Appendix and Supplementary Material for "Group Testing in Mediation Analysis"

Andriy Derkach, Steve Moore, Simina M. Boca, Joshua N. Sampson

1 High dimensional settings (i.e. $p > n$)

1.1 Notation

We classify the m biomarkers into q predefined disjoint sets, where the m_s biomarkers of set $s \in \{1, ..., q\}$ are indexed by $G_s = \{s_1, ..., s_{m_s}\} \subset \{1, ..., m\}$, and define $s(j)$ to be the set containing biomarker j. We denote the MLE by $\hat{\boldsymbol{\beta}}_s = (\hat{\beta}_{s_1}, ..., \hat{\beta}_{s_{m_s}})'$ and $\hat{\boldsymbol{\gamma}}_s = (\hat{\gamma}_{s_1}, ..., \hat{\gamma}_{s_{m_s}})'$. Both vectors have asymptotically distributed as multivariate normal,

$$\hat{\boldsymbol{\beta}}_s \sim N(\boldsymbol{\beta}_s, \Sigma_{\beta_s}) \text{ and} \tag{1.1}$$

$$\hat{\boldsymbol{\gamma}}_s \sim N(\boldsymbol{\gamma}_s, \Sigma_{\gamma_s}), \tag{1.2}$$

with $corr(\hat{\beta}_j, \hat{\gamma}_{j'}) = 0$ for all $j, j' \in G_s$. Here, we consider three types of null hypotheses:

$$H^0_{Es}: \quad \boldsymbol{\beta}_s = \mathbf{0},$$

$$H^0_{Ys}: \quad \boldsymbol{\gamma}_s = \mathbf{0}$$

and

$$H^0: \quad \beta_{s_j}\gamma_{s_j} = 0 \text{ for } j = 1, ..., m_s.$$

1.2 Group tests

Both marginal ($MARG$) and two-step (TS) procedures are based on variance component tests for no association with a set G_s [2, 10, 13]. For testing the null hypothesis H^0_{Es} of no association between the exposure E_i and the set s of biomarkers $\mathbf{M}_{is} = (M_{is_1}, ..., M_{im_s})'$, we considered two test statistics. The first test statistic is

$$T^s_{E,\gamma} = \sum_{j \in G_s} (\hat{\gamma}_j \hat{\beta}_j)^2, \tag{1.3}$$

where $\hat{\gamma}_j$ for $j \in G_s$ are treated as fixed weights. The second statistic is

$$T^s_{E,\mathbf{1}} = \sum_{j \in G_s} \hat{\beta}^2_j. \tag{1.4}$$

Similarly, for testing the null hypothesis (H^0_{Ys}) of no association between the outcome Y_i and the set of biomarkers, $\mathbf{M}_{is}$, we considered two test statistics. The weighted test statistic is

$$T^s_{Y,\beta} = \sum_{j \in G_s} (\hat{\gamma}_j \hat{\beta}_j)^2, \tag{1.5}$$

with weights fixed at $\hat{\beta}_j$ and the unweighted test statistic is

$$T^s_{Y,\mathbf{1}} = \sum_{j \in G_s} \hat{\gamma}^2_j. \tag{1.6}$$

Let $p^s_{E,\gamma}$, $p^s_{E,\mathbf{1}}$, $p^s_{Y,\beta}$ and $p^s_{Y,\mathbf{1}}$ be corresponding p-values for test statistics $T^s_{E,\gamma}$, $T^s_{E,\mathbf{1}}$, $T^s_{Y,\beta}$ and $T^s_{Y,\mathbf{1}}$. These p-values are calculated from distributions corresponding to a linear combination of chi-squared random variables. [1, 13]. We also let $C(\alpha, \Sigma, \boldsymbol{w})$ be a critical value, such that $P\left(T > C(\alpha, \Sigma, \boldsymbol{w})\right) = \alpha$, where $T = \sum_j w_j S^2_j$ and $\boldsymbol{S} = (S_1, ..., S_J) \sim N(0, \Sigma)$. Under models (1.1) and (1.2), the test statistics $T^s_{E,\gamma}$, $T^s_{E,\mathbf{1}}$, $T^s_{Y,\beta}$ and $T^s_{Y,\mathbf{1}}$ also have the following properties

Proposition 1, *Independence*:

1) Under H^0_{Es}: $T^s_{E,\gamma}$ and $T^s_{Y,\mathbf{1}}$ are independent.

2) Under H^0_{Ys}: $T^s_{Y,\beta}$ and $T^s_{E,\mathbf{1}}$ are independent.

3) $T^s_{Y,\mathbf{1}}$ and $T^s_{E,\mathbf{1}}$ are independent.

Proof. Proof of statement 3 is trivial given the independence between biomarker specific estimates $\hat{\beta}_s$ and $\hat{\gamma}_s$. Proof of statement 1 (i.e. similarly for statement 2) of proposition follows

$$P(p^s_{E,\gamma} < t_1, p^s_{Y,\mathbf{1}} < t_2) = E\left[P(p^s_{E,\gamma} < t_1, p^s_{Y,\mathbf{1}} < t_2 | \hat{\gamma}_s)\right] \tag{1.7}$$

$$= E\left[P(p^s_{E,\gamma} < t_1 | \hat{\gamma}_s) P\left(p^s_{Y,\mathbf{1}} < t_2\right)\right] \overset{H^0_{E^s}}{=} t_1 P\left(p^s_{Y,\mathbf{1}} < t_2\right). \tag{1.8}$$

□

We note that $T^s_{E,\gamma}$ and $T^s_{Y,\beta}$ are not independent as the rejection region $\left\{p^s_{E,\gamma} < t_1, p^s_{Y,\beta} < t_2\right\}$ depends jointly on $\hat{\boldsymbol{\beta}}_s$ and $\hat{\boldsymbol{\gamma}}_s$. Thus, MARG uses $T^s_{E,\mathbf{1}}$ and $T^s_{Y,\mathbf{1}}$ when testing the associations between group of biomarkers and E, Y.

1.3 Post-selection adjustment of p-values

Post-selection testing following aggregated association tests has been developed in the context of meta-analysis and gene based testing [4, 5]. Here, we extend their post-selection testing approach (see Theorem 3.1 of [5]) to recalculate $p^C_{E,j}$ for $j \in G_s$, the p-values conditioned on the first selecting set (i.e. G_s) with $p^s_{E,\gamma} < 0.025$ and $p^s_{Y,\mathbf{1}} < 0.1$ (with an equivalent proposition for $p^C_{Y,j}$).

Proposition 2, *Post-selection distribution*

1. For $j \in G_s$,

$$\hat{\beta}_j | \left\{T^s_{E,\gamma} > C(0.025, \Sigma_{\boldsymbol{\beta}_s}, \hat{\boldsymbol{\gamma}}_s), T^s_{Y,\mathbf{1}} > C\left(0.1, \Sigma_{\boldsymbol{\gamma}_s}, \mathbf{1}\right), \hat{\boldsymbol{\gamma}}_s, \boldsymbol{V}_{\beta,s}(j)\right\} \sim TN\left(\beta_j, \hat{\sigma}^2_{\beta j}, A_\beta\left(\boldsymbol{V}_{\beta,s}(j)\right), B_\beta\left(\boldsymbol{V}_{\beta,s}(j)\right)\right)$$

and

2. For $j \in G_s$,

$$\hat{\gamma}_j | \left\{T^s_{Y,\beta} > C(0.025, \Sigma_{\boldsymbol{\gamma}_s}, \hat{\boldsymbol{\beta}}_s), T^s_{E,\mathbf{1}} > C(0.1, \Sigma_{\boldsymbol{\beta}_s}, \mathbf{1}), \hat{\boldsymbol{\beta}}_s, \boldsymbol{V}_{\gamma,s}(j)\right\} \sim TN\left(\gamma_j, \hat{\sigma}^2_{\gamma j}, A_\gamma\left(\boldsymbol{V}_{\gamma,s}(j)\right), B_\gamma\left(\boldsymbol{V}_{\gamma,s}(j)\right)\right),$$

where $TN(\mu, \sigma^2, a, b)$ is normal distribution with mean μ and variance σ^2 truncated at $(-\inf, a] \cup [b, +\inf)$. Orthogonal vectors $\boldsymbol{V}_{\beta,s}(j)$, $\boldsymbol{V}_{\gamma,s}(j)$ (i.e. independent of $\hat{\beta}_j$ and $\hat{\gamma}_j$) and $A_\beta\left(\boldsymbol{V}_{\beta,s}(j)\right)$, $B_\beta\left(\boldsymbol{V}_{\beta,s}(j)\right)$, $A_\gamma\left(\boldsymbol{V}_{\gamma,s}(j)\right)$ and $B_\gamma\left(\boldsymbol{V}_{\gamma,s}(j)\right)$, functions of $\boldsymbol{V}_{\beta,s}(j)$ and $\boldsymbol{V}_{\gamma,s}(j)$ are defined in Theorem 3.1 of Heller et al. [5].

Proof. For $\hat{\beta}_j$ (and similarly for $\hat{\gamma}_j$), the conditional distribution is simplified to

$$\hat{\beta}_j | \left\{ T^s_{E,\gamma} > C(0.025, \Sigma_{\boldsymbol{\beta}_s}, \hat{\boldsymbol{\gamma}}_s), T^s_{Y,\mathbf{1}} > C\left(0.1, \Sigma_{\boldsymbol{\gamma}_s}, \mathbf{1}\right), \hat{\boldsymbol{\gamma}}_s, \boldsymbol{V}_{\beta,s}(j) \right\} = \hat{\beta}_j | \left\{ T^s_{E,\gamma} > C(0.025, \Sigma_{\boldsymbol{\beta}_s}, \hat{\boldsymbol{\gamma}}_s), \hat{\boldsymbol{\gamma}}_s, \boldsymbol{V}_{\beta,s}(j) \right\},$$

because $T^s_{Y,\mathbf{1}} = \sum_{j \in G_s} \hat{\gamma}_j^2$. The remainder of the proof follows from Theorem 3.1 of Heller et al. [5] and independence of $\hat{\boldsymbol{\gamma}}_s$ and $\hat{\boldsymbol{\beta}}_s$. □

We denote the conditional p-values for testing $\beta_j = 0$ and $\gamma_j = 0$ for $j \in G_s$ by $p^C_{E,j}$ and $p^C_{Y,j}$.

1.4 Confidence intervals for the estimated mediation effects

TS method does not only select potentially mediating sets but also it determines biomarkers that are driving associations. Under certain assumptions [7, 8, 9, 12] and assumption of Theorem 1, natural indirect effect of mediator j is equal to $NIE(j) = \beta_j \gamma_j$ for the continuous outcome and rare binary outcome model. We estimate $NIE(j)$ for selected biomarker by product of the MLEs $\hat{\beta}_j$ and $\hat{\gamma}_j$ from models (3) and (5). We note that conditional distributions of $\hat{\beta}_j$ and $\hat{\gamma}_j$ on selection by group tests (see equation 1 and 2 in Supplemental Section 1.3) are not independent because weighting in $T^s_{E,\gamma}$ and $T^s_{Y,\beta}$. Thus, exact methods defined in [5] are not applicable for modeling the joint distribution of $\hat{\beta}_j$ and $\hat{\gamma}_j$. We propose a new numerical approach to calculate confidence intervals while adjusting for selection of a set by group tests (i.e. $p^s_{E,\gamma} < 0.025$ and $p^s_{Y,1} < 0.1$; $p^s_{E,1} < 0.1$ and $p^s_{E,\gamma} < 0.025$. Let $\boldsymbol{V}_{\beta,s}(j)$ and $\boldsymbol{V}_{\gamma,s}(j)$ be orthogonal vector (i.e independent of) to $\hat{\beta}_j$ and $\hat{\gamma}_j$ effect estimates of the j^{th} biomarker of selected set. Estimates $\hat{\boldsymbol{\beta}}_s$ and $\hat{\boldsymbol{\gamma}}_s$ can be rewritten in terms of $\boldsymbol{V}_{\beta,s}(j)$ and $\boldsymbol{V}_{\gamma,s}(j)$

$$\hat{\boldsymbol{\beta}}_s = \hat{\beta}_j + c_\beta \boldsymbol{V}_{\beta,s}(j) \text{ and } \hat{\boldsymbol{\gamma}}_s = \hat{\gamma}_j + c_\gamma \boldsymbol{V}_{\gamma,s}(j), \tag{1.9}$$

where c_β and c_γ are functions of covariance matrix of $\hat{\boldsymbol{\beta}}_s$ and $\hat{\boldsymbol{\gamma}}_s$ and they are considered to be fixed (see [5]). We use these relationships to regenerate new vectors $\boldsymbol{\beta}_s^r$ and $\boldsymbol{\gamma}_s^r$ by simulating effects β_j^r and γ_j^r for a single SNP j from normal distributions with observed means and variances and keeping observed values $V_{\beta,s}(j)$ and $\boldsymbol{V}_{\gamma,s}(j)$ constant. Then, confidence intervals are calculated from β_j^r and γ_j^r in subset of replicates with group tests p-values satisfying our selection criteria, i.e. $p^s_{E,\gamma} < 0.025$, $p^s_{Y,1} < 0.1$, $p^s_{E,1} < 0.1$ and $p^s_{E,\gamma} < 0.025$. Similar procedure was proposed by [5], which is a direct consequence of our Proposition 2. We note that proposed approach can be extended to biomarkers selected by LIN method (linear). Lastly, this approach does not adjust for multiplicity testing (i.e. winner's curse).

1.5 FWER control

Here, we follow similar steps to [11] and show that Theorem 1 holds. We start by defining four sets of biomarkers, w_{00}, w_{01}, w_{10} and w_{11}, where $w_{xy} = \{j : sign(|\beta_j|) = x, sign(|\gamma_j|) = y\}$. We let $w_{\cdot 0} = w_{00} \bigcup w_{10}$, $w_{0\cdot} = w_{00} \bigcup w_{01}$, $w_{\cdot 1} = w_{01} \bigcup w_{11}$, $w_{1\cdot} = w_{10} \bigcup w_{11}$ and $w_{\emptyset} = w_{00} \bigcup w_{10} \bigcup w_{01}$. We also define the sets of selected sets at the first step $S_E = \{s : p^s_{E,\gamma} < 0.025, p^s_{Y,\mathbf{1}} < 0.1\}$, the set of selected biomarkers $G_{E1} = \bigcup_{s \in S_E} G_s$ and the set of candidate biomarkers $G_{E2} = \{j \in G_{E1} : p_{Y,j} < 0.025\}$ (and similarly for Y, $S_Y = \{s : p^s_{Y,\gamma} < 0.025, p^s_{E,\mathbf{1}} < 0.1\}$, $G_{Y1} = \bigcup_{s \in S_Y} G_s$ and $G_{Y2} = \{j \in G_{Y1} : p_{E,j} < 0.025\}$). Furthermore, we set $W = 1$ if $p_{E,j} < 0.025$ for $\forall j \in w_{1\cdot}$ and $p_{Y,j} < 0.025$ for $\forall j \in w_{\cdot 1}$.

In the second step of our procedure, for biomarkers in the sets G_{E1} and G_{Y1}, we define the marker specific thresholds $\alpha^*_{E,j}$ for $j \in G_{E1}$ and $\alpha^*_{Y,j}$ for $j \in G_{Y1}$. We reject the null hypothesis of no association with E if $p^C_{E,j} \leq \alpha^*_{E,j}$ and the null hypothesis of no association with Y if $p^C_{Y,j} \leq \alpha^*_{Y,j}$. Here, we propose to define $\alpha^*_{E,j} = \left(\hat{\gamma}_j^2 / \sum_{j \in G_{E2}} \hat{\gamma}_j^2\right) \alpha/2 = m_{E,j}\alpha/2$ for $j \in G_{E2}$ and $\alpha^*_{E,j} = 0$ otherwise. Similarly, $\alpha^*_{Y,j} = \left(\hat{\beta}_j^2 / \sum_{j \in G_{Y2}} \hat{\beta}_j^2\right) \alpha/2 = m_{Y,j}\alpha/2$ for $j \in G_{Y2}$ and $\alpha^*_{Y,j} = 0$ otherwise. We note that these thresholds are functions of $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_1, ..., \hat{\boldsymbol{\gamma}}_q)'$ or $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_q)'$ for the candidate biomarkers. We define the number of falsely discovered sets as $\sum_{s \in H^0} r_s$, where $r_s = 1$ if TS rejects H^0 for set s and 0 otherwise. Now, in this section, we demonstrate asymptotic control of FWER (i.e. $P(\sum_{s \in H_0} r_s > 0)$), under the following assumption.

Assumption 1, *Block Independence* For any $s, s' \in \{1, ..., q\}$, $\boldsymbol{M}_s$ is independent of $\boldsymbol{M}_{s'}$ given E_i.

Theorem 1, *FWER Control*

If assumption 1 holds then $\lim_{n\to\infty} FWER \leq \alpha$.

Proof. Clearly, $P(W=1) \to 1$ as sample size $n \to \infty$. The number of falsely discovered biomarkers R is calculated as sum of rejected tests in selected groups:

$$R = \sum_{s=1}^{q} \left\{ I(p^s_{Y,\beta} < 0.025, p^s_{E,\mathbf{1}} < 0.1) I(p^s_{E,\gamma} < 0.025, p^s_{Y,\mathbf{1}} < 0.1) \sum_{j \in G_s \cap w_\emptyset} I(p^C_{E,j} < \alpha^*_{E,j}, p^C_{Y,j} < \alpha^*_{Y,j}) \right\} = \sum_{s=1}^{q} R_s \tag{1.10}$$

We can then demonstrate asymptotic control of FWER, $P(\sum_{s\in H_0} r_s > 0) \leq P(\sum_{s\in H_0} R_s > 0) \leq P(R > 0)$, using assumption 1 and propositions 1 and 2. It follows from the Bonferroni inequality that

$$FWER = P(R>0) \leq E\left[\sum_{s=1}^{q} \left\{ I(p^s_{Y,\beta} < 0.025, p^s_{E,\mathbf{1}} < 0.1) I(p^s_{E,\gamma} < 0.025, p^s_{Y,\mathbf{1}} < 0.1) \sum_{j \in G_s \cap w_\emptyset} I(p^C_{E,j} < \alpha^*_{E,j}, p^C_{Y,j} < \alpha^*_{Y,j}) \right\} \right]$$

Then, we split this expectations into two parts and bound it using our observations that $P(W=1) \to 1$ as sample size $n \to \infty$:

$$= E\left[\sum_{s=1}^{q} \left\{ I(p^s_{Y,\beta} < 0.025, p^s_{E,\mathbf{1}} < 0.1) I(p^s_{E,\gamma} < 0.025, p^s_{Y,\mathbf{1}} < 0.1) \sum_{j \in G_s \cap w_\emptyset} I(p^C_{E,j} < \alpha^*_{E,j}, p^C_{Y,j} < \alpha^*_{Y,j}) \right\} | W=1 \right] P_w$$

$$+E\left[\sum_{s=1}^{q} \left\{ I(p^s_{Y,\beta} < 0.025, p^s_{E,\mathbf{1}} < 0.1) I(p^s_{E,\gamma} < 0.025, p^s_{Y,\mathbf{1}} < 0.1) \sum_{j \in G_s \cap w_\emptyset} I(p^C_{E,j} < \alpha^*_{E,j}, p^C_{Y,j} < \alpha^*_{Y,j}) \right\} | W=0 \right] (1-P_w)$$

$$< E\left[\sum_{s=1}^{q} \left\{ I(p^s_{Y,\beta} < 0.025, p^s_{E,\mathbf{1}} < 0.1) I(p^s_{E,\gamma} < 0.025, p^s_{Y,\mathbf{1}} < 0.1) \sum_{j \in G_s \cap w_\emptyset} I(p^C_{E,j} < \alpha^*_{E,j}, p^C_{Y,j} < \alpha^*_{Y,j}) \right\} | W=1 \right] + \epsilon,$$

where $\epsilon = o(1)$ as $n \to \infty$. Next, we split the above expectation into three components (i.e. $w_\emptyset =$

$w_{01} \cup w_{10} \cup w_{00})$

$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{Y,\beta}<0.025, p^s_{E,\mathbf{1}}<0.1)I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_\emptyset} I(p^C_{E,j}<\alpha^*_{E,j}, p^C_{Y,j}<\alpha^*_{Y,j})\right\}|W=1\right] =$$
$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{Y,\beta}<0.025, p^s_{E,\mathbf{1}}<0.1)I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} I(p^C_{E,j}<\alpha^*_{E,j}, p^C_{Y,j}<\alpha^*_{Y,j})\right\}|W=1\right] +$$
$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{Y,\beta}<0.025, p^s_{E,\mathbf{1}}<0.1)I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{10}} I(p^C_{E,j}<\alpha^*_{E,j}, p^C_{Y,j}<\alpha^*_{Y,j})\right\}|W=1\right] +$$
$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{Y,\beta}<0.025, p^s_{E,\mathbf{1}}<0.1)I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{00}} I(p^C_{E,j}<\alpha^*_{E,j}, p^C_{Y,j}<\alpha^*_{Y,j})\right\}|W=1\right].$$

For set w_{01} (and similarly for set w_{10}), we bound expectation

$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{Y,\beta}<0.025, p^s_{E,\mathbf{1}}<0.1)I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} I(p^C_{E,j}<\alpha^*_{E,j}, p^C_{Y,j}<\alpha^*_{Y,j})\right\}|W=1\right]$$
$$\leq E\left[\sum_{s=1}^{q}\left\{I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} I(p^C_{E,j}<\alpha^*_{E,j})\right\}|W=1\right]$$
$$\leq E\left[\sum_{s=1}^{q} I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} E\left[I(p^C_{E,j}<\alpha^*_{E,j})|p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1, \boldsymbol{P}_{E,\gamma}, \boldsymbol{P}_{Y,1}, \hat{\boldsymbol{\gamma}}\right]|W=1\right]$$

where $\boldsymbol{P}_{E,\gamma} = (p^1_{E,\gamma}, ..., p^q_{E,\gamma})$ and $\boldsymbol{P}_{Y,1} = (p^1_{Y,1}, ..., p^q_{Y,1})$. Because conditional p-values are uniformly distributed given all group test p-values (see Assumption 1 and Proposition 2)

$$E\left[\sum_{s=1}^{q} I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} E\left[I(p^C_{E,j}<\alpha^*_{E,j})|p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1, \boldsymbol{P}_{E,\gamma}, \boldsymbol{P}_{Y,1}, \hat{\boldsymbol{\gamma}}\right]|W=1\right]$$
$$\leq E\left[\sum_{s=1}^{q} I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} \alpha^*_{E,j}|W=1\right].$$

Similarly for the second set w_{10}, we bound

$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{Y,\beta}<0.025,p^s_{E,\mathbf{1}}<0.1)I(p^s_{E,\gamma}<0.025,p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{10}}I(p^C_{E,j}<\alpha^*_{E,j},p^C_{Y,j}<\alpha^*_{Y,j})\right\}|W=1\right]$$
$$\leq E\left[\sum_{s=1}^{q}I(p^s_{Y,\beta}<0.025,p^s_{E,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{10}}\alpha^*_{Y,j}|W=1\right]\leq\alpha/2 \text{ as } n\to\infty,$$

with the last inequality coming from the definition of $\alpha^*_{Y,j}$ (i.e. $\alpha^*_{Y,j}=\left(\hat\beta^2_j/\sum_{j\in G_{Y2}}\hat\beta^2_j\right)\alpha/2$). For equation (4.8), we derive similar inequalities,

$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{Y,\beta}<0.025,p^s_{E,\mathbf{1}}<0.1)I(p^s_{E,\gamma}<0.025,p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{00}}I(p^C_{E,j}<\alpha^*_{E,j},p^C_{Y,j}<\alpha^*_{Y,j})\right\}|W=1\right]\leq$$
$$E\left[\sum_{s=1}^{q}\left\{I(p^s_{E,\gamma}<0.025,p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{00},p_{Y,j}<0.025}I(p^C_{E,j}<\alpha^*_{E,j})\right\}|W=1\right]\leq$$
$$E\left[\sum_{s=1}^{q}I(p^s_{E,\gamma}<0.025,p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{00},p_{Y,j}<0.025}\alpha^*_{E,j}|W=1\right].$$

By adding inequalities for w_{01} and w_{00} we observe that

$$E\left[\sum_{s=1}^{q}I(p^s_{E,\gamma}<0.025,p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}}\alpha^*_{E,j}|W=1\right]$$
$$+E\left[\sum_{s=1}^{q}I(p^s_{E,\gamma}<0.025,p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{00},p_{Y,j}<0.025}\alpha^*_{E,j}|W=1\right]\leq$$
$$E\left[\sum_{s=1}^{q}I(p^s_{E,\gamma}<0.025,p^s_{Y,\mathbf{1}}<0.1)\left(\sum_{j\in G_s\cap w_{00},p_{Y,j}<0.025}\alpha^*_{E,j}+\sum_{j\in G_s\cap w_{01}}\alpha^*_{E,j}\right)|W=1\right]\leq\alpha/2,$$

with $n\to\infty$ the last inequality coming from the definition of $\alpha^*_{E,j}$. Final inequality concludes the proof. □

2 Multivariate settings (i.e. $n > p$)

2.1 Notation

Again, we classify the m biomarkers into q predefined disjoint sets, where the m_s biomarkers of set $s \in \{1, ..., q\}$ are indexed by $G_s = \{s_1, ..., s_{m_s}\} \subset \{1, ..., m\}$, and define $s(j)$ to be the set containing biomarker j. Moreover, let $\boldsymbol{M}_{i\backslash j} = \{M_{ij'} : j' \neq j\}$, a vector of all biomarkers without j^{th}. We will say that a biomarker j is a *mediator* if M_{ij} is associated with E_i and, conditional on both E_i and other bomarkers $\boldsymbol{M}_{i\backslash j}$, M_{ij} is associated with Y_i. To formalize this statement, we define the two null hypotheses

$$H^j_{0E} : E_i \perp M_{ij} \tag{2.1}$$

$$H^j_{0Y} : Y_i \perp M_{ij} | E_i, \boldsymbol{M}_{i\backslash j}, \tag{2.2}$$

and say that a biomarker j is a mediator if and only if the two null hypotheses are false. Under assumption of $n > p$, we denote the MLE for m bioamrkers obtained from joint model with all biomarkers $\boldsymbol{M}$, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, ..., \hat{\beta}_m)'$ and $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_1, ..., \hat{\gamma}_m)'$. Both vectors asymptotically distributed as multivariate normal,

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \Sigma_\beta) \text{ and} \tag{2.3}$$

$$\hat{\boldsymbol{\gamma}} \sim N(\boldsymbol{\gamma}, \Sigma_\gamma), \tag{2.4}$$

with $corr(\hat{\beta}_j, \hat{\gamma}_{j'}) = 0$ for all $j, j' = 1, ..., m$. Here, we consider three types of null hypotheses for each set $s \in \{1, ..., q\}$:

$$H^0_{Es} : \quad \boldsymbol{\beta}_s = \mathbf{0},$$

$$H^0_{Ys} : \quad \boldsymbol{\gamma}_s = \mathbf{0}$$

and

$$H^0_s : \quad \beta_{s_j}\gamma_{s_j} = 0 \text{ for } j = 1, ..., m_s.$$

We use the same group tests (MIN, LIN, QUAD, MARG and TS) defined in the Supplemental Section (1.2) with $\hat{\boldsymbol{\beta}}_s = (\hat{\beta}_{s1}, ..., \hat{\beta}_{sm})$ and $\hat{\boldsymbol{\gamma}}_s = (\hat{\gamma}_{s1}, ..., \hat{\gamma}_{sm})$, $s_j \in G_s$ and $\hat{\beta}_{s_j}$ and $\hat{\gamma}_{s_j}$ are MLE estimates of (2.3 and 2.4). Similarly to the Supplemental Section 1.2, under models (2.3) and (2.4), the test statistics $T^s_{E,\gamma}$, $T^s_{E,\mathbf{1}}$, $T^s_{Y,\beta}$ and $T^s_{Y,\mathbf{1}}$ also have the following properties Proposition 3, *Independence under joint model*:

1) Under H^0_{Es}: $T^s_{E,\gamma}$ and $T^s_{Y,\mathbf{1}}$ are independent.

2) Under H^0_{Ys}: $T^s_{Y,\beta}$ and $T^s_{E,\mathbf{1}}$ are independent.

3) $T^s_{Y,\mathbf{1}}$ and $T^s_{E,\mathbf{1}}$ are independent.

2.2 Post-selection adjustment of p-values

In this section, we relax our assumption of independence between any two vectors $\boldsymbol{M}_{is}$ and $\boldsymbol{M}_{is'}$ conditioning on E_i. Here, we extend our post-selection testing approach (see Section 1.3) to recalculate $p^C_{E,j}$ for $j \in G_s$, the p-values conditioned on all selected sets (i.e. G_ss) with $p^s_{E,\gamma} < 0.025$ and $p^s_{Y,\mathbf{1}} < 0.1$ (with an equivalent proposition for $p^C_{Y,j}$).

Proposition 4, *Post-selection distribution under joint model*

1. For $j \in G_s$,

$$\hat{\beta}_j | \left\{ T^{s'}_{E,\gamma} > C(0.025, \Sigma_{\boldsymbol{\beta}_{s'}}, \hat{\boldsymbol{\gamma}}_{s'}), T^{s'}_{Y,\mathbf{1}} > C\left(0.1, \Sigma_{\boldsymbol{\gamma}_{s'}}, \mathbf{1}\right), \hat{\boldsymbol{\gamma}}_{s'}, \boldsymbol{V}_{\beta,s'}(j) \text{ for } s' \in \{1, ..., q\} \right\} \sim TN\left(\beta_j, \hat{\sigma}^2_{\beta j}, C_\beta\left(\boldsymbol{V}_\beta(j)\right)\right)$$

and

2. For $j \in G_s$,

$$\hat{\gamma}_j | \left\{ T^{s'}_{Y,\beta} > C(0.025, \Sigma_{\boldsymbol{\gamma}_{s'}}, \hat{\boldsymbol{\beta}}_{s'}), T^{s'}_{E,\mathbf{1}} > C\left(0.1, \Sigma_{\boldsymbol{\beta}_{s'}}, \mathbf{1}\right), \hat{\boldsymbol{\beta}}_{s'}, \boldsymbol{V}_{\gamma,s'}(j) \text{ for } s' \in \{1, ..., q\} \right\} \sim TN\left(\gamma_j, \hat{\sigma}^2_{\gamma j}, C_\gamma\left(\boldsymbol{V}_\gamma(j)\right)\right)$$

where $TN(\mu, \sigma^2, c)$ is truncated normal distribution with mean μ and variance σ^2 with only possible values from the set c. Orthogonal vectors $\boldsymbol{V}_{\beta,s}(j)$, $\boldsymbol{V}_{\gamma,s}(j)$ (i.e. independent of $\hat{\beta}_j$ and $\hat{\gamma}_j$) and

$$C_\beta\left(\boldsymbol{V}_\beta(j)\right) = \cap_{s' \in \{1,...,q\}}(-\inf, A_\beta\left(\boldsymbol{V}_{\beta,s'}(j)\right)] \ \cup \ [B_\beta\left(\boldsymbol{V}_{\beta,s'}(j)\right), +\inf).$$

Values $A_\gamma\left(\boldsymbol{V}_{\gamma,s'}(j)\right)$ and $B_\gamma\left(\boldsymbol{V}_{\gamma,s'}(j)\right)$ are functions of $\boldsymbol{V}_{\beta,s'}(j)$ and $\boldsymbol{V}_{\gamma,s'}(j)$ defined in Theorem 3.1 of Heller et al. [5] and Section 1.3.

Proof. For $\hat{\beta}_j$ (and similarly for $\hat{\gamma}_j$), the conditional distribution is adjusted for all selected sets with $p^s_{E,\gamma} < 0.025$ and $p^s_{Y,\mathbf{1}} < 0.1$

$$\hat{\beta}_j | \left\{ T^{s'}_{E,\gamma} > C(0.025, \Sigma_{\boldsymbol{\beta}_{s'}}, \hat{\boldsymbol{\gamma}}_{s'}), \hat{\boldsymbol{\gamma}}_{s'}, \boldsymbol{V}_{\beta,s'}(j) \text{ for } s' \in \{1, ..., q\} \right\},$$

as $T^s_{Y,1} = \sum_{j \in G_s} \hat{\gamma}^2_j$. Event $\left\{ T^{s'}_{E,\gamma} > C(0.025, \Sigma_{\boldsymbol{\beta}_{s'}}, \hat{\boldsymbol{\gamma}}_{s'}), \hat{\boldsymbol{\gamma}}_{s'}, \boldsymbol{V}_{\beta,s'}(j) \text{ for } s' \in \{1, ..., q\} \right\}$ is equivalent to

$$\cap_{\text{Selected } s'} \left\{ T^{s'}_{E,\gamma} > C(0.025, \Sigma_{\boldsymbol{\beta}_{s'}}, \hat{\boldsymbol{\gamma}}_{s'}), \hat{\boldsymbol{\gamma}}_{s'}, \boldsymbol{V}_{\beta,s'}(j) \right\} = \cap_{\text{Selected } s'} (-\inf, A_\beta\left(\boldsymbol{V}_{\beta,s'}(j)\right)] \cup [B_\beta\left(\boldsymbol{V}_{\beta,s'}(j)\right), +\inf).$$

Values $A_\beta\left(\boldsymbol{V}_{\beta,s'}(j)\right)$ and $B_\beta\left(\boldsymbol{V}_{\beta,s'}(j)\right)$ are defined in the Section 1.3. □

We denote the conditional p-values for testing $\beta_j = 0$ and $\gamma_j = 0$ for $j \in G_s$ by $p^C_{E,j}$ and $p^C_{Y,j}$. Note, derived post-selection inference is also valid when sets are disjoint.

2.3 FWER control under joint model

We define the number of falsely discovered sets as $\sum_{s \in H^0} r_s$, where $r_s = 1$ if TS rejects H^0 for set s and 0 otherwise. Now, in this section, we demonstrate asymptotic control of FWER (i.e. $P(\sum_{s \in H_0} r_s > 0)$), under general settings.

Theorem 2 *FWER Control under joint model*: $\lim_{n \to \infty} FWER^{TS} \leq \alpha$

Proof. The proof follows the same steps of Theorem 1, with slight modification. For example, for the set

w_{01} (and similarly for other sets), we bound expectation

$$
\begin{aligned}
& E\left[\sum_{s=1}^{q}\left\{I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} I(p^C_{E,j}<\alpha^*_{E,j})\right\}|W=1\right] \\
& \leq E\left[\sum_{s=1}^{q} I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} E\left[I(p^C_{E,j}<\alpha^*_{E,j})|p^s_{E,\gamma}<0.025, \boldsymbol{P}_{E,\gamma}, \boldsymbol{P}_{Y,1}, \hat{\boldsymbol{\gamma}}\right]|W=1\right] \\
& \leq E\left[\sum_{s=1}^{q} I(p^s_{E,\gamma}<0.025, p^s_{Y,\mathbf{1}}<0.1)\sum_{j\in G_s\cap w_{01}} \alpha^*_{E,j}|W=1\right] \leq \alpha/2.
\end{aligned}
$$

where $\boldsymbol{P}_{E,\gamma} = (p^1_{E,\gamma}, ..., p^q_{E,\gamma})$, $\boldsymbol{P}_{Y,1} = (p^1_{Y,1}, ..., p^q_{Y,1})$ and $p^C_{E,j}$ is conditional p-value on all selected sets. The last equality holds because conditional p-values are uniformly distributed given all selected sets (see Proposition 4). □

3 Additional simulation studies

3.1 Sensitivity analysis for presence of latent confounder

We compared the performance of five previously defined procedures (MIN, LIN, QUAD, MARG, TS) for testing sets of biomarkers under null model without and with latent confounder. Simualtion model without confounders is described in the Section 3 of the main paper. For the sensitivity analysis of the confounder effects, we conducted the following set of simulations to estimate the FWER to detect a mediating set of biomarkers:

1. Latent confounder C effecting exposure E and outcome Y, Figure S2A.
2. Latent confounder C effecting E and all biomarkers M , Figure S2B.
3. Latent confounder Ceffecting Y and all biomarkers M, Figure S2C.
4. Latent confounder C effecting all biomarkers M, Figure S2D.

In all these simulation scenarios, we used the same parametric model and parameters as in the main paper without mediating set (i.e., $q_m = 0$ and see Table 1 and Section 3.1). For sensitivity analysis, we assumed

that the size of the confounder effect on E, Y or M is equal to 0.065 in the simulations with the continuous outcome and 0.045 in the simulations with the binary outcome. We generated 10000 simulations per scenario to estimate empirical FWER at a nominal level of $FWER = \alpha = 0.05$.

3.2 Power comparisons with varying proportions of mediators in a set and proportions of variation in outcome explained by mediators

We evaluated the properties of the five testing procedures when the overall effect of exposure on mediators and mediators on outcome was constant while the proportion of mediators in a set was varied. We used the same simulation model as in the main paper with various values of γ_j and β_j and m_M, the number of mediators in a mediating set (with m_s biomarkers). Specifically, we kept the sum of squared effects constant,

$$EV_1 = \sum_{j=1}^{m_M} \beta_j^2 = m_M\beta^2 = 0.01,\ EV_2 = \sum_{j=1}^{m_M} \gamma_j^2 = m_M\gamma^2 = 0.01 \text{ for continious outcome}$$

and

$$EV_1 = m_M\beta^2 = 0.005,\ EV_2 = \sum_{j=1}^{m_M} \gamma_j^2 = m_M\gamma^2 = 0.015 \text{ for binary outcome,}$$

Specific values of γ_j and β_j were obtained from above equations. It was previously shown that sum $\sum_{j=1}^{m_M} \beta_j^2$ is equal to the proportion of variation in outcome explained by m_M uncorrelated biomarkers [2, 3]. This quantity with m_M and m_s determine the power of group tests studied here [2, 3].

In the second set of scenarios, we evaluated the properties of the five testing methods when the proportion of mediators in the set was constant and overall effects (i.e. EV_1 and EV_22) were varied. We used the same simulation model as in the main paper with various values of γ_j and β_j, while proportion of mediators in a mediating set, m_M/m_s was set to 0.1. Overall effects EV_1 and EV_2 were set to be the same in simulations with continuous outcome and the relationship $EV_2 = 3EV_1$ was assumed in simulations with binary outcome.

We generated 1000 simulations per scenario to estimate relationship between power, EV_1, EV_2 and $p_M = n_M/m_s$ at nominal level $FWER = \alpha = 0.05$.

3.3 FWER and power comparisons with $JTV - comp$ method

We compared the FWER and power of four previously defined procedures (MIN, LIN, MARG, TS) and $JTV - comp$, a joint significance test of variance components for composite null [6]. We evaluated proprieties of this test by applying $JTV - comp$ code to the same data simulated as in a main paper (continuous outcome only). $JTV - comp$ is new method that is similar to our MARG procedure. First, it conducts group testing with E and Y using two marginal variance component tests. Then, composite test is conducted on marginal p-values. However, this test is a marginal test, and it does not ensure that there is a common set of mediating biomarkers associated with both exposure and outcome.

3.4 Results

We describe the sensitivity of the procedures to the presence of an unmeasured (latent) confounder affecting biomarkers, exposure, and outcome in Supplemental Figures S4-S7 and S9-S12. The key result is that for all scenarios, TS, MIN, LIN, and QUAD control FWER in the presence of confounder. MARG does not control FWER when one-dimensional sets are present, because confounder transforms these sets into two-dimensional sets.

We describe the power comparison of the procedures when the overall effects of exposure on mediators and mediators on the outcome are constant in Supplemental Figures S17 and S18. We note that the power of TS and MIN decreases as the proportion of mediators in a set increases, and MARG and LIN have constant power. Both TS and MIN test individual markers in a set, thus decrease of effects due to the increase of the proportion of mediators (note: the overall effect of the set is constant) Finally, we use these simulations to compare these procedures with JTV-comp [6] in Supplemental Figures S21 and S22. We note that JTV-comp does not control FWER in the presence of one and two-dimensional sets. This may be due to the small number of sets in the analysis ($q = 20$) making procedure unreliable. JTV-comp has the best power in many simulation settings, but control of FWER makes these results questionable.

References

[1] Chuang, L.-L. and Shih, Y.-S. (2012). Approximated distributions of the weighted sum of correlated chi-squared random variables. *Journal of Statistical Planning and Inference*, 142(2):457–472.

[2] Derkach, A., Lawless, J. F., and Sun, L. (2014). Pooled association tests for rare genetic variants: A review and some new results. *Statistical Science*, 29(2):302–321.

[3] Derkach, A., Zhang, H., and Chatterjee, N. (2017). Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. *Bioinformatics*, 34(9):1506–1513.

[4] Heller, R., Chatterjee, N., Krieger, A., and Shi, J. (2018). Post-selection inference following aggregate level hypothesis testing in large-scale genomic data. *Journal of the American Statistical Association*, 113(524):1770–1783.

[5] Heller, R., Meir, A., and Chatterjee, N. (2019). Post-selection estimation and testing following aggregate association tests. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(3):547–573.

[6] Huang, Y.-T. (2019). Variance component tests of multivariate mediation effects under composite null hypotheses. *Biometrics*.

[7] Imai, K., Keele, L., and Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334.

[8] Imai, K., Keele, L., and Yamamoto, T. (2010b). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71.

[9] MacKinnon, D. P. and Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17(2):144–158.

[10] Neale, B., Rivas, M. A., Voight, B. F., Altshuler, D., Devlin, B., Orho-Melander, M., Kathiresan, S., Purcell, S. M., Roeder, K., and Daly, M. J. (2011). Testing for an unusual distribution of rare variants. *PLoS Genetics*, 7(3):e1001322.

[11] Sampson, J. N., Boca, S. M., Moore, S. C., Heller, R., and Wren, J. (2018). Fwer and fdr control when testing multiple mediators. *Bioinformatics*, 1:7.

[12] Valeri, L., Lin, X., and VanderWeele, T. J. (2014). Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model. *Statistics in Medicine*, 33(28):4875–4890.

[13] Wu, M., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, 89(1):82–93.

4 Supplemental Figures and Tables

Here, we present following additional figures and tables:

Figue S1: Type of sets used in the simulation study.

Figue S2: Simulation models with latent confounder.

Figue S3: FWER comparisons for continuous outcome and no latent confounders.

Figue S4: FWER comparisons for continuous outcome and latent confounder effecting E and Y.

Figue S5: FWER comparisons for continuous outcome and latent confounder effecting E and $\boldsymbol{M}$.

Figue S6: FWER comparisons for continuous outcome and latent confounder effecting $\boldsymbol{M}$ and Y.

Figue S7: FWER comparisons for continuous outcome and latent confounder effecting $\boldsymbol{M}$.

Figue S8: FWER comparisons for binary outcome and no latent confounders.

Figue S9: FWER comparisons for binary outcome and latent confounder effecting E and Y.

Figue S10: FWER comparisons for binary outcome and latent confounder effecting E and $\boldsymbol{M}$.

Figue S11: FWER comparisons for binary outcome and latent confounder effecting $\boldsymbol{M}$ and Y.

Figue S12: FWER comparisons for binary outcome and latent confounder effecting $\boldsymbol{M}$.

Figue S13: Simulations for continuous outcome with $\gamma_j = \beta_j = 0.075$ and $q = 100$ disjoint sets as the baseline.

Figue S14: Simulations for binary outcome with $\gamma_j = 0.1$, $\beta_j = 0.05$ and $q = 100$ disjoint sets as the baseline.

Figue S15: Simulations for continuous outcome with $\gamma_j = \beta_j = 0.065$ and $m_M = 5$ mediating biomarkers as the baseline.

Figue S16: Simulations for binary outcome with $\gamma_j = 0.06$, $\beta_j = 0.04$ and $m_M = 5$ mediating biomarkers as the baseline.

4.1 Graphical description of simulation models

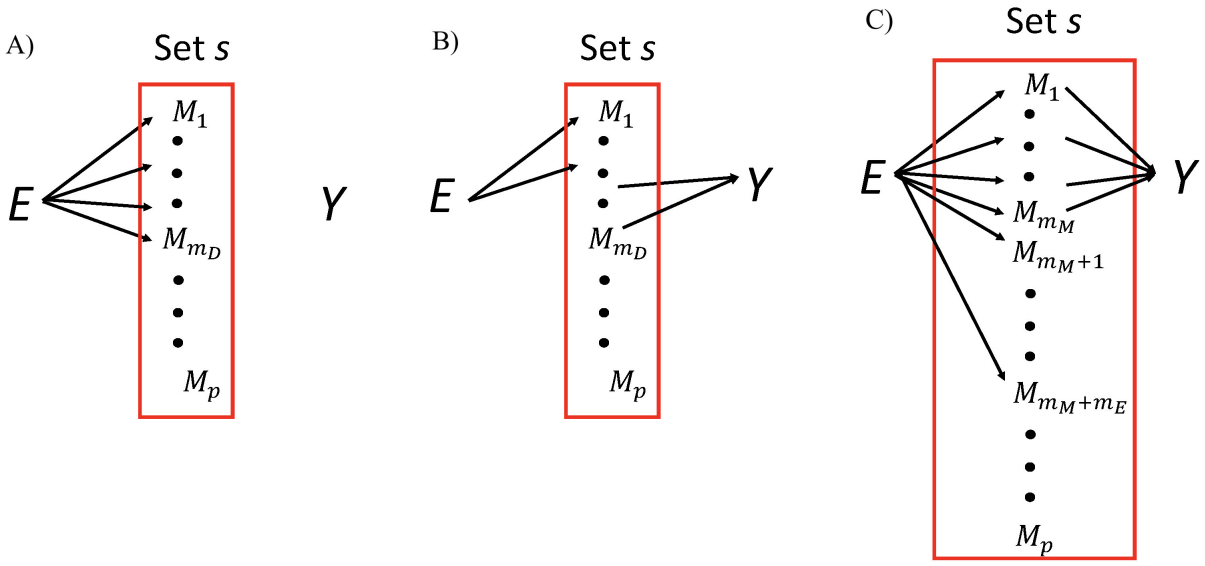

Figure S1: **Type of sets used in the simulation study.** A) *one-dimensional* set with m_D biomarkers associated with exposure; B) *two-dimensional* set with $m_D/2$ biomarkers associated with exposure and $m_D/2$ biomarkers associated with outcome; C) *mediating* set with m_E "noise" biomarkers associated with only exposure.

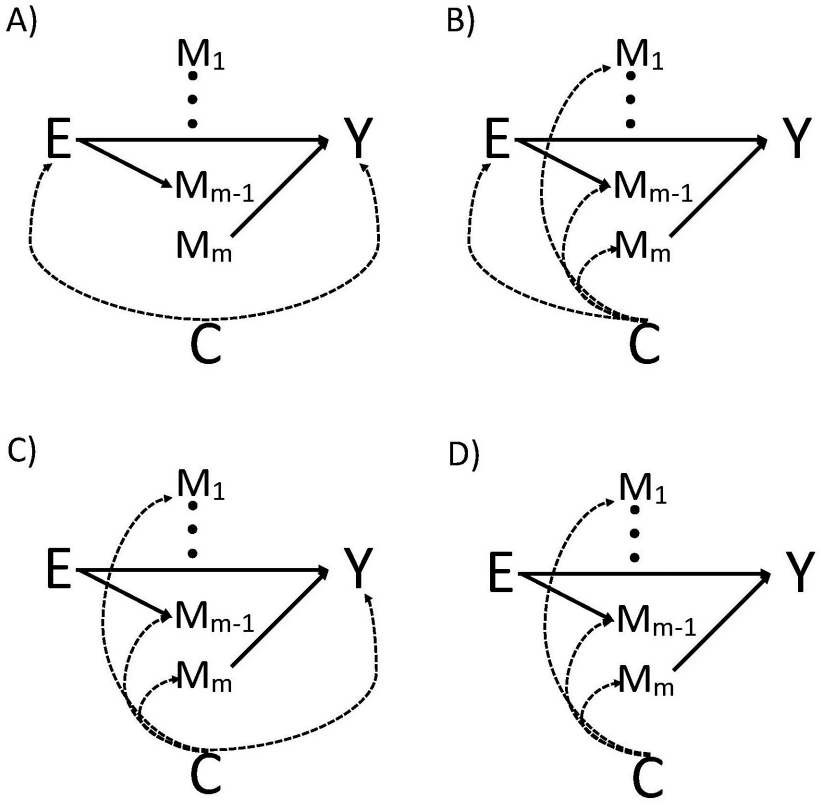

Figure S2: **Simulation models with latent confouder C.** A) confounder affects the exposure and outcome; B) confounder affects the the exposure and all biomarkers; C) confounder affects all biomarkers and the outcome; D) confounder affects all biomarkers.

4.2 Empirical FWER Studies

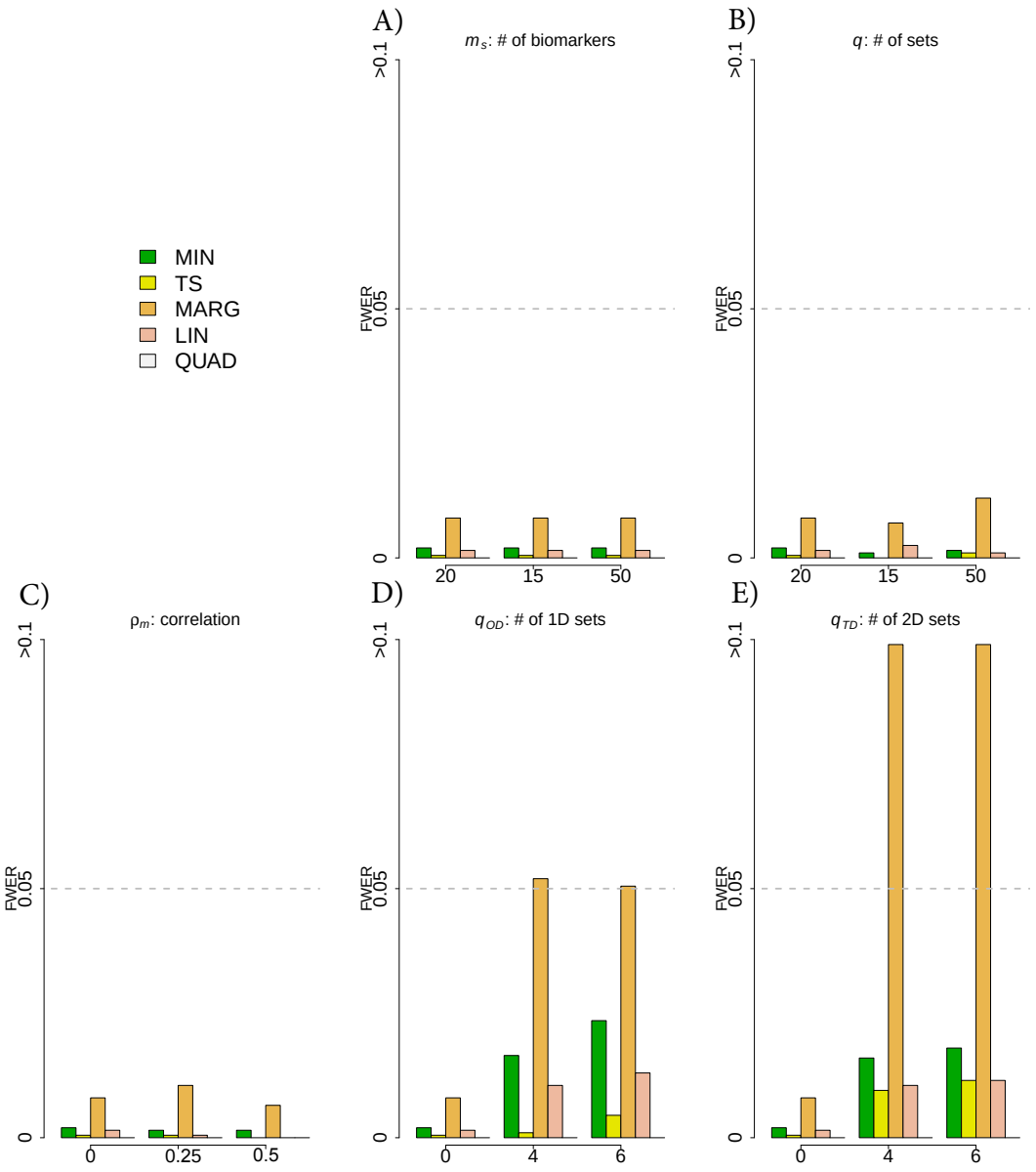

Figure S3: **FWER comparisons for continuous outcome and no latent confounders.** The barplots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

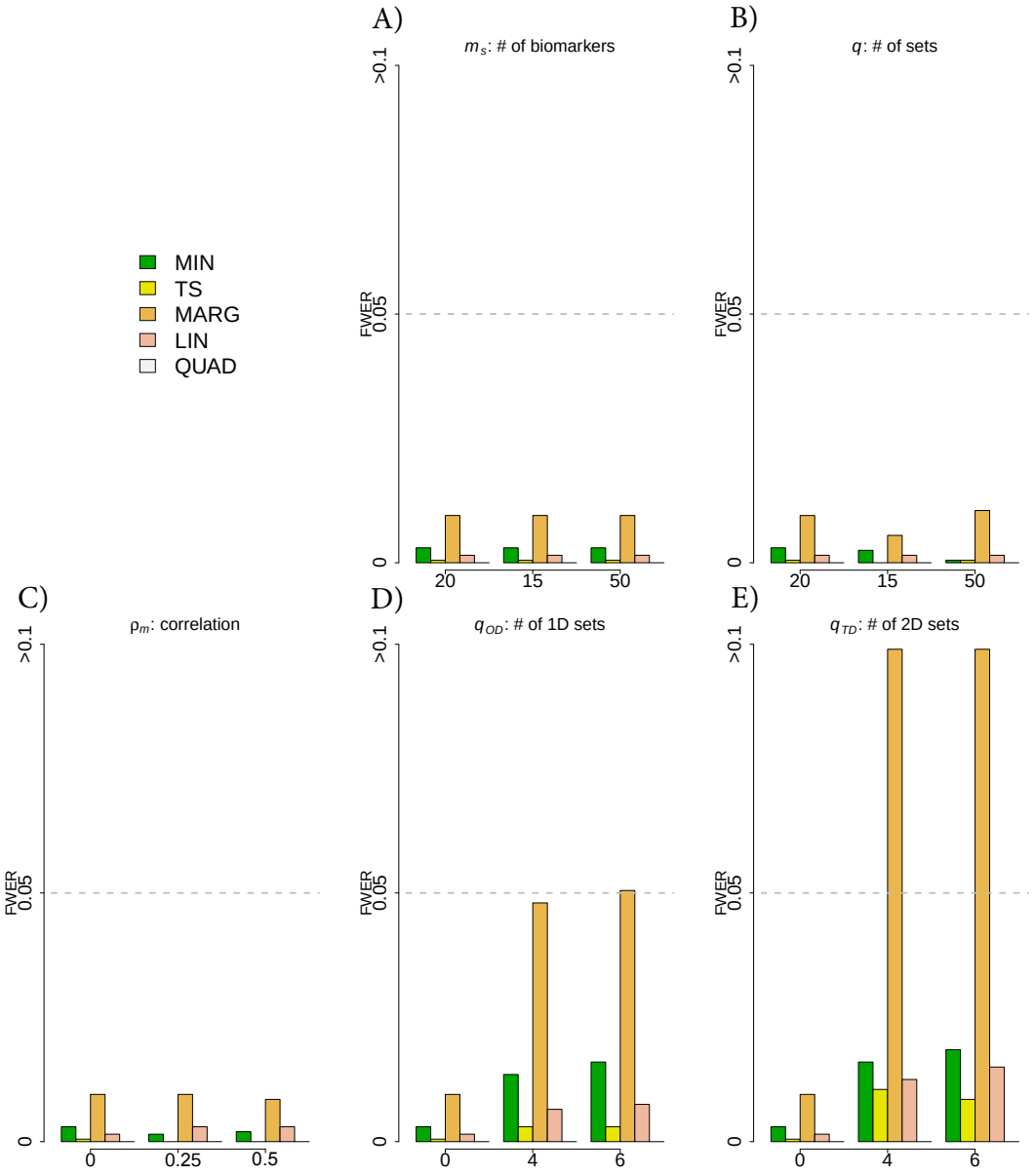

Figure S4: **FWER comparisons for continuous outcome and latent confounder effecting E and Y.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

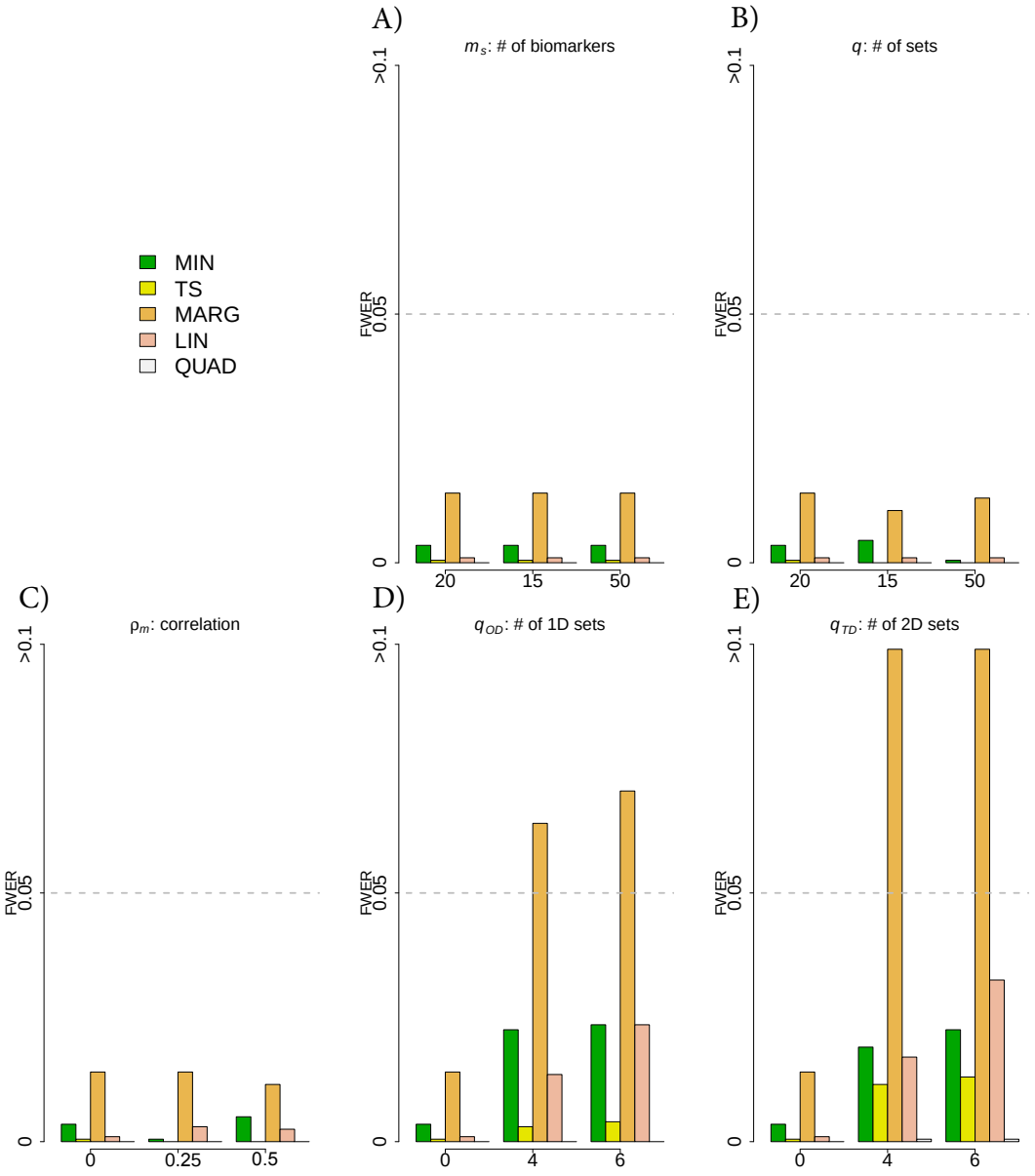

Figure S5: **FWER comparisons for continuous outcome and latent confounder effecting E and M.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

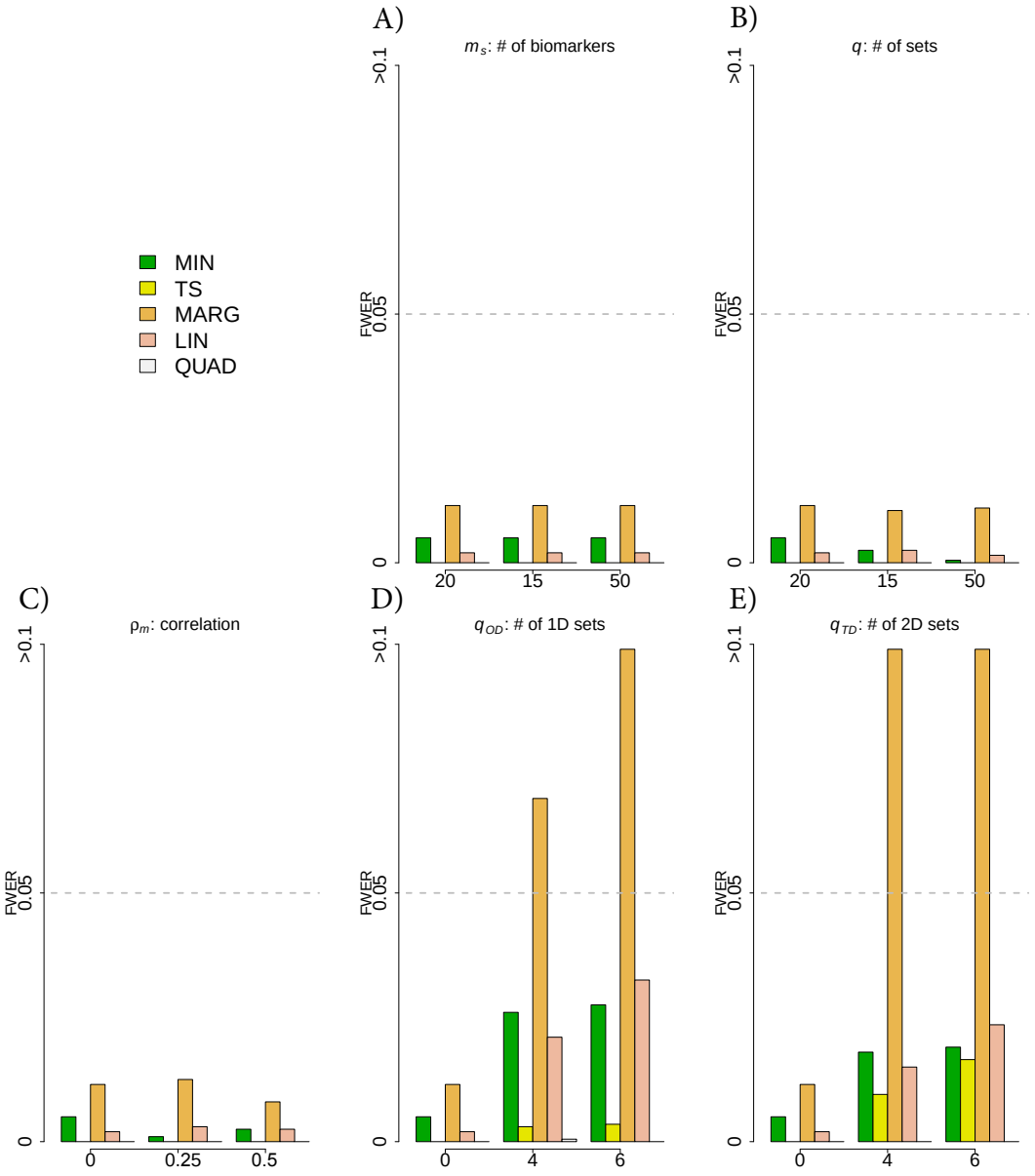

Figure S6: **FWER comparisons for continuous outcome and latent confounder effecting M and Y.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

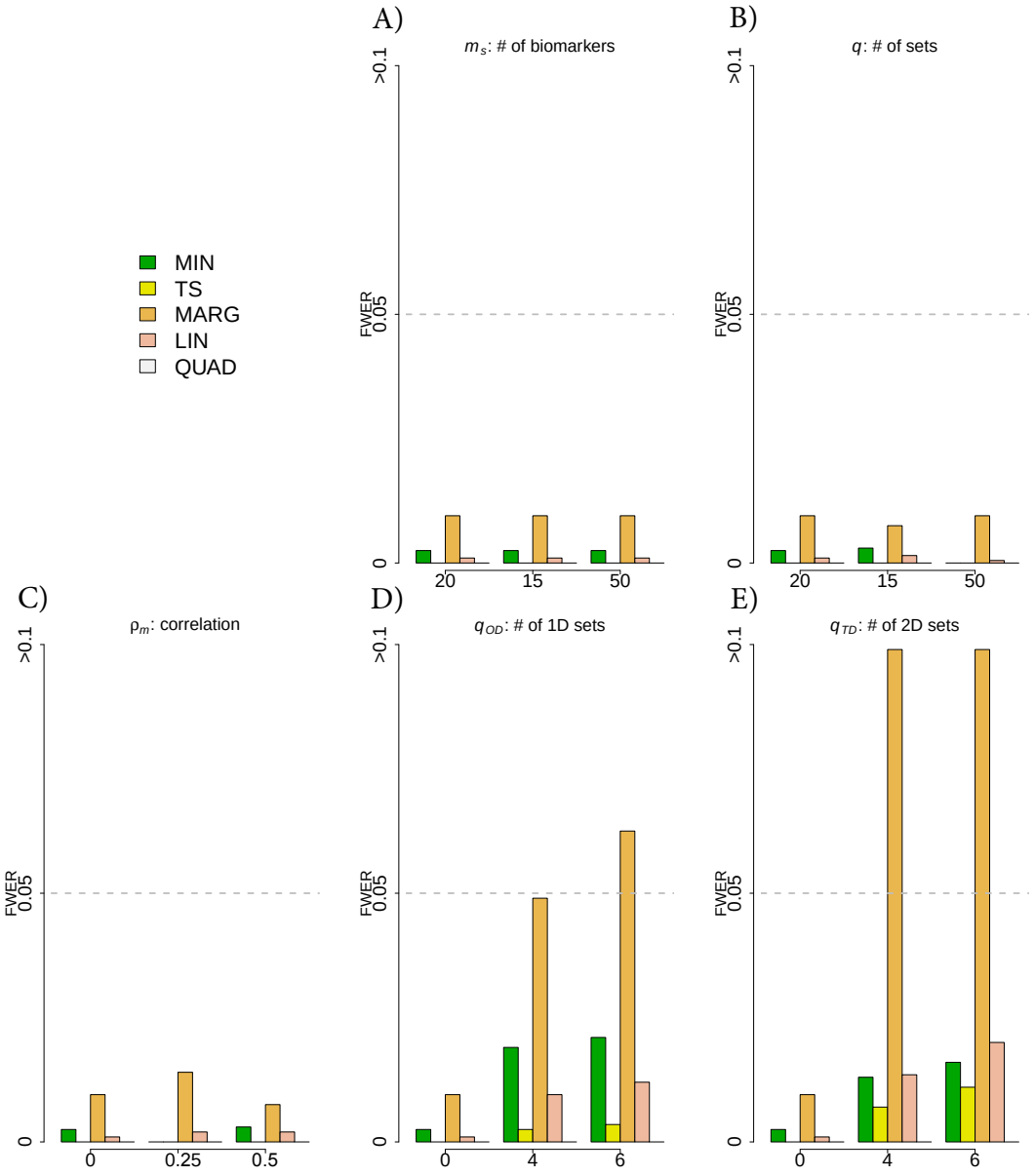

Figure S7: **FWER comparisons for continuous outcome and latent confounder effecting M.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

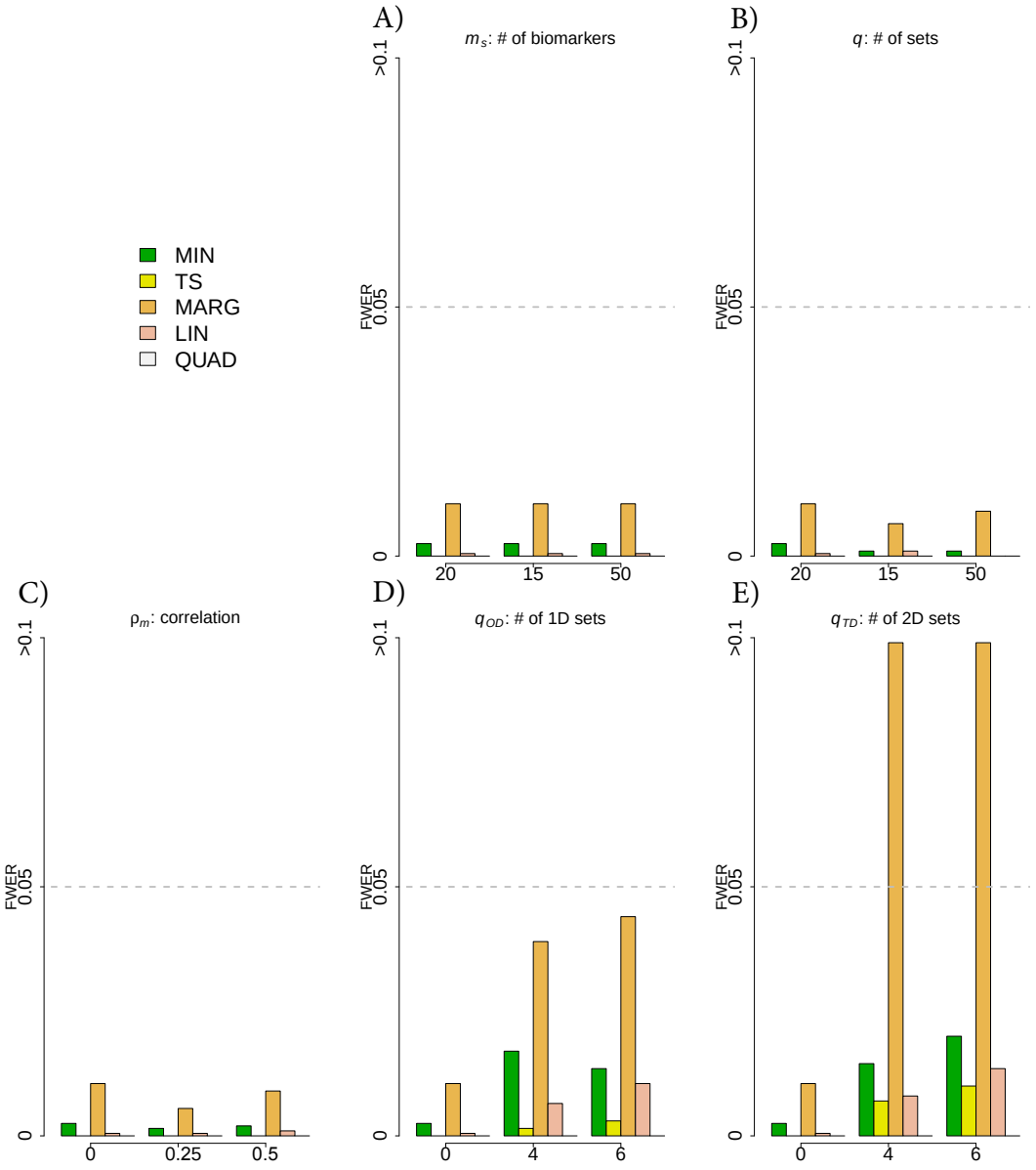

Figure S8: **FWER comparisons for binary outcome and no latent confounders.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

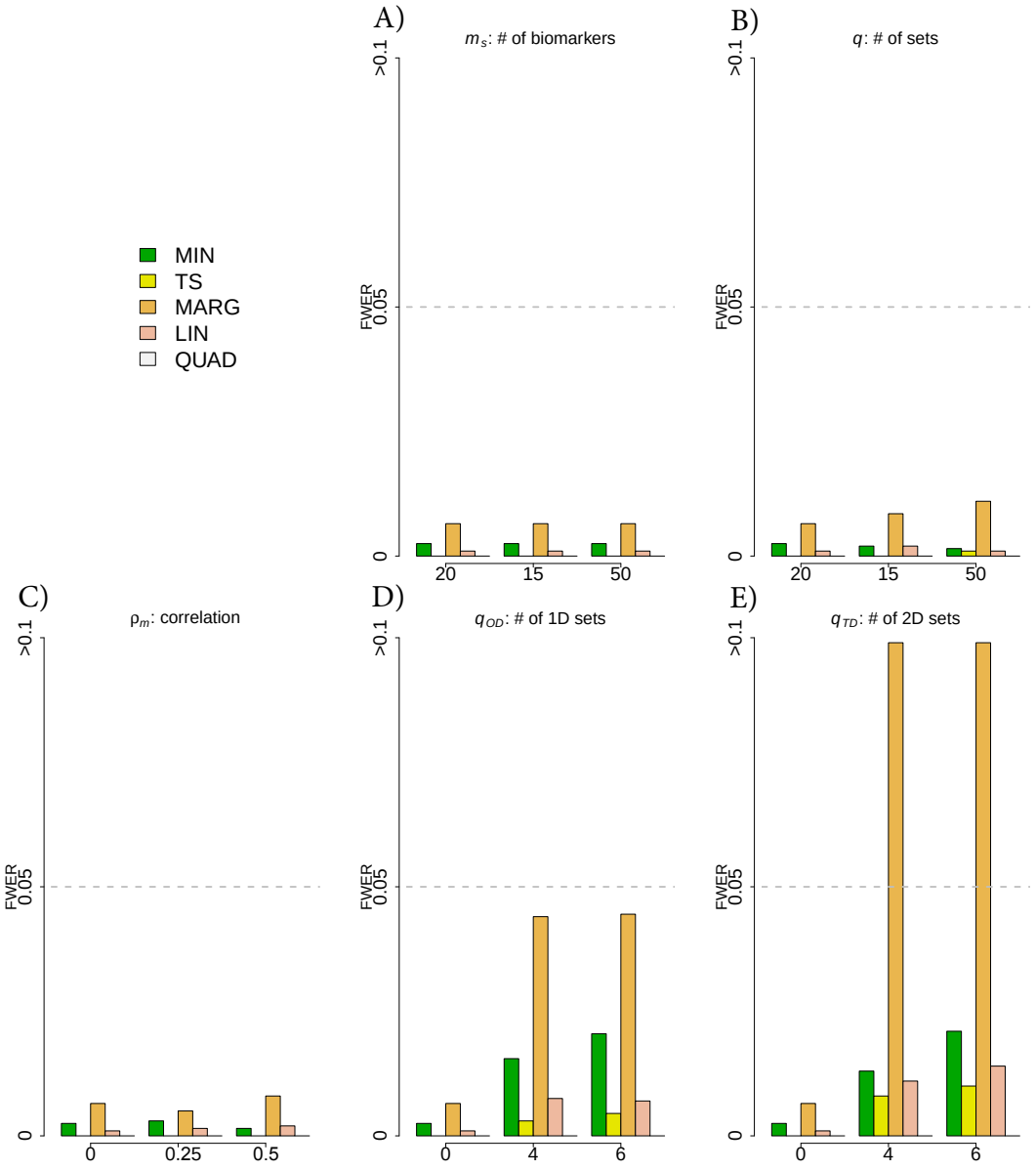

Figure S9: **FWER comparisons for binary outcome and latent confounder effecting E and Y.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

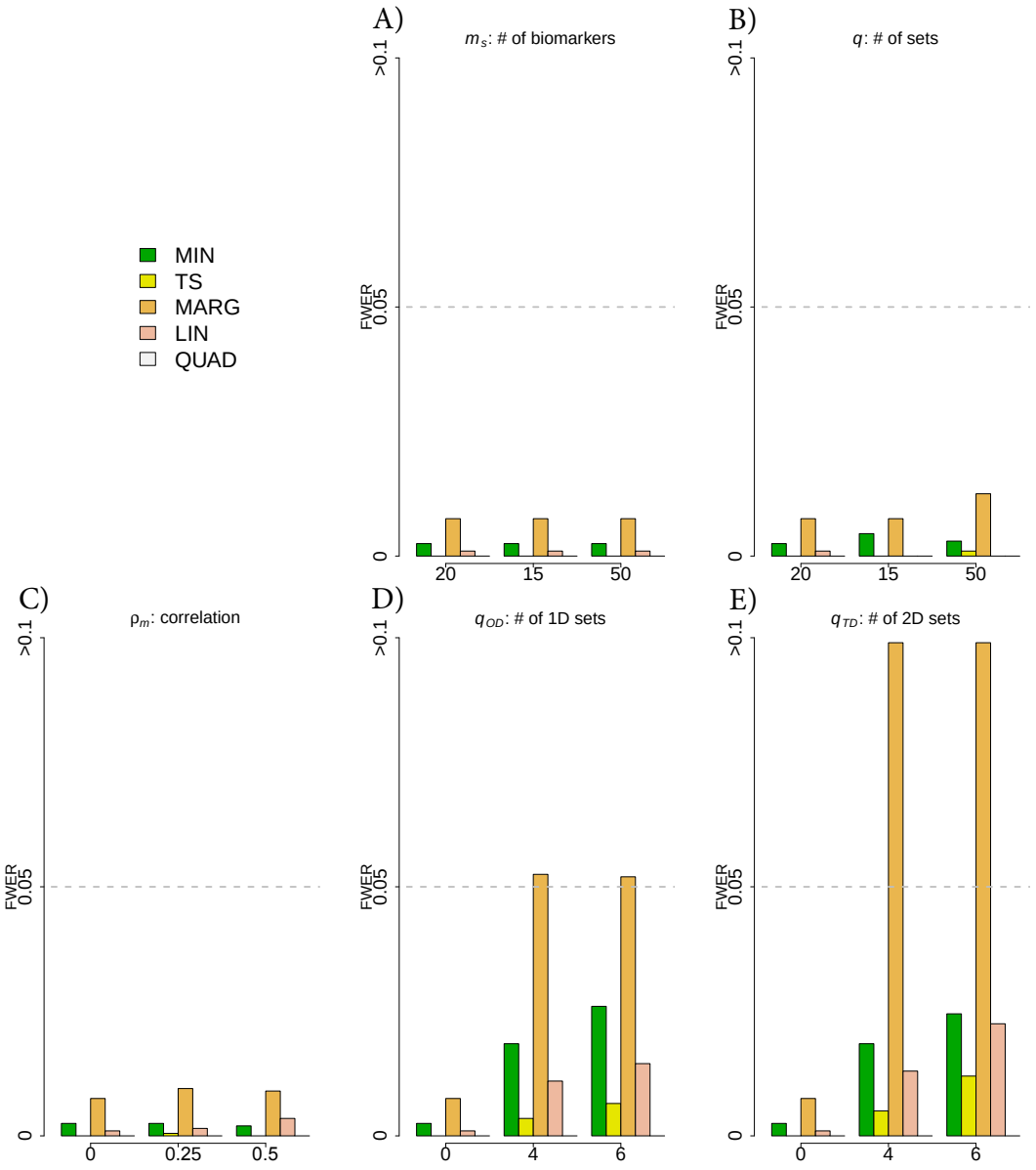

Figure S10: **FWER comparisons for binary outcome and latent confounder effecting E and M.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

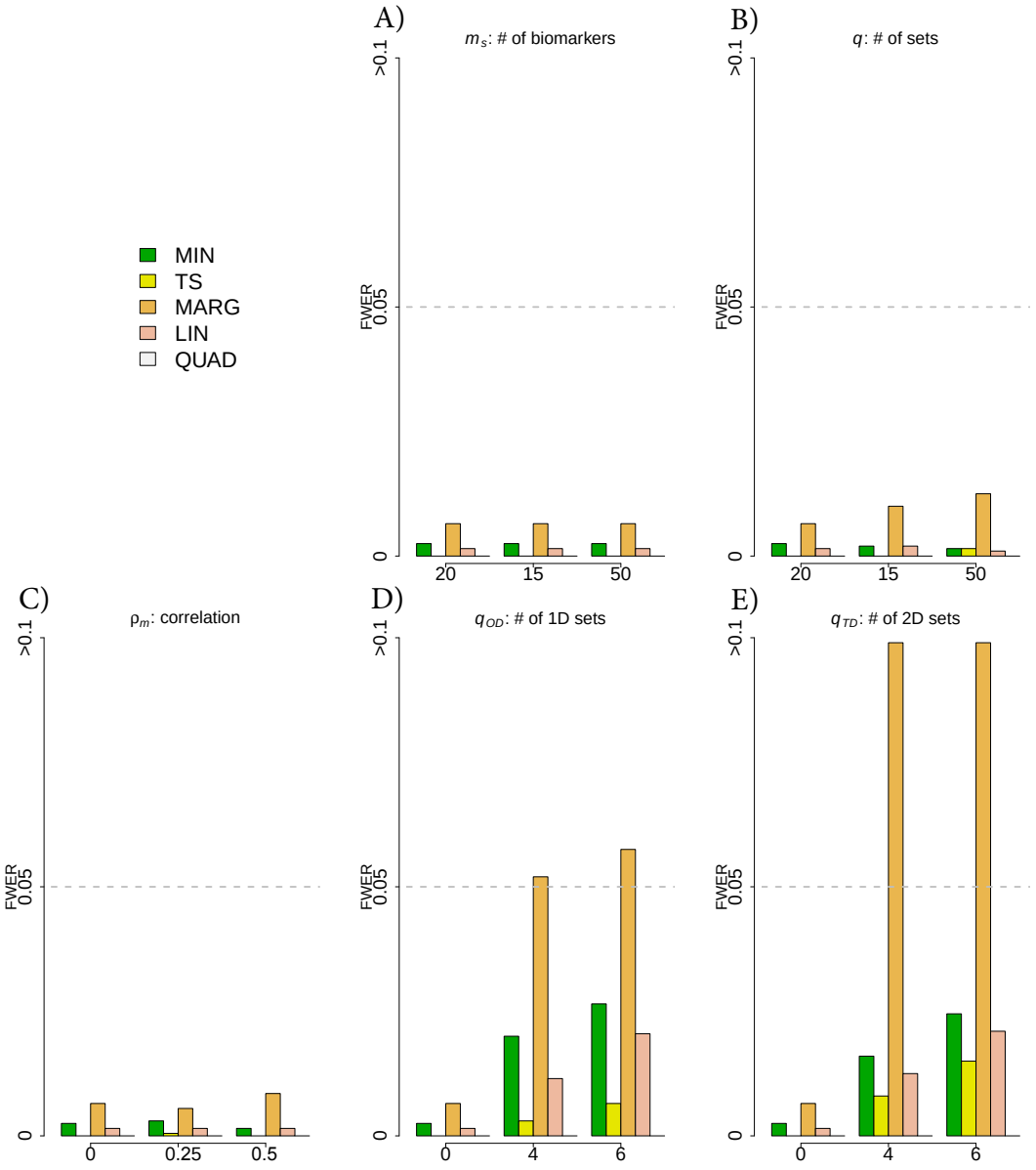

Figure S11: **FWER comparisons for binary outcome and latent confounder effecting M and Y.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

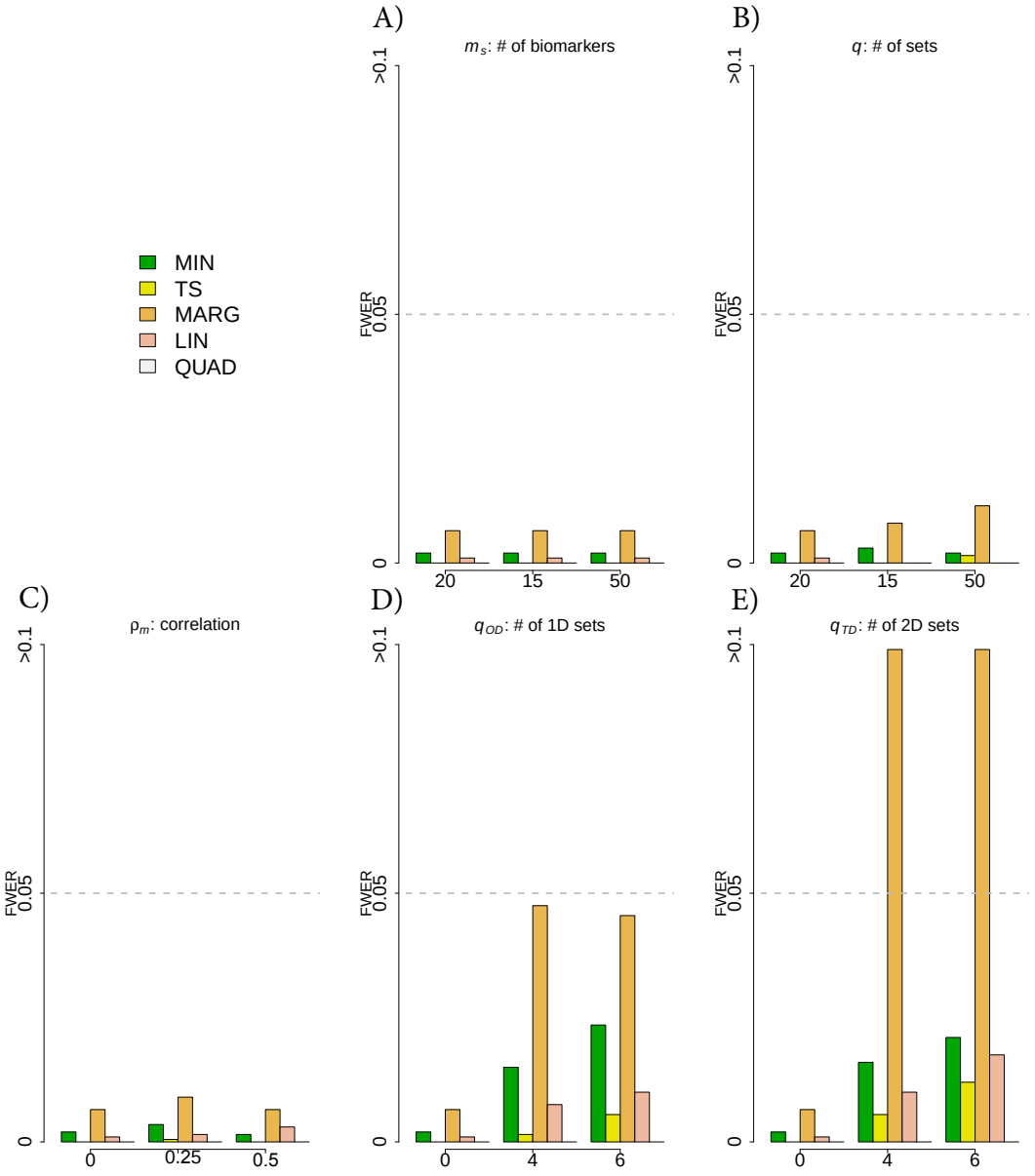

Figure S12: **FWER comparisons for binary outcome and latent confounder effecting M.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

4.3 Empirical Power Studies

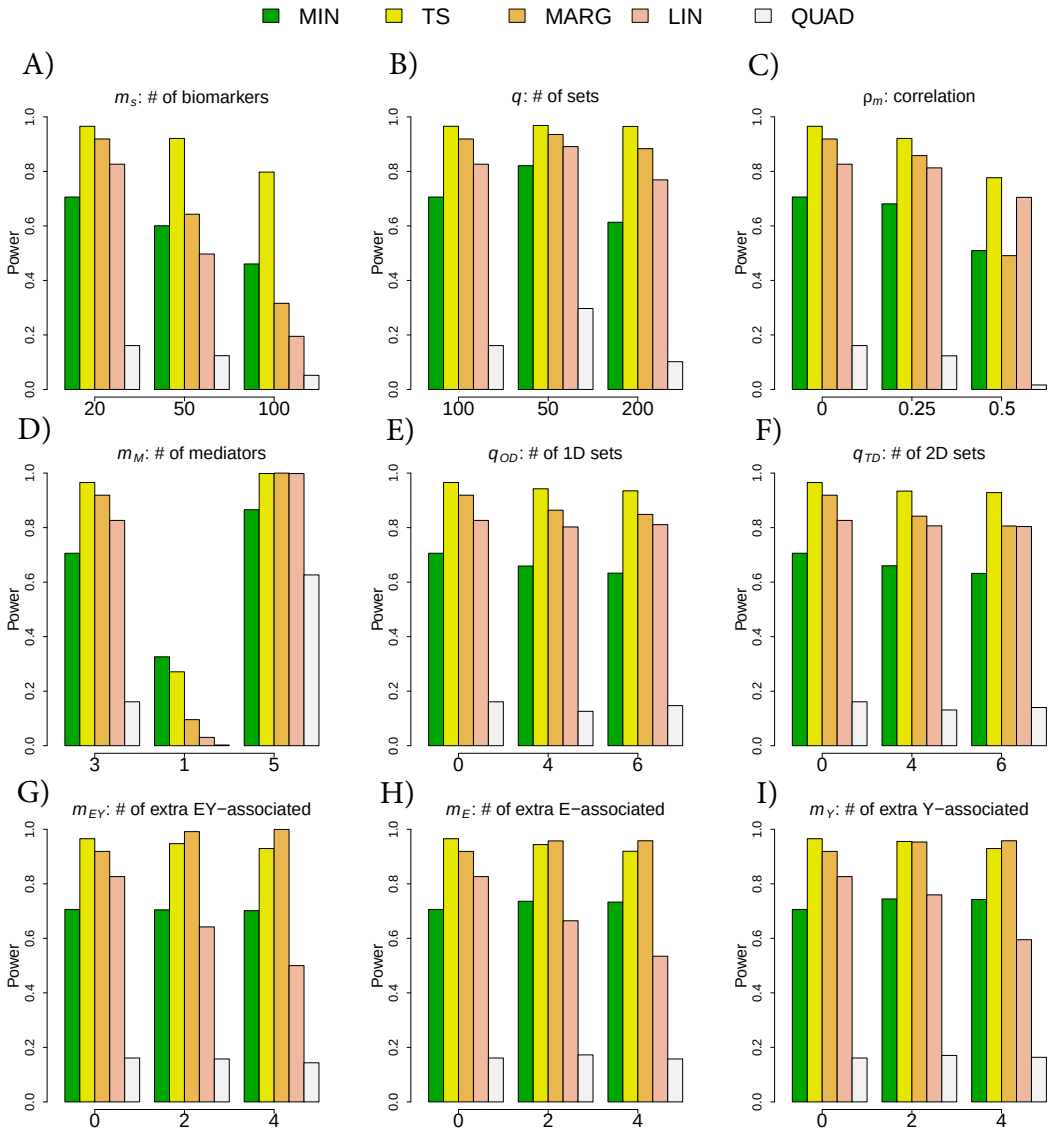

Figure S13: **Simulations for continuous outcome with $\gamma_j = \beta_j = 0.075$ and $q = 100$ disjoint sets as the baseline.** The bar-plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_M = 3$ mediating biomarkers, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

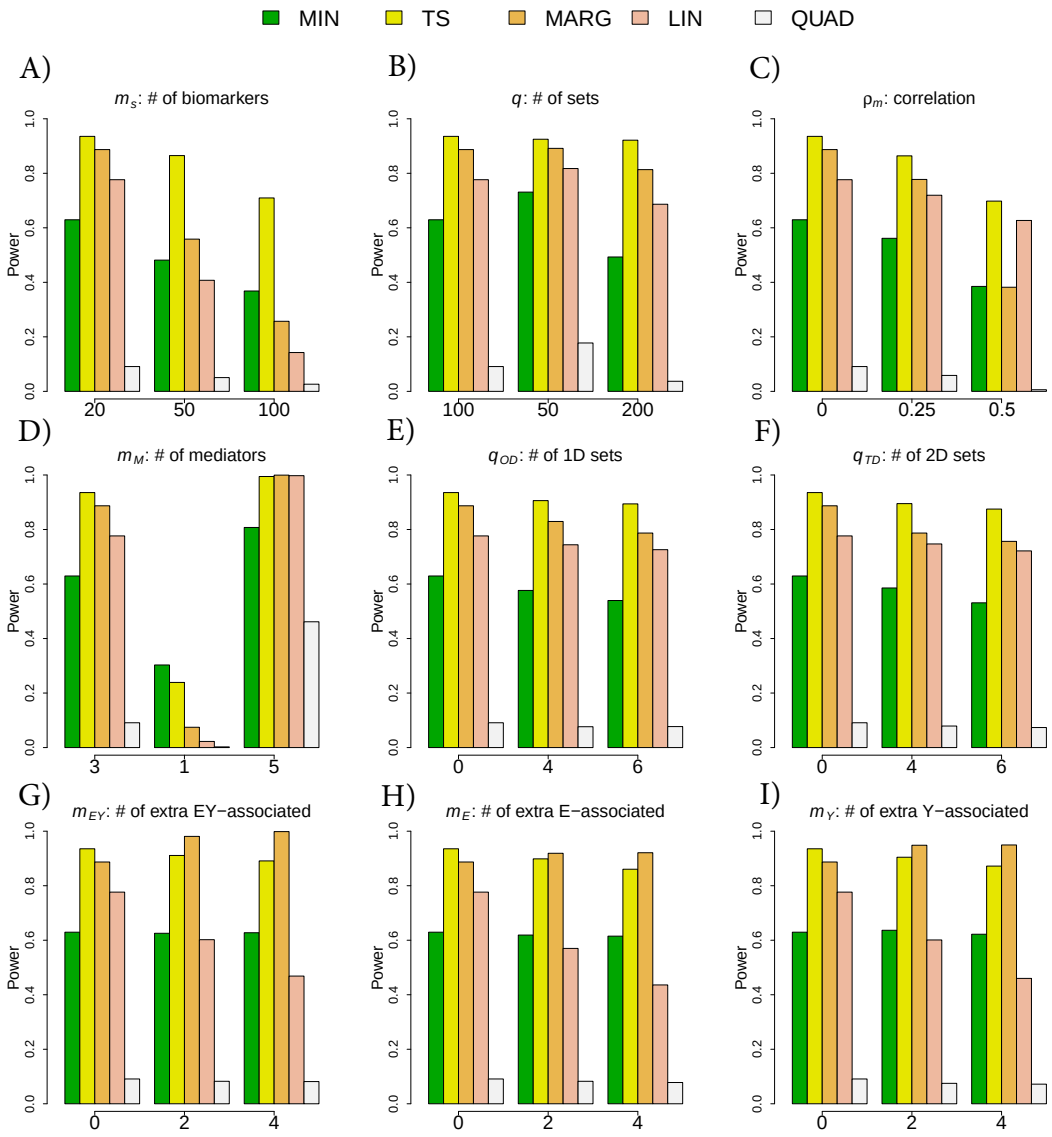

Figure S14: **Simulations for binary outcome with $\gamma_j = 0.1$, $\beta_j = 0.05$ and $q = 100$ disjoint sets as the baseline.** The bar-plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_M = 3$ mediating biomarkers, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

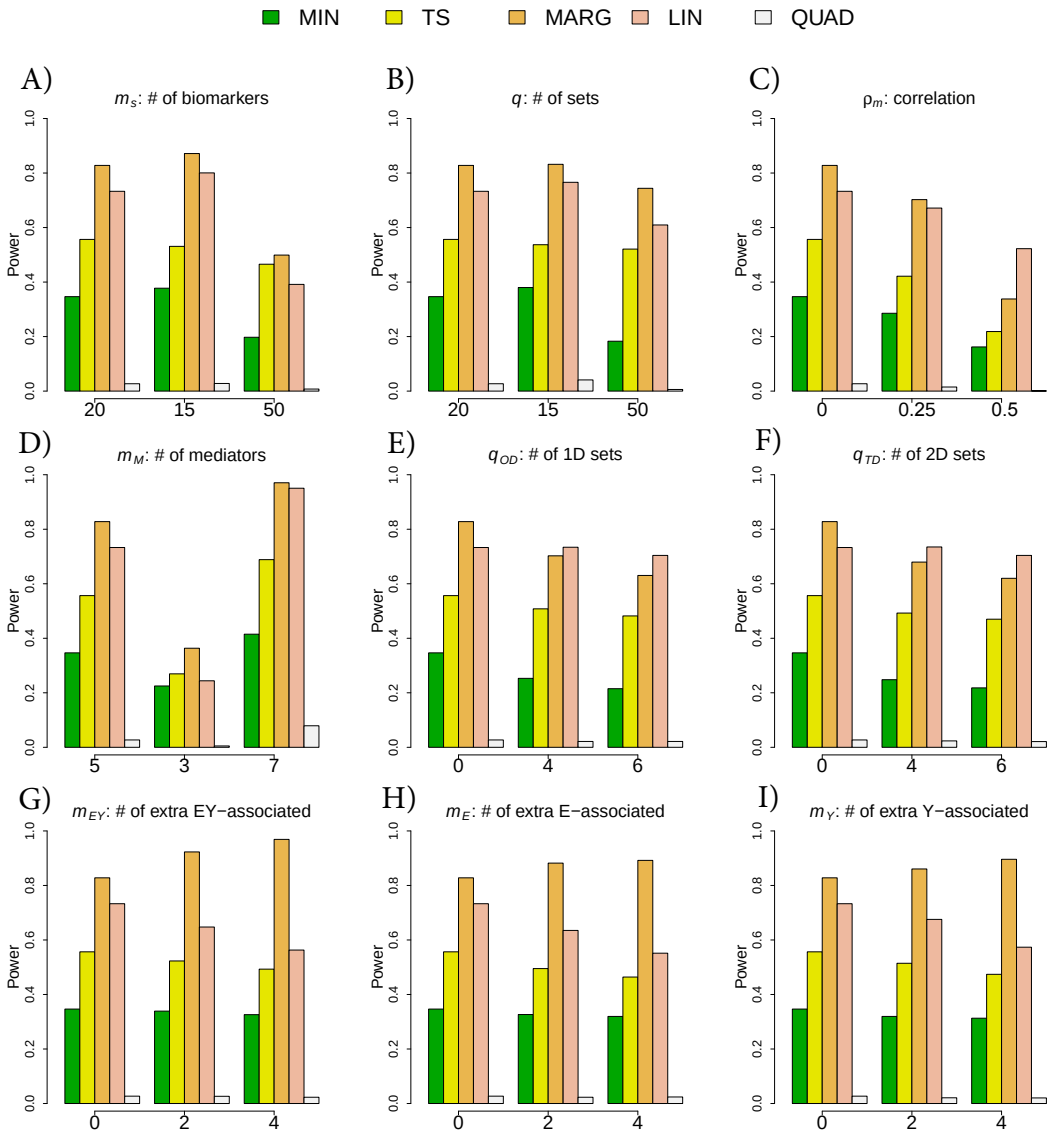

Figure S15: **Simulations for continuous outcome with $\gamma_j = \beta_j = 0.065$ and $m_M = 5$ mediating biomarkers as the baseline.** The bar-plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets as the baseline, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

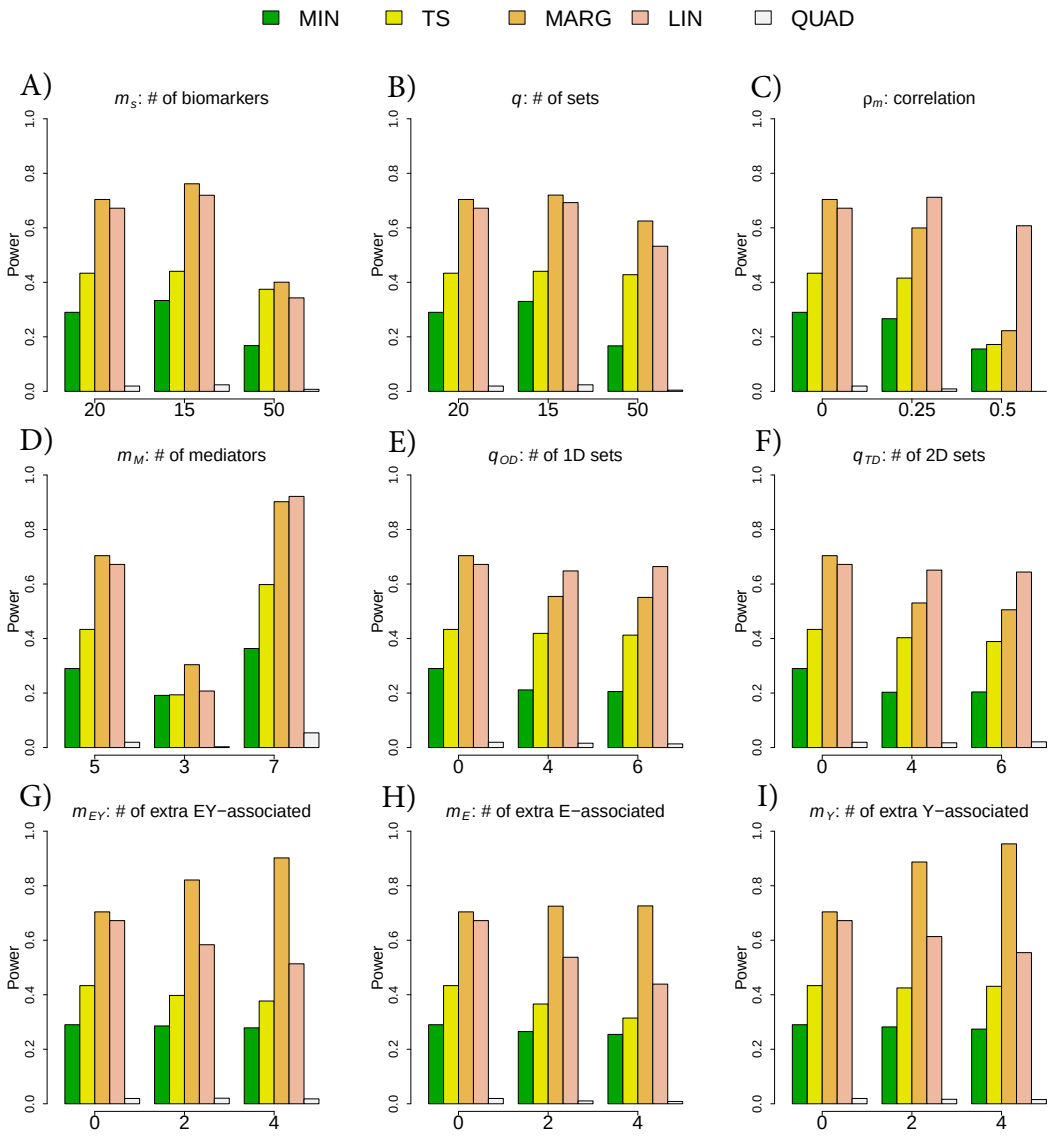

Figure S16: **Simulations for binary outcome with $\gamma_j = 0.06$, $\beta_j = 0.04$ and $m_M = 5$ mediating biomarkers as the baseline.** The bar-plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets as the baseline, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$.

4.4 Power comparisons with varying proportions of mediators in a set

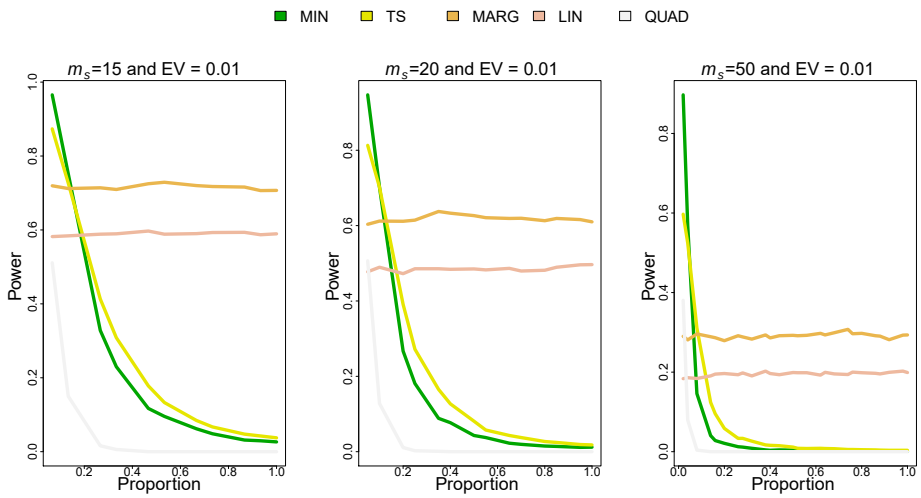

Figure S17: **Simulations for continuous outcome with $m_s = 15$, $m_s = 20$ and $m_s = 50$ biomarkers per set and constant proportion of variation explained by a set.** The plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $q = 20$ disjoint sets as the baseline, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. Overall effects of E on M and M on Y were set to $EV = 0.01$.

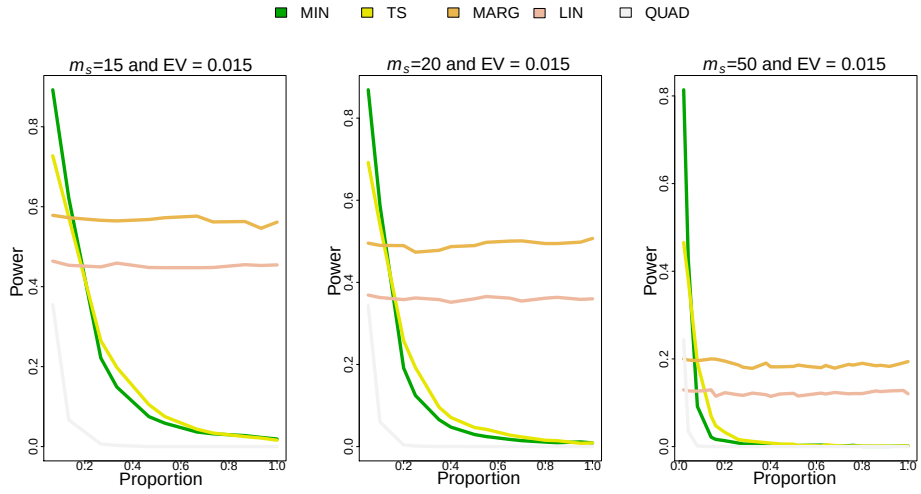

Figure S18: **Simulations for binary outcome with $m_s = 15$, $m_s = 20$ and $m_s = 50$ biomarkers per set and constant proportion of variation explained by a set.** The plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $q = 20$ disjoint sets as the baseline, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. Overall effects of E on M and M on Y were set to $EV_1 = 0.005$ and $EV_2 = 0.015$.

4.5 Power comparisons with varying proportion of variation explained by mediators in a set (EV)

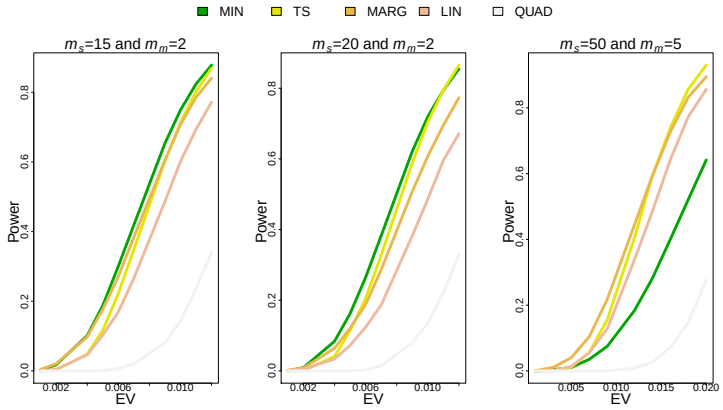

Figure S19: **Simulations for continuous outcome with $m_s = 15$, $m_s = 20$ and $m_s = 50$ biomarkers per set and constant number of mediators per set (i.e. m_m).** The plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $q = 20$ disjoint sets as the baseline, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. Overall effects of E on M and M on Y were set to the EV.

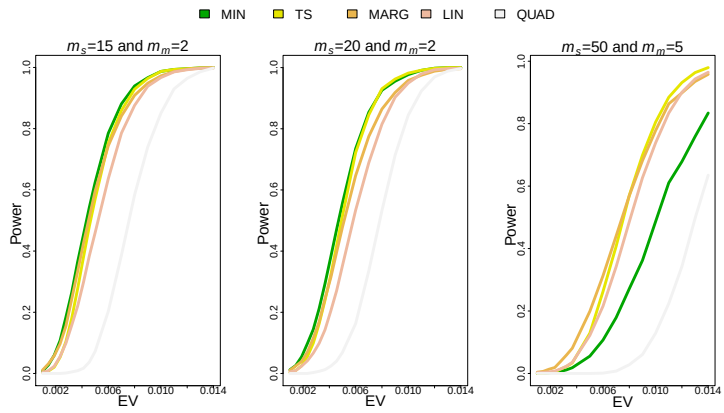

Figure S20: **Simulations for binary outcome with $m_s = 15$, $m_s = 20$ and $m_s = 50$ biomarkers per set and constant number of mediators per set (i.e. m_m).** The plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and QUAD (brown) procedures. The baseline scenario includes $q = 20$ disjoint sets as the baseline, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. Overall effects of E on M and M on Y were set to the $EV_1 = EV$ and $EV_2 = 3EV$.

4.6 FWER and power comparison with $JTV - comp$

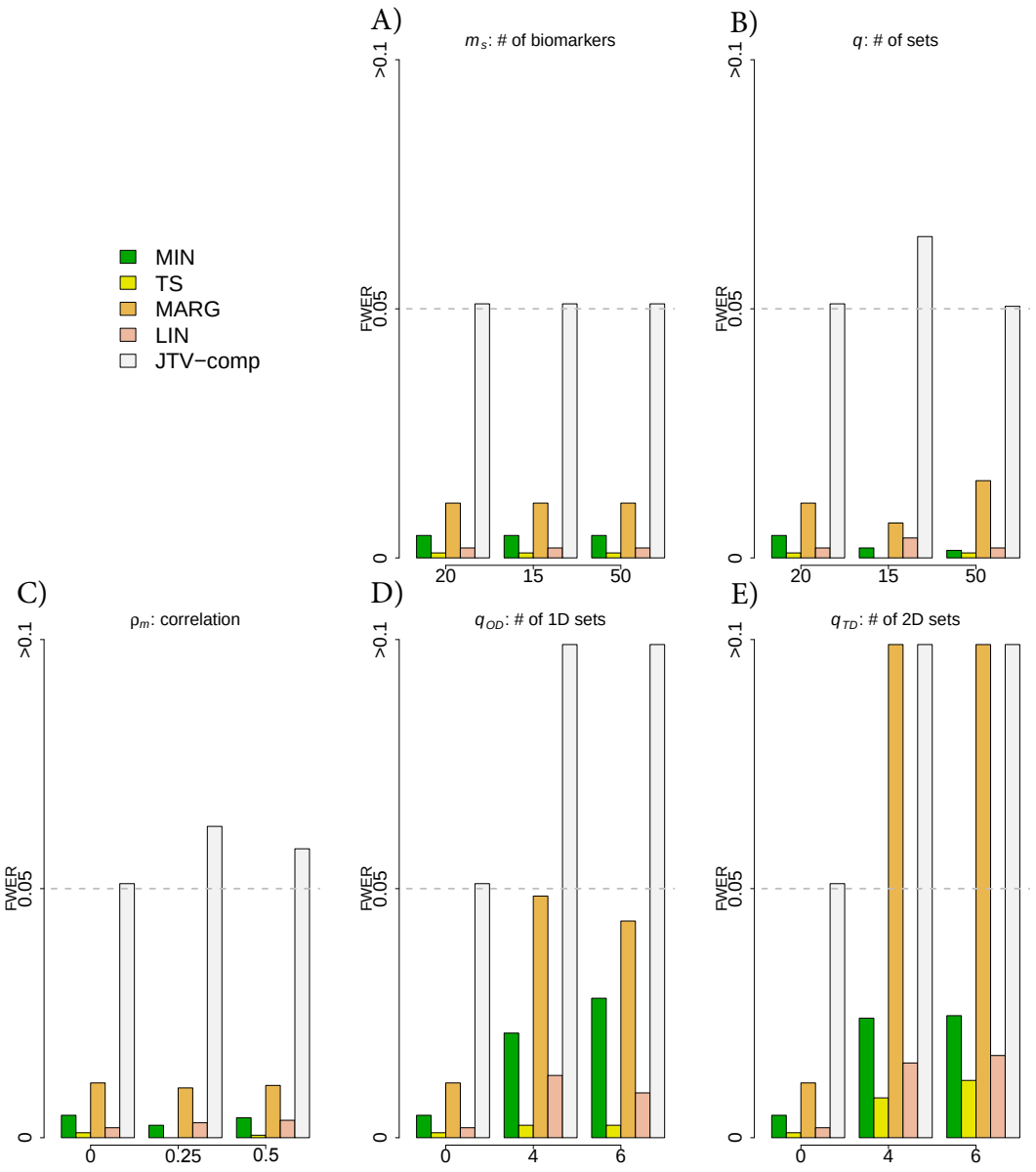

Figure S21: **FWER comparisons with $JTV - comp$ for continuous outcome and no latent confounders.** The bar-plots show the FWER when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and JTV-comp (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

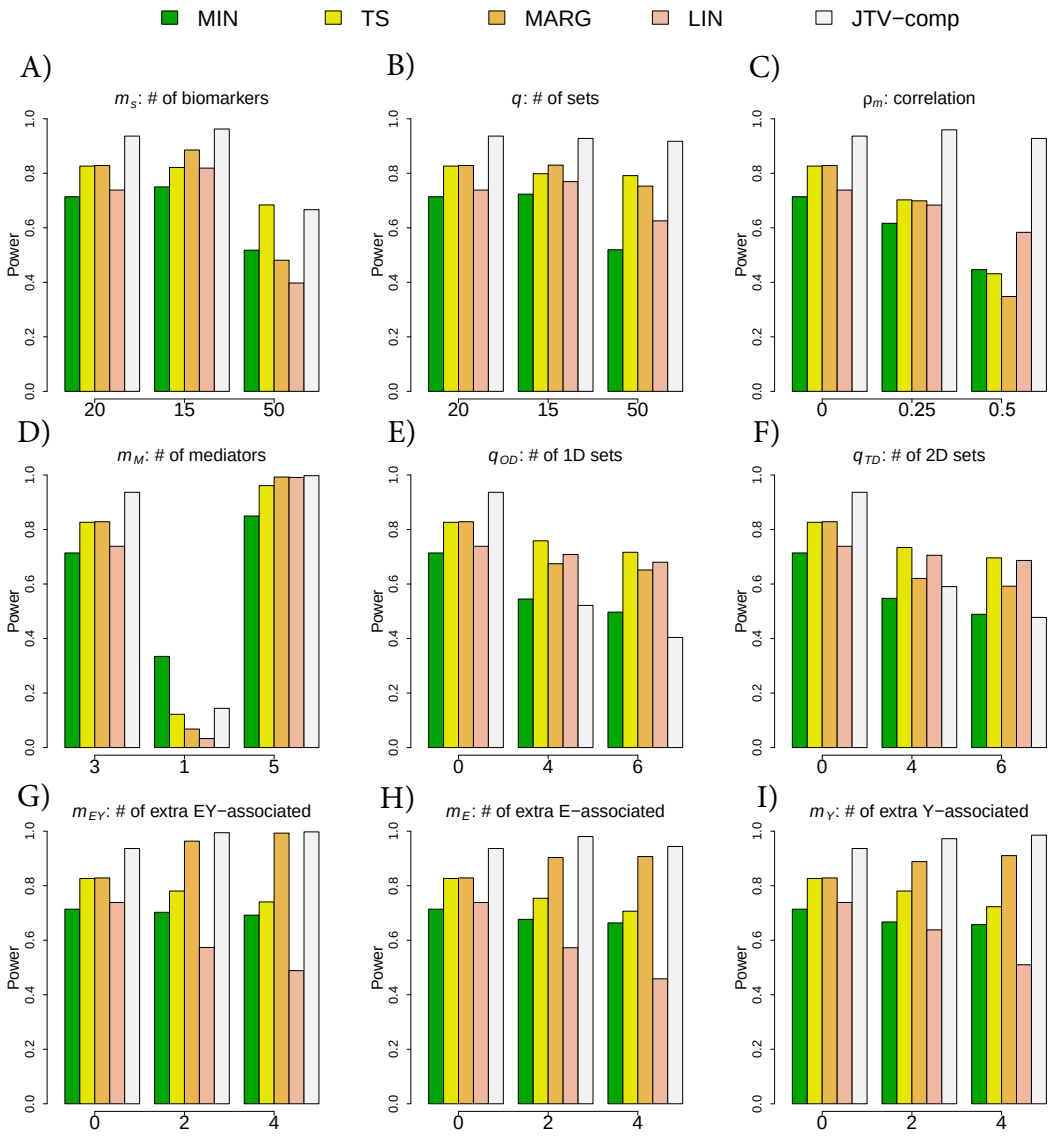

Figure S22: **Power comparisons with $JTV - comp$ under continuous outcome simulations.** The bar-plots show the power to detect the mediating set when using the TS (yellow), MIN (green), MARG (orange), LIN (red), and JTV-comp (brown) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

4.7 Detection of individual mediators in a mediation set

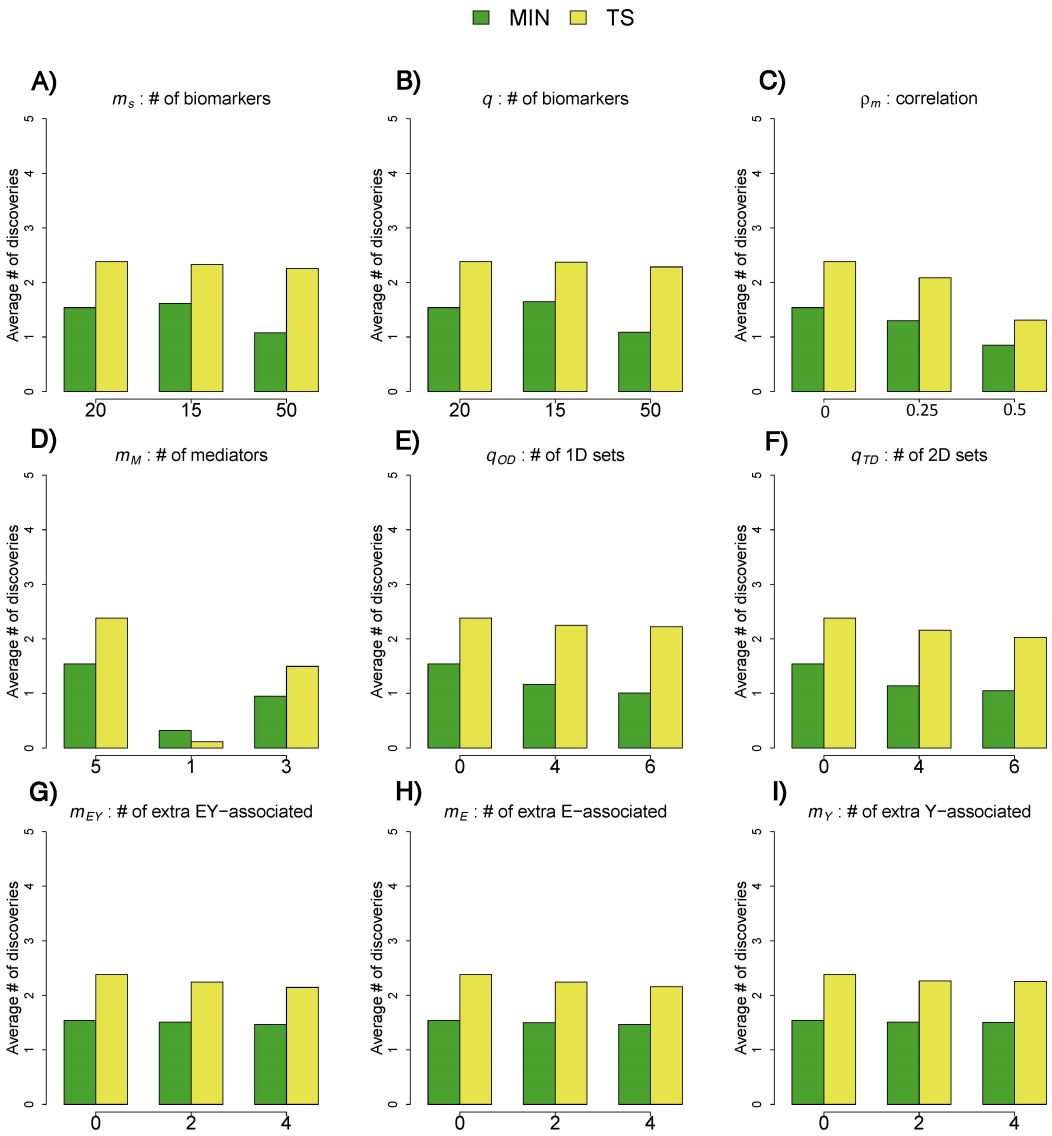

Figure S23: **Average number of biomarkers detected in continuous outcome simulations.** The bar-plots show the average number of mediators detected the mediating set when using the TS (yellow) and MIN (green) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, q= 20 disjoint sets, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_M = 3$ mediating biomarkers, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.

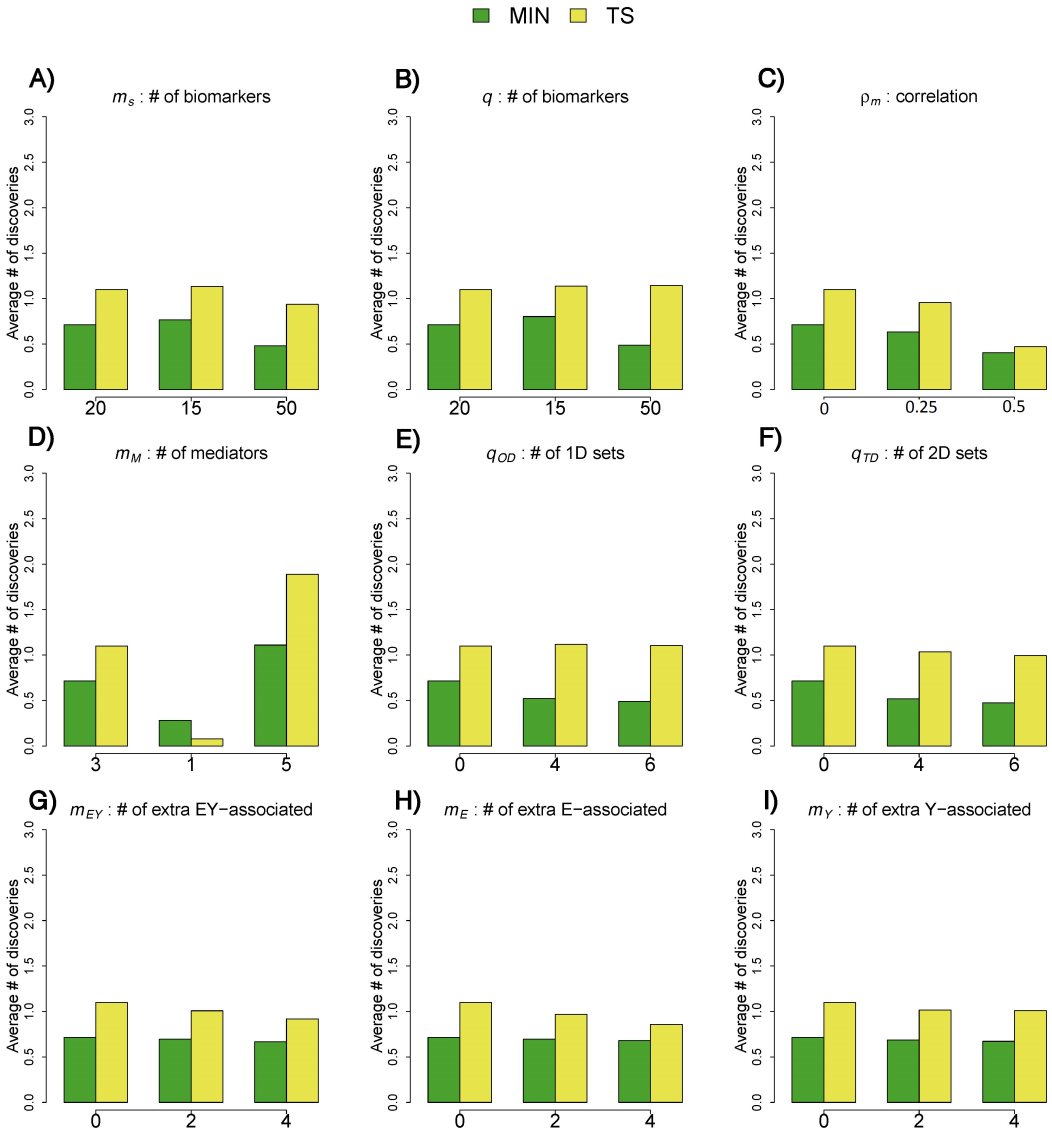

Figure S24: **Average number of biomarkers detected in binary outcome simulations.** The barplots show the average number of mediators detected the mediating set when using the TS (yellow) and MIN (green) procedures. The baseline scenario includes $m_s = 20$ biomarkers per set, $q = 20$ disjoint sets, $q_m = 1$ mediating set, $q_{OD} = 0$ one-dimensional sets, $q_{TD} = 0$ two-dimensional sets, $m_M = 3$ mediating biomarkers, $m_E = 0$ noise-biomarkers, and a correlation of $\rho_M = 0$. We evaluate the effect of varying a single parameter, keeping all other parameters set to their baseline value.