Supplementary Materials for "Optimal Tests of Treatment Effects for the Overall Population and Two Subpopulations in Randomized Trials, using Sparse Linear Programming"

Table of Contents **Pages**

A Extensions to Other Outcomes, Estimated Variances, and Two-sided Tests

We first generalize to outcomes other than normally distributed outcomes. Similar to Section 3.1, we assume that for each subject i, conditioned on his/her subpopulation $k \in \{1, 2\}$ and study arm assignment $a \in \{0, 1\}$, his/her data $Y_{ka,i}$ is a random draw from an unknown distribution Q_{ka}, and that this draw is independent of the data of all the other subjects. Instead of restricting each Q_{ka} to be a normal distribution as in Section 3.1, we allow Q_{ka} to be any distribution on $\mathbb{R}$. This allows, for example, the outcome to be binary valued, count valued, or continuous valued. Define μ_{ka} and σ_{ka}^2 to be the mean and variance, respectively, of Q_{ka}, which we assume to be finite. Let $\Delta_k = \mu_{k1} - \mu_{k0}$ for each $k \in \{1, 2\}$. The null hypotheses are as in Section 3.1 except using the above definition of Δ_k. The z-statistics are as in Section 3.1, except using the outcomes $Y_{ka,i}$ which are distributed as Q_{ka}. The setup in Section 3.1 is then a special case of the above setup, if we let each Q_{ka} be a normal distribution with mean μ_{ka} and variance σ_{ka}^2.

We assume each $\sigma_{ka}^2 > 0$. Then by the multivariate central limit theorem, the joint distribution of $(Z_1 - EZ_1, Z_2 - EZ_2, Z_C - EZ_C)$ converges to a zero mean, multivariate normal distribution with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho_1 \\ 0 & 1 & \rho_2 \\ \rho_1 & \rho_2 & 1 \end{pmatrix},$$

as sample size n goes to infinity (holding p_1, p_2 and each Q_{ka} fixed). The covariance matrix Σ is the same as that for the case of normally distributed outcomes in Section 3.1. Therefore, the optimal multiple testing procedures constructed above for normally distributed outcomes can be expected to perform similarly for the more general case above, at large enough sample sizes. By a similar argument, when replacing variances by sample variances in the definition of the z-statistics, one expects the same to hold at large sample sizes. The sample size at which the normal approximation is a good one depends on features of the data generating distributions Q_{ka}. It is an area of future work to explore the impact of, e.g., skewed and heavy-tailed distributions Q_{ka} on the performance of our multiple testing procedures.

A more formal argument would require consideration of local alternatives, that is, sequences of data generating distributions $Q_{ka}^{(n)}$ with Δ_k of order $1/\sqrt{n}$; this is because at fixed alternatives Q_{ka} with $\Delta_k \neq 0$, the absolute value of the non-centrality parameter δ_k converges to infinity and so all reasonable procedures have power converging to 1. Extending our results under local alternatives is an area of future work.

We now consider null hypotheses related to two-sided tests. For each $k \in \{1, 2\}$, define H'_{0k} to be the null hypothesis $\Delta_k = 0$, i.e., that treatment is equally as effective as control, on average, for subpopulation k; define H'_{0C} to be the null hypothesis $p_1\Delta_1 + p_2\Delta_2 = 0$, i.e., that treatment is equally as effective as control, on average, for the combined population. Our general method can be applied to these hypotheses. The main changes would be that one may decide to specify a power constraint as in (4) not only at $(\delta_1^{\min}, \delta_2^{\min})$, but also at $(-\delta_1^{\min}, -\delta_2^{\min})$; one may also decide to assign weight in the prior to alternatives (δ_1, δ_2) with δ_1 and/or δ_2 negative. We note that the boundaries of the null hypotheses $\mathcal{H}$ and the boundaries of the null hypotheses $\{H'_{01}, H'_{02}, H'_{0C}\}$ are identical; therefore, it may be appropriate to set discretized constraints $G' \subseteq G$.

B Extending the Rejection Regions of a Solution to the Discretized Problem

We restricted to the class of multiple testing procedures $\mathcal{M}_B \subset \mathcal{M}$ that reject no hypotheses outside the region $B = [-b, b] \times [-b, b]$ for a fixed integer $b > 0$. The reason was to make our solution computationally feasible. We now show how to iteratively augment the optimal solution to the discretized problem within the class $\mathcal{M}_{\mathcal{R}}$, which we denote by M_B^*, to allow rejection of null hypotheses outside $B = [-b, b] \times [-b, b]$. Let $B' = [-b', b'] \times [-b', b']$ for an integer $b' > b$. Define $\mathcal{R}' = \{R_{k,k'} : k, k' \in \mathbb{Z}, R_{k,k'} \subset B' \setminus B\}$. Let $\mathcal{M}'$ denote the class of multiple testing procedures $M \in \mathcal{M}_{B'}$ such that for any $u \in [0,1]$, we have (i) $M(z_1, z_2, u) = M^*_B(z_1, z_2, u)$ for any $(z_1, z_2) \in B$, and (ii) for any rectangle $r \in \mathcal{R}'$ we have $M(z_1, z_2, u) = M(z_1', z_2', u)$ whenever (z_1, z_2) and (z_1', z_2') are both in r. This can be expressed as a sparse linear program and can be solved using the algorithm in Section 9. This can be iterated for a sequence of increasing values b'. However, in the examples from Section 5.1 there is little room for improving the Bayes risk by such an extension, since as shown in Section 6, the Bayes risk of the optimal solution in $\mathcal{M}_{\mathcal{R}} \subseteq \mathcal{M}_B$ is within 0.005 of the optimal Bayes risk over the class of procedures $\mathcal{M}$ for the original problem.

C Discretization of Familywise Type I Error and Power Constraints

We show that when restricting to the discretized procedures $\mathcal{M}_{\mathcal{R}}$ and familywise Type I error constraints in G', the constraints (3) and (4) in the constrained Bayes optimization problem are the linear functions of $\mathbf{m}$ given in (11) and (12), respectively. Each familywise Type I error constraint (3) can be expressed as

$$
\begin{aligned}
& P_{\delta_1,\delta_2}[M \text{ rejects any null hypothesis in } \mathcal{H}_{\mathrm{TRUE}}(\delta_1,\delta_2)] \\
= & \sum_{r\in\mathcal{R}} P_{\delta_1,\delta_2}[(Z_1,Z_2)\in r] P[M \text{ rejects any null hypothesis in } \mathcal{H}_{\mathrm{TRUE}}(\delta_1,\delta_2) | (Z_1,Z_2)\in r] \\
= & \sum_{r\in\mathcal{R}} P_{\delta_1,\delta_2}[(Z_1,Z_2)\in r] \sum_{s\in\mathcal{S}: s\cap\mathcal{H}_{\mathrm{TRUE}}(\delta_1,\delta_2)\neq\emptyset} m_{rs} \leq \alpha,
\end{aligned}
$$

where the last line equals (11). A similar argument shows the power constraint (4) can be expressed as the linear function of $\mathbf{m}$ given in (12).

D Additional Examples of Priors and Loss Functions

We present several examples using the setup from Section 5, but with different priors and loss functions. Define prior $\Lambda' = \sum_{j=1}^{4} w'_j \lambda'_j$, where $\mathbf{w}' = (w'_1, w'_2, w'_3, w'_4)$ is a weight vector and $\lambda'_1, \lambda'_2, \lambda'_3, \lambda'_4$ are bivariate normal distributions with mean vectors $(0,0), (\delta_1^{\min}, 0), (0, \delta_2^{\min})$, and $(\delta_1^{\min}, \delta_2^{\min})$, respectively, and each with covariance matrix having diagonal entries $((\delta_1^{\min}/2)^2, (\delta_2^{\min}/2)^2)$ and zero off-diagonal entries. Define the modified symmetric case to be as in Section 5, except with prior Λ' under weight vector $\mathbf{w}' = \mathbf{w}^{(1)} = (1/4, 1/4, 1/4, 1/4)$; similarly, define the modified asymmetric case as in Section 5 except with prior Λ' under weight vector $\mathbf{w}' = \mathbf{w}^{(2)} = (0.2, 0.35, 0.1, 0.35)$.

The solutions in each case at $1 - \beta = 0.88$ are given in Figures 5a and 5b. We used the lower level of discretization $\tau = (0.1, 0.1)$ to highlight an interesting phenomenon. The yellow dots, which occur sporadically on the boundaries between rejection regions, are where the optimal procedure is a randomized procedure that rejects either the set of null hypotheses on one or the other side of the corresponding boundary, with probabilities that sum to 1; this is an artifact of the discretization and diminishes when a finer discretization is used on these boundaries.

Define the following loss function where the penalty for failing to reject each subpopulation null hypothesis H_{0k} when $\Delta_k \geq \Delta^{\min}$ is proportional to Δ_k:

$$\tilde{L}'(s; \delta_1, \delta_2) = \sum_{k=1}^{2} \min\{\delta_k, \bar{b}\} 1[\delta_k \geq \delta_k^{\min}, H_{0k} \notin s],$$

for $\bar{b} = 10$; here $\bar{b}$ denotes the maximum penalty allowed, which we incorporate so that the loss function is bounded. The optimal rejection regions in the modified symmetric and asymmetric cases, except using $\tilde{L}'$ in place of $\tilde{L}$, are given in Figures 5c and 5d. There is little difference between the optimal rejection regions under $\tilde{L}'$ versus under $\tilde{L}$.

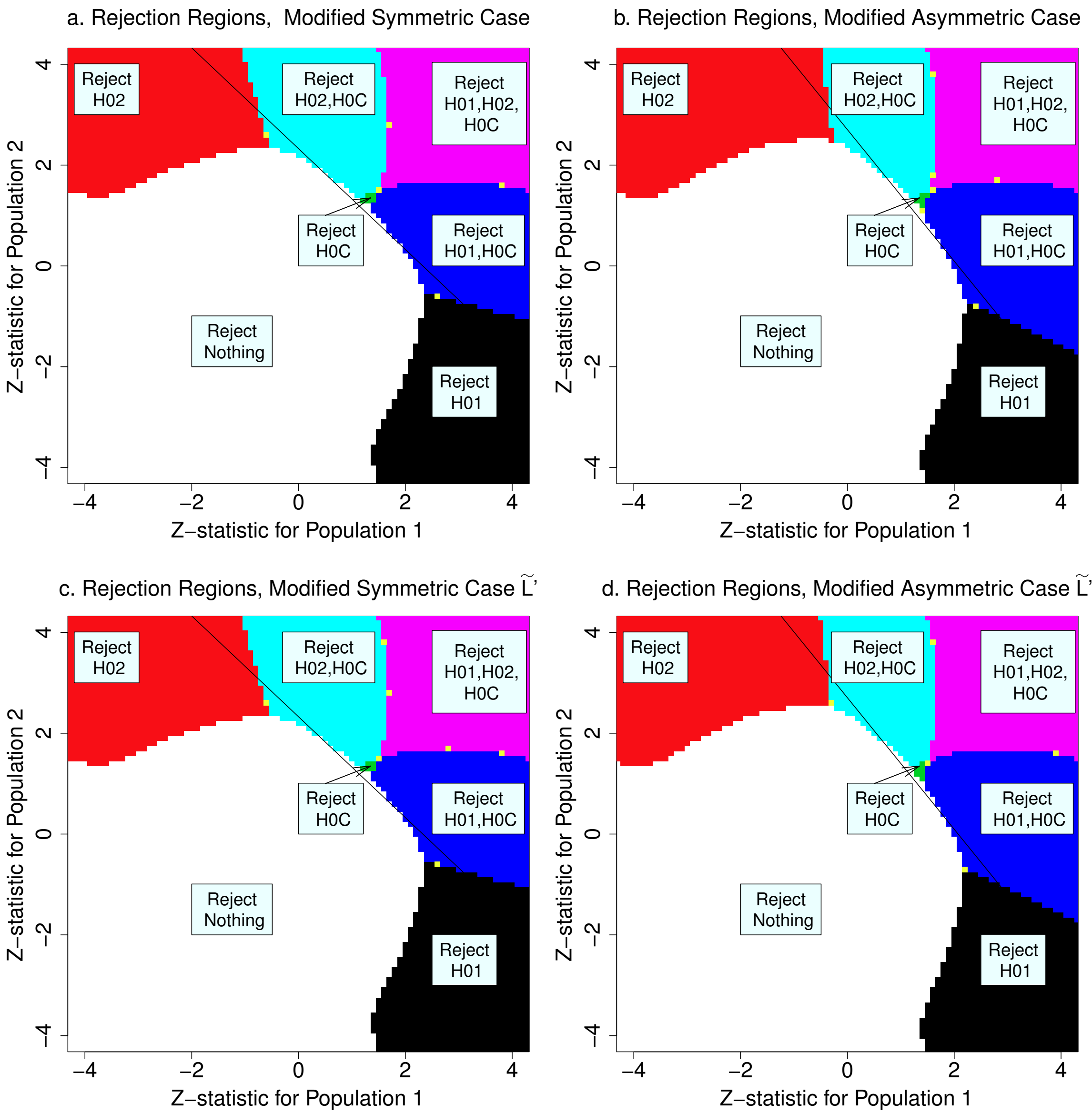

Figure 5: Optimal multiple testing procedures under prior Λ', for (a) the modified symmetric case and (b) the modified asymmetric case, at coarse discretization $\tau = (0.1, 0.1)$. Plots (c) and (d) are analogs of (a) and (b), respectively, except that the loss function $\tilde{L}'$ is used in place of $\tilde{L}$. In each plot, the black line is the boundary of R_{UMP}. Yellow dots are where the optimal procedure is a randomized procedure that rejects either the set of null hypotheses on one or the other side of the boundary, with probabilities that sum to 1.

E Definitions of Multiple Testing Procedures from Prior Work in Section 5.3

E.1 Procedure of Bergmann and Hommel (1988)

As described by Hommel and Bernhard (1999), the procedure of Bergmann and Hommel (1988) involves first specifying which subsets of elementary null hypotheses are "exhaustive." For any index set $J \subseteq \{1, 2, C\}$, the subset $\{H_{0j}, j \in J\}$ is defined to be exhaustive if there exists a data generating distribution under which all and only the null hypotheses in this subset are true. In our problem, all subsets are exhaustive except $\{H_{01}, H_{02}\}$ and the singleton $\{H_{0C}\}$, since whenever H_{01}, H_{02} are both true also H_{0C} is true, and whenever H_{0C} is true at least one of H_{01}, H_{02} is true. The procedure of Bergmann and Hommel (1988), when applied to our set of null hypotheses $\mathcal{H}$, rejects the null hypotheses with indices $\{1, 2, C\} \setminus A$, where A is defined as the union of all subsets $J \subseteq \{1, 2, C\}$ that satisfy:

$$\{H_{0j}, j \in J\} \text{ is exhaustive and } \max\{Z_j : j \in J\} < \Phi^{-1}(1 - 0.05/|J|).$$

E.2 Procedure based on Song and Chi (2007) and Alosh and Huque (2009)

Song and Chi (2007) and Alosh and Huque (2009) designed multiple testing procedures involving the overall population and a single, prespecified subpopulation, which we refer to as subpopulation 1. Here, in contrast, we are interested in the larger family of hypotheses including that for subpopulation 2. To tailor the procedure of Song and Chi (2007) to our context, we augment it to additionally allow rejection of H_{02}, without any loss in power for the overall population or for subpopulation 1, and while maintaining strong control of the familywise Type I error rate. We denote the augmented procedure by M^{SC}, which, for prespecified thresholds $\alpha_0, \alpha_1, \alpha_2$ satisfying $0 \leq \alpha_0 < 0.05 < \alpha_1 \leq 1$, and $0 \leq \alpha_2 \leq 1$, is defined as follows:

If $Z_C > \Phi^{-1}(1-\alpha_0)$, reject H_{0C} as well as each subpopulation null hypothesis H_{0k}, $k \in \{1,2\}$, for which $Z_k > \Phi^{-1}(1-0.05)$. If $\Phi^{-1}(1-\alpha_0) \geq Z_C > \Phi^{-1}(1-\alpha_1)$ and $Z_1 > \Phi^{-1}(1-\alpha_2)$, then reject H_{01}, and if in addition $Z_C > \Phi^{-1}(1-0.05)$ then reject H_{0C}.

The original procedure of Song and Chi (2007) is the same as above except it does not allow rejection of H_{02}, since it was designed in the context of testing only H_{0C} and H_{01}. Their procedure has similar performance to a procedure of Alosh and Huque (2009), so we only include the former in our comparison in Section 5.3. We chose $\alpha_0 = 0.045$ and $\alpha_1 = 0.1$, which is used in an example of Song and Chi (2007). We then used the method of Song and Chi (2007) to compute, for the symmetric case, the largest α_2 (which depends on p_1) such that the above procedure strongly controls the familywise Type I error rate at level 0.05.

We next prove M^{SC} strongly controls the familywise Type I error rate at level 0.05, using the closed testing principle of Marcus et al. (1976). We first define local tests of each intersection null hypothesis in $\mathcal{H}$. The local test of $H_{01} \cap H_{0C}$ is as in Song and Chi (2007), that is, the test that rejects if $Z_C > \Phi^{-1}(1-\alpha_0)$, or if both $\Phi^{-1}(1-\alpha_0) \geq Z_C > \Phi^{-1}(1-\alpha_1)$ and $Z_1 > \Phi^{-1}(1-\alpha_2)$. This is shown to have Type I error at most 0.05 by Song and Chi (2007). We use the same local test for $H_{01} \cap H_{02}$ as just given for $H_{01} \cap H_{0C}$. It follows that this test has Type I error at most 0.05, since $H_{01} \cap H_{02} \subseteq H_{01} \cap H_{0C}$. We set the local test for $H_{02} \cap H_{0C}$ to reject if $Z_C > \Phi^{-1}(1-0.05)$; it follows immediately that this has Type I error at most 0.05. We set the local test of H_{0C} to reject if $Z_C > \Phi^{-1}(1-0.05)$. For subpopulation $k \in \{1,2\}$, we set the local test of each H_{0k} to reject if $Z_k > \Phi^{-1}(1-0.05)$. Applying the closed testing principle, with the above local tests, results in the procedure M^{SC}. By the closed testing principle, this procedure strongly controls the familywise Type I error rate at level 0.05.

F Optimal Solutions for Discretized Problem for Sample Size Greater than $n_{\min}$

We present the rejection regions for the optimal solution $\mathbf{m}^*_{\text{SS}}(n)$ to the constrained Bayes optimization problem from Section 5.3, for $n = 1.03n_{\min}$ and $n = 1.06n_{\min}$, in Figure 6. Recall that at $n = n_{\min}$, we have $\mathbf{m}^*_{\text{SS}}(n) = \mathbf{m}^*_{\text{sym}}(0.9)$, given in Figure 1a.

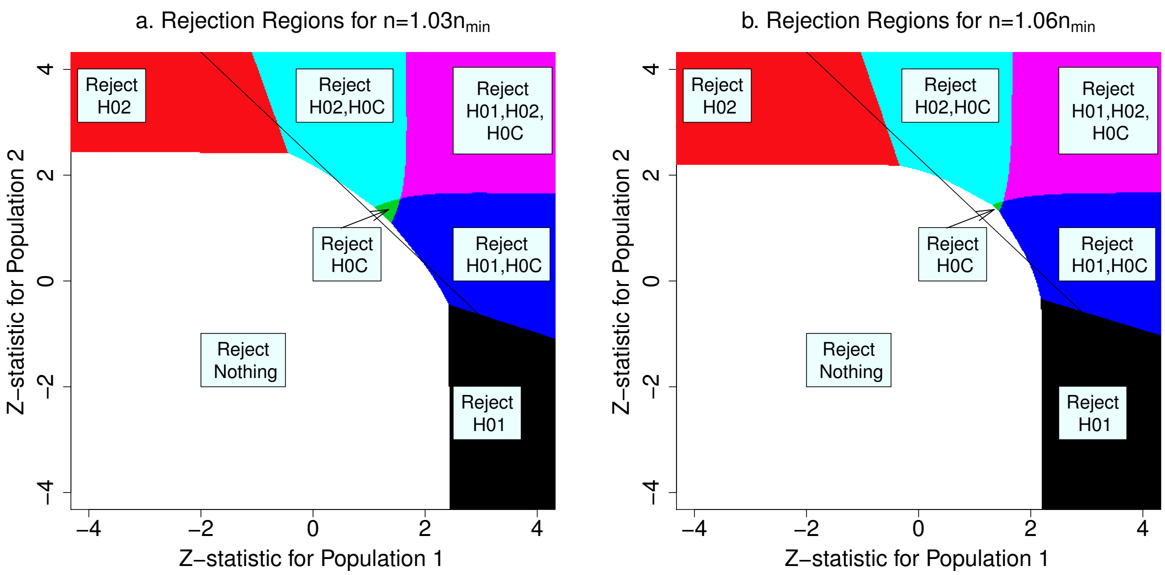

Figure 6: Rejection regions for $\mathbf{m}^*_{\text{SS}}(n)$, for symmetric case, i.e., $p_1 = p_2 = 1/2$ and prior Λ_1, (a) for $n = 1.03n_{\min}$ and (b) for $n = 1.06n_{\min}$.

We next consider the minimum sample size required to achieve a desired power x to detect treatment effects in each subpopulation, while maintaining 90% power for H_{0C} and strongly controlling the familywise Type I error rate. Consider the case where $p_1 = p_2 = 1/2$, with $n_{\min}$ as defined in Section 5.1 at $1 - \beta = 0.9$. For each value of x on the horizontal axis, Figure 7 plots the minimum value of $n/n_{\min}$ such that at sample size n there exists a multiple testing procedure with power at least x for H_{01} at $(\delta_1^{\min}, 0)$, power at least x for H_{02} at $(0, \delta_2^{\min})$, power at least 0.9 for H_{0C} at $(\delta_1^{\min}, \delta_2^{\min})$, and that strongly controls the familywise Type I error rate at level 0.05. These were computed by solving the constrained Bayes optimization problem at a sequence of values $n/n_{\min}$ using the prior Λ defined in Section 5.1

with weight vector $\mathbf{w} = (0, 1/2, 1/2, 0)$; this prior puts weight only on the alternatives $(\delta_1^{\min}, 0)$ and $(0, \delta_2^{\min})$. The solution to the optimization problem in each case was observed to have power for H_{01} at $(\delta_1^{\min}, 0)$ equal to the power for H_{02} at $(0, \delta_2^{\min})$. It follows that at each value of $n/n_{\min}$ we considered, no multiple testing procedure can have greater power than our optimal solution to reject H_{01} at $(\delta_1^{\min}, 0)$ and to reject H_{02} at $(\delta_1^{\min}, 0)$, while satisfying the power constraint (4) at $1 - \beta = 0.9$ and the familywise Type I error constraints (3).

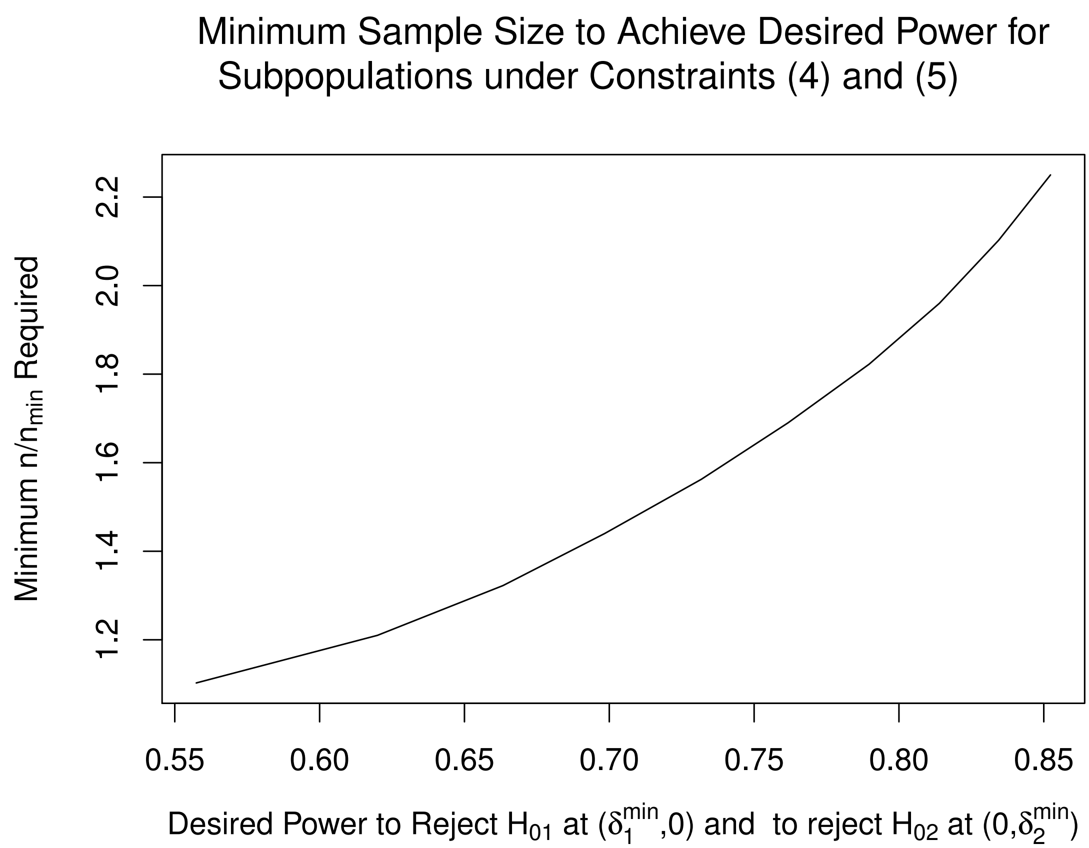

Figure 7: Minimum value of $n/n_{\min}$ to achieve a desired power for both H_{01} and H_{02}, under power constraint (4) at $1 - \beta = 0.9$ and the familywise Type I error constraints (3).

G Using Only the Familywise Type I Error Constraint at the Global Null

Consider what would happen if we only impose the familywise Type I error constraint at the global null hypothesis. We also impose the power constraint (4). The optimal solutions to the corresponding discretized problem at $1 - \beta = 0.88$ are shown in Figure 8, for the symmetric and asymmetric cases from Section 5. All the null hypotheses $\mathcal{H}$ are true at

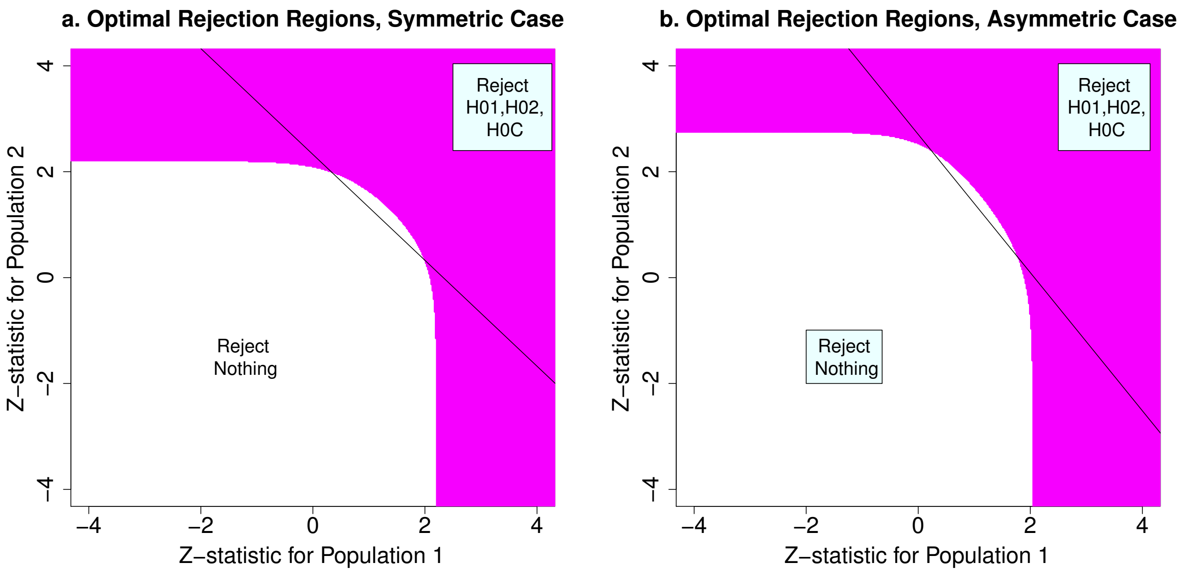

Figure 8: Optimal rejection regions for the discretized problem omitting all familywise Type I error constraints except at the global null hypothesis in (a) symmetric case and (b) the asymmetric case. In each plot, the black line is the boundary of R_{UMP}, which depends on p_1 so differs in the two plots.

the global null hypothesis; therefore a familywise Type I error occurs under the global null hypothesis if any nonempty subset of $\mathcal{H}$ is rejected. Since the loss function $\tilde{L}$ penalizes for failure to reject null hypotheses, it is optimal, at any realization of (Z_1, Z_2), to either reject no null hypothesis or reject all the null hypotheses, as is the case in Figure 8. The familywise Type I error rate at $(\delta_1^{\min}, 0)$ and at $(0, \delta_2^{\min})$ both equal 0.54 in the symmetric case, and equal 0.69 and 0.32, respectively, in the asymmetric case. This shows the importance of including in G' a close approximation to each active familywise Type I error constraint.

H Verifying the Optimal Solution to the Discretized Problem Satisfies All Constraints of the Original Problem

Consider the set of multiple testing procedures $\mathcal{M}^*$ defined in Section 5.2, which are the solutions to the discretized problems in Section 5.1. By definition, each $M^* \in \mathcal{M}^*$ satisfies the familywise Type I error constraints (11) of the corresponding discretized problem. We now verify each $M^* \in \mathcal{M}^*$ satisfies all constraints (3) of the original problem.

In Section H.1, we prove Theorem 1. This will be used to reduce the problem of verifying the constraints (3) for each $(\delta_1, \delta_2) \in \mathbb{R}^2$, to the more manageable problem of verifying these constraints for each $(\delta_1, \delta_2) \in G$ (the boundaries of the null spaces). To solve the latter problem, we use a combination of a grid search and an analytic bound on its approximation error, as shown in Sections H.2 and H.3.

To upper bound (15) by 0.05 using the above approach, it was necessary to solve the discretized linear programs defined in Section 5.1 at $\alpha = 0.05 - 10^{-4}$. The solution to each discretized problem that we reported in the paper (e.g., those given in Section 5.1) was computed at $\alpha = 0.05 - 10^{-4}$. To show this reduction from 0.05 had a negligible effect on the Bayes risk of the resulting procedures, for each of the four discretized problems in Section 5.1, we computed the Bayes risk of the optimal solution at $\alpha = 0.05 - 10^{-4}$ and at $\alpha = 0.05$. For each problem, the difference in the Bayes risk between these two solutions was never more than 0.001.

H.1 Proof of Theorem 1

For any $M \in \mathcal{M}_{det}$, denote the familywise Type I error rate of M at a given (δ_1, δ_2) by $F_M(\delta_1, \delta_2) = P_{\delta_1,\delta_2}[M(Z_1, Z_2) \cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1, \delta_2) \neq \emptyset]$. We first prove part (a) of Theorem 1, where we set $R = B$. Consider any $M \in \mathcal{M}_{det} \cap \mathcal{M}_B$ that satisfies monotonicity conditions (a)-(d) with respect to $R = B$. We will prove $\sup_{(\delta_1,\delta_2)\in\mathbb{R}^2} F_M(\delta_1, \delta_2) = \sup_{(\delta_1,\delta_2)\in G} F_M(\delta_1, \delta_2)$.

To show this, for any given $(\delta_1, \delta_2) \in \mathbb{R}^2$, we will exhibit a point $(\delta_1', \delta_2') \in G$ for which $F_M(\delta_1, \delta_2) \leq F_M(\delta_1', \delta_2')$. We consider 5 cases, each corresponding to (δ_1, δ_2) being in a different region of $\mathbb{R}^2$. Define $\eta(x_1, x_2) = \exp\left\{-x_1^2/2 - x_2^2/2\right\}/(2\pi)$. We use the notation $1[\cdot]$ for indicator variables, as defined in Section 5.1.

Case 1: $\delta_1 > 0$ and $\delta_2 > 0$. Then we have $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \emptyset$, and so $F_M(\delta_1, \delta_2) = 0$. For this case, define $(\delta_1', \delta_2') = (0, 0) \in G$. Then $F_M(\delta_1, \delta_2) = 0 \leq F_M(\delta_1', \delta_2')$. This completes Case 1.

In each of Cases 2-5, for a given (δ_1, δ_2), we will define a point $(\delta_1', \delta_2') \in G$ such that $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \mathcal{H}_{\text{TRUE}}(\delta_1', \delta_2')$. We then define subsets A and A' of R (each of which depends on δ_1, δ_2, but we suppress this for notational clarity), and prove the following:

$$\begin{aligned}
& \int_{(z_1,z_2)\in(A\cup A')} 1[M(z_1, z_2) \cap \mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) \neq \emptyset]\eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
= & \int_{(z_1,z_2)\in(A\cup A')} 1[M(z_1, z_2) \cap \mathcal{H}_{\text{TRUE}}(\delta_1', \delta_2') \neq \emptyset]\eta(z_1 - \delta_1', z_2 - \delta_2')dz_1\, dz_2, \qquad (19)
\end{aligned}$$

and

$$\begin{aligned}
& \int_{(z_1,z_2)\in\mathbb{R}^2\setminus(A\cup A')} 1[M(z_1, z_2) \cap \mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) \neq \emptyset]\eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
\leq & \int_{(z_1,z_2)\in\mathbb{R}^2\setminus(A\cup A')} 1[M(z_1, z_2) \cap \mathcal{H}_{\text{TRUE}}(\delta_1', \delta_2') \neq \emptyset]\eta(z_1 - \delta_1', z_2 - \delta_2')dz_1\, dz_2. \qquad (20)
\end{aligned}$$

Once (19) and (20) are shown, it follows that

$$\begin{aligned}
F_M(\delta_1, \delta_2) &= \int_{(z_1,z_2)\in(A\cup A')} 1[M(z_1,z_2)\cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1,\delta_2)\neq\emptyset]\eta(z_1-\delta_1, z_2-\delta_2)dz_1\,dz_2 \quad (21)\\
&\quad + \int_{(z_1,z_2)\in\mathbb{R}^2\setminus(A\cup A')} 1[M(z_1,z_2)\cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1,\delta_2)\neq\emptyset]\eta(z_1-\delta_1, z_2-\delta_2)dz_1\,dz_2\\
&\leq \int_{(z_1,z_2)\in(A\cup A')} 1[M(z_1,z_2)\cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1',\delta_2')\neq\emptyset]\eta(z_1-\delta_1', z_2-\delta_2')dz_1\,dz_2\\
&\quad + \int_{(z_1,z_2)\in\mathbb{R}^2\setminus(A\cup A')} 1[M(z_1,z_2)\cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1',\delta_2')\neq\emptyset]\eta(z_1-\delta_1', z_2-\delta_2')dz_1\,dz_2\\
&= F_M(\delta_1', \delta_2'). \quad (22)
\end{aligned}$$

Case 2: $\delta_1 \leq 0$, $\delta_2 > 0$, and $\rho_1\delta_1 + \rho_2\delta_2 > 0$. For this case, define $(\delta_1', \delta_2') = (0, \delta_2) \in G$. Then we have $\mathcal{H}_{\mathrm{TRUE}}(\delta_1,\delta_2) = \mathcal{H}_{\mathrm{TRUE}}(\delta_1',\delta_2') = \{H_{01}\}$. Define the vertical line $D = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 = \delta_1/2\}$, which is the perpendicular bisector of the line segment between (δ_1, δ_2) and (δ_1', δ_2'). Define the function $h(z_1, z_2) = (h_1(z_1, z_2), h_2(z_1, z_2)) = (-z_1 + \delta_1, z_2)$, which maps any point $(z_1, z_2) \in \mathbb{R}^2$ to its reflection across the line D.

Define $A = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 < \delta_1/2 \text{ and } H_{01} \in M(z_1, z_2)\}$, and let $A' = h(A) = \{h(z_1, z_2) : (z_1, z_2) \in A\}$, i.e., the reflection of the set A across the line D. It follows that A and A' are disjoint sets, and h is a bijection from A onto A'.

Since we assumed $M \in \mathcal{M}_B$, we have $A \subseteq B$. By this and the definition of A, we have $A \subseteq [((-\infty, \delta_1/2) \times \mathbb{R}) \cap B]$. We next show $A' \subseteq B$, by considering two subcases. The first subcase is where $\delta_1 \leq -2b$. Then $[((-\infty, \delta_1/2) \times \mathbb{R}) \cap B] = \emptyset$, which implies $A = \emptyset$. This implies $A' = \emptyset$ and so $A' \subseteq B$. Consider the remaining subcase where $\delta_1 > -2b$. For any $(z_1, z_2) \in A$, we have $-b \leq z_1 < \delta_1/2$, which implies $-z_1 + \delta_1 > -\delta_1/2 + \delta_1 = \delta_1/2$. Therefore $-b < \delta_1/2 < -z_1 + \delta_1 \leq b + \delta_1 \leq b$ (where the last inequality follows from the assumption above that $\delta_1 \leq 0$). Therefore, $h(z_1, z_2) = (-z_1 + \delta_1, z_2) \in B$. This shows $A' \subseteq B$ in this subcase.

We next show that monotonicity property (a) implies for any $(z_1', z_2') \in A'$, we have $H_{01} \in M(z_1', z_2')$. Consider any $(z_1', z_2') \in A'$. Since we showed above that $A' \subseteq B$, we have $(z_1', z_2') \in A' \subseteq B$. By the definition of A', there exists a point $(z_1, z_2) \in A$ for which $h(z_1, z_2) = (z_1', z_2')$. By the definition of the set A, $z_1 < \delta_1/2$ and $H_{01} \in M(z_1, z_2)$. This implies $\delta_1 - 2z_1 \geq 0$. Since $(z_1', z_2') = h(z_1, z_2) = (z_1 + (\delta_1 - 2z_1), z_2)$, monotonicity property (a) implies $H_{01} \in M(z_1 + (\delta_1 - 2z_1), z_2) = M(z_1', z_2')$, as desired.

We next show the equality (19). Since $\mathcal{H}_{\mathrm{TRUE}}(\delta_1, \delta_2) = \mathcal{H}_{\mathrm{TRUE}}(\delta_1', \delta_2') = \{H_{01}\}$ (for this case), we have $1[M(z_1, z_2) \cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1, \delta_2) \neq \emptyset] = 1[H_{01} \in M(z_1, z_2)]$. This implies the integral on the left side of the equality in (19) equals

$$
\begin{aligned}
& \int_{(z_1,z_2)\in A\cup A'} 1[H_{01} \in M(z_1, z_2)]\eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
= & \int_{(z_1,z_2)\in A\cup A'} \eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 && (23)\\
= & \int_{(z_1',z_2')\in h^{-1}(A\cup A')} \eta(h_1(z_1', z_2') - \delta_1, h_2(z_1', z_2') - \delta_2)dz_1'\, dz_2' && (24)\\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(-z_1', z_2' - \delta_2)dz_1'\, dz_2' && (25)\\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(-(z_1' - \delta_1'), z_2' - \delta_2')dz_1'\, dz_2' && (26)\\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(z_1' - \delta_1', z_2' - \delta_2')dz_1'\, dz_2', && (27)\\
= & \int_{(z_1',z_2')\in A\cup A'} 1[H_{01} \in M(z_1', z_2')]\eta(z_1' - \delta_1', z_2' - \delta_2')dz_1'\, dz_2', && (28)
\end{aligned}
$$

where (23) follows from $H_{01} \in M(z_1, z_2)$ for any $(z_1, z_2) \in A \cup A'$; (24) follows by the change of variables $(z_1', z_2') = h^{-1}(z_1, z_2)$, for which the absolute value of the Jacobian determinant equals 1; (25) follows from $h^{-1}(A \cup A') = A \cup A'$ and the definition of h; (26) follows from the definition of (δ_1', δ_2'); (27) follows from $\eta(-x, y) = \eta(x, y)$ for any $(x, y) \in \mathbb{R}^2$; and, (28) follows from $H_{01} \in M(z_1', z_2')$ for any $(z_1', z_2') \in A \cup A'$. This proves (19) for this case.

To show (20), consider any $(z_1, z_2) \in \mathbb{R}^2 \setminus A$ such that $H_{01} \in M(z_1, z_2)$. Then we have $z_1 - \delta_1/2 \geq 0$, and after multiplying both sides by $-2\delta_1$ (which is nonnegative since $\delta_1 \leq 0$), we have $-2z_1\delta_1 + \delta_1^2 \geq 0$. Therefore, $(z_1 - \delta_1')^2 = z_1^2 \leq z_1^2 - 2z_1\delta_1 + \delta_1^2 = (z_1 - \delta_1)^2$, which implies $\eta(z_1 - \delta_1, z_2 - \delta_2) \leq \eta(z_1 - \delta_1', z_2 - \delta_2')$. We have shown for any $(z_1, z_2) \in \mathbb{R}^2 \setminus A$,

$$1[H_{01} \in M(z_1, z_2)]\eta(z_1 - \delta_1, z_2 - \delta_2) \leq 1[H_{01} \in M(z_1, z_2)]\eta(z_1 - \delta_1', z_2 - \delta_2'). \tag{29}$$

This implies (20). Having shown (19) and (20), the above argument (21)-(22) implies $F_M(\delta_1, \delta_2) \leq F_M(\delta_1', \delta_2')$, which completes Case 2.

Case 3: $\delta_1 \leq 0$, $\delta_2 > 0$ and $\rho_1\delta_1 + \rho_2\delta_2 \leq 0$. This implies $\delta_1 < 0$. For this case, define $(\delta_1', \delta_2') = (-\rho_2\delta_2/\rho_1, \delta_2)$. Then we have $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \mathcal{H}_{\text{TRUE}}(\delta_1', \delta_2') = \{H_{01}, H_{0C}\}$. Define the vertical line $D = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 = (\delta_1 - \rho_2\delta_2/\rho_1)/2\}$, which is the perpendicular bisector of the line segment between (δ_1, δ_2) and (δ_1', δ_2').

Define the function $h(z_1, z_2) = (h_1(z_1, z_2), h_2(z_1, z_2)) = (-z_1 + \delta_1 - \rho_2\delta_2/\rho_1, z_2)$, which maps any point $(z_1, z_2) \in \mathbb{R}^2$ to its reflection across the line D. Define $A = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 < (\delta_1 - \rho_2\delta_2/\rho_1)/2 \text{ and } \{H_{01}, H_{0C}\} \cap M(z_1, z_2) \neq \emptyset\}$. Define $A' = h(A) = \{h(z_1, z_2) : (z_1, z_2) \in A\}$, i.e., the reflection of the set A across the line D. It follows that A and A' are disjoint sets, h is a bijection from A onto A', and since we assumed $M \in \mathcal{M}_B$ we have $A \subseteq B$. By a similar argument as in Case 2, we have $A' \subseteq B$.

We next show that monotonicity properties (a) and (c) imply $\{H_{01}, H_{0C}\} \cap M(z_1', z_2') \neq \emptyset$ for any $(z_1', z_2') \in A'$. Consider any $(z_1', z_2') \in A'$. Then $(z_1', z_2') \in B$, since as argued above $A' \subseteq B$. By definition, there exists a point $(z_1, z_2) \in A$ for which $h(z_1, z_2) = (z_1', z_2')$. By the definition of the set A, $z_1 < (\delta_1 - \rho_2\delta_2/\rho_1)/2$ and $\{H_{01}, H_{0C}\} \cap M(z_1, z_2) \neq \emptyset$. Therefore $(\delta_1 - \rho_2\delta_2/\rho_1) - 2z_1 > 0$. By monotonicity properties (a) and (c),

$\{H_{01}, H_{0C}\} \cap M(z_1 + (\delta_1 - \rho_2\delta_2/\rho_1 - 2z_1), z_2) \neq \emptyset$. Combining this with

$$z_1' = h(z_1, z_2) = -z_1 + \delta_1 - \rho_2\delta_2/\rho_1 = z_1 + (\delta_1 - \rho_2\delta_2/\rho_1 - 2z_1),$$

we have $\{H_{01}, H_{0C}\} \cap M(z_1', z_2') \neq \emptyset$.

We next show (19). Since $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \mathcal{H}_{\text{TRUE}}(\delta_1', \delta_2') = \{H_{01}, H_{0C}\}$ (for this case), the integral on the left side of the equality in (19) equals

$$\begin{aligned}
& \int_{(z_1,z_2)\in A\cup A'} 1[\{H_{01}, H_{0C}\} \cap M(z_1, z_2) \neq \emptyset]\eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
= & \int_{(z_1,z_2)\in A\cup A'} \eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 && (30) \\
= & \int_{(z_1',z_2')\in h^{-1}(A\cup A')} \eta(h_1(z_1', z_2') - \delta_1, h_2(z_1', z_2') - \delta_2)dz_1'\, dz_2' && (31) \\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(-z_1' - \rho_2\delta_2/\rho_1, z_2' - \delta_2)dz_1'\, dz_2' && (32) \\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(-(z_1' + \rho_2\delta_2/\rho_1), z_2' - \delta_2')dz_1'\, dz_2' && (33) \\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(z_1' - \delta_1', z_2' - \delta_2')dz_1'\, dz_2', && (34) \\
= & \int_{(z_1',z_2')\in A\cup A'} 1[\{H_{01}, H_{0C}\} \cap M(z_1', z_2') \neq \emptyset]\eta(z_1' - \delta_1', z_2' - \delta_2')dz_1'\, dz_2', && (35)
\end{aligned}$$

where (30) follows from $\{H_{01}, H_{0C}\} \cap M(z_1, z_2) \neq \emptyset$ for any $(z_1, z_2) \in A \cup A'$; (31) follows by the change of variables $(z_1', z_2') = h^{-1}(z_1, z_2)$, for which the absolute value of the Jacobian determinant equals 1; (38) follows from $h^{-1}(A \cup A') = A \cup A'$ and the definition of h; (33) follows from the definition of (δ_1', δ_2'); (34) follows from $\eta(-x, y) = \eta(x, y)$ for any $(x, y) \in \mathbb{R}^2$; and, (35) follows from $\{H_{01}, H_{0C}\} \cap M(z_1', z_2') \neq \emptyset$ for any $(z_1', z_2') \in A \cup A'$. This shows (19).

To show (20), consider any $(z_1, z_2) \in \mathbb{R}^2 \setminus A$ such that $\{H_{01}, H_{0C}\} \cap M(z_1, z_2) \neq \emptyset$. Then we have $z_1 \geq (\delta_1 - \rho_2\delta_2/\rho_1)/2$, which implies $2(z_1 - \delta_1) \geq -\delta_1 - \rho_2\delta_2/\rho_1$. Multiplying both sides of the previous inequality by $\delta_1 + \rho_2\delta_2/\rho_1$ (which is nonpositive by the assumption above that $\rho_1\delta_1 + \rho_2\delta_2 \le 0$) and moving all terms to the left side, we have $2(z_1 - \delta_1)(\delta_1 + \rho_2\delta_2/\rho_1) + (\delta_1 + \rho_2\delta_2/\rho_1)^2 \le 0$. Therefore,

$$
\begin{aligned}
(z_1 - \delta_1')^2 &= (z_1 - \delta_1 + (\delta_1 + \rho_2\delta_2/\rho_1))^2 \\
&= (z_1 - \delta_1)^2 + 2(z_1 - \delta_1)(\delta_1 + \rho_2\delta_2/\rho_1) + (\delta_1 + \rho_2\delta_2/\rho_1)^2 \\
&\le (z_1 - \delta_1)^2.
\end{aligned}
$$

The result in the above display implies $\eta(z_1 - \delta_1, z_2 - \delta_2) \le \eta(z_1 - \delta_1', z_2 - \delta_2')$. We have shown for any $(z_1, z_2) \in \mathbb{R}^2 \setminus A$,

$$
1[\{H_{01}, H_{0C}\} \cap M(z_1, z_2) \neq \emptyset]\eta(z_1 - \delta_1, z_2 - \delta_2) \le 1[\{H_{01}, H_{0C}\} \cap M(z_1, z_2) \neq \emptyset]\eta(z_1 - \delta_1', z_2 - \delta_2').
$$

This implies (20). Having shown (19) and (20), the above argument (21)-(22) implies $F_M(\delta_1, \delta_2) \le F_M(\delta_1', \delta_2')$, which completes Case 3.

Case 4: $\delta_1 > 0$ and $\delta_2 \le 0$. Analogous arguments as in cases 2 and 3 (which together handle $\delta_1 \le 0, \delta_2 > 0$) apply for this case, by exchanging the subscripts 1 and 2 and making corresponding modifications throughout the arguments.

Case 5: $\delta_1 \le 0$ and $\delta_2 \le 0$. Let $x = \min\{-\delta_1, -\delta_2\}$, which is nonnegative. For this case, define $(\delta_1', \delta_2') = (\delta_1 + x, \delta_2 + x)$, which is in G since at least one of δ_1', δ_2' equals 0. We have that both δ_1' and δ_2' are nonpositive, since for each $s \in \{1, 2\}$, $\delta_s' = \delta_s + x \le \delta_s + (-\delta_s) = 0$. Therefore, $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \mathcal{H}_{\text{TRUE}}(\delta_1', \delta_2') = \{H_{01}, H_{02}, H_{0C}\}$. Also, it follows from each $\delta_s \le 0$ and the definition of x that $\delta_1 + \delta_2 + x \le 0$.

Define the line $D = \{(z_1, z_2) : z_2 + z_1 = \delta_1 + \delta_2 + x\}$, which is the perpendicular bisector of the line segment between (δ_1, δ_2) and (δ_1', δ_2'). Define the function $h(z_1, z_2) = (h_1(z_1, z_2), h_2(z_1, z_2)) = (-z_2 + \delta_1 + \delta_2 + x, -z_1 + \delta_1 + \delta_2 + x)$, which maps any point $(z_1, z_2) \in \mathbb{R}^2$ to its reflection across the line D.

Define $A = \{(z_1, z_2) \in \mathbb{R}^2 : z_1 + z_2 < \delta_1 + \delta_2 + x \text{ and } M(z_1, z_2) \neq \emptyset\}$. Define $A' = h(A) = \{h(z_1, z_2) : (z_1, z_2) \in A\}$, i.e., the reflection of the set A across the line D. It follows that A and A' are disjoint sets, and h is a bijection from A onto A'.

Since we assumed $M \in \mathcal{M}_B$, we have $A \subseteq B$. We next show $A' \subseteq B$. For any $(z_1, z_2) \in A$, we have $z_1 + z_2 < \delta_1 + \delta_2 + x$. Since $A \subseteq B$, we have $z_1 \geq -b, z_2 \geq -b$. These inequalities imply $h_1(z_1, z_2) = -z_2 + \delta_1 + \delta_2 + x > z_1 \geq -b$ and $h_2(z_1, z_2) = -z_1 + \delta_1 + \delta_2 + x > z_2 \geq -b$. For each $s \in \{1, 2\}$, since $z_s \geq -b$ we have $h_{3-s}(z_1, z_2) = -z_s + \delta_1 + \delta_2 + x \leq b + \delta_1 + \delta_2 + x \leq b$ (where we used $\delta_1 + \delta_2 + x \leq 0$). Therefore, $h(z_1, z_2) \in B$. This shows $A' \subseteq B$.

We next show that monotonicity property (d) implies $M(z_1', z_2') \neq \emptyset$ for any $(z_1', z_2') \in A'$. Consider any $(z_1', z_2') \in A'$. Then $(z_1', z_2') \in B$, since we showed that $A' \subseteq B$. By definition, for some $(z_1, z_2) \in A, h(z_1, z_2) = (z_1', z_2')$. By the definition of the set A, we have $z_1 + z_2 < \delta_1 + \delta_2 + x \leq 0$ and $M(z_1, z_2) \neq \emptyset$. Let $d = -z_1 - z_2 + \delta_1 + \delta_2 + x$. It follows that $d > 0$ and $z_1' = z_1 + d, z_2' = z_2 + d$. By monotonicity property (d), $M(z_1 + d, z_2 + d) \neq \emptyset$, and therefore $M(z_1', z_2') \neq \emptyset$.

We next show (19). Since $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \mathcal{H}_{\text{TRUE}}(\delta_1', \delta_2') = \{H_{01}, H_{02}, H_{0C}\}$ (for this case), we have $1[M(z_1, z_2) \cap \mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) \neq \emptyset] = 1[M(z_1, z_2) \neq \emptyset]$. This implies the integral on the left side of the equality in (19) equals

$$
\begin{aligned}
& \int_{(z_1,z_2)\in A\cup A'} 1[M(z_1, z_2) \neq \emptyset]\eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
= & \int_{(z_1,z_2)\in A\cup A'} \eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 && (36) \\
= & \int_{(z_1',z_2')\in h^{-1}(A\cup A')} \eta(h_1(z_1', z_2') - \delta_1, h_2(z_1', z_2') - \delta_2)dz_1'\, dz_2' && (37) \\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(-z_2' + \delta_2 + x, -z_1' + \delta_1 + x)dz_1'\, dz_2' && (38) \\
= & \int_{(z_1',z_2')\in A\cup A'} \eta(-(z_2' - \delta_2'), -(z_1' - \delta_1'))dz_1'\, dz_2' && (39)
\end{aligned}
$$

$$= \int_{(z_1',z_2')\in A\cup A'} \eta(z_1'-\delta_1', z_2'-\delta_2')dz_1'\,dz_2', \tag{40}$$

$$= \int_{(z_1',z_2')\in A\cup A'} 1[M(z_1,z_2)\neq\emptyset]\eta(z_1'-\delta_1', z_2'-\delta_2')dz_1'\,dz_2', \tag{41}$$

where (36) follows from $M(z_1, z_2) \neq \emptyset$ for any $(z_1, z_2) \in A \cup A'$; (37) follows by the change of variables $(z_1', z_2') = h^{-1}(z_1, z_2)$, for which the absolute value of the Jacobian determinant equals 1; (38) follows from $h^{-1}(A \cup A') = A \cup A'$ and the definition of h; (39) follows from the definition of (δ_1', δ_2'); (40) follows from $\eta(x, y) = \eta(-y, -x)$ for any $(x, y) \in \mathbb{R}^2$; and, (41) follows from $M(z_1', z_2') \neq \emptyset$ for any $(z_1', z_2') \in A \cup A'$. This shows (19).

To show (20), consider any $(z_1, z_2) \in \mathbb{R}^2 \setminus A$ such that $M(z_1, z_2) \neq \emptyset$. We must then have $z_1 + z_2 - \delta_1 - \delta_2 \geq x$, and after multiplying both sides by $(-2x)$ (which is nonpositive since $x \geq 0$), we have $-2x(z_1 + z_2 - \delta_1 - \delta_2) \leq -2x^2$. Therefore, for such (z_1, z_2),

$$\begin{aligned}
(z_1-\delta_1')^2 + (z_2-\delta_2')^2 &= (z_1-\delta_1-x)^2 + (z_2-\delta_2-x)^2 \\
&= (z_1-\delta_1)^2 + (z_2-\delta_2)^2 - 2x(z_1-\delta_1+z_2-\delta_2) + 2x^2 \\
&\leq (z_1-\delta_1)^2 + (z_2-\delta_2)^2 - 2x^2 + 2x^2 \\
&= (z_1-\delta_1)^2 + (z_2-\delta_2)^2,
\end{aligned}$$

which implies $\eta(z_1 - \delta_1, z_2 - \delta_2) \leq \eta(z_1 - \delta_1', z_2 - \delta_2')$. We have shown for any $(z_1, z_2) \in \mathbb{R}^2 \setminus A$,

$$1[M(z_1,z_2)\neq\emptyset]\eta(z_1-\delta_1, z_2-\delta_2) \leq 1[M(z_1,z_2)\neq\emptyset]\eta(z_1-\delta_1', z_2-\delta_2'). \tag{42}$$

This implies (20), which completes the proof that in Case 5, $F_M(\delta_1, \delta_2) \leq F_M(\delta_1', \delta_2')$.

Since Cases 1-5 cover all $(\delta_1, \delta_2) \in \mathbb{R}^2$, we showed for any $(\delta_1, \delta_2) \in \mathbb{R}^2$ that there exists $(\delta_1', \delta_2') \in G$ for which $F_M(\delta_1, \delta_2) \leq F_M(\delta_1', \delta_2')$. This completes the proof of part (a) of Theorem 1. Part (b) follows by an analogous argument, since the condition $M \in \mathcal{M}_B$ was only used above to show that $A \cup A' \subseteq B$; when $R = \mathbb{R}^2$, this is not needed. $\square$

We next show monotonicity properties (a)-(c) are not sufficient to imply the conclusion of part (a) of Theorem 1, i.e., that (15) holds for $R = B$. We give a (pathological) counterexample below to demonstrate this. This shows some additional property is required, such as monotonicity property (d), in order for this conclusion to follow. It is an area of future work to devise weaker sets of conditions under which the conclusion of the theorem holds.

Define $A_1 = \{(z_1, z_2) \in \mathbb{R}^2 : -5 \leq z_1 \leq 5, -5 \leq z_2 \leq -4, z_1 > z_2\}$ and $A_2 = \{(z_1, z_2) \in \mathbb{R}^2 : -5 \leq z_2 \leq 5, -5 \leq z_1 \leq -4, z_2 > z_1\}$. Consider the multiple testing procedure that rejects H_{01} whenever $(Z_1, Z_2) \in A_1$, and rejects H_{02} whenever $(Z_1, Z_2) \in A_2$. Though this is a very unrealistic procedure, it does provide a counterexample, as we show next. It is straightforward to verify the procedure satisfies monotonicity properties (a)-(c) with respect to $R = B = [-b, b] \times [-b, b]$ for $b = 5$. Also, the procedure is in $\mathcal{M}_{det}$. We verified that the worst-case familywise Type I error rate over G is 0.38, which is achieved at $(\delta_1, \delta_2) = (-4.50, 0)$; this was done using the method in Section H.3 below. However, at $(\delta_1, \delta_2) = (-3.99, -3.99) \notin G$, the familywise Type I error rate is 0.46, which exceeds the maximum familywise Type I error rate over G. The above example shows that monotonicity properties (a)-(c) do not suffice to show (15). (The above results were rounded to two decimal places.)

H.2 Method for Bounding (3) based on the Mean Value Theorem

For any $(\delta_1, \delta_2) \in \mathbb{R}^2, (\delta_1', \delta_2') \in \mathbb{R}^2$, $\lambda \in [0, 1]$, and $s \in \{1, 2\}$, define the function $\tilde{\delta}_s(\lambda) = \lambda\delta_s + (1-\lambda)\delta_s'$. The set of values $\{(\tilde{\delta}_1(\lambda), \tilde{\delta}_2(\lambda)) : \lambda \in [0, 1]\}$ is the line segment connecting (δ_1', δ_2') to (δ_1, δ_2). The following lemma bounds the difference between the familywise Type I error at (δ_1, δ_2) and at (δ_1', δ_2'), as a function of $|\delta_1 - \delta_1'|$ and $|\delta_2 - \delta_2'|$.

Lemma 1: Assume there exists a set $H \subseteq \mathcal{H}$ such that for all $\lambda \in [0,1]$, $\mathcal{H}_{TRUE}(\tilde{\delta}_1(\lambda), \tilde{\delta}_2(\lambda)) = H$. Then for any $M \in \mathcal{M}_{det}$, we have $|F_M(\delta_1, \delta_2) - F_M(\delta_1', \delta_2')| \leq \sqrt{2/\pi} \sum_{s=1}^{2} |\delta_s - \delta_s'|$.

Proof: Consider any $(\delta_1, \delta_2) \in \mathbb{R}^2, (\delta_1', \delta_2') \in \mathbb{R}^2$, and $H \subseteq \mathcal{H}$ such that the condition of the lemma holds. Consider any $M \in \mathcal{M}_{det}$. We then have

$$
\begin{aligned}
& \left| \frac{d}{d\lambda} F_M(\tilde{\delta}_1(\lambda), \tilde{\delta}_2(\lambda)) \right| \\
= & \left| \frac{d}{d\lambda} \int_{(z_1,z_2)\in\mathbb{R}^2} 1[M(z_1,z_2) \cap H \neq \emptyset] \eta(z_1 - \tilde{\delta}_1(\lambda), z_2 - \tilde{\delta}_2(\lambda)) dz_1 \, dz_2 \right| \\
\leq & \int_{(z_1,z_2)\in\mathbb{R}^2} \left| 1[M(z_1,z_2) \cap H \neq \emptyset] \frac{d}{d\lambda} \eta(z_1 - \tilde{\delta}_1(\lambda), z_2 - \tilde{\delta}_2(\lambda)) \right| dz_1 dz_2 \qquad (43) \\
\leq & \int_{(z_1,z_2)\in\mathbb{R}^2} \left| \frac{d}{d\lambda} \eta(z_1 - \tilde{\delta}_1(\lambda), z_2 - \tilde{\delta}_2(\lambda)) \right| dz_1 dz_2 \\
= & \int_{(z_1,z_2)\in\mathbb{R}^2} \left| \sum_{s=1}^{2} \{z_s - \tilde{\delta}_s(\lambda)\} \frac{d}{d\lambda} \tilde{\delta}_s(\lambda) \right| \eta(z_1 - \tilde{\delta}_1(\lambda), z_2 - \tilde{\delta}_2(\lambda)) dz_1 dz_2 \\
\leq & \sum_{s=1}^{2} \frac{1}{\sqrt{2\pi}} \int_{z_s \in \mathbb{R}} |\{z_s - \tilde{\delta}_s(\lambda)\}(\delta_s - \delta_s')| \exp\left\{ -(z_s - \tilde{\delta}_s(\lambda))^2/2 \right\} dz_s \qquad (44) \\
= & \sqrt{2/\pi} \sum_{s=1}^{2} |\delta_s - \delta_s'|,
\end{aligned}
$$

where the exchange of order of differentiation and integration in (43) is justified by Fubini's theorem (which uses that $M \in \mathcal{M}_{det} \subset \mathcal{M}$, and by definition all procedures in $\mathcal{M}$ are measurable functions), and where (44) follows from $\frac{d}{d\lambda}\{\tilde{\delta}_s(\lambda)\} = \delta_s - \delta_s'$. We then have

$$
\begin{aligned}
|F_M(\delta_1, \delta_2) - F_M(\delta_1', \delta_2')| &= \left| F_M(\tilde{\delta}_1(1), \tilde{\delta}_2(1)) - F_M(\tilde{\delta}_1(0), \tilde{\delta}_2(0)) \right| \\
&\leq |1 - 0| \sup_{\lambda \in [0,1]} \left| \frac{d}{d\lambda} F_M(\tilde{\delta}_1(\lambda), \tilde{\delta}_2(\lambda)) \right| \qquad (45) \\
&\leq \sqrt{2/\pi} \sum_{s=1}^{2} |\delta_s - \delta_s'|,
\end{aligned}
$$

where (45) follows by the mean value theorem. This completes the proof of Lemma 1. $\square$

We next define and prove correctness of a search algorithm used in Section H.3 below. The inputs to the algorithm include a procedure $M \in \mathcal{M}_{det}$, and the two distinct points $(\delta_{1,start}, \delta_{2,start}) \in \mathbb{R}^2$ and $(\delta_{1,end}, \delta_{2,end}) \in \mathbb{R}^2$. Let $\gamma = \left(\sum_{s=1}^{2} |\delta_{s,start} - \delta_{s,end}|\right)^{-1}$, and for each $s \in \{1,2\}$, define $\underline{\delta}_s(x) = x\delta_{s,start} + (1-x)\delta_{s,end}$.

Search algorithm:

1. Initialize $k = 1$, and $x_1 = 1$.
2. Let $f_k = F_M(\underline{\delta}_1(x_k), \underline{\delta}_2(x_k))$.

   If $f_k \geq 0.05$, stop; else, set $x_{k+1} = \max\{0, x_k - \gamma(0.05 - f_k)\}$.
3. If $x_{k+1} = 0$, increment k by 1 and stop; else, increment k by 1 and go to step 2.

Let k^* denote the value of k when the algorithm terminates (and let $k^* = \infty$ if the algorithm never terminates). We prove the following lemma below:

Lemma 2: Assume there exists an $H \subseteq \mathcal{H}$ such that for any $x \in (0,1]$, $\mathcal{H}_{TRUE}(\underline{\delta}_1(x), \underline{\delta}_2(x)) = H$. Then we have ($k^* < \infty$ and $x_{k^*} = 0$) if and only if $\sup_{x\in(0,1]} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$.

The lemma implies that when the above algorithm terminates with $x_{k^*} = 0$, we can conclude the constraints (3) hold for each (δ_1, δ_2) in the set $\{(\underline{\delta}_1(x), \underline{\delta}_2(x)) : x \in (0,1]\}$. Furthermore, if $\sup_{x\in(0,1]} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$, the algorithm is guaranteed to eventually terminate with $x_{k^*} = 0$. The advantage of the above algorithm, compared to a simpler search where $F_M(\underline{\delta}_1(x), \underline{\delta}_2(x))$ is computed over a grid of equally spaced points $x \in (0,1]$, is that the above algorithm can be much faster. This is because, roughly speaking, for regions where $F_M(\underline{\delta}_1(x), \underline{\delta}_2(x))$ is not close to 0.05, the above algorithm evaluates this quantity at less densely spaced points than in regions where $F_M(\underline{\delta}_1(x), \underline{\delta}_2(x))$ is close to 0.05.

Proof of Lemma 2: We show by induction on k that the following claim holds: for any integer $k : 1 \leq k \leq k^*, \sup_{\{x:x_k<x\leq 1\}} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$ (where if the set in the subscript of the supremum is empty, we let the value on the left side of the inequality be $-\infty$ by convention). The base case, $k = 1$, holds since the set in the subscript of the supremum is empty. For the inductive step, consider any integer $k : 1 \leq k < k^*$. Assume the inductive hypothesis, i.e., $\sup_{\{x:x_k<x\leq 1\}} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$. It cannot be that $f_k \geq 0.05$, since if so, the algorithm would have stopped at step 2 in iteration k, leading to $k = k^*$, and contradicting $k < k^*$. Also, it cannot be that $x_k = 0$, since if so, the algorithm would have stopped at step 3 in the previous iteration, so that $k = k^*$, contradicting $k < k^*$. Therefore, $f_k < 0.05$ and $x_k > 0$, which implies $x_{k+1} = \max\{0, x_k - \gamma(0.05 - f_k)\} < x_k$. For any $x \in (x_{k+1}, x_k]$, by Lemma 1 applied at $(\delta_1, \delta_2) = (\underline{\delta}_1(x), \underline{\delta}_2(x)), (\delta_1', \delta_2') = (\underline{\delta}_1(x_k), \underline{\delta}_2(x_k))$,

$$
\begin{aligned}
|F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) - F_M(\underline{\delta}_1(x_k), \underline{\delta}_2(x_k))| &\leq \sqrt{2/\pi} \sum_{s=1}^{2} |(\delta_{s,start} - \delta_{s,end})(x - x_k)| \\
&\leq \sqrt{2/\pi} \sum_{s=1}^{2} |(\delta_{s,start} - \delta_{s,end})(x_{k+1} - x_k)| \\
&\leq \sqrt{2/\pi}\{\gamma(0.05 - f_k)\} \sum_{s=1}^{2} |(\delta_{s,start} - \delta_{s,end})| \\
&< 0.05 - f_k \\
&= 0.05 - F_M(\underline{\delta}_1(x_k), \underline{\delta}_2(x_k)),
\end{aligned}
$$

which implies $F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$. Combining this with the inductive hypothesis, we have shown $\sup_{\{x:x_{k+1}<x\leq 1\}} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$. This completes the proof by induction that for any integer $k : 1 \leq k \leq k^*, \sup_{\{x:x_k<x\leq 1\}} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$. If $k^* < \infty$ and $x_{k^*} = 0$, then by the previous inequality at $k = k^*$, we have $\sup_{x\in(0,1]} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$.

To show the converse, assume $\sup_{x\in(0,1]} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$. Let $\psi = 0.05 - \sup_{x\in(0,1]} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) > 0$. The algorithm can never terminate in step 2, since at step 2 the value of x_k is always in $(0,1]$, and by assumption $\sup_{x\in(0,1]} F_M(\underline{\delta}_1(x), \underline{\delta}_2(x)) < 0.05$. At each iteration k of step 2, we have $0 \leq x_{k+1} \leq \max\{0, x_k - \gamma\psi\}$. Therefore, after at most $\lceil 1/(\gamma\psi) \rceil$ iterations, we must have $x_{k+1} = 0$. This implies the algorithm terminates in step 3 with $x_{k^*} = 0$ after at most $\lceil 1/(\gamma\psi) \rceil$ iterations. This completes the proof of Lemma 2. $\square$

H.3 Verifying Familywise Type I error constraints (3) for each $M^* \in \mathcal{M}^*$

For each $M^* \in \mathcal{M}^*$, we verified the conditions of part (a) of Theorem 1. Therefore, to verify M^* satisfies all constraints (3), it suffices to verify the constraints corresponding to each $(\delta_1, \delta_2) \in G$.

Consider any $M^* \in \mathcal{M}^*$. Define $B' = [-7, 7] \times [-7, 7]$. We partition G into the following sets: $V_1 = \{(0, 0)\}$; $V_2 = G \setminus B'$; $V_3 = \{(\delta_1, \delta_2) \in (B' \setminus V_1) : \delta_1 = 0\}$;
$V_4 = \{(\delta_1, \delta_2) \in (B' \setminus V_1) : \delta_2 = 0\}$; $V_5 = \{(\delta_1, \delta_2) \in (B' \setminus V_1) : \rho_1\delta_1 + \rho_2\delta_2 = 0\}$. For each $i \in I = \{1, 2, 3, 4, 5\}$, define $v_i = \sup_{(\delta_1,\delta_2)\in V_i} F_{M^*}(\delta_1, \delta_2)$. For each $i \in \{1, 2, 3, 4, 5\}$, we consider the case of $(\delta_1, \delta_2) \in V_i$ and show $v_i \leq 0.05$.

Case $i = 1$: To show (3) is satisfied for $(\delta_1, \delta_2) = (0, 0)$, we computed $F_{M^*}(0, 0)$, which was less than or equal to 0.04991 (rounded to five decimal places) for each $M^* \in \mathcal{M}^*$. Therefore, $v_1 < 0.05$.

Case $i = 2$: For any $(\delta_1, \delta_2) \in G \setminus B'$, we have either $|\delta_1| > 7$ or $|\delta_2| > 7$. Consider the subcase where $|\delta_1| > 7$. Then $\{(z_1, z_2) \in \mathbb{R}^2 : |z_1 - \delta_1| \leq 2\} \cap B = \emptyset$. Since each $M^* \in \mathcal{M}_B$,

$$
\begin{aligned}
F_{M^*}(\delta_1, \delta_2) &= \int_{(z_1,z_2)\in B} 1[M^*(z_1, z_2) \cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1, \delta_2) \neq \emptyset]\eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
&\leq \int_{(z_1,z_2)\in B} \eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
&= \int_{(z_1,z_2)\in B:|z_1-\delta_1|>2} \eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 \\
&\leq \int_{(z_1,z_2)\in \mathbb{R}^2:|z_1-\delta_1|>2} \eta(z_1 - \delta_1, z_2 - \delta_2)dz_1\, dz_2 = 2\Phi(-2) < 0.05.
\end{aligned}
$$

This shows $F_{M^*}(\delta_1, \delta_2) < 0.05$ for the subcase $|\delta_1| > 7$. A similar argument shows this for the subcase $|\delta_2| > 7$. Therefore, $v_2 < 0.05$.

Case $i = 3$**:** Consider any $(\delta_1, \delta_2) \in V_3$. We consider two subcases. The first is that $\delta_2 < 0$. Then $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \{H_{01}, H_{02}, H_{0C}\}$. We ran the search algorithm from Section H.2 (implemented in R) for each $M^* \in \mathcal{M}^*$ using inputs $M = M^*$; $(\delta_{1,start}, \delta_{2,start}) = (0, -7)$; $(\delta_{1,end}, \delta_{2,end}) = (0, 0)$. For each $M^* \in \mathcal{M}^*$, the search algorithm stopped with $x_{k^*} = 0$. The condition of Lemma 2 holds for $H = \{H_{01}, H_{02}, H_{0C}\}$. Therefore, Lemma 2 implies $\sup_{(\delta_1,\delta_2)\in[V_3\cap\{(\delta_1,\delta_2):\delta_2<0\}]} F_{M^*}(\delta_1, \delta_2) < 0.05$.

The second subcase is that $\delta_2 > 0$. Then $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \{H_{01}\}$. We ran the search algorithm from Section H.2 for each $M^* \in \mathcal{M}^*$ using inputs $M = M^*$; $(\delta_{1,start}, \delta_{2,start}) = (0, 7)$; $(\delta_{1,end}, \delta_{2,end}) = (0, 0)$. For each $M^* \in \mathcal{M}^*$, the search algorithm stopped with $x_{k^*} = 0$. The condition of Lemma 2 holds for $H = \{H_{01}\}$. Therefore, Lemma 2 implies $\sup_{(\delta_1,\delta_2)\in[V_3\cap\{(\delta_1,\delta_2):\delta_2>0\}]} F_{M^*}(\delta_1, \delta_2) < 0.05$. Since by definition, for any $(\delta_1, \delta_2) \in V_3$ we have $\delta_2 \neq 0$, the condition in one of the subcases holds. The results of the two subcases imply $v_3 < 0.05$.

Case $i = 4$**:** The argument is similar to that in Case $i = 3$, but we give it here for completeness. Consider any $(\delta_1, \delta_2) \in V_4$. We consider two subcases. The first is that $\delta_1 < 0$. Then $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \{H_{01}, H_{02}, H_{0C}\}$. We ran the search algorithm from Section H.2 for each $M^* \in \mathcal{M}^*$ using inputs $M = M^*$; $(\delta_{1,start}, \delta_{2,start}) = (-7, 0)$; $(\delta_{1,end}, \delta_{2,end}) = (0, 0)$. For each $M^* \in \mathcal{M}^*$, the search algorithm stopped with $x_{k^*} = 0$. The condition of Lemma 2 holds for $H = \{H_{01}, H_{02}, H_{0C}\}$. Therefore, Lemma 2 implies $\sup_{(\delta_1,\delta_2)\in[V_4\cap\{(\delta_1,\delta_2):\delta_1<0\}]} F_{M^*}(\delta_1, \delta_2) < 0.05$.

The second subcase is that $\delta_1 > 0$. Then $\mathcal{H}_{\text{TRUE}}(\delta_1, \delta_2) = \{H_{02}\}$. We ran the search algorithm from Section H.2 for each $M^* \in \mathcal{M}^*$ using inputs $M = M^*$; $(\delta_{1,start}, \delta_{2,start}) = (7, 0)$; $(\delta_{1,end}, \delta_{2,end}) = (0, 0)$. For each $M^* \in \mathcal{M}^*$, the search algorithm stopped with

$x_{k^*} = 0$. The condition of Lemma 2 holds for $H = \{H_{02}\}$. Therefore, Lemma 2 implies $\sup_{(\delta_1,\delta_2)\in[V_4\cap\{(\delta_1,\delta_2):\delta_1>0\}]} F_{M^*}(\delta_1, \delta_2) < 0.05$. Since by definition, for any $(\delta_1, \delta_2) \in V_4$ we have $\delta_1 \neq 0$, the condition in one of the subcases holds. The results of the two subcases imply $v_4 < 0.05$.

Case $i = 5$**:** We consider two subcases. The first is that $\delta_1 < 0$. We ran the search algorithm from Section H.2 for each $M^* \in \mathcal{M}^*$ using inputs $M = M^*$; $(\delta_{1,start}, \delta_{2,start}) = (-7, 7\rho_1/\rho_2)$; $(\delta_{1,end}, \delta_{2,end}) = (0, 0)$. For each $M^* \in \mathcal{M}^*$, the search algorithm stopped with $x_{k^*} = 0$. The condition of Lemma 2 holds for $H = \{H_{01}, H_{0C}\}$. Therefore, Lemma 2 implies $\sup_{(\delta_1,\delta_2)\in[V_5\cap\{(\delta_1,\delta_2):\delta_1<0\}]} F_{M^*}(\delta_1, \delta_2) < 0.05$.

The second subcase is that $\delta_1 > 0$. We ran the search algorithm from Section H.2 for each $M^* \in \mathcal{M}^*$ using inputs $M = M^*$; $(\delta_{1,start}, \delta_{2,start}) = (7, -7\rho_1/\rho_2)$; $(\delta_{1,end}, \delta_{2,end}) = (0, 0)$. For each $M^* \in \mathcal{M}^*$, the search algorithm stopped with $x_{k^*} = 0$. The condition of Lemma 2 holds for $H = \{H_{02}, H_{0C}\}$. Therefore, Lemma 2 implies $\sup_{(\delta_1,\delta_2)\in[V_5\cap\{(\delta_1,\delta_2):\delta_1>0\}]} F_{M^*}(\delta_1, \delta_2) < 0.05$. Since by definition, for any $(\delta_1, \delta_2) \in V_5$ we have $\delta_1 \neq 0$, the condition in one of the subcases holds. The results of the two subcases imply $v_5 < 0.05$.

This completes the verification that for any $M^* \in \mathcal{M}^*$, and any $(\delta_1, \delta_2) \in G$, we have $F_{M^*}(\delta_1, \delta_2) \leq 0.05$. By the arguments above, this implies each $M^* \in \mathcal{M}^*$ obeys the family-wise Type I error constraints (3) of the original problem, at $\alpha = 0.05$.

I Using the Dual to Lower Bound the Bayes Risk of the Original Problem

The following multiple testing procedure is a minimizer of the unbounded problem (17):

$$
\begin{aligned}
M_{\mathrm{u}}^*(z_1, z_2) \quad = \quad \arg\min_{s \in \mathcal{S}} \Bigg[& \int L(s; \delta_1, \delta_2)\phi(z_1 - \delta_1)\phi(z_2 - \delta_2) d\Lambda(\delta_1, \delta_2) \\
& -1[H_{0C} \in s]\nu_p^* \phi(z_1 - \delta_1^{\min})\phi(z_2 - \delta_2^{\min}) \\
& + \sum_{j \in C_{\mathrm{FWER}}} 1[s \cap \mathcal{H}_{\mathrm{TRUE}}(\delta_{1,j}, \delta_{2,j}) \neq \emptyset]\nu_j^* \phi(z_1 - \delta_{1,j})\phi(z_2 - \delta_{2,j}) \Bigg],
\end{aligned}
$$

where ϕ is the density of the standard normal distribution.

For any pair (z_1, z_2), one can compute $M_u^*(z_1, z_2)$ by separately evaluating the term in brackets above at each value of $s \in \mathcal{S}$, using numerical integration with respect to Λ, and selecting s corresponding to the minimum value obtained (with ties broken arbitrarily). Using this as a subroutine, one can compute the minimum value of (17) by evaluating the expression in brackets in (17) at $M = M_{\mathrm{u}}^*$, using numerical integration. These numerical integrations are implemented in R using code given in the Supplementary Materials.

J Minimax Optimization Problems; Handling Multiple Optimization Criteria

In Section J.1, we consider the minimax objective function (5), and solve minimax versions of the problems from Section 5.1. In Section J.2, we briefly discuss the problem mentioned in Section 10 of finding a multiple testing procedure that simultaneously has good performance under multiple (Bayes) optimization criteria.

J.1 Solutions to Minimax Versions of Problems from Section 5.1

We can replace the Bayes objective function (2) by the minimax objective function (5), in which the maximum is taken over a finite set of alternatives $\mathcal{P}$. The resulting optimization problem can be solved by binary search over candidate values v for (5), where at each step we compute whether there exists a solution to the set of constraints (3)-(4) plus the additional constraints $E_{\delta_1,\delta_2} L(M(Z_1, Z_2, U); \delta_1, \delta_2) \leq v$ for each $(\delta_1, \delta_2) \in \mathcal{P}$. For values v where a solution exists, (5) must be less than or equal to v; conversely, if no solution exists, (5) must be greater than v. Each step of the binary search is done using the discretized constraints (11)-(14) from Section 4 plus the additional constraint

$$\sum_{r\in\mathcal{R},s\in\mathcal{S}} L(s; \delta_1, \delta_2) P_{\delta_1,\delta_2}[(Z_1, Z_2) \in r] m_{rs} \leq v, \tag{46}$$

for each $(\delta_1, \delta_2) \in \mathcal{P}$. Determining whether a solution exists to a set of sparse linear constraints can be solved using a similar algorithm as in Section 9. We applied this to minimax versions of the problems from Section 5.1, as described next.

For the remainder of this subsection, we replace (2) by the minimax objective function (5), in which the maximum is taken over $\mathcal{P} = \{(\delta_1^{\min}, 0), (0, \delta_2^{\min}), (\delta_1^{\min}, \delta_2^{\min})\}$. We apply the binary search described above to the discretized version of this problem at $1 - \beta = 0.88$.

For the symmetric case, the optimal value of the minimax objective function (5) is 0.49, which is the risk of the minimax optimal procedure at $(\delta_1^{\min}, 0)$ and at $(0, \delta_2^{\min})$. Recall that at these alternatives, respectively, the risk equals one minus the probability the minimax optimal procedure rejects H_{0k} for $k = 1, 2$. One minus 0.49 equals, up do two decimal places, the values in column 2, rows 2-3 of Table 1, which are the corresponding rejection probabilities for the constrained Bayes optimal solution $\mathbf{m}^*_{\text{sym}}(0.88)$. Thus, the minimum risk over $\mathcal{P}$ of the constrained Bayes optimal solution is close to the minimax risk over $\mathcal{P}$.

For the asymmetric case, the optimal value of the minimax objective function (5) is 0.58, which is the risk of the minimax optimal procedure at $(0, \delta_2^{\min})$. Unlike the symmetric case, here the minimum risk over $\mathcal{P}$ of the corresponding constrained Bayes optimal solution

$\mathbf{m}^*_{\text{asym}}(0.88)$ is 12% above (i.e., worse than) the minimax risk over $\mathcal{P}$. This is because $\mathbf{m}^*_{\text{asym}}(0.88)$ has 67% power for H_{01} at $(\delta_1^{\min}, 0)$ but only 30% power for H_{02} at $(0, \delta_2^{\min})$. By trading off power for the former to gain power for the latter, the minimax optimal procedure improves the minimum risk over $\mathcal{P}$.

J.2 Finding a Procedure that Simultaneously has Good Performance under Multiple Optimization Criteria

We briefly discuss the problem mentioned in Section 10 of finding a multiple testing procedure that simultaneously has good performance under multiple optimization criteria. Consider the problem where one is given a finite set of optimization criteria $\{(L^{(i)}, \Lambda^{(i)})\}_{i=1}^k$, where the ith optimization criterion consists of a loss function $L^{(i)}$ and a prior $\Lambda^{(i)}$, and the goal is to find $M \in \mathcal{M}$ that minimizes the following extension of the Bayes objective function (2):

$$\max_{i \le k} \int E_{\delta_1,\delta_2} \left\{ L^{(i)}(M(Z_1, Z_2, U); \delta_1, \delta_2) \right\} d\Lambda^{(i)}(\delta_1, \delta_2), \tag{47}$$

under the constraints (3) and (4).

The above optimization problem can be solved by binary search over candidate values v for (47), using a similar approach as described in the previous subsection. The only difference is that at each step in the binary search, instead of the additional constraints (46), one includes the following additional constraint for each $i \in \{1, \dots, k\}$:

$$\sum_{r \in \mathcal{R}, s \in \mathcal{S}} \left\{ \int L^{(i)}(s; \delta_1, \delta_2) P_{\delta_1,\delta_2}[(Z_1, Z_2) \in r] d\Lambda^{(i)}(\delta_1, \delta_2) \right\} m_{rs} \le v.$$

K Definition of G'_{new} for Each Example in Section 5.1

As described in Section 6.2, we first solved the discretized problems in Section 5.1 at relatively coarse discretizations, to identify the general vicinities of the active constraints. We then defined modified sets of active constraints $G'_{new} \subset G$, with points concentrated in these vicinities, defined below.

For any two distinct points $q \in \mathbb{R}^2, q' \in \mathbb{R}^2$, and any real number $d > 0$, let $\text{Seq}[q, q', d]$ denote the set of equally spaced points on the line segment connecting q to q' with adjacent points distance d apart, and including the endpoints q, q'. E.g., $\text{Seq}[(0,0),(0,0.9),0.3] = \{(0,0),(0,0.3),(0,0.6),(0,0.9)\}$. For each active constraint in the solution to the discretized problem at the relatively coarse discretization, we included a corresponding set $\text{Seq}[q, q', d]$ in G'_{new} roughly centered at this active constraint, and spanning several points in the coarse discretization. The one exception is that we did not include a set of points around $(0,0)$ (which was an active constraint in all the discretized problems), and just included that point itself; it turned out that in each example, the resulting solution at the finer discretization (defined in Section 6.2) was very close to optimal (as described in Section 6.3) and satisfied all constraints of the original problem as verified in Section H.

For the symmetric case with $1 - \beta = 0.9$, we set $d = 0.005$ and

$$G'_{new} = \text{Seq}[(1.9,0),(2.1,0),d] \cup \text{Seq}[(0,1.9),(0,2.1),d] \cup \{(0,0)\}.$$

For this case, there are 83 constraints in G'_{new}, which combined with the power constraint gives a total of 84 constraints for the corresponding discretized problem.

For the asymmetric case with $1 - \beta = 0.9$, we set $d = 0.005$ and

$$G'_{new} = \text{Seq}[(2.2,0),(2.5,0),d] \cup \text{Seq}[(0,1.6),(0,1.9),d] \cup \{(0,0)\}.$$

For this case, there are 123 constraints in G'_{new}, which combined with the power constraint gives a total of 124 constraints for the corresponding discretized problem.

For the symmetric case with $1-\beta = 0.88$, we set $d = 0.01$ and

$$\begin{aligned} G'_{new} &= \mathrm{Seq}[(1.9,0),(2.2,0),d] \cup \mathrm{Seq}[(0,1.9),(0,2.2),d] \\ &\cup \mathrm{Seq}\left[(-1.2,1.2),(-1,1),d\sqrt{2}\right] \cup \mathrm{Seq}\left[(1,-1),(1.2,-1.2),d\sqrt{2}\right] \cup \{(0,0)\}. \end{aligned}$$

For this case, there are 105 constraints in G'_{new}, which combined with the power constraint gives a total of 106 constraints for the corresponding discretized problem.

For the asymmetric case with $1-\beta = 0.88$, where $\rho_1 = \sqrt{0.63}, \rho_2 = \sqrt{0.37}$, we set $d = 0.01$ and

$$\begin{aligned} G'_{new} &= \mathrm{Seq}[(2.3,0),(2.6,0),d] \cup \mathrm{Seq}[(0,1.7),(0,2.0),d] \\ &\cup \mathrm{Seq}\left[(-1.3,1.3\rho_1/\rho_2,d),(-1,\rho_1/\rho_2),d\sqrt{1+(\rho_1/\rho_2)^2}\right] \\ &\cup \mathrm{Seq}\left[(0.6,-0.6\rho_1/\rho_2),(1,-\rho_1/\rho_2),d\sqrt{1+(\rho_1/\rho_2)^2}\right] \cup \{(0,0)\}. \end{aligned}$$

For this case, there are 135 constraints in G'_{new}, which combined with the power constraint gives a total of 136 constraints for the corresponding discretized problem.

L Solving Constrained Bayes Optimization Problem when Active Constraints are Known

Even if the set of active constraints for our problems from Section 5.1 were somehow known or correctly guessed, the problems could still be challenging to solve using standard optimization methods such as Lagrange multipliers. For the sake of illustration, assume the set of active familywise Type I error constraints for the (non-discretized) constrained Bayes optimization problem from Section 3.2 were known, and denote it by C^*. Let k denote the number of elements in C^*, and let $\mathbb{R}_+$ denote the nonnegative reals. We consider an example of a standard approach using Lagrange multipliers to solve the constrained optimization problem from Section 3.2, which is equivalently expressed as the minimization problem (16). A first step is to construct the following unconstrained problem, for any given $\bar{\nu}^* \in \mathbb{R}_+^{k+1}$:

$$\inf_{M\in\mathcal{M}} \left[\int E_{\delta_1,\delta_2} L(M(Z_1,Z_2,U);\delta_1,\delta_2) d\Lambda(\delta_1,\delta_2) + \bar{\nu}_p^* \left\{ 1-\beta - P_{\delta_1^{\min},\delta_2^{\min}}(M \text{ rejects } H_{0C}) \right\} \right.$$
$$\left. + \sum_{j\in C^*} \bar{\nu}_j^* \left\{ P_{\delta_{1,j},\delta_{2,j}} \left(M \text{ rejects any null hypothesis in } \mathcal{H}_{\text{TRUE}}(\delta_{1,j},\delta_{2,j})\right) - \alpha \right\} \right]. \quad (48)$$

This is similar to (17) except the vector ν^* is replaced by $\bar{\nu}^*$ which represents a vector of Lagrange multipliers with unknown values, and we replace C_{FWER} by the (assumed to be known) active constraints C^*. For any $\bar{\nu}^* \in \mathbb{R}_+^{k+1}$, denote the minimizer over $M \in \mathcal{M}$ of (48) by $M^*(\bar{\nu}^*)$. For any given $\bar{\nu}^* \in \mathbb{R}_+^{k+1}$, $M^*(\bar{\nu}^*)$ can be computed using numerical integration as described in Section I. If one could find a vector $\bar{\nu}^* \in \mathbb{R}_+^{k+1}$ such that at $M = M^*(\bar{\nu}^*)$ the value of each expression in curly braces in (48) equals zero, and if $M^*(\bar{\nu}^*)$ satisfies all of the familywise Type I error constraints (3) (which, intuitively, one may expect to occur since C^* is assumed to be the set of active familywise Type I error constraints of the original problem), then

$$
\begin{aligned}
& \int E_{\delta_1,\delta_2} L(M^*(\bar{\nu}^*)(Z_1, Z_2, U); \delta_1, \delta_2) d\Lambda(\delta_1, \delta_2) && (49)\\
\geq \;& \inf_{M \in \mathcal{M}_c} \int E_{\delta_1,\delta_2} L(M(Z_1, Z_2, U); \delta_1, \delta_2) d\Lambda(\delta_1, \delta_2) && (50)\\
\geq \;& \inf_{M \in \mathcal{M}} \Bigg[\int E_{\delta_1,\delta_2} L(M(Z_1, Z_2, U); \delta_1, \delta_2) d\Lambda(\delta_1, \delta_2) + \bar{\nu}_p^* \left\{ 1 - \beta - P_{\delta_1^{\min},\delta_2^{\min}}(M \text{ rejects } H_{0C}) \right\} \\
& \quad + \sum_{j \in C^*} \bar{\nu}_j^* \left\{ P_{\delta_{1,j},\delta_{2,j}} (M \text{ rejects any null hypothesis in } \mathcal{H}_{\mathrm{TRUE}}(\delta_{1,j}, \delta_{2,j})) - \alpha \right\} \Bigg], && (51)\\
= \;& \int E_{\delta_1,\delta_2} L(M^*(\bar{\nu}^*)(Z_1, Z_2, U); \delta_1, \delta_2) d\Lambda(\delta_1, \delta_2), && (52)
\end{aligned}
$$

where (50) follows from $M^*(\bar{\nu}^*) \in \mathcal{M}_c$; (51) follows from the argument in Section 6; and, the last line follows from $M^*(\bar{\nu}^*)$ being a minimizer over $M \in \mathcal{M}$ of (48) and the assumption that at $M = M^*(\bar{\nu}^*)$ the value of each expression in curly braces in (48) equals zero. Because (49) and (52) are equal, we have that all the inequalities in the above display are equalities. This implies $M^*(\bar{\nu}^*)$ is an optimal solution to the constrained optimization problem (50), which is identical to (16). However, finding a vector $\bar{\nu}^*$ with the above properties (if one exists) could be computationally challenging for the problems from Section 5.1, each of which has 5 or 6 active constraints. This would require searching over $\bar{\nu}^* \in \mathbb{R}_+^{k+1}$ for $k+1$ equal to 5 or 6, to find a vector $\bar{\nu}^*$ such that at $M = M^*(\bar{\nu}^*)$ the value of each expression in curly braces in (48) equals zero, and such that $M^*(\bar{\nu}^*)$ satisfies all of the familywise Type I error constraints (3). We emphasize that in general the set of active familywise Type I error constraints will be unknown, so the above approach (given for the sake of illustration, to compare to standard methods) would not be directly applicable in our problems.

M Generalization and Encoding of Monotonicity Properties

We generalize the monotonicity properties (a)-(d) from Section 5.2 to the randomized multiple testing procedures $\mathcal{M}$. We then show how these properties can be encoded as sparse constraints in the discretized problem. Since none of these ideas were used in the results in the main paper (and we only mentioned these ideas in Section 10 as a potential direction for future research), we only give a rough sketch of these ideas below.

We next generalize the monotonicity properties (a)-(d) from Section 5.2 to the randomized multiple testing procedures $\mathcal{M}$. For any $M \in \mathcal{M}$, $H \subseteq \mathcal{H}$, and $(z_1, z_2) \in \mathbb{R}^2$, define $p(M, H, z_1, z_2) = \int_0^1 1[H \cap M(z_1, z_2, u) \neq \emptyset] du$. Intuitively, $p(M, H, z_1, z_2)$ represents the probability that $M(Z_1, Z_2, U)$ rejects at least one null hypothesis in H, conditioned on $Z_1 = z_1, Z_2 = z_2$. For any $M \in \mathcal{M}$ and any $R \subseteq \mathbb{R}^2$, define the following monotonicity properties with respect to R: for any $(z_1, z_2) \in R$,

- a'. $p(M, \{H_{01}\}, z_1, z_2) \leq p(M, \{H_{01}\}, z_1', z_2)$ for any $(z_1', z_2) \in R$ for which $z_1' \geq z_1$;
- b'. $p(M, \{H_{02}\}, z_1, z_2) \leq p(M, \{H_{02}\}, z_1, z_2')$ for any $(z_1, z_2') \in R$ for which $z_2' \geq z_2$;
- c'. $p(M, \{H_{0C}\}, z_1, z_2) \leq p(M, \{H_{0C}\}, z_1', z_2')$ for any $(z_1', z_2') \in R$ for which $z_1' \geq z_1, z_2' \geq z_2$;
- d'. $p(M, \{H_{01}, H_{02}, H_{0C}\}, z_1, z_2) \leq p(M, \{H_{01}, H_{02}, H_{0C}\}, z_1 + x, z_2 + x)$ for any $x > 0$ such that $(z_1 + x, z_2 + x) \in R$.

In the special case that M is a deterministic procedure, i.e., $M \in \mathcal{M}_{det}$, the above properties are equivalent to monotonicity properties (a)-(d) in Section 5.2. This follows since for any $M \in \mathcal{M}_{det}$, $H \subseteq \mathcal{H}$, $(z_1, z_2) \in \mathbb{R}^2$, we have $p(M, H, z_1, z_2) = 1[H \cap M(z_1, z_2) \neq \emptyset] \in \{0, 1\}$; this implies each of the above properties reduces to the corresponding monotonicity property in Section 5.2.

We next sketch the argument that in the discretized problem, the above monotonicity properties can be encoded as sparse constraints. Set the region R in the definition of the monotonicity properties above to be B. We assume $\tau_1 = \tau_2$ and b/τ_1 is an integer. (To handle that the upper boundary and right boundary of B are not covered by $\mathcal{R}$, due to the rectangles being defined using half-open intervals, i.e., $R_{k,k'} = [k\tau_1, (k+1)\tau_1) \times [k'\tau_2, (k'+1)\tau_2)$, we redefine the rectangles on these borders to be closed, i.e., $[k\tau_1, (k+1)\tau_1] \times [k'\tau_2, (k'+1)\tau_2]$; this ensures that the rectangles in $\mathcal{R}$ completely cover B. This change has no impact on the solutions to the optimization problems due to these boundaries having measure zero with respect to Lebesgue measure in $\mathbb{R}^2$; we set this so that our encoded constraints below imply the monotonicity properties hold for the entire region B.)

First consider monotonicity property (a'). For any rectangle $R_{k,k'} \in \mathcal{R}$, we have that $R_{k+1,k'}$ is the rectangle immediately to the right of $R_{k,k'}$. Define the following constraints: For any $r = R_{k,k'} \in \mathcal{R}$ such that $r' = R_{k+1,k'} \in \mathcal{R}$, $\sum_{s \in \mathcal{S}: H_{01} \in s} m_{rs} \leq \sum_{s \in \mathcal{S}: H_{01} \in s} m_{r's}$. These constraints encode that for each rectangle $r \in \mathcal{R}$ for which the rectangle r' immediately to the right is in $\mathcal{R}$ (i.e., r is not on the right side boundary of the region B), the probability of rejecting H_{01} conditional on $(Z_1, Z_2) \in r'$ is at least the probability of rejecting H_{01} conditional on $(Z_1, Z_2) \in r$. Any multiple testing procedure $M = \{m_{rs}\}_{r \in \mathcal{R}, s \in \mathcal{S}}$ in $\mathcal{M}_{\mathcal{R}}$ that satisfies these constraints must also satisfy monotonicity property (a') with respect to $R = B$. To show this, consider any $(z_1, z_2) \in B$. Let r be the rectangle in $\mathcal{R}$ that contains (z_1, z_2). Consider any $(z_1', z_2) \in B$ for which $z_1' \geq z_1$. Then either (z_1', z_2) is in r or is in a rectangle $\tilde{r} \in \mathcal{R}$ that can be reached from r by a sequence of moves from each rectangle to the adjacent rectangle immediately to the right. By the above constraints, each rectangle $\tilde{r}' \in \mathcal{R}$ that can be reached by such a sequence of moves to the right satisfies

$\sum_{s\in\mathcal{S}:H_{01}\in s} m_{rs} \le \sum_{s\in\mathcal{S}:H_{01}\in s} m_{\tilde{r}'s}$. Then we have

$$p(M,\{H_{01}\},z_1,z_2) = \sum_{s\in\mathcal{S}:H_{01}\in s} m_{rs} \le \sum_{s\in\mathcal{S}:H_{01}\in s} m_{\tilde{r}s} = p(M,\{H_{01}\},z_1',z_2),$$

which shows monotonicity property (a') holds with respect to $R = B$.

Each of the above constraints is linear in $\{m_{rs}\}_{r\in\mathcal{R},s\in\mathcal{S}}$. Also, each of the above constraints is sparse, since for any $r \in \mathcal{R}$, the corresponding constraint involves at most $2|\mathcal{S}|$ variables among $\{m_{rs}\}_{r\in\mathcal{R},s\in\mathcal{S}}$. Analogous constraints can be constructed to encode monotonicity property (b'). Monotonicity property (c') can be encoded by considering, for each rectangle $r \in \mathcal{R}$, the rectangle immediately to its right and the rectangle immediately above it. Define the following constraints: For any $r = R_{k,k'} \in \mathcal{R}$: if $r' = R_{k+1,k'} \in \mathcal{R}$ then $\sum_{s\in\mathcal{S}:H_{0C}\in s} m_{rs} \le \sum_{s\in\mathcal{S}:H_{0C}\in s} m_{r's}$, and if $r'' = R_{k,k'+1} \in \mathcal{R}$ then $\sum_{s\in\mathcal{S}:H_{0C}\in s} m_{rs} \le \sum_{s\in\mathcal{S}:H_{0C}\in s} m_{r''s}$. A similar argument as above can be used to show any multiple testing procedure $M = \{m_{rs}\}_{r\in\mathcal{R},s\in\mathcal{S}}$ in $\mathcal{M}_{\mathcal{R}}$ that satisfies these constraints must also satisfy monotonicity property (c') with respect to $R = B$.

Monotonicity property (d') is slightly more challenging to encode. Recall we had set $\tau_1 = \tau_2$, so that all the rectangles in $\mathcal{R}$ are squares. We partition each square $R_{k,k'} \in \mathcal{R}$ into a lower right triangle $T^{(l)}_{k,k'} = \{[k\tau_1,(k+1)\tau_1) \times [k'\tau_1,(k'+1)\tau_1)\} \cap \{(x,x') : x'-x < (k'-k)\tau_1\}$ and an upper left triangle $T^{(u)}_{k,k'} = \{[k\tau_1,(k+1)\tau_1) \times [k'\tau_1,(k'+1)\tau_1)\} \cap \{(x,x') : x'-x \ge (k'-k)\tau_1\}$. Let $\mathcal{T}$ denote the set of lower right and upper left triangles resulting from partitioning each $R_{k,k'} \in \mathcal{R}$. We can define variables $\bar{m}_{ts}$ analogous to m_{rs} for each $s \in \mathcal{S}$, except now for each triangle $t \in \mathcal{T}$ instead of for each rectangle $r \in \mathcal{R}$. It is straightforward to reformulate the discretized optimization problem from Section 4 using these variables.

Monotonicity property (d') can be encoded by setting two types of constraints. The first is that for each lower right triangle $t = T^{(l)}_{k,k'} \in \mathcal{T}$, the sum of variables for each $s \in \mathcal{S} \setminus \{\emptyset\}$

is at most the corresponding sum for $T^{(u)}_{k+1,k'}$, which is the upper (left) triangle immediately to the right of t (which shares a boundary with t). The second type of constraint is that for each upper left triangle $t = T^{(u)}_{k,k'} \in \mathcal{T}$, the sum of variables for each $s \in \mathcal{S} \setminus \{\emptyset\}$ is at most the corresponding sum for $T^{(l)}_{k,k'+1}$, which is the lower right triangle immediately above t (which shares a boundary with t). By similar arguments as above, any multiple testing procedure $M = \{\bar{m}_{ts}\}_{t \in \mathcal{T}, s \in \mathcal{S}}$ in $\mathcal{M}_B$ that satisfies these constraints must also satisfy monotonicity property (d') with respect to $R = B$.

It is straightforward to adapt the encodings of properties (a')-(c') given above to the finer discretization using variables $\bar{m}_{ts}$ that correspond to the triangles $\mathcal{T}$. This can be done by converting each constraint for a pair of adjacent rectangles into corresponding constraints for the triangles that make up these rectangles. E.g., each constraint given above for monotonicity property (a') could be converted to the following constraints: For each $t^{(u)} = T^{(u)}_{k,k'} \in \mathcal{T}$, let $t^{(l)} = T^{(l)}_{k,k'}$, and set the constraint

$$\sum_{s \in \mathcal{S}: H_{01} \in s} \bar{m}_{t^{(u)}s} \leq \sum_{s \in \mathcal{S}: H_{01} \in s} \bar{m}_{t^{(l)}s};$$

For each $t^{(l)} = T^{(l)}_{k,k'} \in \mathcal{T}$, let $t^{(u)'} = T^{(u)}_{k+1,k'}$, and if $t^{(u)'} \in \mathcal{T}$ set the constraint

$$\sum_{s \in \mathcal{S}: H_{01} \in s} \bar{m}_{t^{(l)}s} \leq \sum_{s \in \mathcal{S}: H_{01} \in s} \bar{m}_{t^{(u)'}s}.$$

We conjecture that the proof in Section H.1 can be extended to show the following generalization of Theorem 1:

(a.) For any $M \in \mathcal{M} \cap \mathcal{M}_B$ that satisfies (a')-(d') with respect to $R = B$, we have

$$\begin{aligned} &\sup_{(\delta_1,\delta_2)\in\mathbb{R}^2} P_{\delta_1,\delta_2}[M(Z_1, Z_2, U) \cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1, \delta_2) \neq \emptyset] \\ = &\sup_{(\delta_1,\delta_2)\in G} P_{\delta_1,\delta_2}[M(Z_1, Z_2, U) \cap \mathcal{H}_{\mathrm{TRUE}}(\delta_1, \delta_2) \neq \emptyset]. \end{aligned} \tag{53}$$

(b.) For any $M \in \mathcal{M}$ that satisfies (a')-(d') with respect to $R = \mathbb{R}^2$, (53) holds.

Proving the above result is an open problem.

(Only references for the Supplementary Materials that are not given in the main paper are included below.)

References

Alosh, M. and Huque, M. F. (2009). A flexible strategy for testing subgroups and overall population. *Statistics in Medicine*, 28(1):3–23.

Hommel, G. and Bernhard, G. (1999). Bonferroni procedures for logically related hypotheses. *Journal of Statistical Planning and Inference*, 82(1-2):119 – 128.

Marcus, R., Peritz, E., and Gabriel, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63:655–660.