Supplementary materials for *Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models*

July 11, 2016

S1 Proof of Theorem 1

The Moore-Penrose inverse of $\mathbf{B}_i$ can be computed from its eigen-decomposition. Let $b \leq n_i$ denote the rank of $\mathbf{B}_i$. Let $\mathbf{\Lambda}$ be the $b \times b$ diagonal matrix of the positive eigenvalues of $\mathbf{B}_i$ and $\mathbf{V}$ be the $n_i \times b$ matrix of corresponding eigen-vectors, so that $\mathbf{B}_i = \mathbf{V\Lambda V}'$. Then $\mathbf{B}_i^+ = \mathbf{V\Lambda}^{-1}\mathbf{V}'$ and $\mathbf{B}_i^{+1/2} = \mathbf{V\Lambda}^{-1/2}\mathbf{V}'$. Now, observe that

$$\begin{aligned}
\ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{A}_i\left(\mathbf{I}-\mathbf{H_X}\right)_i\mathbf{\Phi}\left(\mathbf{I}-\mathbf{H_X}\right)_i'\mathbf{A}_i'\mathbf{W}_i\ddot{\mathbf{R}}_i &= \ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{D}_i\mathbf{B}_i^{+1/2}\mathbf{B}_i\mathbf{B}_i^{+1/2}\mathbf{D}_i'\mathbf{W}_i\ddot{\mathbf{R}}_i \\
&= \ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{D}_i\mathbf{V}\mathbf{V}'\mathbf{D}_i'\mathbf{W}_i\ddot{\mathbf{R}}_i.
\end{aligned} \tag{1}$$

Because $\mathbf{D}_i$, and $\mathbf{\Phi}$ are positive definite and $\mathbf{B}_i$ is symmetric, the eigen-vectors $\mathbf{V}$ define an orthonormal basis for the column span of $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. We now show that $\ddot{\mathbf{U}}_i$ is in the column space of $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. Let $\mathbf{Z}_i$ be an $n_i \times (r+s)$ matrix of zeros. Let $\mathbf{Z}_k = -\ddot{\mathbf{U}}_k\mathbf{L}^{-1}\mathbf{M}_{\ddot{\mathbf{U}}}^{-1}$, for $k \neq i$ and take $\mathbf{Z} = \left(\mathbf{Z}_1', ..., \mathbf{Z}_m'\right)'$. Now observe that $\left(\mathbf{I}-\mathbf{H_T}\right)\mathbf{Z} = \mathbf{Z}$. It follows that

$$\begin{aligned}
\left(\mathbf{I}-\mathbf{H_X}\right)_i\mathbf{Z} &= \left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i\left(\mathbf{I}-\mathbf{H_T}\right)\mathbf{Z} = \left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i\mathbf{Z} \\
&= \mathbf{Z}_i - \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\sum_{k=1}^m \ddot{\mathbf{U}}_k'\mathbf{W}_k\mathbf{Z}_k \\
&= \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\left(\sum_{k\neq i}\ddot{\mathbf{U}}_k'\mathbf{W}_k\ddot{\mathbf{U}}\right)\mathbf{L}^{-1}\mathbf{M}_{\ddot{\mathbf{U}}}^{-1} = \ddot{\mathbf{U}}_i.
\end{aligned}$$

Thus, there exists an $N \times (r+s)$ matrix $\mathbf{Z}$ such that $\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{X}}}\right)_i\mathbf{Z} = \ddot{\mathbf{U}}_i$, i.e., $\ddot{\mathbf{U}}_i$ is in the column span of $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. Because $\mathbf{D}_i\mathbf{W}_i$ is positive definite and $\ddot{\mathbf{R}}_i$ is a sub-matrix of $\ddot{\mathbf{U}}_i$, $\mathbf{D}_i\mathbf{W}_i\ddot{\mathbf{R}}_i$ is also in the column span of $\left(\mathbf{I}-\mathbf{H_X}\right)_i$. It follows that

$$\ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{D}_i\mathbf{V}\mathbf{V}'\mathbf{D}_i'\mathbf{W}_i\ddot{\mathbf{R}}_i = \ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{\Phi}_i\mathbf{W}_i\ddot{\mathbf{R}}_i. \tag{2}$$

Substituting (2) into (1) demonstrates that $\mathbf{A}_i$ satisfies the generalized BRL criterion (Eq. 6 of the main paper).

Under the working model, the residuals from cluster i have mean $\mathbf{0}$ and variance

$$\text{Var}\left(\ddot{\mathbf{e}}_i\right) = \left(\mathbf{I}-\mathbf{H_X}\right)_i\mathbf{\Phi}\left(\mathbf{I}-\mathbf{H_X}\right)_i',$$

It follows that

$$\begin{aligned}
\mathrm{E}\left(\mathbf{V}^{CR2}\right) &= \mathbf{M}_{\ddot{\mathbf{R}}}\left[\sum_{i=1}^{m} \ddot{\mathbf{R}}_i'\mathbf{W}_i\mathbf{A}_i\left(\mathbf{I}-\mathbf{H}_{\mathbf{X}}\right)_i \boldsymbol{\Phi}\left(\mathbf{I}-\mathbf{H}_{\mathbf{X}}\right)_i' \mathbf{A}_i\mathbf{W}_i\ddot{\mathbf{R}}_i\right]\mathbf{M}_{\ddot{\mathbf{R}}} \\
&= \mathbf{M}_{\ddot{\mathbf{R}}}\left[\sum_{i=1}^{m} \ddot{\mathbf{R}}_i'\mathbf{W}_i\boldsymbol{\Phi}_i\mathbf{W}_i\ddot{\mathbf{R}}_i\right]\mathbf{M}_{\ddot{\mathbf{R}}} \\
&= \mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)
\end{aligned}$$

S2 Proof of Theorem 2

From the fact that $\ddot{\mathbf{U}}_i'\mathbf{W}_i\mathbf{T}_i = \mathbf{0}$ for $i = 1, ..., m$, it follows that

$$\begin{aligned}
\mathbf{B}_i &= \mathbf{D}_i\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i\left(\mathbf{I}-\mathbf{H}_{\mathbf{T}}\right)\boldsymbol{\Phi}\left(\mathbf{I}-\mathbf{H}_{\mathbf{T}}\right)'\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i'\mathbf{D}_i' \\
&= \mathbf{D}_i\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}-\mathbf{H}_{\mathbf{T}}\right)_i\boldsymbol{\Phi}\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}-\mathbf{H}_{\mathbf{T}}\right)_i'\mathbf{D}_i' \\
&= \mathbf{D}_i\left(\boldsymbol{\Phi}_i - \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\ddot{\mathbf{U}}_i' - \mathbf{T}_i\mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\right)\mathbf{D}_i'
\end{aligned}$$

and

$$\mathbf{B}_i^{+} = \left(\mathbf{D}_i'\right)^{-1}\left(\boldsymbol{\Phi}_i - \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\ddot{\mathbf{U}}_i' - \mathbf{T}_i\mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\right)^{+}\mathbf{D}_i^{-1}. \tag{3}$$

Let $\boldsymbol{\Psi}_i = \left(\boldsymbol{\Phi}_i - \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\ddot{\mathbf{U}}_i'\right)^{+}$. Using a generalized Woodbury identity (Henderson and Searle, 1981),

$$\boldsymbol{\Psi}_i = \mathbf{W}_i + \mathbf{W}_i\ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\left(\mathbf{M}_{\ddot{\mathbf{U}}} - \mathbf{M}_{\ddot{\mathbf{U}}}\ddot{\mathbf{U}}_i'\mathbf{W}_i\ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\right)^{+}\mathbf{M}_{\ddot{\mathbf{U}}}\ddot{\mathbf{U}}_i'\mathbf{W}_i.$$

It follows that $\boldsymbol{\Psi}_i\mathbf{T}_i = \mathbf{W}_i\mathbf{T}_i$. Another application of the generalized Woodbury identity gives

$$\begin{aligned}
\left(\boldsymbol{\Phi}_i - \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\ddot{\mathbf{U}}_i' - \mathbf{T}_i\mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\right)^{+} &= \boldsymbol{\Psi}_i + \boldsymbol{\Psi}_i\mathbf{T}_i\mathbf{M}_{\mathbf{T}}\left(\mathbf{M}_{\mathbf{T}} - \mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\boldsymbol{\Psi}_i\mathbf{T}_i\mathbf{M}_{\mathbf{T}}\right)^{+}\mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\boldsymbol{\Psi}_i \\
&= \boldsymbol{\Psi}_i + \mathbf{W}_i\mathbf{T}_i\mathbf{M}_{\mathbf{T}}\left(\mathbf{M}_{\mathbf{T}} - \mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\mathbf{W}_i\mathbf{T}_i\mathbf{M}_{\mathbf{T}}\right)^{+}\mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\mathbf{W}_i \\
&= \boldsymbol{\Psi}_i.
\end{aligned}$$

The last equality follows from the fact that

$$\mathbf{T}_i\mathbf{M}_{\mathbf{T}}\left(\mathbf{M}_{\mathbf{T}} - \mathbf{M}_{\mathbf{T}}\mathbf{T}_i'\mathbf{W}_i\mathbf{T}_i\mathbf{M}_{\mathbf{T}}\right)^{-}\mathbf{M}_{\mathbf{T}}\mathbf{T}_i' = \mathbf{0}$$

because the fixed effects are nested within clusters. Substituting into (3), we then have that $\mathbf{B}_i^{+} = \left(\mathbf{D}_i'\right)^{-1}\boldsymbol{\Psi}_i\mathbf{D}_i^{-1}$. But

$$\tilde{\mathbf{B}}_i = \mathbf{D}_i\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i\boldsymbol{\Phi}\left(\mathbf{I}-\mathbf{H}_{\ddot{\mathbf{U}}}\right)_i'\mathbf{D}_i' = \mathbf{D}_i\left(\boldsymbol{\Phi}_i - \ddot{\mathbf{U}}_i\mathbf{M}_{\ddot{\mathbf{U}}}\ddot{\mathbf{U}}_i'\right)\mathbf{D}_i' = \mathbf{D}_i\boldsymbol{\Psi}_i^{+}\mathbf{D}_i',$$

and so $\mathbf{B}_i^{+} = \tilde{\mathbf{B}}_i^{+}$. It follows that $\mathbf{A}_i = \tilde{\mathbf{A}}_i$ for $i = 1, ..., m$.

S3 Basic difference-in-differences example

Consider a simple difference-in-differences design with m clusters and $n = 2$ time periods. Suppose that the first m_0 clusters remain untreated in the second time period and the remaining $m_1 = m - m_0$ clusters are treated in the second time period. The basic difference-in-differences model for this design is then

$$y_{it} = \alpha_i + \beta_t + \delta T_{it} + e_{it}, \tag{4}$$

where $T_{i1} = 1$ for $i = m_0 + 1, ..., m$, $T_{it} = 0$ otherwise, and δ is the average treatment effect.

Estimating δ by OLS is exactly equivalent to taking first differences and then calculating the mean difference between treated and untreated clusters. Let $d_i = y_{i1} - y_{i0}$ for $i = 1, ..., m$, $\bar{d}_0 = \sum_{i=1}^{m_0} d_i/m_0$, and $\bar{d}_1 = \sum_{i=m_0+1}^{m} d_i/m_1$. Then $\hat{\delta} = \bar{d}_1 - \bar{d}_0$. In this simplified representation of the model, it is clear that the null hypothesis $\delta = 0$ may be tested using a simple two-sample t-test on the difference scores, while allowing for unequal variances. The sampling variance of $\hat{\delta}$ can be estimated from the difference scores as

$$V_\Delta = \frac{1}{m_0(m_0-1)} \sum_{i=1}^{m_0} \left(d_i - \bar{d}_0\right)^2 + \frac{1}{m_1(m_1-1)} \sum_{i=m_0+1}^{m} \left(d_i - \bar{d}_1\right)^2.$$

Under a "working homoskedasticity" model, the degrees of freedom corresponding to V_Δ are

$$\nu_\Delta = \frac{m^2(m_0-1)(m_1-1)}{m_0^2(m_0-1) + m_1^2(m_1-1)}$$

(Imbens and Kolesar, 2015).

We shall now consider the variance estimator and degrees of freedom generated by the CR2 correction as applied to the full difference-in-differences model (4), while estimating δ after absorbing the cluster- and period-specific effects. We use the "working independence" model for deriving the CR2 adjustment matrices and degrees of freedom. Following the notation of the main paper, this design has

$$\mathbf{R}_i = \begin{bmatrix} 0 \\ T_{i1} \end{bmatrix} \qquad \mathbf{S}_i = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \mathbf{T}_i = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} I(i=1) & I(i=2) & \cdots & I(i=m) \end{bmatrix}$$

After absorption, $\ddot{\mathbf{R}}_i = \left(T_{i1} - m_1/m\right)/2 \begin{bmatrix} -1 \\ 1 \end{bmatrix}'$, $\mathbf{M}_{\ddot{\mathbf{R}}} = 2m/(m_0 m_1)$, and

$$\mathbf{e}_i = \frac{d_i - \bar{d}_0}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{for} \quad i = 1, ..., m_0, \qquad \mathbf{e}_i = \frac{d_i - \bar{d}_1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \text{for} \quad i = m_0+1, ..., m.$$

If the CR2 adjustment matrices are calculated based on the absorbed model only, then

$$\mathbf{A}_i = \left(\mathbf{I}_i - \ddot{\mathbf{R}}_i \mathbf{M}_{\ddot{\mathbf{R}}} \ddot{\mathbf{R}}_i'\right)^{+1/2} = \begin{bmatrix} 1 + a_i & -a_i \\ -a_i & 1 + a_i \end{bmatrix},$$

where

$$\begin{aligned} a_i &= \frac{1}{2}\left(\sqrt{\frac{m_0 m}{m_0 m - m_1}} - 1\right) \qquad & i = 1, ..., m_0 \\ a_i &= \frac{1}{2}\left(\sqrt{\frac{m_1 m}{m_1 m - m_0}} - 1\right) \qquad & i = m_0+1, ..., m. \end{aligned}$$

Using these adjustment matrices yields the variance estimator

$$V_{\ddot{\mathbf{R}}} = \frac{1}{m_0(m_0 - m_1/m)} \sum_{i=1}^{m_0} \left(d_i - \bar{d}_0\right)^2 + \frac{1}{m_1(m_1 - m_0/m)} \sum_{i=m_0+1}^{m} \left(d_i - \bar{d}_1\right)^2,$$

which will be slightly smaller than V_Δ, with Satterthwaite degrees of freedom

$$\nu_{\ddot{\mathbf{R}}} = \frac{\left(\dfrac{m_0-1}{m_0(m_0 - m_1/m)} + \dfrac{m_1-1}{m_1(m_1 - m_0/m)}\right)^2}{\dfrac{1}{m_0(m_0 - m_1/m)} + \dfrac{1}{m_1(m_1 - m_0/m)}},$$

which will be slightly larger than ν_Δ.

Now consider calculating the adjustment matrices using the full design matrix, as recommended in the paper. Theorem 2 implies that the adjustment matrices can be calculated from $\ddot{\mathbf{U}}$, ignoring the cluster-specific effects. We then have

$$\mathbf{A}_i = \left(\mathbf{I}_i - \ddot{\mathbf{U}}_i \mathbf{M}_{\ddot{\mathbf{U}}} \ddot{\mathbf{U}}_i'\right)^{+1/2} = \begin{bmatrix} 1 + b_i & -b_i \\ -b_i & 1 + b_i \end{bmatrix},$$

where

$$b_i = \frac{1}{2}\left(\sqrt{\frac{m_0}{m_0 - 1}} - 1\right) \qquad i = 1, ..., m_0$$
$$b_i = \frac{1}{2}\left(\sqrt{\frac{m_1}{m_1 - 1}} - 1\right) \qquad i = m_0 + 1, ..., m.$$

It can be verified that using these adjustment matrices yields a variance estimator that is exactly equivalent to V_Δ, with degrees of freedom equal to ν_Δ.

S4 Details of simulation study

This section provides further details regarding the design of the simulations reported in Section 4 of the main text. The simulations examined six distinct study designs. Outcomes are measured for n units (which may be individuals, as in a cluster-randomized or block-randomized design, or time-points, as in a difference-in-differences panel) in each of m clusters under one of three treatment conditions. Suppose that there are G sets of clusters, each of size m_g, where the clusters in each set have a distinct configuration of treatment assignments. Let n_{ghi} denote the number of units at which cluster i in configuration g is observed under condition h, for $i = 1, ..., m$, $g = 1, ..., G$, and $h = 1, 2, 3$. Table 1 summarizes the cluster-level sample sizes and unit-level patterns of treatment allocation for each of the six designs. The simulated designs included the following:

1. A balanced, block-randomized design, with an un-equal allocation within each block. In the balanced design, the treatment allocation is identical for each block, so $G = 1$.
2. An unbalanced, block-randomized design, with two different patterns of treatment allocation ($G = 2$).
3. A balanced, cluster-randomized design, in which units are nested within clusters and an equal number of clusters are assigned to each treatment condition.
4. An unbalanced, cluster-randomized design, in which units are nested within clusters but the number of clusters assigned to each condition is not equal.
5. A balanced difference-in-differences design with two patterns of treatment allocation ($G = 2$), in which half of the clusters are observed under the first treatment condition only and the remaining half are observed under all three conditions.
6. An unbalanced difference-in-differences design, again with two patterns of treatment allocation ($G = 2$), but where 2/3 of the clusters are observed under the first treatment condition only and the remaining 1/3 of clusters are observed under all three conditions.

Table 1: Study designs used for simulation

Study design	Balance	Configuration	Clusters	Treatment allocation
Randomized Block	Balanced	1	$m_1 = m$	$n_{11i} = n/2, n_{12i} = n/3, n_{13i} = n/6$
Randomized Block	Unbalanced	1	$m_1 = m/2$	$n_{11i} = n/2, n_{12i} = n/3, n_{13i} = n/6$
		2	$m_2 = m/2$	$n_{21i} = n/3, n_{22i} = 5n/9, n_{23i} = n/9$
Cluster-Randomized	Balanced	1	$m_1 = m/3$	$n_{11i} = n$
		2	$m_2 = m/3$	$n_{22i} = n$
		3	$m_3 = m/3$	$n_{33i} = n$
Cluster-Randomized	Unbalanced	1	$m_1 = m/2$	$n_{11i} = n$
		2	$m_2 = 3m/10$	$n_{22i} = n$
		3	$m_3 = m/5$	$n_{33i} = n$
Difference-in-Differences	Balanced	1	$m_1 = m/2$	$n_{11i} = n$
		2	$m_2 = m/2$	$n_{21i} = n/2, n_{22i} = n/3, n_{23i} = n/6$
Difference-in-Differences	Unbalanced	1	$m_1 = 2m/3$	$n_{11i} = n$
		2	$m_2 = m/3$	$n_{21i} = n/2, n_{22i} = n/3, n_{23i} = n/6$

S5 Additional simulation results

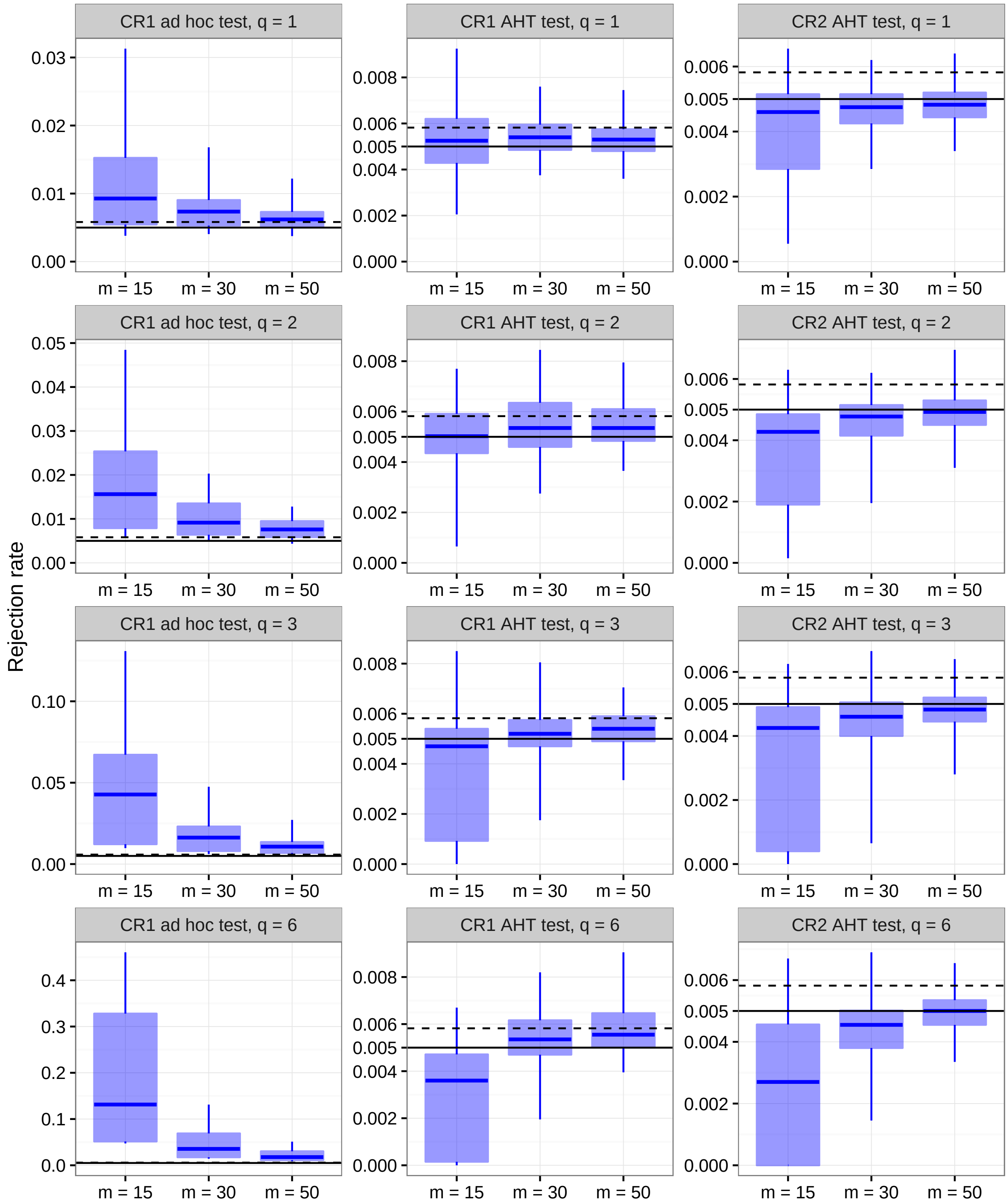

Figure S1: Rejection rates of ad hoc and AHT tests for $\alpha = .005$, by dimension of hypothesis (q) and sample size (m).

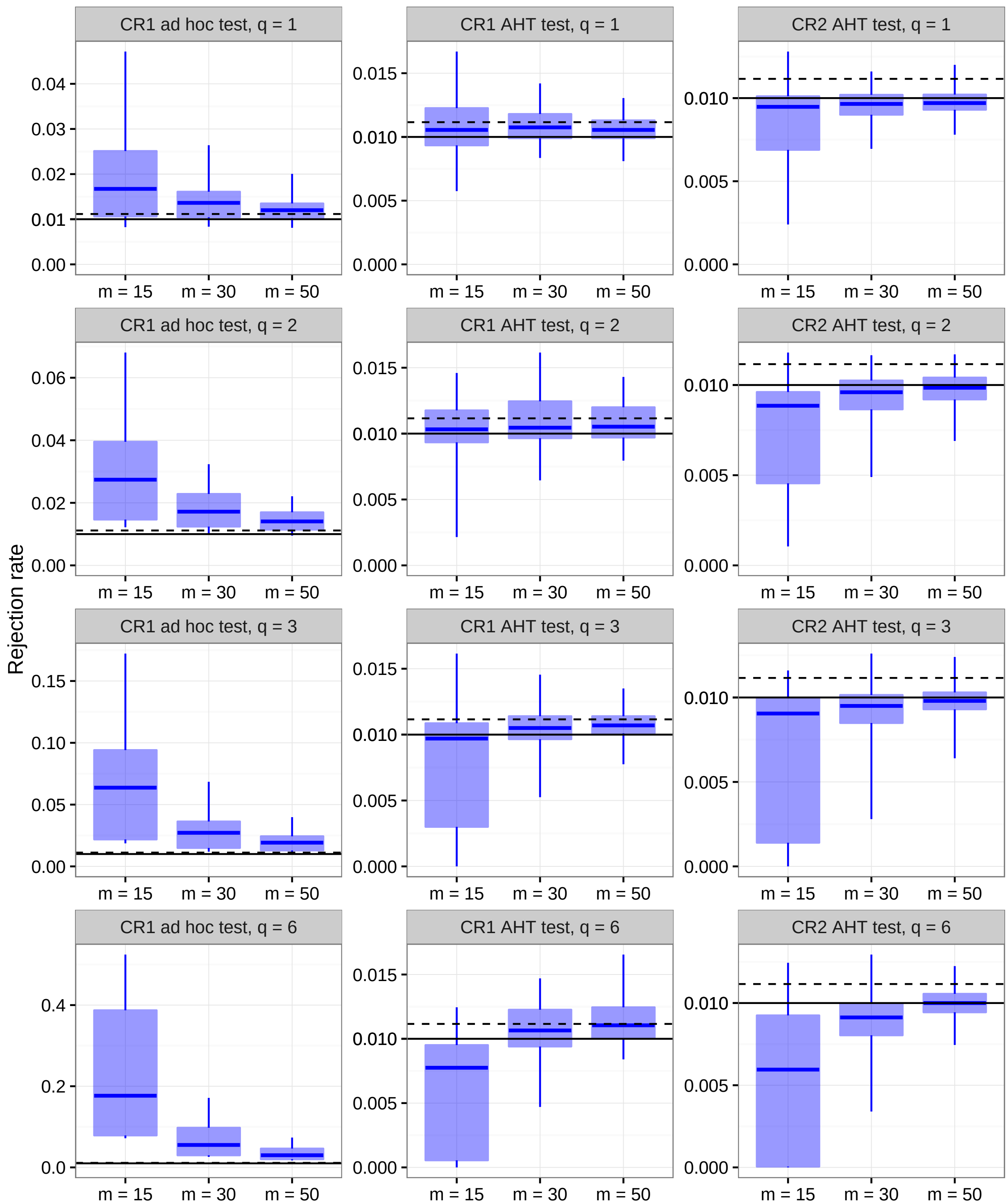

Figure S2: Rejection rates of ad hoc and AHT tests for $\alpha = .01$, by dimension of hypothesis (q) and sample size (m).

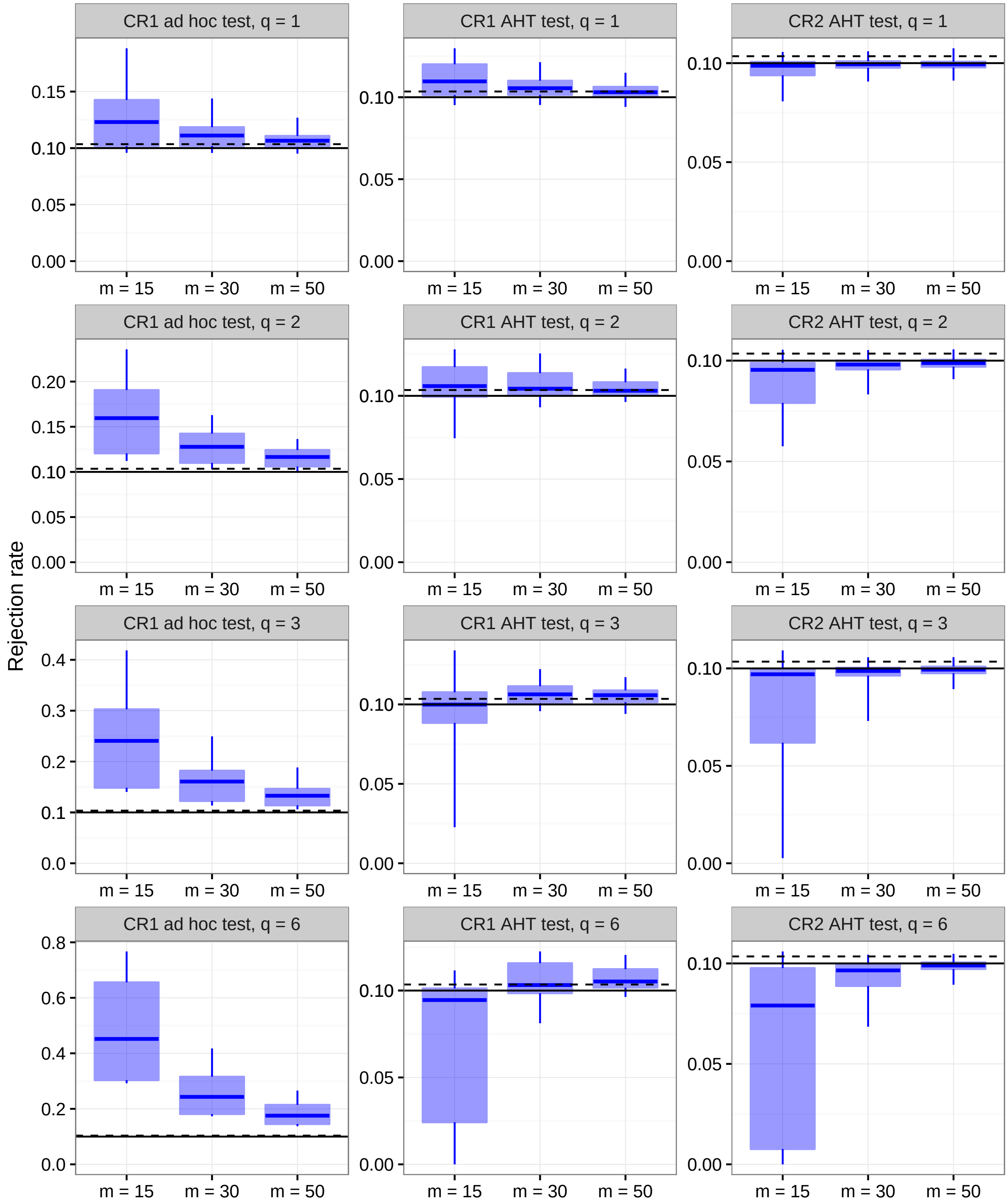

Figure S3: Rejection rates of ad hoc and AHT tests for $\alpha = .10$, by dimension of hypothesis (q) and sample size (m).

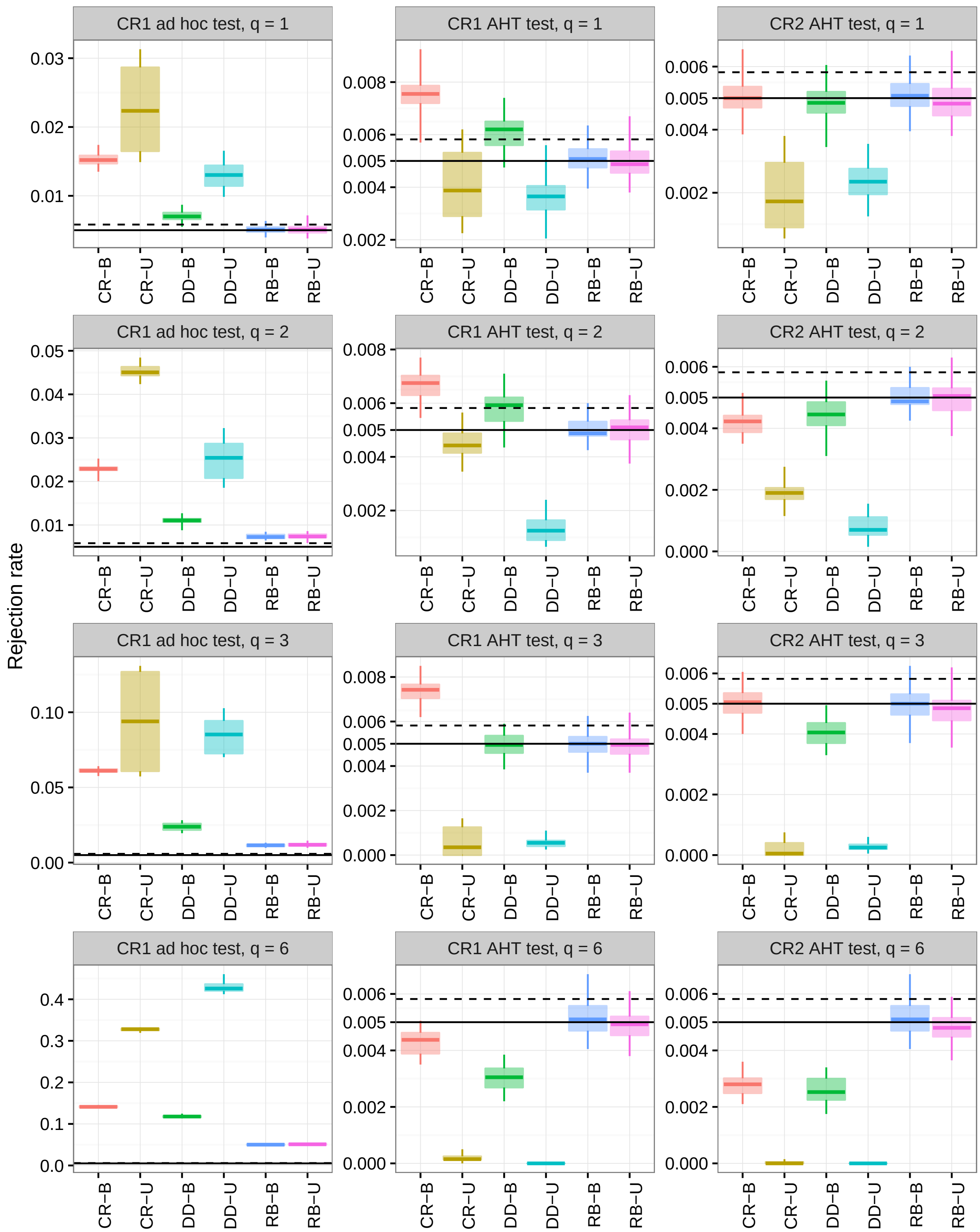

Figure S4: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .005$ and $m = 15$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

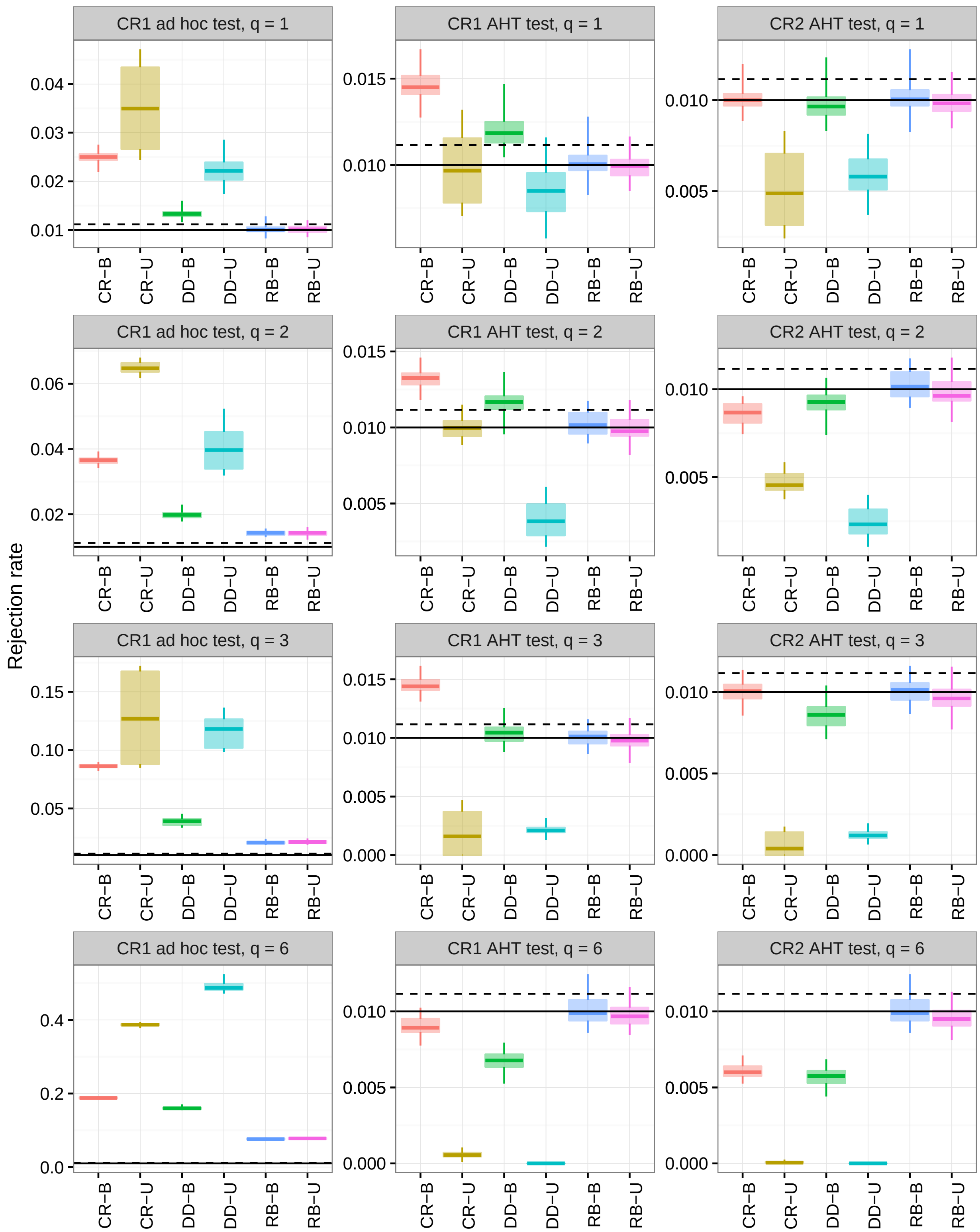

Figure S5: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .01$ and $m = 15$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

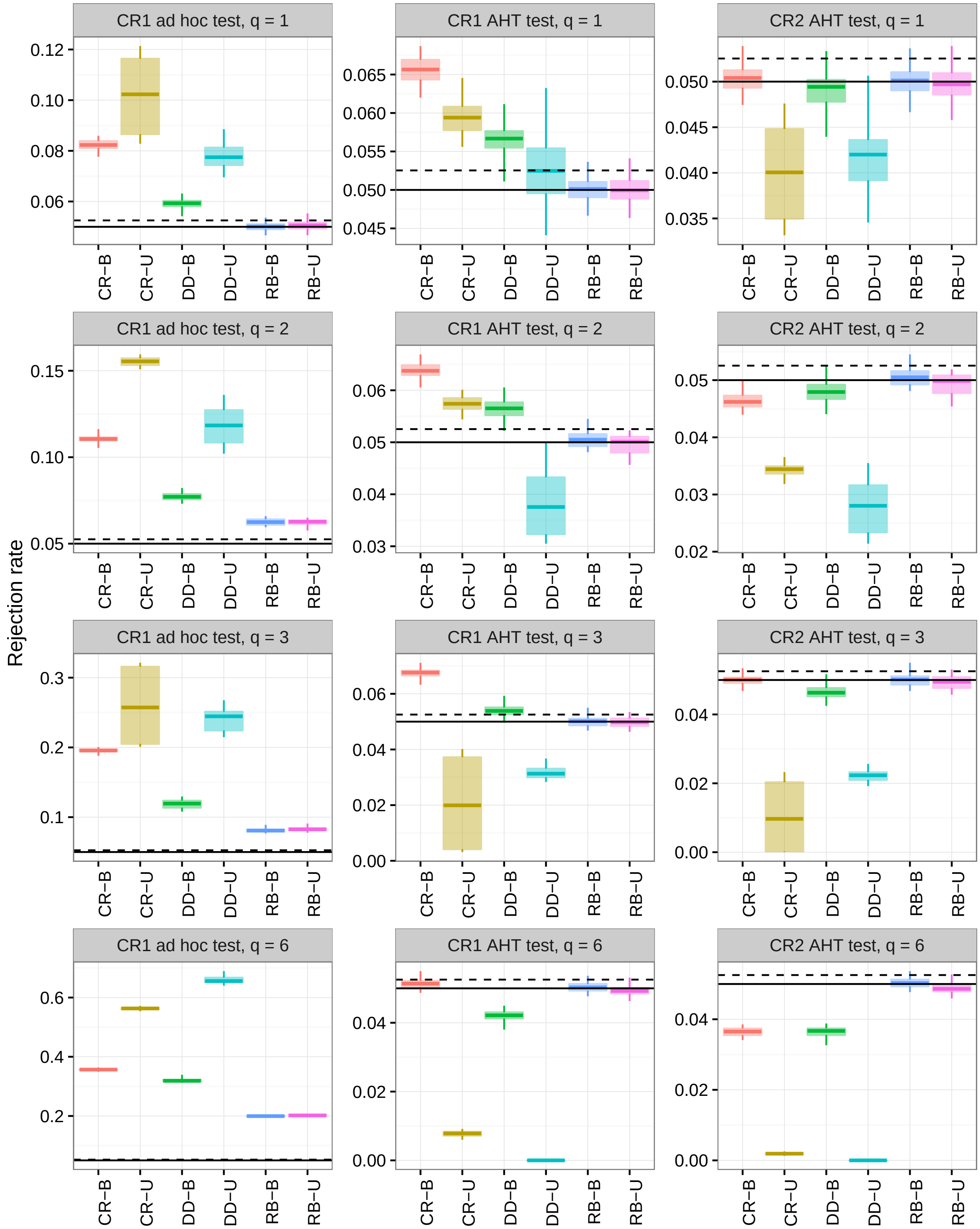

Figure S6: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .05$ and $m = 15$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

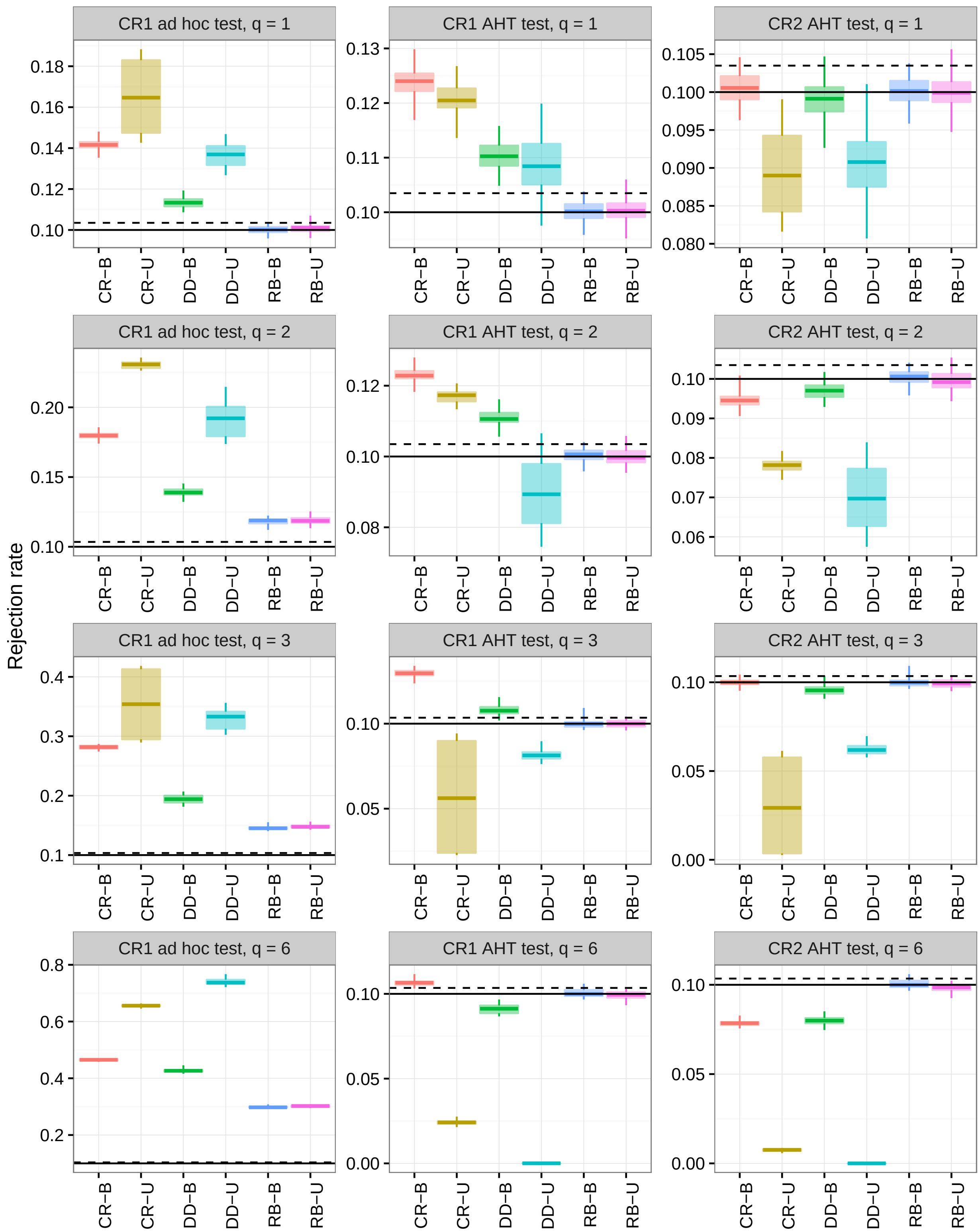

Figure S7: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .10$ and $m = 15$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

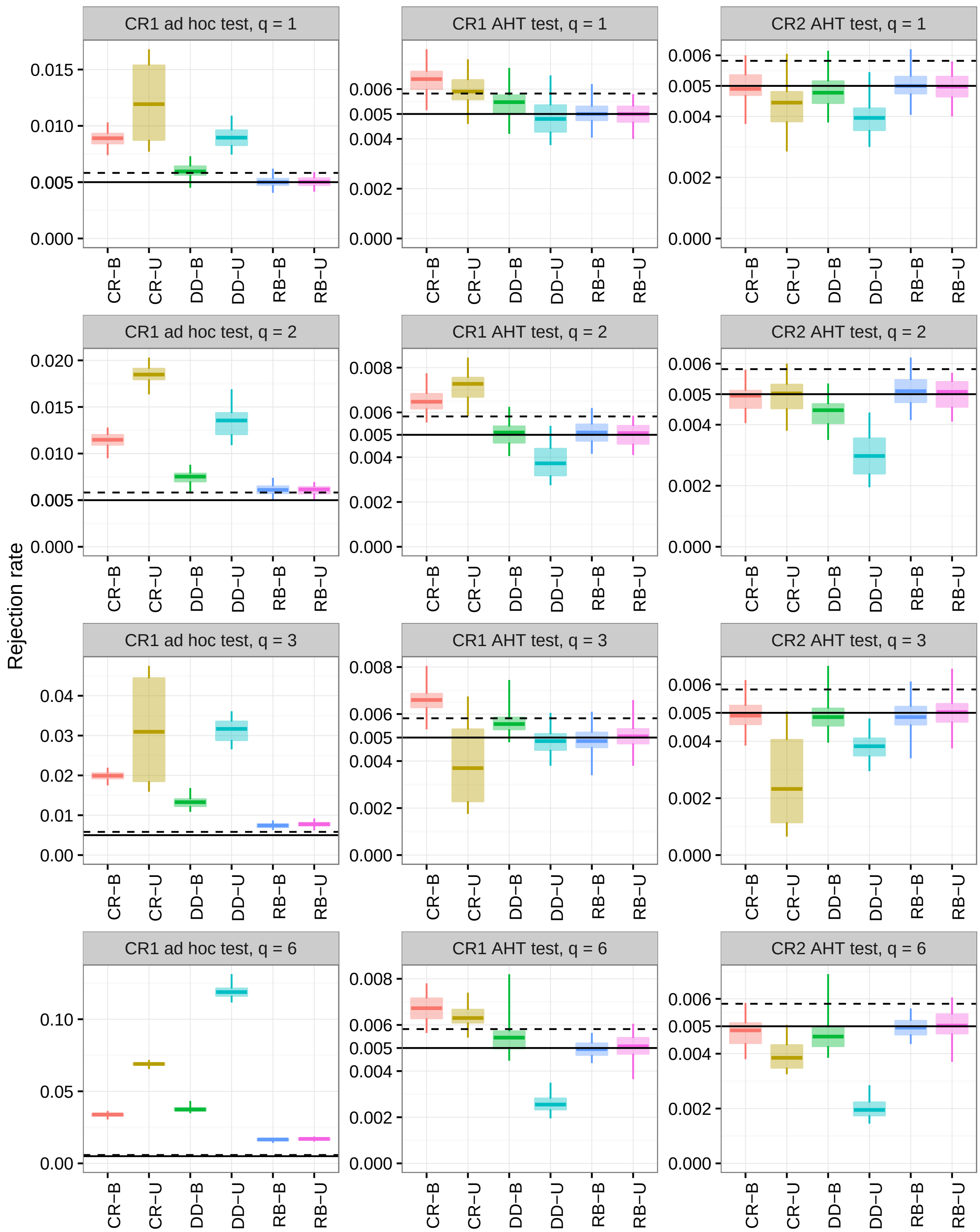

Figure S8: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .005$ and $m = 30$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

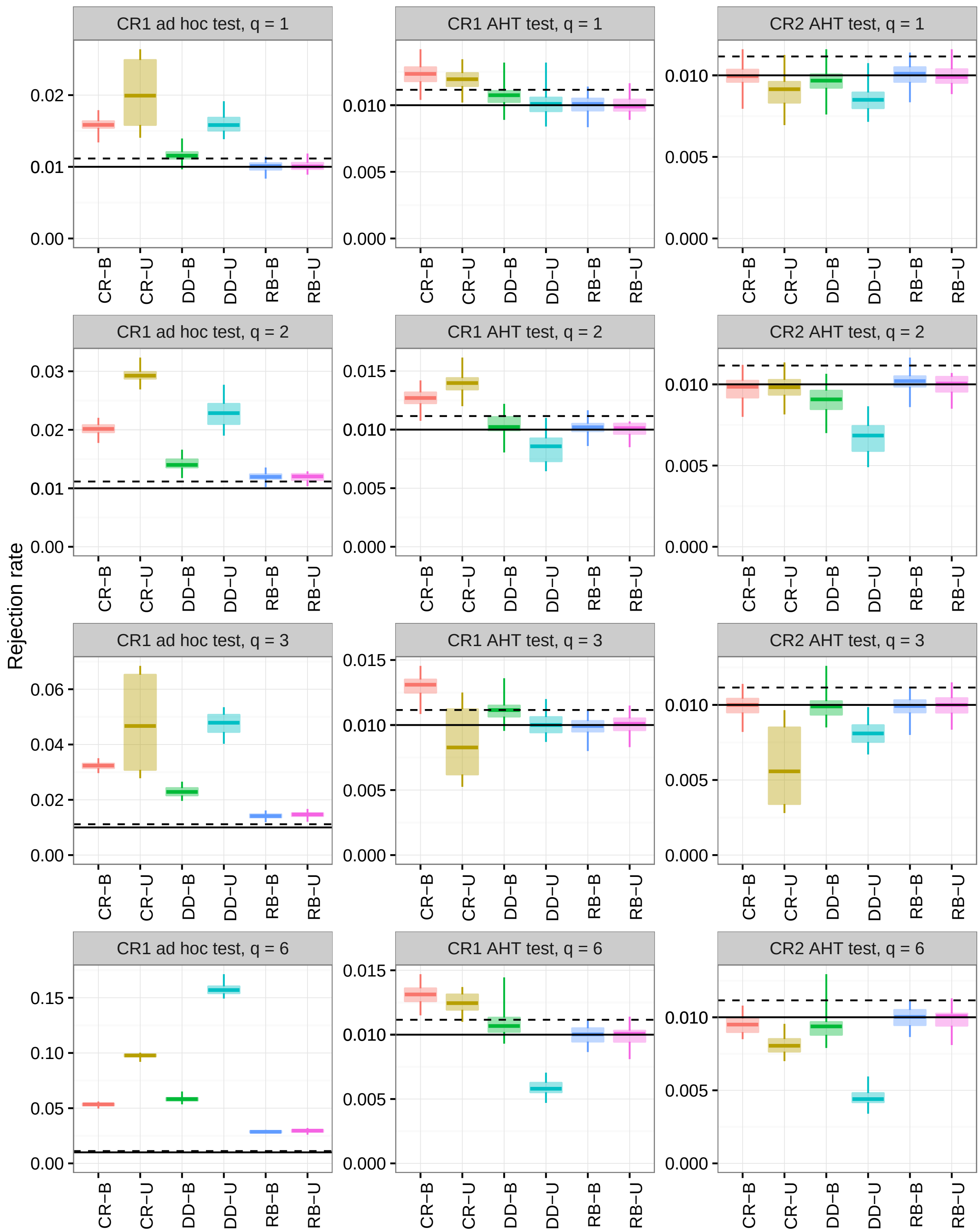

Figure S9: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .01$ and $m = 30$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

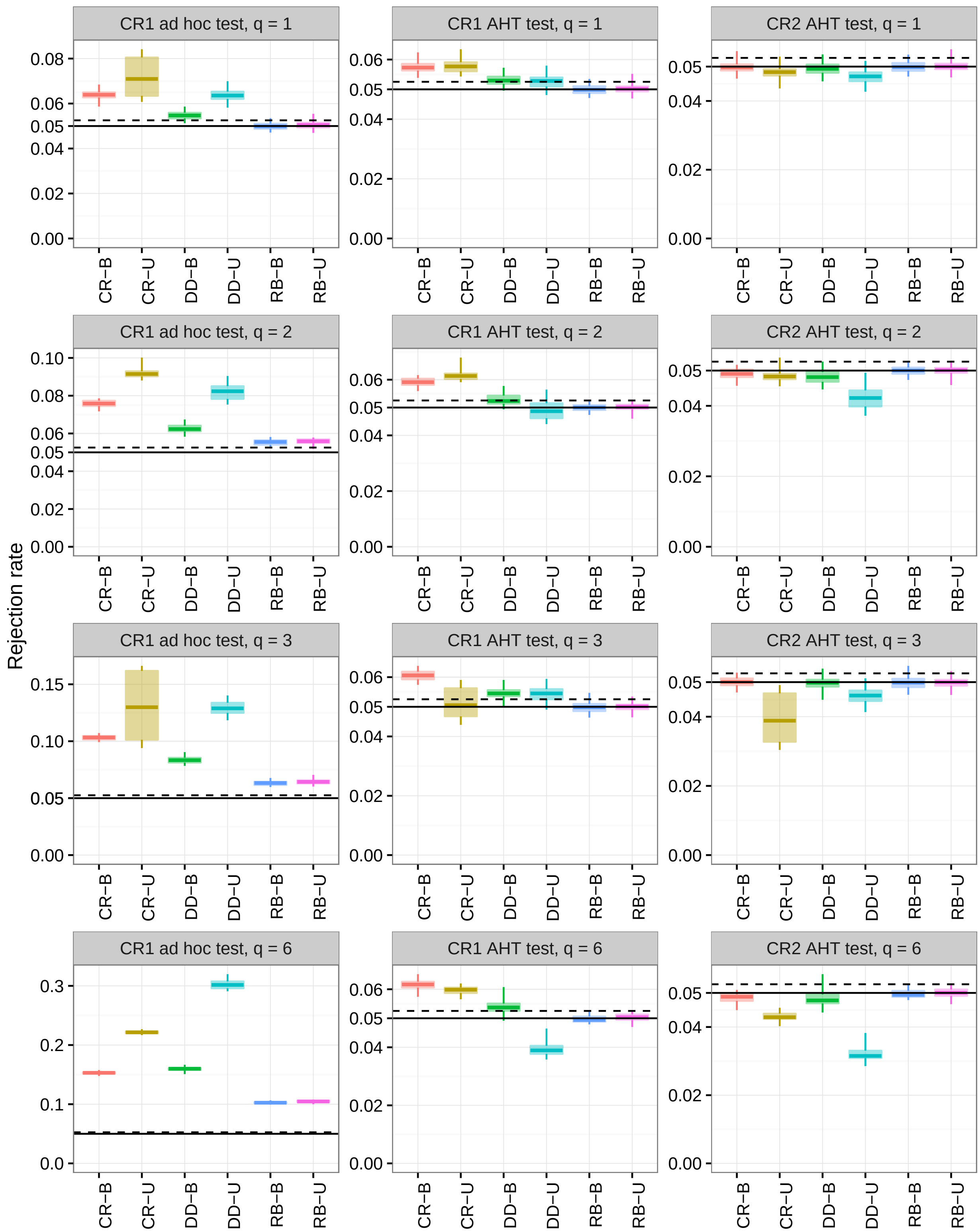

Figure S10: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .05$ and $m = 30$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

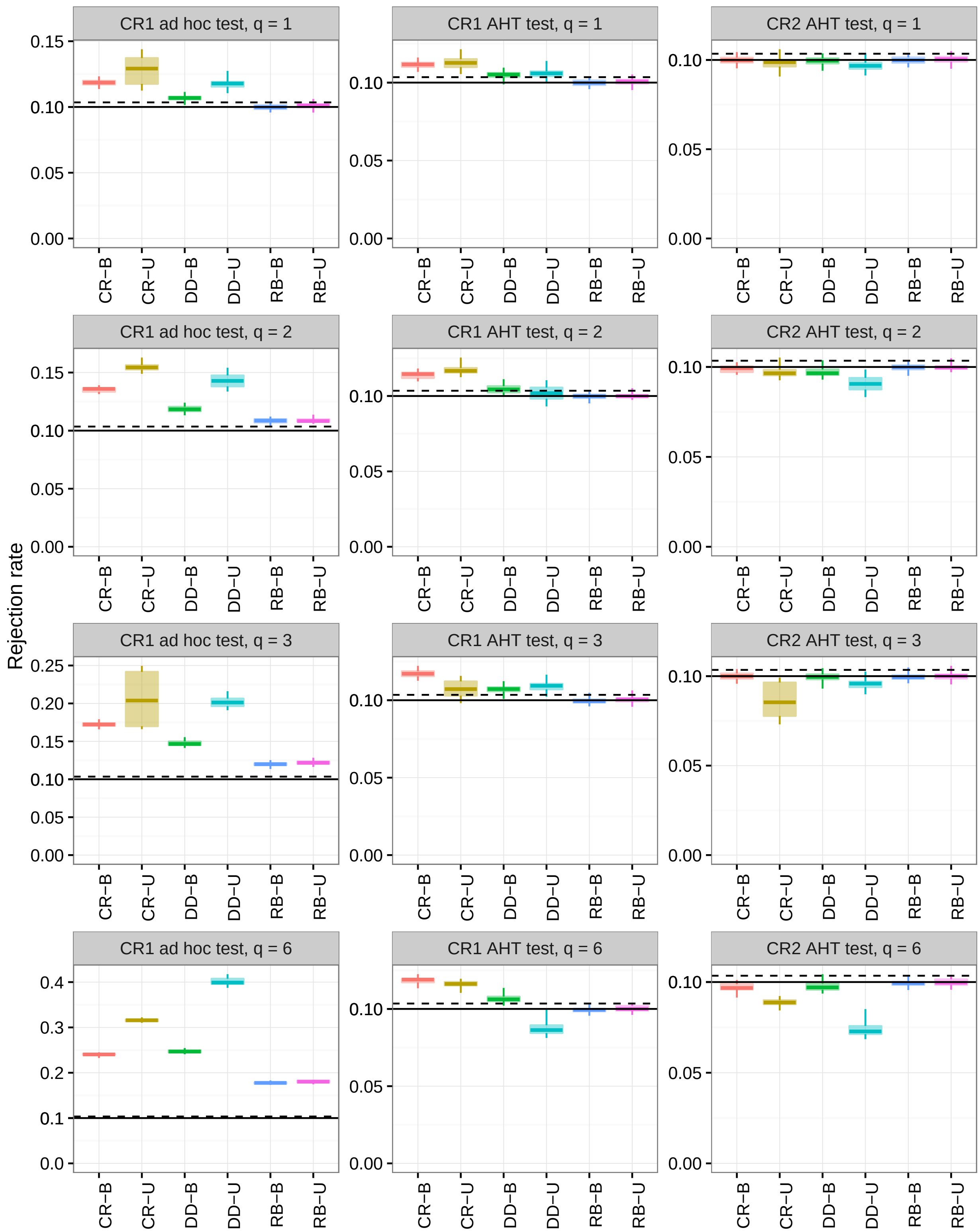

Figure S11: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .10$ and $m = 30$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

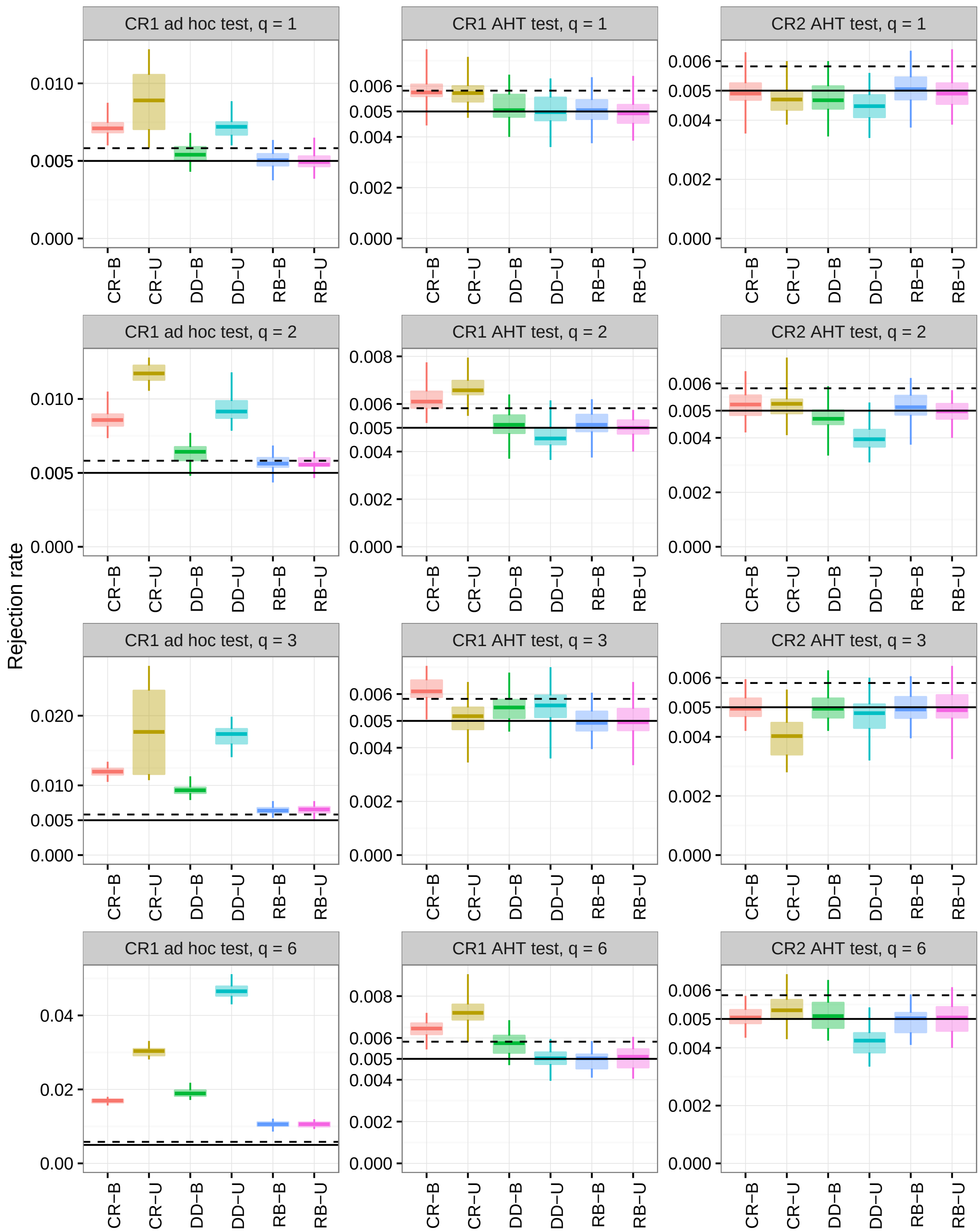

Figure S12: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .005$ and $m = 50$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

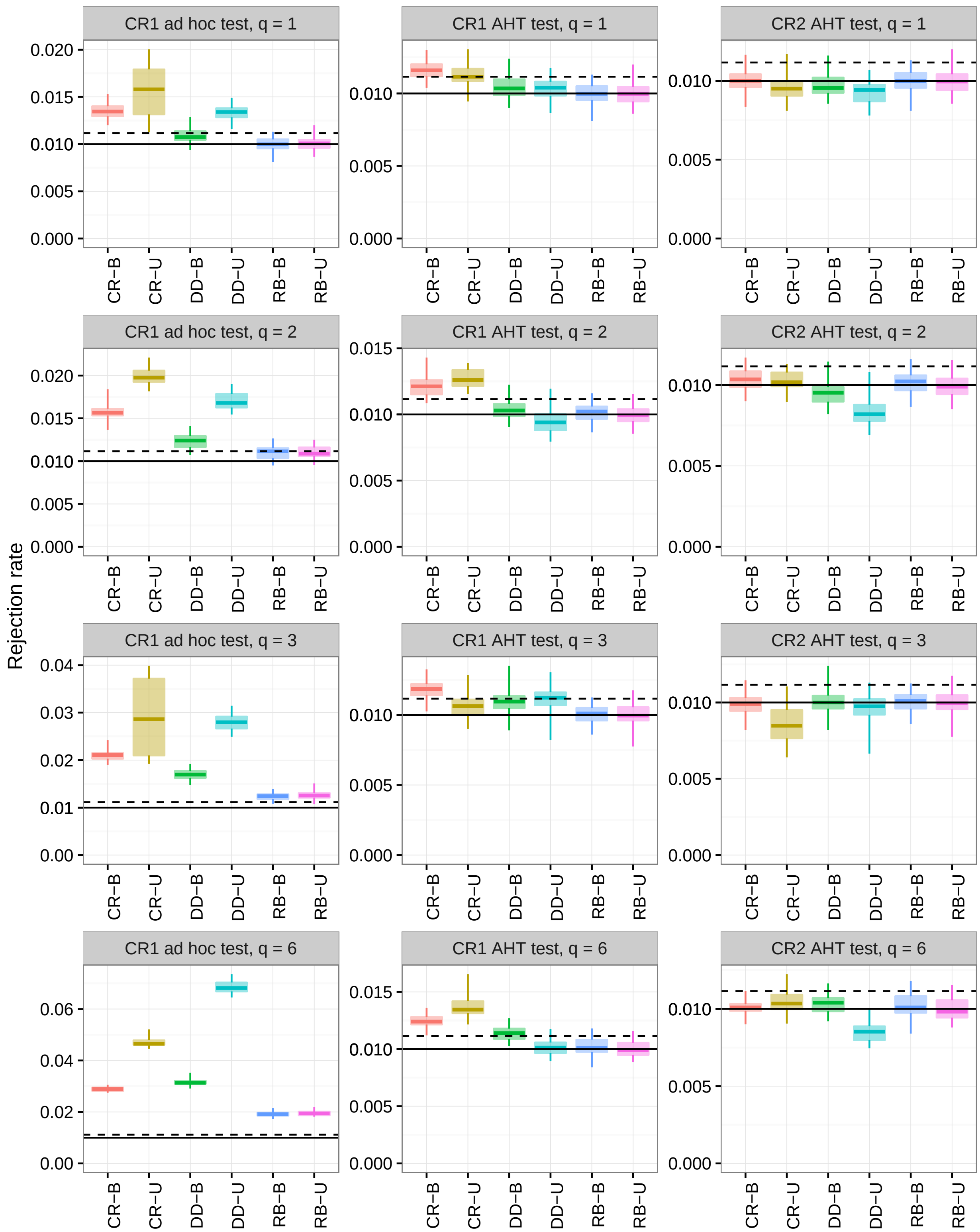

Figure S13: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .01$ and $m = 50$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

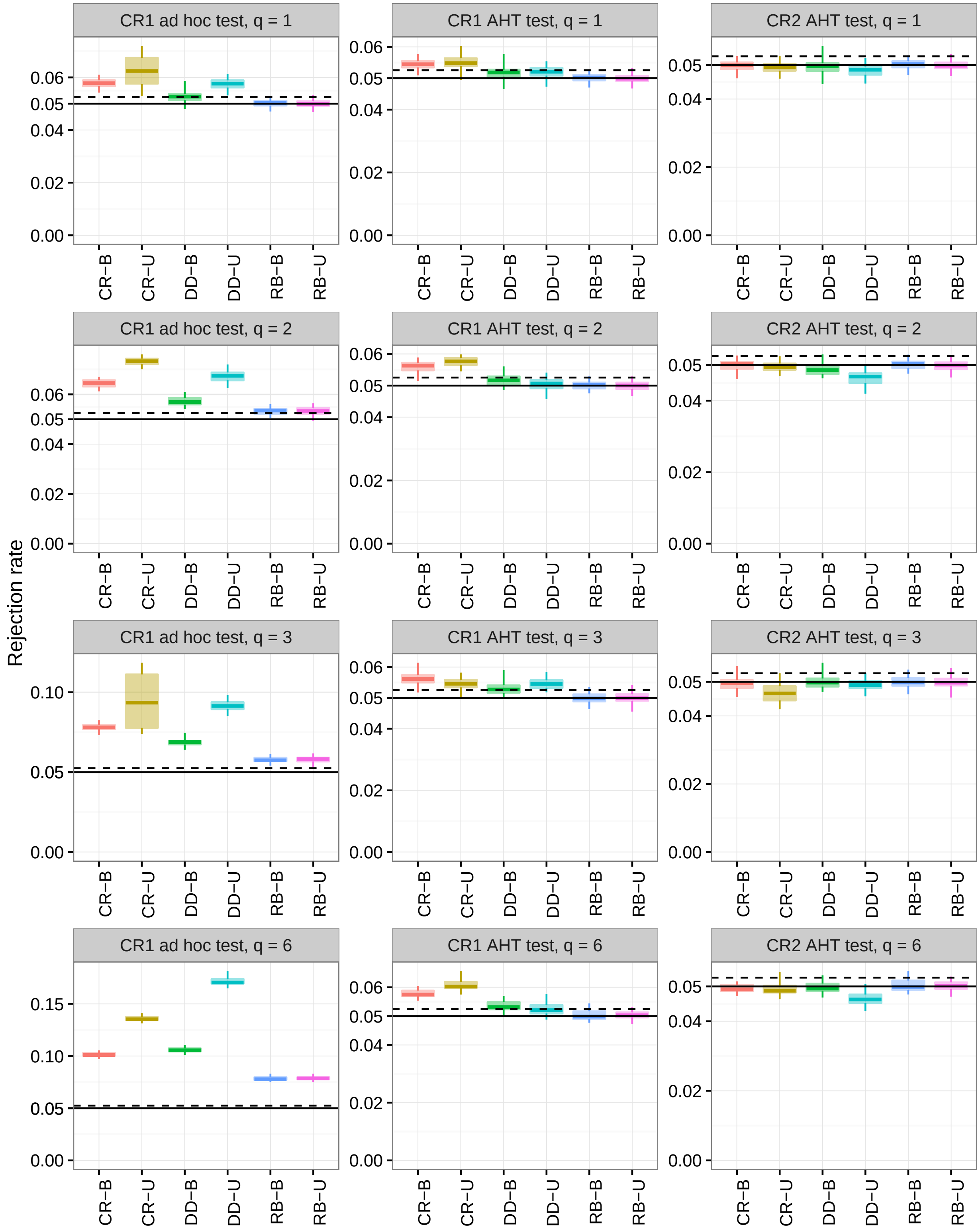

Figure S14: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .05$ and $m = 50$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

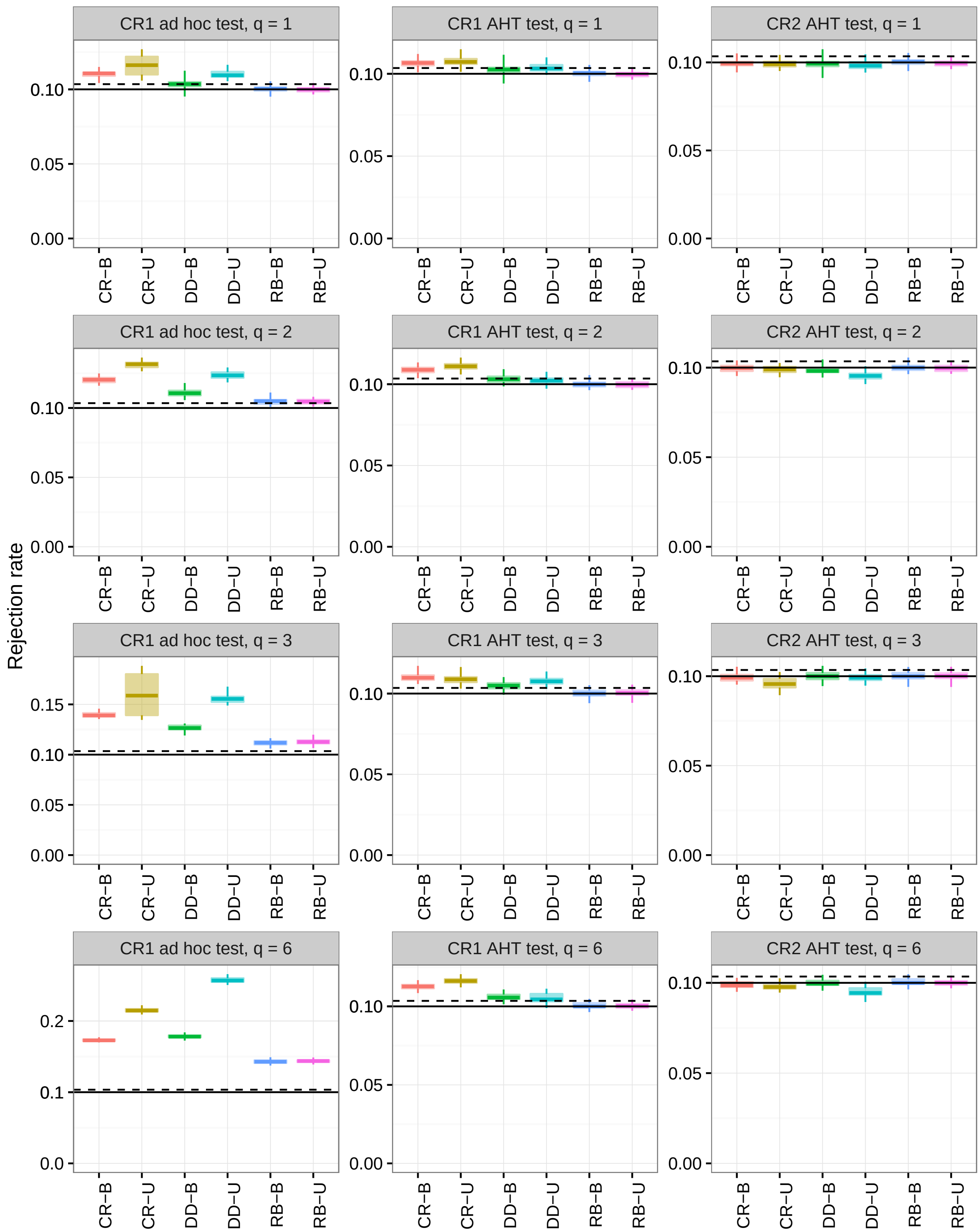

Figure S15: Rejection rates of ad hoc and AHT tests, by study design and dimension of hypothesis (q) for $\alpha = .10$ and $m = 50$. CR = cluster-randomized design; DD = difference-in-differences design; RB = randomized block design; B = balanced; U = unbalanced.

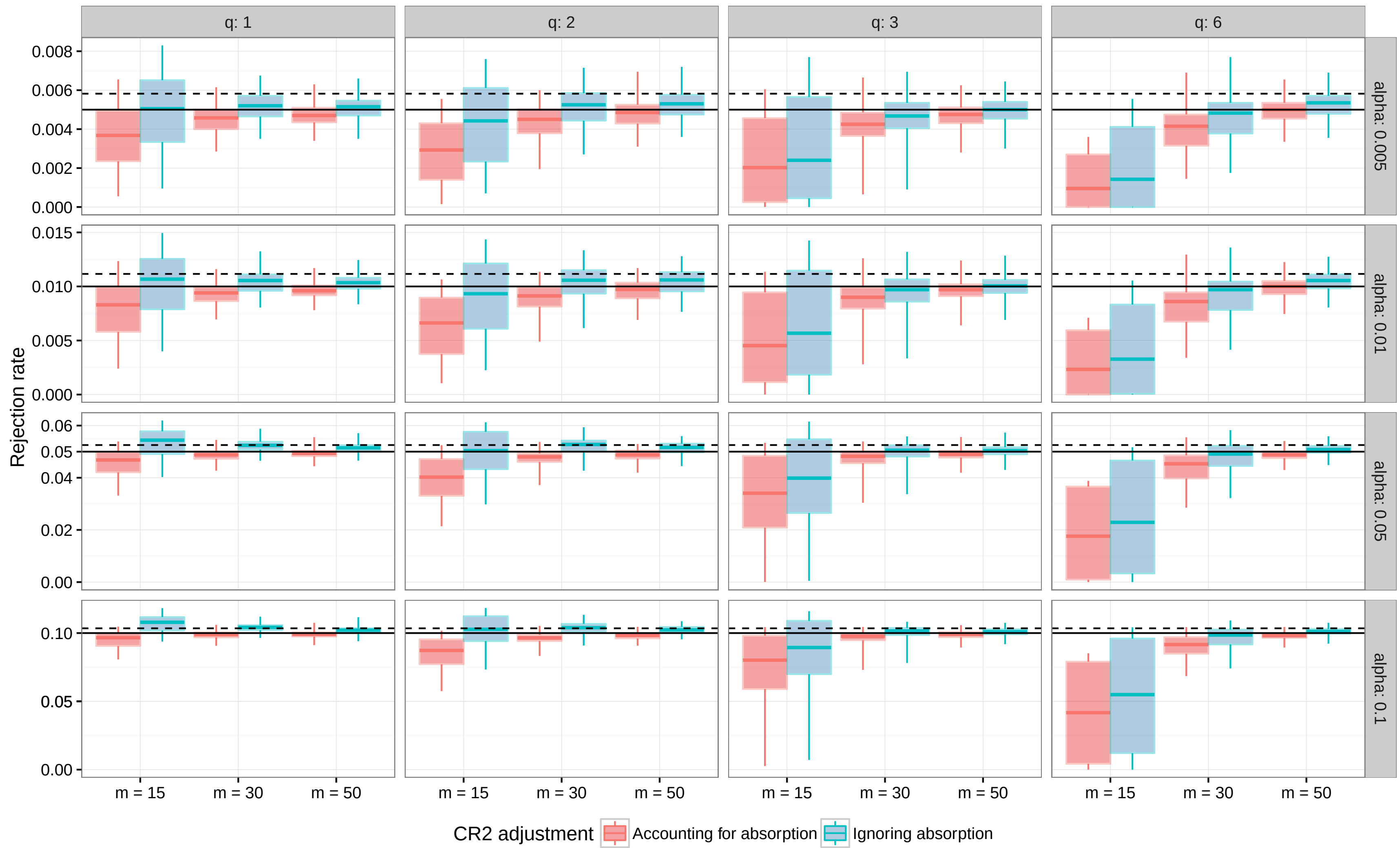

Figure S16: Rejection rates of AHT test using CR2, calculated with and without accounting for absorption of fixed effects, by sample size (m), dimension of hypothesis (q), and α-level. Results for the balanced and unbalanced randomized block designs are excluded because accounting for absorption of fixed effects has no consequence for these designs.

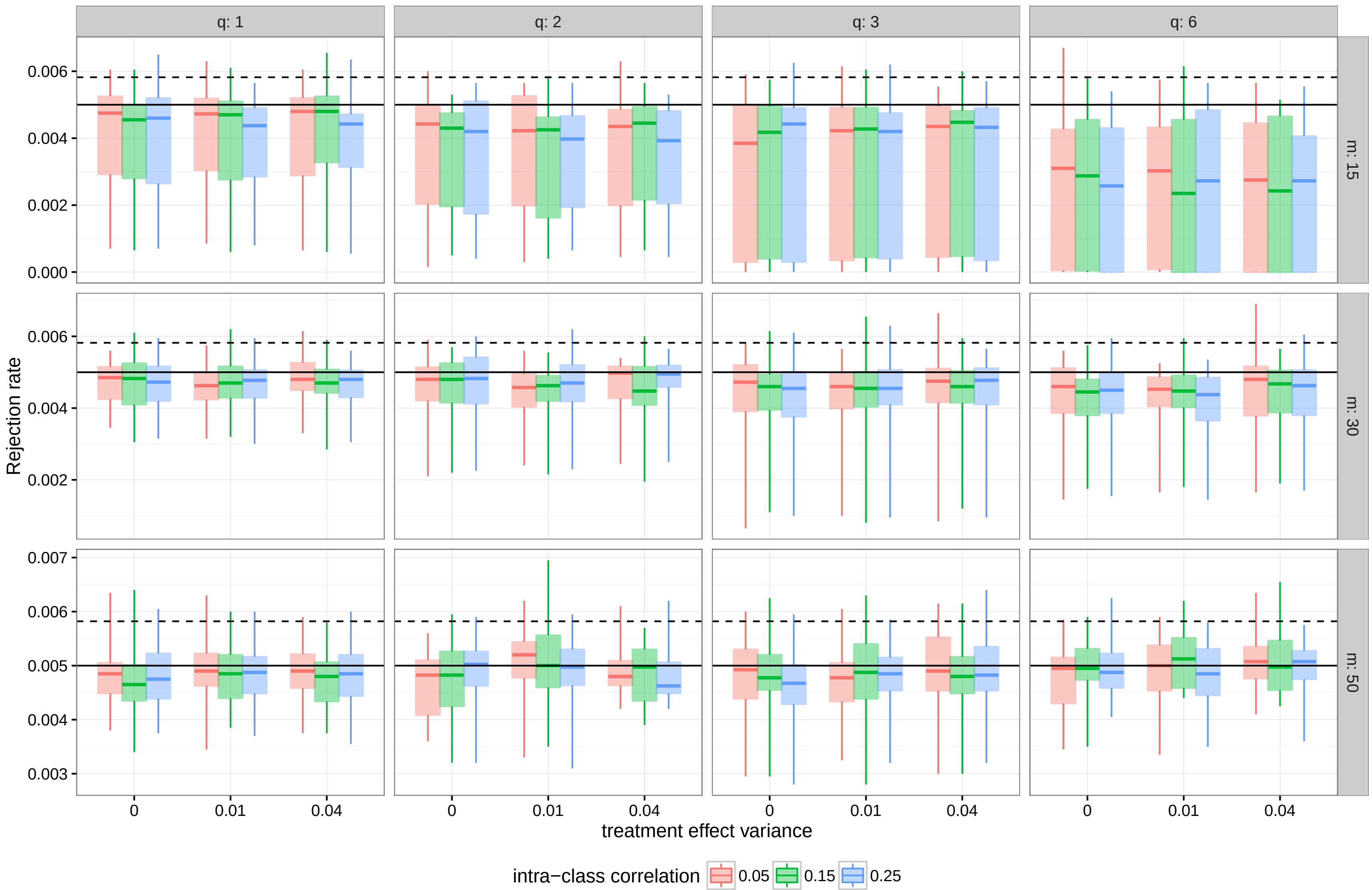

Figure S17: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for $\alpha = .005$.

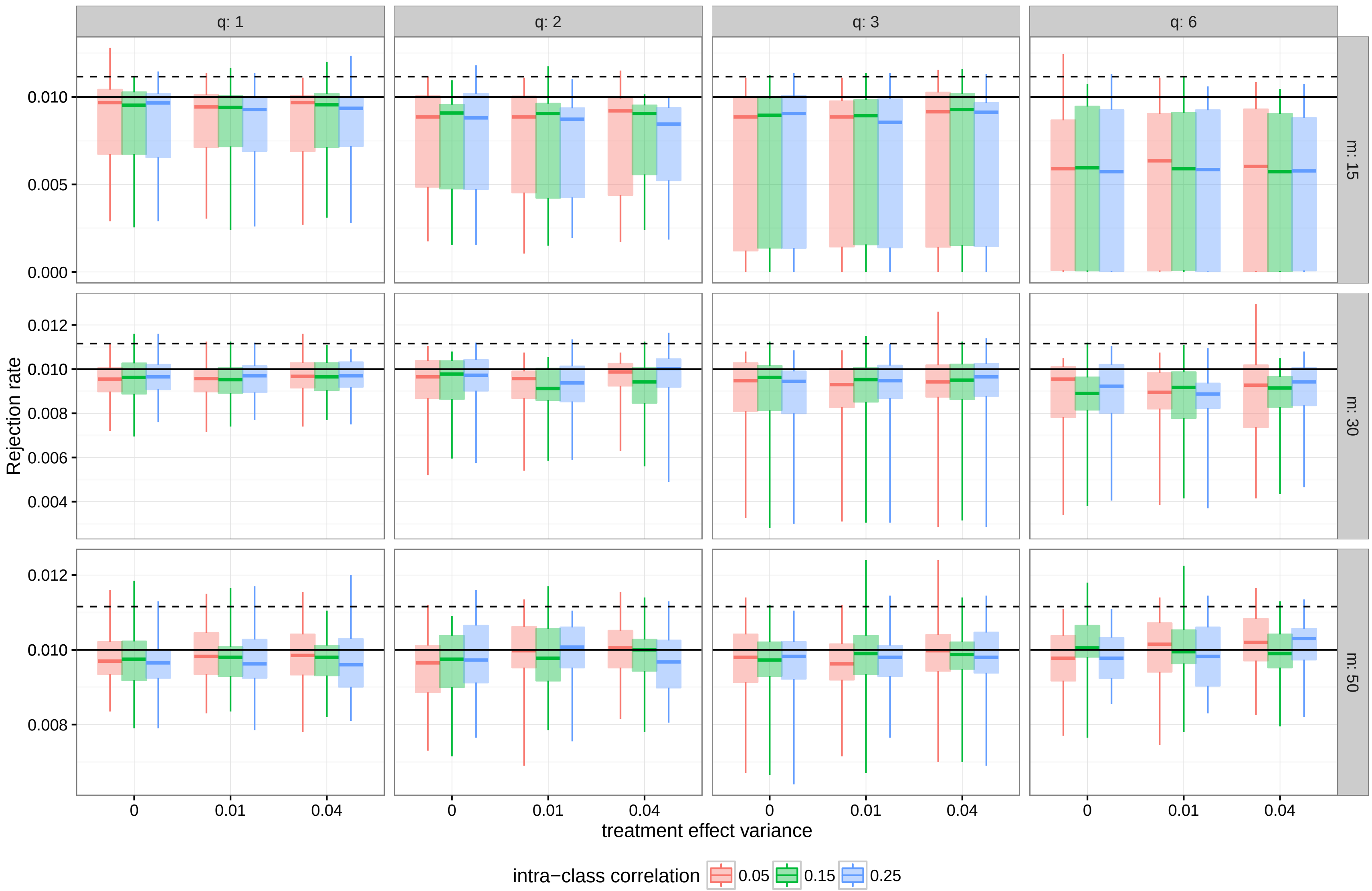

Figure S18: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for $\alpha = .01$.

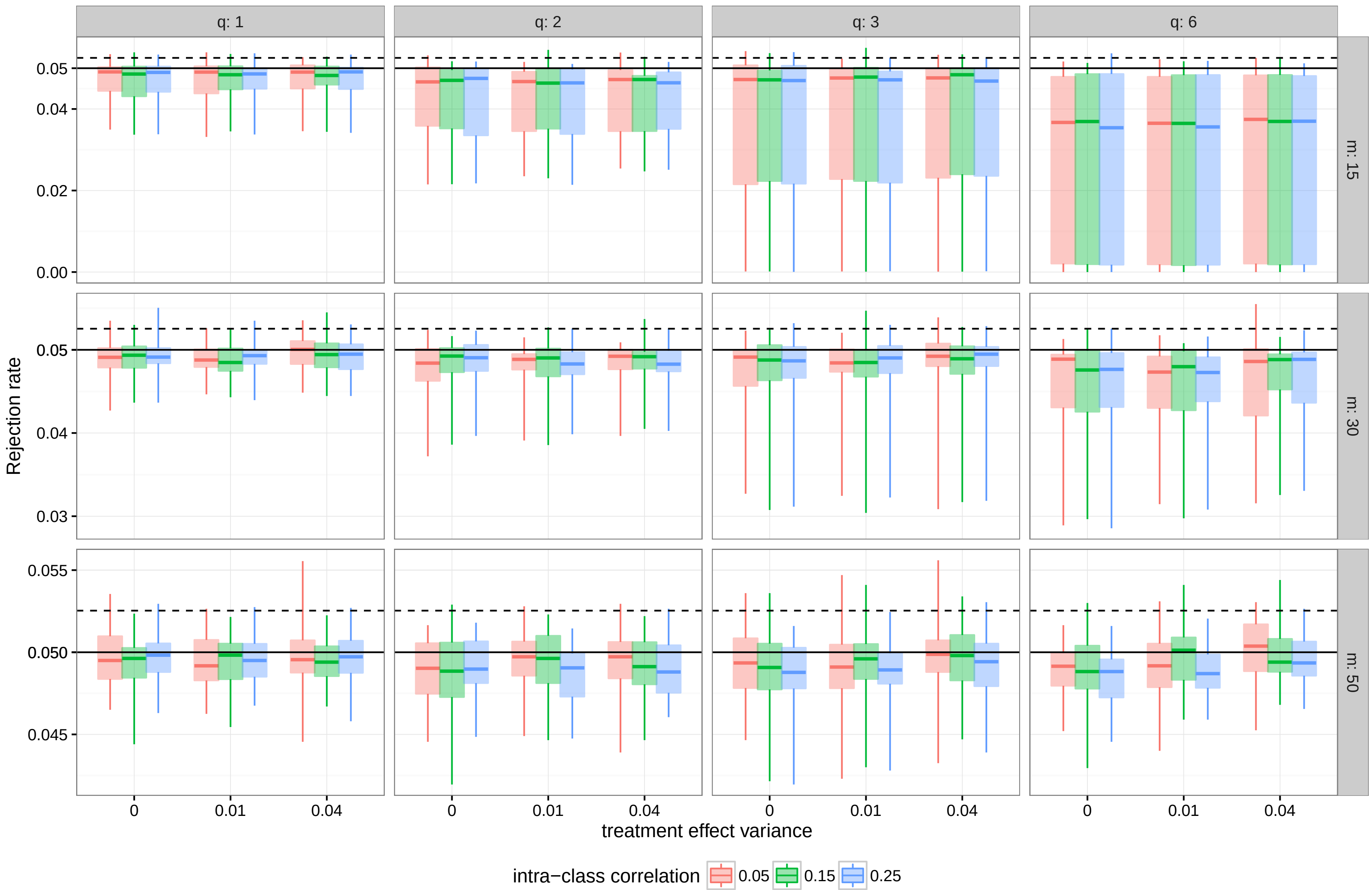

Figure S19: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for $\alpha = .05$.

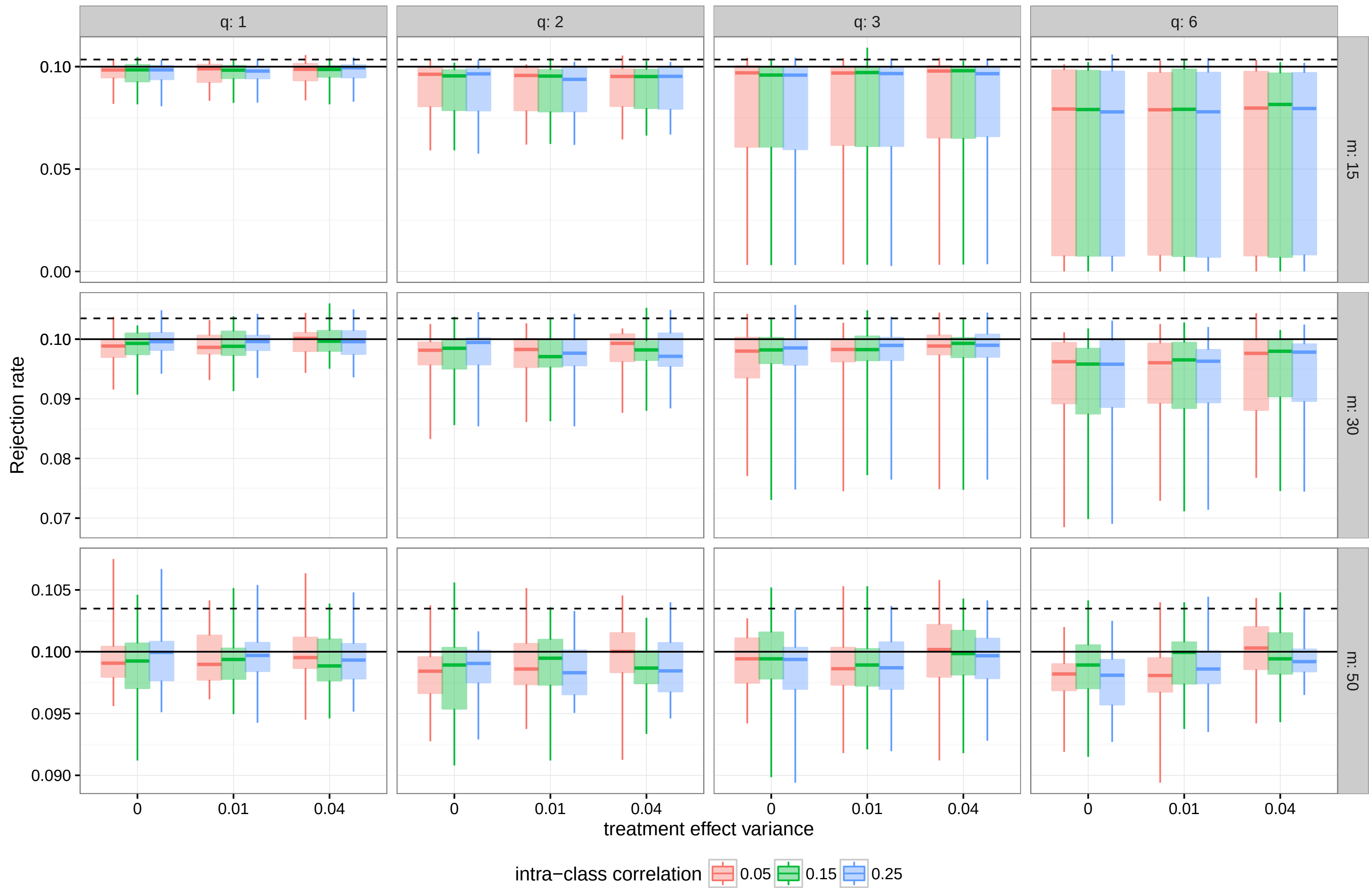

Figure S20: Rejection rates of CR2 AHT test, by treatment effect variance and intra-class correlation for $\alpha = .10$.

References

Henderson, H. V. and Searle, S. R. (1981), 'On deriving the inverse of a sum of matrices', *Siam Review* **23**(1), 53–60.

Imbens, G. W. and Kolesar, M. (2015), Robust Standard Errors in Small Samples: Some Practical Advice. **URL:** *https://www.princeton.edu/ mkolesar/papers/small-robust.pdf*