



HAL
open science

Testing Fraternal Birth Order Effects and Antagonistic Effects for Homosexual Men: Power Comparison of Various Methods

Michel Raymond, Daniel Turek, Pierre-André Crochet

► **To cite this version:**

Michel Raymond, Daniel Turek, Pierre-André Crochet. Testing Fraternal Birth Order Effects and Antagonistic Effects for Homosexual Men: Power Comparison of Various Methods. Archives of Sexual Behavior, 2024, 10.1007/s10508-024-02820-w . hal-04490296

HAL Id: hal-04490296

<https://hal.science/hal-04490296v1>

Submitted on 5 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Testing Fraternal Birth Order Effects and Antagonistic Effects for Homosexual Men: Power
Comparison of Various Methods

Michel Raymond,¹ Daniel Turek,² and Pierre-André Crochet³

¹ Institut des Sciences de l'Evolution, University of Montpellier, CNRS, EPHE, IRD, place E.
Bataillon, 34095 Montpellier, France; e-mail:michel.raymond@umontpellier.fr

² Department of Mathematics, Lafayette College, Easton, PA, USA

³ CEFE, CNRS, University of Montpellier, EPHE, IRD, Montpellier, France

ACKNOWLEDGEMENTS

This is contribution 2024.043 of the Institut des Sciences de l'Evolution de Montpellier (UMR CNRS 5554). This work has been realized with the support of MESO@LR-Platform at the University of Montpellier, and benefited from the ISEM computing facility at ISI service.

Abstract

Research on the biological determinants of male homosexual preference has long realized that the older brother effect (FBOE, i.e. a higher fraternal birth rank of homosexuals) and the antagonist effect (AE, i.e. more fertile women have a higher chance of having a homosexual son) can both generate family data where homosexual men have more siblings and more older siblings than heterosexual men. Various statistical approaches were proposed in the recent literature to evaluate whether the action of FBOE or AE could be discriminated from empirical data, by controlling for the other effect. Here, we used simulated data to formally compare all the approaches that we could find in the relevant literature for their ability to reject the null hypothesis in presence of a specified alternative hypothesis (tests based on regression, Bayesian modeling, or contingency tables). When testing for the FBOE, the relative performance of the different tests was different depending on the specific function generating the older brother effect. Even if no tests were found to always perform better than the others, some tests performed systematically poorly, and some tests displayed a systematic high rate of type-I error. For testing the AE, the relative performance of the tests was generally not changed across all parameter values assayed, providing a clear ranking of the various proposed approaches. Pros and cons for each candidate test are discussed, taking into consideration power and the rate of type-I error but also practicability, the possibility to control for confounding variables and to consider alternatives hypotheses.

Key words: homosexuality, sexual orientation, fraternal birth order, older brothers, female fecundity, balancing selection

Introduction

Male homosexual orientation, i.e., preferential attraction of male subjects to same-sex partners for sexual intercourse and/or romantic relationships, is an evolutionary enigma as it is partially heritable and associated with a fertility cost. Its origin has long been discussed, with several evolutionary hypotheses being proposed, mostly involving kin selection or antagonistic pleiotropy (see Apostolou, 2020b; Barthes et al., 2015; Gavrillets & Rice, 2006; Raymond & Crochet, 2022 for reviews).

One intriguing feature of homosexual men is their higher fraternal birth rank (i.e., they have more older brothers on average) compared to heterosexual men. This can be explained by two non-exclusive mechanisms. Firstly, an antagonistic effect (AE), implying that more fertile women have a higher chance of having a homosexual son, has been proposed (Camperio-Ciani et al., 2004, 2009; Iemmola & Camperio-Ciani, 2009): homosexual men would then be overrepresented in large families and would thus have a higher mean birth rank than the population average. Secondly, a fraternal birth effect (FBOE), where each additional older brother increases the chances for a male embryo to develop later a homosexual orientation due to an immunoreactivity process, has been repeatedly put forward (Blanchard, 2018a; Bogaert et al., 2018; Bogaert & Skorska, 2011).

Whether the FBOE and AE are both at play in human populations or not remains a difficult question to tackle, as their effects can be difficult to distinguish when analyzing empirical data (Raymond et al., 2023). Indeed, higher fertility of mothers of homosexual men implies that when sampling homosexuals from a population, the mean birth order of homosexuals will be higher, on average, than the mean birth order of heterosexuals. If sex-ratio is unbiased, this will result in higher fraternal birth order of homosexual men.

Conversely, if fertility varies within a population independently of the occurrence of homosexuality, and if FBOE is at play, then sampling higher birth ranks (as is the case when sampling homosexuals in the presence of the FBOE) will generate a sample biased towards offspring from high-fertility mothers.

This is especially problematic in family data generated from population samples typically used to study these questions: homosexual and heterosexual men are sampled and they report the composition of their sibship. The importance of controlling for family size when studying birth order has long been recognized (e.g. Blanchard, 2014; Slater, 1962), although the importance of controlling for birth order when studying fertility has been often overlooked (e.g. Camperio-Ciani et al., 2004). Recently, in order to separate these AE and FBOE effects from empirical family data, several statistical methods have proposed to test for the presence of an FBOE when controlling for fertility, or to test the presence of AE when controlling for birth rank.

To test for the presence of the FBOE, controlling statistically for fertility variation, several methods have been proposed (Table 1). First, various regression models have been used, such as a Poisson regression, where the number of older brothers is the dependent variable, and sexual orientation is the variable of interest, with sibship size as a control variable (e.g., Nila et al., 2019); or a binomial regression on sexual orientation, with the number of older brothers being the variable of interest, and fertility being controlled by either one (sum of the other sibs, e.g., Vanderlaan et al., 2017), or several variables (older sisters, younger brothers, younger sisters, e.g., Blanchard & Bogaert, 1996); or sibship size, older sibs, and younger brothers, see Ablaza et al., 2022). Second, both FBOE and AE have been explicitly modeled in a Bayesian analysis using a hierarchical model, and support of the

FBOE in the presence of AE is estimated (Raymond et al., 2023). These methods rely on individual data so that control variables could be added to take into account confounding effects, such as variation of fecundity with the year of birth, variation of sibship composition with age, etc. Two methods have been proposed to test for the presence of FBOE on aggregated data, i.e., after pooling family data by sexual orientation for all individuals. Firstly, the total number of older brothers, relatively to older sisters, or relatively to all other sib categories, can be compared between homosexual and heterosexual men. The measure of the FBOE is an odd-ratio, named “odds of homosexuality” (see study 1 of Blanchard & Skorska, 2022), and OBOR (or Older Brother Odds Ratio) (Blanchard, 2018a; Blanchard et al., 2021), respectively, and the corresponding tests are Fisher exact test on a 2x2 contingency table. Another possible approach to control for fertility is to consider only individuals with a specific number of sibs, for example only one brother, and to contrast the sexual orientation of those with one older or one younger sib, using a 2x2 contingency table (Blanchard & Lippa, 2021; Khovanova, 2020).

Similarly, several approaches have been used to study variations in female fecundity associated with male homosexuality, controlling for a direct effect of birth rank (Table 1). First, a regression on the number of sibs, controlling for male birth rank, and sexual orientation as the dependent variable of interest (Raymond et al., 2023). Second, a binomial regression on sexual orientation, with the number of sibs being the variable of interest, and birth rank controlled with three variables (older sibs, older brothers, and younger brothers, see Ablaza et al., 2022). Third, a Bayesian modeling considering both FBOE and AE, as above, and support of the AE in the presence of FBOE is estimated (Raymond et al., 2023). Another possible approach to control for birth rank is to consider only individuals with a specific birth

rank, e.g., only first-born or second-born, and contrast their sexual orientation between large and small families (e.g., Blanchard, 2012; Blanchard et al., 2020; Khovanova, 2020).

These various methods have been used in different studies, and each one has some advantages and limitations. However, they have never been formally compared, and their relative power to reject the null hypothesis in presence of alternative hypotheses is not known. In this paper, we compare the power of these methods using simulated data. Specifically, we generate numerous family data with a specified level of FBOE and/or AE, then sample heterosexual and homosexual men and perform the various tests presented above. This process is repeated multiple times, allowing the computation of the power for each test, i.e., the proportion of rejection of the null hypothesis in presence of an alternative hypothesis (FBOE and/or AE)

Material and Methods

Simulating Family Data

A total of 6000 families were generated, with fecundity drawn from a Poisson distribution of parameter λ . For each offspring, sex was randomly assigned from the outcome of a binomial of parameter $\frac{1}{2}$, and birth order was recorded. FBOE was explicitly modeled by two different functions. First, by considering that the probability P to be homosexual is $P = p_0 + a \cdot ob$, where p_0 is the probability to be homosexual for a man without older brothers, ob is the number of older brothers, and $a > 0$ (corresponding to function f5 of Table S1 in Raymond et al., 2023). Second, by considering that $P = p_0$ for $ob = 0$, and $P = p_0 + a_1$ for $ob > 0$ (function f7 of Table S1 in Raymond et al., 2023). In all cases, FBOE was affecting only male offspring, so that an eventual older sister effect was not generated. AE was simulated using parameters α and β (both ≥ 0), and considering that for $\frac{1}{4}$ of the families, p_0 and fertility are higher, $p_0(1+\alpha)$ and

$\lambda(1+\beta)$, respectively. From the whole offspring population, a random sample of 200 heterosexual and 200 homosexual men was drawn, thus matching most empirical data sets where balanced numbers of homosexual and heterosexual subjects are sampled from the populations, before performing the various tests for detecting the presence of FBOE or AE and storing the corresponding p-values. This was replicated at least 200 times for a given set of parameters (a or a_1 , α , β). A value of $p_0 = 0.05$ and $\lambda = 4$ was considered throughout.

The various tests compared for their power relatively to various alternative hypotheses are indicated in Table 1, and detailed below.

Testing for FBOE

Regressions

R1. Generalized linear model with the number of older brothers as the response variable, which tested whether males of different sexual orientations had more or fewer older brothers (Nila et al., 2019). The variable of interest was the sexual orientation of sampled men (non-ordinal qualitative variable), and the control variable was the number of siblings (quantitative variable) of sampled men. Generalized linear regression was performed, using a quasiPoisson error structure. The significance of the variable of interest was calculated by removing it and comparing the resulting variation in deviance using the χ^2 test, as done by the function *Anova* from the *car* R package. Significant p-values indicative of FBOE corresponded to cases where more older brothers were associated with an homosexual orientation, relatively to heterosexual men (i.e., a positive slope estimate of the variable of interest).

R2-R4. Generalized linear model with sexual orientation as the response variable, which tested whether sexual orientation was explained by the number of older brothers. The

variable of interest was the numbers of older brothers of sampled men (quantitative variable). For R2 (Ablaza et al., 2022), the control variables were the number of siblings, the number of younger brothers, and the number of older sibs (all quantitative variables). For R3 (Blanchard & Bogaert, 1996), the control variables were the number of older sisters, the number of younger brothers, and the number of younger sisters (all quantitative variables). For R4 (Vanderlaan et al., 2017), the control variable was the number of other sibs (quantitative variable). Generalized linear regression was performed, using a binomial error structure. The significance of the variable of interest was calculated by removing it and comparing the resulting variation in deviance using the χ^2 test, as done by the function *Anova* from the *car* R package. Significant p-values indicative of FBOE corresponded to cases where a homosexual orientation was associated with more older brothers (i.e., a positive slope estimate of the variable of interest).

Goodness of fit for R1-R4 was measured using the McFadden pseudo-R-squared.

Bayesian modeling

B1. Both the FBOE and AE were explicitly and simultaneously modeled in a Bayesian analysis using a hierarchical model, as proposed by (Raymond et al., 2023), using the nimble R package (de Valpine et al., 2017). The birth order effect, i.e., the probability of displaying a homosexual preference according to the number of older siblings was modeled using function *f5* or *f7* (see above). Fertility was assumed to follow a Poisson distribution with rate parameter λ . AE was modeled as the possible presence of a subgroup displaying higher fertility and simultaneously a larger value of p_0 . This effect was modeled as an increase in λ by a factor $(1+\beta)$, conjointly with an increase in p_0 by a factor $(1+\alpha)$, with α and $\beta \geq 0$. For each individual, the probability that the AE applies followed a Bernoulli distribution with

parameter ϕ_{ae} . We implemented the model in a Bayesian framework by assigning uninformative (flat) prior distributions for all model parameters. Support for the FBOE, in the presence of the AE, was computed using Reversible-Jump Markov chain Monte Carlo (or RJMCMC, Green, 1995). RJMCMC is an extension of the standard MCMC methodology that allows the simulation of the posterior distribution on spaces of varying dimensions. One indicator variable dictated the presence or absence of the FBOE parameter (a or a_1 , depending of the FBOE function considered). RJMCMC was run at least 200,000 iterations, discarding the initial 10,000 iterations as burn-in, and only convergent chains (according to Geweke, 1991) were considered. The mean of the posterior distribution of the binary inclusion variable was used as an estimate of the support of the FBOE. Analyses were run in R 4.0.2 (R Core Team, 2020) using version 0.12.2 of the nimble package (de Valpine et al., 2020).

Contingency tables on aggregated data

T1. A 2x2 contingency table, contrasting sexual orientation and the total number of older brothers and older sisters (Blanchard & Skorska, 2022). Significant p-values for an excess of older brothers, relatively to older sisters, for homosexual men, was considered as indicative of FBOE.

T2. A 2x2 contingency table, contrasting sexual orientation and the total number of older brothers and other sibs (Blanchard, 2018b; Blanchard et al., 2021). Significant p-values for an excess of older brothers, relatively to other sibs, for homosexual men, were considered as indicative of FBOE.

T3. A 2x2 contingency table, contrasting sexual orientation for individuals with one older or one younger brother, for aggregated data restricted to individuals with only one brother (sibship size = 2) (Khovanova, 2020). Significant p-values for an excess of

homosexual men with one older brother, relatively to heterosexual men, were considered as indicative of FBOE.

For all 2x2 contingency tables, a Fisher's exact probability test was performed.

Testing AE

Regressions

R5. A generalized linear model with the number of siblings as the response variable tested whether males of different sexual orientation had more or fewer siblings (Raymond et al., 2023). The variable of interest was the sexual orientation of sampled men (non-ordinal qualitative variable), and the control variable was the male birth rank (quantitative variable) of sampled men. Generalized linear regression was performed, using a quasiPoisson error structure. The significance of the variable of interest was calculated by removing it and comparing the resulting variation in deviance using the χ^2 test, as done by the function *Anova* from the *car* R package. Significant p-values indicative of AE corresponded to cases where a higher sib numbers were associated with a homosexual orientation, relatively to heterosexual men (i.e., a positive slope estimate of the variable of interest).

R6. Generalized linear model with the sexual orientation as the response variables, which tested whether sexual orientation was explained by the number of siblings (Ablaza et al., 2022). The variable of interest was the number of siblings of sampled men (quantitative variable). The control variables were the number of older brothers, the number of younger brothers, and the number of older sibs (all quantitative variables), see R2. The significance of the variable of interest was calculated by removing it and comparing the resulting variation in deviance using the χ^2 test, as done by the function *Anova* from the *car* R package. Significant p-values indicative of AE corresponded to cases where a homosexual orientation was

associated with more siblings (i.e., a positive slope estimate of the variable of interest).

Goodness of fit for R5 and R6 was measured using the McFadden pseudo-R-squared.

Bayesian modeling

B2. The same Bayesian model as B1 was used, except now including an indicator variable controlling the inclusion or exclusion of AE. RJMCMC was again used for model fitting, where the indicator variable dictates the presence or absence of the two AE parameters (alpha and beta). RJMCMC was run at least 200,000 iterations, discarding the initial 10,000 iterations as burn-in, and only convergent chains (according to Geweke, 1991) were considered. The mean of the posterior distribution of the binary inclusion variable was used as an estimate of the support of the AE. Analyses were run in R 4.0.2 (R Core Team, 2020) using version 0.12.2 of the nimble package (de Valpine et al., 2020).

Contingency tables on aggregated data

T4. A Fisher's exact probability test on a 2x2 contingency table, contrasting sexual orientation for individuals with a higher or lower number of sibs than the population mean, for aggregated data restricted to individuals with no older brother (Khovanova, 2020). Significant p-values for an excess of homosexual men with a higher mean sib number, relatively to heterosexual men, were considered as indicative of AE.

Estimating Power and False Discovery Rate

For each statistical test (frequentist models), the proportion of p-values less than or equal to the rejection rate (0.05) was calculated. These provide an estimation of the power of each test (when an alternative hypothesis was included in the simulated data) or of the false discovery rate, or risk of type I error, when an alternative hypothesis was absent. We only counted the number of p-values supporting the AE or FBOE alternative hypotheses (i.e.,

excess of older brothers, or excess of sibs for homosexual men) so the expected false discovery rate was 2.5% (one-sided test). The power-equivalent for Bayesian modeling was taken to be the proportion of the replicated experiments for which the posterior mean of the RJMCMC indicator variable was above 50%. This approach considers a type-2 error as occurring when the RJMCMC posterior distribution suggests the effect being considered is less than 50% likely to have been present.

Results

Testing FBOE

When data were generated without FBOE (i.e., no alternative hypothesis), the power of each test corresponds to the rate of type I error or false positive. When AE was absent, all tests displayed a rate of type-I error not significantly higher than the expected value of 2.5 %. The only exception was T2, which displayed a rate of type-I error of 6.3 % (Table 2). When AE was present, the rate of type-I error increased for all tests, particularly for R3 and R4, reaching 19.0 % and 26.2 %, respectively (Table 2). The rate of type-I error was significantly higher than 2.5 % for R1 (5.2 %), T2, R3, and R4. The only exception was T3 for which the rate of type-I error decreased.

When data were generated in presence of only FBOE, the power to reject the null hypothesis increased with the intensity of FBOE and was affected by the specific function used to generate the FBOE. When function f7 was used to generate the FBOE, five tests (B1, R1, R3, R4, and T2) provided a higher power than the other ones (R2, T1, and T3), for all values of the parameter a_1 describing the strength of FBOE (Fig. 1A). For the best test (B1), the power was (Mean \pm SEM) 50.5 % \pm 3.5 for $a_1 = 0.025$, and 91.5 % \pm 2.0 for $a_1 = 0.05$. For the worst test (T3), these values were 2.0 % \pm 1.0 and 4.5 % \pm 1.5, respectively. For this test,

data reduction (men having not exactly one brother are discarded), across all parameter values, was on average 96.4 %. When function f5 was used to generate the FBOE, R1, R3, R4, and T2 provided a higher power than the other ones (B1, R2, T3, and T4), for all values of the parameter a describing the strength of FBOE (Fig. 1C). Nevertheless, all tests provided a power higher than 80 % for parameters values $a \geq 0.05$. The only exception was test T3, displaying a relatively low power (< 20 %) for all parameter values tested. For this test, data reduction, across all parameter values, was on average 96.7%. For R1-R4, across all parameter values, the mean McFadden's pseudo-R-squared was 22.91 %, 6.55 %, 6.55 %, and 6.16 %, respectively.

When AE was also present (with parameters $\alpha = \beta = 1$) in addition to FBOE (generated with function f7 or f5), the power decreased for all tests, although their relative ranking remained globally similar (Fig. 1B, 1D). The only exceptions were tests R3 and R4, for which the power increased, e.g., for the FBOE intensity $a = 0.0125$, from $65.5 \% \pm 3.4$ to $73.5 \% \pm 1.4$, and from $84.5 \% \pm 2.6$ to $90.0 \% \pm 2.1$, for R3 and R4, respectively. For T3, data reduction was on average 98.1% for function f7, and 98.1% for function f5. For R1-R4, across all parameter values, the mean McFadden's pseudo-R-squared was 27.05 %, 7.03 %, 7.03 %, and 6.64 %, respectively.

Testing AE

When data were generated without AE, the power of each test corresponds to the rate of type I error or false positive. When FBOE was absent, or when FBOE was present, either using function f7 or f5, all tests displayed a rate of type-I error not significantly higher than the expected value of 2.5 % (Table 3).

When population data were generated in presence of only AE, the power to reject the

null hypothesis increased with the intensity of AE. The most powerful test, for all parameter values, was B2, followed by R5, T4, and R6 (Fig. 2A). For AE corresponding to a two-fold increase of both the probability to be born homosexual and female fertility ($\alpha = \beta = 1$), the power was $100 \% \pm 0$, $45.5 \% \pm 3.5$, $18.5 \% \pm 2.8$, and $17.0 \% \pm 2.7$, for B2, R5, T4, and R6, respectively. For R5 and R6, across all parameter values, the mean McFadden's pseudo-R-squared was 21.31 % and 1.96 %, respectively. For T4, across all parameter values, data reduction (men with at least one older brother are discarded) was on average 64.5 %.

When FBOE was also present, generated with either function f7 ($a_1 = 0.1$) or function f5 ($a_1 = 0.1$), the power decreased for all tests, although their relative ranking remained globally similar (Fig. 2B and 2C). For R5 and R6, across all parameter values, the mean McFadden's pseudo-R-squared was 21.83 % and 3.34 %, respectively.

Discussion

Research on the biological determinants of male homosexual preference has long realized that the older brother effect (FBOE) and the antagonist effect (AE) can both generate family data where homosexual men have more siblings and more older siblings than heterosexual men. Various statistical approaches have been proposed in the recent literature to evaluate whether the action of FBOE or AE could be discriminated from empirical data, by controlling for the other effect. In this paper, we used simulated data to formally compare all the approaches that we could find in the relevant literature for their ability to reject the null hypothesis in presence of a specified alternative hypothesis.

Regressions

The various types of regression proposed differ for the dependent variable and the variable of interest (Table 1). For example, to test for the presence of FBOE, the dependent

variable could be the number of older brothers (R1, Poisson regression), or the sexual orientation (R2-R4, binomial regression): variables of interest are reversed, being sexual orientation (R1) or the number of older brothers (R2-R4), respectively. In all cases, control variables were present to control for birth rank (when testing AE), or to control for fertility (when testing FBOE).

Curiously, when testing for the presence of the FBOE, controlling for fertility variation, tests R3 and R4 displayed a relatively high rate of type-I error when AE was present (19 and 26%, respectively), thus incorrectly rejecting the null hypothesis (i.e., the absence of FBOE). This bias was not apparent when AE was absent, suggesting that these tests do not control adequately for the collinearity created between birth rank and number of sibs by variations in fertility as AE generates heterogeneity for fertility in the sample which is not captured easily by simple regression. This probably explains why all regressions tested displayed an increase in the rate of type-I error when AE was present, although these increases were modest for R1 and R2. The large increase observed for R3 and R4 remains to be fully understood.

The relatively modest performance of tests R2 and R6 is possibly explained by the presence of multiple correlated explanatory variables. The four explanatory variables of this regression model are the number of older sibs, older brothers, sibs, and younger brothers. The principle of this type of regression, according to the authors, is “to isolate the independent influences of siblings’ sex, birth order, and sibship size” (Ablaza et al., 2022). For test R2, the β coefficient associated with the number of older brothers is interpreted by the authors as the increase in the dependent variable (here, the probability of homosexuality) when there is an additional older brother, holding all other predictor variables constant (total siblings, older

siblings, and younger brothers): this additional older brother, according to the authors, must therefore be associated with the loss of an older sister. Similarly, for test R6, the β coefficient associated with the number of sibs is interpreted by the authors as the increase of the dependent variable when an additional sib is present, with all the other predictor variables constant (older sibs, older brothers, and younger brothers): this additional sib must thus be a younger sister according to the authors. This procedure has been proposed and used on a very large data set (Ablaza et al., 2022), but not formally evaluated until now, particularly for the consequences of multicollinearity. In a regression model, multicollinearity is known to lower the ability to detect significant coefficients, as it inflates standard errors of the estimates (e.g., Daoud, 2017). This could explain the relatively low power of these tests, particularly for R6, for which the explanatory variable (number of sibs) displayed a VIF (Variance Inflation Factor, Zuur et al., 2010) value higher than 5, a situation usually considered problematic, in 25% of the samples. The collinearity of R2 has been deliberately constructed in order to test for FBOE while controlling for fertility and SBOE. Thus type-I error is probably never inflated for those alternative hypotheses, although this is at the cost of a low power for the sample size considered here (200 homosexual and 200 heterosexual men). The behavior of R2, in terms of power and type-I error for alternative hypotheses and sample sizes not explored here, remain to be formally established. General linear regression methods are easily accessible (implemented in all statistical packages). The inclusion of confounding variables allowing a statistical control of some type of variation (such as age, socioeconomic level, etc.) is easy, although non-linear effects could be hard to implement when using pre-programmed current packages. However, they cannot fully account for the collinearity generated between birth rank and number of sibs by fertility variation.

Bayesian Modeling

Bayesian modeling using a hierarchical model can accommodate complex non-linear effects, although explicit expression of all effects is required. When FBOE was modeled using function f7 to generate the data, test B1 (coded with function f7 to detect FBOE) performed best relative to the other tests including regressions (discounting R3 and R4, see above), whether or not AE was present (Fig. 1A and B). However, when function f5 was used to generate the data, the performance of test B1 (coded with function f5 to detect FBOE) was unremarkable compared to the other ones, for low values of FBOE. Unfortunately, the rate at which additional older brothers increase the probability of homosexuality is not known (see discussion in Raymond et al., 2023). Until this function has been deciphered, it is uncertain whether or not this type of Bayesian modeling will perform best to detect FBOE.

To detect female fecundity associated with male homosexuality (AE), Bayesian modeling (test B2) displayed a higher power in all conditions evaluated, including in the presence of an FBOE (Fig. 2).

Contingency Tables

The performances of the various contingency tests were very different, some displaying a relatively high power (T2), and other performing poorly (T1, T3, and T4).

Under the null hypothesis, the power of Fisher's exact test on a 2x2 contingency table is 5%, if the rejection rate is 0.05. This is true if the variances of the variables for each cell are binomial (Bennett & Hsu, 1960). This is the case for T1, where the number of older brothers (*ob*) and older sisters (*os*) are compared across sexual orientation: for a given number of older sibs (n), the number of *ob* and *os* is drawn from a binomial $B(n, p)$, with $p = 1/2$; and the sum

of repeated binomials sampling (with the same p parameter) is also a binomial. As expected, the power under the null hypothesis (the rate of type-I error) for T1 is not different from the rejection rate (Table 2). If the variances of the variables for each cell are not binomial, type-I error could be inflated. This is the case when cells of the contingency table are drawn from a quasi-binomial distribution (Appendix 1, Fig. S1). Interestingly, the number of older or younger sibs, for a given sib size, follows a uniform discrete distribution (and not a binomial, as these sib classes are ordered), with a higher variance than binomial. A Fisher's exact test on a 2x2 contingency test of two independent samples contrasting older and younger sibs displays an inflated rate of type-I error, whether the test is performed with one draw for each line, or with the sum of multiple draws (Appendix 2, Fig. S2). Test T2 contrasts *ob* and *younger+os* over the sum of multiple draws, thus not corresponding to a simple known distribution. Indeed, the sum of several independent discrete uniform distributions is not uniform, with no simple formal description. The non-binomial distribution of the cell counts for T2 explains its inflated type-I error, suggesting that the use of this test should be discontinued.

The poor performance of T3 was probably due to the drastic data reduction required, as only individuals with exactly one brother (either older or younger) were retained, corresponding here to more than 95% data loss. The expected loss is a function of population fertility λ , as the proportion of individuals without exactly one brother is $1 - \lambda \cdot e^{-\lambda}/2$ (derived from Appendix 2 in Raymond et al., 2023). The minimum data loss is thus 81.6% (for $\lambda = 1$), although the presence of FBOE alters slightly this figure. Less drastic data loss was reported, from 12 to 68% (average 30%) (Blanchard et al., 2020), although these values were computed without excluding sisters (R. Blanchard, pers. communication), and thus

represent large underestimates. From the Canadian data published by Blanchard and Bogaert (1996), individuals with only one brother (and no sisters) represent 14.9 % of the data, thus the data loss for the T3 test is 85.1%. Similarly, from the Greek data from Apostolou (2020a) and the Indonesian data from Nila et al. (2019), data loss would be 74.1 % and 92.6 %, respectively.

Similarly, the poor performance of test T4 is also attributable to data reduction, as only firstborn were considered, corresponding here to 64.5 – 72.6 % data loss. For the Canadian, Greek, and Indonesian data, the corresponding data loss would be 55.6, 47.8, and 73.1%, respectively. The probability to sample a latter born is $1 - (1 - e^{-\lambda})/\lambda$, (derived from Appendix 1 in Raymond et al., 2023), giving an increasing function for the proportion of data loss in function of population fertility λ . Unless very large datasets are collected from low fertility populations, the T3 and T4 tests might be unable to reject the null hypothesis from a single sample, mostly if the FBOE (for T3) or AE (for T4) are of moderate intensity.

A contingency table test on aggregate data can incorporate aggregate confounding variables by switching to a regression equivalent test, allowing a statistical control of some type of variation (such as mean age for homosexual and heterosexual men).

Recommendations

When testing for the FBOE, some tests turned out to be inadequate due to their systematic high rate of type-I error (T2), or high rate of type-I error in presence of AE (R3 and R4). The performance of B1 was different depending on the specific function generating the older brother effect (best or medium performance when FBOE was modeled using function f7 or f5, respectively). Test R1 is a good compromise between practicability and efficiency, although it displays a slight increase in type-I error in presence of AE.

For testing the AE, the relative performance of the tests was generally not changed across all parameter values assayed, with or without FBOE, providing a clear ranking of the various proposed approaches (B2 > R5 > T4 > R6).

In real life, we don't know which effect is present and a good method should be able to identify both FBOE and AE if they are present and, most importantly, should have a low rate of type-I error. Bayesian modeling thus appears as the best option as it never suffers from an increased rate of type-I error (low false discovery rate) and has good power in most situations (only moderate power if the probability to be homosexual increases with the number of older brothers as modeled by function f7).

However, we acknowledge that applying Bayesian modeling requires computational skills that will probably restrain its adoption in the field. As an alternative, we recommend regressions R1 and R5, i.e., the simplest models where only the number of older brothers and the number of sibs are included in the model. Some care should be taken though as they do not entirely account for collinearity between birth rank and number of sibs; a good practice is probably to decrease α from 0.05 to 0.01 for the evaluation of the significance of the p-value. More complex models where several correlated variables are included (e.g., R2, R3, R4; R6) should be avoided as they suffer from either a high false discovery rate (R3, R4) or a low power (R2, R6).

Contingency tables are simple to use and could constitute an appealing alternative to regressions. Yet, they consistently have a low power: "null" results (absence of AE or FBOE) should be interpreted cautiously as these approaches probably require very large datasets to detect AE or FBOE in human population samples. Additionally, they cannot account for confounding variables. Thus, they do not offer any obvious advantage compared to regression

approaches. In any case, if using contingency tables, T2 should be avoided as it suffers from an increased false discovery rate.

As a word of caution, the current results and conclusions pertaining to the statistical tests are applicable exclusively to the alternative hypotheses that have been considered. We note that the results of statistical tests presented herein are interpretable only relative to the alternative hypothesis considered in each test. Specifying a different alternate hypothesis for any given test will generally alter the resulting power and/or type-I error rate. In such cases, the relative ranking of these tests could also change. For instance, a sororal birth order effect (SBOE) has not been examined here as a plausible alternative hypothesis, despite suggestions that, alongside the fraternal birth order effect (FBOE), a genuine SBOE exists in certain datasets (Ablaza et al., 2022; Blanchard & Lippa, 2007; 2021, but see Raymond et al., 2023), albeit to a lesser extent than the FBOE. An exploratory analysis was conducted to assess the FBOE under two conditions: 1) when an SBOE was also present (with the SBOE effect being one-third of the FBOE effect), a significant drop in the power of some tests occurred (e.g., R2) (see Appendix 3); and 2) when only an SBOE was present (without an FBOE), a substantial increase in the type-I error above the false discovery rate was observed for some tests (e.g., R4), while other tests remained unaffected (e.g., R2, T1, T3). Some tests exhibited varying outcomes (R1, B1, T2) depending on the function used to generate the SBOE (see Appendix 4). R2 was always conservative (no increase in type-I error) in presence of SBOE but at the cost of a relatively low power for the sample size considered (see above). Similarly, considering larger sample sizes could reduce power differences between the tests, and also could increase type-I error for some tests. Thus, for alternative hypotheses not thoroughly

considered here, and for different sample sizes, the above recommendation need to be re-evaluated.

Limits and Future Directions

We acknowledge several limitations of this study. First, a limited number of parameter values were explored, which limits the generality of the results, especially for lower values of population fecundity (λ). Second, the sample size was limited: tests T3 and T4, which require extensive data pruning, would likely perform better with a larger sample size. More generally, for very large samples, or for very large effect sizes of FBOE or AE, the power of all tests should converge to 100% : however, in some societies the sample size could be limited, even if a complete demographic survey is conducted (e.g., Camperio-Ciani et al., 2016), or the effect size could be modest (e.g., FBOE when the mean fecundity is low, Fig. 2A of Raymond et al. 2023). Third, the shape of the FBOE was modeled using only two functions (f5 and f7), which yielded different results for some tests: including others functions could potentially change the results. However, when the true function is known, it will be easier to compare the power of the various tests proposed. Fourth, all possible alternative hypotheses have not been considered, particularly the presence of a SBOE. In absence of SBOE, our conclusions remain valid. In presence of SBOE, or of any other alternative hypotheses not considered here, type-I error and power comparison should be undertaken to identify the best test. Fifth, the literature on AE and FBOE has not been exhaustively reviewed, so there may be alternative tests not considered here. Nevertheless, this study shows that power comparison is a useful step when designing a new test for FBOE or AE.

Declarations

Funding. The authors declare that they have received no specific funding for this study.

Conflicts of interest. The authors have no conflicts of interest to declare.

Availability of data and material. Not applicable.

Code availability. Script s and code are available online DOI : [10.5281/zenodo.7857346](https://doi.org/10.5281/zenodo.7857346)

Ethical approval: Not applicable.

Informed consent : Not applicable.

References

- Ablaza, C., Kabátek, J., & Perales, F. (2022). Are sibship characteristics predictive of same sex marriage? An examination of fraternal birth order and female fecundity effects in population-level administrative data from the Netherlands. *Journal of Sex Research*, 59(6), 671–683. <https://doi.org/10.1080/00224499.2021.1974330>
- Apostolou, M. (2020a). Does fraternal birth order predict male homosexuality, bisexuality, and heterosexual orientation with same-sex attraction? Evidence from a greek-speaking sample from Greece. *Archives of Sexual Behavior*, 49(2), 575–579. <https://doi.org/10.1007/s10508-019-01466-3>
- Apostolou, M. (2020b). *The evolution of same-sex attraction*. Springer.
- Barthes, J., Crochet, P.-A., & Raymond, M. (2015). Male homosexual preference: Where, when, why? *PLOS ONE*, 10(8), e0134817. <https://doi.org/10.1371/journal.pone.0134817>
- Bennett, B. M., & Hsu, P. (1960). On the power function of the exact test for the 2x2 contingency table. *Biometrika*, 47, 393–398.
- Blanchard, R. (2012). Fertility in the Mothers of Firstborn Homosexual and Heterosexual Men. *Archives of Sexual Behavior*, 41(3), 551–556. <https://doi.org/10.1007/s10508-011-9888-0>
- Blanchard, R. (2014). Detecting and correcting for family size differences in the study of sexual orientation and fraternal birth order. *Archives of Sexual Behavior*, 43(5), 845–852. <https://doi.org/10.1007/s10508-013-0245-3>
- Blanchard, R. (2018a). Fraternal birth order, family size, and male homosexuality: Meta-analysis of studies spanning 25 years. *Archives of Sexual Behavior*, 47(1), 1–15. <https://doi.org/10.1007/s10508-017-1007-4>

- Blanchard, R. (2018b). Older brothers and older sisters odds ratios in 36 samples of homosexual males. *Archives of Sexual Behavior*, *47*(4), 829–832.
<https://doi.org/10.1007/s10508-018-1160-4>
- Blanchard, R., Beier, K. M., Gómez Jiménez, F. R., Grundmann, D., Krupp, J., Semenyna, S. W., & Vasey, P. L. (2021). Meta-analyses of fraternal and sororal birth order effects in homosexual pedophiles, hebephiles, and teleiophiles. *Archives of Sexual Behavior*, *50*, 779–796. <https://doi.org/10.1007/s10508-020-01819-3>
- Blanchard, R., & Bogaert, A. F. (1996). Homosexuality in men and number of older brothers. *American Journal of Psychiatry*, *153*, 27–31. file:///D:/CDOCS-VALERIE/Valerie/Bibliographie/CPDF_evol/6163_blanchard_1996_am_jour_psychiatry.pdf
- Blanchard, R., Krupp, J., VanderLaan, D. P., Vasey, P. L., & Zucker, K. J. (2020). A method yielding comparable estimates of the fraternal birth order and female fecundity effects in male homosexuality. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1923), 20192907. <https://doi.org/10.1098/rspb.2019.2907>
- Blanchard, R., Lippa, R. A. (2007). Birth order, sibling sex ratio, handedness, and sexual orientation of male and female participants in a BBC internet research project. *Archives of Sexual Behavior*, *36*, 163–176. <https://doi.org/10.1007/s10508-006-9159-7>
- Blanchard, R., & Lippa, R. A. (2021). Reassessing the effect of older sisters on sexual orientation in men. *Archives of Sexual Behavior*, *50*, 797–805.
<https://doi.org/10.1007/s10508-020-01840-6>
- Blanchard, R., & Skorska, M. N. (2022). New data on birth order in homosexual men and women and a reply to Vilsmeier et al. (2021a, 2021b). *Archives of Sexual Behavior*.
<https://doi.org/10.1007/s10508-022-02362-z>

- Bogaert, A. F., & Skorska, M. (2011). Sexual orientation, fraternal birth order, and the maternal immune hypothesis: A review. *Frontiers in Neuroendocrinology*, *32*, 247–254.
- Bogaert, A. F., Skorska, M. N., Wang, C., Gabriele, J., MacNeil, A. J., Hoffarth, M. R., VanderLaan, D. P., Zucker, K. J., & Blanchard, R. (2018). Male homosexuality and maternal immune responsivity to the Y-linked protein NLGN4Y. *Proceedings of the National Academy of Sciences*, *115*(2), 302–306.
<https://doi.org/10.1073/pnas.1705895114>
- Camperio-Ciani, A., Battaglia, U., & Liotta, M. (2016). Societal norms rather than sexual orientation influence kin altruism and avuncularity in tribal Urak-Lawoi, Italian, and Spanish adult males. *The Journal of Sex Research*, *53*(2), 137–148.
<https://doi.org/10.1080/00224499.2014.993748>
- Camperio-Ciani, A., Corna, F., & Capiluppi, C. (2004). Evidence for maternally inherited factors favouring male homosexuality and promoting female fecundity. *Proceedings of the Royal Society London, B*, *271*, 2217–2221.
- Camperio-Ciani, A., Iemmola, F., & Blecher, S. R. (2009). Genetic factors increase fecundity in female maternal relatives of bisexual men as in homosexuals. *The Journal of Sexual Medicine*, *6*(2), 449–455. <https://doi.org/10.1111/j.1743-6109.2008.00944.x>
- Daoud, J. I. (2017). Multicollinearity and regression analysis. *Journal of Physics: Conference Series*, *949*(1), 012009. <https://doi.org/10.1088/1742-6596/949/1/012009>
- de Valpine, P., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn, C., Rodríguez, A., Temple Lang, D., Zhang, W., Paganin, S., & Hug, J. (2020). *Nimble: MCMC, particle filtering, and programmable hierarchical modeling* (0.10.1) [Computer software]. <https://CRAN.R-project.org/package=nimble>
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model

- structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413. <https://doi.org/10.1080/10618600.2016.1172487>
- Gavrilets, S., & Rice, W. R. (2006). Genetic models of homosexuality: Generating testable predictions. *Proceedings of the Royal Society London, B*, 273, 3031–3038.
- Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics* (Vol. 4, pp. 169–193). Clarendon Press. <https://doi.org/10.21034/sr.148>
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711–732. <https://doi.org/10.1093/biomet/82.4.711>
- Iemmola, F., & Camperio-Ciani, A. (2009). New evidence of genetic factors Influencing sexual orientation in men: Female fecundity increase in the maternal line. *Archives of Sexual Behavior*, 38, 393–399.
- Khovanova, T. (2020). On the mathematics of fraternal birth order effect and the genetics of homosexuality. *Archives of Sexual Behavior*, 49, 551–555. <https://doi.org/doi.org/10.1016/j.jtbi.2004.04.021>
- Nila, S., Crochet, P.-A., Barthes, J., Rianti, P., Juliandi, B., Suryobroto, B., & Raymond, M. (2019). Male homosexual preference: Femininity and the older brother effect in Indonesia. *Evolutionary Psychology*. <https://doi.org/doi.org/10.1177/1474704919880701>
- R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. <https://www.R-project.org/>
- Raymond, M., & Crochet, P.-A. (2022). Carving non-proximal explanations for same-sex sexual orientation. *Archives of Sexual Behavior*. <https://doi.org/10.1007/s10508-022-02497-z>

- Raymond, M., Turek, D., Durand, V., Nila, S., Suryobroto, B., Vadez, J., Barthes, J., Apostolou, M., & Crochet, P.-A. (2023). Increased birth rank of homosexual males: Disentangling the older brother effect and sexual antagonism hypothesis. *Peer Community Journal*, 3, e22. <https://doi.org/10.24072/pcjournal.247>
- Slater, E. (1962). Birth order and maternal age of homosexuals. *Lancet*, 1, 69–71.
- Vanderlaan, D. P., Blanchard, R., Zucker, K. J., Massuda, R., Fontanari, A. M. V., Borba, A. O., Costa, A. B., Schneider, M. A., Mueller, A., Soll, B. M. B., Schwarz, K., Da Silva, D. C., & Lobato, M. I. R. (2017). Birth order and androphilic male-to-female transexualism in Brazil. *Journal of Biosocial Science*, 49(4), 527–535. <https://doi.org/10.1017/S0021932016000584>
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>

Table 1. Statistical tests used to evidence FBOE, controlling for fertility, or to evidence AE, controlling for birth rank. The variables describing sibship composition, ob, os, yb, ys, sib, and older_sib are the number of, respectively, older brothers, older sisters, younger brothers, younger sisters, all brothers and sisters, and older sibs. “so” is the binary variable describing sexual orientation. Subscripts H and h designate homosexual and heterosexual men, respectively. Ob, Os, and Othersibs are aggregated variables of ob, os, and os+younger sibs, respectively, across all sampled individuals. For the regression models or the regression modeling, the variable of interest is underlined.

| Name | Test type | Test principle | Effect size | References |
|---|--|---|---|--|
| Testing FBOE | | | | |
| R ₁ | Poisson regression | ob ~ <u>so</u> + sib | Model estimates | Nila et al. 2019 |
| R ₂ | Binomial regression | so ~ <u>ob</u> +sib+older_sib+yb | Model estimates | Ablaza et al. 2022 |
| R ₃ | Binomial regression | so ~ <u>ob</u> +os+yb+ys | Model estimates | Blanchard & Bogaert 1996 |
| R ₄ | Binomial regression | so ~ <u>ob</u> +other_sib | Model estimates | VanderLaan et al. 2017 |
| B ₁ | Bayesian modeling | data ~ <u>FBOE</u> + AE | Mean of posterior distribution of the RJMCMC binary inclusion variable | Raymond et al. 2023 |
| T ₁ | Fisher’s exact test on 2x2 contingency table | $\begin{array}{cc} & \text{H} & \text{h} \\ \text{Ob} & \cdot & \cdot \\ \text{Os} & \cdot & \cdot \end{array}$ | $\text{Odd Ratio} = \frac{\frac{Ob_H}{Os_H}}{\frac{Ob_h}{Os_h}}$ | Blanchard & Skorska, 2022 |
| T ₂ | Fisher’s exact test on 2x2 contingency table | $\begin{array}{cc} & \text{H} & \text{h} \\ \text{Ob} & \cdot & \cdot \\ \text{Other sibs} & \cdot & \cdot \end{array}$ | $\text{OBOR} = \frac{\frac{Ob_H}{Othersibs_H}}{\frac{Ob_h}{Othersibs_h}}$ | Blanchard, 2018a; Blanchard et al., 2021 |
| T ₃ | Fisher’s exact test on 2x2 contingency table | $\begin{array}{cc} & \text{H} & \text{h} \\ \text{ob}=1 & \cdot & \cdot \\ \text{yb}=1 & \cdot & \cdot \end{array}$ | Odd ratio | Khovanova, 2020 |
| (data restriction: only men with 1 brother) | | | | |
| Testing AE | | | | |

| R ₅ | Poisson regression | $sib \sim \underline{so} + ob$ | Model estimates | Raymond et al. 2023 | | | | | | | | | |
|----------------|--|--|--|---------------------|---|-----------|---|---|-----------|---|---|-----------|-----------------|
| R ₆ | Binomial regression | $so \sim ob + \underline{sib} + older_sib + yb$ | Model estimates | Ablaza et al. 2022 | | | | | | | | | |
| B ₂ | Bayesian modeling | Data \sim FBOE + <u>AE</u> | Mean of posterior distribution of the RJMCMC binary inclusion variable | Raymond et al. 2023 | | | | | | | | | |
| T ₄ | Fisher's exact test on 2x2 contingency table | <table border="1" style="display: inline-table; vertical-align: middle;"> <thead> <tr> <th></th> <th>H</th> <th>h</th> </tr> </thead> <tbody> <tr> <td>sibs>Mean</td> <td>.</td> <td>.</td> </tr> <tr> <td>sibs<Mean</td> <td>.</td> <td>.</td> </tr> </tbody> </table> <p>(data restriction: only men without ob)</p> | | H | h | sibs>Mean | . | . | sibs<Mean | . | . | Odd ratio | Khovanova, 2020 |
| | H | h | | | | | | | | | | | |
| sibs>Mean | . | . | | | | | | | | | | | |
| sibs<Mean | . | . | | | | | | | | | | | |

Table 2. Rate of type-I error for FBOE tests. All tests are performed on $N = 600$ independent samples, from data generated without FBOE, and with or without AE, and the percentage of false rejection is indicated (Rate). This percentage is compared to the expected rate of type-I error for one-sided tests (2.5%), using a binomial exact test (p-value). Bold characters indicates significant values (p-value < 0.05) for higher values than the expected one (2.5%).

| Type of test | Without AE | | With AE | |
|--------------|------------|----------------------------------|----------|-------------------------------------|
| | Rate (%) | p-value | Rate (%) | p-value |
| R1 | 3.2 | 0.29 | 5.2 | $2 \cdot 10^{-4}$ |
| R2 | 2.7 | 0.79 | 3.3 | 0.19 |
| R3 | 3.2 | 0.29 | 19.0 | $< 10^{-4}$ |
| R4 | 3.7 | 0.09 | 26.2 | $< 10^{-4}$ |
| B1 | 0.5 | $3 \cdot 10^{-4}$ | 2.3 | 0.90 |
| T1 | 2.2 | 0.70 | 3.0 | 0.43 |
| T2 | 6.3 | $< 10^{-4}$ | 7.2 | $< 10^{-4}$ |
| T3 | 1.5 | 0.15 | 0.2 | $< 10^{-4}$ |

Table 3. Rate of type-I error for AE tests. All tests are performed on $N = 600$ independent samples, from data generated without AE and FBOE, and on $N = 200$ independent samples from data with FBOE using function f5 or f7. The percentage of false rejection is indicated (Rate), and is compared to the expected rate of type-I error for one-sided tests (2.5%), using a binomial exact test (p-value). No significant values ($p < 0.05$) for higher values than the expected are present.

| Type of test | Without FBOE | | With FBOE (f7) | | With FBOE (f5) | |
|--------------|--------------|-------------|----------------|---------|----------------|---------|
| | Rate (%) | p-value | Rate (%) | p-value | Rate (%) | p-value |
| B2 | 0.7 | 2.10^{-3} | 1.0 | 0.25 | 2.5 | 1.0 |
| R5 | 2.5 | 1.0 | 1.5 | 0.50 | 2.0 | 0.82 |
| R6 | 2.0 | 0.51 | 1.0 | 0.25 | 1.5 | 0.50 |
| T4 | 1.3 | 0.07 | 4.5 | 0.10 | 4.0 | 0.17 |

FIGURES

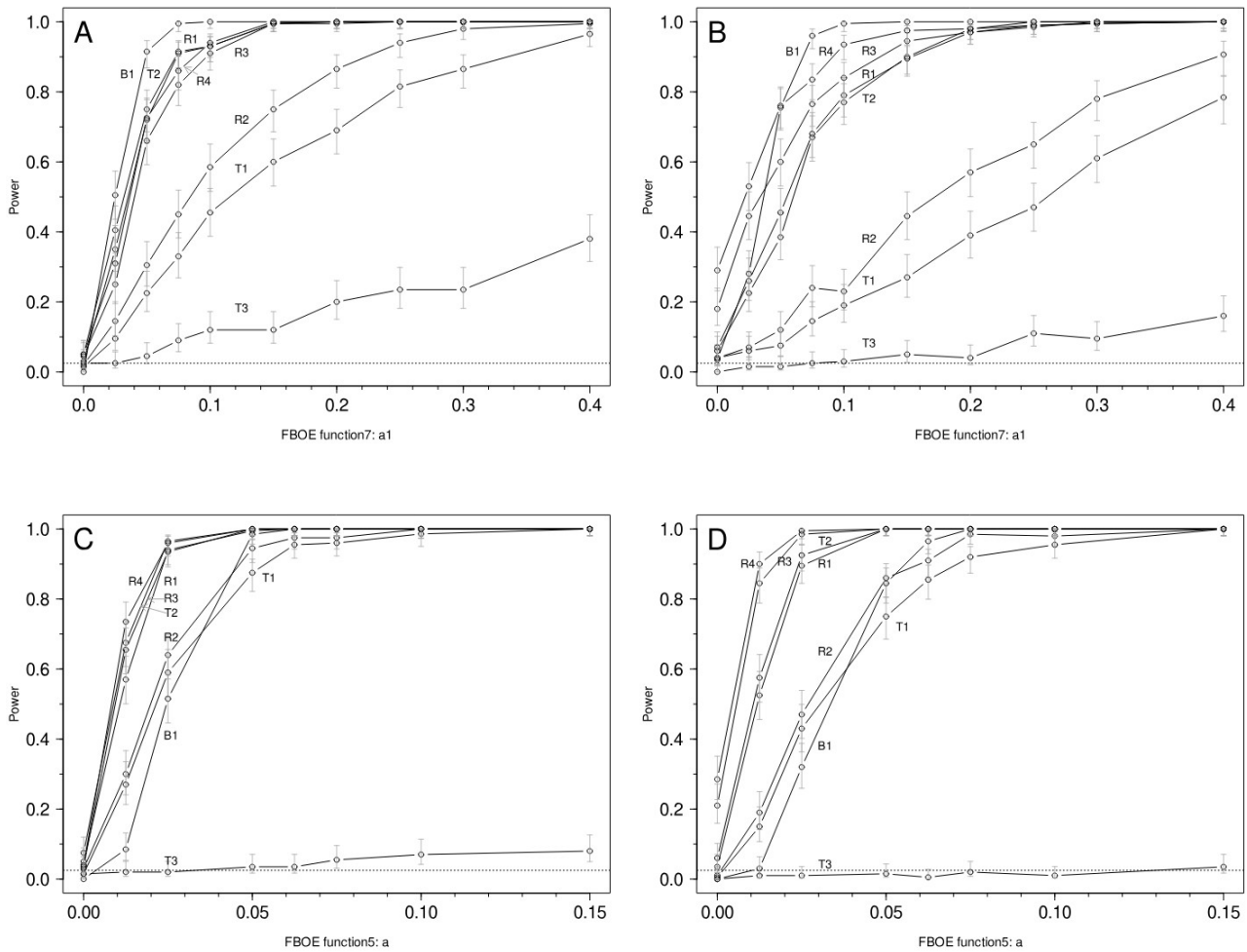


Figure 1. Power comparison for testing FBOE. Data are generated with FBOE of various strength, described by the x-axis (a_1 from function f7 for panels A and B, or a from function f5 for panels C and D). AE is either absent (panels A and C), or present with $\alpha = \beta = 1$ (panels B and D). The name of each test is indicated for each curve. The bar-interval indicates the 95% confidence interval. The dotted horizontal line indicates the false rejection rate of 2.5%.

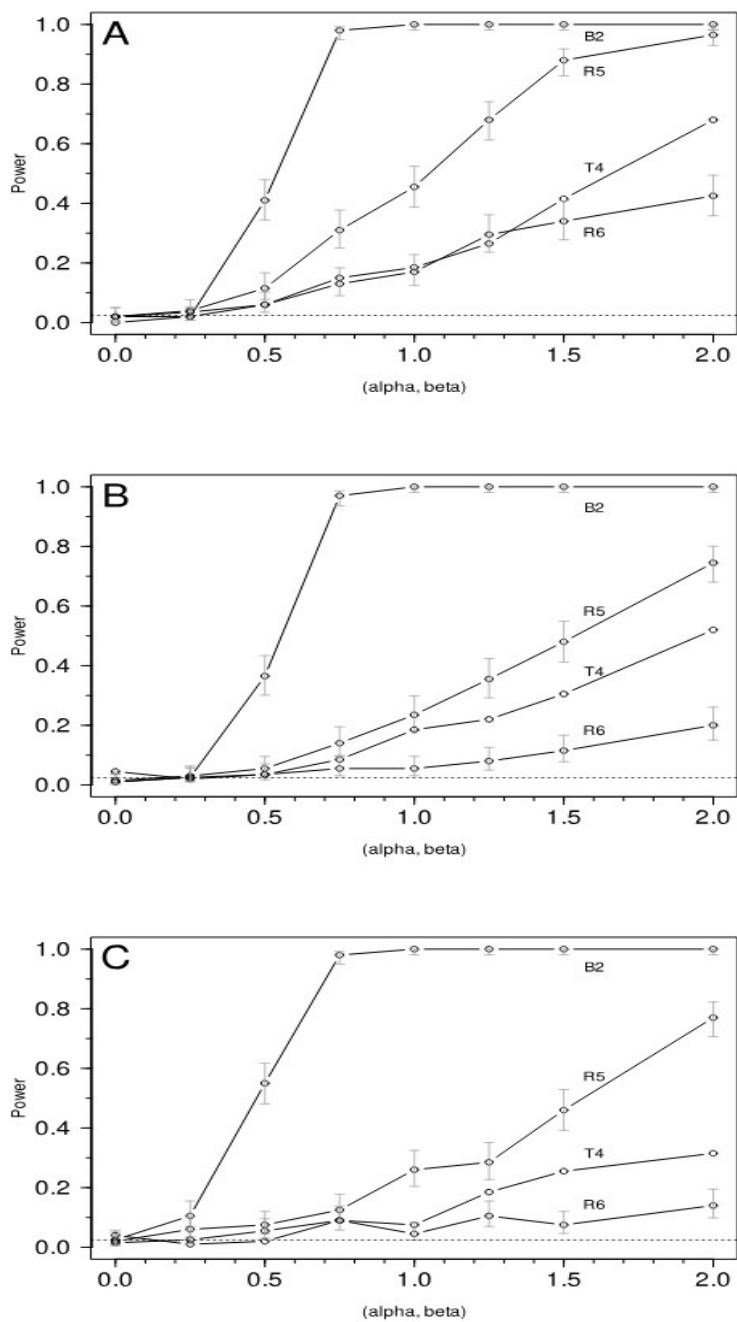


Figure 2. Power comparison for testing AE. Data are generated with AE of various strength, described by the x-axis (with $\alpha = \beta$). FBOE is either absent (A), or present with function f7, $a_1 = 0.1$ (B), or with function f5, $a_1 = 0.1$ (C). The name of each test is indicated for each

curve. The arrows indicates the 95% confidence interval. The dotted horizontal line indicates the false rejection rate (2.5%).

Appendix 1

Rate of type-I error of the Fisher test on 2x2 contingency table draw from a quasi-binomial sampling.

Simulations were used to compute the rate of type-I error of Fisher's test on 2x2 contingency table, when cells were drawn from a quasi-binomial distribution (QBD). A quasi-binomial distribution of type II (QBD II) was used (Consul & Mittal, 1975; Mishra et al., 1992), with parameter n , p , and α , with n the number of experiments, p the probability of success, and $0 \leq \alpha < (1-p)/n$. When $\alpha = 0$, QBD II becomes a classical binomial, $B(n, p)$. When $0 < \alpha < (1-p)/n$, the variance of QBD II is higher than the binomial variance (Mishra et al., 1992).

A 2x2 contingency table was generated under the null hypothesis (independence of rows and columns). In an infinite bag of A and B objects, with frequency p and $1-p$ respectively, n objects were drawn, resulting in a_1 objects A, and b_1 objects B, with $a_1 + b_1 = n$. In the same bag, another set of n objects were independently drawn, resulting in a_2 objects A, and b_2 objects B. Each draw was performed using QBD II, with parameter $n = 800$, $p = 0.25$, and α . The resulting contingency table was,

| | A | B | Total |
|-------------|-------|-------|-------|
| First draw | a_1 | b_1 | n |
| Second draw | a_2 | b_2 | n |

and the two sided p-value of the Fisher's exact test on this contingency table was calculated. This process was repeated 1000 times, allowing to compute the proportion of p-values lower than 0.05 (i.e., the rate of type-I error associated with this set of parameters). This calculation was repeated for α varying from 0 to 0.0009, by step of 0.0001.

When $\alpha = 0$, the rate of type-I error was not different than the rejection rate (Fig. S1), as expected when cells of the contingency table display a binomial distribution. When $\alpha > 0$, the rate of type-I error was significantly higher than the rejection rate, increasing with increasing values of α (Fig. S1).

References

- Consul, P. C., & Mittal, S. P. (1975) A new urn model with predetermined strategy. *Biometrische Zeitschrift*, 17, 67–75. <https://doi.org/10.1002/bimj.19750170202>
- Mishra, A., Tiwary, D., & Singh, S. K. (1992) A class of quasi-binomial distributions. *The Indian Journal of Statistics, series B*, 54, 67–76.

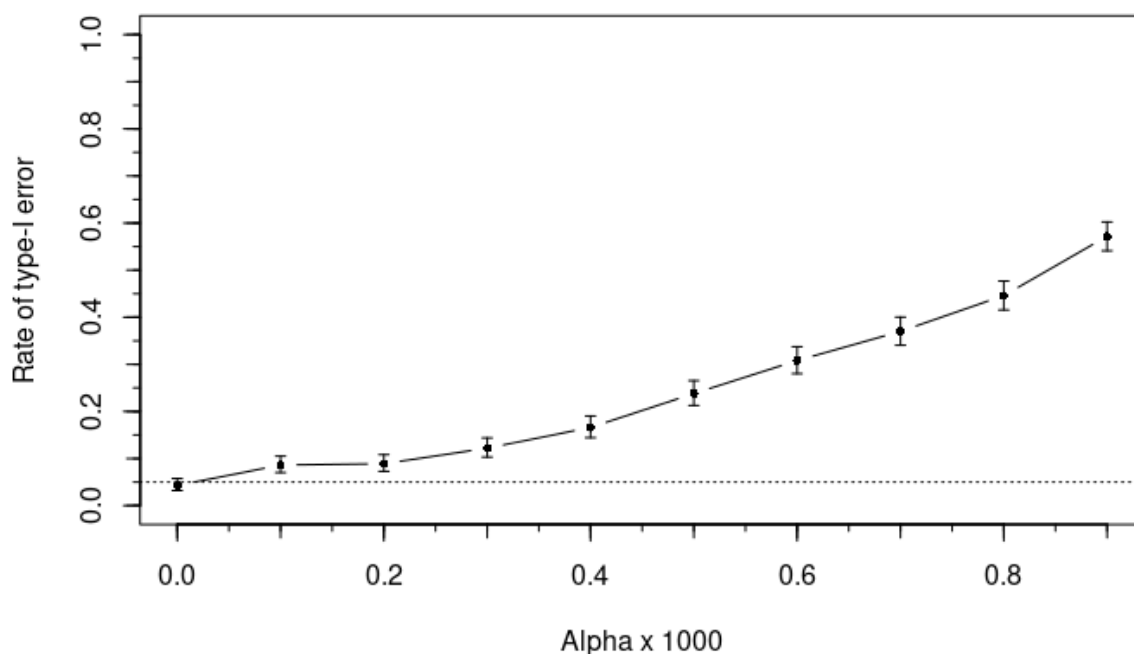


Figure S1. Rate of type-I error for a 2x2 contingency table with cells drawn from a quasi-binomial distribution, for various values of α . For each α value, the rate of type-I error of the test is indicated, based on 1000 replicates. Bars indicate 95% confidence intervals. The dotted line indicates the rejection rate (0.05). $\alpha = 0$ corresponds to a binomial distribution.

Appendix 2

Rate of type-I error of the Fisher test on 2x2 contingency table for binomial, discrete uniform, and mixed sampling.

Simulations were used to compute the rate of type-I error of Fisher's test on 2x2 contingency table. Two independent samples are compared. First, each sample corresponded to one draw from a large family of $n = 200$ sibs. Older sibs (*older*) were drawn from a uniform distribution $U[1, 200]$, and older brothers (*ob*) were drawn from a binomial $B(\text{older}, \frac{1}{2})$. Younger sibs (*younger*) and older sisters (*os*) were derived as $\text{younger} = n - \text{older}$, and $\text{os} = \text{older} - \text{ob}$. Three tests on 2x2 contingency tables were performed. For the OR test (or test T1 of Table 1), *ob* and *os* were compared across two independent samples, and were thus derived from a binomial sampling. For the test comparing *older* and *younger* across two independent samples (OY test), cell counts were derived from a uniform distribution. For the OBOR test, *ob* and $\text{younger} + \text{os}$ were compared across two independent samples, and were derived from a mixed distribution: *ob* and *os* are from a binomial distribution $B(\text{older}, 1/2)$, an *younger* is from a uniform distribution (Fig. S2 A). Second, each sample corresponded to the sum 200 draws from a families of 4 sibs. The same three tests (OR, OY, and OBOR) were performed. For the OR test, the sum of *ob* and *os* were compared across two independent samples, and were thus derived from a binomial sampling, as the sum of binomials (with the same p parameter) is also a binomial: $B(n_1, p) + B(n_2, p) = B(n_1+n_2, p)$. For the OY test, the sum of *older* and *younger* follows a complex distribution, as the sum of independent discrete uniform distribution is not uniform, with no simple formal description (except for the sum of just 2 or 3 uniform distributions). For the OBOR test, the sum of *ob* follows a binomial

distribution, and the sum of *younger* + *os* follows a complex distribution, resulting in a distribution not described easily by a simple formula (Fig. S2 B).

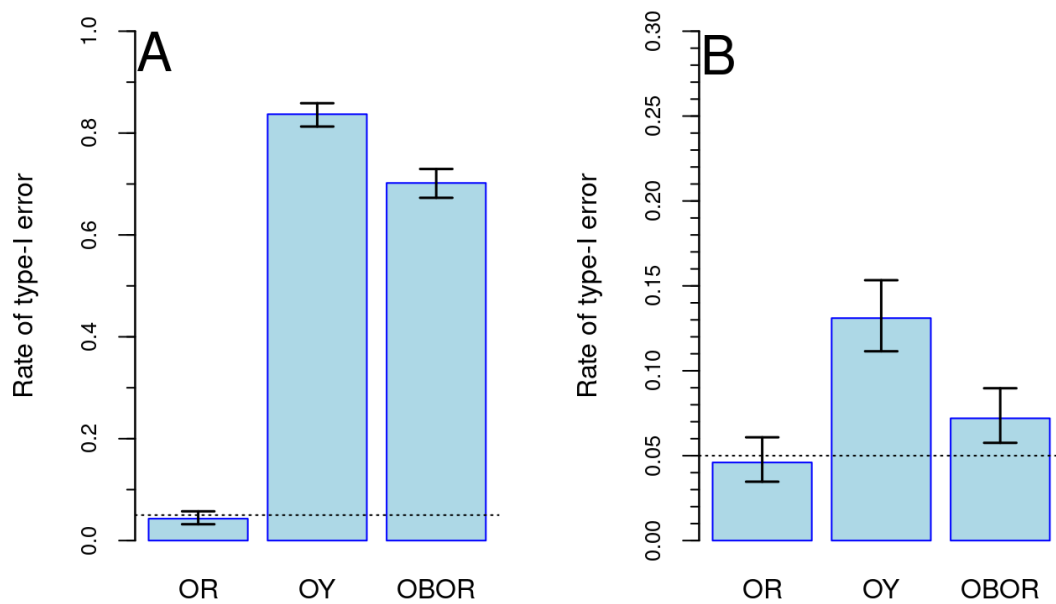


Figure S2. Rate of type-I error for a 2x2 contingency table. Two independent samples are compared. A. Each line of the contingency table corresponds to one draw from a large family (200 sibs). B. Each line of the contingency table corresponds to sum of 200 draws from families of 4 sibs. The dotted line indicates the rejection rate (0.05).

Appendix 3

Power comparison for testing FBOE, when and SBOE is also present.

Families were generated as described above. FBOE and SBOE were explicitly modeled by two different functions. First, by considering that the probability P to be homosexual is $P = p_0 + a \cdot ob + (a/sb) \cdot os$, where p_0 is the probability to be homosexual for a man without older sibs, ob and os are the number of older brothers and older sisters, respectively, with $sb = 3$, and $a \geq 0$ (corresponding to function f5 of Table S1 in Raymond et al., 2023). Second, by considering that $P = p_0$ for $ob + os = 0$, and $P = p_0 + a_1$ for $ob > 0$ and $os = 0$, $P = p_0 + a_1/sb$ for $os > 0$ and $ob = 0$, and $P = p_0 + a_1 + a_1/sb$ for $ob > 0$ and $os > 0$, with $sb = 3$, and $a_1 \geq 0$ (corresponding to function f7 of Table S1 in Raymond et al., 2023). AE was not considered here (i.e., $\alpha = \beta = 0$). From the whole offspring population, a random sample of 200 heterosexual and 200 homosexual men was drawn as described above, and the various tests for detecting the presence of FBOE were performed (B1, R1-R4, T1-T3). This was replicated at least 200 times for a given set of parameters (a or a_1), allowing the computation of the power of each test for each value of a or a_1 (Fig. S3).

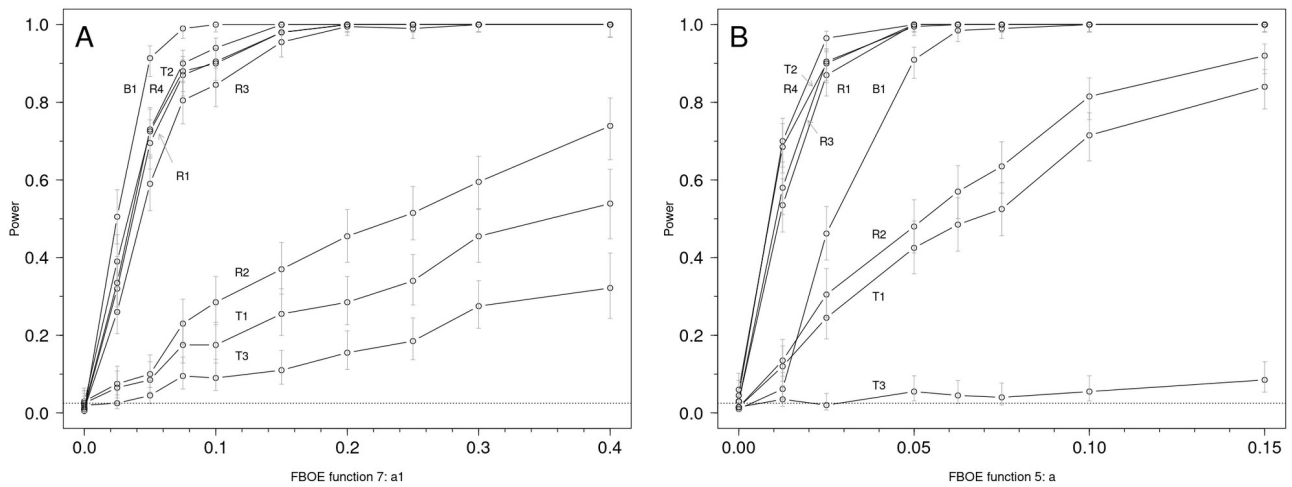


Figure S3. Power comparison for testing FBOE, when SBOE is also present. Data are generated with FBOE of various strength, described by the x-axis (a_1 from function f7 for panel A, or a from function f5 for panel B). SBOE is three times weaker than FBOE, for each a or a_1 value. The name of each test is indicated for each curve. The bar-interval indicates the 95% confidence interval. The dotted horizontal line indicates the false rejection rate of 2.5%.

Appendix 4

Type-I error comparison for testing FBOE, when SBOE is present.

Families were generated as described above. SBOE were explicitly modeled by two different functions. First, by considering that the probability P to be homosexual is $P = p_0 + a \cdot os$, where p_0 is the probability to be homosexual for a man without older sibs, os is the number of older sisters, and $a \geq 0$ (corresponding to function f5 of Table S1 in Raymond et al., 2023). Second, by considering that $P = p_0$ for $os = 0$, and $P = p_0 + a_1$ for $os > 0$, and $a_1 \geq 0$ (corresponding to function f7 of Table S1 in Raymond et al., 2023). AE was not considered here (i.e., $\alpha = \beta = 0$). From the whole offspring population, a random sample of 200 heterosexual and 200 homosexual men was drawn as described above, and the various tests for detecting the presence of FBOE were performed (B1, R1-R4, T1-T3). This was replicated at least 200 times for a given set of parameters (a or a_1), allowing the computation of the type-I error rate of each test for each value of a or a_1 (Fig. S4).

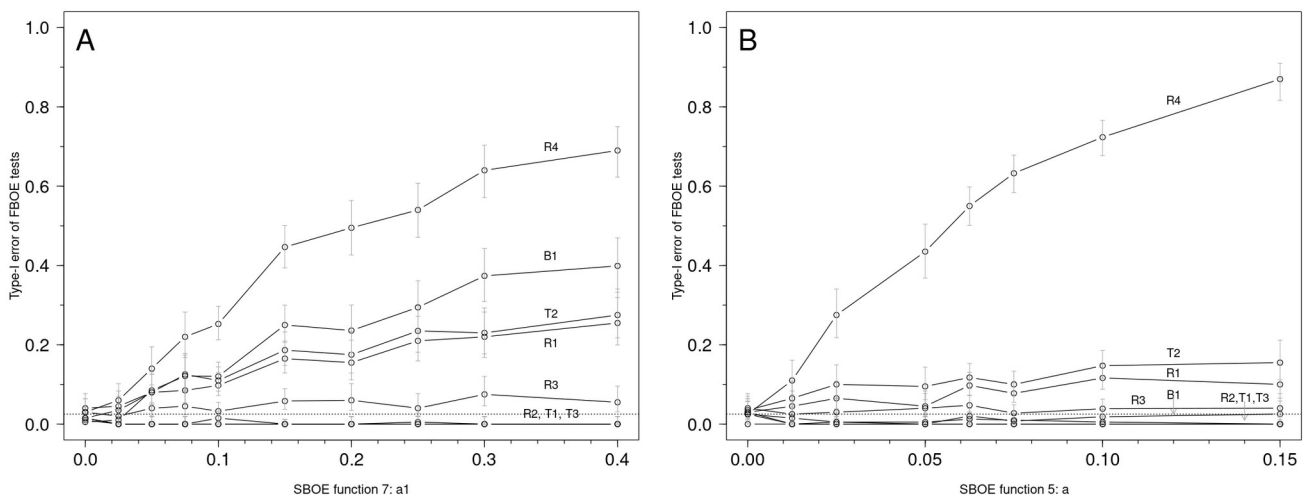


Figure S4. Type-I error rate for testing FBOE, when SBOE is present. Data are generated with SBOE of various strength, described by the x-axis (a_1 from function f7 for panel A, or a from function f5 for panel B). No FBOE was generated. The name of each test is indicated for each curve. The bar-interval indicates the 95% confidence interval. The dotted horizontal line indicates the false rejection rate of 2.5%.