Published in final edited form as:

Genet Epidemiol. 2009 February ; 33(2): 136. doi:10.1002/gepi.20365.

A hybrid design: case-parent triads supplemented by control-mother dyads

S.H. Vermeulen¹, M. Shi², C.R. Weinberg², and D.M. Umbach²

¹Department of Endocrinology, Department of Epidemiology, Biostatistics and HTA (133) & Department of Human Genetics, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands ²National Institute of Environmental Health Sciences, MD A3-03, P.O. Box 12233, Research Triangle Park, NC 27709

Abstract

Hybrid designs arose from an effort to combine the benefits of family-based and population-based study designs. A recently proposed hybrid approach augments case-parent triads with population-based control-parent triads, genotyping everyone except the control offspring. Including parents of controls substantially improves statistical efficiency for testing and estimating both offspring and maternal genetic relative risk parameters relative to using case-parent triads alone. Moreover, it allows testing of required assumptions. Nevertheless, control fathers can be hard to recruit, whereas control offspring and their mothers may be readily available. Consequently, we propose an alternative hybrid design where offspring-mother pairs, instead of parents, serve as population-based controls. We compare the power of our proposed method with several competitors and show that it performs well in various scenarios, though it is slightly less powerful than the hybrid design that uses control parents. We describe approaches for checking whether population stratification will bias inferences that use controls and whether the mating symmetry assumption holds. Surprisingly, if mating symmetry is violated, even though mating-type parameters cannot be directly estimated using control-mother dyads alone, and maternal effects cannot be estimated using case-parent triads alone, combining both sources of data allows estimation of all the parameters. This hybrid design can also be used to study environmental influences on disease risk and gene-by-environment interactions.

Keywords

genetic relative risk; maternal effect; Single Nucleotide Polymorphism (SNP); association studies; family-based design; population-based design; Poisson regression; early-onset disease

Introduction

Genetic factors may increase risk for congenital disorders via direct effects of the inherited genotype of the offspring or via the genotype of the mother. Maternal effects can arise because the genetic variants carried by the mother influence the prenatal environment in which the fetus develops, thereby increasing disease risk. Under such a mechanism, the genotype distribution in the case mothers is different from that in the control population, whereas transmission of the genetic variant to the offspring conforms to Mendelian expectation. Disentangling offspring

from maternal genetic effects can increase insight into the etiology of disorders with early onset in life, such as congenital malformations and childhood cancers.

The case-parents design, comprising affected offspring and their two parents, permits testing and estimation of offspring- and maternally-mediated genetic effects [Weinberg et al., 1998; Wilcox et al., 1998]. Offspring effects are evaluated by the apparent over-transmission (compared to the Mendelian expectation of 0.5) of deleterious alleles from heterozygous parents to affected offspring. Maternal genotype effects are assessed through deviations from genotype mating symmetry in the case-parents data under the assumption of mating symmetry in the source population. That is, a deleterious allele that acts via the mother will be more prevalent in mothers than in fathers of affected offspring. The key mating-symmetry assumption cannot, however, be evaluated using case-parent triads alone.

Another design that can be used to disentangle offspring and maternal genetic effects ascertains affected children and their mothers as well as a random sample of unaffected children and their mothers. In such a case-mother/control-mother design, the offspring-mother pairs are the units for analysis. Both offspring and maternal effects are assessed through case-control comparisons; no mating symmetry assumption is required. Unlike the case-parents design, however, this design is vulnerable to bias due to the existence of genetically distinct subpopulations. Weinberg and Umbach [2005] describe the strengths and weaknesses of both the population-based case-mother/control-mother and the family-based case-parents design.

To bring together the advantages of both family-based and population-based approaches, Nagelkerke et al. [2004], Epstein et al. [2005] and Weinberg and Umbach [2005] have proposed various hybrid designs that combine both kinds of data. One hybrid approach [Weinberg and Umbach, 2005] enrolls case-parent and control-parent triads and genotypes case-parent triads but only the parents of control offspring (case-parent triad/control parents design). This hybrid design greatly improves power for the evaluation of offspring and maternally-mediated genetic effects. It allows testing for bias from population stratification; and, if bias is detected, confounding is avoided by using only the case-parent triads. In addition, the assumption of parental mating symmetry in the population at large can be tested, and maternal genetic effects can be validly estimated even if mating symmetry is rejected [Weinberg and Umbach, 2005].

The case-parent triad/control parents hybrid design calls for both control parents to be genotyped. Fathers, however, are often hard to recruit while mothers and unaffected offspring may already be available. This paper presents an alternative hybrid design where control-mother dyads replace parents of controls. Exposure information would still be collected for both the case offspring and the unrelated control offspring. In the hybrid design that genotypes control parents, one can directly estimate the mating type frequencies in the source population. In the proposed design that genotypes control-mother dyads instead, the information on those frequencies is indirect because the genotype of the control child acts as a surrogate for the father's genotype. This feature introduces some challenges. We demonstrate that, even in scenarios with missing genotype data, this alternative hybrid design has greater power than the case-parents or the case-mother/control-mother designs but has slightly reduced power compared to the hybrid design that uses control parents. We also show that this alternative hybrid design retains the ability to test for bias due to population stratification and, thereby, to examine whether case-parents and control data can be safely combined. Finally, we describe procedures for checking the assumption of mating symmetry and for making valid inference on maternal effects even when that assumption fails.

Analysis of case-parent triad/control-mother dyad data

Our alternative hybrid approach starts with a random sample of affected individuals and a random sample of unaffected individuals. Cases and their parents are enrolled and genotyped

but only mothers of controls and the controls themselves are genotyped. We are interested in testing for association between disease risk and the offspring and maternal genotypes at a diallelic autosomal locus. We assume that the disease is rare and that Mendelian transmission probabilities hold for that locus in the underlying population, and hence among controls. Validly combining information from case and control families requires an assumption that any population structure is benign with respect to bias, an assumption that can be probed with the data at hand. Neither Hardy-Weinberg equilibrium nor random mating is required for validity.

Let p denote the frequency of the minor or 'variant' allele. Which allele is designated as the 'variant' has no effect on estimation or testing beyond the mathematical inversions. Let M , F and C represent the number of variant alleles (0,1,2) carried by the mother, father and child, respectively. D is an indicator variable for disease status, which is 1 for case families and 0 for control families.

Following Schaid and Sommer [1993], we define nine different mating types based on the number of variant copies carried by the mother and the father. These mating types along with their possible offspring genotypes lead to 15 possible (M, F, C) categories, and, consequently, one can imagine two 15-cell multinomial distributions of offspring and parental genotypes, one for control triads and one for case triads. In hybrid designs, typically the full (M, F, C) data are recorded for case families whereas only partial data are collected from control families; here, only (M, C) data. Initially we assume that any genotyping called for by the design is complete, but missing-data methods can be employed if some genotypes are missing [Weinberg, 1999a].

For control families, expected counts in the 15-cell multinomial can be modeled using Mendelian transmission probabilities and mating type parameters (μ_{mf}), which are proportional to the frequencies of mother-father pairs with $M=m$ and $F=f$ in the source population. The expected counts for control-mother dyads (the last 7 lines of Table I) arise from the 15-cell multinomial by summing counts across the genotypes of possible fathers. The distribution of control-mother dyads has two noteworthy features: First, the (M, C) cells (0,2) and (2,0) are not possible, leaving only seven cells with non-zero expected counts. Second, and less obviously, the following relationship is a consequence of Mendelian transmission alone: when $M=1$, the expected count for $C=1$ is the sum of the expected counts for $C=0$ and $C=2$. This constraint reduces the available degrees-of-freedom contributed by the seven control-dyad cells from six to five.

For case families, expected counts in the 15-cell multinomial involve not only mating-type parameters and Mendelian probabilities but also four genetic relative risk parameters. We denote these relative risk parameters as follows: R_1 (R_2) is the relative risk for offspring carrying 1 (respectively, 2) copies of the variant compared to offspring carrying none; S_1 (S_2) is the relative risk for offspring whose mother carries 1 (respectively, 2) copies of the variant allele compared to offspring whose mother carries none. Combining the 15 cells for case-parent triads with the seven cells for control-mother dyads yields a 22-cell multinomial for the proposed design (Table I).

For the case-parents design and for the hybrid design that uses parents of controls, the multinomial expected cell counts are all products of parameters and can be fitted using log-linear Poisson regression. Because the expected counts for control-mother dyads involve sums of parameters, however, the expected counts in Table I for the 22-cell multinomial are not themselves log-linear. A straightforward way to proceed with fitting either model is to regard the 22-cell multinomial as a version of the full 30-cell multinomial (15 cells each for case families and control families) that is missing genotype data for fathers of controls by design. Thus, one can use missing-data methods like the Expectation-Maximization (EM) algorithm

[Dempster et al., 1977] in conjunction with a log-linear model for the 30-cell multinomial. Use of the EM algorithm also allows inclusion of any case-parent triads or control-mother dyads with missing genotypes that may arise through genotyping failure or incomplete ascertainment. For valid analysis, one must be able to assume that these genotype data are missing at random conditional on disease status and the observed genotypes. When control fathers are missing only by design, this assumption is satisfied without doubt because all of them are missing.

Assuming a multiplicative model for risk and no bias from population structure, a log-linear model for the full 30-cell multinomial (corresponding to Table I, column 5) would be:

$$\ln [E(\text{count} | M=m, F=f, C=c, D=d)] = \ln(\mu_{mf}) + \beta_1 dI_{(c=1)} + \beta_2 dI_{(c=2)} + \alpha_1 dI_{(m=1)} + \alpha_2 dI_{(m=2)} + \gamma d + \ln(\text{Off}_{mfc}) \quad (1)$$

$I_{()}$ is an indicator function which takes a value of 1 if the parenthetical condition is met and 0 otherwise. β_1 and β_2 denote the natural logarithms of the offspring genetic relative risks, R_1 and R_2 , respectively; and α_1 and α_2 denote the natural logarithms of maternal genetic relative risks, S_1 and S_2 , respectively. γ corresponds to the natural logarithm of the normalizing factor B. The offset, denoted Off_{mfc} is the constant multiplier (1, $\frac{1}{2}$ or $\frac{1}{4}$) given in Table I, column 5. In this model, the nine mating-type parameters, μ_{mf} , are common to both cases and controls and can be interpreted as proportional to the mating type frequencies in the source population.

This model is fundamental to the analysis of data from our proposed design. Modifications of the model by omitting or including certain additional terms allow one to construct likelihood ratio tests (LRTs) of hypotheses about the genetic relative risk parameters or to test the assumptions about bias from population stratification or about mating symmetry. In addition, models addressing maternal-fetal incompatibility can be constructed by including two additional relative risk parameters [Sinsheimer et al., 2003]. All tests of assumptions that we subsequently develop under the four risk parameters in model (1) can be developed and applied without difficulty when these two additional risk parameters are present.

Likelihood ratio tests in this missing data situation must be based on the observed-data likelihood, not the pseudo-complete-data likelihood. We have used the program LEM [van den Oord and Vermunt, 2000] in our subsequent analyses. LEM was designed to fit log-linear models with missing data via the EM algorithm. Consequently, dealing with missing fathers among controls or with other patterns of missing data does not require special programming. The program allows incorporation of the Mendelian constraints and returns valid test statistics as well as valid estimates for risk parameters and their standard errors. Examples of the LEM scripts that we used are available at <http://www.niehs.nih.gov/research/atniehs/labs/bb/staff/weinberg/index.cfm#downloads>.

Association of offspring or maternal genotypes with risk

An LRT statistic for any particular subset of the four relative risk parameters can be calculated by computing twice the change in maximized observed-data log likelihood between a version of model (1) that estimates all parameters and a version that fixes the subset to be tested at their null values (most often, zero). For calculating a p value, this LRT statistic is referred to a χ^2 distribution whose degrees of freedom (df) equal the number of parameters in the subset. Model (1) can be modified, if desired, to accommodate specific modes of inheritance like recessive, dominant, or log-additive.

To this point, we focused on model (1) with nine mating type parameters (Table I, column 5). We call this model the mating-asymmetry model because it places no constraints on the mating-

type parameters. Mating symmetry is frequently assumed [Schaid and Sommer, 1993; Wilcox et al., 1998] and is required for inference on maternal genetic effects with a case-parents design. Mating symmetry means that the probability of parents with $M=m$ and $F=f$ is the same as the probability of parents with $M=f$ and $F=m$ in the source population. In terms of the μ_{mf} parameters, mating symmetry implies three constraints, namely: $\mu_{01} = \mu_{10}$, $\mu_{02} = \mu_{20}$, and $\mu_{12} = \mu_{21}$, in effect reducing nine mating types to six (Table I, column 6). The mating-symmetry model can also be described algebraically using model (1) with the understanding that the mating-type parameters and offsets are changed to those of Table I, column 7. Any tests that can be carried out on the relative risk parameters under the original nine-mating-type model (mating asymmetry) can also be carried out under the six-mating-type model (mating symmetry). When mating symmetry holds in the general population, tests under the mating-symmetry model will be more powerful than tests under the mating-asymmetry model because the former model involves fewer parameters. Using either nine or six mating types, model (1) employs the same mating-type parameters for both cases and controls and, thereby, relies on an assumed absence of bias from population structure, as does any case-control study.

Checking assumptions: absence of bias from population structure

For assessing the contribution of variant alleles to risk, the key to the power gains possible with this hybrid design is that control-mother dyads contribute to the estimation of the mating-type parameters. The two sets of mating-type parameters will truly be equal only if there is no bias due to population structure. The existence of such bias requires not only that genetic population structure be present but also that the risk in noncarriers be correlated with allele frequency across subpopulations. Whenever this bias is present, model (1) is not valid for combining data from case and control families.

A test for bias from population structure is simply a test of whether model (1) with common mating-type parameters for cases and controls fits the data as well as an extended model with distinct mating-type parameters for cases and controls. Such a test could be conducted under either mating asymmetry or mating symmetry; and, in principle, the extended model would involve eight or five additional parameters, respectively. The ability to estimate separate mating-type parameters for cases and controls is constrained, however, by use of control-mother dyads and, under mating asymmetry, by the complete aliasing of certain mating-types with maternal effects. Consequently, the extended model has only three additional df under mating asymmetry and only four under mating symmetry. An alternative approach that will sometimes be more powerful under mating symmetry than the four-df LRT is to employ a test for a particular linear trend in the disparity between case and control mating-type parameters. This one-df LRT compares model (1) to a model that includes the single additional term $\theta(M + F)d$ where θ , the unknown trend slope, is zero in the absence of bias from population structure. A monotone trend would arise, for example, if a population consisted of two distinct subpopulations, each in HWE at the SNP in question, and their respective baseline risks and allele frequencies were correlated. If under either of the testing strategies the fit is statistically significantly improved by the inclusion of the separate mating type parameter(s), case and control data cannot be validly combined, and estimation of relative risk parameters should be based on the case-parent triads only.

Checking assumptions: mating symmetry

Unlike the case-parents design where mating symmetry is required for assessing maternal effects but cannot be checked, the proposed hybrid design allows one to assess the assumption and to impose it or not accordingly. Enforcing mating symmetry when it holds will enhance power, but enforcing it when it fails could bias estimates of genetic effects, particularly maternal effects. If bias from population structure is absent, the three constraints on the nine

μ_{mf} implied by mating symmetry can be tested in full. The appropriate three-df LRT is constructed by comparing model (1) under mating asymmetry (nine mating-type parameters) to model (1) under mating symmetry (six mating-type parameters). If mating symmetry is rejected, then the mating asymmetry version of model (1) is used to test and estimate relative risk parameters.

Ideally, one might like to examine mating symmetry in the source population without concern for population structure. Unfortunately, the control-mother dyads do not supply enough information about the nine mating-type parameters under asymmetry to fully probe the three mating-symmetry constraints. A partial examination of these constraints is possible, however. Under mating symmetry, the sum of expected counts in (M,C) cells (0,1) and (1,2) equals the sum in cells (1,0) and (2,1) (Table I, column 7). Re-expressed in terms of the μ_{mf} , this constraint becomes: $\frac{1}{2}\mu_{01} + \frac{1}{2}\mu_{12} + \mu_{02} = \frac{1}{2}\mu_{10} + \frac{1}{2}\mu_{21} + \mu_{20}$. Thus, a one-df LRT can be constructed by comparing a model in which the two sums are constrained to be equal to a model without that constraint. This test probes mating symmetry only in part because the two sums can be the same even when mating symmetry fails. This test, though limited, would be important in settings where the presence of bias from population structure forced one to rely on the case-parents component of the hybrid design for inference about risk parameters, but a check on the assumption required for examining maternal effects was desired. Interestingly, this test is closely related to the 1-TDT [Sun et al., 1999]. When applied to case-mother dyads, the 1-TDT tests for offspring genetic effects assuming both mating symmetry and the absence of maternal genetic effects; when applied to control-mother dyads instead, the 1-TDT is asymptotically equivalent to our one-df LRT for mating symmetry.

Power comparisons: relative risk parameters

We compared the power of several study designs for detecting offspring and maternal genetic associations with disease: the family-based case-parents design; the population-based case-mother/control-mother design; and two hybrid designs (case-parent triads/control parents and case-parent triads/control-mother dyads). For all these designs, we assessed power for a four-df LRT of the null hypothesis that $R_1=R_2=S_1=S_2=1$ under several alternative risk scenarios. For the two hybrid designs, we investigated the tests under both the mating-symmetry and mating-asymmetry models. We used traditional logistic regression for the case-mother/control-mother design; we used log-linear Poisson regression for the other designs.

Our power calculations are based on the noncentrality parameter for a four-df chi-squared LRT. We calculated the noncentrality parameter as the LRT statistic constructed by treating expected counts under the specified alternative hypothesis as data [Agresti, 1990]. Values of the noncentrality parameter can be translated to power values using the noncentral χ^2 distribution with four degrees of freedom. To calculate expected counts, we considered populations with allele frequencies ranging over the interval from zero to one where parental mating-type frequencies obeyed Hardy-Weinberg equilibrium and, consequently, exhibited mating symmetry. (Note Hardy-Weinberg equilibrium is simply a convenience here; it is not needed for the validity of our analyses.) We based all our calculations on 150 case families and 150 control families (the case-parents design used only the 150 case families). We plotted the noncentrality parameters as a function of allele prevalence and included horizontal reference lines corresponding to specific levels of power for a four-df LRT at alpha level 0.05. When the noncentrality parameter exceeds a given reference line, the LRT's power exceeds the specified power. Transformation of the noncentrality parameter to a different planned sample size with the same case:control ratio (say, for K cases) is accomplished by multiplying the noncentrality parameter values from our figure by $K/150$. For any alpha level, the ratio of two noncentrality parameters corresponds to the relative efficiencies of the corresponding two designs, *i.e.*, the ratio of the sample sizes needed to achieve the same power.

We considered four distinct risk scenarios. The first scenario included a gene-dose effect of the fetal genotype but no effect of the maternal genotype, specifically, R_1, R_2, S_1, S_2 , were 2, 3, 1, 1, respectively. The second scenario included a gene-dose effect of the maternal genotype and no effect of the fetal genotype, specifically, R_1, R_2, S_1, S_2 , were 1, 1, 2, 3, respectively. The third involved recessive effects of both the fetal and maternal genotypes (R_1, R_2, S_1, S_2 were 1, 2, 1, 3, respectively). The final scenario included a recessive effect of the fetal genotype and a dominant effect of the maternal genotype (R_1, R_2, S_1, S_2 , were 1, 3, 2, 2, respectively). In all four scenarios, either hybrid design exhibited better power across the range of allele frequencies than both the case-parent triads and the case-mother/control-mother design (Figure 1). In addition, the hybrid using parents of controls always performed somewhat better than the hybrid using control-mother dyads, reflecting that the former design provides more mating-type information from control families. Analyses of the hybrid designs that enforced mating symmetry generally provided more power than those that did not, but the difference in power depended on the scenario. In the scenario with maternal but without offspring effects (Figure 1, panel *b*), the power of hybrid designs when mating symmetry was not enforced was the same or close to that of the case-mother/control-mother design. The relative efficiency of the case-parents design compared to the case-mother/control-mother design depended on the particular scenario.

We also examined the same set of four scenarios when each individual genotype in every triad or dyad independently had a 20% chance being missing (Figure 1, right column). Thus, for example, a triad could be complete, or have exactly one or two or all three of its three genotypes missing. The missing genotypes could be the mother, father, or offspring. Our analyses used all families except those where every individual had a missing genotype. For these analyses, we used the EM algorithm to recover information from families with only partial genotype data. For the case-mother/control-mother design that used logistic regression, invoking the EM algorithm entails exploiting the well-known equivalence between logistic models with discrete covariates and log-linear models for multi-way tables [Agresti, 1990]. Although missing data lowered all the curves compared to their complete-data counterparts (Figure 1), the relative efficiencies among curves representing the various designs and models were much the same. The one exception was the case-parents design; in the scenarios with offspring effects, its efficiency was more reduced by missing data than was that of the other designs.

Power for checking assumptions

We examined the power of our tests for bias due to population structure for a configuration where mating-type parameters differed between cases and controls. We created this configuration by mixing two subpopulations, each in Hardy-Weinberg equilibrium, under a null relative risk scenario [Weinberg and Umbach, 2005]: the first had allele frequency 0.1 and baseline disease risk (risk among those with 0 copies of the variant) 0.001; the second had allele frequency 0.5 and baseline disease risk 0.003. This mixture induced a trend in the discrepancy between the separate case and the control mating-type parameters.

With 150 case families and 150 control families, this population structure posed a serious challenge to the proposed hybrid design under model (1) by producing biased relative risk estimates and an inflated type I error rate for testing genetic effects (actual α -level = 0.43 at nominal $\alpha = 0.05$). Testing for the presence of bias due to population structure under mating symmetry at $\alpha = 0.10$, the power was 0.42 and 0.54 for the four- and the one-df test, respectively, corresponding to noncentrality parameters of 3.71 and 3.03. Doubling the sample size doubles the noncentrality parameters, yielding power 0.69 and 0.79, respectively. All these values were near those achieved with the hybrid design using control parents where the analogous tests yielded power of 0.41, 0.53, 0.68 and 0.79, respectively [Weinberg and Umbach, 2005].

We examined our ability to detect mating asymmetry for the proposed hybrid design using a sample size of 150 case-families and 150 control families from a population with no population structure and null relative risks. For both the one- and the three-df tests, we verified that noncentrality parameters were zero under mating symmetry, as expected. We considered two specific scenarios where mating was not symmetric: one where only the three-df test would have power; a second where both the one- and the three-df tests would (Table II). The patterns of asymmetry that we introduced for purposes of illustration were arbitrary and possibly unrealistic; we had no population-based data on mating asymmetry to guide us. For the mating-asymmetry scenario of Table II, column 3, the asymmetry-induced bias for detecting relative risks using a model incorrectly assuming symmetry was small (actual α -level = 0.055 at nominal α = 0.05). The one-df test had no power, as constructed; but the power of the three-df test was 0.90. For the asymmetry scenario of Table II, column 4, the asymmetry-induced bias for detecting relative risks using a model incorrectly assuming symmetry was large (actual α -level = 0.57 at nominal α = 0.05). The three-df test for asymmetry had power 0.86 whereas the one-df test had power 0.41.

Discussion

The popularity of family-based designs is due, in part, to their robustness to the insidious bias that genetic population structure can induce in population-based case-control comparisons. Population-based case-control designs, however, are superior for studying exposure effects on risk. The original motivation for combining family-based and population-based components was to use information from both sources to increase power for studying offspring genetic effects [Nagelkerke et al., 2004]. Family-based designs achieve robustness by conditioning on parental genotypes but thereby sacrifice information parental genotypes can offer about genetic risks. Any allele that increases risk in the offspring should be over-represented in the parents of cases compared to the parents of controls. Hybrid designs use their population-based component to provide additional information that can improve inference on genetic risks. The possibility that genetic population structure might bias a combined analysis could be addressed at best obliquely when the population-based component included only control subjects [Epstein et al., 2005]. Hybrid designs that use control parents or control-mother dyads to enhance power offer a more straightforward ability to check for bias due to population structure while preserving the option of dropping back to the case-parents component when population structure is detected.

The primary motivation for considering a hybrid design that uses control-mother dyads instead of a hybrid design that uses parents of controls is a practical one: recruiting mothers of controls is easier than recruiting both mothers and fathers. Thus, larger sample sizes should be easier to achieve and possible bias from selective refusals of fathers (or of mothers with concerns about paternity) can be reduced. On the other hand, control-mother pairs provide less direct information about mating-type parameters in the underlying population than do control parent pairs. This feature is reflected in our results on power. The hybrid with control parents provided slightly more power for testing offspring and maternal genetic effects than did the hybrid with control-mother dyads.

In practice, the best design might be a hybrid of the two hybrid designs considered here. One could plan to recruit the parents of controls but, when the father is unavailable, try to recruit the control-mother dyad instead. Again, missing-data methods like the EM algorithm can be used for inference. Presumably the power for such a design would correspond to noncentrality parameters that fall between the curves of the two hybrid designs (Figure 1). This hybrid-of-hybrids design would be tantamount to a design where case-parent triads are augmented by control-parent triads but missing data is dealt with in the analysis. The only advantage of

genotyping control offspring when both parents are available, however, is the capacity to test the basic assumption of Mendelian transmission, an assumption that is not often questioned.

For studying both offspring and maternal genetic effects on disease risk, hybrid designs that use control parents or control-mother dyads offer additional benefits over the case-parents design. The case-parents design is capable of studying the maternal genetic effects under the assumption of mating symmetry in the source population but offers no way to check that assumption. Hybrid designs provide information not only for checking that assumption but also for studying maternal genetic effects even when mating symmetry is not satisfied. These features were delineated previously for the hybrid that uses control parents [Weinberg and Umbach, 2005], and we have shown here that they are maintained when control-mother pairs are substituted for control parents. Two different phenomena can masquerade as maternal genetic effects if not properly accounted for: mating asymmetry in the underlying population; and imprinting, where the effect of a transmitted allele depends on the parent of origin. After detecting a maternal effect, the investigator should ideally exclude the possibility that it was wholly or partly attributable to one of these other sources. With a case-parents design, mating symmetry in the underlying population must remain an assumption. The case-mother/control-mother design implicitly tolerates mating asymmetry. With either hybrid design, mating symmetry can be checked; and, if rejected, maternal effects can be studied under a model that accommodates mating asymmetry. Imprinting can be accommodated by the case-parents design using a log-linear model that incorporates parent-of-origin effects [Weinberg, 1999b]. Although we have not considered it here, this same modeling tactic could be adapted for the hybrid designs or for the case-mother/control-mother design as well.

An appealing feature of hybrid designs that involve either parents of controls or control-mother pairs is the ability to check key assumptions. The corresponding weakness is that these tests, particularly the key test for bias from population structure, appear to have limited power. For example, with 150 case and control families, we had little power to detect bias from population structure under a scenario which, if it went undetected, would induce substantial bias in inference. Power seemed somewhat better for detecting mating asymmetry in our examples; but we could certainly produce other examples where power was more limited. Larger sample sizes would help, of course. With a hybrid design, a simple practical precaution is to compare risk estimates under models that do and do not make a particular assumption. With concern about population structure, comparing relative risk estimates from the hybrid design with those from its case-parent triad component alone might provide reassurance. Similarly, with concern about mating symmetry, one could compare estimates from the mating-symmetry and the mating-asymmetry models.

The hybrid designs that we have considered will be useful for studies of complex phenotypes such as a birth defect where environmental factors will likely also play a role. The log-linear model for the 30-cell multinomial for case triads and control triads together can in principle be extended to incorporate environmental factors assessed categorically. Consequently, environmental risks and gene-by-environment interactions can be evaluated under either hybrid design. Postulating appropriate models for checking bias from population structure might be challenging, however, because exposure prevalence could change across genetically distinct subpopulations.

We have also demonstrated the potential increase in power that the hybrid designs can achieve when bias from population stratification is absent. The relative gain in efficiency (as measured by the ratio of the curves) depends on the underlying scenario. We based these power comparisons on equal numbers of case families in each design reflecting that the number of case families available is often a limitation when designing a study. The hybrid designs considered here require genotyping five individuals per case compared to only three or four,

respectively, for the case-parents or the case-mother/control-mother designs. To compare designs when each of them genotyped 750 individuals (the number measured in the hybrid designs in Figure 1), one would modify Figure 1 by multiplying the noncentrality curves for the case-parents design by 5/3 and those for case-mother/control-mother design by 5/4 (curves for hybrid designs remain the same). On that per-genotype basis, hybrid designs were not the most efficient in every scenario that we considered; however, they were more efficient than case-mother/control-mother designs in general and more efficient than the case-parents design in realistic scenarios where genotypes were missing at random. In our view, any disadvantage in power per genotype of hybrid designs compared to the case-parents design is more than offset by the greater flexibility that hybrid designs offer for checking assumptions and evaluating effects of exposures. The cost of genotyping should not be the investigator's sole concern.

In summary, we have studied a hybrid design that combines data from case-parent triads and control-mother dyads. It provides the same flexibility as, though slightly less power than, a hybrid design that was introduced previously, one that uses the parents of controls instead of control-mother dyads. When needed but testable assumptions are met, either of these hybrid designs has better power for studying both offspring and maternal genetic effects on disease risk than would a family-based case-parents design or a population-based case-mother/control-mother design.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (NIH Z01-ES040007) and by the Ter Meulen Fund. We thank Drs. Abee Boyles and Gregg Dinse for their helpful comments.

References

- Agresti, A. *Categorical Data Analysis*. New York: John Wiley & Sons; 1970.
- Dempster AP, Laird NM, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Statist Soc B* 1977;39:1–38.
- Epstein MP, Veal CD, Trembath RC, Barker JN, Li C, Satten GA. Genetic association analysis using data from triads and unrelated subjects. *Am J Hum Genet* 2005;76:592–608. [PubMed: 15712104]
- Nagelkerke NJD, Hoebee B, Teunis P, Kimman TG. Combining the transmission disequilibrium test and case-control methodology using generalized logistic regression. *Eur J Hum Genet* 2004;12:964–970. [PubMed: 15340361]
- Sinsheimer JS, Palmer CGS, Woodward JA. Detecting genotype combinations that increase risk for disease: the maternal-fetal genotype incompatibility test. *Genet Epidemiol* 2003;24:1–13. [PubMed: 12508251]
- Sun F, Flanders WD, Yang Q, Khoury MJ. Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. *Am J Epidemiol* 1999;150:97–104. [PubMed: 10400559]
- Van den Oord EJ, Vermunt JK. Testing for linkage disequilibrium, maternal effects, and imprinting with (in)complete case-parent triads, by use of the computer program LEM. *Am J Hum Genet* 2000;66:335–338. [PubMed: 10631165]
- Weinberg CR, Wilcox AJ, Lie RT. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 1998;62:969–978. [PubMed: 9529360]
- Weinberg CR. Allowing for missing parents in genetic studies of case-parent triads. *Am J Hum Genet* 1999a;64:1186–1193. [PubMed: 10090904]
- Weinberg CR. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am J Hum Genet* 1999b;65:229–235. [PubMed: 10364536]
- Weinberg CR, Umbach DM. A hybrid design for studying genetic influences on risk of diseases with onset early in life. *Am J Hum Genet* 2005;77:627–636. [PubMed: 16175508]

Wilcox AJ, Weinberg CR, Lie RT. Distinguishing the effects of maternal and offspring genes through studies of “case-parent triads”. *Am J Epidemiol* 1998;148:893–901. [PubMed: 9801020]

Figure 1.

Chi-squared noncentrality parameters and power as a function of allele prevalence for various designs and risk scenarios. Vertical axes: left, noncentrality parameter for a four-df likelihood ratio test of the null hypothesis $R_1=R_2=S_1=S_2=1$ based on 150 case families and 150 control families; right, power of corresponding test when $\alpha=0.05$. Left column (panels a–d): no missing genotypes; right column (panels e–h): 20% missing genotypes. First row (panels a, e): $R_1=2$, $R_2=3$, $S_1=1$, $S_2=1$; second row (panels b, f): $R_1=1$, $R_2=1$, $S_1=2$, $S_2=3$; third row (panels c, g): $R_1=1$, $R_2=2$, $S_1=1$, $S_2=3$; fourth row (panels d, h): $R_1=1$, $R_2=3$, $S_1=2$, $S_2=2$. Designs: hybrid with control-mother dyads under mating symmetry (gray solid) or mating asymmetry (gray dash); hybrid with control parents under mating symmetry (black solid) or mating asymmetry (black dash).

(black dash); case-parent triads (black dot); case-mother/control-mother (black dot-dash). Each panel has curves that nearly or completely coincide: panel *a* (black dot, black dot-dash); panels *a*, *e* (black solid, black dash) and (gray solid, gray dash); panels *b*, *f* (black dash, gray dash, black dot, black dot-dash); panels *c*, *g* (black dash, gray dash); panels *d*, *h* (gray solid, gray dash).

Table 1
Expected counts of case-parent triads and control-mother dyads under mating asymmetry or mating symmetry.

MF ^a	C ^b	affected	mating asymmetry		mating symmetry	
			mating types	expected count ^c	mating types	expected count ^c
00	0	yes	00	B μ_{00}	00	B μ_{00}
01	0	yes	01	B $\mu_{01}/2$	01/10	B $\mu_{01}/4$
01	1	yes	01	B $\mu_{01}R_1/2$	01/10	B $\mu_{01}R_1/4$
10	0	yes	10	B $\mu_{10}S_1/2$	01/10	B $\mu_{10}S_1/4$
10	1	yes	10	B $\mu_{10}R_1S_1/2$	01/10	B $\mu_{01}R_1S_1/4$
02	1	yes	02	B $\mu_{02}R_1$	02/20	B $\mu_{02}R_1/2$
20	1	yes	20	B $\mu_{20}R_1S_2$	02/20	B $\mu_{02}R_1S_2/2$
11	0	yes	11	B $\mu_{11}S_1/4$	11	B $\mu_{11}S_1/4$
11	1	yes	11	B $\mu_{11}R_1S_1/2$	11	B $\mu_{11}R_1S_1/2$
11	2	yes	11	B $\mu_{11}R_2S_1/4$	11	B $\mu_{11}R_2S_1/4$
12	1	yes	12	B $\mu_{12}R_1S_1/2$	12/21	B $\mu_{12}R_1S_1/4$
12	2	yes	12	B $\mu_{12}R_2S_1/2$	12/21	B $\mu_{12}R_2S_1/4$
21	1	yes	21	B $\mu_{21}R_1S_2/2$	12/21	B $\mu_{12}R_1S_2/4$
21	2	yes	21	B $\mu_{21}R_2S_2/2$	12/21	B $\mu_{12}R_2S_2/4$
22	2	yes	22	B $\mu_{22}R_2S_2$	22	B $\mu_{22}R_2S_2$
0- ^d	0	no	00 or 01	$\mu_{00} + \mu_{01}/2$	00 or 01/10	$\mu_{00} + \mu_{01}/4$
0-	1	no	01 or 02	$\mu_{01}/2 + \mu_{02}$	01/10 or 02/20	$\mu_{01}/4 + \mu_{02}/2$
1-	0	no	10 or 11	$\mu_{10}/2 + \mu_{11}/4$ ^e	01/10 or 11	$\mu_{01}/4 + \mu_{11}/4$ ^e
1-	1	no	10 or 11 or 12	$\mu_{10}/2 + \mu_{11}/2 + \mu_{12}/2$ ^e	01/10 or 11 or 12/21	$\mu_{01}/4 + \mu_{11}/2 + \mu_{12}/4$ ^e
1-	2	no	11 or 12	$\mu_{11}/4 + \mu_{12}/2$ ^e	11 or 12/21	$\mu_{11}/4 + \mu_{12}/4$ ^e
2-	1	no	20 or 21	$\mu_{20} + \mu_{21}/2$	02/20 or 12/21	$\mu_{02}/2 + \mu_{12}/4$
2-	2	no	21 or 22	$\mu_{21}/2 + \mu_{22}$	12/21 or 22	$\mu_{12}/4 + \mu_{22}$

^aCopies of variant allele carried by the mother and the father.

^bCopies of variant allele carried by the child.

^c“B” is a normalizing factor whose inclusion ensures that the total fitted count for case triads matches the corresponding total observed count and, similarly, for control dyads. Under mating symmetry, stratification parameter μ_{mf} includes the stratum with m and f reversed.

$d_{i..}$ is used to denote a missing paternal genotype.

^eWhen $M=1$, the expected counts in cells where $C=0$ and $C=2$ sum to the expected count in the cell where $C=1$.

Table II

Arbitrary mating-type configurations used to examine the power to detect mating asymmetry.

Mating-type Parameter	Parameter value: proportion of families in each mating type.		
	Under mating symmetry: HWE with $p=0.5$ ^a	Under arbitrary mating asymmetry	
		$\frac{1}{2}\mu_{01}+\frac{1}{2}\mu_{12}+\mu_{02} = \frac{1}{2}\mu_{10}+\frac{1}{2}\mu_{21}+\mu_{20}$	$\frac{1}{2}\mu_{01}+\frac{1}{2}\mu_{12}+\mu_{02} \neq \frac{1}{2}\mu_{10}+\frac{1}{2}\mu_{21}+\mu_{20}$
μ_{00}	0.0625	0.0625	0.0625
μ_{01}	0.1250	0.0950	0.0950
μ_{10}	0.1250	0.1550	0.1550
μ_{02}	0.0625	0.0875	0.0375
μ_{20}	0.0625	0.0375	0.0875
μ_{11}	0.2500	0.2500	0.2500
μ_{12}	0.1250	0.1050	0.1050
μ_{21}	0.1250	0.1450	0.1450
μ_{22}	0.0625	0.0625	0.0625

^aHardy-Weinberg equilibrium with allele frequency $p=0.5$.