occurs with probability at most α. So in all three cases, $\tau^a = \tau^b$ or $\tau^a < \tau^b$ or $\tau^a > \tau^b$, we have $\Pr\left(\mathcal{I}_v^- \supseteq \tau_{\min}\right) \geq 1 - \alpha$, proving (ii). ■

By a parallel argument, we obtain analogous $1 - \alpha$ upper intervals, $\mathcal{I}_c^+$ and $\mathcal{I}_v^+$, of the form $(-\infty, \widetilde{\tau})$ for τ if $\tau^a = \tau^b = \tau$ or without restrictions for $\tau_{\max}$. Taking the intersections, $\mathcal{I}_c = \mathcal{I}_c^- \cap \mathcal{I}_c^+$ and $\mathcal{I}_v = \mathcal{I}_v^- \cap \mathcal{I}_v^+$, of two one-sided $1 - \alpha/2$ intervals yields analogous two-sided $1 - \alpha$ intervals for τ if $\tau^a = \tau^b = \tau$ or without restrictions for the interval $[\tau_{\min}, \tau_{\max}]$. In most cases, $\mathcal{I}_v$ can be constructed by taking the union of $\mathcal{I}_c$ and the two two-sided intervals constructed from the matched sets using each separate version of control. When this union is disjoint, $\mathcal{I}_v$ is the shortest interval that contains all three intervals.

In case (ii), the proof above that $\Pr\left(\mathcal{I}_v \supseteq \tau_{\min}\right) \geq 1-\alpha$ is similar to, but not quite identical to, results in Lehmann (1952), Berger (1982) and Laska and Meisner (1989). These authors proposed tests that would invert to yield as a confidence interval the shortest interval $\mathcal{I}_*$ containing $\left\{\tau_0 : P^a_{\tau_0} > \alpha \text{ or } P^b_{\tau_0} > \alpha\right\}$, whereas $\mathcal{I}_v$ is the shortest interval containing $\left\{\tau_0 : P_{\tau_0} > \alpha \text{ or } P^a_{\tau_0} > \alpha \text{ or } P^b_{\tau_0} > \alpha\right\}$, thereby ensuring $\mathcal{I}_v \supseteq \mathcal{I}_c$. Of course, $\mathcal{I}_v \supseteq \mathcal{I}_*$, but unlike $\mathcal{I}_*$, our method ensures that $\mathcal{I}_v$ and $\mathcal{I}_c$ both simultaneously cover $\tau^a = \tau^b = \tau$ at rate $1 - \alpha$ when there is actually only a single version of treatment. Because $\mathcal{I}_c$ is built using all of the data and under stronger assumptions, it is unlikely that $\mathcal{I}_*$ will be much shorter than $\mathcal{I}_v$; however, this logical possibility is the price for reporting the usual interval, $\mathcal{I}_c$, without multiplicity correction.

Why not report a single robust interval like $\mathcal{I}_*$ instead of two intervals, $\mathcal{I}_c$ and $\mathcal{I}_v$, that have slightly more nuanced coverage properties? A simple example may help illustrate the advantage of reporting $\mathcal{I}_c$ and $\mathcal{I}_v$ over $\mathcal{I}_*$. Suppose that $\mathcal{I}_*$ and $\mathcal{I}_v$ both contain zero but $\mathcal{I}_c$ does not. If we choose to report $\mathcal{I}_*$ then we have little additional information to determine why $\mathcal{I}_*$ contains zero – is Fisher's sharp null true or is the smaller of the the two versions of