Sex difference tests

There are several ways to identify sex differences in stratified associations (Winkler et al. 2017). However, some tests are not valid because they violate simulated estimations of type I error rates, and some others are underpowered to detect differences in specific scenarios. Briefly, Winkler et al. 2017 tested for seven different approaches to detect sex differences: one that tests for sex differences in all associations, and six that test for sex differences in a subset that passes a filtering criteria first. From those six that filter particular associations, half consider the filtering and sex difference test applied to the same samples (*one-stage approach*), while the other half split the dataset into one that it is used to filter and another that it is used to test for sex differences (*two-stage approach*).

The test used to identify sex differences in associations is the *difference test*:

$$Z_{diff} = \frac{\beta_1 - \beta_2}{\sqrt{se_1^2 + se_2^2}}$$

The filtering tests are the following

- *Overall association*:

$$Z_{overall} = \frac{\frac{\beta_1}{se_1^2} + \frac{\beta_2}{se_2^2}}{\sqrt{\frac{1}{se_1^2} + \frac{1}{se_2^2}}}$$

- *Stratified association*:

$$Z_1 = \frac{\beta_1}{se_1} \quad \text{or} \quad Z_2 = \frac{\beta_2}{se_2}$$

- *Alternative joint association*:

$$C_{joint} = (\frac{\beta_1}{se_1})^2 + (\frac{\beta_2}{se_2})^2$$

The are two different α to determine whether a result is significant or not, or passes a predefined criteria. α_{Diff} is a Bonferroni corrected alpha based on M number of tests, therefore $0.05/M$. On the other hand, α_{Filter} is a predefined threshold. The authors evaluated different thresholds for the filtering to assess differences in power, which are summarized at the end. The seven different approaches evaluated are the following:

- [Diff$_{\alpha_{\text{Diff}}}$]: Difference test applied to all possible associations.
- [Overall$_{\alpha_{\text{Filter}}} \to$ Diff$_{\alpha_{\text{Diff}}}$]: one-stage approach that uses an overall filter before testing for differences.
- [Strat$_{\alpha_{\text{Filter}}} \to$ Diff$_{\alpha_{\text{Diff}}}$]: one-stage approach that uses a stratified filter before testing for differences.
- [Joint$_{\alpha_{\text{Filter}}} \to$ Diff$_{\alpha_{\text{Diff}}}$]: one-stage approach that uses a joint filter before testing for differences.

- $[\text{Overall}_{\alpha_{\text{Filter}}}] \to [\text{Diff}_{\alpha_{\text{Diff}}}]$: two-stage approach that uses an overall filter in one dataset, and test for differences in another.
- $[\text{Strat}_{\alpha_{\text{Filter}}}] \to [\text{Diff}_{\alpha_{\text{Diff}}}]$: two-stage approach that uses an stratified filter in one dataset, and test for differences in another.
- $[\text{Joint}_{\alpha_{\text{Filter}}}] \to [\text{Diff}_{\alpha_{\text{Diff}}}]$: two-stage approach that uses an overall filter in one dataset, and test for differences in another.

The article evaluates type I error rates and power. Therefore, the best approach is identified as the one that maintains type I error and shows the best power. The first result is whether these different approaches maintain valid type I error rates in simulated data. For the approach without filtering, $[\text{Diff}_{\alpha_{\text{Diff}}}]$, the empirical type I error is close to 5%. For the filtering one-stage approaches, $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$, also maintains a type I error close to 5%. However, $[\text{Strat}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$ and $[\text{Joint}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$, severely violate type I error rates (from 8% to 49%). For the filtering two-stage approaches, all approaches maintain correct type I error rates (close to 5%).

Note that because one-stage and two-stage overall filtering approaches, $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$ and $[\text{Overall}_{\alpha_{\text{Filter}}}] \to [\text{Diff}_{\alpha_{\text{Diff}}}]$, kept correct type I error rates, and because there is no hypothesis of why one should have better power, the authors only tested $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$ for subsequent power estimations.

In terms of power estimations, it largely depends on the type of sex difference (quantitative, qualitative, or pure). For qualitative differences, the $[\text{Diff}_{\alpha_{\text{Diff}}}]$ test, and the two-stage approaches, $[\text{Strat}_{\alpha_{\text{Filter}}}] \to [\text{Diff}_{\alpha_{\text{Diff}}}]$ and $[\text{Joint}_{\alpha_{\text{Filter}}}] \to [\text{Diff}_{\alpha_{\text{Diff}}}]$, perform better than the overall filtering approach $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$. However, the $[\text{Diff}_{\alpha_{\text{Diff}}}]$ approach always outperforms any other filtering approach under the different scenarios. On the other hand, when the differences point in the same direction, that is, for quantitative or pure differences, the overall association, $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$, shows a better power than any other approach.

In terms of varying filtering thresholds, α_{Filter}, for qualitative differences $[\text{Diff}_{\alpha_{\text{Diff}}}]$ outperforms any other approach irrespective of filtering. For pure differences, $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$ has the best power irrespective of the filtering threshold, showing the highest power when α_{Filter} is 0.05 to 1×10^{-4}. For quantitative differences, $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$ also shows the best power, with increasing power when decreasing the filtering threshold.

In terms of differences in sample size between sexes (balanced versus unbalanced designs), for qualitative differences, the $[\text{Diff}_{\alpha_{\text{Diff}}}]$ approach has the best power for balanced and slightly unbalanced designs (one group 5 times larger than the other), while the $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$ approach is better for extremely unbalanced designs (one group 5 times larger than the other). For pure differences, the best approach is the $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$ when the effect is in the larger group, but $[\text{Diff}_{\alpha_{\text{Diff}}}]$ when the effect is in the smaller group. Finally, for quantitative differences, the best approach is $[\text{Overall}_{\alpha_{\text{Filter}}} \to \text{Diff}_{\alpha_{\text{Diff}}}]$

irrespective of the scenario.

Recommendations

When there is no hypothesis on the type of sex difference, two approaches should be used in parallel:

- An association-wide sex difference test [$\text{Diff}_{\alpha_{\text{Diff}}}$], and
- An overall filtering first approach, [$\text{Overall}_{\alpha_{\text{Filter}}} \rightarrow \text{Diff}_{\alpha_{\text{Diff}}}$], with α_{Filter} set to 10^{-5}.

References

Winkler, Thomas W., Anne E. Justice, L. Adrienne Cupples, Florian Kronenberg, Zoltán Kutalik, Iris M. Heid, and GIANT consortium. 2017. "Approaches to Detect Genetic Effects That Differ between Two Strata in Genome-Wide Meta-Analyses: Recommendations Based on a Systematic Evaluation." PloS One 12 (7): e0181038. https://doi.org/10/gbqbnm.