Adoption Studies

Adoption usually refers to the rearing of a nonbiological child in a family. This practice is commonplace after wars, which leave many children orphaned, and is moderately frequent in peacetime. Approximately 2% of US citizens are adoptees.

Historically, adoption studies have played a prominent role in the assessment of genetic variation in human and animal traits [10]. Most early studies focused on cognitive abilities [9], but there is now greater emphasis on psychopathology [5] and physical characteristics, such as body mass [20]. Adoption studies have made major substantive contributions to these areas, identifying the effects of genetic factors where they were previously thought to be absent [3, 12, 15].

In recent years the adoption study has been overshadowed by the much more popular twin study [17] (see Twin Analysis). Part of this shift may be due to the convenience of twin studies and the complex ethical and legal issues involved in the ascertainment and sampling of adoptees. Certain Scandinavian countries—especially Denmark, Sweden and Finland [8, 13, 14]—maintain centralized databases of adoptions and thus have been able to mount more representative and larger adoption studies than elsewhere.

The adoption study is a "natural experiment" that mirrors cross-fostering designs used in genetic studies of animals, and therefore has a high face validity as a method to resolve the effects of genes and environment on individual differences. Unfortunately, the adoption study also has many methodological difficulties. First, is the need to maintain confidentiality, which can be a problem even at initial ascertainment, as some adoptees do not know that they are adopted. Recent legal battles for custody fought between biological and adoptive parents make this a more critical issue than ever. Secondly, in many substantive areas, e.g. psychopathology, there are problems with sampling, in that neither the biological nor the adoptive parents can be assumed to be a random sample of parents in the population. For example, poverty and its sequelae may be more common among biological parents who have their children adopted into other families than among parents who rear their children themselves. Conversely, prospective adoptive parents are, on average, and through self-selection, older and less fertile than biological parents. In addition, they
are often carefully screened by adoption agencies, and may be of higher socio-economic status than nonadoptive parents. Statistical methods (see below) may be used to control for these sampling biases if a random sample of parents is available. Some studies indicate that adoptive and biological parents are quite representative of the general population for demographic characteristics and cognitive abilities [19], so this potential source of bias may not have substantially affected study results.

Thirdly, selective placement is a common methodological difficulty. For statistical purposes, the ideal adoption study would have randomly selected adoptees placed at random into randomly selected families in the population. Often there is a partial matching of the characteristics of the adoptee (e.g. hair and eye color, religion and ethnicity) to those of the adopting family. This common practice may improve the chances of successful adoption. Statistically, it is necessary to control for the matching as far as possible. Ideally, the matching characteristics used should be recorded and modeled. Usually, such detailed information is not available, so matching is assumed to be based on the variables being studied and modeled accordingly (see below). In modern adoption studies, these methods are used often [18, 19].

Types of Adoption Study

Nuclear families in which at least one member is not biologically related to the others offer a number of potential comparisons that can be genetically informative (see Table 1). Of special note are monozygotic (MZ) twins reared apart (MZ\textsubscript{A}) (see Zygosity Determination). Placed into uncorrelated environments, the correlation between MZ twins directly estimates the proportion of variation due to all genetic sources of variance ("broad heritability"). Estimation of heritability in this way is statistically much more powerful than, e.g. the classical twin study that compares MZ and dizygotic (DZ) twins reared together (MZ\textsubscript{T} and DZ\textsubscript{T}). With MZ\textsubscript{A} twins the test for heritability is a test of the null hypothesis that the correlation is zero, whereas the comparison of MZ\textsubscript{T} and DZ\textsubscript{T} is a test of a difference between correlations. Environmental effects shared by members of a twin pair (known as "common", "shared" or "family" environment or "C") are excluded by design. If this source of variation is of interest, then additional groups of relatives, such as unrelated individuals reared together, are needed to estimate it. Similar arguments may be made about across-generational sources of resemblance. Heath & Eaves [11] compared the power to detect genetic and environmental transmission across several twin-family (twins and their parents or twins and their children) adoption designs.

<table>
<thead>
<tr>
<th>Relationship</th>
<th>$V_A$</th>
<th>$V_D$</th>
<th>$V_{AA}$</th>
<th>$V_{AD}$</th>
<th>$V_{DD}$</th>
<th>$E_S$</th>
<th>$E_P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP-BC</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BP-AC</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AP-AC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AC-BC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BC-BC\textsubscript{T}</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BC-BC\textsubscript{A}</td>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MZ\textsubscript{T}-MZ\textsubscript{T}</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MZ\textsubscript{A}-MZ\textsubscript{A}</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1 Coefficients of genetic and environmental variance components quantifying resemblance between adopted and biological relatives, assuming random sampling, mating and placement.

Variables: $V_A$ = additive genetic; $V_D$ = dominance genetic; $V_{AA}$ = additive × additive interaction; $V_{AD}$ = additive × dominance interaction; $V_{DD}$ = dominance × dominance interaction; $E_S$ = environment shared by siblings; $E_P$ = environment shared or transmitted between parent and child. Relationships are: MZ = monozygotic twin; DZ = dizygotic twin; BP = biological parent; BC = biological child; AP = adoptive parent; AC = adopted child. The subscripts T and A refer to reared together and reared apart, respectively.

Methods of Analysis

Most modern adoption study data are analyzed with Structural Equation Models (SEM) [2, 17]. SEM is an extension of multiple linear regression analysis that involves two types of variable: observed variables that have been measured, and latent variables that have not. Two variables may be specified as causally related or simply correlated from unspecified effects. It is common practice to represent the variables and their relationships in a path diagram (see Path Analysis in Genetics), where single-headed arrows indicate causal relationships, and double-headed arrows represent correlations. By convention, observed variables are shown as squares and latent variables are shown as circles.
Figure 1 shows the genetic and environmental transmission from biological and adoptive parents to three children. Two of the children are offspring of the biological parents (sibs reared together) while the third is adopted. This diagram may also be considered as multivariate, allowing for the joint analysis of multiple traits. Each box and circle then represents a vector of observed variables. Multivariate analyses (see Multivariate Analysis, Overview) are particularly important when studying the relationship between parental attributes and outcomes in their offspring. For example, harsh parenting may lead to psychiatric disorders. Both variables should be studied in a multivariate genetically informative design such as an adoption or twin study to distinguish between the possible direct and indirect genetic and environmental pathways.

From the rules of path analysis [22, 23] we can derive predicted covariances among the relatives, in terms of the parameters of the model in Figure 1. These expectations may, in turn, be used in a structural equation modeling program such as Mx [16] to estimate the parameters using maximum likelihood or some other goodness-of-fit function. Often, simpler models than the one shown will be adequate to account for a particular set of data.

A special feature of the diagram in Figure 1 is the dotted lines representing delta-paths [21]. These represent the effects of two possible types of selection: assortative mating, in which husband and wife correlate; and selective placement, in which the adoptive and biological parents are not paired at random. The effects of these processes may be deduced from the Pearson–Aiken selection formulas [1]. These formulas are derived from linear regression under the assumptions of multivariate linearity and homoscedasticity. If we partition the variables into selected variables, $X_S$, and unselected variables $X_N$.
then it can be shown that changes in the covariance of $X_N$ lead to changes in covariances among $X_N$ and the cross-covariances ($X_S$ with $X_N$). Let the original (pre-selection) covariance matrix of $X_S$ be $A$, the original covariance matrix of $X_N$ be $C$, and the covariance between $X_N$ and $X_S$ be $B$. The preselection matrix may be written

$$\begin{pmatrix} A & B \\ B' & C \end{pmatrix}.$$

If selection transforms $A$ to $D$, then the new covariance matrix is given by

$$\begin{pmatrix} D \\ B' A^{-1} B \\ C - B' (A^{-1} - A^{-1} D A^{-1}) B \end{pmatrix}.$$

Similarly, if the original means are $(x_0 : x_0)'$ and selection modifies $x_0$ to $\bar{x}_0$, then the vector of means after selection is given by

$$(x_0 : x_0 + A^{-1} B (x_0 - \bar{x}_0) )'.$$

These formulas can be applied to the covariance structure of all the variables in Figure 1. First, the formulas are applied to derive the effects of assortative mating, and secondly, they are applied to derive the effects of selective placement. In both cases, only the covariances are affected, not the means. An interesting third possibility would be to control for the effects of nonrandom selection of the biological and adoptive relatives, which may well change both the means and the covariances.

**Selected Samples**

A common approach in adoption studies is to identify members of adoptive families who have a particular disorder, and then examine the rates of this disorder in their relatives (see *Ascertainment*). These rates are compared with those from control samples. Two common starting points for this type of study are (i) the adoptees (the adoptees’ families method) and (ii) the biological parents (the adoptee study method). For rare disorders, this use of selected samples may be the only practical way to assess the impact of genetic and environmental factors.

One limitation of this type of method is that it focuses on one disorder, and is of limited use for examining co-morbidity between disorders. This limitation is in contrast to the population-based sampling approach where many characteristics – and their covariances or co-morbidity – can be explored simultaneously.

A second methodological difficulty is that ascertained samples of the disordered adoptees or parents may not be representative of the population. For example, those attending a clinic may be more severe or have different risk factors than those in the general population who also meet criteria for diagnosis, but do not attend the clinic.

**Genotype $\times$ Environment Interaction**

The natural experiment of an adoption study provides a straightforward way to test for gene-environment interaction. In the case of a continuous phenotype, interaction may be detected with linear regression on

1. the mean of the biological parents’ phenotypes (which directly estimates heritability)
2. the mean of the adoptive parents’ phenotypes
3. the product of points 1 and 2.

Significance of the third term would indicate significant $G \times E$ interaction. With binary data such as psychiatric diagnoses, the rate in adoptees may be compared between subjects with biological or adoptive parents affected, vs. both affected. $G \times E$ interaction has been found for alcoholism [7] and substance abuse [6].

**Logistic regression** is a popular method to test for genetic and environmental effects and their interaction on binary outcomes such as psychiatric diagnoses. These analyses lack the precision that structural equation modeling can bring to testing and quantifying specific hypotheses, but offer a practical method of analysis for binary data. Analysis of binary data can be difficult within the framework of SEM, requiring either very large sample sizes for asymptotic weighted least squares [4] or integration of the multivariate normal distribution over as many dimensions as there are relatives in the pedigree, which is numerically intensive.

**References**

Adverse Selection

Some health care providers (e.g., clinicians, hospitals, or Health Maintenance Organizations (HMOs)) and insurers serve populations that are substantially sicker or more difficult to care for than average. Such a provider or insurer is said to experience adverse selection; one with healthier than average patients has favorable selection. A provider suffering adverse selection will tend to have higher costs per patient and worse health outcomes per patient. Such providers may not be paid enough to cover costs of the care they provide in systems where providers are paid a fixed price for each patient, or where they are penalized for expending more resources than their peers for patient care. Action taken by the provider to achieve favorable selection is called “skimming” or “cream skimming”. Biased selection (either favorable or adverse) is not necessarily a problem if such differences are recognized and accounted for when paying for care or holding providers accountable for their patients’ utilization or health outcomes.

Bibliography