^n.$$

Suppose we have a family of tests $\{A_\theta\}$ and a group $\mathcal{G}$ on R^n. The family is *equivariant under* $\mathcal{G}$ if

$$A_{g(\theta)} = g(A_\theta), \quad \forall \theta \in R^n, \ \forall g \in \mathcal{G}.$$

Conventional Hyperrectangular Regions

$c_\alpha \equiv F_{(1+(1-\alpha)^{1/n})/2}$, where F_p is pth quantile of F.

For location model, conventional α-level acceptance regions are hypercubes centered at the hypothesized parameter values:

$$B_\theta \equiv \bigtimes_{j=1}^{n} [\theta_j - c_\alpha, \theta_j + c_\alpha].$$

Unbiased if the density of F is symmetric and unimodal.

Equivariant under permutations of the coordinates, reflections around the coordinate axes, translations.

Duality between Tests and Confidence Sets

Suppose $\theta \in R^n$, 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

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

is a $1 - \alpha$ simultaneous confidence set for $\boldsymbol{\mu}$

Inverting standard tests

Inverting

$$B_\theta \equiv \bigtimes_{j=1}^{n} [\theta_j - c_\alpha, \theta_j + c_\alpha]$$

gives

$$S(\mathbf{X}) = \bigtimes_{j=1}^{n} [X_j - c_\alpha, X_j + c_\alpha].$$

Simple because of symmetry: whether $\mathbf{X}$ is in B_θ depends only on $\max_j |X_j - \theta_j|$.

What are our goals?

- Want confidence intervals for all components of $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$.
- Want simultaneous $1 - \alpha$ coverage
- Want interval for each μ_j to contain values of only one sign
- Want intervals to be short.
- When it's easy to tell the signs of the components, don't want to lose any precision compared with standard intervals.

What are our goals?

- Want confidence intervals for all components of $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$.
- Want simultaneous $1 - \alpha$ coverage
- Want interval for each μ_j to contain values of only one sign
- Want intervals to be short.
- When it's easy to tell the signs of the components, don't want to lose any precision compared with standard intervals.

What are our goals?

- Want confidence intervals for all components of $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$.
- Want simultaneous $1 - \alpha$ coverage
- Want interval for each μ_j to contain values of only one sign
- Want intervals to be short.
- When it's easy to tell the signs of the components, don't want to lose any precision compared with standard intervals.

What are our goals?

- Want confidence intervals for all components of $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$.
- Want simultaneous $1 - \alpha$ coverage
- Want interval for each μ_j to contain values of only one sign
- Want intervals to be short.
- When it's easy to tell the signs of the components, don't want to lose any precision compared with standard intervals.

What are our goals?

- Want confidence intervals for all components of $\boldsymbol{\mu} = (\mu_j)_{j=1}^n$.
- Want simultaneous $1 - \alpha$ coverage
- Want interval for each μ_j to contain values of only one sign
- Want intervals to be short.
- When it's easy to tell the signs of the components, don't want to lose any precision compared with standard intervals.

Determining Signs

Suppose θ_0 and θ_1 differ in the sign of their jth component.

$S_A(\mathbf{X})$ won't determine the sign of the jth component of $\boldsymbol{\mu}$ if $\mathbf{X} \in A_{\theta_0} \cap A_{\theta_1}$.

To determine the signs of the components as frequently as possible, confine A_θ as nearly as possible to the orthant containing θ.

More general hyperrectangular acceptance regions

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

$$\mathbf{P}_\theta\{\mathbf{X} \notin H\} \leq \alpha$$

and a side-length constraint

$$\ell_j(\theta) + u_j(\theta) \leq C, \ j = 1, \ldots, n.$$

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 $2C$.

QC Acceptance Regions

$\mathcal{Z}(\theta) \equiv \{j : \theta_j = 0\}$.

$\mathcal{N}(\theta) \equiv \{j : \theta_j \neq 0\}$.

QC acceptance A_θ for $\theta \geq \mathbf{0}$:

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.

Reduce protrusion of A_θ into orthants other than the one θ belongs to by lengthening the sides for large components of θ and allowing A_θ to be centered at a point other than θ.

For θ not in the positive orthant, the QC acceptance region A_θ is defined by reflecting the negative components about their coordinate axes. E.g.,

$$\ell_j((\theta_1, \ldots, -\theta_j, \ldots, \theta_n)) = u_j((\theta_1, \ldots, \theta_j, \ldots, \theta_n)).$$

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

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

and

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

QC complications

- Not translation-equivariant.
- Biased except when coincides with standard hypercubes.
- Non-unique: can have $\theta \neq \eta$ with $A_\theta = A_\eta$.
- Much harder to invert than hypercubes: Whether $X \in A_\theta$ depends on more than just $\max_j |X_j - \theta_j|$.
- But they determine signs better!

QC complications

- Not translation-equivariant.
- Biased except when coincides with standard hypercubes.
- Non-unique: can have $\theta \neq \eta$ with $A_\theta = A_\eta$.
- Much harder to invert than hypercubes: Whether $X \in A_\theta$ depends on more than just $\max_j |X_j - \theta_j|$.
- But they determine signs better!

QC complications

- Not translation-equivariant.
- Biased except when coincides with standard hypercubes.
- Non-unique: can have $\theta \neq \boldsymbol{\eta}$ with $A_\theta = A_{\boldsymbol{\eta}}$.
- Much harder to invert than hypercubes: Whether $X \in A_\theta$ depends on more than just $\max_j |X_j - \theta_j|$.
- But they determine signs better!

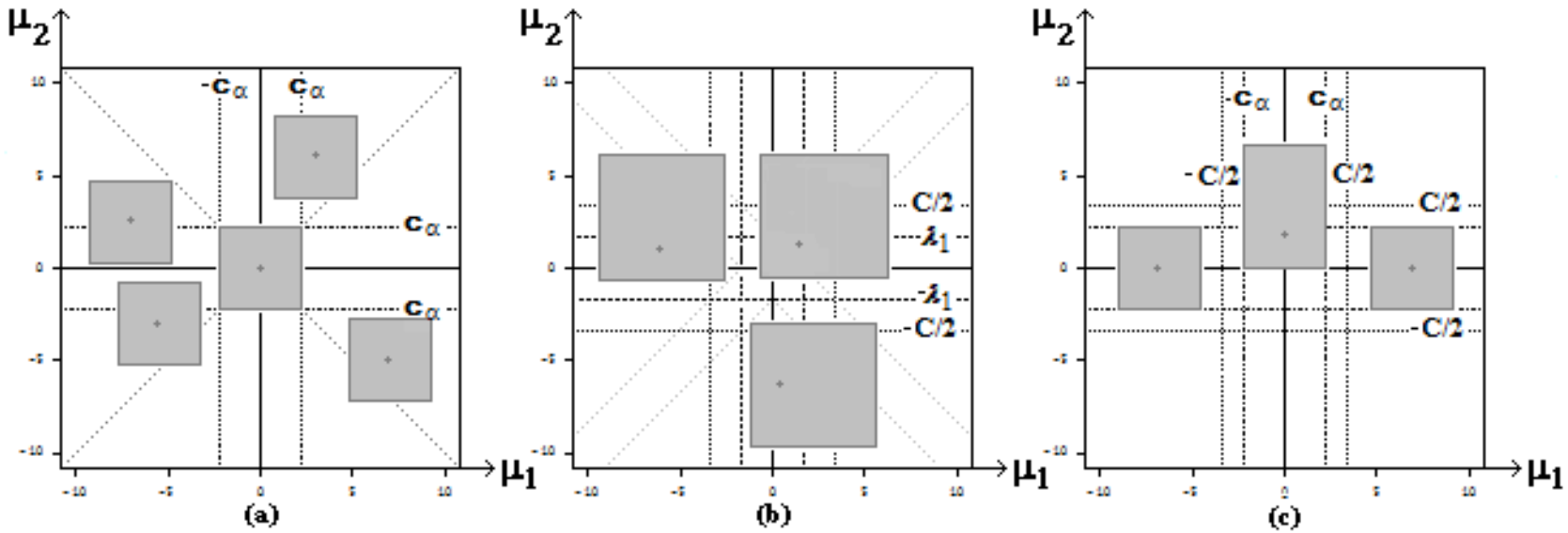
QC complications

- Not translation-equivariant.
- Biased except when coincides with standard hypercubes.
- Non-unique: can have $\theta \neq \boldsymbol{\eta}$ with $A_\theta = A_{\boldsymbol{\eta}}$.
- Much harder to invert than hypercubes: Whether $X \in A_\theta$ depends on more than just $\max_j |X_j - \theta_j|$.
- But they determine signs better!

QC complications

- Not translation-equivariant.
- Biased except when coincides with standard hypercubes.
- Non-unique: can have $\theta \neq \boldsymbol{\eta}$ with $A_\theta = A_{\boldsymbol{\eta}}$.
- Much harder to invert than hypercubes: Whether $X \in A_\theta$ depends on more than just $\max_j |X_j - \theta_j|$.
- But they determine signs better!

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, centered at θ in one coordinate when $\min\{|\theta_1|, |\theta_2|\} < c_\alpha$ and $\left||\theta_2| - |\theta_1|\right| \geq C/2 - \lambda_1$ (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 $\left||\theta_2| - |\theta_1|\right| < C/2 - \lambda(\theta) \leq C/2 - \lambda_1$ (top right). (c) Rectangles when one component of θ is zero.

QC Confidence set for μ

QC Acceptance regions are equivariant under reflection, so confidence intervals are too: assume wlog $\mathbf{X} \geq 0$.

Construct other cases by reflecting about the coordinate axes of those components of $\mathbf{X}$ that are negative.

Define

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

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

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

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

and

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

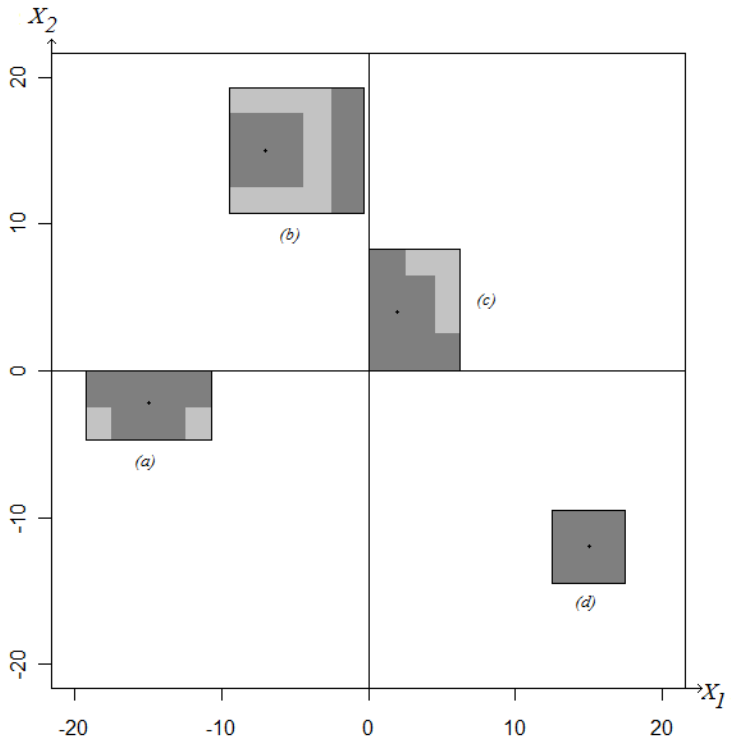
Upper Confidence Bounds

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.

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

Bivariate QC confidence sets and confidence intervals



${}_L\mathbf{X}_U$ is 95% confidence interval around the estimator X. $C/2 = 1.8c_\alpha$. (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 $\mathcal{I}_2 = {}_{10.97}\mathbf{15}_{19.03}$ (c) $\mathcal{I}_1 = {}_{0.00}\mathbf{1.98}_{6.01}$ and $\mathcal{I}_2 = {}_{0.00}\mathbf{4}_{8.03}$ (d) $\mathcal{I}_1 = {}_{12.76}\mathbf{15}_{17.24}$ and $\mathcal{I}_2 = {}_{-14.23}-12_{-9.77}$.

Almost too good to be true

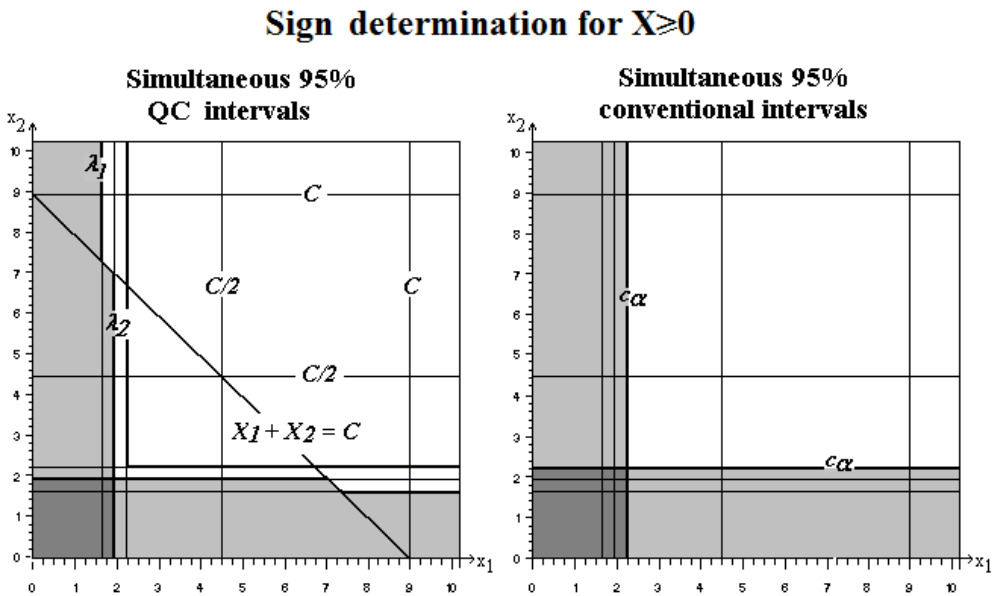
The set of data values for which QC confidence intervals determine the sign of at least one component of $\boldsymbol{\mu}$ strictly includes the set for which conventional intervals do.

For $C/2 = 1.8c_\alpha$, if one component of $\mathbf{X}$ is large, signs of both parameters determined when the smaller component of $\mathbf{X}$ is larger in magnitude than $\lambda_1 = 1.65$.

Comparable to 1.645, the threshold to determine sign using a one-sided regular interval—with a pre-determined direction.

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 separately.

Sign determinations: QC vs conventional simultaneous



(Left) Data for which 95% QC intervals determine 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 for which 95% conventional intervals determine sign of one or both components of $\boldsymbol{\mu}$. The white regions are data for which both components of $\boldsymbol{\mu}$ are determined to be nonnegative, the light gray regions are data for which one component is determined to be nonnegative, and the dark gray regions are data for which neither component is determined to be nonnegative.

Women's Health Initiative (WHI) randomized trial

Estrogen plus Progestin hormone therapy for postmenopausal women.

Included 161,808 generally healthy postmenopausal women.

Primary endpoint for success was a decrease in Coronary Heart Disease (CHD); the primary adverse endpoint was Invasive Breast Cancer (IBC); combined endpoint "Global Health Index" (GHI), which combined risks and benefits.

Larger values indicate worse health.

Trial stopped early (after 11y) because treatment unexpectedly increased CHD and increased IBC beyond a predetermined threshold. The GHI indicated that, overall, risk outweighed benefit.

WHI Confidence Intervals

Reported simultaneous and non-simultaneous confidence intervals.

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.

QC 95% intervals support the clinical recommendations

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]

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 WHI study of hormone-replacement therapy. Uses normal approximation to log odds ratio.

Coffee Consumption and Mortality

Lopez-Garcia et al., 2008. Ann. Internal Medicine.

Observational study of association between coffee consumption and mortality from CVD, cancer, and all causes.

18 years follow-up in 41,736 men; 24 years in 41 736 men and 86,214 women.

Raw results show positive association of coffee intake and mortality for all causes.

After adjustments for age, smoking status, alcohol consumption, BMI, Cox proportional hazard model shows weak negative association of RRs, no multiplicity adjustments.

Coffee Consumption and Mortality: Confidence Intervals

Consumption	$<$4 c/week	5–7 c/week	2–3 c/day	4–5 c/day	$\geq$ 6 c/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]

Reference group: $<$ 1 c/month. 95% Confidence intervals. QC uses $C/2 = 1.8c_\alpha$. I don't vouch for the analysis—Cox proportional hazard model, adjustments for age, smoking status, alcohol consumption, BMI, etc.