Comparative assessment of different familial aggregation methods in the context of large and unstructured pedigrees

Christian X. Weichenberger*, Johannes Rainer, Cristian Pattaro, Peter P. Pramstaller and Francisco S. Domingues

Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy

*To whom correspondence should be addressed.

Associate Editor: Oliver Stegle

Received on February 28, 2018; revised on May 10, 2018; editorial decision on June 26, 2018; accepted on July 10, 2018

Abstract

Motivation: Familial aggregation analysis is an important early step for characterizing the genetic determinants of phenotypes in epidemiological studies. To facilitate this analysis, a collection of methods to detect familial aggregation in large pedigrees has been made available recently. However, efficacy of these methods in real world scenarios remains largely unknown. Here, we assess the performance of five aggregation methods to identify individuals or groups of related individuals affected by a Mendelian trait within a large set of decoys. We investigate method performance under a representative set of combinations of causal variant penetrance, trait prevalence and number of affected generations in the pedigree. These methods are then applied to assess familial aggregation of familial hypercholesterolemia and stroke, in the context of the Cooperative Health Research in South Tyrol (CHRIS) study.

Results: We find that in some situations statistical hypothesis testing with a binomial null distribution achieves performance similar to methods that are based on kinship information, while kinship based methods perform better when information is available on fewer generations. Potential case families from the CHRIS study are reported and the results are discussed taking into account insights from the performance assessment.

Availability and implementation: The familial aggregation analysis package is freely available at the Bioconductor repository, http://www.bioconductor.org/packages/FamAgg.

Contact: christian.weichenberger@eurac.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

Familial aggregation (FA) analysis is classically the first step in genetic epidemiology studies (Khoury et al., 1993). Aim of FA analysis is to identify groups of related individuals carrying the same symptoms or disease due to some underlying shared mechanism. A wide range of diseases has been previously analyzed with FA methods (Naj and Beaty, 2017), including cancer (Tokuhata and Lilienfeld, 1963), cardiovascular disease (Feinleib et al., 1977) and more recently autoimmune disease (Kuo et al., 2015) and insomnia (Jarrin et al., 2017). The increasing availability of large pedigrees for epidemiological research provides both an opportunity and a challenge for FA analysis, especially considering that such large pedigrees are often incomplete and with no regular structure. When pedigrees lack a regular structure, kinship-based methods have proven to be successful in modeling aggregation in terms of distance between individuals (Hill, 1980; Savica et al., 2016; Sveinbjornsottir et al., 2000). Even though research in this field initiated decades ago, little effort has been made to disseminate aggregation methods and make them available as software tools for use by the scientific community. We therefore have recently developed FamAgg (Rainer
et al., 2016), an open source R Bioconductor package providing both established and novel approaches to investigate FA for binary traits in large pedigrees. An exemplary application has demonstrated the potential of kinship-based methods in detecting FA. Nevertheless, a formal comparison between those methods in the context of large and unstructured pedigrees has not yet been made, making it unclear under which scenarios these methods may perform better or worse.

Here, we report on the performance assessment of different kinship-based approaches for FA. Formal comparisons were based on the structure of the pedigrees from the Minnesota Breast Cancer Family study (MBC) (Sellers et al., 1999), which are publicly available and allow reproducible tests. Using such pedigree structure as a backbone, we simulated FA scenarios by generating randomly disease cases in families under a dominant model of inheritance. Methods were then tested for their capability to identify these disease cases in case families among a set of control families with randomly assigned diseased individuals. Finally, we provide applications of the methods to a real context by assessing the aggregation of familial hypercholesterolemia and stroke in the context of large and unstructured pedigrees has not yet been made, making it unclear under which scenarios these methods may perform better or worse.

Here, we report on the performance assessment of different kinship-based approaches for FA. Formal comparisons were based on the structure of the pedigrees from the Minnesota Breast Cancer Family study (MBC) (Sellers et al., 1999), which are publicly available and allow reproducible tests. Using such pedigree structure as a backbone, we simulated FA scenarios by generating randomly disease cases in families under a dominant model of inheritance. Methods were then tested for their capability to identify these disease cases in case families among a set of control families with randomly assigned diseased individuals. Finally, we provide applications of the methods to a real context by assessing the aggregation of familial hypercholesterolemia and stroke in the Cooperative Health Research in South Tyrol (CHRIS) study (Pattaro et al., 2015), the former representing a disease that is still largely under-diagnosed and under-treated (Abul-Husn et al., 2016; Nordestgaard et al., 2013).

## 2 Materials and methods

### 2.1 Dataset

The performance assessment analysis is based on the pedigree structure of the publicly available MBC dataset (Sellers et al., 1999), which consists of genealogic information on 426 independent families whose founders entered a longitudinal study on breast cancer in the state of Minnesota (USA) in 1944. After removing 10,060 individuals with no gender information or no relatives (singleton), 18,021 related subjects were available. We further dropped 11 families for different practical reasons, leaving 415 families with 16,719 individuals available for analysis. These families are organized into two to five generations with a median family size of 36 individuals.

### 2.2 Performance assessment

Assessment is set up such that aggregation methods are required to identify families characterized by a Mendelian disease (case families) among control families with an unspecific random disease background. The choice of a Mendelian trait presents a suitable test setup, as it reflects a real-world scenario where kinship plays an important role among affected individuals.

We create case families in the MBC dataset by running RarePedSim software (Li et al., 2015) to randomly generate disease cases under a Mendelian autosomal dominant model of inheritance (briefly, Mendelian) with full penetrance and no phenocopy effects. The result is at least one affected founder (one individual case) in each case family, who passes on the trait to descendants according to the aforementioned model.

Three parameters are allowed to vary in order to generate a range of scenarios: the disease penetrance (Q), the disease prevalence (R) and the number of generations (G). Penetrance Q describes the fraction of cases that also express the phenotype. We test scenarios under Q = 100% (full penetrance), 60% (moderate) and 30% (low). For a case family, penetrance Q < 100%, is achieved by randomly turning (100 – Q)% of the affected individuals to non-affected. Prevalence R is defined as the simple proportion of disease cases on the total number of subjects in the pedigrees. The term prevalence is used for convenience and it is not referred to any reference population. For control families, we randomly choose affected individuals by simulating a Bernoulli trial with success probabilities P = 1/10 (10%), 1/16 (6.25%), 1/25 (4%), 1/50 (2%) and 1/80 (1.25%), such that the proportion of affected individuals in the overall population corresponds to prevalence R = P. Ultimately, the third parameter controls the number of generations (G) that are affected by the Mendelian trait, and mimics a spontaneous mutation or a mutation that is introduced into the family by marriage. We start with a case family as generated by RarePedSim, and keep the disease status of a sub-branch with G generations. For the remaining family members the disease status is cleared and reassigned according to the prevalence model described above. This procedure is carried out for G = 2 and G = 3 generations.

### 2.3 Familial aggregation analysis

The assessment evaluates the performance of several tests for FA without stratification provided by the FamAgg Bioconductor package (Rainer et al., 2016), which utilize the kinship coefficient (Malécot, 1948), a measure that quantifies relationship between two individuals in a pedigree by computing the probability that an allele is shared identical-by-descent at a given locus. These methods have been previously described in detail (Rainer et al., 2016). Briefly, each method calculates a specific test statistic from the observed family data, and a sampling distribution under the null hypothesis is generated empirically by random sampling 50,000 times the same number of affected individuals from the pool of all 415 families. The genealogical index of familiality (IF) test calculates the mean kinship coefficient between all affected individuals of a family (Hill, 1980). For an individual i, the kinship sum (KS) test computes the sum of kinship coefficients for all other affected family members.

| Table 1. Number of generations used in the performance assessment for each penetrance/prevalence pair |
|---------------------------------|-----------------|-----------------|-----------------|
| Penetrance (Q)                  | Prevalence (R)  |
| 100% (full)                     | 60% (moderate)  | 30% (low)       |
| 10%                             | 3/all           | 3/all           | —               |
| 6.25%                           | 3/all           | 3/all           | —               |
| 4%                              | 2/3/all         | 2/3/all         | 2/3/all         |
| 2%                              | 2/3/all         | 2/3/all         | 2/3/all         |
| 1.25%                           | 2/3/all         | 2/3/all         | —               |

*No assessment was performed for penetrance/prevalence pairs indicated by a dash.*
Kinship group tests utilize the most distant relative of a specific affected individual to define a group of individuals more closely related to the affected individual. The group ratio (GR) test simply counts the number of affected persons in that group, whereas the group closest relative (GC) test records the kinship coefficient of the closest affected relative in the group. We have added a binomial test (PB) that estimates the probability to detect by chance a family of size \( n \) with \( \geq k \) affected members given trait prevalence \( R \), i.e. we perform hypothesis testing with a random variable \( X \sim B(n, R) \). This binomial test has been now integrated into the FamAgg package. Tests are assessed on their ability to identify the case family based on computed \( P \) values. This can readily be done for the IF and PB tests, which report \( P \) values on families only. For the KS, GR and GC tests that report a \( P \) value for each affected individual, we take the lowest \( P \) value within a family. We adjust for multiple hypothesis testing employing Benjamini-Hochberg correction at a false discovery rate (FDR) of 0.05. We notice that distinct unadjusted \( P \) values may cluster within the same FDR level. Therefore, unless otherwise noted, ranking is performed on unadjusted \( P \) values to preserve order and avoid ties. In addition, significance is reported at the aforementioned FDR. We use the notations \( P_{\text{raw}} \) and \( P_{\text{adj}} \) for unadjusted and adjusted \( P \) values, respectively.

### 2.4 Performance measures

For evaluation of a given aggregation method with fixed parameters \( G, Q \) and \( R \), results obtained for all 415 family sets are combined, giving rise to 415 cases and \( 415 \times 414 = 171 \, 810 \) controls. Given a \( P \) value cutoff \( t \), a binary classification is established based on the knowledge of cases and controls, defining true positives (TP) and false negatives (FN) (cases with \( P \) value \( \leq t \) and \( > t \), respectively), and false positives and true negatives (controls with \( P \) value \( \leq t \) and \( > t \), respectively) (Baldi et al., 2000). Since this scenario defines a heavily imbalanced dataset with on average one case in 415 families, we analyze test performance by precision/recall curves, which put emphasis on detectability of true positives (Davis and Goadrich, 2006). These two performance measures are defined as follows: precision = \( \frac{\text{TP}}{\text{TP} + \text{FP}} \) and recall = \( \frac{\text{TP}}{\text{TP} + \text{FN}} \), whereas precision is also known as the positive predictive value, and recall as the true positive rate or sensitivity.

### 2.5 Application to the CHRIS study

We apply FA tests to data from the CHRIS study, a population-based study carried out in an alpine Italian valley (Pattaro et al., 2015). The recruitment area includes several municipalities, each one characterized by a main center and small villages and settlements. Recruitment is performed municipality-wise, and the pedigree structure reflects the geographical landscape of the area. All participants underwent blood drawing following overnight fasting. Medical history was reconstructed via standardized computer assisted questionnaires. Drugs used for treatment were barcoded scanned and classified according to the Anatomical Therapeutic Chemical (ATC) classification code. The study received ethical approval by the Ethical Committee of the Healthcare System of the Autonomous Province of Bolzano. All participants gave written informed consent.

After removing singletons, the dataset includes 4373 phenotyped individuals who reported two generations of ancestors resulting in 186 pedigrees with a total of 9024 individuals, out of which 4651 remain unphenotyped, but are used for establishing kinship (Noce et al., 2017). One exceptionally large family (termed ‘XL’) includes 3676 phenotyped and 3997 unphenotyped individuals. However, this family’s tree exhibits a highly horizontal structure with the furthest two individuals being separated by 65 individuals along the shortest path within only five generations. Excluding family XL, family sizes range between 3 and 45 members.

Pedigree information is used to identify familial clusters of hypercholesterolemia (HC), a condition with high serum levels of total cholesterol or low density lipoprotein (LDL), which presents a risk factor for arteriosclerosis and development of coronary heart disease (Carmena et al., 2004). We define HC as either total cholesterol level \( >290 \text{mg/dl} \) or low-density lipoprotein cholesterol level \( >190 \text{mg/dl} \) (Bertolini et al., 2017; Simon Broome Register Group, 1991), provided that triglyceride levels are \( <200 \text{mg/dl} \) (National Heart Lung and Blood Institute, 2002). We identify 296 HC cases out of 4979, corresponding to a disease prevalence of \( R = 5.94\% \). When removing singletons, we find 265 affected out of 4373, which gives \( R = 6.06\% \). Information on prescription of statins (ATC classification code C10) is not considered a case of HC, but is used for descriptive purposes during family analysis.

In the CHRIS study, stroke is assessed based on the Jackson Heart Study (Sempos et al., 1999) screening questionnaire. We utilize here the question ‘Have you ever been told by a doctor that you had a stroke?’, resulting in 54 cases of self-reported incidences of stroke (\( R = 1.23\% \)) (see Section 5 of the Supplementary Material for more details).

### 3 Results

#### 3.1 Performance assessment

Each FA test was applied to a set of 415 families, where one particular family (case family) included affected individuals following an autosomal dominant inheritance mode. Each family was assigned as case family and combined with the remaining 414 control families, resulting in 415 family sets. Case families were parameterized by trait penetrance \( Q \) and number of affected generations \( G \), and controls were generated according to prevalence \( R \). Aggregation tests were challenged to identify cases in 415 family sets for an extensive combination of parameters \( G, Q \) and \( R \). Certain parameter combinations were not examined, as they either resulted in cases indistinguishable from the background (e.g. high prevalence, low penetrance and two affected generations) or represented trivial cases (see Table 1).

#### 3.1.1 Assessment by family set

Two questions need to be addressed when analyzing the performance of an aggregation test for a given family set: first, what is the rank of the case family and second, is the reported case statistically significant? As mentioned in the Materials and methods section, ranking is performed on unadjusted \( P \) values, and significance is reported at a FDR of 0.05.

Figure 1 shows the assessment results for moderately and fully penetrant traits restricted to two and three generations. Generally, for moderate penetrance and high prevalence scenarios the aggregation tests are operating in the limit of detectability (Fig. 1a and b). For fully penetrant scenarios the performance of the different methods improves considerably (Fig. 1c and d), as expressed by a manifold increase of top three ranking cases (black and dark gray bars in Fig. 1). Furthermore, the performance is lower for \( G = 2 \) than when three generations are considered (Fig. 1b and d versus a and c). Tests PB and KS were frequently the best performing methods regarding the number of top ranking case families (Fig. 1, crosses and dots).
For a more challenging $G = 2$, KS test had an advantage over PB test. It also turned out that GC and GR tests reported a high number of significant cases among top three results, but usually they result in no further cases among the insignificant top three hits. However, for other parameter combinations the number of top three ranking case families found by GC and GR tests were similar to those identified by KS and PB tests. (See Supplementary Figures S1–S9 for all parameter sets.) We note in advance that despite their good performance in ranking case families in the top three, tests GC and GR are characterized by a high number of false positives, such that results should be treated cautiously. A detailed analysis of cases differentiated by rank, significance and number of affected individuals is given by Supplementary Figures S10–S41.

The overlap of top ranking cases among the different methods is provided in Figure 2. The agreement is lower for higher prevalence ($R = 6.25\%$), with a considerable number of unique top rank cases identified by PB, followed by IF and KS. The agreement between methods increased for lower prevalence ($R = 2\%$). Overall, with decreasing difficulty, expressed by higher $G$, lower $R$ or higher $Q$, we observed higher agreement in ranking cases on the top (we refer to Supplementary Figures S42–S73 for a comprehensive set of overlap analysis plots).

### 3.1.2 Assessment by binary classifier

In the previous section, we have compared aggregation tests by investigating the rank of the affected case family within a family set in combination with $P_{adj} \leq 0.05$ cutoff for assessing significant aggregation. In a second analysis, we have formulated the question of overall test performance as a binary classification problem based on $P_{adj}$, where for a fixed parameter set $G$, $Q$ and $R$ all 415 family sets were treated as a single set with 415 cases and 415 × 414 controls.
For selected parameters, Figure 3 summarizes the results by means of precision/recall curves where threshold $P_{\text{adj}} = 0.05$ is marked by a circle. The performance curves for the GC and GR tests start with precision value well below 1.0 and remain at constant precision until a certain recall value. This is due to a mixture of cases and controls that were both reported with the same, small $P_{\text{raw}}$ value, implying that these two tests report false positives even at moderate recall and high significance level. Frequently, we observe a short recovery phase indicated by increased precision, during which many cases are detected (e.g. Fig. 3a). After reaching a local maximum, precision drops due to an excess of false positives.

The PB and KS methods stood out by performing consistently better than the other three tests regarding precision and recall. Noteworthy, even though at low recall, tests PB, KS and IF are characterized by reasonably high precision values at the significance cut-off of $P_{\text{adj}} = 0.05$, distinguishing them as acceptable methods for FA analysis under the scenarios evaluated. The IF test usually lacks performance when compared to PB and KS tests, but showed some advantages when applied to small subfamilies with two generations or low to moderate trait penetrance (cf. Supplementary Figures S75–S96).

3.2 Familial aggregation in the CHRIS population study

We have applied different methods to detect FA to the CHRIS study in order to identify potential cases of familial hypercholesterolemia and stroke. We note that CHRIS pedigrees are characterized by nuclear families that were joined by a varying number of non-participating parents and grandparents.

3.2.1 Detection of familial cases of hypercholesterolemia

In the CHRIS study, we found the HC threshold levels (defined in Materials and methods) of total cholesterol, LDL cholesterol, and triglyceride as the 97th, 93rd, and 94th percentile, respectively. Overall, applying the proposed HC criteria resulted in a disease prevalence $R$ of 5.94% in the study.

Due to the large family XL, analyzing the CHRIS pedigrees represented a challenge, especially for methods based on families rather than individuals (IF and PB), as they might not have been able to fully capture the finer inheritance patterns between affected participants within that family. Family XL contained 236 out of all 265 affected participants. Interestingly, this family was identified by the IF test with $P_{\text{adj}} = 1.41 \times 10^{-3}$. The PB test did not report any significant family ($P_{\text{raw}}$(family XL) = 0.18), which may be indicative that the proportion of affected individuals found in family XL is in
Table 2. Top scoring family groups for HC with associated results of familial aggregation tests

<table>
<thead>
<tr>
<th>Group</th>
<th>N&lt;sup&gt;a&lt;/sup&gt;</th>
<th>N&lt;sub&gt;eff&lt;/sub&gt;&lt;sup&gt;d&lt;/sup&gt;</th>
<th>N&lt;sub&gt;eff&lt;/sub&gt;&lt;sup&gt;c&lt;/sup&gt;</th>
<th>P&lt;sub&gt;raw&lt;/sub&gt;</th>
<th>P&lt;sub&gt;adj&lt;/sub&gt;</th>
<th>Rank</th>
<th>P&lt;sub&gt;raw&lt;/sub&gt;</th>
<th>P&lt;sub&gt;adj&lt;/sub&gt;</th>
<th>Rank</th>
<th>P&lt;sub&gt;raw&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>5</td>
<td>5</td>
<td>2.00 × 10^-4</td>
<td>1.50 × 10^-3</td>
<td>1</td>
<td>2.61 × 10^-3</td>
<td>5.88 × 10^-3</td>
<td>4</td>
<td>5.88 × 10^-3</td>
</tr>
<tr>
<td>B</td>
<td>67</td>
<td>8</td>
<td>9</td>
<td>5.23 × 10^-4</td>
<td>1.55 × 10^-3</td>
<td>2</td>
<td>1.69 × 10^-2</td>
<td>6.60 × 10^-2</td>
<td>7</td>
<td>1.95 × 10^-2</td>
</tr>
<tr>
<td>C</td>
<td>63</td>
<td>7</td>
<td>8</td>
<td>2.05 × 10^-3</td>
<td>1.18 × 10^-3</td>
<td>6.5</td>
<td>1.47 × 10^-2</td>
<td>1.80 × 10^-2</td>
<td>5</td>
<td>3.60 × 10^-2</td>
</tr>
<tr>
<td>D</td>
<td>82</td>
<td>6</td>
<td>8</td>
<td>2.05 × 10^-3</td>
<td>1.87 × 10^-3</td>
<td>6.5</td>
<td>1.81 × 10^-2</td>
<td>1.30 × 10^-3</td>
<td>3</td>
<td>1.24 × 10^-3</td>
</tr>
<tr>
<td>E</td>
<td>114</td>
<td>9</td>
<td>10</td>
<td>2.19 × 10^-4</td>
<td>2.70 × 10^-4</td>
<td>10</td>
<td>3.92 × 10^-3</td>
<td>5.80 × 10^-3</td>
<td>15</td>
<td>1.54 × 10^-3</td>
</tr>
</tbody>
</table>

<sup>a</sup>Table is sorted by P<sub>raw</sub> of KS test and refers to all individuals with P<sub>adj</sub> = 0.0528 (the best P<sub>adj</sub> observed for KS test).
<sup>b</sup>Isolated group of individuals within family XL that maximizes the number of affected people sharing kinship.
<sup>c</sup>Number of phenotyped individuals in the group.
<sup>d</sup>Number of affected individuals that share kinship within the group.
<sup>e</sup>Overall number of affected individuals in the group, irrespective of kinship.
<sup>f</sup>Rank denotes the position in the specific aggregation test result list. In case of ties, ranks are averaged.
<sup>g</sup>Result as would be returned by the PB test for a family with N members and N<sub>eff</sub> affected family members and Bernoulli trial success probability P = 265/4373, which is the prevalence of HC in the CHRIS study excluding singletons.

 Familial aggregation tests were performed to identify groups of affected individuals. The table above shows the top scoring family groups for hypercholesterolemia (HC) with associated results of familial aggregation tests. Each group is represented by a set of affected individuals, and the table includes statistics for each group, such as the number of affected individuals (N), the number of affected individuals that share kinship (N<sub>eff</sub>), the P-values of the tests (P<sub>raw</sub> and P<sub>adj</sub>), and their rankings. The groups are sorted based on the P<sub>raw</sub> values of the Kolmogorov-Smirnov (KS) test.

The family tree in Figure 4 illustrates the relationships among the family members. The tree includes symbols to denote the affected status and age groups of the individuals. Non-phenotyped relatives are represented by empty symbols, while affected members are marked with black symbols. The tree is constructed to include the complete sibship plus spouses of all five affected individuals.

The performance assessment of the method for detecting familial cases of stroke is described in Section 3.2.2. The method was tested on the CHRIS study participants, and the results showed high precision in identifying families with stroke aggregation. The supplementary material provides further information on the performance assessment.

### 3.2.2 Detection of familial cases of stroke

When performing FA analysis of stroke within the CHRIS study, we have identified a nuclear family consisting of an affected mother (age group 80s) and both of her sons (age group 60s), all affected by stroke. In the performance assessment, this situation is best described by a fully penetrant trait in two generations with disease prevalence R = 1.25% (cf. Fig. 1d). All kinship-based methods identified this trio as a significant cluster (P<sub>adj</sub> < 0.05). Section 5 of the Supplementary Material provides further information that this family exhibits a high level of stroke aggregation.

### 4 Discussion

We have completed a performance assessment in order to investigate the ability of different methods to detect trait aggregation in large families, covering a range of trait prevalence and penetrance parameters, as well as the number of affected generations for a given set of pedigrees with phenotyped individuals. Overall, the PB test consistently displays high performance. Remarkably, this test was able to...
rank on top more cases for more challenging high trait prevalence ($R = 10\%$ and $R = 6.25\%$) scenarios than the other tests based on kinship. Conversely, the PB test was inferior to KS for incomplete penetrance and very low prevalence ($R \leq 2\%$) in up to three affected generations. Under such circumstances, kinship-based tests were more selective in detecting cases. In several scenarios with only two affected generations the IF test has demonstrated small advantages. For low and moderate penetrance, the KS test performed slightly better than the PB test. The ability of KS to identify individual cases in arbitrarily large families (in contrast to PB and IF), makes it a very good all-round aggregation test. Tests GC and GR clearly overestimate the significance of reported results, and should only be taken as an indicator to support hypothesis building based on the other tests’ results.

By looking at the overlap of the methods’ top ranking cases (Fig. 3, Supplementary Figures S42–S73), we identified several tendencies. First, tests GC and GR tend to report redundant results, as they frequently reported the same cases on top. Second, for low penetrance ($R \leq 4\%$), there was considerable agreement between the different tests. Most importantly however, for many parameter combinations, PB is the only test being able to identify many cases. Therefore, the current best strategy is to combine the sensitivity of PB with the kinship awareness of KS, and which should inspire future development of more powerful aggregation methods.

Analysis of the performance assessment suggests a strategy to efficiently exploit the advantages of each aggregation test, as we have demonstrated in the HIC and stroke examples. The results indicate that it is advisable to run all tests from the FamAgg package and compare the top ranking families. Significance is assigned by choosing a false discovery rate threshold $\alpha_{FDR}$ preferably $\leq 0.05$, since this allows to better interpret the findings by comparison with the performance assessment results. The higher the number of agreement between tests PB, KS and IF, the higher the confidence that can be put on the findings. The KS test should always list the family in question as top or very high ranking, and if the family size is not too large, the PB test should confirm the KS results. Hits also detected by the IF test are further strengthening the overall analysis, especially when they are significant at a false discovery rate $\leq 0.05$, since this test showed very high precision at this level of significance. (The IF test reported significant clustering of stroke and HIC cases in family XL.) On the other hand, due to the low precision of GC and GR aggregation methods, a positive hit from these tests should only be considered supportive for other tests’ results.

It should be recalled that in the performance assessment we observe many large families (median family size is 36 members) with all generations phenotyped, whereas the CHRIS study contains one large, the PB test should confirm the KS results. Hits also detected by the IF test are further strengthening the overall analysis, especially when they are significant at a false discovery rate $\leq 0.05$, since this showed very high precision at this level of significance. (The IF test reported significant clustering of stroke and HIC cases in family XL.) On the other hand, due to the low precision of GC and GR aggregation methods, a positive hit from these tests should only be considered supportive for other tests’ results.

Funding

The research and the CHRIS study were funded by the Department of Innovation, Research, Development and Cooperatives of the Autonomous Province of Bolzano-South Tyrol. The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/ Bolzano for covering the Open Access publication costs.

Conflict of Interest: none declared.

References


