

Estimation of Trait-Model Parameters in a MOD Score Linkage Analysis

Markus Brugger · Susanne Rospleszcz · Konstantin Strauch

Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, and Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany

Keywords

Parametric linkage analysis · MOD scores · Trait-model parameters · Identifiability · Estimation bias

Abstract

Background/Aims: Theoretically, the trait-model parameters (disease allele frequency and penetrance function) can be estimated without bias in a MOD score linkage analysis. We aimed to practically evaluate the MOD score approach regarding its ability to provide unbiased trait-model parameters for various pedigree-type and trait-model scenarios. We further investigated the ability of the MOD score approach to detect imprinting using affected sib pairs (ASPs) and affected half-sib pairs (AHSPs) when all parental genotypes are missing. **Methods:** Simulated pedigree data were analyzed using the GENEHUNTER-MODSCORE software package. Parameter estimation performance in terms of bias and variability was evaluated with regard to trait-model type and pedigree complexity. **Results:** Generally, parameters were estimated with lower bias and variability with increasing pedigree complexity, especially for recessive and over-dominant models. However, dominant and additive models could hardly be distinguished even when using 3-genera-

tion pedigrees. Imprinting could clearly be detected for mixtures of mainly ASPs and only few AHSPs with the common parent of the imprinted sex, even though no parental genotypes were available. **Conclusion:** Our results provide guidance to researchers regarding the possibility to estimate trait-model parameters by a MOD score analysis, including the degree of imprinting, with certain types of pedigrees.

© 2017 The Author(s)
Published by S. Karger AG, Basel

Introduction

Trait Inheritance and Pedigree Analysis

The inheritance of a trait is defined as the mechanism by which the joint phenotypic distribution of the particular trait in pedigree members can explicitly be described [1]. A pedigree can be considered as a discrete unit of a population for which the relationship connecting any pair of pedigree members is unambiguously known. There is hence no other individual for which a relationship to any of these pedigree members can be established. Under the assumption that pedigrees implicitly contain information about details of the mode of inheritance of a trait through the covariation and cosegregation of the

trait characteristics among its members, collecting and analyzing samples of pedigrees can be used to study the trait inheritance. In genetics, inference about trait inheritance by pedigree analysis is made assuming that the main factors underlying the inheritance are genes. Mathematical-genetic models can then be used to describe the trait inheritance, and these models are tested using pedigree samples drawn from the population.

If the genetic model of trait inheritance is inferred on the basis of the pedigree sample, which contains the necessary information through the joint phenotypic cosegregation in the pedigree members, such an analysis is called “segregation analysis” [1]. If the purpose of the analysis is to map the putative disease gene(s), whose existence may have been previously established by segregation analysis to specific chromosomal segments by investigation of cosegregation of DNA marker alleles and the trait phenotype, such an analysis is called “linkage analysis” [1]. Nowadays, pure segregation analysis is of less practical importance than it has been a few decades ago. With increasing availability of DNA marker maps and rapid and cost-effective DNA genotyping techniques, linkage analysis has become the state-of-the-art technique of pedigree analysis. In addition, association analysis can be performed with pedigrees as well as samples of unrelated individuals. However, software packages for segregation analysis like PAP [2], S.A.G.E. [3], and MORGAN [4, 5] continue to be available and provide great flexibility with respect to fitting the model for the mode of inheritance (see also e.g. Kriszt et al. [6] for a recent publication using complex segregation analysis with keratoconus pedigrees).

Linkage Analysis

In earlier times, linkage analysis has been used to map genes that were already known to exist. In the meantime, linkage analysis serves 2 purposes: (1) to prove the existence of a disease gene and (2) to map it [7]. Linkage analysis methods can be distinguished as model-based or model-free [8]. The former is also known as parametric or LOD score linkage analysis for which a certain set of trait-model parameters regarding the segregation of the disease is explicitly assumed in the genetic likelihood. The latter, which is also known as nonparametric linkage analysis, proceeds without such explicit models. These 2 types of linkage analysis are, however, closely related to each other. It can be shown that certain nonparametric and parametric linkage tests are equivalent for any type of pedigree [9, 10] and can be considered as different ways to parametrize the allele-sharing probabilities, i.e., the probabilities of allele(s) shared identical-by-descent

(IBD) by affected pedigree members, in the genetic likelihood.

Mode of Inheritance and Trait-Model Parameters

A crucial factor in linkage analysis is the true mode of inheritance. Under the term “mode of inheritance,” 2 concepts are often subsumed that need, however, to be distinguished. The first concept is the genetic mechanism of the disease involving the number of loci, the number of alleles at each locus, and the segregation parameters including the recombination fraction among the trait loci as well as between them and any marker(s) [11]. The second concept is the genotype-phenotype relation, which is defined by the penetrance function, i.e., the probability that an individual with a certain number of copies of the disease allele is affected by the disease. The genetic mechanism of the disease, apart from the recombination fraction, is assumed to be known for linkage analysis. In the case of a binary trait governed by a single diallelic autosomal locus, which is assumed throughout this paper, the disease allele frequency p and the 3 penetrances f_0, f_1 , and f_2 , with f_i denoting the probability that an individual with i copies of the disease allele is affected by the disease, can be subsumed under the term “trait-model parameters.” In the case of parametric linkage analysis, trait-model parameters can either be prespecified according to results from previous segregation analyses or maximized along with the recombination fraction in a joint segregation and linkage (JSL) analysis. A specific type of this approach is the MOD score analysis, which was first proposed by Risch [12]. If the genetic mechanism of the trait is not modelled correctly, however, which is expected in practice due to the large number of possible inheritance modes, parameter estimates obtained from a MOD score analysis will be asymptotically biased [11, 13].

Likelihood and Sample Space

In pedigree analysis, the likelihood given a particular sample of pedigrees can be defined as the probability to observe the data available for the individuals in the pedigree, constructed under a certain genetic model. In fact, any formulation that is proportional to this probability can be used as the likelihood. The pedigree samples used for pedigree analysis are collected from what is called the “real” population that is defined on the basis of usually unknown factors like the population’s origin and history. This real population is mapped into a set of disjoint pedigrees by the use of those relationships between members of the real population that can unambiguously be established [1]. These disjoint pedigrees are then further deter-

mined by the predefined sampling design, which partitions the pedigrees into substructures of certain inheritance relations, e.g., sibships with all other relationships outside sibships being ignored. The resulting structures are called “true pedigrees.” As described in Ginsburg et al. [1], pedigree analysis is performed on sampled pedigrees collected from the set of true pedigrees. The subset of pedigrees that in principle can be sampled according to the sampling design is called the “sample space.” The sampling procedure involves the pedigree ascertainment (primary selection), the intrafamilial extension (inclusion of additional relatives), and the selective inclusion in the analysis (censoring).

In the following, we will assume that ascertainment takes place through probands. For each true pedigree, there are members who could “potentially” become probands due to prespecified proband characteristics, e.g., geographic area, age, sex, but independently of their phenotypes. This subset of potential probands in the true pedigree, including both their relationships and phenotypes, is called the “proband sampling frame” (PSF, [14]). It can be shown that assuming the wrong mode of inheritance and/or the wrong model for the sampling procedure leads to asymptotically biased trait-model parameters and nuisance parameters of the sampling model when performing maximum likelihood estimation [15]. In order to obtain unbiased parameter estimates, the pedigree likelihood is defined as the probability of the particular pedigree data having been sampled (ascertained, extended, and included in the analysis) on the sample space generated by the sampling procedure under the given mode of inheritance [1]. The sample space for the given sampling procedure is the probability that at least 1 pedigree is sampled from the set of true pedigrees [1]. In this general form, however, the pedigree likelihood cannot be calculated using only the sampled data [1]. This would demand knowledge about the distribution of possible PSFs to calculate the sample space on which the likelihood is defined. Therefore, pedigree likelihoods are conditioned on specific parts of the sampled data to circumvent this problem and – by the same token – to retain unbiasedness of parameter estimates. In the following sections, pedigree likelihoods, which are conditioned on specific parts of the sampled data, are briefly introduced in the context of JSL analysis.

Sampling Model-Based Likelihood

As was explained in the previous section, the pedigree likelihood provides consistent estimates of the trait-model parameters if it is conditioned on the pedigree having

been sampled, i.e., ascertained, extended, and included in the sample under analysis [16]. This also holds true for JSL analysis. In parametric JSL analysis, which is the main focus of this paper, the likelihood is formulated using the trait-model parameters, i.e., the disease allele frequency p and the penetrances f_0, f_1 , and f_2 , as well as the marker allele frequencies and the recombination fractions – and, if applicable, linkage disequilibria (LD) between loci. These parameters can be subsumed under the term “joint trait-marker inheritance parameters” [16]. In addition, information about the following aspects must also be included in the likelihood: (1) the whole PSF structure and its population distribution, which is relevant for ascertainment, (2) the pedigree extension procedure, and (3) the conditions relevant to inclusion, which could be specific marker genotypes of certain pedigree members. Since the population distribution of the PSF structure is unknown, the pedigree likelihood can be conditioned on the substructure of the pedigree that is “relevant to sampling” (RS), in order to make the likelihood calculable and to properly take the sampling procedure into account. The structure RS corresponds to all PSF members of the true pedigree under study – i.e., the part of the pedigree “relevant to ascertainment” (RA) – and those pedigree members responsible for the inclusion of the pedigree in the sample. Importantly, the likelihood is only conditioned on the structure RS but not on the phenotypes of the corresponding pedigree members. Since the likelihood includes explicit details of the sampling procedure, it is termed “sampling model-based (SMB) likelihood” [16]. The SMB likelihood provides asymptotically unbiased estimates of all joint trait-marker inheritance parameters, including the mode of inheritance, as well as of the parameters determining the ascertainment, extension, and inclusion procedure [1].

Sampling Model-Free Likelihood

A sampling model-free (SMF) likelihood can be formulated using a more robust procedure initially proposed by Ewens and Shute [17] in the context of segregation analysis, in which uncertainties about the ascertainment procedure are controlled by conditioning the likelihood on that part of the pedigree data RA. The latter approach is called “ascertainment assumption-free” (AAF) and can readily be extended to be SMF, if the likelihood is also conditioned on that part of the data RS [16]. The part of the data RS is the data RA and that part of the data relevant to inclusion, which could be, e.g., certain parental marker genotypes. In contrast to the SMB likelihood, which is conditioned only on the structure RS, the SMF

likelihood is conditioned on the data RS, i.e., structure as well as marker and trait values RS. This SMF likelihood provides asymptotically unbiased estimates of all joint trait-marker inheritance parameters, including the mode of inheritance, as well as of the extension parameter [1].

Likelihood in a MOD Score Analysis

The question arises which kind of likelihood underlies a JSL analysis using the MOD score, and if it is in principle possible to obtain unbiased parameters from this procedure. As shown by Clerget-Darpoux et al. [18] and later also by Elston [11], maximizing the LOD score in the context of a MOD score analysis is equivalent to maximizing the likelihood of the marker data, conditional on the pedigree structure and conditional on all the trait data, i.e., not only on that part RS. This conditional likelihood – from now on referred to as “MOD score likelihood” – does not depend on the ascertainment scheme, provided that the sampling of pedigrees is independent of marker data. Hence, this means that selective inclusion of pedigrees based on marker genotypes (i.e., marker-dependent sampling) is not controlled in the MOD score likelihood, because it does not contain the inclusion parameter. As a consequence, the MOD score will yield biased estimates of the joint trait-marker inheritance parameters if there is association between disease and marker alleles ($LD > 0$), because ascertainment is no longer marker-independent in that case [19].

The following conditions must be satisfied to obtain unbiased estimates of the joint trait-marker inheritance parameters from a MOD score analysis [1, 19]: (i) the marker locus must be truly linked to the trait locus, (ii) the genetic mechanism of the trait (number of loci and number of alleles at each locus) is known, (iii) sampling is marker-independent, (iv) the model for the pedigree extension procedure is known, and either (v) trait values are available for all members of the PSF, which has to be completely known, or (vi) the ascertainment is proband-independent (PI) or single in the sense described by Hodge and Vieland [20], i.e., all pedigrees have equal probabilities of being ascertained, independent of pedigree size or structure, or (vii) the joint probability of the unobserved trait phenotypes of the members of the PSF, conditional on the trait and marker phenotypes of all the observed pedigree members, does not depend on the marker phenotypes. Condition (v) reflects that the MOD score likelihood can be derived from the SMB likelihood by conditioning the latter on the trait values of all individuals, including all PSF members, in addition to the structure RS. Condition (vi) is due to the fact that the

MOD score likelihood does not include an ascertainment parameter as opposed to the SMB likelihood, which contains such a parameter. The probability of ascertainment, however, actually depends on the joint trait-marker inheritance parameters, if sampling is not PI or single [21]. Only with PI or single ascertainment, the probability of ascertainment no longer depends on these parameters and can, therefore, be omitted in the likelihood without influencing the estimates of the parameters [1]. Without specifying details of the sampling procedure, parameter estimates are also consistent when missing trait values of the PSF members do not depend on marker phenotypes (condition [vii]). However, this only holds in the case of no LD and no linkage between trait and marker locus, or if the trait phenotype unambiguously defines the trait genotype [19].

The MOD score likelihood differs from the SMF likelihood by the fact that it is conditioned on all trait values (i.e., not only of the PSF members) in addition to the data RS, and that it assumes PI or single ascertainment as well as marker-independent sampling, rather than specifying some value for the ascertainment probability in the likelihood. This is why the MOD score likelihood can be considered to be somewhere between SMB and SMF. If sampling is marker-independent, but conditions (i), (ii), and (iv) are not simultaneously satisfied, parameter estimates obtained from MOD score analyses will be biased. If conditions (i)–(iv) hold, but neither condition (v), (vi), nor (vii) is met, the estimate of the recombination frequency will only slightly be biased [1]. In this case, it is of note that estimates of the recombination fraction are biased even when trait-model parameters are fixed at their true values [22].

Summary of Conditions to Obtain Unbiased Parameter Estimates from a MOD Score Analysis

The pedigree likelihood of the MOD score approach delivers asymptotically unbiased estimates of the joint trait-marker inheritance parameters (recombination fraction, allele frequency, and penetrances, but not the LD parameter), if the following conditions are satisfied (see also Malkin and Elston [19]):

- i The marker is truly linked.
AND
- ii The genetic mechanism of the trait (number of loci and number of alleles at each locus) is known.
AND
- iii Sampling (ascertainment, extension, inclusion) of pedigrees is independent from marker data.
AND

- iv The model of extension is known.
AND
- At least 1 of the following 3 conditions is satisfied:
- v All members of the pedigree PSF must have measured trait values (if not sampled, information on trait values can be gathered using a questionnaire as proposed by Ginsburg et al. [16]).
OR
- vi The ascertainment procedure is PI or single in the sense of Hodge and Vieland [20].
OR
- vii The joint probability of the unobserved trait phenotypes of the members of the PSF, conditional on the trait and marker phenotypes of all the observed pedigree members, does not depend on the marker phenotypes.

Hence, unbiased estimates of the joint trait-marker inheritance parameters can in principle be obtained without explicitly formulating the ascertainment and inclusion procedures. It should further be noted that the likelihood correction in a MOD score analysis directly follows from the AAF method proposed in Ewens and Shute [17]. Whereas conditions (i)–(iii) are crucial, conditions (v)–(vii) may be of minor impact on the bias of parameter estimates in practice [20, 23]. With respect to condition (v), if members of the pedigree PSF are not sampled and trait values cannot be gathered using a questionnaire, an approximate likelihood using the sample mean of the trait value can be constructed [1]. Condition (iv) could be satisfied as follows. PI sampling implies that fixed pedigree structures are sampled, which renders a specification of the extension parameter pointless. With single ascertainment, the pedigree extension model could be chosen to be trait-independent, such that any initially sampled subpedigree is further extended using all available relatives, regardless of their phenotypes and with a random, trait-independent stopping rule. If this holds true, an extension parameter does not have to be formulated in the likelihood. Despite being hard to achieve in practice, conditions (iv)–(vii) can in theory be fulfilled. If not, the resulting bias in parameter estimates is argued to be small [20], but numerical quantification of the bias of the joint trait-marker inheritance parameters obtained from a MOD score analysis under many different sampling schemes is not available so far. This would demand an extensive simulation study to prove that the MOD score approach is robust with regard to its ability to estimate parameters, even if some necessary assumptions do not hold. Even if all necessary conditions are satisfied, a bias of maximum likelihood estimates can nevertheless occur for finite sample sizes. In addition,

variances of the obtained estimates are expected to be rather large using the MOD score likelihood due to a loss of pedigree information by conditioning not only on the pedigree structure but also on the trait data of all individuals [24].

The focus of the present paper is the proof-of-principle of the ability of a MOD score analysis to obtain asymptotically unbiased joint trait-marker inheritance parameters in practice, given that conditions (i)–(iv) and at least one of (v)–(vii) are satisfied. In particular, the identifiability (see also next section) of these parameters using various pedigree types and realistic sample sizes will be investigated.

Identifiability of Inheritance Parameters

Even if the conditions under which the MOD score provides unbiased estimates of the joint trait-marker inheritance parameters are fulfilled, the identifiability of these parameters is restricted by the type(s) of pedigrees in a given sample. In a model-based linkage analysis, such as a MOD score analysis, the penetrances, disease allele frequency, and the recombination fraction represent a reparametrization of the truly underlying allele-sharing classes [9, 10, 25, 26]. In other words, allele-sharing probabilities (classes) of a given pedigree type can be expressed in terms of the joint trait-marker inheritance parameters. In the case of an affected sib pair (ASP), these allele-sharing classes are z_0 , z_1 , and z_2 that an ASP shares 0, 1, or 2 allele(s) IBD with restrictions to genetically possible models [27]. With $z_2 = 1 - z_0 - z_1$ and restrictions $z_1 \leq 0.5$ and $2 \times z_0 \leq z_1$, the allele-sharing classes of ASPs form a 2-dimensional parameter space – the so-called “possible triangle” [27]. Hence, as there are only $3 - 1 = 2$ free parameters that can be estimated from ASP data, there will be many sets of f_0, f_1, f_2, p , and the recombination fraction θ that correspond to the estimated \hat{z}_0, \hat{z}_1 , and \hat{z}_2 . With larger pedigrees, and hence more allele-sharing classes, the degree to which the trait-model parameters can be correctly determined should be higher. However, the corresponding allele-sharing configurations have hitherto only been formulated for unilineal, affected relative pairs (e.g., affected half-sib pairs [AHSPs] [10]), ASPs [27], and affected sib triplets (ASTs) [28]. The parameter space for AHSPs is degenerated to a single line [10]. Hence, many different sets of trait-model parameters correspond to the same point on this so-called “possible line.”

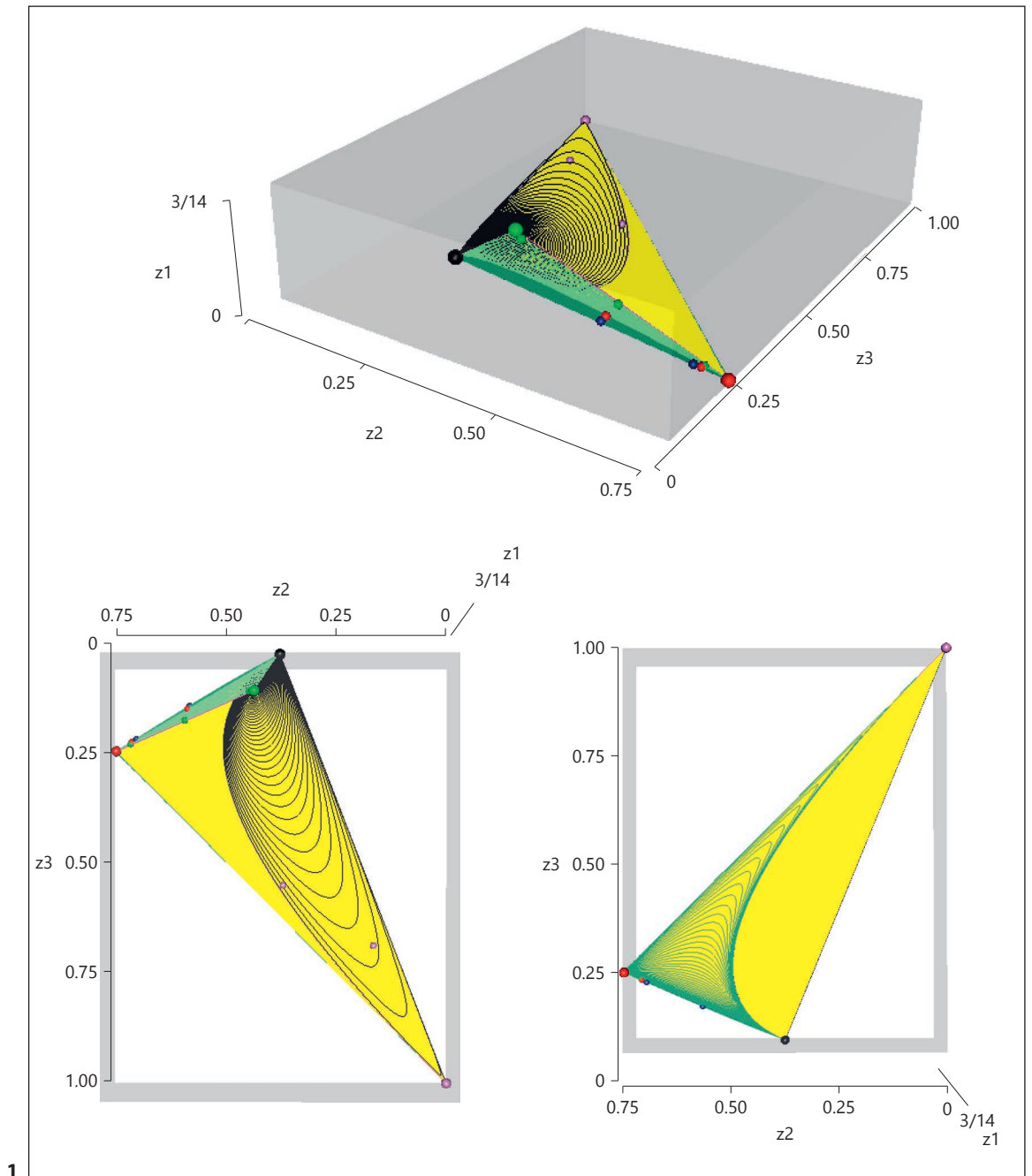
Using the formulas in Knapp [28], it is possible to draw the 3-dimensional parameter space for ASTs with empirically assessed restrictions for genetically possible models (Fig. 1). However, the parameter restrictions have not been derived in closed form so far. The parameter spaces for

larger pedigrees involve a larger number of dimensions, and the corresponding restrictions for genetically possible models are expected to have an even more complicated form [10, 28]. It is of note that for any type of affecteds-only analysis, the absolute values of penetrances cannot be determined, because multiplication of all penetrances by the same factor does not change the result. However, their ratios are not defined if the penetrance in the denominator

of the ratio is estimated to be 0. Additionally, the ratio is subject to the estimation variance of both the penetrance in the numerator and in the denominator.

Imprinting

Genomic imprinting implies dependence of an individual's liability to develop a disease on the parental origin of the mutated allele(s), leads to a deviation from the



(For legend see next page.)

classic Mendelian assumption of equal contribution of parental genomes to the progeny and is, therefore, called a “parent-of-origin effect” [29]. In the context of a parametric linkage analysis, imprinting can be modelled using a 4-penetrance formulation distinguishing the heterozygotes according to the parental origin of the disease allele: $f = (f_0, f_{1,pat}, f_{1,mat}, f_2)$, as implemented in the program GENEHUNTER-MODSCORE (GHM) [25, 30–32], which is a further development of GENEHUNTER-IMPRINTING [33]. In the nonparametric context, the allele-sharing class z_1 of an ASP is split up into $z_{1,pat}$ and $z_{1,mat}$ according to the parental origin of the shared allele. The corresponding parameter space of ASPs, hence, extends to a 3-dimensional tetrahedron which accounts for disease models with $z_{1,pat} \neq z_{1,mat}$, i.e., for imprinting [34]. In the case of AHSPs, the allele-sharing class z_1 is distinguished as either being $z_{1,pat}$ or $z_{1,mat}$ depending on the sex of the common parent, i.e., male or female, respectively. Although the information contained in AHSPs on all trait-model parameters is limited, the information for imprinting may be high, such that parameter estimates for $f_{1,pat}$ and $f_{1,mat}$ using a sample of AHSPs having a common father and of AHSPs having a common mother should indicate imprinting if it was really present. In the case of an informative marker, this even holds if parental genotypes are missing.

In contrast, imprinting information contained in ASPs with untyped parents is 0, even in the case of a fully informative marker, because alleles shared IBD through the father cannot be distinguished from those shared IBD through the mother. However, we hypothesize that the information on linkage and imprinting gained from AHSPs can be combined with the pure linkage information contained in ASPs in the analysis to compensate for missing parental marker genotypes. If there is sufficient evidence for linkage, this pedigree scenario should lead to trait-model parameter estimates reflecting at least some degree of imprinting. Using GHM, imprinting can be quantified by looking at the imprinting index I [35], calculated from the estimated penetrances. The imprinting index equals the difference between the 2 heterozygote penetrances, normalized by the difference of the homozygote penetrances in order to properly take the case of a non-0 phenocopy rate or reduced penetrance into account:

$$I = \frac{f_{1,pat} - f_{1,mat}}{f_2 - f_0}.$$

An imprinting index of $I = 1$, therefore, indicates complete maternal imprinting (cmi), whereas $I = -1$ indicates complete paternal imprinting (cpi). If penetrances are not restricted to $f_0 < f_1 < f_2$ in the analysis, the penetrances $f_{1,pat}$

Fig. 1. Graphical depiction of the allele-sharing parameter space for affected sib triplets (ASTs). The axes notations are defined as follows (see also Knapp [28]). Axis z_1 : allele-sharing class z_1 with range $\{0; 3/14\}$. Axis z_2 : z_2 with range $\{0; 0.75\}$. Axis z_3 : z_3 with range $\{0; 1\}$. The panels top and at the left correspond to “top view.” The boundary of the parameter space, which is defined by the genetically possible models, was empirically determined by varying the trait-model parameters $\{f_0, f_1, f_2, p\}$ in the formulas given in

Knapp [28]. p , disease allele frequency; f_i , penetrances, with f_i denoting the probability that an individual with i copies of the disease allele is affected by the disease. Light green, dark green, and black lines were drawn by varying p between 0 and 1. For more details, see table below. Figures were drawn using rgl: 3D Visualization Using OpenGL, R package version 0.95.1441 (2016) by Adler, Murdoch, and others.

Boundary region Specific point	Trait model type
Light green “protruding” region	$f_1 > f_0, f_2$ or $f_1 < f_0, f_2; f_0 \neq f_2$
Dark green “bottom” region	$f_2 > f_1 > f_0$ or $f_0 > f_1 > f_2; f_1^2 > f_0 f_2; f_0$ or $f_2 = 0$
Black “top” region	$f_1^2 < f_0 f_2; 0 \leq f_0, f_2 \leq 1; f_1 = 0$
Yellow plane (reached from the “top”)	$f_1^2 < f_0 f_2; f_0 f_2 \rightarrow f_1^2$
Yellow plane (reached from the “bottom”)	$f_1^2 > f_0 f_2; f_0 f_2 \rightarrow f_1^2$
Large black sphere	$H_0(z_1, z_2, z_3) = (0.1875, 0.375, 0.1875)$
Large blue sphere	(0, 0.75, 0.25) genetically strongest additive/dominant model
Large violet sphere	(0, 0, 1) genetically strongest recessive model
Large green sphere	(3/14, 3/7, 1/7) genetically strongest overdominant model
Violet spheres	Recessive models R3, with $p = 0.01$ closer to H_0
Blue spheres	Additive models A3, with $p = 0.1$ closer to H_0
Red spheres	Dominant models D3, with $p = 0.1$ closer to H_0
Green spheres	Overdominant models U3, with $p = 0.01$ near the red sphere, and U4 near the green sphere

and $f_{1,mat}$ can, therefore, be estimated to be $<f_0$ and $>f_2$. Thus, the imprinting index may exceed 1 or fall below -1 . In the case of $f_0 = f_2$, the imprinting index is defined to be 0. In a work by Haghghi and Hodge [36], it was shown that asymptotically unbiased estimates of parent-of-origin effects can be obtained using a likelihood formulation for segregation analysis without including an ascertainment parameter when ascertainment is single. The same should hold true for the method by Strauch et al. [33] applied in this paper in the context of parametric linkage analysis according to the arguments given by Ginsburg et al. [1] and Malkin and Elston [19], provided that the formulation with 4 penetrances correctly reflects the genetic mechanism of genomic imprinting.

Aims of the Present Study

The aim of the present study was to evaluate how accurately penetrances, or penetrance ratios in the case of affecteds-only analyses, and the disease allele frequency of a monogenic, dichotomous trait can be estimated in a MOD score analysis. To this end, we performed a simulation study to determine the bias and variability of trait-model parameter estimation for 6 pedigree types (AHSP, ASP, AST, discordant sib triplets [DST], discordant sib quadruplets [DSQ], and 3-generation [3-G] pedigrees) and 4 types of generic models (recessive, dominant, additive, and overdominant) as well as an imprinting model. A single marker locus linked with $\theta = 0$ to the disease locus was considered. It is of note that we did not consider the estimation of the recombination fraction θ or any LD parameter in our analysis. That is because the primary focus of this paper is on the estimation of trait-model parameters, which do not include the recombination fraction. However, the recombination fraction is confounded with the trait-model parameters, especially for smaller pedigree types, like the ones considered in our study, having only a limited number of allele-sharing classes (see also “Identifiability of Inheritance Parameters” above). In addition, LD parameters cannot be estimated using GHM so far.

We avoided the problem of an additional bias due to a possible misspecification of the sampling model for the likelihood correction. This was done by designing the simulation study in a way that conditions (i)–(iv) and (vi) mentioned above to obtain asymptotically unbiased parameter estimates from a MOD score analysis were satisfied as follows (note that only one of conditions [v]–[vii] needs to be fulfilled):

- i The marker was truly linked ($\theta = 0$).
- ii A diallelic autosomal binary trait locus, which is usually assumed as the mode of inheritance in a MOD

score analysis, was used for the simulation of pedigree data.

- iii Sampling of pedigrees was marker-independent.
- iv Extension of pedigrees was trait-independent.
- v –
- vi Ascertainment was single in the sense of Hodge and Vieland [20].
- vii –

Hence, the questions we aimed to answer in our study were:

1. For each pedigree type, can the MOD score approach differentiate between the trait-model types? That is, are, for example, recessive models recognized as being recessive, irrespective of the accuracy of the individual parameter estimates?
 2. How does the estimation accuracy change from ASP to AST, i.e., when adding an affected sibling?
 3. How does the estimation differ between an analysis using only affecteds vs. both affecteds and unaffecteds?
 4. How does the estimation accuracy change from DST to DSQ, i.e., when adding a second unaffected sibling?
 5. How does the estimation accuracy change when more complex pedigrees are considered?
 6. How well can imprinting be detected and estimated in a sample of AHSPs and in a mixture sample of AHSPs and ASPs when parental genotypes are missing?
- The answers to these questions are summarized in the Results section.

Methods

Nomenclature

Parameters written in capital letters ($P, D, F_0, F_{1,pat}, F_{1,mat}, F_1, F_2, I$) denote theoretical parameters and the parameters that were used for simulation (“true” parameters). Lowercase letters ($p, d, f_0, f_{1,pat}, f_{1,mat}, f_1, f_2, i$) denote the parameters that were estimated from simulated data.

Data Generation

The 5 pedigree types shown in Figure 2 (top and middle row) were chosen for the simulations. We used a sample size of 500 families for each pedigree type to ensure sufficient power to detect linkage while maintaining reasonable computation times. For certain trait-model scenarios, we performed additional analyses with a sample size of 1,000 families to assess the degree by which parameter estimates are biased due to finite sample sizes. Disease and marker locus genotypes were simulated using FastSLINK [37–39]. For each pedigree-type-trait-model scenario, we simulated 1,000 replicates. Affection statuses were assumed to be unknown for all founders. Nonfounders were either affected or unaffected (Fig. 2).

Recessive, additive, dominant, and overdominant trait models were considered in the simulations. An overview of the simulated trait models is given in Table 1. Trait models were named accord-

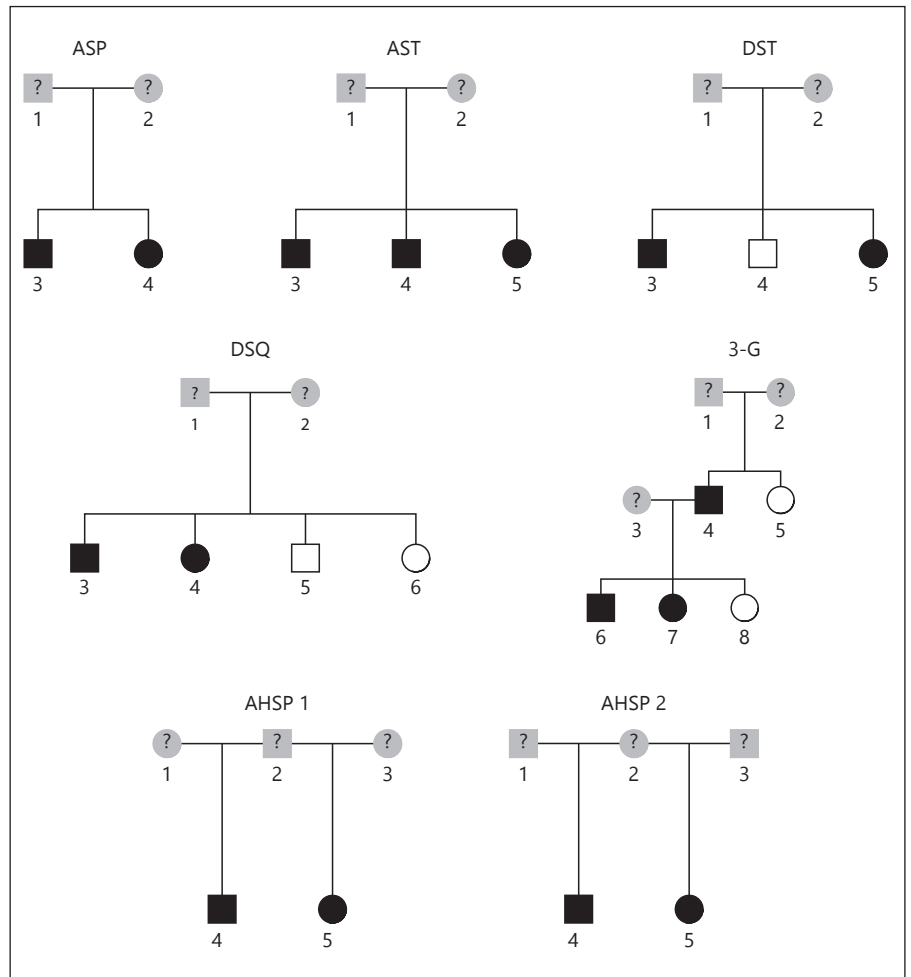


Fig. 2. Pedigree types used for the simulations. ASP, affected sib pair; AST, affected sib triplet; DST, discordant sib triplet; DSQ, discordant sib quadruplet; 3-G, three-generation pedigree; AHSP 1, affected half-sib pair with common father; AHSP 2, affected half-sib pair with common mother; ?, unknown phenotype; filled symbols, affected; empty symbols, unaffected.

ing to their generic type, i.e., “R” for a recessive model, “D” for a dominant model, “A” for an additive model, and “U” for an overdominant model. For each of the 4 generic types, 3 trait models with a particular combination of penetrances were simulated (trait-model names 1–3; Table 1). The setup of the trait-model parameters was inspired by Xing and Elston [40]. Each of the 3 trait models was simulated with a disease allele frequency $P = 0.1$ or 0.01 . For the lower disease allele frequency $P = 0.01$, an additional trait model was simulated with a sample size of 1,000 families per replicate for each of the 4 generic types (Table 1). This amounts to 28 simulated scenarios. Furthermore, an overdominant model with a different combination of penetrances was simulated (model U4). For the recessive, dominant, and additive trait models, 2 further models similar to those in Flaquer and Strauch [41] were considered (models preceded by “AF” in Table 1). One of these models was simulated with sample sizes 500 and 1,000, whereas the other model was simulated with sample size 500 only. The total number of simulated scenarios, therefore, amounts to 38.

We furthermore analyzed AHSP and ASP pedigrees under a model of cpi or cmi. Differing from the scenarios in Table 1, samples contained a mixture of 2 pedigree types. Three scenarios were

considered. In the first scenario, each replicate simulated under the cpi model contained 100 AHSPs who had a common father and 100 AHSPs who had a common mother (Fig. 2, bottom row). In the second scenario, each replicate simulated under the cpi model contained 100 AHSPs who had a common mother and 100 ASPs (Fig. 2, bottom row). In the third scenario, 20 AHSPs who had a common mother and 180 ASPs were simulated under the cmi model. Again, 1,000 replicates were simulated for each scenario (see Table 2 for an overview of the imprinting simulations). Imprinting was simulated using the SLINK extension SLINK Imprinting [42].

For the imprinting model, all founder genotypes were removed after data generation. The rationale behind this approach is the following: if the founder genotypes of AHSPs and ASPs are unknown, information about imprinting can only be inferred from AHSPs, with ASPs contributing only information about linkage. As a reference for comparison, a corresponding no imprinting (ni) model was considered.

Data Analysis

We used GHM version 3.1 [25] for MOD score calculation and trait-model parameter estimation. In particular, we used the GHM

Table 1. Overview of the simulated scenarios using trait models of the generic types “recessive,” “dominant,” “additive,” and “overdominant”

Model type	Name	P	F_0	F_1	F_2	Sample size
Recessive	R1	0.01; 0.1	0.01	0.01	0.2	500 and 1,000; 500
	R2	0.01; 0.1	0.01	0.01	0.5	500
	R3	0.01; 0.1	0.01	0.01	0.8	500
	AFR1	0.2	0.04	0.04	0.2	500 and 1,000
	AFR2	0.25	0.003	0.05	0.5	500
Dominant	D1	0.01; 0.1	0.01	0.2	0.2	500 and 1,000; 500
	D2	0.01; 0.1	0.01	0.5	0.5	500
	D3	0.01; 0.1	0.01	0.8	0.8	500
	AFD1	0.05	0.04	0.2	0.2	500 and 1,000
	AFD2	0.25	0.003	0.5	0.5	500
Additive	A1	0.01; 0.1	0.01	0.1	0.2	500 and 1,000; 500
	A2	0.01; 0.1	0.01	0.2	0.5	500
	A3	0.01; 0.1	0.01	0.5	0.8	500
	AFA1	0.1	0.03	0.13	0.23	500 and 1,000
	AFA2	0.5	0.003	0.25	0.5	500
Overdominant	U1	0.01; 0.1	0.01	0.2	0.01	500
	U2	0.01; 0.1	0.01	0.5	0.01	500
	U3	0.01; 0.1	0.01	0.8	0.01	500 and 1,000; 500
	U4	0.35	0.01	0.9	0.01	500

P , disease allele frequency; F_0, F_1, F_2 , penetrances with F_i denoting the probability that an individual with i copies of the disease allele is affected by the disease.

Table 2. Overview of simulated trait models with imprinting and corresponding no imprinting model

	Pedigree structure	Model name	P	F_0	$F_{1,pat}$	$F_{1,mat}$	F_2
Model with imprinting	1. 100 AHSPs with a common father + 100 AHSPs with a common mother	cpi	0.01	0	0	1	1
	2. 100 AHSPs with a common mother + 100 ASPs	cpi	0.01	0	0	1	1
	3. 20 AHSPs with a common mother + 180 ASPs	cmi	0.01	0	1	0	1
Comparison model	All structures	ni	0.01	0	0.5	0.5	1

AHSP, affected half-sib pair; ASP, affected sib pair; P , disease allele frequency; F_0, F_1, F_2 , penetrances with F_i denoting the probability that an individual with i copies of the disease allele is affected by the disease; $F_{1,pat}, F_{1,mat}$, heterozygote penetrances distinguished by the parental origin of the disease allele (*pat*: paternally inherited, *mat*: maternally inherited); cpi, complete paternal imprinting; cmi, complete maternal imprinting; ni, no imprinting.

options “modcalc single,” “penetrance restriction off,” “allfreq restriction off,” “maximization dense,” and “dimensions 4” or “dimensions 5” for ni models and imprinting models, respectively. “modcalc single” enables a separate maximization for each genetic position. “penetrance restriction off” allows for over- and under-dominant models, i.e., allows heterozygote penetrance(s) to be varied freely between 0 and 1 during the maximization. This also affects the dominance index, which is defined as

$$D = \frac{F_{1,pat} + F_{1,mat} - F_0 - F_2}{F_2 - F_0}.$$

$D = 1$ indicates a fully dominant model, whereas $D = -1$ indicates a fully recessive model. However, if the penetrances are not restricted to $F_0 < F_1 < F_2$, the dominance index may also exceed 1 or fall below -1 . Note that the dominance index is defined to be 0 for models with $F_2 = F_0$, i.e., strictly overdominant or strictly underdominant models. “allfreq restriction off” allows the disease allele frequency to be estimated >0.5 . “maximization dense” indicates that the MOD score is calculated for a greater number of predefined models before the fine maximization than in the standard setting. “dimensions 4” or “dimensions 5” allows all parameters (disease allele frequency plus 3 penetrances in the case of ni models or disease allele frequency plus 4 penetrances in the case of imprinting models) to be varied simultaneously in the maximization. For the models with imprinting, we ran 2 analyses. For the first, “imprinting” was set to “off” and “dimensions” to “4,” and for the second, they were set to “on” and “5,” to obtain ni and imprinting MOD scores, respectively. Estimates of trait-model parameters were obtained from the model yielding the highest MOD score in the analysis.

Results

Estimated values of the trait-model parameters of each simulation scenario are reported as medians based on 1,000 replicates. Sometimes, penetrances of a given replicate were estimated to be exactly 0, rendering penetrance ratios undetermined. In this case, penetrance ratios were either set to a very large number (10^6) or to 1, in case both the numerator and the denominator of the penetrance ratio were 0. Hence, no information for the estimation of the median was lost. To facilitate the comparison of the quality of estimation across pedigree types, we constructed graphics that display all 5 pedigree types using various trait models. Bias was defined as the deviation of the median estimate of a parameter from its expected value. The corresponding measure of variability is the median absolute deviation (MAD). In general, a good estimation shows both small bias and MAD (high efficiency). Impact of bias can be considered of minor importance when MAD is high. In addition to absolute penetrances, the corresponding evaluation of bias and MAD of penetrance ratios for ASPs and ASTs will be given in a dedicated sec-

tion. MOD scores for each model and pedigree type are displayed in Table 3. Parameter estimation result tables for each model and pedigree type can be found in the Appendix.

Recessive Models

The parameter estimation results for recessive models can be found in Figure 3 and Appendix Tables A1, A5, and A9. With regard to recessive models, bias and MAD were often higher for ASPs compared to ASTs (Fig. 3). This is due to the fact that only 2 out of 4 parameters (3 penetrances and the disease allele frequency) are identifiable. With ASTs, 3 out of 4 parameters should be identifiable. It is of note that it is impossible in the case of affecteds-only analysis to estimate absolute penetrance values; here, only penetrance ratios, which correspond to genotype relative risks, are identifiable in the best case. Consider, for example, the 2 sets of penetrances resulting in the same MOD score: $f_0, f_1, f_2 = 0.1, 0.5, 1$ and $f_0, f_1, f_2 = 10^{-3}, 0.005, 0.01$, with the first set being more likely to be evaluated in the analysis due to the predefined trait models initially tested by GHM before the fine maximization. Generally, for all types of models (recessive, dominant, additive, and overdominant), higher MOD scores were obtained for ASTs compared to ASPs.

With ASPs, most recessive models were recognized as such, indicated by a median dominance index $d < 0$. Only R1, a model with an extremely reduced penetrance, was estimated as being additive (median $d = 0$) for $P = 0.01$. This is due to the fact that affected persons are more likely to be phenocopies in the context of a strongly reduced penetrance and a small disease allele frequency, which reduces the amount of allele sharing among affected siblings. An equivalent explanation for this can be found in Figure 4, which shows the projection of the estimated trait-model parameters for ASP pedigrees on the triangular parameter space as described in the Introduction (subsection “Identifiability of Inheritance Parameters”). For all models, the estimated values scattered around the true values without systematic deviation. However, the true value for model R1 with $P = 0.01$ lies close to the point of no linkage in the upper right corner of the triangle. In the proximity of this point, all types of generic models (recessive, dominant, additive, and overdominant) accumulate and are hard to distinguish from each other.

For ASTs, all recessive models were clearly recognized as such. Estimation accuracy of the dominance index D improved from ASPs to ASTs for most recessive models. Intriguingly, ASTs even showed the best parameter estimation performance in terms of small bias and MAD

Table 3. MOD scores of the simulated trait-model scenarios for all pedigree types

Model name	P	Estimated median MOD (MAD)				
		ASP	AST	DST	DSQ	
R1	0.1	33.07 (5.51)	134.47 (11.19)	32.41 (5.32)	31.25 (5.36)	37.85 (5.45)
R1	0.01	0.07 (0.11)	2.89 (1.69)	0.23 (0.32)	0.27 (0.35)	0.49 (0.49)
R1n1000	0.01	0.09 (0.14)	5.58 (2.31)	0.26 (0.36)	0.31 (0.38)	0.52 (0.5)
R2	0.1	100.76 (8.31)	207.56 (12.31)	108.46 (9.06)	114.38 (9.76)	89.51 (7.01)
R2	0.01	1.08 (0.97)	116.89 (11.4)	1.15 (0.88)	1.0 (0.88)	1.84 (1.1)
R3	0.1	127.37 (9.16)	225.05 (14.13)	162.37 (10.9)	189.64 (11.69)	158.31 (9.82)
R3	0.01	5.37 (2.29)	283.25 (15.7)	4.76 (2.03)	3.86 (1.87)	8.35 (2.72)
AFR1	0.2	2.69 (1.53)	13.02 (3.53)	2.81 (1.49)	2.71 (1.52)	3.78 (1.69)
AFR1n1000	0.2	5.26 (2.19)	25.87 (5.26)	5.13 (2.07)	5.08 (1.95)	7.08 (2.39)
AFR2	0.25	40.43 (5.97)	62.95 (7.65)	50.02 (6.3)	59.07 (6.24)	33.99 (4.87)
D1	0.1	19.28 (3.72)	36.43 (5.36)	20.64 (3.87)	21.95 (4.14)	29.17 (4.97)
D1	0.01	25.19 (4.2)	91.2 (7.44)	24.93 (4.28)	24.45 (4.18)	65.87 (6.25)
D1n1000	0.01	49.92 (5.82)	182.62 (9.84)	49.26 (5.95)	48.61 (5.91)	131.43 (8.67)
D2	0.1	24.43 (4.28)	43.74 (6.06)	34.74 (4.87)	45.41 (5.99)	56.48 (6.89)
D2	0.01	47.85 (5.07)	124.71 (7.95)	54.98 (5.39)	60.9 (6.34)	117.79 (7.59)
D3	0.1	25.98 (4.38)	45.93 (6.11)	67.63 (6.82)	115.52 (9.44)	133.19 (10.3)
D3	0.01	54.44 (5.11)	134.19 (7.84)	90.84 (6.9)	125.33 (9.12)	206.27 (10.94)
AFD1	0.05	4.04 (1.9)	15.8 (3.67)	4.19 (1.81)	4.1 (1.76)	9.0 (2.66)
AFD1n1000	0.05	7.94 (2.51)	31.42 (5.15)	7.92 (2.59)	7.85 (2.55)	17.45 (3.51)
AFD2	0.25	8.94 (2.74)	12.37 (3.22)	15.77 (3.77)	24.13 (4.46)	21.30 (4.21)
A1	0.1	13.58 (3.25)	28.6 (4.72)	14.13 (3.4)	14.37 (3.44)	15.05 (3.4)
A1	0.01	7.45 (2.48)	48.12 (6.2)	7.36 (2.44)	7.17 (2.46)	28.56 (4.57)
A1n1000	0.01	14.85 (3.48)	96.12 (8.53)	14.62 (3.25)	13.99 (3.14)	57.27 (6.14)
A2	0.1	21.84 (3.93)	40.65 (5.34)	24.22 (4.08)	26.54 (4.43)	23.93 (4.22)
A2	0.01	25.46 (4.21)	90.54 (7.48)	25.42 (4.22)	25.01 (4.17)	62.34 (6.53)
A3	0.1	24.43 (3.88)	41.74 (5.63)	37.47 (5.02)	50.58 (6.05)	56.32 (6.83)
A3	0.01	47.70 (5.18)	122.61 (8.05)	55.18 (5.67)	61.37 (6.5)	116.46 (7.99)
AFA1	0.1	3.95 (1.79)	11.81 (3.3)	4.22 (1.82)	4.15 (1.72)	5.81 (2.27)
AFA1n1000	0.1	7.69 (2.43)	23.10 (4.54)	8.12 (2.42)	8.06 (2.42)	10.96 (3.12)
AFA2	0.5	2.33 (1.39)	3.28 (1.63)	3.74 (1.73)	5.38 (2.16)	3.7 (1.69)
U1	0.1	21.22 (3.97)	46.18 (5.88)	22.59 (3.98)	23.91 (4.33)	51.61 (5.54)
U1	0.01	25.33 (4.09)	93.75 (7.37)	24.99 (4.09)	24.53 (4.17)	69.53 (6.35)
U2	0.1	27.37 (4.3)	56.77 (6.8)	36.69 (5.19)	46.41 (6.11)	88.48 (6.96)
U2	0.01	48.26 (5.15)	128.75 (7.97)	55.03 (5.79)	60.94 (6.36)	123.23 (7.73)
U3	0.1	29.17 (4.32)	59.87 (6.86)	68.25 (6.92)	109.36 (8.95)	153.82 (9.67)
U3	0.01	54.91 (4.99)	139.16 (7.7)	90.54 (6.97)	124.03 (9.13)	208.56 (10.64)
U3n1000	0.01	109.07 (6.81)	277.17 (11.52)	180.17 (10.03)	247.99 (13.17)	416.58 (15.56)
U4	0.35	12.4 (3.3)	19.99 (4.24)	66.74 (6.53)	136.54 (9.11)	149.85 (8.58)

MAD, median absolute deviation, adjusted by a constant (1.4826) for asymptotically normal consistency; ASP, affected sib pair; AST, affected sib triplet; DST, discordant sib triplet; DSQ, discordant sib quadruplet; 3-G, 3-generation pedigree; P, true value for the disease allele frequency.

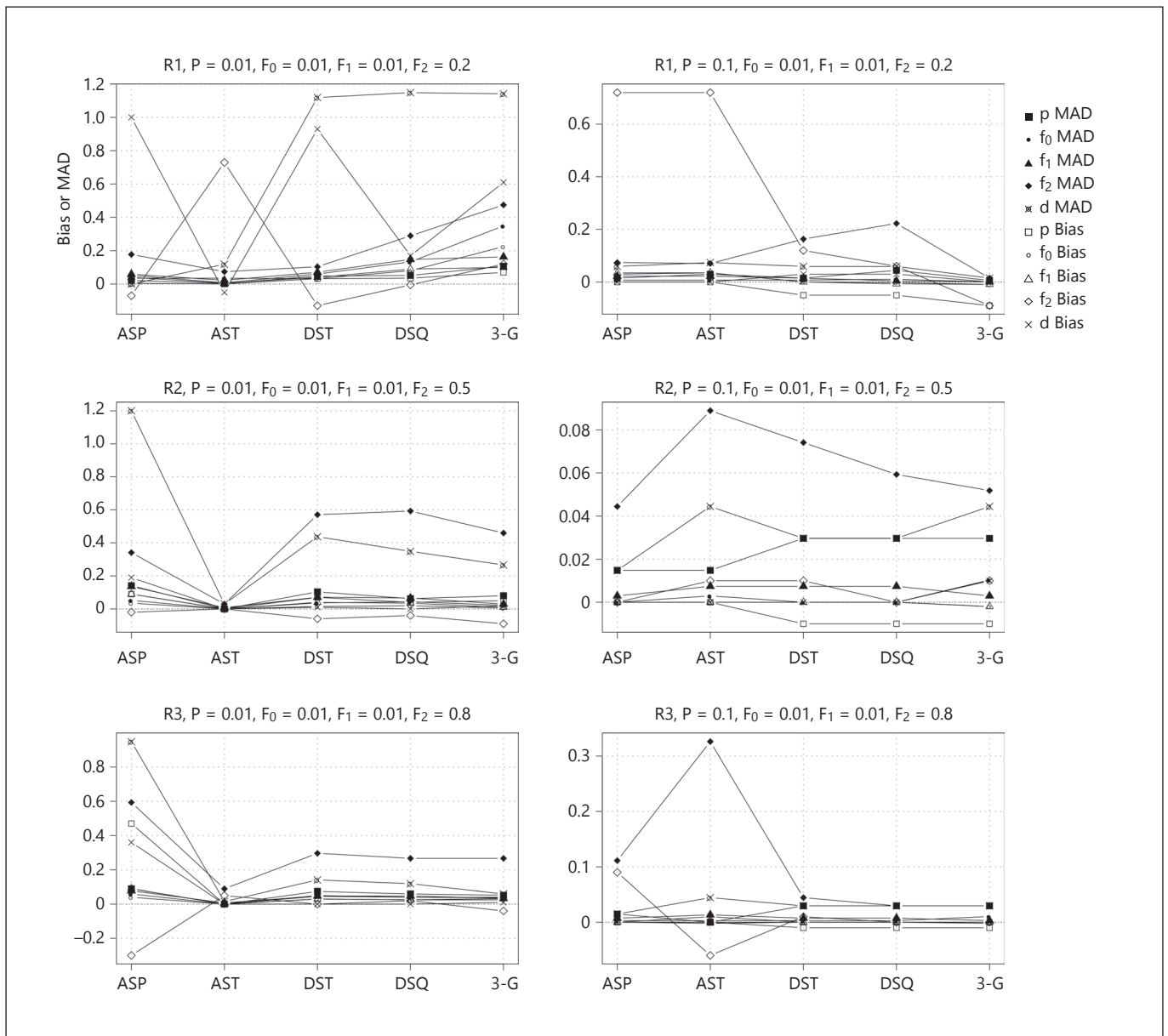


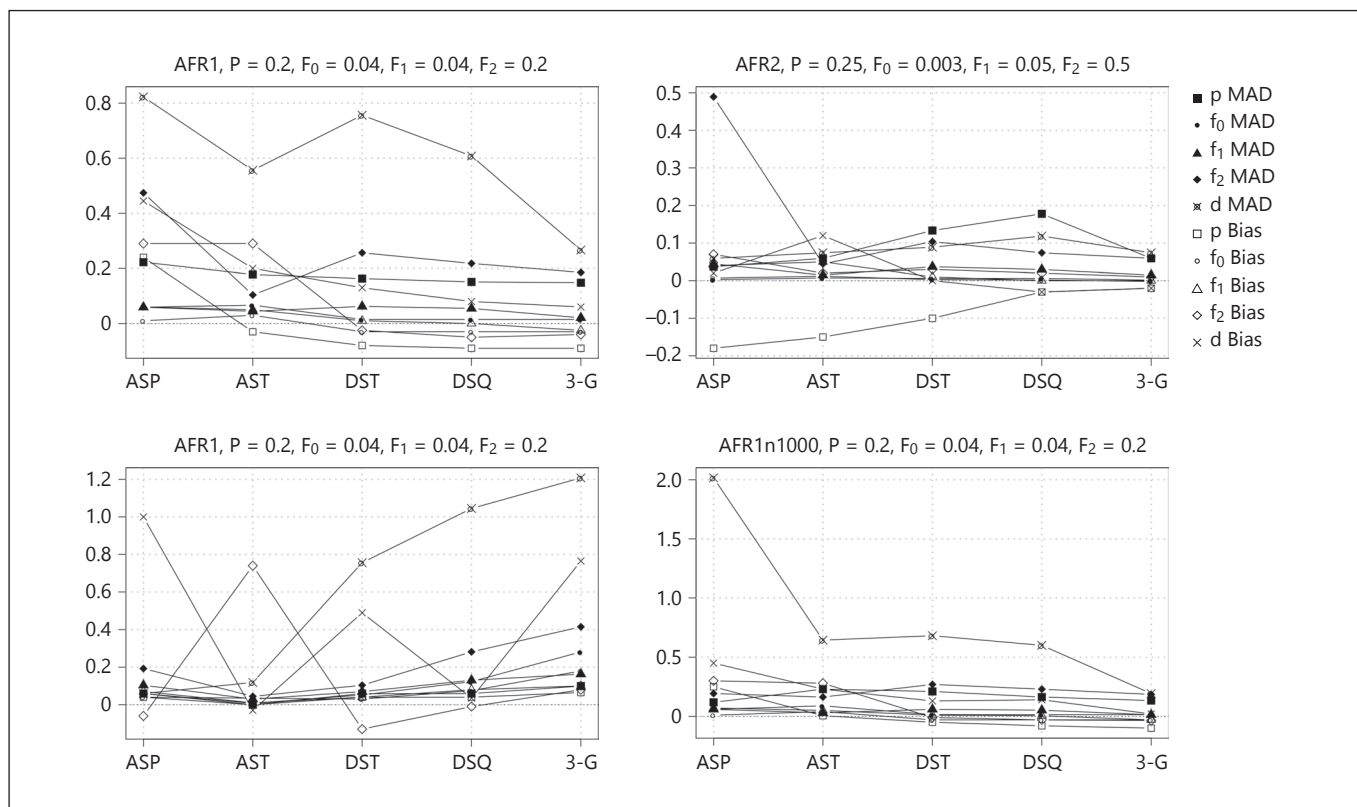
Fig. 3. Illustration of bias and variability of the parameter estimation for recessive models using different pedigree types. The trait-model parameters used for the simulations are given above the panels for each trait model. Estimations of the individual parameters are depicted by five unique symbols. For each parameter, the median absolute deviation (MAD) and bias, defined as $bias = me-$

$dian$ (true parameter value – estimated value), are plotted. Pedigree types (for details, see Fig. 2 and its legend) are displayed on the x-axis with increasing complexity, i.e., ASPs are located on the very left side and 3-G pedigrees are located on the very right side. p , disease allele frequency; f_i , penetrances; d , dominance index.

(Figure continued on next page.)

across all investigated pedigrees for models R2 and R3 both with $P = 0.01$ (Fig. 3). This might be explained as follows: although only penetrance ratios can in theory be estimated using ASTs, the corresponding set of absolute values of the penetrances resulting in such high ratios

(F_2/F_1 : 50 for R2 and 80 for R3) is limited in a maximization starting with a fixed grid of genetically plausible values (the genotype relative risk of model R1 with $P = 0.01$ obviously was too low to show the aforementioned effect). Further, despite the small disease allele frequency, a



low phenocopy rate together with a high penetrance ensures enough information for the estimation of F_2 in relation to F_0 and F_1 in the context of ASTs. In addition, the number of degrees of freedom in an AST MOD score analysis is lower compared to an analysis with pedigrees containing healthy individuals, which can lead to a higher power of an affecteds-only analysis (see also Flaquer and Strauch [41]) and hence to a more efficient parameter estimation for some model types (up to a constant factor multiplied to all penetrances).

With regard to DSTs and DSQs, all models were correctly classified as being recessive, and the median dominance index was mostly close to its expected value. In most cases, median estimates of all parameter values were similar for the 2 pedigree types. When the true disease allele frequency was small ($P = 0.01$), it was always overestimated. When it was large ($P \geq 0.1$), it was always underestimated. Penetrances F_0 and F_1 were estimated with high accuracy for models R1–R3 with $P = 0.1$ and model AFR2. For models R1–R3 with $P = 0.01$, F_0 and F_1 were overestimated. In the case of model AFR1, F_0 was underestimated; however, F_1 was estimated with good accuracy. Median estimates of F_2 were close to their ex-

pected values for most models, with higher accuracy for DSQs compared to DSTs. In general, F_2 could be estimated more accurately for stronger genetic models, which is the case for the investigated recessive models with higher penetrance and disease allele frequency. MOD scores were comparable for DSTs and DSQs (Table 3), except for models R2 ($F_2 = 0.5$) and R3 ($F_2 = 0.8$) with $P = 0.1$ as well as model AFR2 ($F_2 = 0.5$). This is due to the fact that an additional healthy individual increases linkage information only if penetrance and genotype relative risk are sufficiently high ($F_2 \gg F_0, F_1$ for a recessive model).

Using 3-G pedigrees, median estimated dominance indices were all close to their expected values except for model R1 with $P = 0.01$. The estimation of the disease allele frequency was accurate for models R2 and R3 both with $P = 0.1$ and AFR2 with $P = 0.25$. The median F_0 and F_1 penetrances were estimated with good accuracy for models R3 with $P = 0.1$, R2, and AFR2. The homozygous mutant penetrance F_2 was estimated with good accuracy for models R2 with $P = 0.1$, R3, and AFR2. However, in all other cases, the estimated median F_2 was still larger than the corresponding medians for F_0 and F_1 .

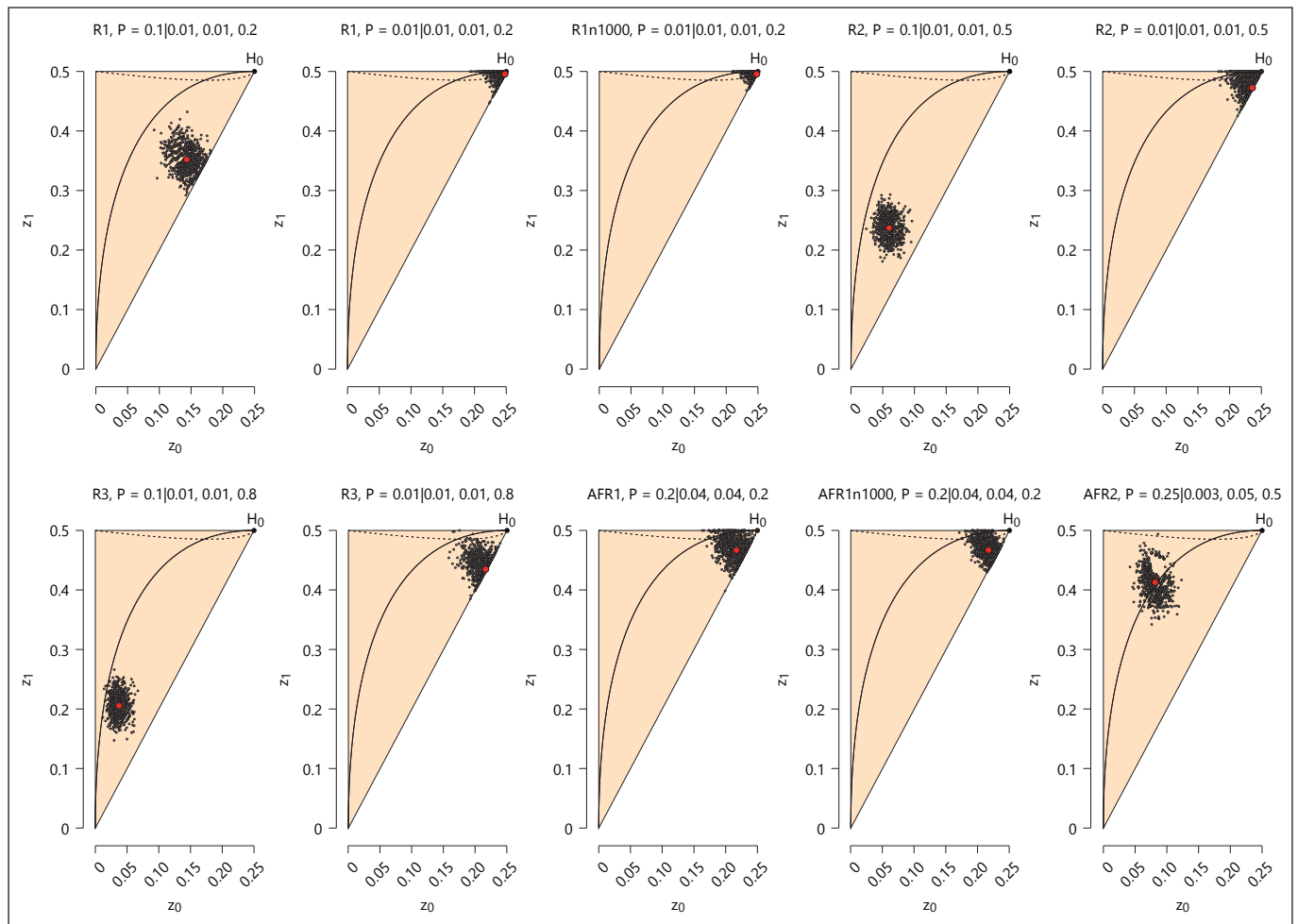


Fig. 4. Recessive models: projections of trait-model parameter estimates on the possible triangle parameter space of affected sib pairs (ASPs). The trait-model parameters used for the simulation are given above the panels for each trait model, and its projection in terms of allele-sharing is depicted by a red dot. z_0 , allele-sharing

probability that an ASP shares no allele identical-by-descent (IBD); z_1 , allele-sharing probability that an ASP shares 1 allele IBD; trait-model parameters used for the simulation: disease allele frequency P and penetrances F_0, F_1, F_2 .

With regard to DSTs, DSQs, and 3-G pedigrees, bias was often smaller than MAD across all models, yet especially large MADs were obtained for penetrance F_2 and models R2 and R3 both with $P = 0.01$ as well as for AFR1. Values for bias and MAD did not consistently decrease when moving from DSTs over DSQs to 3-G pedigrees, except for the weak genetic model R1 with $P = 0.01$ and AFR2 (Fig. 3). Better parameter identifiability when moving from ASTs to DSTs as measured by a reduction in bias, especially of the F_2 penetrance, could only be observed for models R1 with $P = 0.1$ and AFR1.

Dominant Models

Parameter estimation results for dominant models are given in Figure 5 and Appendix Tables A2, A6, and A10. The estimation of individual parameters for ASPs and ASTs was not very accurate, which is in line with the fact that exact penetrances cannot be estimated for affecteds-only pedigrees, as explained above. The median dominance index was underestimated for all models, some of which were even misclassified as being additive. In the case of ASPs, this can be explained by the proximity of both model classes in the triangular parameter space (Fig. 6, 7). In particular, dominant models without phenocopies are represented by the dashed line, whereas ad-

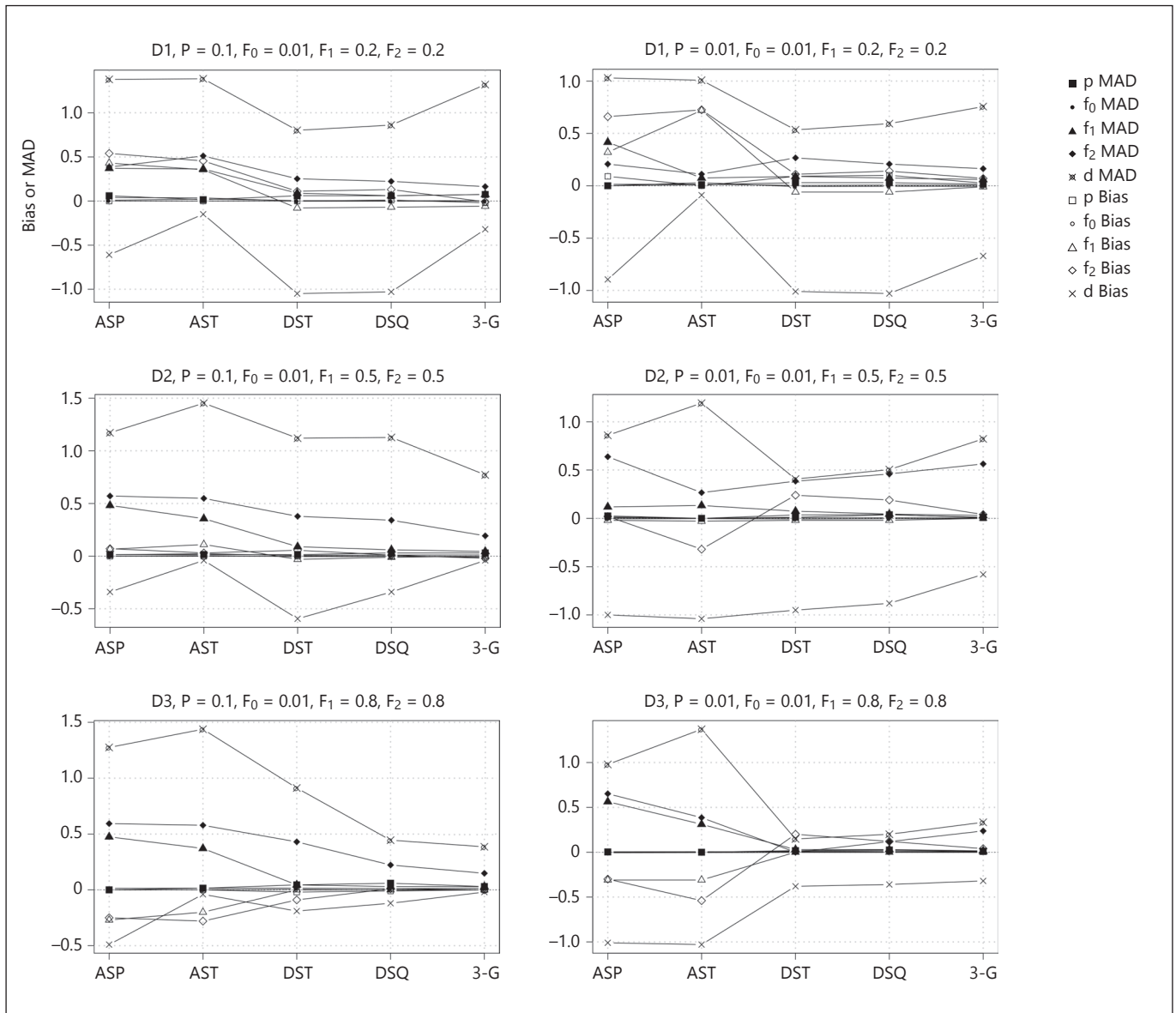
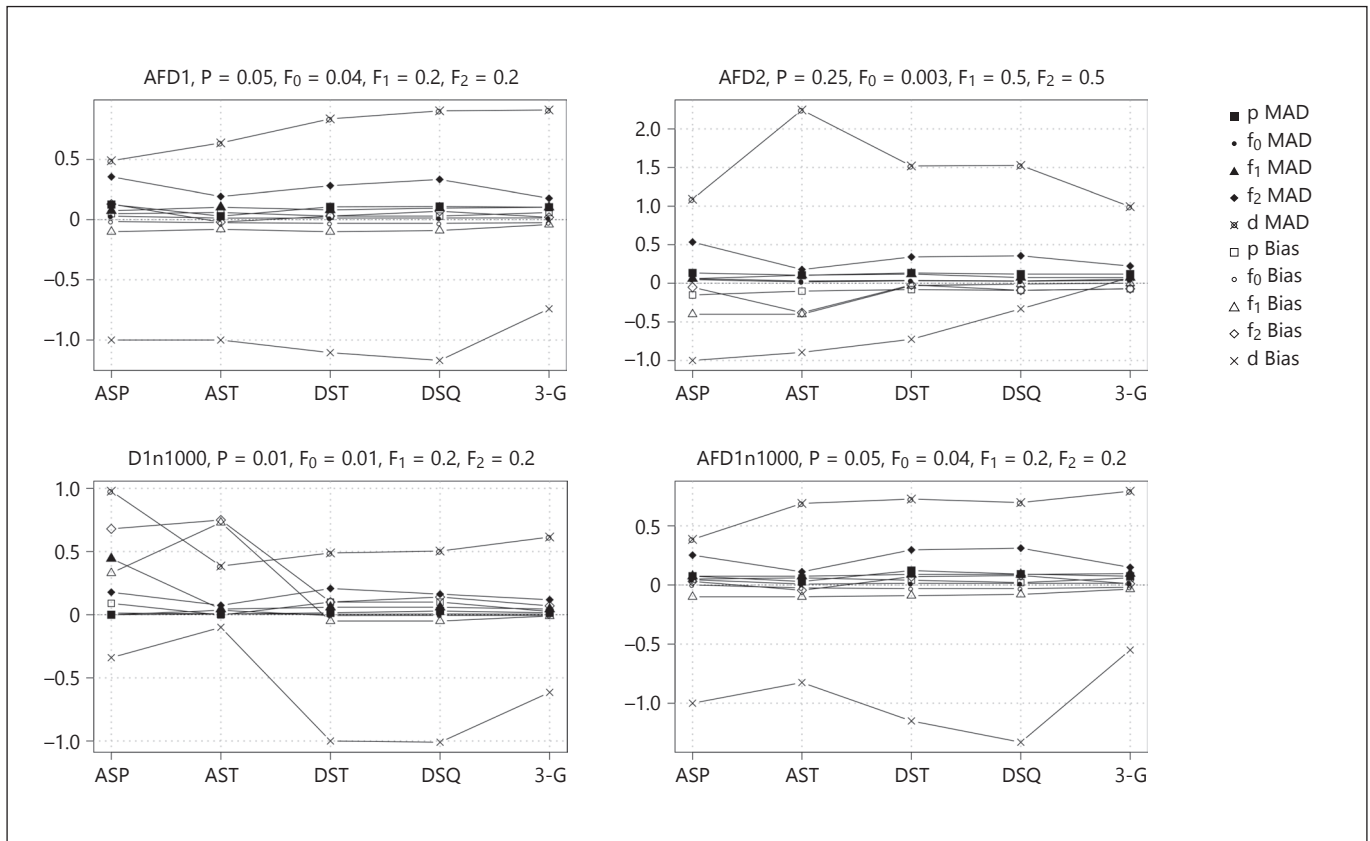


Fig. 5. Illustration of bias and variability of the parameter estimation for dominant models using different pedigree types. For more details, see Figure 3.

(Figure continued on next page.)

ditive models lie on the upper edge of the triangle. Hence, models D1–D3 with $P = 0.01$ and AFD1, which are located closest to the upper edge of the triangle, showed a median estimated dominance index d close to 0, corresponding to an additive model. The estimation of the dominance index improved when analyzing ASTs instead of ASPs for most models. The same holds for the disease allele frequency, albeit to a lesser degree.

For DSTs and DSQs, many models were misclassified as rather additive for both pedigree types when looking at their corresponding dominance indices. Only the median dominance index d for model D3 with $P = 0.1$ clearly pointed to dominance ($d = 0.81$ for DSTs and $d = 0.88$ for DSQs). Otherwise, median dominance indices for models D2, D3, and AFD2 were all positive but clearly below 1 for both pedigree types. Models D1 and AFD1 even showed median d values around 0 and below 0, re-



5

spectively. The disease allele frequency was estimated accurately for models D1–D3 with $P = 0.1$, overestimated for models D1–D3 with $P = 0.01$ and the AFD1 model, and underestimated for the AFD2 model. Estimates of P were comparable between both pedigree types. Penetrance F_0 was mostly underestimated for models D1–D3 and AFD1 using both pedigree types. With regard to F_1 , models D2, D3, and AFD2 showed good accuracy for both pedigree types, whereas it was underestimated for models D1 and AFD1. F_2 was often overestimated. Similar to recessive models, MOD scores were comparable between DSTs and DSQs (Table 3), except for models D2 ($F_1, F_2 = 0.5$), D3 ($F_1, F_2 = 0.8$), and AFD2 ($F_1, F_2 = 0.5$). As before, this is due to the fact that an additional healthy individual increases linkage information only if penetrance and genotype relative risk are sufficiently high ($F_1, F_2 \gg F_0$ for a dominant model). Only in this case, penetrance estimation is also improved for DSQs compared to DSTs.

In the case of 3-G pedigrees, median d values pointed towards dominance for all models. Median dominance indices were close to their expected values for models D2

and D3 with $P = 0.1$ as well as model AFD2. Estimates of the disease allele frequency showed good accuracy for models D1 with $P = 0.1$, D2, and D3. Estimates for F_0 were mostly close to the expected value. Estimates for F_1 and F_2 were very close to their expected values, with the highest accuracy for models D2 and D3.

With respect to dominant models, bias and MAD decreased when moving from ASPs over ASTs, DSTs, and DSQs to 3-G pedigrees for models D1, D2, and D3 all with $P = 0.1$ (Fig. 5). Median bias for F_2 seemed to be unduly small for ASPs for model D2 with $P = 0.01$. This can be explained by looking at the corresponding parameter distribution for ASPs (data not shown), which showed that F_2 was mostly estimated near 0 (<0.1 in 25.3% of the replicates) or 1 (>0.9 in 36.6% of the replicates). This is also reflected in the high MAD of F_2 (Fig. 5). Generally, for all dominant models, bias and MAD mostly decreased when moving from affecteds-only pedigrees over DSTs and DSQs to 3-G pedigrees. Only for model AFD1, the results were similar across all pedigree types. Bias was mostly smaller than MAD across all models for DSTs, DSQs, and 3-G pedigrees.

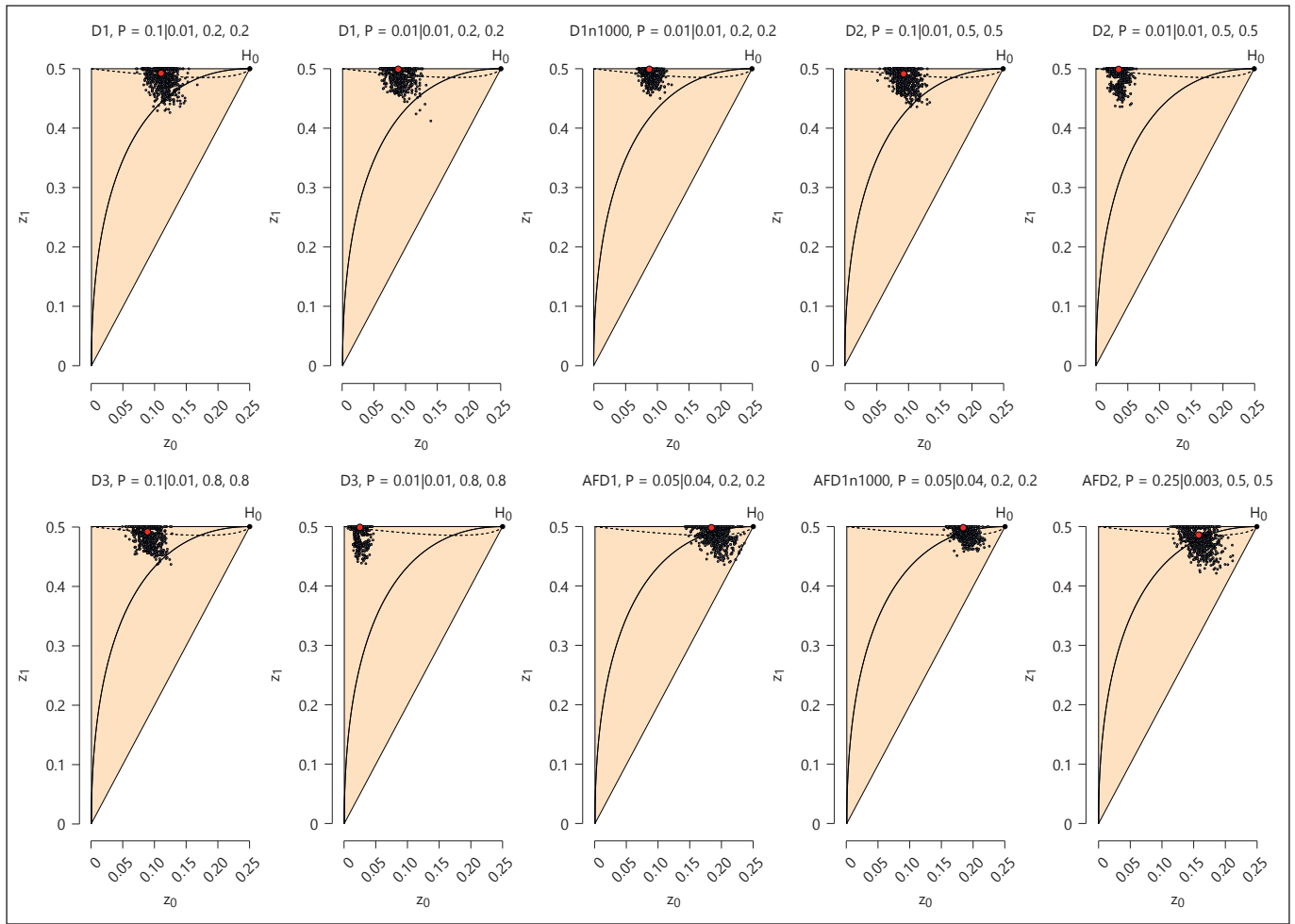


Fig. 6. Dominant models: projections of trait-model parameter estimates on the possible triangle parameter space of ASPs. For more details, see Figure 4.

Additive Models

Parameter estimation results for additive models are depicted in Figure 8 and Appendix Tables A3, A7, and A11. For ASPs, the projection of estimated trait-model parameters on the triangular parameter space, as displayed in Figure 7, illustrates that all additive models are very close to the upper edge of the triangle. Model AFA2, which has the weakest genetic effect among the investigated additive models, shows the largest distance to strictly dominant models (dashed line in Figure 7) within the allele-sharing parameter space of ASPs. The median estimated dominance indices d were close to their expected values for both ASPs and ASTs, except for model A2, which showed deviation towards dominance, and model A3. For most models and both pedigree types, the median

estimated disease allele frequency p was also close to the expected value. Again, the estimation of individual penetrances for ASPs and ASTs was not very accurate, given that these pedigree types contain only affected individuals.

For DSTs and DSQs, the median dominance indices tended towards their expected values, but were not accurate for most models. The estimation of the disease allele frequency was comparable between DSTs and DSQs and showed good accuracy for models A1 with $P = 0.01$ as well as models A2 and A3 both with $P = 0.1$. Otherwise, models with $P = 0.01$ showed an overestimated disease allele frequency (A2, A3), whereas for models with $P \geq 0.1$ it was underestimated (A1, AFA1, AFA2). Penetrances F_0 and F_1 were estimated accurately for all models and

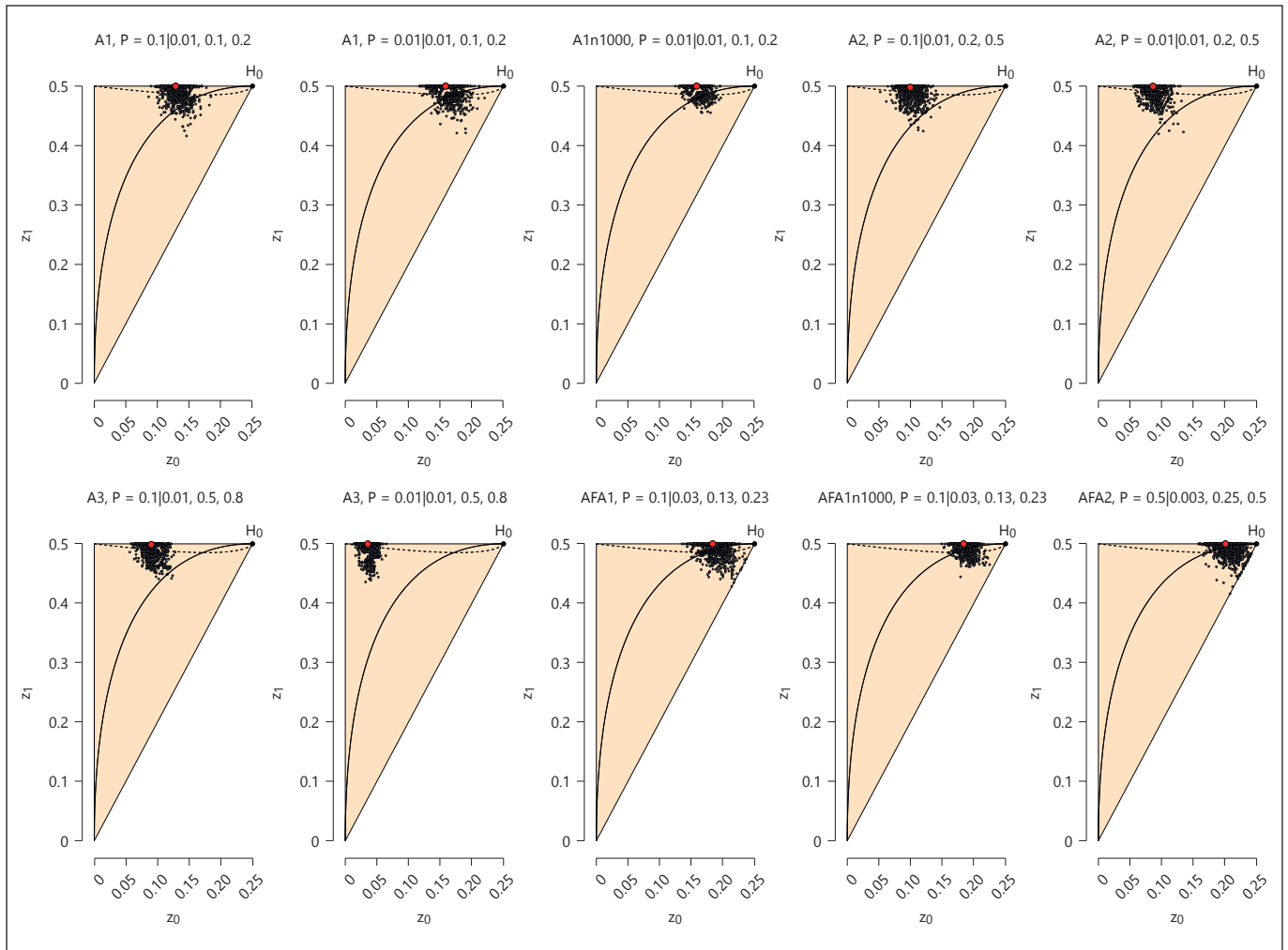


Fig. 7. Additive models: projections of trait-model parameter estimates on the possible triangle parameter space of ASPs. For more details, see Figure 4.

both pedigree types, with a slight underestimation in some cases. F_2 was estimated with acceptable accuracy for both pedigree types; however, it was always underestimated, most prominently for model A2 with $P = 0.01$ ($F_2 = 0.5$; $f_2 = 0.335$ for DSTs and $f_2 = 0.35$ for DSQs). The parameter estimation did not substantially improve when using DSQs instead of DSTs (Fig. 8). This is in line with the MOD scores in Table 3, which were comparable between DSTs and DSQs, with only a slight increase for models A3 and AFA2. As before, this is due to the fact that models A3 and AFA2 show the highest penetrance and genotype relative risk among the investigated models, such that an additional healthy individual can contribute at least some extra linkage information in the analysis.

The accuracy of median d values for additive models was not very high when using 3-G pedigrees in the analysis. However, most dominance indices still pointed to additivity. The results for the disease allele frequency showed good accuracy for models A1 and A2, each with $P = 0.1$, A3, and AFA1. The estimates for penetrance F_0 showed good accuracy for most models. Median estimates for F_1 were mostly identical to their expected value. Penetrance F_2 was estimated with good accuracy for models A1 and A2, each with $P = 0.1$, A3, AFA1, and AFA2.

The results for the additive models in Figure 8 showed a general trend towards less bias when moving from affecteds-only pedigrees over DSTs and DSQs to 3-G pedigrees, except for model AFA1. When moving from DSTs over DSQs to 3-G pedigrees, MAD slightly decreased ex-

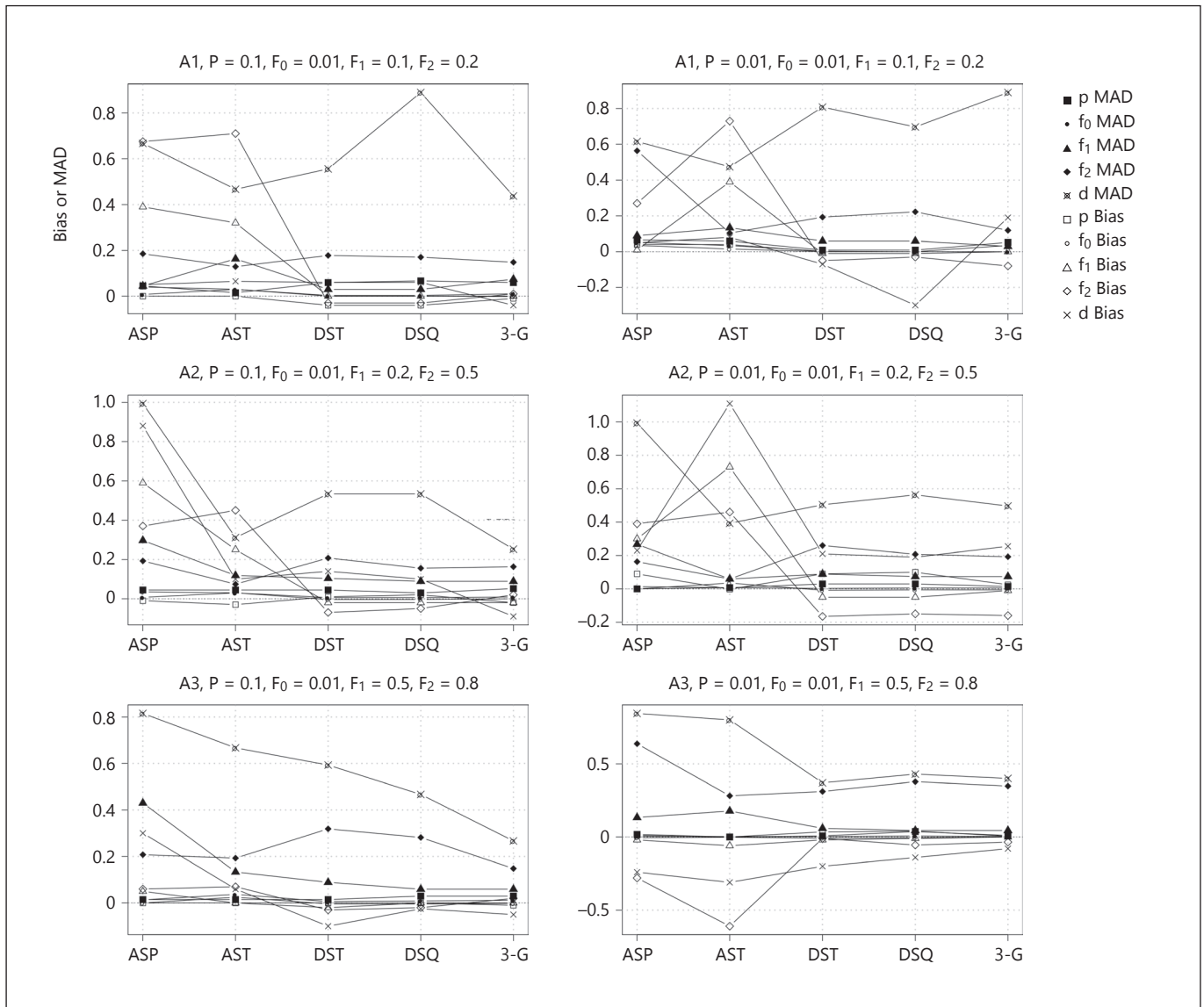


Fig. 8. Illustration of bias and variability of the parameter estimation for additive models using different pedigree types. For more details, see Figure 3.

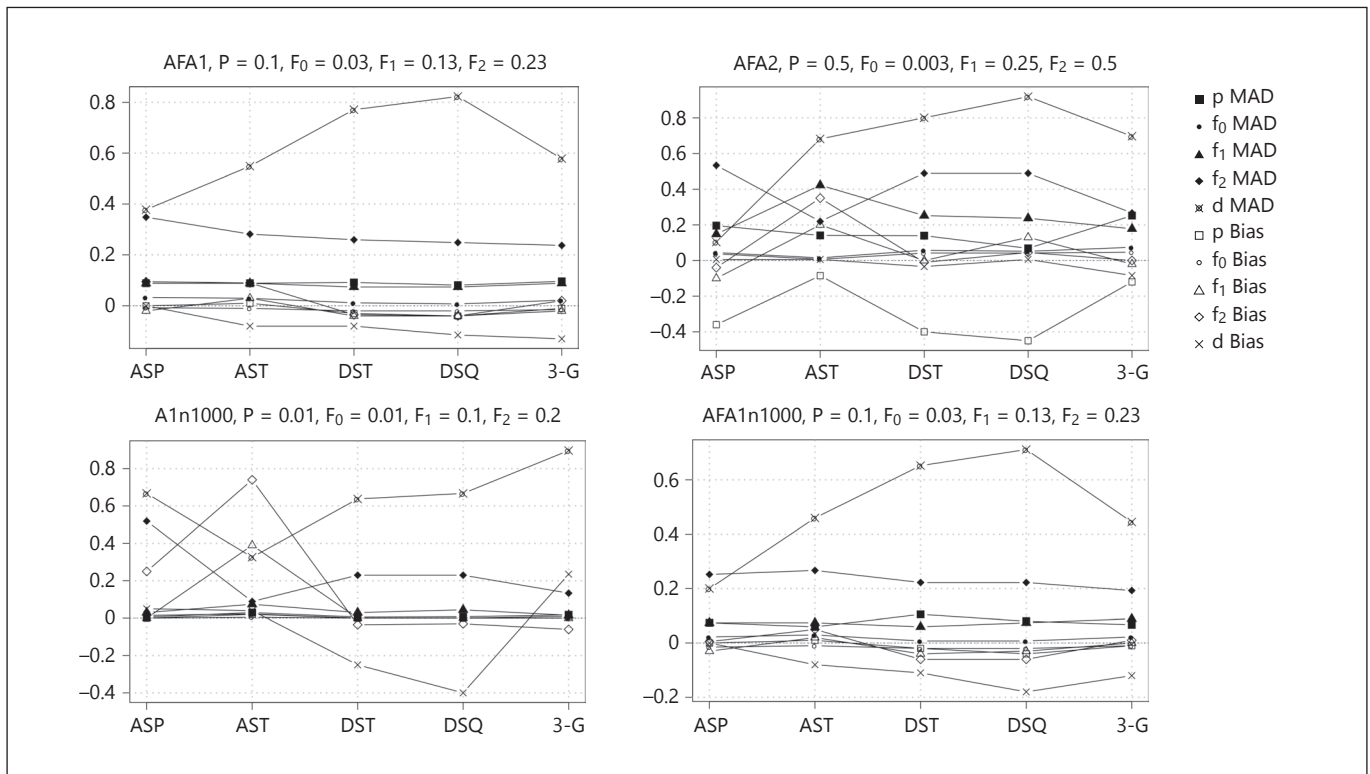
(Figure continued on next page.)

cept for models A3 with $P = 0.01$ and AFA1. Bias was mostly smaller than MAD across all models for DSTs, DSQs, and 3-G pedigrees.

Overdominant Models

Parameter estimation results for overdominant models are given in Figure 9 and Appendix Tables A4, A8, and A12. As already mentioned above, the dominance index D is defined to be 0 for models with $F_0 = F_2$, because the denominator would be 0. Therefore, D cannot serve as a performance measure for the analyzed overdominant

models. For ASPs and most models, the median disease allele frequency p was estimated close to the expected value. Overdominance, i.e., $F_0 < F_1$ and $F_2 < F_1$, was correctly assessed for models U1–U3 with $P = 0.1$ and U3 with $P = 0.01$ and a sample size of 1,000 pedigrees (Appendix Table A4). All other models were classified as rather additive (e.g., U1 with $P = 0.01$) or dominant (e.g., U2 with $P = 0.01$). The projections of the estimated trait-model parameters in the parameter space of ASPs are shown in Figure 10. The allele-sharing estimates of particular models were not evenly distributed around the true point (e.g.,



8

U2 with $P = 0.1$ and U3 with $P = 0.1$), which might be due to peculiarities of the parameter space. The true point for model U1 with $P = 0.01$ lies between the upper edge of the triangle, which corresponds to additive models, and the dashed line, representing dominant models. The location and distribution of the estimates for this model resembled those of the additive model A3 with $P = 0.1$ depicted in Figure 7. Indeed, the median estimates for U1 with $P = 0.01$ and A3 with $P = 0.1$ were similar for all penetrances as well as the disease allele frequency.

With regard to ASTs, the median disease allele frequency p was estimated close to the true value for all models. Overdominant models could be better distinguished from other model types when using ASTs instead of ASPs, because the corresponding allele-sharing values form a unique, separated compartment of the 3-dimensional parameter space (Fig. 1). Hence, overdominance was correctly assessed for all models except model U4 (Fig. 9, ASTs). Why model U4 was so difficult to be classified as overdominant for ASPs and ASTs can be explained as follows. As can be seen from Figures 1 (ASTs) and 10 (ASPs), model U4 occupies a distinct part of the parameter space as compared to models U1–U3. For both pedigree types, however, it can be shown that this distinct part of the pa-

parameter space can as well be occupied by underdominant models, i.e., $F_0 > F_1$ and $F_2 > F_1$, which is reflected by the corresponding median penetrance estimates for ASTs (Appendix Table A4).

For DSTs and DSQs, estimates of the disease allele frequency for models U1–U3 with $P = 0.1$, model U3 with $P = 0.01$, and model U4 showed good accuracy; otherwise, it was clearly overestimated. Median F_0 penetrances were estimated around their expected value (0.01) for both pedigrees, albeit slightly underestimated. Estimations of the median F_1 penetrance were accurate for all models, except model U1. Estimating penetrance F_2 proved to be difficult, since only models U3 with $P = 0.1$ and U4 showed values that were near their expectations. Generally, an estimation of F_2 is difficult when the disease allele frequency is low, because only a few individuals of the dataset actually have a homozygous mutant genotype and can contribute information to the estimation of F_2 . Therefore, the relations $F_0 < F_1$ and $F_2 < F_1$ were only identified for models U2 and U3 both with $P = 0.1$ and U4 with $P = 0.35$ for both pedigree types. As explained above, the additional healthy individual in DSQs can increase the MOD score only if the penetrance and the genotype relative risk are sufficiently high, which is the case for models U2, U3

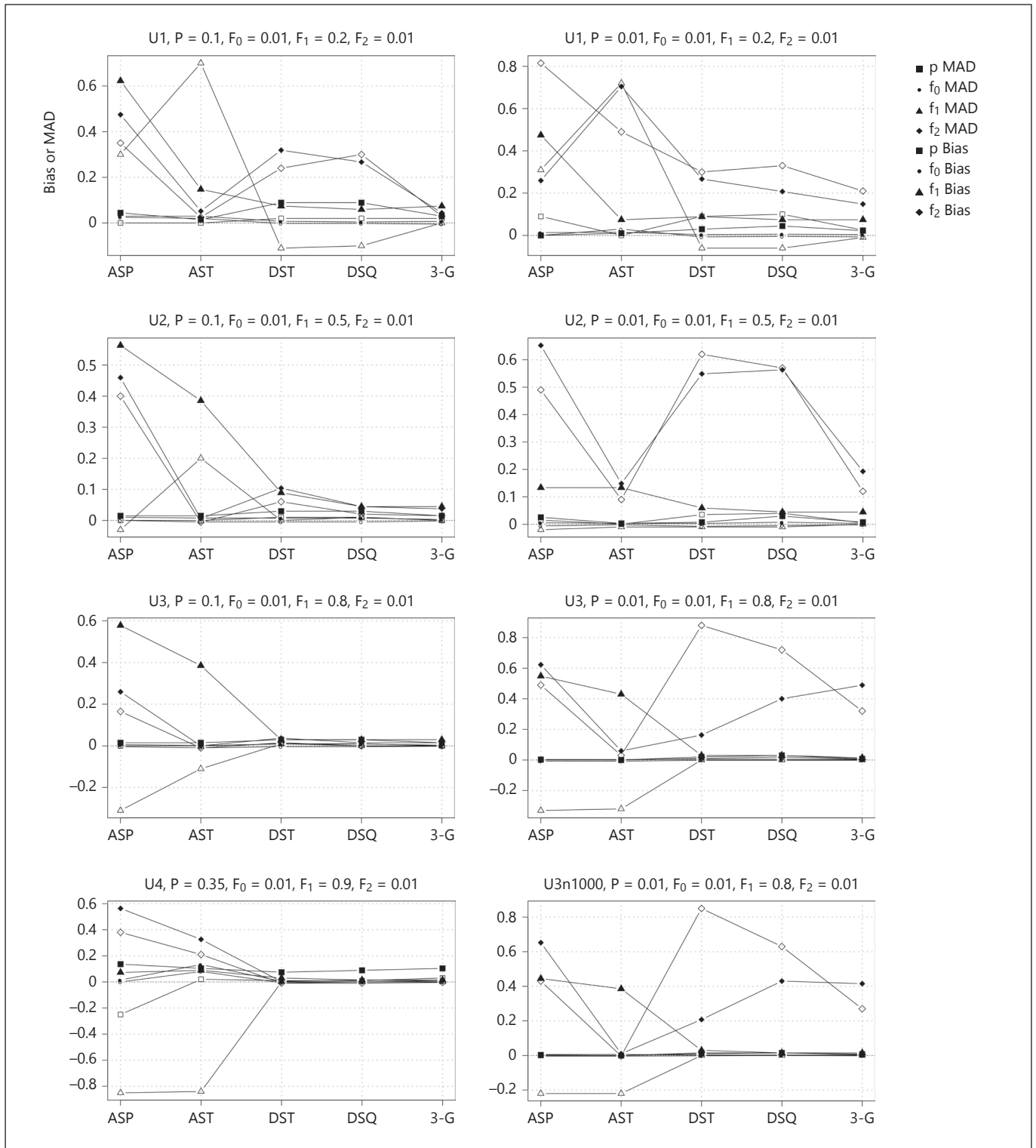


Fig. 9. Illustration of bias and variability of the parameter estimation for overdominant models using different pedigree types. For more details, see Figure 3.

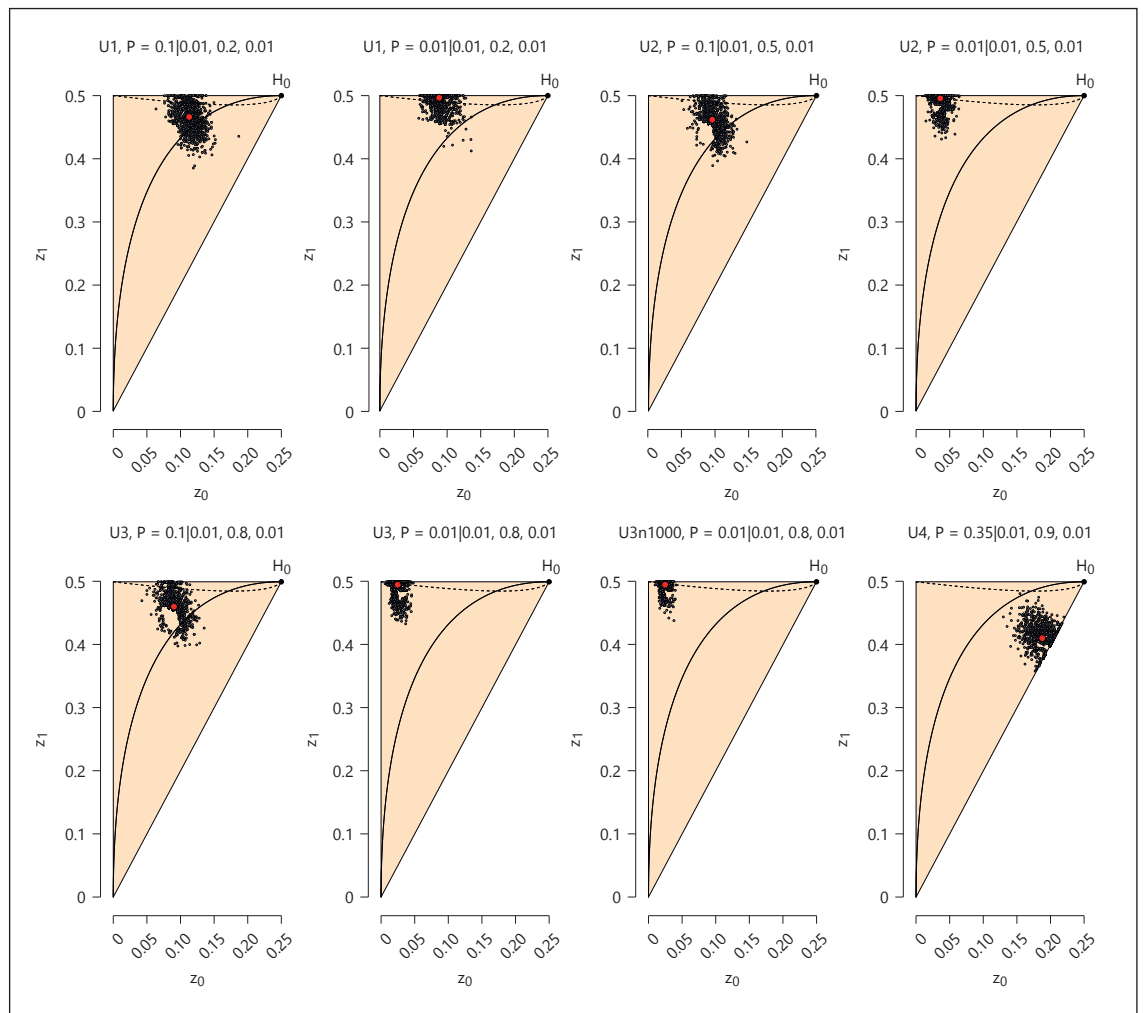


Fig. 10. Overdominant models: projections of trait-model parameter estimates on the possible triangle parameter space of ASPs. For more details, see Figure 4.

and U4 (Table 3). By the same token, when adding a second healthy individual, penetrance estimation was also improved for model U3 with $P = 0.01$, which pointed to overdominance only for DSQs but not for DSTs (Appendix Table A8).

Using 3-G pedigrees, the estimation of the disease allele frequency showed good accuracy for most models, especially for models U1–U3 with $P = 0.1$ and model U4. The median penetrances F_0 were estimated near their expected value (0.01) for all models, albeit slightly underestimated in most cases. Penetrance F_1 was estimated with very high accuracy, with all but one medians estimated exactly at the expected value. The accuracy of the estimation of F_2 depended on the disease allele frequency – models with $P \geq 0.1$ showed good accuracy, where-

as F_2 was always overestimated for models with $P = 0.01$. As mentioned above, when the disease allele frequency is low, the dataset contains too few individuals with a homozygous mutant genotype that can contribute to the estimation of F_2 . However, median estimates of F_2 were significantly lower than those of F_1 , which clearly indicates overdominance, except for model U1 with $P = 0.01$.

For models U1–U3 with $P = 0.01$, median bias of F_2 was high, especially for ASPs, DSTs, and DSQs (Fig. 9). This is due to the fact that ASPs, DSTs, and DSQs contain only 2 affected individuals, compared to ASTs and 3-G pedigrees having 3 affected individuals. The additional affected individual results in a larger number of mutant alleles per pedigree and hence in a larger number of ho-

mozygous mutant individuals. Bias and MAD decreased when moving from DSTs over DSQs to 3-G pedigrees for most models. Better identifiability of parameters as measured by a reduction in bias when using pedigrees with unaffected individuals could only be observed for models U1 with $P = 0.01$, U3 with $P = 0.1$, and U4. In DSTs, DSQs, and 3-G pedigrees, bias was often larger than MAD for F_2 . As can be seen from Figure 9, parameter estimation results were best for models with $P \geq 0.1$, especially when using 3-G pedigrees.

Penetrance Ratios for ASPs and ASTs

As already mentioned above, the exact numerical values for trait-model parameters cannot be obtained from affecteds-only analyses. However, the corresponding penetrance ratios can in principle be estimated. In Table 4, we present the estimation of pairwise ratios of the penetrances F_0 , F_1 , F_2 for all models in our affecteds-only analyses with ASPs and ASTs. Generally, the variability (as measured by MAD in our case) for penetrance ratios is expected to be higher than for the corresponding individual penetrances, especially when the expected penetrance ratio is high.

For recessive models and ASPs, the 3 penetrance ratios (F_1/F_0 ; F_2/F_0 ; F_2/F_1) were estimated with best accuracy for models R1, R2, and R3 with the larger disease allele frequency $P = 0.1$. The ratio between F_1 and F_0 , which equals 1 for all recessive models except model AFR2, was usually well recognized, whereas F_2/F_0 and F_2/F_1 were underestimated for models with $P = 0.01$. There was a clear improvement in the estimation of the penetrance ratios F_2/F_0 and F_2/F_1 when using ASTs for the models with disease allele frequency $P = 0.01$ and the AFR1 model. Only the models R1, R2, and R3, each with $P = 0.1$, as well as AFR1 showed a smaller bias than MAD for both ASPs and ASTs and for all penetrance ratios. While bias of penetrance ratios often decreased when using ASTs instead of ASPs, the corresponding MAD was often higher, especially for models with $P = 0.01$ (Table 4).

For dominant models, the penetrance ratio that was close to 1, i.e., F_2/F_1 , was overestimated for ASPs, albeit only slightly for models D1–D3 with $P = 0.1$ and D1 with $P = 0.01$. The ratios F_1/F_0 and F_2/F_0 were mostly underestimated for models D1–D3 with $P = 0.1$ and AFD2, or mostly overestimated for models D1–D3 with $P = 0.01$ and AFD1. The estimation of ratios improved with ASTs compared to ASPs only for models D1–D3 with $P = 0.01$. In the case of ASPs, bias was smaller than MAD for all penetrance ratios and models, except for AFD1n1000 and AFD2. For ASTs, in addition to AFD1n1000 and AFD2,

higher bias than MAD was also obtained for models D2 and D3, each with $P = 0.1$.

For additive models, penetrance ratios were estimated best for models AFA1 and A1, which were strictly additive or close to strictly additive, respectively. In general, the benefit for the accuracy of the estimation of penetrance ratios when using ASTs instead of ASPs was not as clear-cut as for the other models. Here, the estimation mostly improved for one ratio and deteriorated for another one. For ASPs and ASTs, bias was smaller than MAD for all penetrance ratios and models, except for models A2 and AFA2, and, in the case of ASTs, model A3 with $P = 0.1$.

Results for the overdominant models and ASPs showed that the penetrance ratio F_1/F_0 was underestimated for models U1 and U2 with $P = 0.1$ as well as model U4 with $P = 0.35$, and overestimated for models with $P = 0.01$. The other 2 ratios, F_2/F_0 and F_2/F_1 , were always overestimated, even to a higher degree for models with $P = 0.01$. The penetrance ratios for model U4 could not be estimated accurately, for neither ASPs nor ASTs, due to the confounding of over- and underdominant models, as explained above. In most other cases, there was a clear improvement in estimation accuracy of all 3 penetrance ratios when using ASTs compared to ASPs. For both pedigree types, bias was mostly smaller than MAD for all penetrance ratios and models.

Summary of Trait-Model Parameter Estimation

Results

The results are summarized as answers to questions (1)–(5) given in the Introduction section.

(1) The ability of the MOD score approach to differentiate between the trait-model types (recessive, dominant, additive, and overdominant) was limited by the underlying parameter spaces of the corresponding pedigrees in the analysis. Among the recessive models, a stronger genetic effect provided a better discrimination from other model types across all sorts of investigated pedigrees. Adding one unaffected individual to an ASP pedigree was mostly sufficient to identify and correctly estimate the parameters of the recessive model. Additive and dominant models were generally hard to discriminate using affecteds-only data due to their spatial proximity in the corresponding allele-sharing parameter space. The discrimination between additive and dominant models improved by adding unaffected individuals and when using 3-G pedigrees. The correct classification of overdominant models substantially improved from ASPs to ASTs. With 3-G pedigrees, trait-model parameters of overdominant

Table 4. Estimation of penetrance ratios for ASPs and ASTs

Model name	P	F_1/F_0	Estimated median f_1/f_0 (MAD)		F_2/F_0	Estimated median f_2/f_0 (MAD)		F_2/F_1	Estimated median f_2/f_1 (MAD)	
			ASP	AST		ASP	AST		ASP	AST
R1	0.1	1	1.0 (0.74)	1.0 (0.66)	20	21.67 (3.95)	20.27 (4.84)	20	20.22 (9.54)	20.89 (8.85)
R1	0.01	1	1.0 (0.69)	0.78 (1.15)	20	1.09 (1.55)	19.74 (4.75)	20	1.0 (0.8)	18.6 (26.71)
R1n1000	0.01	1	1.0 (0.67)	0.76 (1.13)	20	1.0 (1.36)	19.8 (2.11)	20	1.16 (0.89)	21.33 (22.36)
R2	0.1	1	1.0 (0.37)	1.0 (0.89)	50	51.0 (8.9)	51.25 (18.16)	50	51.0 (17.05)	51.0 (42.4)
R2	0.01	1	1.59 (1.62)	1.0 (0.74)	50	8.0 (10.41)	50.0 (2.97)	50	2.0 (2.85)	51.0 (26.19)
R3	0.1	1	1.0 (0.74)	1.6 (1.85)	80	90.0 (29.65)	82.0 (34.1)	80	83.67 (40.46)	60.33 (61.28)
R3	0.01	1	1.6 (2.19)	1.0 (0.74)	80	2.5 (3.35)	83.0 (10.38)	80	6.11 (8.73)	80.0 (34.59)
AFR1	0.2	1	2.2 (2.92)	0.75 (0.98)	5	9.8 (13.05)	4.88 (6.85)	5	4.13 (5.11)	5.33 (2.69)
AFR1n1000	0.2	1	2.22 (2.97)	0.67 (0.83)	5	10.44 (14.0)	4.5 (6.33)	5	3.55 (4.33)	5.0 (2.58)
AFR2	0.25	16.67	10.0 (5.93)	6.0 (2.97)	166.67	65.67 (36.08)	33.33 (13.1)	10	8.41 (2.05)	5.6 (0.89)
D1	0.1	20	16.8 (8.43)	15.37 (9.18)	20	20.0 (15.12)	17.45 (17.58)	1	1.14 (1.14)	0.98 (0.79)
D1	0.01	20	47.0 (43.88)	22.5 (5.19)	20	50.0 (64.2)	22.22 (28.66)	1	1.14 (1.24)	1.01 (0.44)
D1n1000	0.01	20	50.0 (47.15)	20.67 (2.35)	20	62.71 (73.82)	20.0 (14.66)	1	1.11 (1.2)	1.01 (0.19)
D2	0.1	50	31.0 (29.65)	22.22 (19.01)	50	30.83 (35.63)	22.22 (31.46)	1	1.09 (1.25)	0.86 (0.74)
D2	0.01	50	95.0 (66.72)	59.0 (21.87)	50	113.13 (167.72)	62.75 (92.22)	1	1.96 (1.42)	1.1 (1.58)
D3	0.1	80	50.0 (49.42)	27.71 (27.15)	80	49.0 (66.42)	22.88 (33.47)	1	1.08 (1.31)	0.84 (0.71)
D3	0.01	80	95.0 (44.48)	84.0 (38.55)	80	113.75 (168.65)	70.0 (103.78)	1	1.91 (1.85)	1.01 (1.49)
AFD1	0.05	5	5.0 (4.45)	5.5 (1.73)	5	10.0 (10.04)	10.0 (5.93)	1	1.88 (0.88)	1.68 (1.01)
AFD1n1000	0.05	5	5.0 (2.8)	5.33 (0.99)	5	9.0 (5.07)	9.33 (3.95)	1	1.82 (0.74)	1.55 (0.84)
AFD2	0.25	166.67	8.83 (5.07)	4.75 (2.8)	166.67	14.92 (18.16)	3.0 (4.45)	1	1.9 (2.0)	1.17 (1.68)
A1	0.1	10	10.2 (1.19)	10.5 (4.98)	20	18.0 (5.19)	19.53 (10.3)	2	1.88 (0.8)	1.84 (0.88)
A1	0.01	10	9.4 (3.11)	12.35 (4.97)	20	17.0 (11.86)	22.22 (29.32)	2	1.88 (0.76)	1.72 (0.75)
A1n1000	0.01	10	10.0 (1.98)	10.8 (1.78)	20	18.2 (13.39)	20.0 (25.87)	2	1.89 (1.32)	1.77 (0.55)
A2	0.1	20	18.8 (10.91)	10.47 (4.16)	50	20.0 (19.71)	22.22 (19.45)	2.5	1.11 (1.14)	2.13 (0.95)
A2	0.01	20	49.0 (46.26)	21.67 (4.02)	50	56.0 (71.17)	22.22 (13.19)	2.5	1.19 (1.17)	1.03 (0.19)
A3	0.1	50	39.67 (33.31)	18.8 (15.12)	80	47.0 (56.19)	24.19 (32.3)	1.6	1.12 (1.24)	1.41 (0.8)
A3	0.01	50	92.75 (61.9)	55.0 (18.29)	80	111.25 (164.94)	81.0 (97.11)	1.6	1.96 (1.43)	1.44 (1.69)
AFA1	0.1	4.33	5.0 (3.9)	4.89 (1.91)	7.67	10.0 (9.14)	8.89 (5.35)	1.77	1.89 (0.73)	1.83 (0.99)
AFAn1000	0.1	4.33	5.0 (2.67)	4.5 (1.07)	7.67	9.0 (3.56)	8.4 (3.56)	1.77	1.8 (0.36)	1.89 (0.87)
AFAn2	0.5	83.33	5.0 (4.94)	8.58 (12.19)	166.67	10.0 (10.91)	24.54 (36.38)	2	1.85 (0.67)	1.78 (1.16)
U1	0.1	20	16.0 (10.5)	25.0 (17.3)	1	15.0 (21.99)	1.19 (1.66)	0.05	1.05 (1.56)	0.07 (0.1)
U1	0.01	20	46.0 (43.0)	23.5 (6.67)	1	49.0 (61.68)	22.22 (30.87)	0.05	1.14 (1.24)	0.75 (0.51)
U2	0.1	50	40.0 (50.41)	118.75 (146.41)	1	29.33 (39.78)	1.0 (1.48)	0.02	0.59 (0.87)	0.01 (0.01)
U2	0.01	50	99.17 (72.15)	71.67 (33.36)	1	113.75 (168.65)	17.75 (26.32)	0.02	1.96 (2.09)	0.27 (0.4)
U3	0.1	80	80.0 (107.49)	272.5 (375.99)	1	33.33 (47.94)	1.0 (1.48)	0.0125	0.38 (0.57)	0.0 (0.0)
U3	0.01	80	98.0 (47.07)	112.5 (67.95)	1	116.25 (172.35)	11.13 (16.49)	0.0125	1.92 (2.85)	0.16 (0.24)
U3n1000	0.01	80	90.0 (33.36)	100.0 (37.07)	1	100.0 (148.26)	2.75 (4.08)	0.0125	1.16 (1.72)	0.03 (0.04)
U4	0.35	90	1.67 (2.4)	0.19 (0.28)	1	16.08 (23.08)	1.2 (1.16)	0.0111	13.33 (19.64)	0.05 (11.78)

Ratios with both the numerator and denominator being exactly 0 were set to 1, ratios with only the denominator being exactly 0 were set to the arbitrarily chosen high number 10^6 to include their information for the calculation of the median.

MAD, median absolute deviation, adjusted by a constant (1.4826) for asymptotically normal consistency; P , true value for the disease allele frequency; ASP, affected sib pair; AST, affected sib triplet; f_1/f_0 , estimated penetrance ratio; F_1/F_0 , true value for penetrance ratio.

Table 5. Results of the imprinting models

Model name	P	F	GHM analysis with imprinting						GHM analysis without imprinting						
			Estimated median parameters (MAD)						Estimated median parameters (MAD)						
			MOD	p	f ₀	f _{1,pat}	f _{1,mat}	f ₂	i	MOD	p	f ₀	f ₁	f ₂	
cpi (I = -1)	0.01	0; 0; 1; 1	ped 1	26.0 (2.66)	0.008 (0.0003)	0.0002 (0.00003)	0.0003 (0.00004)	0.94 (0.089)	0.94 (0.089)	-1 (0.0)	9.26 (2.42)	0.1 (0.0)	0.01 (0.0)	0.54 (0.615)	0.5 (0.638)
cpi (I = -1)	0.01	0; 0; 1; 1	ped 2	39.61 (2.81)	0.003 (0.0004)	0.0 (0.0)	0.95 (0.074)	0.555 (0.645)	-0.1 (1.28)	39.49 (2.93)	0.002 (0.0)	0.0001 (0.0001)	0.43 (0.163)	1.0 (0.0)	
cmi (I = 1)	0.01	0; 1; 0; 1	ped 3	24.17 (2.63)	0.008 (0.0003)	0.0003 (0.00004)	0.9 (0.148)	0.001 (0.002)	0.7 (0.445)	0.93 (0.13)	18.81 (2.78)	0.05 (0.01)	0.0003 (0.0004)	0.12 (0.059)	0.51 (0.415)
ni	0.01	0; 0.5; 0.5; 1	ped 1	43.89 (4.0)	0.01 (0.0)	0.0002 (0.0003)	0.71 (0.423)	0.76 (0.356)	0.94 (0.089)	0	43.86 (3.99)	0.01 (0.0)	0.0 (0.0)	0.2 (0.274)	0.9 (0.148)
			ped 2	34.45 (3.48)	0.008 (0.0003)	0.0005 (0.00007)	0.38 (0.561)	0.43 (0.415)	0.92 (0.119)	0	34.35 (3.4)	0.008 (0.0)	0.0005 (0.0004)	0.48 (0.356)	0.83 (0.252)
			ped 3	27.06 (3.17)	0.008 (0.0003)	0.0004 (0.00006)	0.001 (0.001)	0.44 (0.474)	0.67 (0.489)	-0.035 (0.85)	26.89 (3.12)	0.008 (0.01)	0.0003 (0.0004)	0.48 (0.334)	0.85 (0.222)

cpi, complete paternal imprinting; cmi, complete maternal imprinting; ni, no imprinting; MAD, median absolute deviation, adjusted by a constant (1.4826) for asymptotically normal consistency; I, true value for the imprinting index; P, true value for the disease allele frequency; F, true values for the penetrances (F₀; F_{1,pat}; F_{1,mat}; F₂); ped, pedigree structure; MOD, MOD score; p, estimated disease allele frequency; f₀, estimated penetrances; i, estimated imprinting index.

models were mostly estimated with good accuracy, whereas DST and DSQ data sometimes showed larger bias than MAD for specific parameters.

(2) As was expected, the estimation of trait-model parameters and penetrance ratios improved when adding an affected sibling to an ASP, resulting in an AST. The identifiability of the trait-model type depended on the true point of allele-sharing in the corresponding parameter space. The parameter space for ASPs is the possible triangle, whereas the parameter space for ASTs has not been graphically depicted so far. However, using the formulas given by Knapp [28], we were able to empirically draw the parameter space for ASTs (Fig. 1), and hence to hypothesize which model types could be better discriminated using ASTs compared to ASPs. As was expected from the structure of the parameter spaces for both pedigree types, estimation accuracy using ASTs was particularly higher for overdominant models compared to ASPs. Discrimination of additive and dominant models, especially when the genetic effect was small to moderate, remained difficult. Recessive models were generally identified as such using either ASPs or ASTs due to their clear spatial separation in the parameter space from other model types.

(3) In line with our expectations, the identifiability of absolute values of the penetrances instead of pairwise ratios could be achieved when unaffected pedigree members were included in the analysis, i.e., DSTs and DSQs as well as 3-G pedigrees.

(4) Interestingly, the identifiability of trait-model parameters was only slightly better when adding a further unaffected sibling to DSTs, i.e., when using DSQs. The number of allele-sharing classes of DSTs hence seemed to be sufficient for the identification of the trait-model parameters.

(5) With more complex pedigrees, the identifiability of trait-model parameters further improved for some models. While the median estimates were mostly similar, using 3-G pedigrees instead of DSTs or DSQs often led to a reduction in MAD of the parameters.

Imprinting Models

The results of the imprinting scenarios can be found in Table 5. All parental genotypes were removed for both AHSPs and ASPs prior to the analysis.

NI Model

Using pedigree structure 1, i.e., AHSPs with one half of the sample having a common father and the other half having a common mother, the disease allele fre-

quency P and the penetrance F_0 were estimated with high accuracy for the ni model in a MOD score analysis without taking imprinting into account. However, penetrances F_1 and F_2 were both underestimated, with more downward bias for F_1 . It is of note that only 1 free parameter can in principle be identified from AHSP data in a MOD score analysis. In the case of the corresponding analysis taking imprinting into account, P and F_0 were estimated with high accuracy. Penetrance F_2 was estimated close to its expected value; however, the heterozygote penetrances were both clearly overestimated. The median values for the heterozygote penetrances $F_{1,pat}$ and $F_{1,mat}$ were comparable, which was expected for the ni model. The correct imprinting index $I = 0$ was obtained in the analysis of pedigree structure 1 and the ni model. MOD scores were comparable between the 2 analyses, i.e., with and without taking imprinting into account, whereby the imprinting MOD score is per definition always as large as the corresponding ni score. In the case of the ni model, MOD scores were generally highest using pedigree structure 1 and lowest for pedigree structure 3.

Using pedigree structure 2, i.e., 100 ASPs and 100 AHSPs having a common mother, the estimated median disease allele frequency P and penetrances F_0 and F_1 were close to the expected value in the analysis without taking imprinting into account. Penetrance F_2 was underestimated. In the case of the analysis taking imprinting into account, P and F_0 were estimated close to the expected value, whereas the heterozygote penetrances $F_{1,pat}$ and $F_{1,mat}$ as well as F_2 were underestimated. As was with pedigree structure 1, the correct imprinting index $I = 0$ could be obtained from the analysis of pedigree structure 2. MOD scores of both analysis types were comparable.

The corresponding values for the trait-model parameters for pedigree structure 3, i.e., 180 ASPs and 20 AHSPs with a common mother, were comparable to those of pedigree structure 2 for the ni analysis. When imprinting was taken into account in the analysis, penetrances $F_{1,pat}$ and F_2 were estimated lower ($f_{1,pat} = 0.001$; $f_2 = 0.67$) compared to pedigree structure 2 ($f_{1,pat} = 0.38$; $f_2 = 0.92$). Most strikingly, penetrance $F_{1,pat}$ was estimated close to 0, which reflects the unidentifiability between paternal imprinting and ni models when parental genotypes have been removed. It appears counterintuitive at first sight that an apparently stronger indication of paternal imprinting is obtained for pedigree structure 3, which contains only 20 AHSPs, compared to pedigree structure 2, which contains 100 AHSPs (Table 5). However, with a larger number of AHSPs in pedigree structure 2, it is more likely that 2 half-

sibs have received the disease allele from the 2 separate fathers rather than from their joint mother, which reduces the likelihood of a paternal imprinting model. The estimated median imprinting index was estimated close to its expected value, albeit slightly below 0 due to the underestimation of $F_{1,pat}$.

Imprinting Models

In contrast to the ni model, the presentation of the results for the imprinting simulations starts with the MOD scores taking imprinting into account, which are then compared to the ni results. Using pedigree structure 1 and the cpi model, the disease allele frequency and the penetrances were estimated with good accuracy in a MOD score analysis taking imprinting into account. The correct imprinting index $I = -1$ could be obtained as well. With regard to the corresponding ni analysis, the median estimated trait-model parameters of the cpi model were difficult to interpret due to the following: since the ni MOD score analysis assumes the equivalence of parental genomes, i.e., the equivalence of AHSPs having a common father and AHSPs having a common mother, this leads to a reduced likelihood and to bias of trait-model parameter estimates. This is because the truly underlying genetic mechanism, i.e., the imparity of parental genomes, is misspecified in a ni MOD score analysis, which cannot be compensated by maximizing over the ni trait model. If complete imprinting is really present but not modelled in the analysis, only the meioses of those AHSPs with a common parent of the non-imprinted sex contribute linkage information, whereas the other AHSPs point at no linkage. Therefore, the MOD score dropped from 26.0 with imprinting to 9.26 without imprinting taken into account in the analysis, and trait-model parameter estimates for the ni model were distorted.

Using pedigree structure 2, trait-model parameters could be estimated with good accuracy in an imprinting MOD score analysis, except for F_2 , which was clearly underestimated. In fact, F_2 was mostly estimated as either 0 or close to 1 (data not shown). This was most likely due to the fact that a paternal imprinting model with penetrances $(F_0, F_{1,pat}, F_{1,mat}, F_2) = (0;0;1;1)$ can hardly be distinguished from an overdominant model with penetrances $(0;0;1;0)$ using ASP data. This was also reflected by a median imprinting index with a smaller absolute value than expected ($i = -0.1$), because i is defined to be 0 if the estimates of F_0 and F_2 are equal, and a high MAD for the F_2 penetrance (0.645; Table 5). Owing to the AHSPs with a common mother, however, the relation $F_{1,pat} \ll F_{1,mat}$ could mostly be determined. With regard to the corre-

sponding ni analysis, trait-model parameters were estimated with good accuracy, whereby the median heterozygote penetrance f_1 was estimated close to the mean of the penetrances $F_{1,pat}$ and $F_{1,mat}$ that were used for the cpi model simulation. MOD scores of the ni analysis were comparable to those of the imprinting analysis for pedigree structure 2, because assuming strong maternal allele sharing is almost as likely as an additive model, for which allele sharing can take place through parents of both sexes. In other words, maternal allele sharing in AHSPs with a common mother does not imply random (non-excess) paternal allele sharing in ASPs with untyped parents.

Using pedigree structure 3 and the cmi model, the combined sample of 180 ASPs and 20 AHSPs having a common mother led to trait-model parameter estimates reflecting maternal imprinting, albeit with an underestimation of $F_{1,pat}$ and F_2 ($f_{1,pat} = 0.9$; $f_2 = 0.7$). The reason why F_2 was underestimated is the same as it was for pedigree structure 2. In contrast to pedigree structure 2, the imprinting analysis yielded substantially higher MOD scores than the ni analysis, because the non-excess allele sharing of AHSPs with a common mother can only be explained by maternal imprinting, whereas for the ni analysis the non-excess sharing of maternal alleles in AHSPs reduces linkage information. This goes along with distorted trait-model parameter estimates for the combined dataset. The imprinting index I was estimated close to its expected value reflecting maternal imprinting.

Summary of Imprinting Results

The imprinting results are summarized as an answer to question (6) given in the Introduction section.

Imprinting could reliably be detected in samples that include AHSPs having a common father as well as AHSPs with a common mother, even if the parents are untyped (pedigree structure 1). When analyzing an equal mixture of ASPs and AHSPs having a common mother, all with untyped parents, imprinting could in part be declared when looking at the imprinting index I obtained from the imprinting MOD score analysis and the cpi model. However, the difference between the ni and imprinting MOD score seemed to be marginal, such that there was no significant evidence for imprinting. However, using 180 ASPs and 20 AHSPs having a common mother, again with untyped parents, the results for the cmi model clearly showed that information on imprinting can be extracted when adding a few AHSPs with a common parent of the imprinted sex to a sample of ASPs with untyped parents, which only harbor information on linkage, to obtain substantial evidence of imprinting.

Discussion

The ability of a pedigree analysis to estimate parameters of trait inheritance has been extensively discussed in the literature [1, 11, 16, 19–23]. More specifically, the possibility to jointly estimate linkage and segregation parameters in a MOD score analysis has been debated. A MOD score analysis does not perform classical segregation analysis in the sense of determining whether or not there is major gene segregation, but it estimates some segregation-model parameters together with parameters for linkage, which we denote joint trait-marker inheritance parameters (recombination fraction, LD parameters, and trait-model parameters: disease allele frequency and penetrances). Since the publication of the AAF method proposed by Ewens and Shute [17], the MOD score has often been referred to as being AAF, such that it delivers asymptotically unbiased estimates of the trait-model parameters [11]. It is of note that estimates obtained from maximum likelihood techniques are naturally biased for finite sample sizes. However, the problem of ascertainment or sampling was often neglected and most theoretical work on parameter estimation in linkage analysis assumed what is called PI sampling, i.e., sampling of fixed pedigree structures independent of any proband. Hence, if no correction of the likelihood as to the ascertainment procedure is applied, the estimates of the joint marker-trait inheritance parameters will be biased.

Over the years, the problem of ascertainment/sampling for linkage analysis was gradually elaborated [1, 16, 19–23]. Presumably the most comprehensive and most detailed work on these aspects of pedigree analysis is the book by Ginsburg et al. [1], who claimed that unbiased estimates can in fact be obtained from a pedigree analysis (see also Ginsburg et al. [16]). They provided a general likelihood framework that can be used to accommodate the likelihood for many aspects of the sampling procedure and also showed how to accomplish sampling correction in practice. Although their focus was not on the MOD score approach per se, they provided the above-mentioned conditions (i)–(vii), under which the MOD score delivers asymptotically unbiased parameter estimates. Along these lines, one of the goals of the present paper was to investigate the ability of a MOD score analysis to obtain unbiased trait-model parameter estimates in practical situations. To this end, we have thoroughly recapitulated the theoretical background, including conditions under which the parameter estimates should be asymptotically unbiased. We then evaluated the parameter estimation performance of a MOD score analysis in

a simulation study. The first condition of correctly specifying the mode of inheritance referring to the number of loci and the number of alleles at each locus is presumably most crucial. Therefore, a diallelic autosomal binary trait locus was used for the simulation of pedigree data, which is usually assumed as the mode of inheritance in a MOD score analysis. Although complex disorders are expected to follow more complicated modes of inheritance, e.g., involving a larger number of trait loci, the number of possible models, i.e., degrees of freedom, to be tested in a MOD score analysis would be prohibitively large and procedures to avoid inflated type I error rates would presumably diminish power. The second condition of marker-independent sampling is often ignored in practice. However, when performing linkage analysis in the era of densely available markers, this assumption is likely to hold, since the limiting step is often the recruitment of an individual rather than obtaining informative genotypes. Conditions (iii)–(vi) refer to the sampling procedure, which is often assumed to be PI. Admittedly, only few linkage studies are really PI. However, even in this case, parameter estimation remains free of bias if the sampling procedure can be controlled. This is the case if either all members of the PSF have measured trait values (see condition [v]), e.g., by using a questionnaire to include information on potential probands not sampled (see Ginsburg et al. [16]), or sampling is single in the sense of Hodge and Vieland [20] (see condition [vi]), and the model of extension is random (see condition [iv]). Then, the MOD score can readily be used to obtain asymptotically unbiased joint marker-trait inheritance parameter estimates.

Previous work using simulated pedigree data has shown that the maximum LOD score is obtained for the truly underlying genetic model, provided that there is enough power to detect linkage [43]. However, the focus of the aforementioned work was only on strictly dominant ($f_1 = f_2$) and strictly recessive models ($f_1 = 0$) without phenocopies ($f_0 = 0$) and with the disease allele frequency fixed at the true value for the analysis. In addition, maximization was done using a limited set of penetrance values [43]. In our simulation study, the MOD score with a more exhaustive maximization as implemented in GHM was used. Furthermore, we studied a wider range of trait models and pedigree structures. We did not investigate the ability of the MOD score to estimate the recombination fraction and any LD parameters. The recombination fraction is confounded with the trait-model parameters, i.e., with the disease allele frequency p and the 3 penetrances f_0, f_1 , and f_2 , and was hence excluded from the estimation, but rather fixed at the true value of $\theta = 0$. Oth-

erwise, it would not be possible to distinguish confounding of parameters and bias from each other. In the current program version of GHM, LD is not modelled. As stated earlier, to obtain unbiased trait-model parameter estimates, LD between markers and disease locus must in fact be absent, otherwise sampling is no longer marker independent. As noted by Malkin and Elston [19], such a situation is unlikely when using marker panels of densely spaced single nucleotide polymorphisms. However, selective inclusion of only a subset of markers can ensure linkage equilibrium at least between these markers, while still retaining sufficient information for linkage analysis. With such a sparser set of markers, it is also less likely that one of them is in LD with a disease allele. If LD between marker and disease alleles happens to be present, the expected bias in parameter estimates is so far unknown. Further, we did not consider bias of trait-model parameters due to gene-environment interactions, which are usually not controlled in a linkage analysis. In addition, we did not investigate the ascertainment or sampling bias that may occur when recruiting families in practice. Still, the problem of ascertainment or sampling for linkage analysis with estimation of joint trait-marker inheritance parameters has been thoroughly reviewed and discussed in the Introduction section.

Another aspect of estimating trait-model parameters is their identifiability. It has been shown by Strauch [10] that the identifiability of trait-model parameters depends on the truly underlying number of allele-sharing classes. In addition, only penetrance ratios can be estimated from affecteds-only data. The identifiability is expected to increase with larger sibships or more complex pedigrees. Therefore, we were interested in the degree to which the identifiability of trait-model parameters increases when adding affected or unaffected siblings to an ASP or when analyzing a 3-G pedigree.

In this study, we were able to show how trait-model parameters can in principle be estimated in a MOD score linkage analysis and to what extent the identifiability depends on the pedigree types in the dataset. Our findings can provide guidance to researchers aiming to estimate parameters by a MOD score linkage analysis using family data. Parameter estimation generally showed smaller bias and MAD with increasing pedigree complexity for all investigated model types. Identifiability of trait-model parameters increased with (a) more affected siblings in an affecteds-only analysis of nuclear families, although only ratios of parameter values can be identified in this case, (b) adding unaffected siblings to nuclear families, and for some models with (c) adding a generation (3-G pedi-

gress). Penetrance estimation performance was substantially affected by confounding of the trait-model parameters in terms of their proximity or identity in the corresponding nonparametric allele-sharing parameter space. This is equivalent to the more “parametric” notion that the degree of information to accurately estimate parameters given their identifiability still depends on the proportions of disease locus genotypes that are induced by the number of affected and unaffected individuals in a pedigree, together with the truly underlying trait-model parameters. Therefore, especially additive and dominant models can hardly be distinguished, even when analyzing more complex pedigrees. A sufficient number of pedigrees in the sample is a further prerequisite to be able to actually estimate the parameters in practice, according to the identifiability that is theoretically possible with a certain pedigree type. Furthermore, we have shown under which scenarios imprinting can be detected even if all parents have missing genotypes. Imprinting could reliably be estimated in terms of the imprinting index I [35] with the datasets containing both AHSPs having a common father as well as a common mother. We were also able to show that it is possible to combine pure linkage information from ASPs with imprinting-sensitive linkage information from AHSPs having a common mother to obtain substantial evidence for maternal imprinting. This finding indicates that adding AHSPs with a common parent of the imprinted sex draws the trait-model parameter

estimates of the combined ASP/AHSP sample towards the truly underlying imprinting model.

In essence, asymptotically unbiased parameter estimates can be obtained from a MOD score analysis, given that certain conditions are satisfied ([i]–[vii], see Introduction section). In most real-life situations, these conditions can hardly be fulfilled. The extent to which a violation of any of these conditions or a combination of them causes bias is unclear and demands further investigations. Such a subsequent simulation study might reveal situations in which, despite, for example, an incorrect sampling model, the parameter estimates obtained from the analysis are essentially correct, which has been referred to as the “man bites dog” criterion [11]. Along these lines, the results of our present study are an important prerequisite for future investigations on robustness of MOD score-based parameter estimation under various sampling schemes.

Acknowledgements

This work was supported by grants Str643/4-1 and Str643/6-1 of the Deutsche Forschungsgemeinschaft (German Research Foundation). Further, this research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. In addition, we greatly appreciate the reviewers’ thoughtful comments, which have helped to improve the paper.

Appendix

Table A1. Estimation of trait-model parameters for ASPs and ASTs (recessive models)

Model name	P	Estimated median p (MAD)		F_0	Estimated median f_0 (MAD)		F_1	Estimated median f_1 (MAD)		F_2	Estimated median f_2 (MAD)		D	Estimated median d (MAD)	
		ASP	AST		ASP	AST		ASP	AST		ASP	AST		ASP	AST
R1	0.1	0.1 (0.01)	0.1 (0.03)	0.01	0.04 (0.007)	0.045 (0.022)	0.01	0.045 (0.022)	0.045 (0.022)	0.2	0.92 (0.074)	0.92 (0.07)	-1.0	-1.0 (0.06)	-1.0 (0.07)
R1	0.01	0.015 (0.02)	0.01 (0.0)	0.01	0.045 (0.052)	0.045 (0.007)	0.01	0.05 (0.059)	0.015 (0.022)	0.2	0.13 (0.178)	0.93 (0.074)	-1.0	0.0 (0.0)	-1.05 (0.12)
R1n1000	0.01	0.05 (0.06)	0.01 (0.0)	0.01	0.05 (0.059)	0.045 (0.007)	0.01	0.08 (0.104)	0.02 (0.03)	0.2	0.14 (0.193)	0.94 (0.044)	-1.0	0.0 (0.06)	-1.03 (0.12)
R2	0.1	0.1 (0.01)	0.1 (0.01)	0.01	0.01 (0.0)	0.01 (0.003)	0.01	0.01 (0.003)	0.01 (0.007)	0.5	0.5 (0.044)	0.51 (0.089)	-1.0	-1.0 (0.01)	-1.0 (0.04)
R2	0.01	0.1 (0.14)	0.01 (0.0)	0.01	0.045 (0.052)	0.01 (0.0)	0.01	0.1 (0.136)	0.01 (0.007)	0.5	0.48 (0.341)	0.5 (0.03)	-1.0	-0.81 (1.2)	-1.0 (0.03)
R3	0.1	0.1 (0.01)	0.1 (0.0)	0.01	0.01 (0.003)	0.008 (0.003)	0.01	0.01 (0.007)	0.01 (0.013)	0.8	0.89 (0.111)	0.74 (0.326)	-1.0	-1.0 (0.01)	-0.99 (0.04)
R3	0.01	0.48 (0.09)	0.01 (0.0)	0.01	0.05 (0.059)	0.01 (0.0)	0.01	0.1 (0.078)	0.01 (0.007)	0.8	0.5 (0.593)	0.85 (0.089)	-1.0	-0.64 (0.95)	-1.0 (0.01)
AFR1	0.2	0.44 (0.22)	0.17 (0.18)	0.04	0.05 (0.059)	0.07 (0.067)	0.04	0.1 (0.059)	0.09 (0.044)	0.2	0.49 (0.474)	0.49 (0.104)	-1.0	-0.56 (0.82)	-0.80 (0.56)
AFR1n1000	0.2	0.45 (0.12)	0.21 (0.23)	0.04	0.05 (0.059)	0.08 (0.089)	0.04	0.11 (0.059)	0.09 (0.03)	0.2	0.5 (0.193)	0.48 (0.163)	-1.0	-0.55 (2.02)	-0.77 (0.64)
AFR2	0.25	0.07 (0.04)	0.1 (0.06)	0.003	0.01 (0.003)	0.015 (0.007)	0.05	0.09 (0.044)	0.1 (0.015)	0.5	0.57 (0.489)	0.52 (0.044)	-0.81	-0.79 (0.06)	-0.69 (0.07)

MAD, median absolute deviation, adjusted by a constant (1.4826) for asymptotically normal consistency; p , estimated value for the disease allele frequency; f_0, f_1, f_2 , estimated values for penetrances with f_i denoting the probability that an individual with i copies of the disease allele is affected by the disease; d , estimated value for the dominance index; P , true value for the disease allele frequency; F_0, F_1, F_2 , true values for the penetrances; D , true value for the dominance index; ASP, affected sib pair; AST, affected sib triplet.

Table A2. Estimation of trait-model parameters for ASPs and ASTs (dominant models)

Model name	P	Estimated median p (MAD)		F_0	Estimated median f_0 (MAD)		F_1	Estimated median f_1 (MAD)		F_2	Estimated median f_2 (MAD)		D	Estimated median d (MAD)	
		ASP	AST		ASP	AST		ASP	AST		ASP	AST		ASP	AST
D1	0.1	0.1 (0.06)	0.1 (0.01)	0.01	0.05 (0.007)	0.045 (0.022)	0.2	0.63 (0.371)	0.555 (0.363)	0.2	0.74 (0.385)	0.655 (0.511)	1.0	0.39 (1.38)	0.85 (1.39)
D1	0.01	0.1 (0.0)	0.01 (0.01)	0.01	0.01 (0.014)	0.04 (0.015)	0.2	0.52 (0.415)	0.92 (0.074)	0.2	0.86 (0.208)	0.925 (0.111)	1.0	0.11 (1.03)	0.91 (1.01)
D1n1000	0.01	0.1 (0.0)	0.008 (0.01)	0.01	0.01 (0.013)	0.045 (0.007)	0.2	0.53 (0.445)	0.93 (0.044)	0.2	0.88 (0.178)	0.95 (0.074)	1.0	0.66 (0.98)	0.90 (0.39)
D2	0.1	0.1 (0.01)	0.1 (0.01)	0.01	0.01 (0.014)	0.015 (0.022)	0.5	0.565 (0.482)	0.61 (0.356)	0.5	0.57 (0.571)	0.53 (0.549)	1.0	0.66 (1.17)	0.96 (1.45)
D2	0.01	0.025 (0.03)	0.01 (0.0)	0.01	0.005 (0.006)	0.008 (0.003)	0.5	0.48 (0.119)	0.47 (0.133)	0.5	0.52 (0.638)	0.18 (0.267)	1.0	0.0 (0.86)	-0.04 (1.19)
D3	0.1	0.1 (0.0)	0.1 (0.01)	0.01	0.01 (0.013)	0.01 (0.015)	0.8	0.53 (0.474)	0.6 (0.371)	0.8	0.55 (0.593)	0.52 (0.578)	1.0	0.51 (1.28)	0.96 (1.44)
D3	0.01	0.01 (0.0)	0.01 (0.0)	0.01	0.004 (0.005)	0.006 (0.006)	0.8	0.49 (0.563)	0.49 (0.311)	0.8	0.50 (0.652)	0.26 (0.385)	1.0	0.01 (0.98)	-0.03 (1.37)
AFD1	0.05	0.1 (0.13)	0.11 (0.03)	0.04	0.025 (0.03)	0.015 (0.013)	0.2	0.1 (0.074)	0.12 (0.104)	0.2	0.33 (0.356)	0.18 (0.193)	1.0	0.0 (0.49)	0.0 (0.64)
AFD1n1000	0.05	0.1 (0.07)	0.11 (0.03)	0.04	0.04 (0.044)	0.015 (0.007)	0.2	0.1 (0.074)	0.1 (0.074)	0.2	0.23 (0.252)	0.155 (0.111)	1.0	0.0 (0.39)	0.18 (0.69)
AFD2	0.25	0.1 (0.13)	0.15 (0.1)	0.003	0.05 (0.059)	0.02 (0.03)	0.5	0.1 (0.059)	0.1 (0.104)	0.5	0.45 (0.534)	0.12 (0.177)	1.0	0.0 (1.09)	0.11 (2.25)

For more details, see Appendix Table A1.

Table A3. Estimation of trait-model parameters for ASPs and ASTs (additive models)

Model name	P	Estimated median p (MAD)		F_0	Estimated median f_0 (MAD)		F_1	Estimated median f_1 (MAD)		F_2	Estimated median f_2 (MAD)		D	Estimated median d (MAD)	
		ASP	AST		ASP	AST		ASP	AST		ASP	AST		ASP	AST
A1	0.1	0.1 (0.04)	0.1 (0.01)	0.01	0.05 (0.007)	0.04 (0.03)	0.1	0.49 (0.044)	0.42 (0.163)	0.2	0.875 (0.185)	0.91 (0.129)	-0.05	0.0 (0.67)	0.02 (0.47)
A1	0.01	0.05 (0.07)	0.05 (0.06)	0.01	0.045 (0.052)	0.025 (0.034)	0.1	0.11 (0.089)	0.49 (0.133)	0.2	0.47 (0.563)	0.93 (0.104)	-0.05	0.0 (0.62)	0.03 (0.47)
A1n1000	0.01	0.01 (0.01)	0.03 (0.03)	0.01	0.01 (0.015)	0.015 (0.022)	0.1	0.1 (0.03)	0.49 (0.074)	0.2	0.45 (0.519)	0.94 (0.089)	-0.05	0.0 (0.67)	-0.01 (0.33)
A2	0.1	0.09 (0.04)	0.07 (0.04)	0.01	0.045 (0.007)	0.04 (0.03)	0.2	0.79 (0.297)	0.45 (0.119)	0.5	0.87 (0.193)	0.95 (0.074)	-0.22	0.66 (0.99)	-0.12 (0.31)
A2	0.01	0.1 (0.0)	0.008 (0.01)	0.01	0.01 (0.013)	0.045 (0.007)	0.2	0.50 (0.267)	0.93 (0.059)	0.5	0.89 (0.163)	0.96 (0.059)	-0.22	0.01 (0.99)	0.89 (0.39)
A3	0.1	0.1 (0.01)	0.1 (0.01)	0.01	0.01 (0.014)	0.035 (0.037)	0.5	0.55 (0.43)	0.50 (0.133)	0.8	0.86 (0.208)	0.87 (0.193)	0.24	0.54 (0.82)	0.30 (0.67)
A3	0.01	0.02 (0.02)	0.01 (0.0)	0.01	0.005 (0.006)	0.008 (0.003)	0.5	0.48 (0.133)	0.44 (0.178)	0.8	0.52 (0.638)	0.19 (0.282)	0.24	0.0 (0.85)	-0.07 (0.8)
AFA1	0.1	0.1 (0.09)	0.11 (0.09)	0.03	0.023 (0.033)	0.02 (0.03)	0.13	0.11 (0.089)	0.16 (0.089)	0.23	0.325 (0.348)	0.32 (0.282)	0.0	0.0 (0.38)	-0.08 (0.55)
AFA1n1000	0.1	0.1 (0.07)	0.11 (0.06)	0.03	0.015 (0.022)	0.02 (0.03)	0.13	0.1 (0.074)	0.15 (0.074)	0.23	0.235 (0.252)	0.28 (0.267)	0.0	0.0 (0.2)	-0.08 (0.46)
AFA2	0.5	0.14 (0.2)	0.415 (0.14)	0.003	0.04 (0.044)	0.01 (0.015)	0.25	0.15 (0.148)	0.45 (0.423)	0.5	0.46 (0.534)	0.85 (0.219)	<0.01	0.0 (0.1)	0.0 (0.68)

For more details, see Appendix Table A1.

Table A4. Estimation of trait-model parameters for ASPs and ASTs (overdominant models)

Model name	P	Estimated median p (MAD)		F_0	Estimated median f_0 (MAD)		F_1	Estimated median f_1 (MAD)		F_2	Estimated median f_2 (MAD)	
		ASP	AST		ASP	AST		ASP	AST		ASP	AST
U1	0.1	0.1 (0.04)	0.1 (0.01)	0.01	0.035 (0.03)	0.03 (0.03)	0.2	0.5 (0.623)	0.9 (0.147)	0.01	0.36 (0.474)	0.035 (0.052)
U1	0.01	0.1 (0.0)	0.01 (0.01)	0.01	0.01 (0.013)	0.04 (0.015)	0.2	0.51 (0.474)	0.92 (0.074)	0.01	0.825 (0.259)	0.5 (0.704)
U2	0.1	0.1 (0.01)	0.1 (0.01)	0.01	0.01 (0.01)	0.005 (0.007)	0.5	0.47 (0.563)	0.7 (0.385)	0.01	0.41 (0.46)	0.004 (0.006)
U2	0.01	0.025 (0.03)	0.01 (0.0)	0.01	0.004 (0.006)	0.008 (0.003)	0.5	0.48 (0.133)	0.49 (0.133)	0.01	0.50 (0.652)	0.1 (0.148)
U3	0.1	0.1 (0.01)	0.1 (0.01)	0.01	0.006 (0.006)	0.001 (0.001)	0.8	0.49 (0.578)	0.69 (0.385)	0.01	0.175 (0.259)	0.0 (0.0)
U3	0.01	0.01 (0.0)	0.01 (0.0)	0.01	0.003 (0.003)	0.002 (0.003)	0.8	0.47 (0.549)	0.48 (0.43)	0.01	0.50 (0.623)	0.04 (0.059)
U3n1000	0.01	0.01 (0.0)	0.01 (0.0)	0.01	0.006 (0.006)	0.006 (0.006)	0.8	0.58 (0.445)	0.58 (0.385)	0.01	0.44 (0.652)	0.007 (0.01)
U4	0.35	0.1 (0.14)	0.37 (0.1)	0.01	0.01 (0.015)	0.09 (0.133)	0.9	0.05 (0.073)	0.06 (0.089)	0.01	0.39 (0.563)	0.22 (0.326)

For more details, see Appendix Table A1.

Table A5. Estimation of trait-model parameters of DST and DSQ pedigrees (recessive models)

Model name	Estimated median p (MAD)		F_0	Estimated median f_0 (MAD)		F_1	Estimated median f_1 (MAD)		F_2	Estimated median f_2 (MAD)		D	Estimated median d (MAD)	
	DST	DSQ		DST	DSQ		DST	DSQ		DST	DSQ		DST	DSQ
R1	0.1	0.05 (0.01)	0.05 (0.04)	0.01	0.01 (0.007)	0.008 (0.01)	0.01	0.01 (0.015)	0.003 (0.004)	0.2	0.32 (0.163)	0.26 (0.222)	-1.0	-0.97 (0.06)
R1	0.01	0.043 (0.05)	0.045 (0.05)	0.01	0.043 (0.06)	0.09 (0.133)	0.01	0.05 (0.071)	0.1 (0.148)	0.2	0.07 (0.104)	0.195 (0.289)	-1.0	-0.07 (1.12)
R1n1000	0.01	0.05 (0.06)	0.05 (0.06)	0.01	0.04 (0.052)	0.085 (0.123)	0.01	0.05 (0.07)	0.09 (0.13)	0.2	0.07 (0.104)	0.19 (0.282)	-1.0	-0.51 (0.76)
R2	0.1	0.09 (0.03)	0.09 (0.03)	0.01	0.01 (0.0)	0.01 (0.0)	0.01	0.01 (0.007)	0.01 (0.007)	0.5	0.51 (0.074)	0.5 (0.059)	-1.0	-1.0 (0.03)
R2	0.01	0.08 (0.01)	0.05 (0.06)	0.01	0.025 (0.037)	0.03 (0.043)	0.01	0.05 (0.071)	0.045 (0.066)	0.5	0.44 (0.571)	0.46 (0.593)	-1.0	-0.99 (0.44)
R3	0.1	0.09 (0.03)	0.09 (0.03)	0.01	0.01 (0.003)	0.01 (0.003)	0.01	0.01 (0.007)	0.01 (0.007)	0.8	0.81 (0.044)	0.8 (0.03)	-1.0	-1.0 (0.03)
R3	0.01	0.06 (0.07)	0.05 (0.06)	0.01	0.04 (0.046)	0.035 (0.04)	0.01	0.04 (0.047)	0.035 (0.047)	0.8	0.8 (0.297)	0.82 (0.267)	-1.0	-1.0 (0.14)
AFR1	0.2	0.12 (0.16)	0.11 (0.15)	0.04	0.01 (0.015)	0.01 (0.015)	0.04	0.05 (0.062)	0.04 (0.055)	0.2	0.175 (0.256)	0.15 (0.218)	-1.0	-0.87 (0.76)
AFR1n1000	0.2	0.15 (0.21)	0.12 (0.16)	0.04	0.01 (0.015)	0.01 (0.015)	0.04	0.05 (0.059)	0.045 (0.052)	0.2	0.19 (0.27)	0.17 (0.23)	-1.0	-0.87 (0.68)
AFR2	0.25	0.15 (0.13)	0.22 (0.18)	0.003	0.006 (0.006)	0.005 (0.006)	0.05	0.06 (0.037)	0.05 (0.03)	0.5	0.53 (0.104)	0.52 (0.074)	-0.81	-0.81 (0.09)

For more details, see Appendix Table A1. DST, discordant sib triplet; DSQ, discordant sib quadruplet.

Table A6. Estimation of trait-model parameters of DST and DSQ pedigrees (dominant models)

Model name	Estimated median p (MAD)		F_0	Estimated median f_0 (MAD)		F_1	Estimated median f_1 (MAD)		F_2	Estimated median f_2 (MAD)		D	Estimated median d (MAD)	
	DST	DSQ		DST	DSQ		DST	DSQ		DST	DSQ		DST	DSQ
D1	0.1	0.1 (0.06)	0.11 (0.06)	0.01	0.008 (0.004)	0.008 (0.003)	0.2	0.12 (0.089)	0.13 (0.059)	0.2	0.31 (0.252)	0.33 (0.222)	1.0	-0.05 (0.8)
D1	0.01	0.1 (0.03)	0.11 (0.03)	0.01	0.002 (0.003)	0.005 (0.006)	0.2	0.14 (0.089)	0.14 (0.074)	0.2	0.31 (0.267)	0.34 (0.208)	1.0	-0.01 (0.53)
D1n1000	0.01	0.11 (0.01)	0.11 (0.03)	0.01	0.003 (0.003)	0.005 (0.006)	0.2	0.15 (0.059)	0.15 (0.059)	0.2	0.3 (0.208)	0.34 (0.163)	1.0	<0.01 (0.49)
D2	0.1	0.1 (0.01)	0.1 (0.03)	0.01	0.006 (0.007)	0.008 (0.01)	0.5	0.47 (0.089)	0.49 (0.059)	0.5	0.555 (0.378)	0.51 (0.341)	1.0	0.41 (1.12)
D2	0.01	0.045 (0.01)	0.05 (0.04)	0.01	0.003 (0.004)	0.006 (0.006)	0.5	0.48 (0.074)	0.48 (0.044)	0.5	0.74 (0.385)	0.69 (0.46)	1.0	0.05 (0.41)
D3	0.1	0.08 (0.04)	0.09 (0.06)	0.01	0.015 (0.016)	0.008 (0.01)	0.8	0.8 (0.044)	0.8 (0.03)	0.8	0.71 (0.43)	0.81 (0.222)	1.0	0.81 (0.91)
D3	0.01	0.02 (0.02)	0.025 (0.03)	0.01	0.008 (0.004)	0.008 (0.004)	0.8	0.8 (0.03)	0.8 (0.03)	0.8	1.00 (0.0)	0.92 (0.119)	1.0	0.62 (0.15)
AFD1	0.05	0.08 (0.11)	0.08 (0.11)	0.04	0.01 (0.014)	0.01 (0.014)	0.2	0.1 (0.082)	0.11 (0.096)	0.2	0.23 (0.282)	0.27 (0.334)	1.0	-0.11 (0.84)
AFD1n1000	0.05	0.09 (0.12)	0.07 (0.09)	0.04	0.01 (0.014)	0.01 (0.014)	0.2	0.11 (0.089)	0.12 (0.089)	0.2	0.27 (0.297)	0.28 (0.311)	1.0	-0.15 (0.73)
AFD2	0.25	0.17 (0.13)	0.16 (0.12)	0.003	0.035 (0.037)	0.03 (0.03)	0.5	0.47 (0.119)	0.49 (0.074)	0.5	0.48 (0.341)	0.41 (0.356)	1.0	0.28 (1.52)

For more details, see Appendix Tables A1 and A5.

Table A7. Estimation of trait-model parameters of DST and DSQ pedigrees (additive models)

Model name	Estimated median p (MAD)		Estimated median F_0 (MAD)		Estimated median f_0 (MAD)		Estimated median F_1 (MAD)		Estimated median f_1 (MAD)		Estimated median F_2 (MAD)		Estimated median f_2 (MAD)		Estimated median d (MAD)	
	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ
A1	0.1	0.06 (0.06)	0.06 (0.07)	0.01	0.01 (0.003)	0.01 (0.003)	0.1	0.1 (0.03)	0.1 (0.03)	0.1 (0.03)	0.2	0.17 (0.178)	0.17 (0.17)	-0.05	0.01 (0.56)	0.01 (0.89)
A1	0.01	0.01 (0.01)	0.01 (0.01)	0.01	0.01 (0.003)	0.01 (0.0)	0.1	0.09 (0.059)	0.09 (0.059)	0.2	0.15 (0.193)	0.17 (0.222)	-0.05	-0.12 (0.81)	-0.35 (0.7)	
A1n1000	0.01	0.01 (0.01)	0.01 (0.01)	0.01	0.01 (0.0)	0.01 (0.0)	0.1	0.1 (0.03)	0.1 (0.044)	0.2	0.165 (0.23)	0.17 (0.23)	-0.05	-0.30 (0.64)	-0.45 (0.67)	
A2	0.1	0.11 (0.04)	0.12 (0.03)	0.01	0.006 (0.006)	0.006 (0.006)	0.2	0.18 (0.104)	0.18 (0.089)	0.5	0.43 (0.208)	0.45 (0.156)	-0.22	-0.08 (0.53)	-0.12 (0.53)	
A2	0.01	0.1 (0.03)	0.11 (0.03)	0.01	0.002 (0.003)	0.004 (0.005)	0.2	0.15 (0.089)	0.15 (0.074)	0.5	0.335 (0.259)	0.35 (0.208)	-0.22	-0.01 (0.5)	-0.03 (0.56)	
A3	0.1	0.1 (0.01)	0.1 (0.03)	0.01	0.006 (0.006)	0.006 (0.009)	0.5	0.48 (0.089)	0.5 (0.059)	0.8	0.77 (0.319)	0.78 (0.282)	0.24	0.14 (0.59)	0.22 (0.47)	
A3	0.01	0.045 (0.01)	0.05 (0.04)	0.01	0.003 (0.004)	0.005 (0.007)	0.5	0.48 (0.059)	0.49 (0.044)	0.8	0.79 (0.311)	0.745 (0.378)	0.24	0.04 (0.37)	0.1 (0.43)	
AFA1	0.1	0.07 (0.09)	0.06 (0.08)	0.03	0.01 (0.012)	0.01 (0.007)	0.13	0.09 (0.074)	0.09 (0.074)	0.23	0.195 (0.259)	0.19 (0.248)	0.0	-0.08 (0.77)	-0.12 (0.82)	
AFA1n1000	0.1	0.08 (0.11)	0.06 (0.08)	0.03	0.01 (0.007)	0.01 (0.007)	0.13	0.09 (0.059)	0.1 (0.074)	0.23	0.17 (0.222)	0.17 (0.222)	0.0	-0.11 (0.65)	-0.18 (0.71)	
AFA2	0.5	0.1 (0.14)	0.05 (0.07)	0.003	0.045 (0.058)	0.045 (0.052)	0.25	0.25 (0.252)	0.38 (0.237)	0.5	0.49 (0.489)	0.545 (0.489)	<0.01	-0.04 (0.8)	0.0 (0.92)	

For more details, see Appendix Tables A1 and A5.

Table A8. Estimation of trait-model parameters of DST and DSQ pedigrees (overdominant models)

Model name	Estimated median p (MAD)		Estimated median F_0 (MAD)		Estimated median f_0 (MAD)		Estimated median F_1 (MAD)		Estimated median f_1 (MAD)		Estimated median F_2 (MAD)		Estimated median f_2 (MAD)	
	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ	DST	DSQ
U1	0.1	0.12 (0.09)	0.12 (0.09)	0.01	0.008 (0.007)	0.008 (0.004)	0.2	0.09 (0.074)	0.1 (0.059)	0.01	0.25 (0.319)	0.31 (0.267)		
U1	0.01	0.1 (0.03)	0.11 (0.04)	0.01	0.003 (0.004)	0.005 (0.006)	0.2	0.14 (0.089)	0.14 (0.074)	0.01	0.31 (0.267)	0.34 (0.208)		
U2	0.1	0.11 (0.03)	0.11 (0.03)	0.01	0.005 (0.007)	0.005 (0.007)	0.5	0.5 (0.089)	0.51 (0.044)	0.01	0.07 (0.104)	0.03 (0.044)		
U2	0.01	0.045 (0.01)	0.05 (0.03)	0.01	0.002 (0.003)	0.005 (0.007)	0.5	0.49 (0.059)	0.49 (0.044)	0.01	0.63 (0.549)	0.58 (0.563)		
U3	0.1	0.11 (0.03)	0.11 (0.03)	0.01	0.006 (0.009)	0.006 (0.009)	0.8	0.81 (0.03)	0.8 (0.03)	0.01	0.025 (0.037)	0.01 (0.015)		
U3	0.01	0.02 (0.02)	0.03 (0.03)	0.01	0.008 (0.006)	0.008 (0.006)	0.8	0.8 (0.03)	0.8 (0.03)	0.01	0.89 (0.163)	0.73 (0.4)		
U3n1000	0.01	0.02 (0.01)	0.025 (0.01)	0.01	0.008 (0.003)	0.01 (0.003)	0.8	0.8 (0.03)	0.8 (0.015)	0.01	0.86 (0.208)	0.64 (0.43)		
U4	0.35	0.36 (0.07)	0.36 (0.09)	0.01	0.003 (0.004)	0.01 (0.015)	0.9	0.9 (0.03)	0.9 (0.015)	0.01	0.001 (0.001)	0.0 (0.0)		

For more details, see Appendix Tables A1 and A5.

Table A9. Estimation of trait-model parameters of three-generation (3-G) pedigrees (recessive models)

Model name	P	Estimated median p (MAD)	F_0	Estimated median f_0 (MAD)	F_1	Estimated median f_1 (MAD)	F_2	Estimated median f_2 (MAD)	D	Estimated median d (MAD)
R1	0.1	0.01 (0.01)	0.01	0.001 (0.0003)	0.01	0.001 (0.0007)	0.2	0.11 (0.015)	-1.0	-1.0 (0.01)
R1	0.01	0.08 (0.11)	0.01	0.235 (0.348)	0.01	0.11 (0.163)	0.2	0.32 (0.474)	-1.0	-0.39 (1.14)
R1n1000	0.01	0.075 (0.1)	0.01	0.19 (0.281)	0.01	0.11 (0.163)	0.2	0.28 (0.415)	-1.0	-0.24 (1.21)
R2	0.1	0.09 (0.03)	0.01	0.008 (0.01)	0.01	0.008 (0.003)	0.5	0.51 (0.052)	-1.0	-0.99 (0.04)
R2	0.01	0.06 (0.08)	0.01	0.015 (0.022)	0.01	0.02 (0.03)	0.5	0.41 (0.46)	-1.0	-0.98 (0.27)
R3	0.1	0.09 (0.03)	0.01	0.008 (0.01)	0.01	0.01 (0.003)	0.8	0.8 (0.03)	-1.0	-1.0 (0.03)
R3	0.01	0.05 (0.05)	0.01	0.035 (0.037)	0.01	0.043 (0.033)	0.8	0.76 (0.267)	-1.0	-0.99 (0.06)
AFR1	0.2	0.11 (0.15)	0.04	0.01 (0.015)	0.04	0.015 (0.021)	0.2	0.16 (0.185)	-1.0	-0.94 (0.27)
AFR1n1000	0.2	0.1 (0.13)	0.04	0.01 (0.015)	0.04	0.01 (0.013)	0.2	0.165 (0.185)	-1.0	-0.98 (0.19)
AFR2	0.25	0.23 (0.06)	0.003	0.001 (0.001)	0.05	0.05 (0.015)	0.5	0.51 (0.059)	-0.81	-0.83 (0.07)

For more details, see Appendix Table A1.

Table A10. Estimation of trait-model parameters of three-generation (3-G) pedigrees (dominant models)

Model name	P	Estimated median p (MAD)	F_0	Estimated median f_0 (MAD)	F_1	Estimated median f_1 (MAD)	F_2	Estimated median f_2 (MAD)	D	Estimated median d (MAD)
D1	0.1	0.08 (0.07)	0.01	0.01 (0.003)	0.2	0.14 (0.074)	0.2	0.19 (0.163)	1.0	0.68 (1.32)
D1	0.01	0.035 (0.01)	0.01	0.004 (0.005)	0.2	0.19 (0.059)	0.2	0.27 (0.163)	1.0	0.33 (0.76)
D1n1000	0.01	0.035 (0.01)	0.01	0.004 (0.005)	0.2	0.19 (0.044)	0.2	0.27 (0.119)	1.0	0.39 (0.62)
D2	0.1	0.09 (0.03)	0.01	0.008 (0.012)	0.5	0.5 (0.044)	0.5	0.48 (0.193)	1.0	0.96 (0.77)
D2	0.01	0.015 (0.01)	0.01	0.01 (0.003)	0.5	0.5 (0.03)	0.5	0.54 (0.563)	1.0	0.42 (0.82)
D3	0.1	0.1 (0.03)	0.01	0.01 (0.015)	0.8	0.8 (0.148)	0.8	0.8 (0.148)	1.0	0.98 (0.39)
D3	0.01	0.015 (0.01)	0.01	0.008 (0.003)	0.8	0.8 (0.015)	0.8	0.84 (0.237)	1.0	0.68 (0.33)
AFD1	0.05	0.11 (0.1)	0.04	0.015 (0.015)	0.2	0.22 (0.104)	0.2	0.22 (0.178)	1.0	0.26 (0.91)
AFD1n1000	0.05	0.11 (0.07)	0.04	0.02 (0.015)	0.2	0.21 (0.096)	0.2	0.21 (0.148)	1.0	0.45 (0.79)
AFD2	0.25	0.18 (0.12)	0.003	0.04 (0.052)	0.5	0.43 (0.074)	0.5	0.43 (0.222)	1.0	1.08 (0.99)

For more details, see Appendix Table A1.

Table A11. Estimation of trait-model parameters of three-generation (3-G) pedigrees (additive models)

Model name	P	Estimated median p (MAD)	F_0	Estimated median f_0 (MAD)	F_1	Estimated median f_1 (MAD)	F_2	Estimated median f_2 (MAD)	D	Estimated median d (MAD)
A1	0.1	0.09 (0.06)	0.01	0.008 (0.01)	0.1	0.1 (0.074)	0.2	0.21 (0.148)	-0.05	-0.09 (0.44)
A1	0.01	0.045 (0.05)	0.01	0.01 (0.0)	0.1	0.1 (0.03)	0.2	0.12 (0.119)	-0.05	0.14 (0.89)
A1n1000	0.01	0.02 (0.02)	0.01	0.01 (0.0)	0.1	0.1 (0.015)	0.2	0.14 (0.133)	-0.05	0.19 (0.9)
A2	0.1	0.08 (0.05)	0.01	0.01 (0.009)	0.2	0.18 (0.089)	0.5	0.52 (0.163)	-0.22	-0.31 (0.25)
A2	0.01	0.035 (0.01)	0.01	0.004 (0.005)	0.2	0.19 (0.074)	0.5	0.34 (0.193)	-0.22	0.04 (0.5)
A3	0.1	0.09 (0.03)	0.01	0.008 (0.012)	0.5	0.5 (0.059)	0.8	0.82 (0.148)	0.24	0.19 (0.27)
A3	0.01	0.015 (0.01)	0.01	0.01 (0.003)	0.5	0.5 (0.044)	0.8	0.765 (0.348)	0.24	0.16 (0.4)
AFA1	0.1	0.09 (0.1)	0.03	0.015 (0.022)	0.13	0.11 (0.089)	0.23	0.25 (0.237)	0.0	-0.13 (0.58)
AFA1n1000	0.1	0.09 (0.07)	0.03	0.02 (0.022)	0.13	0.13 (0.089)	0.23	0.24 (0.193)	0.0	-0.12 (0.44)
AFA2	0.5	0.38 (0.25)	0.003	0.05 (0.074)	0.25	0.23 (0.178)	0.5	0.5 (0.267)	<0.01	-0.09 (0.7)

For more details, see Appendix Table A1.

Table A12. Estimation of trait-model parameters of three-generation (3-G) pedigrees (overdominant models)

Model name	P	Estimated median p (MAD)	F_0	Estimated median f_0 (MAD)	F_1	Estimated median f_1 (MAD)	F_2	Estimated median f_2 (MAD)
U1	0.1	0.12 (0.03)	0.01	0.005 (0.007)	0.2	0.2 (0.074)	0.01	0.04 (0.044)
U1	0.01	0.035 (0.02)	0.01	0.003 (0.004)	0.2	0.19 (0.074)	0.01	0.22 (0.148)
U2	0.1	0.1 (0.01)	0.01	0.008 (0.003)	0.5	0.5 (0.044)	0.01	0.025 (0.037)
U2	0.01	0.015 (0.01)	0.01	0.008 (0.003)	0.5	0.5 (0.044)	0.01	0.13 (0.193)
U3	0.1	0.1 (0.01)	0.01	0.008 (0.003)	0.8	0.8 (0.03)	0.01	0.01 (0.015)
U3	0.01	0.015 (0.01)	0.01	0.008 (0.003)	0.8	0.8 (0.015)	0.01	0.33 (0.489)
U3n1000	0.01	0.015 (0.01)	0.01	0.008 (0.003)	0.8	0.8 (0.015)	0.01	0.28 (0.415)
U4	0.35	0.38 (0.1)	0.01	0.01 (0.007)	0.9	0.9 (0.015)	0.01	0.006 (0.008)

For more details, see Appendix Table A1.

References

- 1 Ginsburg E, Malkin I, Elston RC: Theoretical Aspects of Pedigree Analysis. Tel Aviv, Ramot Publishing House, 2006.
- 2 Hasstedt SJ: Pedigree Analysis Software, version 7.1. Salt Lake City, Department of Human Genetics, University of Utah, 2009.
- 3 S.A.G.E. 6.4. Statistical Analysis for Genetic Epidemiology. 2016. <http://darwin.cwru.edu/sage> (accessed: October 5, 2017).
- 4 Thompson EA: Monte Carlo in Genetic Analysis. Technical Report No. 294. Seattle, Department of Statistics, University of Washington, 1995.
- 5 Thompson EA: Statistical Inferences from Genetic Data on Pedigrees. NSF-CBMS Regional Conference Series in Probability and Statistics, vol 6. Beachwood, Institute of Mathematical Statistics, 2000.
- 6 Kriszt A, Losonczy G, Berta A, Vereb G, Takács L: Segregation analysis suggests that Keratoconus is a complex non-Mendelian disease. *Acta Ophthalmol* 2014;92:e562–e568.
- 7 Rao DC: CAT scans, PET scans, and genomic scans. *Genet Epidemiol* 1998;15:1–18.
- 8 Elston RC: Methods of linkage analysis – and the assumptions underlying them. *Am J Hum Genet* 1998;63:931–934.
- 9 Knapp M, Seuchter SA, Baur MP: Linkage analysis in nuclear families. 2: Relationship between affected sib-pair tests and lod score analysis. *Hum Hered* 1994;44:44–51.
- 10 Strauch K: MOD-score analysis with simple pedigrees: an overview of likelihood-based linkage methods. *Hum Hered* 2007;64:192–202.
- 11 Elston RC: Man bites dog? The validity of maximizing lod scores to determine mode of inheritance. *Am J Med Genet* 1989;34:487–488.
- 12 Risch N: Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type I diabetes. *Am J Hum Genet* 1984;36:363–386.
- 13 Greenberg DA: Linkage analysis assuming a single-locus mode of inheritance for traits determined by two loci: inferring mode of inheritance and estimating penetrance. *Genet Epidemiol* 1990;7:467–479.
- 14 Elston RC, Sobel E: Sampling considerations in the gathering and analysis of pedigree data. *Am J Hum Genet* 1979;31:62–69.
- 15 Sawyer S: Maximum likelihood estimators for incorrect models, with an application to ascertainment bias for continuous characters. *Theor Popul Biol* 1990;38:351–366.
- 16 Ginsburg E, Malkin I, Elston RC: Sampling correction in linkage analysis. *Genet Epidemiol* 2004;27:87–96.
- 17 Ewens WJ, Shute NC: A resolution of the ascertainment sampling problem. I. Theory. *Theor Popul Biol* 1986;30:388–412.
- 18 Clerget-Darpoux F, Bonaïti-Pellié C, Hochez J: Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 1986;42:393–399.
- 19 Malkin I, Elston RC: Response to letter by Veronica J. Vieland and Susan E. Hodge. *Genet Epidemiol* 2005;28:286–287.
- 20 Hodge SE, Vieland VJ: The essence of single ascertainment. *Genetics* 1996;144:1215–1223.
- 21 Vieland VJ, Hodge SE: Inherent intractability of the ascertainment problem for pedigree data: a general likelihood framework. *Am J Hum Genet* 1995;56:33–43.
- 22 Vieland VJ, Hodge SE: Ascertainment bias in linkage analysis: comments on Ginsburg et al. *Genet Epidemiol* 2005;28:283–285.
- 23 Slager SL, Vieland VJ: Investigating the numerical effects of ascertainment bias in linkage analysis: development of methods and preliminary results. *Genet Epidemiol* 1997;14:1119–1124.
- 24 Liang KY, Rathouz PJ, Beaty TH: Determining linkage and mode of inheritance: mode scores and other methods. *Genet Epidemiol* 1996;13:575–593.
- 25 Brugger M, Strauch K: Fast linkage analysis with MOD scores using algebraic calculation. *Hum Hered* 2014;78:179–194.
- 26 Suarez BK, Rice J, Reich T: The generalized sib pair IBD distribution: its use in the detection of linkage. *Ann Hum Genet* 1978;42:87–94.
- 27 Holmans P: Asymptotic properties of affected-sib-pair linkage analysis. *Am J Hum Genet* 1993;52:362–374.
- 28 Knapp M: A note on linkage analysis with affected sib triplets. *Hum Hered* 2005;59:21–25.
- 29 Falls JG, Pulford DJ, Wylie AA, Jirtle RL: Genomic imprinting: implications for human disease. *Am J Pathol* 1999;154:635–647.
- 30 Mattheisen M, Dietter J, Knapp M, Baur MP, Strauch K: Inferential testing for linkage with GENEHUNTER-MODSCORE: the impact of the pedigree structure on the null distribution of multipoint MOD scores. *Genet Epidemiol* 2008;32:73–83.
- 31 Dietter J, Mattheisen M, Fürst R, Rüschemdorf F, Wienker TF, Strauch K: Linkage analysis using sex-specific recombination fractions with GENEHUNTER-MODSCORE. *Bioinformatics* 2007;23:64–70.
- 32 Strauch K: Parametric linkage analysis with automatic optimization of the disease model parameters. *Am J Hum Genet* 2003;73(suppl 1):A2624.
- 33 Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, Baur MP: Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. *Am J Hum Genet* 2000;66:1945–1957.
- 34 Knapp M, Strauch K: Affected-sib-pair test for linkage based on constraints for identical-by-descent distributions corresponding to disease models with imprinting. *Genet Epidemiol* 2004;26:273–285.
- 35 Strauch K: Gene mapping, imprinting and epigenetics; in Jorde LB, Little PFR, Dunn MJ, Subramaniam S (eds): *Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics*. Hoboken, John Wiley & Sons, 2005.
- 36 Haghighi F, Hodge SE: Likelihood formulation of parent-of-origin effects on segregation analysis, including ascertainment. *Am J Hum Genet* 2002;70:142–156.
- 37 Ott J: Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 1989;86:4175–4178.
- 38 Schäffer AA, Lemire M, Ott J, Lathrop GM, Weeks DE: Coordinated conditional simulation with SLINK and SUP of many markers linked or associated to a trait in large pedigrees. *Hum Hered* 2011;71:126–134.
- 39 Weeks DE, Lehner T, Squires-Wheeler E, Kaufmann C, Ott J: Measuring the inflation of the lod score due to its maximization over model parameter values in human linkage analysis. *Genet Epidemiol* 1990;7:237–243.
- 40 Xing C, Elston RC: Distribution and magnitude of type I error of model-based multipoint lod scores: implications for multipoint mod scores. *Genet Epidemiol* 2006;30:447–458.
- 41 Flaquer A, Strauch K: A comparison of different linkage statistics in small to moderate sized pedigrees with complex diseases. *BMC Res Notes* 2012;5:411.
- 42 Shete S, Zhou X: Parametric approach to genomic imprinting analysis with applications to Angelman's syndrome. *Hum Hered* 2005;59:26–33.
- 43 Greenberg DA: Inferring mode of inheritance by comparison of lod scores. *Am J Med Genet* 1989;34:480–486.