Effect Sizes and Intervention Research

Philip Osteen, PhD, MSW
Charlotte Bright, PhD, MSW
University of Maryland
January 16, 2010
Agenda

- Why are ES important?
- Types of ES
- Estimating Magnitude of ES
- Calculating ES
Effect sizes are easy!

- Unlike many methods workshops, this relies ONLY on simple mathematical formulae and information already available.

- Calculating effect sizes relies on simple but underused methods.

- By the end of this workshop, you will be able to interpret and report effect sizes in your work.
What is an “effect size”?

- An “effect” is
  - A change or changed state occurring as a direct result of action by somebody or something else (Encarta, 2009)

- “Size”
  - The degree of something in terms of how big or small it is (Encarta, 2009)
What is an “effect size”? 

- In statistical inference, the effect size is an estimate of the strength of association between 2 or more variables.

- In and of itself, the effect size is not an indication of causality.
Why are effect sizes important?

- **Power Analysis**
  - Statistical power is the probability of rejecting a false null hypothesis
  - Statistical power is affected by the estimated effect size, α level, and sample size
    - $\uparrow$ ES = $\uparrow$ power
    - $\uparrow$ α level = $\uparrow$ power
    - $\uparrow$ N = $\uparrow$ power
  - Power $\geq 0.80$ is the standard
    - Inaccurate estimation of power may lead to wasted resources
Why are effect sizes important?

- Knowing the magnitude of an effect allows us to ascertain the practical significance of statistical significance
  
  - Does statistical significance always mean practical significance?
    * With sufficient sample size, any effect can reach statistical significance
  
  - Does statistical non-significance always mean practical non-significance?
    * With insufficient sample size, even the largest effects may not be statistically significant
Types of Effect Sizes

- Pearson’s $r$ and $R^2$
  - Primarily used in correlation and regression
  - $r$ is the linear association between 2 continuous variables
    \[ r_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{n(\sum x^2 - (\sum x)^2)} \sqrt{n(\sum y^2 - (\sum y)^2)}} \]
    - Standardized ES
    - Bounded between -1 and +1
    - Interpretation: For each 1 SD change in $x$, there is a “$r$” SD change in $y$
  - $R^2$ (Coefficient of Determination) is the proportion of shared variance between 2 or more variables
    - Standardized ES
    - Bounded between 0 and 1
    - Interpretation: “$R^2 \times 100$” percent of the variance in $y$ can be explained by the variance in $x(s)$
    - Note: May see pseudo-$R^2$ reported for logistic regression, but interpretation is not equivalent
Types of Effect Sizes

- Odds Ratio (OR)
  - Odds of being in one group relative to the odds of being in a different group
  - Used with categorical outcomes (e.g., $\chi^2$, logistic regression)
  - Ranges from 0 to $+\infty$
  - Values $>1$ indicate an increase in odds relative to the reference group
  - Values $<1$ indicate a decrease in odds relative to the reference group
Types of Effect Sizes

- Cohen’s $d$

\[
d = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pooled}}}
\]

- Standardized ES of the difference between 2 means
- Used with t-tests
- $d$ ranges from $-\infty$ to $+\infty$
- Interpretation: The difference in the mean values is “$d$” standard deviation(s)
- Note: Other ES available when appropriate
  - Hedge’s G – corrects for small sample size
  - Glass’s $\Delta$ - corrects for unequal variances
Types of Effect Sizes

- Eta-squared ($\eta^2$) and Partial Eta-squared ($\eta_p^2$)

\[ \eta^2 = \frac{SS_{treatment}}{SS_{total}} \quad \eta_p^2 = \frac{SS_{treatment}}{SS_{treatment} + SS_{Error}} \]

- Standardized ES of the shared variance between a continuous outcome and categorical predictor(s)

- Used with ANOVA family and GLMs

- Bounded between 0 and 1

- Interpretation: "$\eta^2(x \times 100)$" percent of the variance in $y$ can be explained by the variance in $x$

- ES interpretation is therefore consistent with $R^2$ interpretation (Dattalo, 2008)

- Inconsistent recommendations for which one to use
  - ($\eta^2$) is constrained by the size and magnitude of other effects
  - ($\eta_p^2$) is not additive
Types of Effect Sizes

- Phi (\(\phi\)) and Cramer’s Phi (\(\phi_c\)) or \(V\)

\[
\phi = \sqrt{\frac{x^2}{N}} \quad \phi_c = \sqrt{\frac{x^2}{N(k-1)}}
\]

- Standardized ES of association for the chi-square test
  - Can be squared to show how much shared variance is accounted for by the relationship detected by the chi-square

- Phi (\(\phi\)) used for 2 binary variables

- Cramer’s Phi (\(\phi_c\)) can be used with any number of levels

- Bounded between 0 and 1

- Interpreted like Pearson’s \(r\) and \(R^2\)
Types of Effect Sizes

- Cohen’s $f^2$

$$f^2 = \frac{R^2}{1-R^2}$$

- Standardized ES of the proportion of explained variance over unexplained variance

- Rarely reported (lacks intuitive sense) but frequently used in power calculations

- Can be used for multiple regression or ANOVA
Types of Effect Sizes

- **ES by Analysis Summary**
  - Correlation/Regression
    - $r$, $R^2$, Cohen’s $f^2$
  - Logistic Regression
    - Odds Ratios
    - Pseudo- $R^2$
  - Mean Differences
    - Cohen’s $d$, $\eta^2$, $R^2$, Cohen’s $f^2$
  - Crosstabs/Chi-Square
    - Phi/Cramer’s $V$
Magnitude of Effect

- Effect sizes are generally broken down into “small”, “moderate”, or “large”
  - What constitutes a small, moderate, or large effect depends on the type of effect size being considered
  - These terms are arbitrary and relational
  - These are guidelines, not cutoff values
  - Most often cited reference for magnitude of effect is Cohen (1988)
Magnitude of Effect

- For Pearson’s $r$, phi, Cramer’s phi
  - “small” $\approx 0.1$
  - “moderate” $\approx 0.3$
  - “large” $\approx 0.5$
Magnitude of Effect

- For $r^2$, $\eta^2$, and $\eta_p^2$
  - “small” $\approx 0.01$
  - “moderate” $\approx 0.09$
  - “large” $\approx 0.25$
Magnitude of Effect

- For $R^2$
  - “small” $\approx 0.02$
  - “medium” $\approx 0.13$
  - “large” $\approx 0.26$
Magnitude of Effect

- For Cohen’s $d$
  - “small” $\approx \pm 0.2$
  - “moderate” $\approx \pm 0.5$
  - “large” $\approx \pm 0.8$
Magnitude of Effect

- Odds ratios
  - No specified criteria for categorizing magnitude
  - However...
    - Chinn (2000) \[ d = \frac{\ln(OR)}{1.81} \]
    - Can substitute values of \( d \) to get \( \ln(OR) \)
    - And, \( e^{\ln(OR)} = OR \)

- “small” = 1.44
- “moderate” = 2.47
- “large” = 4.25

*note – if you choose to report this, you should provide appropriate references and explain your rationale
Magnitude of Effect

- For Cohen’s $f^2$
  - “small” $\approx 0.02$
  - “moderate” $\approx 0.15$
  - “large” $\approx 0.35$
## Magnitude of Effect Summary Table

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Small</th>
<th>Moderate</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson's $r$</td>
<td>0.10</td>
<td>0.30</td>
<td>0.50</td>
</tr>
<tr>
<td>$r^2$</td>
<td>0.01</td>
<td>0.09</td>
<td>0.25</td>
</tr>
<tr>
<td>$\eta^2$</td>
<td>0.01</td>
<td>0.09</td>
<td>0.25</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02</td>
<td>0.13</td>
<td>0.26</td>
</tr>
<tr>
<td>Cohen's $d$</td>
<td>±0.20</td>
<td>±0.50</td>
<td>±0.80</td>
</tr>
<tr>
<td>Phi/Cramer's $V$</td>
<td>0.10</td>
<td>0.30</td>
<td>0.50</td>
</tr>
<tr>
<td>Cohen's $f^2$</td>
<td>0.02</td>
<td>0.15</td>
<td>0.35</td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>1.44</td>
<td>2.47</td>
<td>4.25</td>
</tr>
</tbody>
</table>
Effect Size Conversions

• Cohen’s $d$ is the most widely reported (and presumably understood) effect size

• It is possible to convert many ES to Cohen’s $d$
Effect Size Conversions

- Calculating Cohen’s $d$ from...
  - Correlation
    \[ d = \frac{2r}{\sqrt{1-r^2}} \]
  - Chi-Square
    - $df = 1$
    \[ d = 2\sqrt{\frac{x^2}{N-x^2}} \]
    - $df > 1$
    \[ d = 2\sqrt{\frac{x^2}{N}} \]
  - Odds Ratio (Chinn, 2000)
    \[ d = \frac{\ln(OR)}{1.81} \]

* See Dunst, et al. (2004) for a more comprehensive list of conversion formulas
Practical Significance

- So we have an ES, we know the magnitude of the effect, and we know whether or not the inferential test it is based on is statistically significant – now what?

  - Mordock (2000) suggests that we have to make a value judgment about the ES
    - Is it important?
    - Is it feasible?
    - Is it practical?
Small Effect Sizes

“Small effect sizes can have substantial practical value. This is particularly the case if a treatment is relatively inexpensive, is easy to execute, is politically feasible, and can be employed on a large scale, thereby affecting large numbers of individuals.” (Litschge, Vaughn, & McCrea, 2010, p. 22).
Reporting Guidelines and Trends

- Reporting effect sizes has three important benefits (APA, 1999):
  - Meta-analysis
  - Informing subsequent research
  - Interpretation and evaluation of results within the context of related literature
Reporting Guidelines and Trends

- What to report (APA, 2010):
  
  ◦ Type of effect size
  
  ◦ Value of the effect size (in original units, such as lbs. or mean difference on a scale, and/or ES statistic)
  
  ◦ Magnitude of the effect size
  
  ◦ Interpretation of the effect size
  
  ◦ Practical significance of the effect size
Examples

Using online calculators:
http://www.uccs.edu/~faculty/lbecker/
http://faculty.vassar.edu/lowry/newcs.html

Examples from calculations used in forthcoming article:
Example 1 – Cohen’s $d$

Difference between two types of group care on length of stay in placement

Effect Size Calculators

Calculate Cohen’s $d$ and the effect-size correlation, $r_{YX}$, using --

- means and standard deviations
- independent groups $t$ test values and $df$

For a discussion of these effect size measures see Effect Size Lecture Notes

Calculate $d$ and $r$ using means and standard deviations

Calculate the value of Cohen’s $d$ and the effect-size correlation, $r_{YX}$, using the means and standard deviations of two groups (treatment and control).

Cohen’s $d = M_1 - M_2 / \sigma_{\text{pooled}}$

where $\sigma_{\text{pooled}} = \sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}$

$r_{YX} = d / \sqrt{(d^2 + 4)}$

Note: $d$ and $r_{YX}$ are positive if the mean difference is in the predicted direction.
Example 1 – Cohen’s $d$

Difference between two types of group care on length of stay in placement
Example 1 – Cohen’s $d$

Difference between two types of group care on length of stay in placement

Effect Size Calculators

Calculate Cohen’s $d$ and the effect-size correlation, $r_{YX}$, using --

- means and standard deviations
- independent groups t test values and df

For a discussion of these effect size measures see Effect Size Lecture Notes

Calculate $d$ and $r$ using means and standard deviations

Calculate the value of Cohen’s $d$ and the effect-size correlation, $r_{YX}$, using the means and standard deviations of two groups (treatment and control).

$\text{Cohen’s } d = \frac{M_1 - M_2}{\sigma_{\text{pooled}}}$

where $\sigma_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} / 2$

$r_{YX} = d / \left(d^2 + 4\right)$

Note: $d$ and $r_{YX}$ are positive if the mean difference is in the predicted direction.
Example 2 – Cramer’s phi (V)

Association between type of care and likelihood of new arrest 12 months post-discharge
Example 2 – Cramer’s phi (V)

Association between type of care and likelihood of new arrest 12 months post-discharge
Example 2 – Cramer’s phi (V)
Association between type of care and likelihood of new arrest 12 months post-discharge
Example 2 – Cramer’s phi (V)
Association between type of care and likelihood of new arrest 12 months post-discharge
SAS and Effect Size Calculations

SAS will automatically output values for phi and Cramer’s V in chi-square analyses:

```sas
proc freq;
  table var1 * var2 / chisq;
run;
```

And for the values of $r$ and $R^2$ in correlation or regression analyses:

```sas
proc corr;
  var1 var2;
run;
```

```sas
proc reg;
  model y = x1 x2...xk;
run;
quit;
```

Also, for odds ratios in logistic regression analyses:

```sas
proc logistic descending;
  model y = x1 x2...xk;
run;
```

SAS will NOT output values for $\eta^2$ (does give sums of squares to simplify hand-calculations)
ES can be obtained through syntax or “options”

- $r$ and $R^2$ automatically provided in regression
- Odds ratios automatically provided in logistic regression
Phi and Cramer’s phi (V) can be selected as part of the output
- “Descriptive Statistics” -> “Crosstabs” -> “Statistics”

$\eta^2$ can be selected as part of the output
- “General Linear Model” -> … -> “Options”
SAS and SPSS

- Cohen’s $d$
  - Cohen’s $d$, Glass’s $\Delta$, Hedge’s $g$ can be calculated using SPSS syntax, available at: http://www.spsstools.net/Syntax/T-Test/StandardizedEffectsSize.txt.
  - Cohen’s $d$ can be calculated in SAS via a somewhat complex process
  - Hand calculation is, for the moment, simpler

- Additional syntax (Meyer, et al., 2003) for this and other processes for calculating effect sizes in SAS and SPSS is available at: www.tandf.co.uk/journals/authors/hjpa/resources/basiceffectssizeguide.rtf.
Effect Size and Power Calculators

- G*Power 3
  - Covers all of the major types of ES
  - [http://wwwpsycho.uni-duesseldorfdede/abteilungen/aap/gpower3/](http://wwwpsycho.uni-duesseldorfdede/abteilungen/aap/gpower3/)

- Optimal Design
  - Developed for hierarchical models
  - [http://sitemakerumichedu/group-based/optimal_design_software](http://sitemakerumichedu/group-based/optimal_design_software)

- StatPages
  - Provides links to more than 400 online statistical calculators
  - [http://statpagesorg/](http://statpagesorg/)


Contact Information

Philip Osteen, PhD
University of Maryland
School of Social Work
525 W. Redwood St.
Baltimore, MD, 21201
410-706-3612
posteen@ssw.umaryland.edu

Charlotte Bright, PhD
University of Maryland
School of Social Work
525 W. Redwood St.
Baltimore, MD, 21201
410-706-3605
cbright@ssw.umaryland.edu