As part of a 2009 Annals of Statistics paper, Gavrilov, Benjamini, and Sarkar report results of simulations that estimated the false discovery rate (FDR) for equally correlated test statistics using a well-known multiple test procedure. In our study we estimate the distribution of the false discovery proportion (FDR) for the same procedure under a variety of correlation structures among multiple dependent variables in a non-Normal context. Specifically, we study the mean (FDR), skewness, kurtosis, and percentiles of the FDP in the distribution of multiple comparisons that give rise to correlated non-normal statistics when results at several time points are being compared to baseline. If the FDR achieves its nominal value, any aspect of the distribution of the FDP depends on the interaction between the effect size and the correlation of test statistics, the proportion of true nulls, and number of dependent variables. We show examples where the mean FDP (the FDR) is 10% as designed, yet there is a surprising probability of having 30% or more false discoveries. Thus, in a real experiment, the proportion of false discoveries could be quite different from the stipulated FDR.

### FDR doesn’t tell the whole story: Joint influence of effect size and covariance structure on the distribution of the false discovery proportions.

**Alan H. Feiveson**, Ph.D., James Feidler, Ph.D., Robert J. Ploutz-Snyder, Ph.D., PStat*  
*NASA Johnson Space Center, Houston, TX *University Space Research Association, Houston, TX

**Abstract**

As part of a 2009 Annals of Statistics paper, Gavrilov, Benjamini, and Sarkar report results of simulations that estimated the false discovery rate (FDR) under equally correlated test statistics using a well-known multiple test procedure. In our study we estimate the distribution of the false discovery proportion (FDR) for the same procedure under a variety of correlation structures among multiple dependent variables in a non-Normal context. Specifically, we study the mean (FDR), skewness, kurtosis, and percentiles of the FDP in the distribution of multiple comparisons that give rise to correlated non-normal statistics when results at several time points are being compared to baseline. If the FDR achieves its nominal value, any aspect of the distribution of the FDP depends on the interaction between the effect size and the correlation of test statistics, the proportion of true nulls, and number of dependent variables. We show examples where the mean FDP (the FDR) is 10% as designed, yet there is a surprising probability of having 30% or more false discoveries. Thus, in a real experiment, the proportion of false detections could be quite different from the stipulated FDR.

### Background and Significance

Gavrilov, Benjamini, and Sarkar (GBS) [1] discuss the pros and cons of controlling the FDR in a multiple-testing scenario with a large number of variables which may be correlated. In particular, they prove the following: if a family of Monte Carlo simulations suggested by Benjamini, Krieger, and Yekutieli (BKY) [2] does indeed control the FDR to a desired level \( q \), when the test statistics are independently and identically distributed. GBS then show results of some simulations with equal correlation and normally distributed test statistics to show that the FDR of this simplified BKY procedure is fairly robust under this model condition.

### False Discovery Proportion and Rate

- **\( \alpha \) = \# of hypotheses rejected**
- **\( V \) = \# of true null hypotheses rejected**
- **\( R \) = \# of rejections**
- **\( m \) = \# of total tests**
- **\( FDP = V/ m \)**
- **\( FDR = V/ R \)**
- **\( q \) = \# of hypotheses rejected**
- **\( (3) \)**

### Study Summary

- **20, 40, 60, 100, 320 variables**
- **1000 simulated experiments per simulation run**
- **Dependence scenarios (DS):**
  - **1:** variables and tests are completely independent
  - **2:** variables are independent, but with dependent multiple comparisons as a result of repeated measures design
  - **3:** general covariance structure between variables (weighted sum of AR1 constant correlation, and two-stage Wishart) arising from multiple comparisons.
- **FDP attributes studied:** mean (FDR), median, skewness, kurtosis

### Simplified BKY Procedure

As defined in [1], the adaptive step-down procedure based on m tests is as follows:

1. Let \( q_0, q_1, \ldots, q_m \) be the required \( p \)-values.
2. Define critical values \( c_0, c_1, \ldots, c_m \) as follows:
   
   \( c_i = \alpha / (m-i+1) \)

3. Let \( q_i = \min \{ q \in (0,1) | \sum_{j=1}^i q_j > c_i \} \)  

4. If \( q_i \leq \alpha \), stop and reject hypotheses with \( p \)-values \( q_0, \ldots, q_i \). Otherwise repeat step (2).

### Simulated Data Model

- **Longitudinal observations of \( k \) variables from \( \beta \) subjects at \( t \) times:**
  
  \( \mathbf{Y} = \mathbf{X} \mathbf{B} + \mathbf{Z \epsilon} \)

- **For \( k \) variables, \( \epsilon \sim N(0, \text{correlation matrix}) \)**

- **Model parameters:**
  
  \( \mathbf{B} \) = \[ \beta_1 \ldots \beta_k \] \( \mathbf{X} \) = design matrix \( \mathbf{Z} \) = random effects matrix

### Results

#### Average Power vs. Simulated FDR level

As expected, power increases with stipulated FDP increase and the increase is sharper for larger effect sizes and a greater percentage of non-nulls.

#### When does the FDR attain its nominal level?

- **\( m_0 / m \) = \# of true null hypotheses**

### Graphs by DS: 50 to 75

- **\( D \) = \[ \alpha \] = 0.01**
- **\( I \) = \# of quartiles**

### FDP Distributions by Dependence

- **When the effect size and \( m_0 / m \) are small, the FDP distribution can have extreme positive skewness, with a high probability of no rejections. In this case the FDR is also below its nominal value (larger FDR deficit). Skewness appears to increase under dependence.**

#### Example of FDP Distributions by Dependence

- **The best FDP distributions tend to arise when \( m_0 / m \) is small (little or no dependence) and especially when there is a small proportion of nulls. Here, the distribution is more symmetric and closer to the nominal FDR (0.01), and also has relatively little spread (small IQR).**

### Inter-quartile Range (IQR)

- **Dependent (Y) vs. independent (X)**

### Association of Dependence with FDP Characteristics

#### Effect Size

<table>
<thead>
<tr>
<th>Size</th>
<th>Mean Value</th>
<th>Median Value</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-1</td>
<td>0.025 - 0.15</td>
<td>0.045 - 0.15</td>
<td>0.16 - 0.2</td>
<td>0.15 - 0.2</td>
</tr>
<tr>
<td>D-2</td>
<td>0.025 - 0.15</td>
<td>0.045 - 0.15</td>
<td>0.16 - 0.2</td>
<td>0.15 - 0.2</td>
</tr>
<tr>
<td>D-3</td>
<td>0.025 - 0.15</td>
<td>0.045 - 0.15</td>
<td>0.16 - 0.2</td>
<td>0.15 - 0.2</td>
</tr>
</tbody>
</table>

### Conclusions

1. **When effect sizes and the proportion of nulls are both small, the actual FDR can be a lot smaller than its nominal value.**

2. **Even when the FDR is close to its nominal value, the FDP distribution can have extreme skewness opening up the possibility of making occasional large FDP in real experiments.**

3. **For fixed \( q \), dependency on \( m_0 / m \), skewness and other characteristics of the FDP distribution can be strongly associated with the degree of dependence between test statistics.**

4. **In exploring the effect of correlated dependent variables on functions of test statistics, one cannot assume that the correlation structure of the original variables is preserved in the derived (or transformed) dependent variables.**

5. **Consider controlling the \( k \)-FWER (probability of \( k \) or more rejections when \( q \) true) in the presence of moderate to extreme dependence, especially when one suspects a large proportion of null cases, but with relatively small effect sizes.**

### Limitations & Future Directions

1. **Our simulations have inherent limitations because it is not possible to investigate all plausible covariance structures.**

2. **In cases where \( H_0 \) was not true, means were set to either a constant or zero.**

3. **We focused on one multiple testing method (BKY), and primarily one value of the nominal test level (0.10).**

### Relating Correlation of Variables to Correlation of Dependent Statistics

- **Y = \mathbf{X} \mathbf{B} + \mathbf{Z \epsilon}**

- **Correlation of test \( \mathbf{Y} \) vs. \( \mathbf{X} \)**

### References


