This is chapter 3: Effect sizes (pp. 19-32) from

Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Hogrefe & Huber.

3
Effect Sizes

In the following sections the type of data used for meta-analysis, the so-called effect sizes, and their statistical characteristics are introduced. The focus here will be laid on measures of effect sizes that are typical research outcomes in the social and behavioral sciences, especially in the field of psychology. Other measures will only be treated in passing. The statistical models to be applied to this form of data will then be presented in detail followed by sections on specific statistical approaches to meta-analysis that can count as the most common applied in research up to date.

The study of a series of research articles in any field of psychology still leaves the impression that the main goal of applying statistical methods is predominantly testing of null hypotheses (Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000). This seems surprising given the high information value ordinarily attached to effect sizes (but see Chow, 1988), and policies articulated by large psychological organizations — like the American Psychological Association — are clearly in favor of reporting effect sizes in research articles (American Psychological Association, 2001; Wilkinson & Task Force on Statistical Inference, 1999). However, a recent study on the editorial policies of the reporting practices has revealed that these policies still have not been fully adopted by editors of major research journals in psychology (Vacha-Haase et al., 2000). Hence, encouragement to report effect sizes is not translated into action. Yet there are also reasons to believe that simple calls for the reporting of effect sizes in publications may not be sufficient to eliminate bias of published results (Lane & Dunlap, 1978).

The lack of reporting effect sizes poses a problem for the meta-analyst because the data to be analyzed are often not readily available from study reports. As a result, effect sizes have to be extracted from research reports when sufficient information is available. There is a host of publications which illustrate that this aim may not always easily be achieved (see, e.g., Olejnik & Algina, 2000; Rosnow & Rosenthal, 1996; Seifert, 1991). Furthermore, design characteristics also have to be taken into account when extracting an effect size, otherwise wrong measures may result (Dunlap, Cortina, Vaslow, & Burke, 1996; Morris & DeShon, 1997). In a reanalysis of 140 studies on psychosocial treatments or prevention studies in psychology Ray and Shadish (1996) have shown that different techniques to extract effect size information, proposed in the literature, lead to different magnitudes of effect sizes. Moreover, Matt (1989) has shown that judgmental factors in extracting effect sizes also play an important role for the establishment of a database for meta-analysis. In sum, there remain several problems in extracting the relevant information for effect sizes in some areas of research. The techniques for aggregation of effect sizes, to be introduced, presume that there is a database of effect sizes already available and problems of the form just described are not of relevance.

The next section provides an overview of certain families of effect size measures that are most common in psychology. The focus here will be laid on the correlation coefficient as an effect size. A second common measure, the standardized mean difference, will also be considered. These two effect size measures are by far not the only available to researchers, but they are those of highest importance for the present purpose of evaluating meta-analytical approaches in psychology.

The effect sizes of interest in the present context belong to two families, the r and the d family (Rosenthal, 1994). In short, they are comprised of correlation coefficients on the one hand and standardized mean differences on the other. They can both be characterized by one of the main features of effect sizes, the provision of a standardized measure for an effect of interest. First, focus will be on the correlation coefficient.

3.1 CORRELATION COEFFICIENTS AS EFFECT SIZES

The sample correlation coefficient r, usually designated as the Pearson product moment correlation, is based on n pairs (x_o, y_o), $o = 1, \ldots, n$, of observations and is given by

$$r = \frac{\sum_{o=1}^{n} (x_o - \overline{x})(y_o - \overline{y})}{\sqrt{\sum_{o=1}^{n} (x_o - \overline{x})^2} \sqrt{\sum_{o=1}^{n} (y_o - \overline{y})^2}}.$$

The corresponding pair of random variables for (x_o, y_o) is (X, Y). Here, and in what follows, it should be noted that the correlation coefficient can also be considered as a random variable based on the variates X and Y. This will occasionally be highlighted in the following by the symbol R, although the symbol r will predominantly be used. It will be clear from the context when r should be understood as a random variable and when it should only be considered as a sample statistic.

The corresponding population correlation coefficient ρ is given by

$$\rho = \frac{E\left(\left(X - E[X]\right)\left(Y - E[Y]\right)\right)}{\sigma_X \sigma_Y}.$$

As is easily seen from these equations, the correlation coefficient can also be regarded as the covariance of standardized variables. Hence, it is extremely useful for measurement at the interval scale level because it is invariant with respect to — not necessarily the same — positive linear transformations of the variables. This is exactly what is most frequently intended by the computation of effect sizes in psychology, to express an effect free from the influence of specific standard deviations of measurement instruments. What is commonly viewed in the behavioral sciences as an advantage of effect sizes, namely to represent the size of an effect irrespective of the scale it is measured on, has also raised questions about the meaning of the resulting scale-free measures (Feinstein, 1995).

The question arises how the variate r may be distributed. Fortunately, it is well-known that r is approximately normally distributed with *large* samples. However, convergence of the distribution is very slow and it is said to be unwise to assume it for $n < 500$ (Stuart, Ord, & Arnold, 1999, p. 481), a case most frequently encountered in practice. The distribution of r is a very complicated statistical topic that cannot be fully dealt with here (for overviews, see Johnson, Kotz, & Balakrishnan, 1995; Stuart & Ord, 1994). The focus of the following presentation will therefore be on aspects of importance to meta-analysis. That is, first, the distribution of r when the pair (X, Y) follows a bivariate normal distribution, and second, point estimation of ρ.

The exact probability density function (PDF) of the distribution of R for r in the interval $[-1, 1]$ is given in the seminal paper by Hotelling (1953, p. 200) as[1]

$$\begin{aligned} p_R(r) = \frac{\mathrm{df}}{\pi\sqrt{2}}(1-r^2)^{\frac{\mathrm{df}}{2}-1}(1-r\rho)^{-\mathrm{df}-\frac{1}{2}}(1-\rho^2)^{\frac{\mathrm{df}+1}{2}} \\ \times \mathrm{B}(\mathrm{df}+1, \tfrac{1}{2})\ {}_2F_1(\tfrac{1}{2}, \tfrac{1}{2}; \mathrm{df}+\tfrac{3}{2}; \tfrac{1}{2}(r\rho+1)) \end{aligned} \tag{3.1}$$

where B denotes the complete Beta function, $\mathrm{df} = n - 2$, and ${}_2F_1$ is the Gaussian hypergeometric function:

$${}_2F_1(a_1, a_2; a_3; a_4) = \sum_{v=0}^{\infty} \frac{\Gamma(a_1+v)\Gamma(a_2+v)\Gamma(a_3)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3+v)} \frac{a_4^v}{v!}.$$

In this formula, Γ represents the Euler Γ function. There are also some different forms of the density to be found in the literature based on different derivations

[1]Note that the exact form given by Hotelling differs somewhat from the form given here which better fits in the notation already introduced. The equivalence between both forms is seen by noting that n is in Hotelling's paper the symbol for the degrees of freedom and by noticing the following equivalencies: $\mathrm{B}(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, $\Gamma(0.5) = \sqrt{\pi}$, and $\Gamma(a+1) = a\Gamma(a)$.

(see Johnson, Kotz, & Balakrishnan, 1995), but the presentation here will focus on the one provided in Equation 3.1.

Obviously, using this distribution for tests of correlation coefficients is not feasible when $\rho \neq 0$. Only for $\rho = 0$ is r distributed as t with $n - 2$ degrees of freedom, and hence tractable. However, interest in the present context also lies on the nonnull distribution of r. For testing purposes, various approximations to the distribution of the correlation coefficient have been proposed. A series of these approximations will be presented as well as evaluated. Results are reported in Subsection 7.5.1 of Chapter 7, so that a discussion of most of these approximations is postponed. Nevertheless, one of these approximations will be introduced at this point because of its high relevance for the approaches of meta-analysis to be presented in Chapter 5.

The most popular approximation was given by Fisher (1921)[2] who also derived the distribution in the bivariate normal case (Fisher, 1915). He suggested the following transformation to be applied to the correlation coefficient

$$z = \tanh^{-1} r = \frac{1}{2} \ln \frac{1+r}{1-r}. \tag{3.2}$$

In analogy to the case of the correlation coefficient, z can be considered as a random variable Z, but it will be denoted by a lowercase z in most of what follows. The corresponding transformation for the population correlation ρ is

$$\zeta = \tanh^{-1} \rho = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}.$$

As an inverse transformation

$$r = \tanh z = \frac{\exp(2z) - 1}{\exp(2z) + 1} \tag{3.3}$$

is specified for the correlation coefficient, and, again, a corresponding transformation for the population correlation given by

$$\rho = \tanh \zeta = \frac{\exp(2\zeta) - 1}{\exp(2\zeta) + 1}.$$

What happens to the correlation coefficients when the Fisher-z transformation is applied? In Figure 3.1, the transformation provided by Fisher is illustrated. As can be seen, the transformation stretches the values in the boundary regions. Furthermore, the possible values of z are not bounded by -1 and 1, as is the case for the correlation coefficient. Instead, they span the whole interval $[-\infty, +\infty]$.

What are the main virtues of applying the Fisher-z transformation to correlation coefficients? First, the transformed correlation coefficient (z) is approximately normally distributed. That is, the result from stretching the values is to

[2]See also Konishi (1978, 1981) for a more concise derivation.

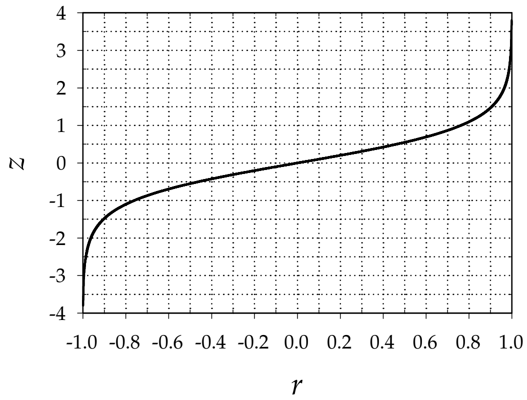

Figure 3.1 The r to Fisher-z transformation.

achieve an approximate normal distribution. In contrast to the distribution of r, the distribution of z converges to normality very much faster.

A second benefit of applying the transformation is stabilization of the variance. This can be seen from Equation 3.4 which gives the approximate variance of Z

$$\sigma_Z^2 \approx \frac{1}{n-3}. \tag{3.4}$$

The variance of z is stable in the sense that it does *not* depend on the parameter ζ but only on the sample size n. As will be seen, this highly desirable feature stands in contrast to the variance of r, which does depend on the population parameter.

The approximate variance of z can easily be computed in practical applications and used for tests as well as for the construction of approximate confidence intervals for ζ, a third beneficial aspect of the transformation. Construction of confidence intervals is easy because it is possible to draw on the normal distribution to find the interval limits. Having found the confidence limits for ζ, it is also possible to transform them to limits for ρ, a procedure that will be outlined in the context of presenting the various statistical approaches to meta-analysis in Chapter 5. The possibility to conduct a statistical test as well as to construct confidence intervals is the main reason why this approximation is so popular in practice.

In sum, by applying the transformation to the correlation coefficient it can be said that one changes spaces. That is, the examination of the linear relationship between two variables starts in the space of r with random variable R and population parameter ρ. The transformation leads to an examination in the space of z with random variable Z and population parameter ζ. The whole purpose of applying the transformation is to make inferences about the population parameter ρ by exploiting desirable properties in the space of z, and thereby avoiding to deal with the complicated PDF.

Apart from examining the distribution of r and its approximation, the question arises whether the Pearson product moment correlation constitutes an unbiased estimator of the effect. That is, whether the equality $E(r) = \rho$ holds for all ρ. This is not the case, and r is therefore a biased estimator of ρ. Hotelling (1953) provided the moments of r about ρ of which the first moment ($\Xi_1 = E(r - \rho)$) is given as

$$\Xi_1 = \left(1-\rho^2\right)\left(-\frac{\rho}{2(n-1)} + \frac{\rho - 9\rho^3}{8(n-1)^2} + \frac{\rho + 42\rho^3 - 75\rho^5}{16(n-1)^3} + \ldots\right)$$

(Hotelling, 1953, p. 212). The bias is usually approximated by truncation of the series, resulting in

$$\Xi_1 = -\frac{\rho(1-\rho^2)}{2(n-1)}.$$

This is the well known formula for the *negative* bias of r as an estimator of positive ρ. To compensate for this bias in r one could apply the following correction

$$r^* = r + \frac{r(1-r^2)}{2(n-1)},$$

which is almost identical to an approximation to the unique minimum variance unbiased (UMVU) estimator by Olkin and Pratt (1958) to be presented below (page 26).

Hotelling (1953) also provided the moments of z about ζ, of which, again, only the first moment is given here

$$\kappa_1 = \frac{\rho}{2(n-1)} + \frac{5\rho + 9\rho^3}{8(n-1)^2} + \frac{11\rho + 2\rho^3 + 3\rho^5}{16(n-1)^3} + \ldots$$

As is obvious, a *positive* bias of Fisher-z for positive ρ is present here. A question that was discussed in the literature of meta-analysis with correlations as effect sizes is which of the biases is smaller in absolute value. Whereas Hunter and Schmidt (1990) claimed to have shown a smaller absolute bias of r in comparison to z, Corey, Dunlap, and Burke (1998) reported results of a Monte Carlo study in which they found the opposite result. Using the formulae given by Hotelling and truncating the series, the biases of the two estimators can be evaluated. For a direct comparison, the biases resulting from the formula for κ_1 were transformed into the space of r by the inverse Fisher-z transformation given in Equation 3.3 and plugging in κ_1 for z. In Figure 3.2 the resulting biases are illustrated.

The bias of both r and z for ρ is shown across different values for ρ as well as sample sizes n in the left panel of Figure 3.2. As can be seen, the bias of both estimators vanishes at $\rho = 0$. The light surface in this graph depicts the the biases of r and the shaded surface those of z. With higher values of ρ the bias continuously increases for z, whereas the bias of r attains its maximum at $\rho \approx .583$ for positive ρ and at $\rho \approx -.583$ for negative ρ. The right panel provides absolute differences in biases with positive values indicating higher biases for z. All values of the difference surface indicate higher bias for z, except for $\rho = 0$, thus z has a larger approximate bias in comparison to r.

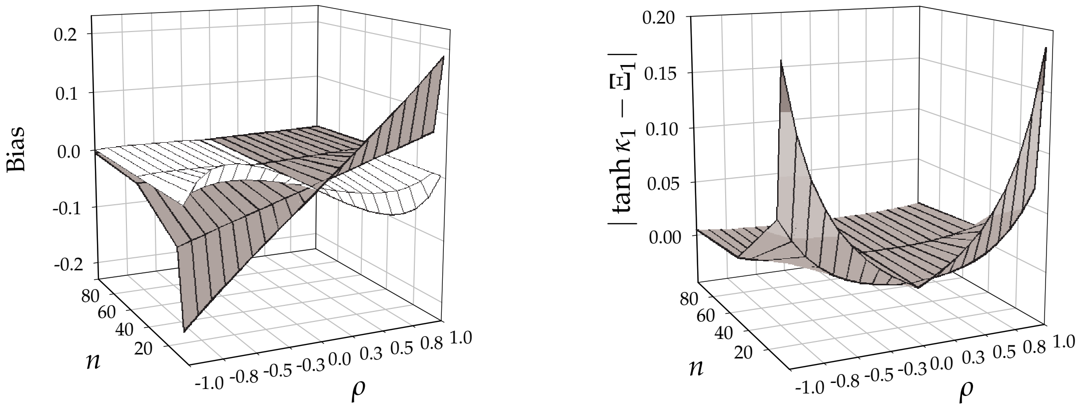

Figure 3.2 Bias of r and z in comparison. The left panel shows the bias for r (light surface) and z (shaded surface). The right panel shows the absolute difference surface for the biases.

In the methodological literature on meta-analysis based on correlational data, the use of Fisher-z versus r as estimators of ρ has attracted considerable attention, especially in the validity generalization literature (e.g., Corey et al., 1998; Law, 1995; Schmidt, Hunter, & Raju, 1988; Silver & Dunlap, 1987), and the bias of these statistics has been quite a controversial issue (see Hunter & Schmidt, 1990; James, Demaree, & Mulaik, 1986). As shown here, it is expected that Fisher-z will exhibit a larger bias from a theoretical point of view. In the Monte Carlo study to be presented in Chapter 7 and 8, it will be examined whether these expectations hold under the conditions of the simulation procedure.

Hotelling proposed several improvements of the Fisher-z transformation. First, he suggested the substraction of $r/(2n-3)$ from z when ρ is unknown to correct for its positive bias (Hotelling, 1953, p. 219). This correction was evaluated in a Monte Carlo study by Paul (1988), who concluded that for the estimation of $\rho < .50$ the modification of Hotelling performed best amongst the estimators he considered, and for $\rho > .50$ Fisher-z performed best. Alexander, Hanges, and Alliger (1985), in contrast, found no substantial differences between these estimators in their Monte Carlo study.

A further improvement was proposed by Hotelling (1953, p. 224) as

$$z^{**} = z - \frac{3z + r}{4(n-1)} - \frac{23z + 33r - 5r^3}{96(n-1)^2},$$

however the quality of this modification has not been sufficiently evaluated to date.

In contrast to these procedures, some authors in the methodological literature of meta-analysis, for example, Erez, Bloom, and Wells (1996, p. 288), and

Overton (1998, p. 358) used the correction

$$r^{\#} = r - \frac{r(1-r^2)}{2n}$$

to compensate for the positive bias in z. This correction was followed by an application of the Fisher-z transformation in both authors' work. Although this procedure obviously lowers the positive bias of z, it is of unclear origin and lacks a clear rationale from a statistical viewpoint. Because at least Erez et al. (1996) attributed the correction to Hotelling (1953), it may be speculated that (a variant of) Ξ_1 was used to correct the bias in z. How this flaw in procedure affects their results is however unclear.

As an important contribution to the statistical literature of estimators of ρ, the UMVU estimator was presented as

$$G = r \times {}_2F_1\left(\tfrac{1}{2}, \tfrac{1}{2}; \tfrac{n-2}{2}; 1-r^2\right) \tag{3.5}$$

by Olkin and Pratt (1958, p. 202). The following formula gives an approximation of G

$$G = r\left(1 + \frac{1-r^2}{2\,(n-1-3)}\right)$$

(Olkin & Pratt, 1958, p. 203). G has the same range and asymptotic distribution as r, but larger variance and smaller mean-squared error in general (Hedges & Olkin, 1985, p. 226). Surprisingly, although this estimator has very desirable properties from a statistical viewpoint, it is not widely used in the literature. This may be due to unawareness or due to statements in the literature that not much can be gained from an application of the correction of r (Hedges, 1989). It is expected from the statistical properties of this estimator that its usage will lead to a minimum bias among the estimators in the Monte Carlo study to be presented.

In addition to the bias of an estimator, its variance is also of great importance for meta-analysis. The variance of r is usually approximated in practice as

$$\sigma_R^2 \approx \frac{\left(1-\rho^2\right)^2}{n-1} \tag{3.6}$$

which is Ξ_2, the second moment about ρ presented by Hotelling (1953, p. 212) truncated after the first term in the series. In practice, this approximation is used by plugging in r for ρ in order to estimate the variance. This may, however, not be a good approximation. The reason for this is not only truncation, but most importantly the very slow convergence of the distribution of r to the normal distribution. As will be noticed, ρ (or in practical applications r) itself is involved in the variance approximation. In Figure 3.3 the dependency of σ_R^2 on ρ is illustrated. As can easily be seen in this figure, the variance is at max-

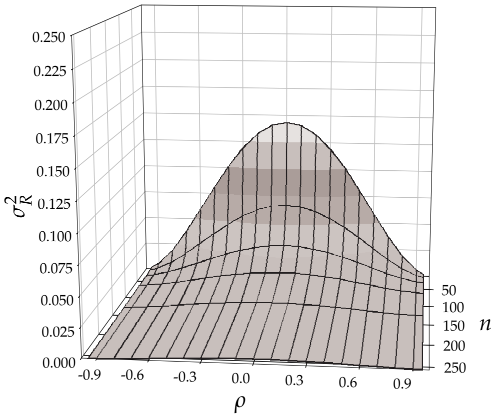

Figure 3.3 Variance of r across different values of ρ and n.

imum when $\rho = 0$, across all values of n. The variance changes maximally at $\rho \approx .577$ for positive values[3] of ρ.

The variance of G, in contrast, can be estimated by

$$\hat{\sigma}_G^2 = G^2 - 1 + \frac{(n-3)(1-r^2)_2F_1\left(1,1;\frac{n}{2};1-r^2\right)}{n-2} \tag{3.7}$$

(Hedges, 1988, p. 198; see also Hedges, 1989, p. 477). Again, the variance of the estimator is dependent on the parameter, though not as apparent as in the previous case. Figure 3.4 illustrates the relationship.

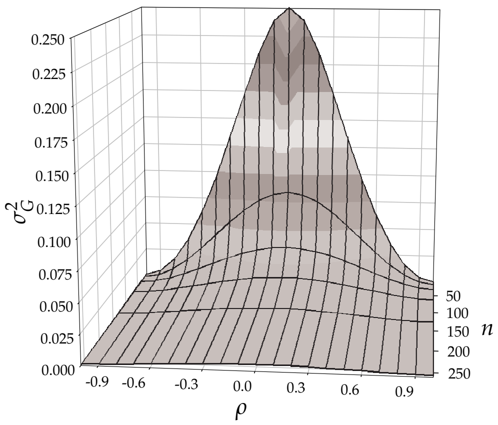

Figure 3.4 Variance of G across different values of ρ and n.

By way of comparison of Figures 3.3 and 3.4 it becomes clear that although the relationships are similar in form they are actually quite different with a stronger change in variance for G. The largest change in variance occurs at $\rho = .347$.

[3]This value results from taking the partial derivative of the variance and finding its minimum.

Up to this point, only the Pearson correlation coefficient has been examined, but there are several other correlation coefficients in the r family available (see Rosenthal, 1994; Rosenthal, Rosnow, & Rubin, 2000). The properties of other indices, like the point-biserial, biserial or rank correlation coefficient, for example, are not of concern here as only the correlation coefficient for the bivariate normal case is under scrutiny. For the distribution theory and examinations of the robustness of the coefficients reported in this book the reader is referred to Johnson, Kotz, and Balakrishnan (1995).

3.2 STANDARDIZED MEAN DIFFERENCES AS EFFECT SIZES

As previously mentioned, a second common effect size measure in the psychological literature is the standardized mean difference. It is mostly used in a situation when two groups of participants are examined and differences of means are of interest. More succinctly,

$$X_{o_1} \sim \mathcal{N}(\mu_1, \sigma^2) \qquad o_1 = 1, \ldots, n_1,$$

and

$$Y_{o_2} \sim \mathcal{N}(\mu_2, \sigma^2) \qquad o_2 = 1, \ldots, n_2.$$

That is, both random variables are assumed to be normally distributed with common standard deviation σ but not necessarily with the same number of observations n. For this case, the effect size — also known as Cohen's d (Cohen, 1988) — is defined as

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}.$$

The estimators proposed in this family are different with respect to the choice of the standard deviation (S). They are all computed by the generic form

$$\frac{\overline{X} - \overline{Y}}{S}$$

and therefore represent a standardized measure of the effect. There are three popular coefficients that are presented here. The first will be denoted by d and results from inserting the pooled estimate of the standard deviation in the denominator of Equation 3.8. The pooled estimate S_{pool} is given by

$$S_{\text{pool}} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}},$$

where S_1 and S_2 are the sample standard deviations for X and Y, respectively. Therefore,

$$d = \frac{\overline{X} - \overline{Y}}{S_{\text{pool}}}. \tag{3.8}$$

Another estimator was proposed by Glass (1976; see also Glass et al., 1981) and will be denoted by d'. It is given by inserting S_{con} in the denominator of Equation 3.8, where S_{con} represents the standard deviation of a control group. The control group is ordinarily chosen as the reference group in a two-group experimental setting.

Both d and d' have a distribution related to the noncentral t distribution (see Hedges, 1981; Hedges & Olkin, 1985). Let

$$\tilde{n} = \frac{n_1 n_2}{n_1 + n_2},$$

then $\sqrt{\tilde{n}}d$ and $\sqrt{\tilde{n}}d'$ follow a noncentral t distribution with noncentrality parameter $\tau = \sqrt{\tilde{n}}\delta$. Bias as well as variance of d are smaller than those of d' (Hedges & Olkin, 1985). The focus will therefore be on d.

The expected value of d is given by Hedges (1981) as

$$E(d) = \frac{\delta}{f(m)},$$

where $m = n_1 + n_2 - 2$ and

$$f(m) = \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{\frac{m}{2}}\Gamma\left(\frac{m-1}{2}\right)}.$$

Hedges (1981) also derived an unbiased estimator d'' of δ by drawing on this result. It is given as an approximation in the following equation

$$d'' = d \times \left(1 - \frac{3}{4(n_1 + n_2) - 9}\right).$$

This is also the UMVU estimator when $n_1 = n_2$ (Hedges, 1981). Further properties of this estimator are not given here as the focus is on the more common estimator d.

The asymptotic distribution of d is normal with expected value δ. The asymptotic variance of the random variable d is given by

$$\sigma_D^2 = \frac{n_1 + n_2}{n_1 n_2} + \frac{\delta^2}{2\left(n_1 + n_2\right)} \tag{3.9}$$

(Hedges & Olkin, 1985, p. 86, Equation 15). Customarily, the variance is estimated by plugging in d for δ in practical applications. For an equal number of persons in both groups, this variance estimate based on Equation 3.9 reduces to

$$\hat{\sigma}_D^2 = \frac{4 + d^2}{n}, \tag{3.10}$$

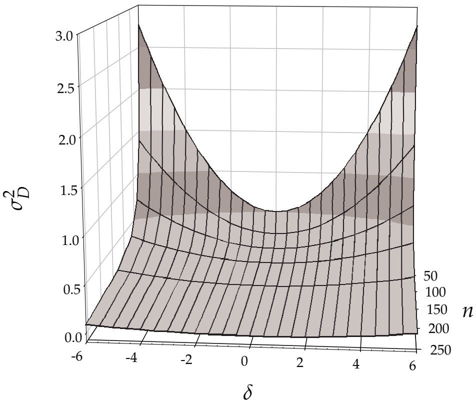

Figure 3.5 Variance of d across different values of δ and n.

where $n_1 + n_2 = n$. As can easily be seen from these equations, the variance (estimate) depends on the population parameter (or d) itself as was the case for the correlation coefficient. To provide an impression of this dependency, consider Figure 3.5.

As can be seen, the relationship between δ and the variance of the estimator is quite strong for large absolute values of δ and different in shape in comparison to the relationships previously examined for correlation coefficients. Since these variances play a central role in meta-analyses using d as an effect size, this may have unwanted effects on the results.

The details on the r and d families of effect sizes necessary for the present purposes are outlined at this point. Discussion will now turn to the question of the relation between r and d measures.

3.3 CONVERSION OF EFFECT SIZES

The conversion of effect sizes is one of the central features of meta-analysis. Effect sizes have always to be converted when the database does not provide coefficients from the same family. For example, it may be the case that one half of available studies reports the results from experiments and therefore d values[4], whereas the other half has observed the bivariate linear relationship between variables of interest and reports r values. The question arises in such cases how different effect size measures may be analyzed in a single meta-analysis.

Conversions of effect sizes are intended to homogenize the database to one single effect size (family). A host of conversion formulae for the various specific effect sizes has been presented to date, that will not be repeated here (see, e.g., Olejnik & Algina, 2000; Rosenthal, 1994). Instead, only the following formulae for the conversion of the Pearson correlation coefficient and the stan-

[4]Depending on the design, other indices than the d as introduced here may be appropriate.

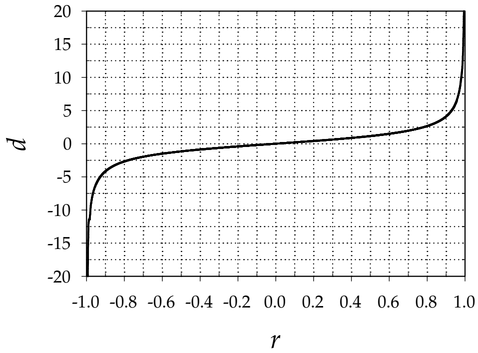

Figure 3.6 The *r* to *d* transformation.

dardized effects size measure *d* will be presented. They are given in various resources for the case of equal group sizes (i.e., $n_1 = n_2$) as

$$r = \sqrt{\frac{d^2}{d^2 + 4}} \tag{3.11}$$

and

$$d = \frac{2r}{\sqrt{(1 - r^2)}} \tag{3.12}$$

(e.g., Cohen, 1988; Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Rosenthal, 1991). The conversion with Equation 3.12 is illustrated in Figure 3.6.

In Figure 3.6 it can be seen that the conversion of *r* to *d* has a similar shape in comparison to the Fisher-*z* transformation presented in Figure 3.1 but is much steeper in the tails. This suggests a normalizing transformation of the correlation coefficient as was the case for the Fisher-*z* transformation but may not result in an equally good normal approximation. Conversely, a transformation of *d* to *r* leads to relatively large differences in the space of *r* for values near zero being relatively close to each other, but large differences far from zero translate into small differences in absolute values of *r*.

Aaron, Kromrey, and Ferron (1998) provided a derivation of the conversion for equal *n* that slightly differs from Equation 3.11

$$r = \sqrt{\frac{d^2}{d^2 + 4 - \frac{8}{n}}},$$

where $n = n_1 + n_2$. For the case of unequal *n* they proposed

$$r = \sqrt{\frac{d^2}{d^2 + \frac{(n_1+n_2)^2 - 2(n_1+n_2)}{n_1 n_2}}}.$$

The authors also showed that discrepancies exist between their corrected formulae and results from Equation 3.11. The reported differences were considered as negligible for the balanced case when $n > 50$.

However, it is not clear from Aaron et al.'s (1998) presentation whether their corrected formulae provide more accurate procedures for *estimating* the correlation coefficient by way of d. As is the case for the standard formula in Equation 3.11, their derivation also draws on the *null* distribution of both effect sizes that is approximately t with $n - 2$ degrees of freedom. Assuming both effect sizes to have an equal distribution seems to be *only* justified in this case. Yet, for the nonnull case (i.e., $\rho \neq 0$ and $\delta \neq 0$), neither the distribution of r nor the distribution of d is exactly (noncentral) t. As a consequence, there is no statistical derivation available for the conversion of r to d or vice versa for the nonnull case. Hence, when there is a lack of a standard for comparison, there is no way to theoretically evaluate the quality of the conversion formulae.

One possibility for evaluation, that is pursued in the following Monte Carlo study, is to apply the conversion proposed in Equation 3.11 to simulated data and study the behavior of converted statistics. This will enable the examination of the implicit assumption, by the widespread application of Equation 3.11, that the conversion itself does not have any influence on the results in meta-analysis.