Supplementary Material for "Confidence bounds for the true discovery proportion based on the exact distribution of the number of rejections"

Friederike Preusse[1], Anna Vesely[2] and Thorsten Dickhaus[1]

[1] *Institute of Statistics, University Bremen*

[2] *Department of Statistical Sciences, University of Bologna*

1 Proofs

Theorem 1

The number of true discoveries m_1 is an unknown integer between 0 and m. Fix a candidate value $\tilde{m}_1 \in \{0,\ldots,m\}$, as well as the probability $\mathbb{P}_{\tilde{m}_1}$ under the corresponding model. From Eq. (2), the value $\gamma_{\tilde{m}_1}$ is such that

$$\mathbb{P}_{\tilde{m}_1}(\tilde{m}_1 \geq R\gamma_{\tilde{m}_1}) \geq 1-\alpha.$$

Then consider the minimum γ^* of these values, as defined in Eq. (3). For any $\tilde{m}_1$, we have that $R\gamma^* \leq R\gamma_{\tilde{m}_1}$, and so

$$\mathbb{P}_{\tilde{m}_1}(\tilde{m}_1 \geq R\gamma^*) \geq \mathbb{P}_{\tilde{m}_1}(\tilde{m}_1 \geq R\gamma_{\tilde{m}_1}) \geq 1-\alpha.$$

This holds in particular for the true value m_1, so that

$$\mathbb{P}_{m_1}(m_1 \geq R\gamma^*) \geq 1-\alpha.$$

Finally, recall that m_1 can take only integer values. Hence

$$m_1 \geq R\gamma^* \quad \Longleftrightarrow \quad m_1 \in \{\lceil R\gamma^* \rceil,\ldots,m\} \quad \Longleftrightarrow \quad m_1 \geq \lceil R\gamma^* \rceil$$

and so

$$\mathbb{P}_{m_1}(m_1 \geq \lceil R\gamma^* \rceil) = \mathbb{P}_{m_1}(m_1 \geq R\gamma^*) \geq 1-\alpha.$$

Therefore $\hat{m}_1 = \lceil r\gamma^* \rceil$ is a lower $(1-\alpha)$-confidence bound for m_1.

Theorem 2

From Eq. (2)

$$\gamma_{\tilde{m}_1} = \max\{\gamma \in [0,1] : \mathbb{P}_{\tilde{m}_1}(R\gamma \leq \tilde{m}_1) \geq 1-\alpha\}.$$

Recall that R is a discrete variable taking values in $\{0,\dots,m\}$, so it is sufficient to study its CDF in these values. First, suppose that $\tilde{m}_1 = 0$. In this case,

$$\mathbb{P}_0(R\gamma \leq 0) = \mathbb{P}_0(R\gamma = 0) = \begin{cases} 1 & \text{if } \gamma = 0 \\ \mathbb{P}_0(R=0) & \text{if } \gamma \in (0,1]. \end{cases}$$

Hence it is sufficient to look for the maximum in $\{0,1\}$:

$$\gamma_0 = \max\{\gamma \in \{0,1\} : \mathbb{P}_0(R\gamma = 0) \geq 1-\alpha\} = \begin{cases} 1 & \text{if } \mathbb{P}_0(R=0) \geq 1-\alpha \\ 0 & \text{otherwise.} \end{cases}$$

Then consider any other value $\tilde{m}_1 \in \{1,\dots,m\}$. Notice that for $\gamma = \tilde{m}_1/m$ we obtain

$$\mathbb{P}_{\tilde{m}_1}(R\gamma \leq \tilde{m}_1) = \mathbb{P}_{\tilde{m}_1}(R \leq m) = 1 \geq 1-\alpha,$$

and so $\gamma_{\tilde{m}_1} \geq \tilde{m}_1/m > 0$. Then we can re-write

$$\begin{aligned}
\gamma_{\tilde{m}_1} &= \max\left\{\gamma \in \left[\frac{\tilde{m}_1}{m}, 1\right] : \mathbb{P}_{\tilde{m}_1}\left(R \leq \frac{\tilde{m}_1}{\gamma}\right) \geq 1-\alpha\right\} \\
&= \max\left\{\gamma \in \left\{\frac{\tilde{m}_1}{m}, \frac{\tilde{m}_1}{m-1}, \dots, \frac{\tilde{m}_1}{\tilde{m}_1}\right\} : \mathbb{P}_{\tilde{m}_1}\left(R \leq \frac{\tilde{m}_1}{\gamma}\right) \geq 1-\alpha\right\} \\
&= \max\left\{\gamma = \frac{\tilde{m}_1}{\ell},\ \ell \in \{\tilde{m}_1,\dots,m-1,m\} : \mathbb{P}_{\tilde{m}_1}(R \leq \ell) \geq 1-\alpha\right\} \\
&= \frac{\tilde{m}_1}{\ell_{\tilde{m}_1}}
\end{aligned}$$

where

$$\ell_{\tilde{m}_1} = \min\{\ell \in \{\tilde{m}_1,\dots,m-1,m\} : \mathbb{P}_{\tilde{m}_1}(R \leq \ell) \geq 1-\alpha\}.$$

From this, it follow that

$$\gamma_{\tilde{m}_1} = \tilde{m}_1/\ell_{\tilde{m}_1}.$$

Proposition 1

Denote by $S = |\mathscr{R} \cap \mathscr{M}_1|$ the random variable corresponding to the number of true discoveries in the rejection set $\mathscr{R}$, and by $V = R - S$ the random variable corresponding to the number of false discoveries. We use that $R = V + S$, then by the law of total probability

$$\begin{aligned}
\mathbb{P}_{\tilde{m}_1}(R=\ell) &= \sum_{j=0}^{\ell} \mathbb{P}_{\tilde{m}_1}(V=j, S=\ell-j) \\
&= \sum_{j=0}^{\ell} \binom{m-\tilde{m}_1}{j}\binom{\tilde{m}_1}{\ell-j} t_\ell^j (F(t_\ell))^{\ell-j} \\
&\qquad \cdot \Psi^{Uni[0,1],\overline{F}}_{m-\tilde{m}_1-j,\tilde{m}_1-\ell+j}(1-t_m,\dots,1-t_{\ell+1}),
\end{aligned}$$

The last equality is due to Roquain & Villers (2011), Section 5.3. The expression for $\mathbb{P}_{\tilde{m}_1}(R \leq \ell)$ given in the proposition follows.

Proposition 2

Because R is discrete, the expected value of R can be computed as

$$E_{\tilde{m}_1}[R] = \sum_{\ell=0}^{m} \ell \cdot \mathbb{P}_{\tilde{m}_1}(R = \ell) = \sum_{\ell=0}^{m} l \cdot (\mathbb{P}_{\tilde{m}_1}(R \leq \ell) - \mathbb{P}_{\tilde{m}_1}(R \leq \ell - 1)),$$

where $\mathbb{P}_{\tilde{m}_1}(R \leq \ell)$ is given in Proposition 1. The expression of $E_{\tilde{m}_1}[R\gamma^*]$ follows.

Similarly, the variance is computed as

$$\begin{aligned}
&Var_{\tilde{m}_1}[R] = E_{\tilde{m}_1}[R^2] - E_{\tilde{m}_1}[R]^2 \\
&= \sum_{\ell=0}^{m} \ell^2 \cdot \mathbb{P}_{\tilde{m}_1}(R = \ell) - \left(\sum_{\ell=0}^{m} \ell \cdot \mathbb{P}_{\tilde{m}_1}(R = \ell) \right)^2 \\
&= \sum_{\ell=0}^{m} \ell^2 \cdot (\mathbb{P}_{\tilde{m}_1}(R \leq \ell) - \mathbb{P}_{\tilde{m}_1}(R \leq \ell - 1)) - \left(\sum_{\ell=0}^{m} \ell \cdot (\mathbb{P}_{\tilde{m}_1}(R \leq \ell) - \mathbb{P}_{\tilde{m}_1}(R \leq \ell - 1)) \right)^2.
\end{aligned}$$

Family of critical vectors defined by Eq. 8

Let $\mathbf{t}(\lambda, \beta) = (t_1(\lambda, \beta), \ldots, t_m(\lambda, \beta))^\top$ be a vector belonging to the family of vectors defined by Eq. (8). In the following we notationally omit dependency on λ and β to improve readability. According to Definition 1, $\mathbf{t}$ is a critical vector of a step up procedure when it is non-decreasing and takes values between 0 and 1.

Since $i \in \{1, \ldots, m\}$, i/m is non-decreasing in i. Therefore, $t_i = \lambda(i/m)^\beta$ is non-decreasing in i if $\lambda \geq 0$ and $\beta \geq 0$. When the first requirement is met, it is enough to show that $t_0 \geq 0$ and $t_m \leq 1$ to fulfill the second requirement. Note that $t_0 = \lambda(0/m)^\beta = 0$ for $\beta \neq 0$. For $\beta = 0$, $t_0 = \lambda \geq 0$. The last inequality is due to the restrictions on λ introduced by the first requirement. Furthermore, $t_m = \lambda$ for any choice of β. Therefore, for $0 \leq \lambda \leq 1$, $\mathbf{t}$ satisfies the second requirement.

Consequently, for $0 \leq \lambda \leq 1$ and $\beta \geq 0$, any vector defined by Eq (8) is a critical vector for a step-up procedure.

2 Algorithms

In this section, we give algorithms for the proposed methodology.

Algorithm 1 computes a lower $(1-\alpha)$-confidence bound $\hat{m}_1$ for the number of true discoveries as in Eq. (4), using results from Theorems 1 and 2 and Proposition 1. Note that observations in form of r are only used in the last step of the algorithm.

Algorithm 2 implements the methods based on thresholds for the p-values to compute an estimate $\hat{\theta}$ of the effect size, as described in Section 5.5 in the main article. The algorithm uses t-values $\mathbf{q} = (q_1, \ldots, q_m)^\top$ and p-values $p = (p_1, \ldots, p_m)^\top$. It selects the t-values for which

Algorithm 1: Algorithm to compute $\hat{m}_1$ such that $\mathbb{P}_{m_1}(m_1 \geq \hat{m}_1) \geq 1-\alpha$ as in Eq. (4).

for $\tilde{m}_1 = 0, \ldots, m$ **do**
  $\ell = -1$;
  $Pr = 0$;
  while $Pr < 1-\alpha$ & $\ell < m$ **do**
    $\ell = \ell + 1$;
    $Pr = Pr + P(R = \ell)$;
  end
  if $\ell < \tilde{m}_1$ **then**
    $\ell_{\tilde{m}_1} = \tilde{m}_1$;
  else
    $\ell_{\tilde{m}_1} = r$
  end
  if $\ell_{\tilde{m}_1} = 0$ **then**
    $\gamma_{\tilde{m}_1} = 1$;
  else
    $\gamma_{\tilde{m}_1} = \frac{\tilde{m}_1}{\ell_{\tilde{m}_1}}$
  end
end
$\gamma^* = \min_{\tilde{m}_1 \in [m]}(\gamma_{\tilde{m}_1})$;
return $\lceil r \cdot \gamma^* \rceil$;

the corresponding p-values do not exceed a given threshold h, and uses those to determine $\hat{\theta}$. The threshold h may be either a fixed value or the threshold based on of a single-step test controlling the FWER at level a, such as the Bonferroni and Šidák corrections. For the Bonferroni correction,

$$h = \frac{a}{m},$$

while for the Šidák correction,

$$h = 1 - (1-a)^{\frac{1}{m}}.$$

Algorithm 2: Algorithm to compute $\hat{\theta}$ based on Eq. (12), using a threshold h for the p-values.

$\nu = N - 1$;
$\mathbf{q}_{\text{sel}} = \mathbf{q}[\mathbf{p} \leq h]$;
$\hat{\mu} = \text{mean}(\mathbf{q}_{\text{sel}}) \cdot \sqrt{\frac{2}{\nu}} \frac{\Gamma(\nu/2)}{\Gamma((\nu-1)/2)}$;
return $\hat{\mu} \cdot \sqrt{\frac{2}{N}}$

3 Additional simulation results

In this section, further details about the simulations of Section 5 in the main article are given. We claim that, for any family of critical vectors indexed by a parameter λ (and eventually other parameters), optimal power is achieved when selecting the largest λ for which $\gamma^* = 1$ in Eq. (3), as mentioned in Section 4 in the main article. Subsequently, we present additional plots that support the claims on results made in Section 5 in the main article. Tables and plots are shown only for some choices of the parameters, but other values lead to the same conclusions.

First, consider the setting under Assumption 1, where p-values are independent and the CDF F is correctly specified. Table 1 displays the sum of the expected values given in Eq. (9), obtained using different critical vectors and for varying effect sizes θ. Eq. (9) is maximized when λ is chosen as the largest value for which $\gamma^* = 1$. Furthermore, Table 2 displays the average confidence bounds of the proposed procedure with different critical vectors for a fixed $\theta = 0.8$ and varying values of m_1. In most cases, the largest confidence bound is obtained for critical vectors with $\gamma^* = 1$, with the only exception of settings with very sparse signal ($m_1 \leq 10$). This justifies the choice of maximizing Eq. (9) to select the critical vector, as in most cases the same critical vector leads to the largest confidence bound.

Figures 1-9 display, for different simulation scenarios, the proportion of iterations for which the lower confidence bound is larger than the true number of discoveries ($m_1 < \hat{m}_1$), as well as the average lower confidence bound $\hat{m}_1$. In all cases, $\hat{m}_1$ is obtained from *GS*, *MR*, *GL* and the proposed method with different families of critical vectors. All results presented are based on $N^{[b]} = 50$. If the proportion of iterations with $\hat{m}_1 > m_1$ is at most α, the bounds are considered valid. Critical vectors of each family have been chosen using the largest $\gamma^* \in \{0.8, 0.9, 0.95, 1\}$

such that the proposed method determines valid confidence bounds. Results are shown for smaller effect sizes ($\theta \leq 1.2$) and $N^{[e]} \in \{0, 50\}$. Since the validity and power of the confidence bounds obtained from any procedure depend on the denseness of the signal, we show results for the extreme values of the denseness of the signal ($m_1 \in \{10, 90\}$).

In particular, Figure 1-5 display results under dependency of the p-values and correct specification of F. The proportions of iterations with $\hat{m}_1 > m_1$ are displayed in Figures 1 and 2 for $\rho = 0.3$ and Figure 3 for $\rho = 0.9$. For low positive dependency ($\rho = 0.3$) and sparse signal, the proposed procedure with $\gamma^* = 1$ returns valid confidence bounds, as illustrated in Figure 1, upper right panel. In this setting, confidence bounds obtained from *MR* are invalid for all considered effect sizes, while *GL* returns valid bounds for low effect sizes. As in the independent setting, the effect size θ for which *GL* returns invalid bounds decreases as $N^{[e]}$ increases. This is illustrated in the upper left and right panel, which show results for $N^{[e]} = 0$ and $N^{[e]} = 50$, respectively. When the signal is dense, the established procedures return valid confidence bounds for all considered effect sizes and $N^{[e]} \in \{0, 50\}$. The proposed procedure with critical vectors Exp_{opt} and $AORC_{opt}$ return invalid confidence bounds for $\theta \leq 0.6$. This is displayed in the lower half of Figure 1 for $N^{[e]} = 0$ (left-hand side) and $N^{[e]} = 50$ (right-hand side). The number of iterations with $\hat{m}_1 > m_1$ decreases as γ^* decreases. This is illustrated in Figure 2 for $\gamma^* \in \{0.9, 0.95\}$ and $\rho = 0.3$. Note that only results for $\theta \geq 0.6$ are displayed, as it was not possible to find critical vectors such that $0 < \gamma^* < 1$ for $\theta = 0.4$. For $\rho = 0.9$, the number of settings for which confidence bounds are invalid increases for sparse signals. That is, *MR* and *GL* return invalid confidence bounds for all considered effect sizes. Confidence bounds obtained from the proposed procedure with critical vectors $AORC_{opt}$ are invalid for small effect sizes ($\theta \leq 0.4$). This is displayed in the upper half of Figure 3. For dense signal (lower half of Figure 3), the established procedures return valid confidence bounds for all considered effect sizes, while the proposed procedure with critical vectors Exp_{opt} and $AORC_{opt}$ return invalid confidence bounds for $\theta \leq 0.4$. For both sparse and dense signal, the size of $N^{[e]}$ has almost no influence on the number of iterations with $\hat{m}_1 > m_1$ for $\rho = 0.9$. This is illustrated on the left-hand side ($N^{[e]} = 0$) and right-hand side ($N^{[e]} = 50$) of Figure 3.

The average $\hat{m}_1$ for $\rho = 0.3$ is illustrated in Figure 4 for $N^{[e]} = 0$ and Figure 5 for $N^{[e]} = 50$. In both figures, the proposed procedure with $\gamma^* = 0.9$ (left-hand side), $\gamma^* = 0.95$ (middle) and $\gamma^* = 1$ (right-hand side) are displayed. While $\gamma^* = 0.9$ corresponds to valid confidence bounds for very small effect sizes ($\theta \geq 0.6$), $\gamma^* = 0.95$ and $\gamma^* = 1$ correspond to valid confidence bounds for slightly larger effect sizes ($\theta \geq 0.8$). As before, for $\gamma^* < 1$, only effect sizes $\theta \geq 0.6$ are displayed. When valid, increasing γ^* leads to an increase in the power of the proposed procedure. For sparse signal and $N^{[e]} = 0$ the proposed procedure is more powerful than the *GS* procedure for low effect sizes ($\theta \leq 0.8$). When *GL* returns valid bounds, it is generally more powerful than the proposed procedure. This is illustrated in the upper half of Figure 4. When the signal is dense, the established procedures are more powerful than the proposed procedure. The established procedures are more powerful than the proposed procedure for both sparse and dense signal for $N^{[e]} = 50$. The exception is sparse signal and very low effect size, where $AORC_{opt}$ is the most powerful procedure, as displayed in Figure 5, upper right panel.

Results under misspecification of F ($\hat{\theta} = \theta + 0.1$) and independence of the p-values are illustrated in Figure 6-7. Note that the performance of the established procedure is identical

to the case of correct specification of F. Figure 6 displays the proportion of iterations with $\hat{m}_1 > m_1$ for both $N^{[e]} = 0$ (left-hand side) and $N^{[e]} = 50$ (right-hand side). The proposed procedure with any family of critical vectors and $\gamma^* = 1$ returns valid confidence bounds for both sparse signal, displayed in the upper half of Figure 6, and dense signal, displayed in the lower half. Figure 7 illustrates the average $\hat{m}_1$ for $N^{[e]} = 0$ (left-hand side) and $N^{[e]} = 50$ (right-hand side). For sparse signal, the established procedures are more powerful than the proposed procedure, for both $N^{[e]} = 0$ and $N^{[e]} = 50$. This is illustrated in the upper half of Figure 7. For dense signal, the proposed procedure with any family of critical vector is more powerful than the established procedures for low effect sizes ($\theta \leq 0.8$) and $N^{[e]} = 0$. This is not the case for $N^{[e]} = 50$. These results are displayed in the lower half of Figure 7.

Lastly, Figure 8-9 illustrate the results under dependency of the p-values and misspecification of F for $\rho = 0.9$ and $\hat{\theta} = \theta + 0.1$. The confidence bounds are invalid for some m_1 and $\rho \geq 0.6$; we choose to show results for $\rho = 0.9$ to make comparison with the setting of correct specification of F possible. Again, the results for the established procedures are identical to the results under correct specification of F. The proposed procedure with any family of critical vector and $\gamma^* = 1$ returns valid confidence bounds for sparse signal, as illustrated in the upper half of Figure 8. For dense signal, the proposed procedure with critical vectors Exp_{opt} and $AORC_{opt}$ return invalid confidence bounds for $\theta \leq 0.4$. This is displayed in the lower half of Figure 8. Figure 9 displays the average $\hat{m}_1$ for $\hat{\theta} = \theta + 0.1$ and $\rho = 0.9$. The proposed procedure with any family of critical vector is more powerful than the established procedures for $\theta = 0.4$ and sparse signal, for both $N^{[e]} = 0$ and $N^{[e]} = 50$. This is illustrated in the upper half of Figure 9. For dense signal, the proposed procedure with any family of critical vectors is less powerful than the established procedures, as displayed in the lower half of Figure 9.

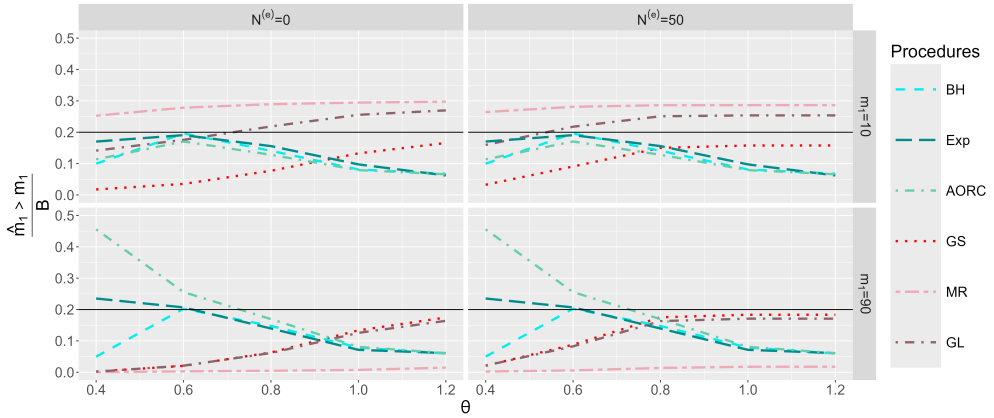

Figure 1: Dependency setting: p-values have correlation $\rho = 0.3$ and effect size θ is known. Proportion of iterations with $\hat{m}_1 > m_1$, where $\hat{m}_1$ is the lower confidence bound for the number of true discoveries $m_1 = \{10, 90\}$ obtained from the proposed procedure with different critical vectors (BH_{opt}, Exp_{opt}, $AORC_{opt}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 0$ (left) and $N^{[e]} = 50$ (right). The total number of iterations is $B = 10{,}000$. The black solid line corresponds to the significance level $\alpha = 0.2$.

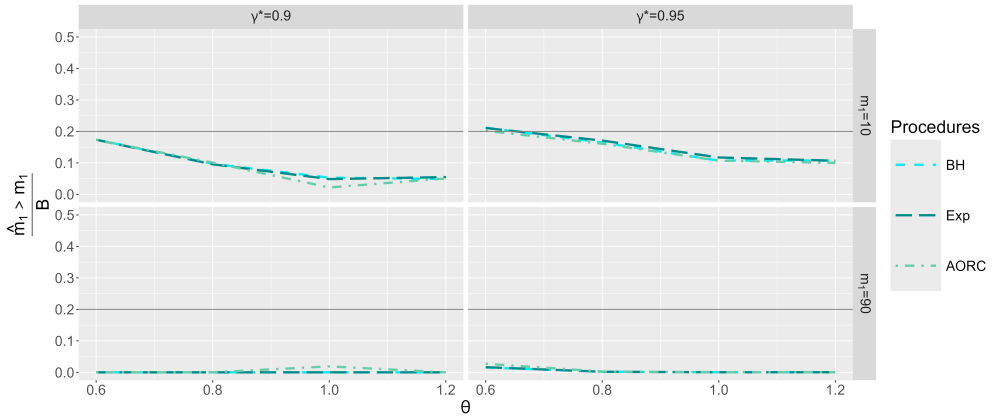

Figure 2: Dependency setting: p-values have correlation $\rho = 0.3$ and effect size θ is known. Proportion of iterations with $\hat{m}_1 > m_1$, where $\hat{m}_1$ is the lower confidence bound for the number of true discoveries $m_1 = \{10, 90\}$ obtained from the proposed procedure with different critical vectors (BH_{γ^*},Exp_{γ^*}, $AORC_{\gamma^*}$). The total number of iterations is $B = 10{,}000$. The black solid line corresponds to the significance level $\alpha = 0.2$.

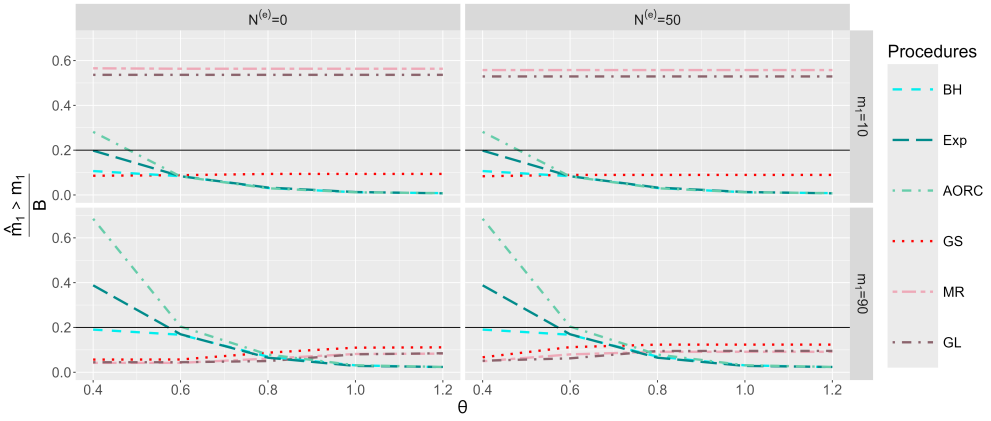

Figure 3: Dependency setting: p-values have correlation $\rho = 0.9$ and effect size θ is known. Proportion of iterations with $\hat{m_1} > m_1$, where $\hat{m_1}$ is the lower confidence bound for the number of true discoveries $m_1 = \{10, 90\}$ obtained from the proposed procedure with different critical vectors (BH_{opt}, Exp_{opt}, $AORC_{opt}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 0$ (left) and $N^{[e]} = 50$ (right). The total number of iterations is $B = 10{,}000$. The black solid line corresponds to the significance level $\alpha = 0.2$.

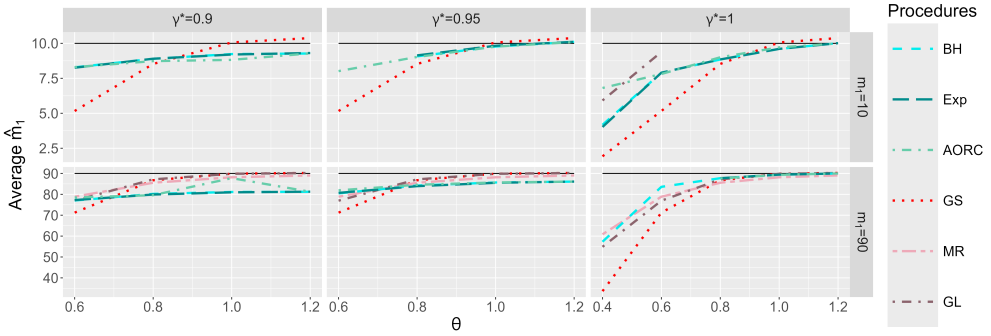

Figure 4: Dependency setting: p-values have correlation $\rho = 0.3$ and effect size θ is known. Average lower confidence bound $\hat{m_1}$ obtained from the proposed procedure with different critical vectors (BH_{γ^*},Exp_{γ^*}, $AORC_{\gamma^*}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 0$. The black solid line corresponds to m_1. Only confidence bounds valid for m_1 are shown.

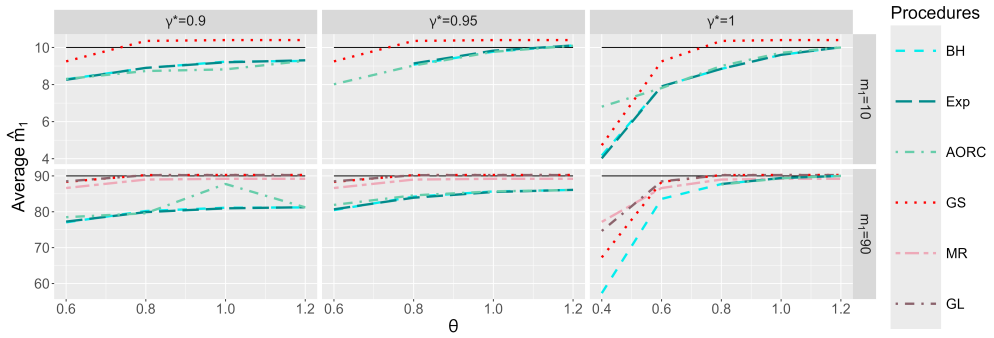

Figure 5: Dependency setting: p-values have correlation $\rho = 0.3$ and effect size θ is known. Average lower confidence bound $\hat{m}_1$ obtained from the proposed procedure with different critical vectors (BH_{γ^*},Exp_{γ^*}, $AORC_{\gamma^*}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 50$. The black solid line corresponds to m_1. Only confidence bounds valid for m_1 are shown.

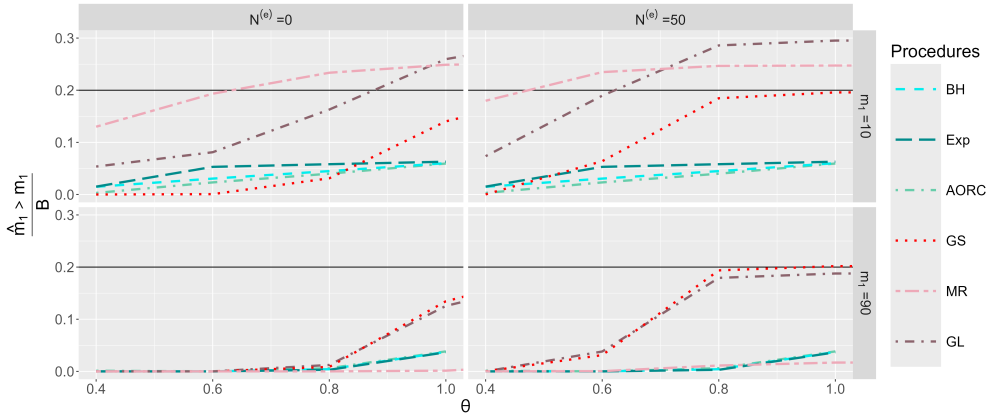

Figure 6: Misspecification of F: p-values have correlation $\rho = 0$ and the true effect size is overestimated ($\hat{\theta} = \theta + 0.1$). Proportion of iterations with $\hat{m}_1 > m_1$, where $\hat{m}_1$ is the lower confidence bound for the number of true discoveries m_1 obtained from the proposed procedure with different critical vectors (BH_{opt}, Exp_{opt}, $AORC_{opt}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 0$. The total number of iterations is $B = 10{,}000$. The black solid line corresponds to the significance level $\alpha = 0.2$.

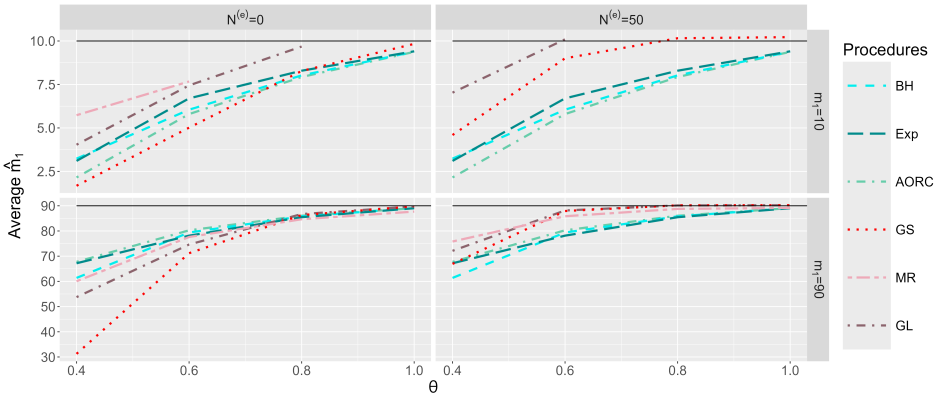

Figure 7: Misspecification of F: p-values have correlation $\rho = 0$ and the true effect size is overestimated ($\hat{\theta} = \theta + 0.1$). Average lower confidence bound $\hat{m_1}$ for the number of true discoveries m_1 obtained from the proposed procedure with different critical vectors (BH_{opt}, Exp_{opt}, $AORC_{opt}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 0$. The black solid line corresponds to m_1. Only confidence bounds valid for m_1 are shown.

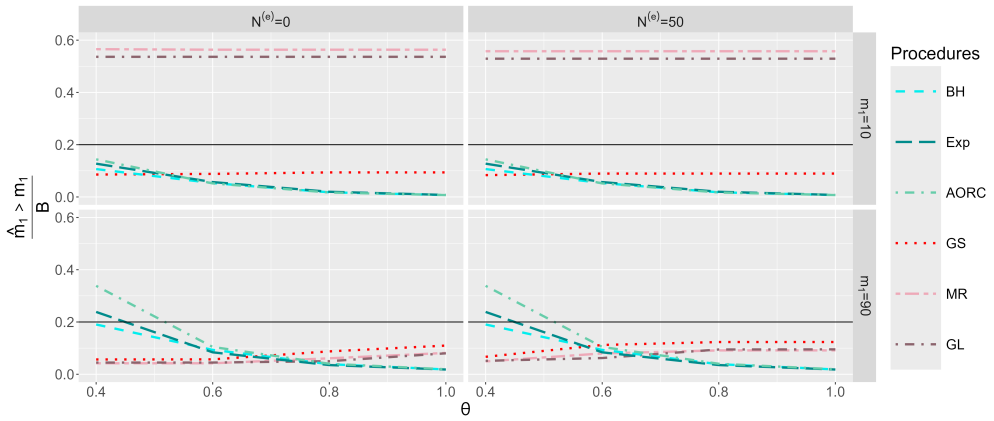

Figure 8: Dependency and misspecification of F: p-values have correlation $\rho = 0.9$ and the true effect size is overestimated ($\hat{\theta} = \theta + 0.1$). Proportion of iterations with $\hat{m_1} > m_1$, where $\hat{m_1}$ is the lower confidence bound for the number of true discoveries m_1 obtained from the proposed procedure with different critical vectors (BH_{opt}, Exp_{opt}, $AORC_{opt}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 0$. The total number of iterations is $B = 10{,}000$. The black solid line corresponds to the significance level $\alpha = 0.2$.

Table 1: Independence setting: p-values have correlation $\rho = 0$ and the effect size θ is known. Sum of the expected values $\mathbb{E}_{\tilde{m}_1}[R\gamma^*]$ over all possible candidate values $\tilde{m}_1 \in \{0,\ldots,m\}$, as given in Eq. (9). Results are obtained from the proposed procedure with different families of critical vectors (BH_{γ^*}, Exp_{γ^*}, $AORC_{\gamma^*}$). Bold values correspond to the highest value of the sum for each θ.

	θ				
Critical vector	0.6	0.8	1	1.2	2
BH_{opt}	4720.33	4910.98	5011.96	**5058.00**	5065.01
$BH_{0.95}$	4535.58	4702.08	4786.21	4817.20	4820.65
$BH_{0.9}$	4372.23	4504.50	4569.98	4582.53	4584.15
$BH_{0.8}$	3928.18	4079.47	4113.83	4110.73	4106.56
BY_{opt}	4720.44	4911.31	5011.99	5057.98	5065.04
$BY_{0.95}$	4540.14	4700.22	4786.94	4817.05	4820.16
$BY_{0.9}$	4365.37	4505.53	4567.08	4583.50	4583.98
$BY_{0.8}$	3921.54	4065.40	4111.10	4106.36	4111.25
Exp_{opt}	4703.07	4909.88	**5012.18**	5057.61	5057.37
$Exp_{0.95}$	4548.57	4686.55	4783.07	4818.85	4878.56
$Exp_{0.9}$	4384.82	4481.32	4556.33	4589.78	4724.74
$Exp_{0.8}$	3939.10	4030.66	4084.48	4119.19	4349.03
$AORC_{opt}$	**4740.10**	**4915.25**	5012.17	**5058.00**	**5065.06**
$AORC_{0.95}$	4604.67	4718.93	4786.76	4815.32	4820.62
$AORC_{0.9}$	4455.21	4531.71	4388.23	4584.36	4586.82
$AORC_{0.8}$	4011.63	4080.76	4101.44	4105.72	4106.77

Table 2: Independence setting: p-values have correlation $\rho = 0$ and the effect size $\theta = 0.8$ is known. Average lower confidence bound $\hat{m}_1$ for the number of true discoveries m_1 obtained from the proposed procedure with different families of critical vectors (BH_{γ^*}, Exp_{γ^*}, $AORC_{\gamma^*}$) and the methods of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GS*) with $N^{[e]} = 0$ and $N^{[e]} = 50$. Bold values correspond to the highest value of $\hat{m}_1$ for each m_1 for $N^{[e]} = 0$. Only valid confidence bounds are displayed.

m_1	10	20	30	40	50	60	70	80	90
GS ($N^{[e]} = 0$)	8.28	16.87	25.78	35.08	44.66	54.46	64.67	75.15	86.31
MR ($N^{[e]} = 0$)	-	17.80	27.07	36.50	46.00	55.56	65.24	74.94	84.66
GL ($N^{[e]} = 0$)	**9.68**	**18.96**	28.34	37.78	47.42	57.16	66.89	76.74	86.72
BH_{opt}	8.87	18.80	28.78	38.85	48.85	58.72	68.60	78.30	87.95
BY_{opt}	8.87	18.80	28.79	38.85	48.85	58.72	68.60	78.30	87.95
Exp_{opt}	9.01	18.92	**28.86**	**38.88**	48.83	58.66	68.50	78.18	87.82
$AORC_{opt}$	8.75	18.66	28.66	38.78	**48.86**	**58.80**	**68.75**	**78.52**	**88.21**
$BH_{0.95}$	9.07	18.70	28.19	37.83	47.34	56.76	66.06	75.40	84.35
$BY_{0.95}$	9.06	18.69	28.17	37.81	47.31	56.74	66.04	75.38	84.33
$Exp_{0.95}$	9.13	18.69	28.10	37.70	47.12	56.55	65.78	75.16	84.07
$AORC_{0.95}$	9.03	18.69	28.22	37.91	47.52	56.96	66.38	75.68	84.71
GS ($N^{[e]} = 50$)	10.16	20.11	30.08	40.07	50.06	60.07	70.09	80.12	90.18
MR ($N^{[e]} = 50$)	-	18.30	28.00	37.72	47.70	57.57	67.84	78.18	88.76
GL ($N^{[e]} = 50$)	-	-	-	-	-	-	-	-	90.15

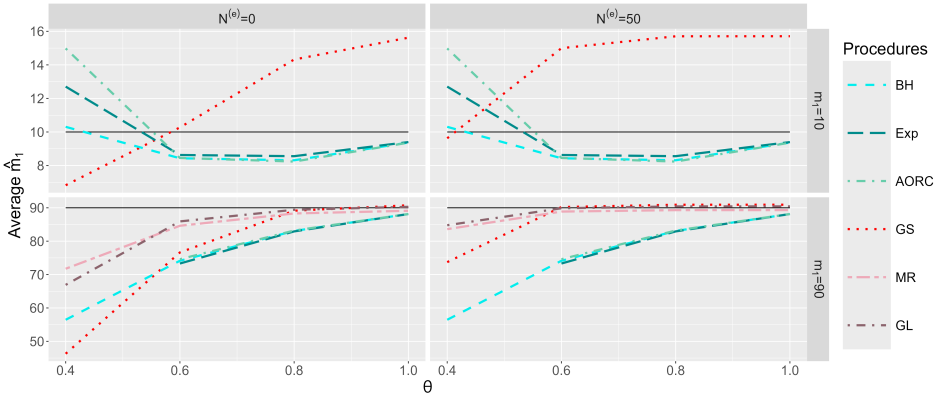

Figure 9: Dependency and misspecification of F: p-values have correlation $\rho = 0.9$ and the true effect size is overestimated ($\hat{\theta} = \theta + 0.1$). Average lower confidence bound $\hat{m_1}$ for the number of true discoveries m_1 obtained from the proposed procedure with different critical vectors (BH_{opt}, Exp_{opt}, $AORC_{opt}$) and the method of Goeman & Solari (2011) (*GS*) Meinshausen & Rice (2006) (*MR*) and Ge & Li (2012) (*GL*) with $N^{[e]} = 0$. The black solid line corresponds to m_1. Only confidence bounds valid for m_1 are shown.

4 Analysis of fMRI data

This section provides additional information on the analysis of fMRI data illustrated in Section 6 in the main article. First, we describe in greater detail the workflow of the analysis. Then we describe how the ROIs for the considered data set have been defined.

Pipeline for the analysis

The general process of the proposed fMRI analysis is illustrated in Figure 10. The goal of the analysis is computing a lower $(1-\alpha)$-confidence bound for the proportion of active voxels (TDP) within a pre-specified ROI. The analysis uses the measured bold time series Y_{ij} and the design matrix X_{ij} for each voxel i and each subject j. The distribution F under the alternative of the p-values obtained from the second level analysis is assumed to be known, i.e., the parameters that characterize F are assumed to be determined prior to the analysis.

Figure 11 further illustrates the last part of the process, where the contrasts obtained from the first level analysis are used to compute the TDP confidence bound. The $N = 140$ subjects are split into two sub-groups; the first is used to estimate the effect size θ, and the second to compute the confidence bound. If reliable prior information about the effect size are available, they can be used to compute the confidence bounds. In this case, sample splitting is unnecessary.

Definition of the ROI

ROIs have been defined using the clusters and peaks of activation reported by Schirmer et al. (2012) and Binder et al. (2000), that investigated human voices vs. non-human sounds. In Schirmer et al. (2012), we have considered only the four clusters containing at least 20 voxels with a total of seven peaks of activation. Additionally, we considered two clusters determined by Binder et al. (2000) with a total of eleven peaks of activation. Then we have defined the ROIs from the selected clusters, as following.

ROIs have been defined as either spheres or cuboids. Spherical ROIs have been defined such that they include all peaks of activation in the cluster and have a similar size in mm^3 as the clusters reported in the corresponding study. Note that only Schirmer et al. (2012) determined the size in mm^3 of the clusters. If the cluster contained more than one peak of activation, the center of the sphere has been defined as the mean voxel coordinate; if the cluster contained only one peak of activation the spherical ROI has been centered around it. If the spherical cluster had a voxel size that differed too much from the one given in Schirmer et al. (2012), a cuboid ROI has been defined, such that each peak of activation was the center of a cube and these cubes were connected. The size of the cubes has been defined such that the resulting size of the ROI was similar to the cluster size given in Schirmer et al. (2012). An overview of the different clusters, including their size and location, is given in Table 3 and illustrated in Figure 12.

We have used the "icbm2tal" transformation in GingerAle (Laird et al. (2010), Lancaster et al. (2007), see also `https://www.brainmap.org/icbm2tal/`) to transform the Talairach coordinates given in Schirmer et al. (2012) to MNI coordinates. We have used FSL (Jenkinson et al., 2012) to create the masks for the ROIs.

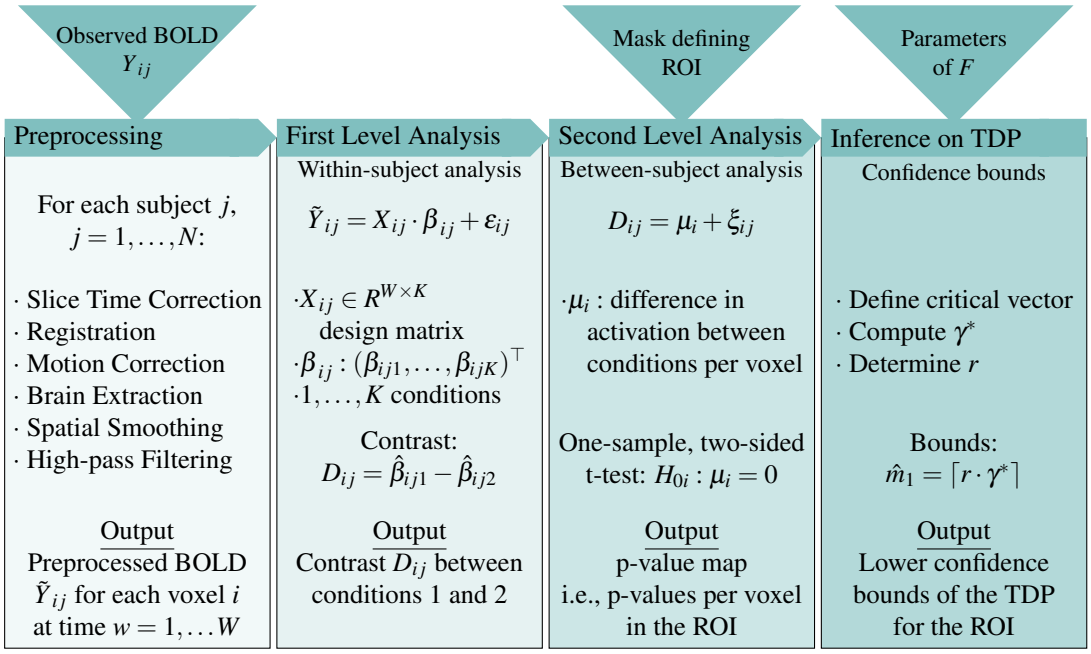

Figure 10: The general steps needed to compute a lower $(1-\alpha)$-confidence bound for the TDP within a given ROI. The inputs are, for each voxel and each subject, the measured BOLD time series Y_{ij} and the design matrix X_{ij}.

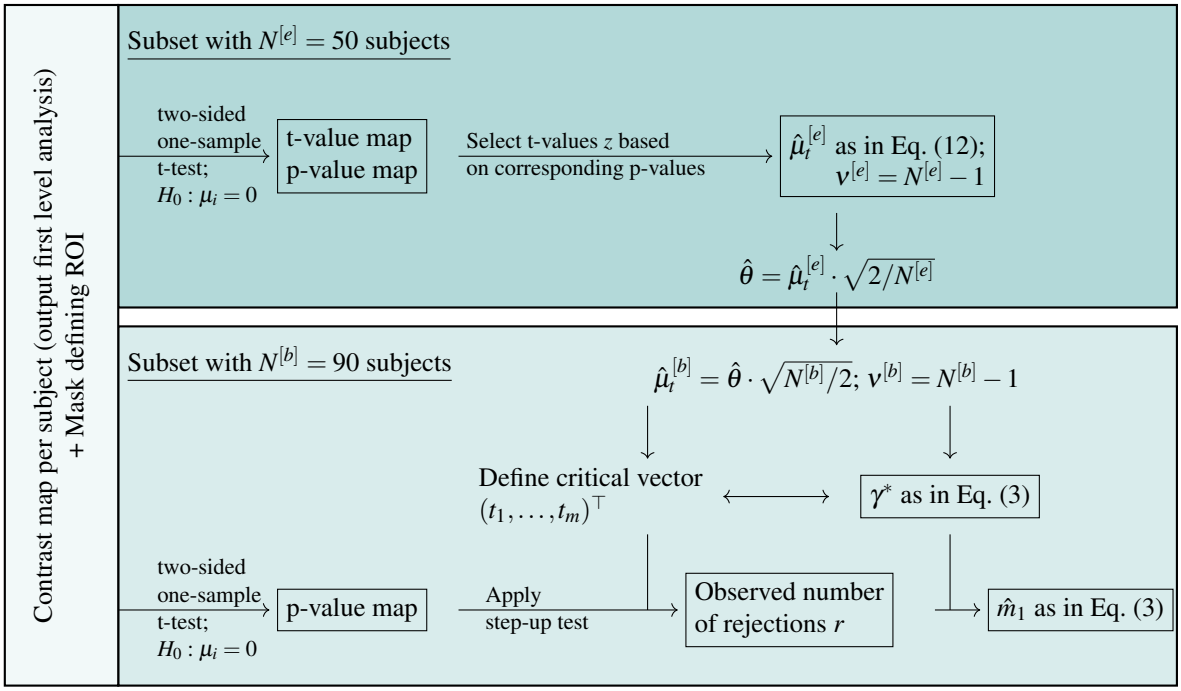

Figure 11: Illustration of the work flow to compute a lower $(1-\alpha)$-confidence bound for the TDP within a given ROI. The sample is split into two sub-samples to estimate the effect size and compute the lower confidence bounds separately. For each sub-sample, the inputs are the contrasts $D_{ij} = \mu_i + \xi_{ij}$ of each voxel i and subject j.

Table 3: Information about the regions of interest. The clusters are named according to the area of the brain in which the peak of activation is located, based on the Talairach Daemon Labels in FSL (Lancaster et al., 2000). Given is the shape of the ROI, the width (for cuboid) or radius (for spheres) of the ROIs, the number of peaks of activation and the size of the ROI in terms of number of voxels. Furthermore, for spherical ROI the voxel coordinates [x,y,z] of the center of each sphere are reported, for the cuboid ROI the coordinates of the peaks of activation are reported. The coordinates are given in the MNI152 space.

Reference	ROI	Shape	Radius/width in mm	Size	|Peaks|	Coordinates x	y	z
Schirmer et al. (2012)	L STG	Cuboid	6	162	3	-38	-38	10
						-44	-30	8
						-50	-20	2
	L AC	Sphere	4	33	2	-60	-18	12
	R AC	Sphere	8	257	1	52	-26	18
	R FG	Sphere	5	81	1	39	-39	-14
Binder et al. (2000)	R MTG	Sphere	8	257	6	64	-4	-8
	R STG	Sphere	9	389	5	58	-30	2

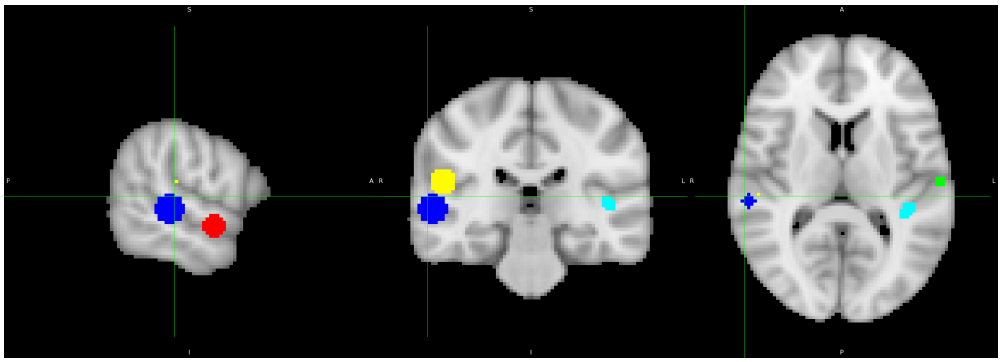

Figure 12: Location of the ROI in the brain. The turquoise region corresponds to the ROI in the Left STG, the green sphere corresponds to the ROI centered in the Left AC, the yellow sphere corresponds to the ROI centered in the Right AC, the blue sphere corresponds to the ROI centered in the Right MTG and the red sphere corresponds to the ROI centered in the Right STG. Note that in the picture on the left the right hemisphere of the brain is shown.

References

Binder, J., Frost, J., Hammeke, T., Bellgowan, P., Springer, J., Kaufman, J. & Possing, E. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, *10*(5), 512-528.

Ge, Y. & Li, X. (2012). Control of the False Discovery Proportion for Independently Tested Null Hypotheses. *Journal of Probability and Statistics*, *2012*, 320425.

Goeman, J. J. & Solari, A. (2011). Multiple Testing for Exploratory Research. *Statistical Science*, *26*(4), 584 – 597.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782-790. (20 YEARS OF fMRI)

Laird, A. R., Robinson, J. L., McMillan, K. M., Tordesillas-Gutiérrez, D., Moran, S. T., Gonzales, S. M., Ray, K. L., Franklin, C., Glahn, D. C., Fox, P. T. & Lancaster, J. L. (2010). Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: Validation of the Lancaster transform. *NeuroImage*, *51*(2), 677-683.

Lancaster, J. L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J. C. & Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*, *28*(11), 1194-1205.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., Kochunov, P. V., Nickerson, D., Mikiten, S. A. & Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120-131.

Meinshausen, N. & Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *The Annals of Statistics*, *34*(1), 373 – 393.

Roquain, E. & Villers, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *The Annals of Statistics*, *39*(1), 584 – 612.

Schirmer, A., Fox, P. M. & Grandjean, D. (2012). On the spatial organization of sound processing in the human temporal lobe: A meta-analysis. *NeuroImage*, *63*(1), 137-147.