Technical Assistance Report:
Mediation Analysis

Research in Prevention Laboratory

Last Revised: October 18, 2000
NIDA Grant # 5 R01 DA09757-04
P. I. David P. MacKinnon
Persons who contributed to this report: Marcia P. Taborga, JeeWon Cheong, and David P. MacKinnon

Table of Contents

Introduction

The purpose of this report is to briefly describe mediation analysis and to provide programs to analyze mediation models. We begin with a description of mediation, follow with a general description of mediation analysis, and end with data analyses programs and results from the SAS and SPSS statistical software packages. We present the analyses and results for a one mediator model first and then a four mediator model.

Mediation

Mediation models explain "how" an effect occurred by hypothesizing a causal sequence. The basic mediation model is a causal sequence in which the independent variable (X) causes the mediator (M) which in turn causes the dependent variable (Y), therefore explaining how X had its' effect on Y (see figure 1 in MacKinnon, 1994 or MacKinnon & Dwyer, 1993). Mediational processes are common in basic and applied psychology. For example, prevention programs in a variety of substantive areas are often developed with the same general underlying model displayed in Figure 1. The prevention program (X) is designed to change mediating variables (M) hypothesized to be causally related to the outcome (Y). If the assumption that the mediating variables are causally related to the outcome is correct, a prevention program that substantially changes the mediating variable will, in turn, change the outcome.

Mediation Analysis

Mediation analysis is accomplished with three steps (Baron & Kenny, 1986; Judd & Kenny, 1981; MacKinnon & Dwyer, 1993). The first step is to determine the effect of the independent variable on the dependent variable. The second step is to determine the effect of the independent variable on the mediator. Lastly, the effect of the mediator on the dependent variable is determined. If there is evidence that the program caused the mediator and the mediator caused the dependent variable, there is evidence for mediation of the relationship between the program and the dependent variable.

Calculation of the mediated effect can be accomplished in two ways that yield identical results when the dependent variable is continuous (MacKinnon & Dwyer, 1993). The first method is more common in epidemiology. It involves two regression equations. In the first equation the dependent variable is regressed on the independent variable as shown in figure 3.

$$Model1: Y_o = b_1 + t X_P + e_1$$

This yields the regression coefficient, ô, which relates the independent variable, X_p, to the dependent variable, Y_o, without taking the mediator into account. The â$_1$ represents the intercept and e_1 represents the error term. In the second equation, the dependent variable, Y_o, is regressed on both the independent variable, X_p, and the mediator, X_m. This yields the coefficient â relating the mediator to the dependent variable and the coefficient ô' relating the independent variable to the dependent variable after it is adjusted for the mediator. Again, the â$_2$ represents the intercept and e_2 represents the error term.

$$Model2: Y_O = b_2 + t' X_P + b X_M + e_2$$

The value of the total effect is ô (from Model 1), the value of the direct effect is ô', and the value of

the indirect, or mediated effect, is ô - ô'.

The second method is more common in the social sciences and is based on path analysis (Baron & Kenny, 1986; Judd & Kenny, 1981; MacKinnon & Dwyer, 1993). It involves two regression equations to calculate the regression coefficients that correspond to the three paths shown in Figure 3. The first equation is Model 2, from above which yields the coefficient â relating the mediator to the dependent variable and the coefficient ô' relating the independent variable to the dependent variable after it is adjusted for the mediator. In the next equation, the mediator, Y_m, is regressed on the independent variable, X_p, yielding coefficient, á. As with the previous two models, the $â_3$ represents the intercept and e_3 represents the error term.

Model 3: $Y_M = b_3 + a X_P + e_3$

The product of áâ is the indirect, or mediated, effect, ô' is the direct effect, and the total effect is calculated as ô = áâ + ô'.

Note that ô' is referred to as the direct effect and both ô - ô' and áâ are referred to as the indirect, or mediated, effect.

Significance Tests of Mediation

The t-test value of the mediated effect can be calculated by dividing the mediated effect by the standard error of the mediated effect (MacKinnon and Dwyer, 1993). The variance of the mediated effect for the ô - ô' method was derived by McGuigan and Langholtz (1988).

$$s^2_{t-t'} = s^2_t + s^2_{t'} - 2s_{tt'}$$

The following is the equation for the significance test of ô - ô'.

$$t = \frac{(t - t')}{\sqrt{s^2_t + s^2_{t'} - 2s_{tt'}}}$$

The variance of áâ was derived by Sobel (1982) using the multivariate delta method. The following is the exact variance of áâ, however, the last term in parenthesis is omitted in Sobel's variance because it is based on first derivatives.

$$s^2_{ab} = a^2 s^2_b + b^2 s^2_a (+ s^2_a s^2_b)$$

The significance test of áâ uses a z distribution although the test may not follow a normal distribution (MacKinnon & Lockwood, under review; MacKinnon, Warsi, & Dwyer, 1995). The following is the equation for the significance test of áâ using Sobel's (1982) first order solution for the variance of the mediated effect.

$$z = \frac{ab}{\sqrt{a^2 s^2_b + b^2 s^2_a}}$$

Sample data set

The adolescent training and learning to avoid steroids (ATLAS) data set will be used to illustrate mediation analysis. For the one-mediator model we use program (no program vs. program) as the independent variable, peers-as-an-information source as the mediator, and intentions-to-use-anabolic androgenic steroids (AAS) as the dependent variable. For the four-mediator model we use the same independent and dependent variables, program and intentions-to-use-AAS. For the mediators we use peers-as-an-information source, resistance to drug offers, normative beliefs about AAS use, and attitudes toward users of AAS.

SPSS and SAS programs and output for this one-mediator and four-mediator model are shown below.

One-Mediator Model

SAS Program

```
/*A mediation model for ATLAS data
group to peer-as-information-source to intent to use
Output includes regressions, se, and correlations
*/

libname old '$dave/datafiles/';
option nocenter ls=80;
data univ; set old.btog(keep=school id
        csgroup dgroup csintent dintet
        csperin dperin waveb wavecs
        waved);

*Selecting only cohort 1;
if waveb=1;
*Combining cs and d wave variables for cohort1 subjects;
data combine; set univ;
if wavecs=1 then dgroup=csgroup;
if wavecs=1 then dperin=csperin;
if wavecs=1 then dintent=csintent;
group=dgroup;
peerinf=dperin;
intent=dintent;

if not (group = . or peerinf = . or intent = .);

proc means data=combine;

proc reg data=combine;
model intent=group;
model peerinf=group;
model intent=group peerinf;
run;

run;
quit;
```

Data
Opening data set "btog" in directory '$dave/datafiles/' and selecting the relevant variables

Variables
Group – group vs. no group (independent variable)
Peerinf – peers-as-information-source (mediator)
Intent – intent-to-use-AAS (dependent variable)

Case selection
Selecting cases with complete data for all three variables.

Regressions
Regressing the intent-to-use-AAS (1st model)and peers-as-information source (2nd model) on group.
Regressing intent-to-use-AAS on both group and peers-as-information-source (3rd model).

Output from SAS

```
Model: MODEL1
Dependent Variable: INTENT
Parameter Estimates
                 Parameter      Standard    T for H0:
Variable  DF      Estimate         Error   Parameter=0    Prob > |T|

INTERCEP   1      1.769631    0.05634485        31.407        0.0001
GROUP      1     -0.166514    0.08261059        -2.016        0.0441
```

Regression coefficient estimate for path ô, the total effect of group on intent-to-use-AAS = -0.167

```
Model: MODEL2
Dependent Variable: PEERINF
Parameter Estimates
                 Parameter      Standard    T for H0:
Variable  DF      Estimate         Error   Parameter=0    Prob > |T|
INTERCEP   1      4.597614    0.07507145        61.243        0.0001
GROUP      1      0.579859    0.11006677         5.268        0.0001
```

Regression coefficient estimate for path á, the effect of group on the mediator, peers-as-information-source = 0.580.

```
Model: MODEL3
Dependent Variable: INTENT
Parameter Estimates
                 Parameter      Standard    T for H0:
Variable  DF      Estimate         Error   Parameter=0    Prob > |T|
INTERCEP   1      2.169626    0.12965967        16.733        0.0001
GROUP      1     -0.116066    0.08341575        -1.391        0.1645
PEERINF    1     -0.087001    0.02543580        -3.420        0.0007
```

Regression coefficient estimate for path â, the effect of peers-as-information-source on intent-to-use-AAS = -0.087 and the regression coefficient estimate for path ôr, the effect of the mediator, peers-as-an-information source, on intent-to-use-AAS = -0.116.

Results:

Mediated effect = á*â = .580 * -.087 = -0.05

Direct effect = ôr = -0.12.

Total effect = (á*â)+ôr or ô = -0.12 + -0.05 = -0.17

SPSS Program

```
get sas data='c:\my documents\btog'.

select if waveb=1.

if (wavecs=1) dgroup=csgroup.
if (wavecs=1) dperin=csperin.
if (wavecs=1) dintent=csintent.

compute group=dgroup.
compute peerinf=dperin.
compute intent=dintent.

variable labels group 'Group'
                peerinf 'Peers as information source'
                intent 'Intent to use'.

select if (group ge 0) and (peerinf ge 0)
                and (intent ge 0).

regression
        /variables=group peerinf intent
        /missing=listwise
        /descriptives=mean stddev
        /statistics=all
        /dependent=intent
        /enter=group
        /enter=peerinf.

regression
        /variables=group peerinf
        /missing=listwise
        /descriptives=mean stddev
        /statistics=all
        /dependent=peerinf
        /enter=group.
```

Data
Opening data set "btog"

Variables
Group – group vs. no group (independent variable)
Peerinf – peers-as-information-source (mediator)
Intent – intent-to-use-AAS (dependent variable)

Case selection
Selecting cases with complete data for all three variables.

Regressions
Regressing intent-to-use-AAS (1st model)on group.
Regressing intent-to-use-AAS on both group and peers-as-information-source (3rd model).

Regressions
Regressing peers-as-information source (2nd model) on group.

Output from SPSS

```
          * * * *   M U L T I P L E   R E G R E S S I O N   * * * *
Equation Number 1    Dependent Variable..   INTENT   Intent to use
Block Number  1.  Method:  Enter      GROUP

------------------------ Variables in the Equation -----------------------
Variable              B        SE B       Beta    SE Beta   Correl Part Cor

GROUP          -.166514     .082611   -.068571    .034019 -.068571 -.068571
(Constant)     1.769631     .056345
```

Regression coefficient estimate for path ô, the total effect of group on intent-to-use-AAS = -0.167

```
          * * * *   M U L T I P L E   R E G R E S S I O N   * * * *

Equation Number 1    Dependent Variable..   INTENT   Intent to use
Block Number  2.  Method:  Enter      PEERINF

------------------------ Variables in the Equation -----------------------

Variable              B        SE B       Beta    SE Beta   Correl Part Cor

GROUP          -.116066     .083416   -.047797    .034351 -.068571 -.047043
PEERINF        -.087001     .025436   -.117494    .034351 -.125945 -.115643
(Constant)     2.169626     .129660
```

Regression coefficient estimate for path â, the effect of peers-as-information-source on intent-to-use-AAS = -0.087 and the regression coefficient estimate for path ôr, the effect of the mediator, peers-as-an-information source, on intent-to-use-AAS = -0.116.

```
          * * * *  M U L T I P L E  R E G R E S S I O N  * * * *

Equation Number 1    Dependent Variable..   PEERINF   Peers as information sourc
Block Number  1.  Method:  Enter      GROUP
------------------------ Variables in the Equation -----------------------

Variable              B        SE B       Beta    SE Beta   Correl Part Cor

GROUP           .579859     .110067    .176815    .033562  .176815  .176815
(Constant)     4.597614     .075071
```

Regression coefficient estimate for path á, the effect of group on the mediator, peers-as-information-source = 0.580.

SAS Program

```
libname old 'c:\my documents\atlas\data';
option nocenter ls=80;
data univ; set old.btog(keep=school id csgroup dgroup csintent
   dintent csperin dperin csnorms dnorms csatuse datuse csresist dresist
   waveb wavecs waved);

*Selecting only cohort 1;
if waveb=1;

*Combining cs and d wave variables for cohort1 subjects;
data combine; set univ;
if wavecs=1 then dgroup=csgroup;
if wavecs=1 then dperin=csperin;
if wavecs=1 then dnorms=csnorms;
if wavecs=1 then dresist=csresist;
if wavecs=1 then datuse=csatuse;
if wavecs=1 then dintent=csintent;
group=dgroup;
peerinf=dperin;
norms=dnorms;
attitude=datuse;
resistan=dresist;
intent=dintent;

if (group ne . and peerinf ne . and norms ne . and attitude ne .
      and resistan ne . and intent ne .);

proc freq data=combine; table group peerinf norms attitude resistan intent
  / norow nopercent;


proc reg data=combine;
model intent=group;
model peerinf=group;
model norms=group;
model attitude=group;
model resistan=group;
model intent=group peerinf norms attitude resistan;

run;
quit;
```

SAS Output

The REG Procedure
Model: MODEL1
Dependent Variable: intent

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.80598	5.80598	4.03	0.0449
Error	853	1227.55660	1.43911		
Corrected Total	854	1233.36258			

Root MSE	1.19963	R-Square	0.0047
Dependent Mean	1.68240	Adj R-Sq	0.0035
Coeff Var	71.30461		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.75930	0.05612	31.35	<.0001
group	1	-0.16520	0.08225	-2.01	0.0449

The SAS System 13:07 Wednesday, October 25, 2000 57

The REG Procedure
Model: MODEL2
Dependent Variable: peerinf

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	72.17577	72.17577	27.91	<.0001
Error	853	2205.78505	2.58591		
Corrected Total	854	2277.96082			

Root MSE	1.60808	R-Square	0.0317
Dependent Mean	4.87836	Adj R-Sq	0.0305
Coeff Var	32.96348		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.60722	0.07522	61.25	<.0001
group	1	0.58248	0.11025	5.28	<.0001

The REG Procedure
Model: MODEL3
Dependent Variable: norms

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.55828	0.55828	0.23	0.6352
Error	853	2114.38409	2.47876		
Corrected Total	854	2114.94237			

Root MSE	1.57441	R-Square	0.0003
Dependent Mean	2.24230	Adj R-Sq	-0.0009
Coeff Var	70.21399		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.21845	0.07365	30.12	<.0001
group	1	0.05123	0.10794	0.47	0.6352

The REG Procedure
Model: MODEL4
Dependent Variable: attitude

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.29163	5.29163	2.60	0.1073
Error	853	1736.28848	2.03551		

```
Corrected Total         854     1741.58012


Root MSE              1.42671    R-Square     0.0030
Dependent Mean        5.31813    Adj R-Sq     0.0019
Coeff Var            26.82734


                        Parameter Estimates

                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|

Intercept     1        5.24471        0.06674      78.59      <.0001
group         1        0.15772        0.09782       1.61      0.1073
```



```
The REG Procedure
Model: MODEL5
Dependent Variable: resistan

                            Analysis of Variance

                                    Sum of           Mean
Source                    DF        Squares         Square    F Value    Pr > F

Model                      1        7.35361        7.35361       4.06    0.0443
Error                    853     1546.65503        1.81319
Corrected Total          854     1554.00864


Root MSE              1.34655    R-Square     0.0047
Dependent Mean        6.00741    Adj R-Sq     0.0036
Coeff Var            22.41481


                        Parameter Estimates

                     Parameter       Standard
Variable     DF       Estimate          Error    t Value    Pr > |t|

Intercept     1        5.92086        0.06299      94.00      <.0001
group         1        0.18592        0.09232       2.01      0.0443
```



```
The REG Procedure
Model: MODEL6
Dependent Variable: intent
```

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	266.74573	53.34915	46.86	<.0001
Error	849	966.61685	1.13854		
Corrected Total	854	1233.36258			

Root MSE	1.06702	R-Square	0.2163
Dependent Mean	1.68240	Adj R-Sq	0.2117
Coeff Var	63.42269		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.36778	0.21127	20.67	<.0001
group	1	-0.07243	0.07444	-0.97	0.3308
peerinf	1	-0.02507	0.02322	-1.08	0.2806
norms	1	0.06117	0.02326	2.63	0.0087
attitude	1	-0.17704	0.02759	-6.42	<.0001
resistan	1	-0.28715	0.02899	-9.90	<.0001

SPSS Program

```
get sas data='c:\my documents\atlas\data\sastrans.por'.

select if waveb=1.

if (wavecs=1) dgroup=csgroup.
if (wavecs=1) dperin=csperin.
if (wavecs=1) datuse=csatuse.
if (wavecs=1) dnorms=csnorms.
if (wavecs=1) dresist=csresist.
if (wavecs=1) dintent=csintent.

compute group=dgroup.
compute peerinf=dperin.
compute attitude=datuse.
compute norms=dnorms.
compute resistan=dresist.
compute intent=dintent.

variable labels group 'Group'
 peerinf 'Peers as information source'
 attitude 'Attitude towards use of AAS'
 norms 'Normative beliefs about AAS use'
 resistan 'Resistance skills'
 intent 'Intent to use'.

select if not(missing(group) and missing(peerinf) and missing(attitude) and missing(norms)
        and missing(resistan) and missing(intent)).

regression
        /variables=group peerinf attitude norms resistan intent
        /missing=listwise
        /descriptives=mean stddev
        /statistics=coeff r anova ses tol cha zpp
        /dependent=intent
        /enter=group
        /enter=peerinf attitude norms resistan.

regression
        /variables=group peerinf
        /missing=listwise
        /descriptives=mean stddev
        /statistics=coeff r anova ses tol cha zpp
        /dependent=peerinf
        /enter=group.
```

```
regression
        /variables=group attitude
        /missing=listwise
        /descriptives=mean stddev
        /statistics=coeff r anova ses tol cha zpp
        /dependent=attitude
        /enter=group.

regression
        /variables=group norms
        /missing=listwise
        /descriptives=mean stddev
        /statistics=coeff r anova ses tol cha zpp
        /dependent=norms
        /enter=group.

regression
        /variables=group resistan
        /missing=listwise
        /descriptives=mean stddev
        /statistics=coeff r anova ses tol cha zpp
        /dependent=resistan
        /enter=group.
```

SPSS output

```
     * * * *  M U L T I P L E  R E G R E S S I O N  * * * *
Equation Number 1    Dependent Variable..   INTENT   Intent to use
Block Number  1.  Method:  Enter      GROUP
Variable(s) Entered on Step Number
   1..    GROUP     Group

------------------ Variables in the Equation ------------------

Variable             B        SE B       Beta         T  Sig T

GROUP          -.165204     .082249   -.068611    -2.009  .0449
(Constant)     1.759300     .056116               31.351  .0000


          * * * * * * * * * * * * * * * * * * * * * * * * * * * *
Block Number  2.  Method:  Enter      PEERINF  ATTITUDE NORMS    RESISTAN
Variable(s) Entered on Step Number
   2..    NORMS     Normative beliefs about AAS use
   3..    ATTITUDE  Attitude towards use of AAS
   4..    PEERINF   Peers as information source
   5..    RESISTAN  Resistance skills

------------------ Variables in the Equation ------------------
Variable             B        SE B       Beta         T  Sig T
GROUP          -.072427     .074436   -.030080     -.973  .3308
PEERINF        -.025067     .023216   -.034067    -1.080  .2806
ATTITUDE       -.177039     .027590   -.210375    -6.417  .0000
NORMS           .061172     .023259    .080104     2.630  .0087
RESISTAN       -.287150     .028991   -.322323    -9.905  .0000
(Constant)     4.367779     .211267              20.674  .0000
          * * * *  M U L T I P L E  R E G R E S S I O N  * * * *

          * * * *  M U L T I P L E  R E G R E S S I O N  * * * *

Equation Number 1    Dependent Variable..   PEERINF   Peers as information sourc
  Descriptive Statistics are printed on Page    5
Block Number  1.  Method:  Enter      GROUP
Variable(s) Entered on Step Number
   1..    GROUP     Group


------------------ Variables in the Equation ------------------
```

```
Variable             B        SE B       Beta         T  Sig T

GROUP           .582477     .110253    .178001     5.283  .0000
(Constant)     4.607221     .075223               61.248  .0000

          * * * *   M U L T I P L E   R E G R E S S I O N   * * * *
Equation Number 1    Dependent Variable..   ATTITUDE   Attitude towards use of A
  Descriptive Statistics are printed on Page    8
Block Number  1.  Method:  Enter      GROUP
Variable(s) Entered on Step Number
   1..    GROUP     Group

------------------ Variables in the Equation ------------------

Variable             B        SE B       Beta         T  Sig T

GROUP           .157717     .097818    .055122     1.612  .1073
(Constant)     5.244712     .066739               78.586  .0000
          * * * *   M U L T I P L E   R E G R E S S I O N   * * * *
Equation Number 1    Dependent Variable..   NORMS   Normative beliefs about AAS
Block Number  1.  Method:  Enter      GROUP
Variable(s) Entered on Step Number
   1..    GROUP     Group

------------------ Variables in the Equation ------------------

Variable             B        SE B       Beta         T  Sig T

GROUP           .051228     .107945    .016247      .475  .6352
(Constant)     2.218454     .073648               30.123  .0000
          * * * *   M U L T I P L E   R E G R E S S I O N   * * * *
Equation Number 1    Dependent Variable..   RESISTAN   Resistance skills
Block Number  1.  Method:  Enter      GROUP
Variable(s) Entered on Step Number
   1..    GROUP     Group

------------------ Variables in the Equation ------------------

Variable             B        SE B       Beta         T  Sig T

GROUP           .185923     .092322    .068790     2.014  .0443
(Constant)     5.920861     .062989               93.998  .0000
```

References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*(5), 602-615.

MacKinnon, D. P. (1993). Analysis of mediating variables in prevention intervention studies. In A. C. a. L. A. Beatty (Ed.), *Scientific Methods in Prevention and Intervention Research* (pp. 127-153). Washington, DC: NIDA, DHHS Pub. 94-3631.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*(2), 144-158.

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*(1), 41-62.

MacKinnon, D. P. L., C. (under review). A distribution of products test for the indirect effect: Power and Type I error rates. *Sociological Methodology*.

McGuigan, K., & Longholtz, B. (1988). *A note on testing mediation paths using ordinary least-squares regression* .

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models, *Sociological Methodology* (pp. 290-313).