

2013

Forest plots in Excel: Moving beyond a clump of trees to a forest of visual information

James H. Derzon

Aaron A. Alford

Follow this and additional works at: <https://scholarworks.umass.edu/pare>

Recommended Citation

Derzon, James H. and Alford, Aaron A. (2013) "Forest plots in Excel: Moving beyond a clump of trees to a forest of visual information," *Practical Assessment, Research, and Evaluation*: Vol. 18 , Article 7.

DOI: <https://doi.org/10.7275/96vm-5c74>

Available at: <https://scholarworks.umass.edu/pare/vol18/iss1/7>

This Article is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Practical Assessment, Research, and Evaluation by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for non-profit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 18, Number 7, April 2013

ISSN 1531-7714

Forest plots in Excel: Moving beyond a clump of trees to a forest of visual information

James H. Derzon & Aaron A. Alford

Battelle Centers for Public Health Research and Evaluation

Forest plots provide an effective means of presenting a wealth of information in a single graphic. Whether used to illustrate multiple results in a single study or the cumulative knowledge of an entire field, forest plots have become an accepted and generally understood way of presenting many estimates simultaneously. This article explores advanced uses of the forest plot with the intent of highlighting the flexibility of Excel in generating both simple and complex forest plots. A step-by-step tutorial is included with specific directions for generating a stratified forest plot and general suggestions for modifying the forest plot to meet the user's specific needs.

Forest plots provide an effective means of presenting a wealth of information in a single graphic. Whether used to illustrate multiple results in a single study or the cumulative knowledge of an entire field, forest plots have become an accepted and generally understood way of presenting many estimates simultaneously. In meta-analysis, the forest plot has become the standard for simultaneously presenting aggregate estimates along with individual estimates. The forest plot is a relatively new graphical tool. Early versions appeared in the 1970s (Freiman, Chalmers, Smith, Kuebler, 1978), were discussed in seminal books on meta-analysis (Hedges & Olkin, 1985) and quickly became a staple of published meta-analyses (Lewis & Ellis, 1982; Lewis & Clarke, 2001).

However, until recently, forest plots were difficult to generate in most statistical programs (Lewis & Clarke, 2001). Currently, forest plot graphing capabilities are built into a wide array of statistics programs (e.g., SPSS, Stata, SAS); however, these options tend not to be flexible. Graphics programs such as Sigma-plot, on the other hand, are exceedingly flexible, but expensive.

Several free or low-cost add-on programs allow users to complete meta-analyses in Excel and include a forest plot function for presenting subsequent results (Mix 2.0, MetaXL). Like most statistical programs focused on meta-analytic procedures, these programs produce a forest plot that reports the findings of a single analysis in a fixed format with limited options for varying the components of a forest plot. Among these options are varying color, line shape and

size, titles, marker shape and size, the number of estimates per line and position of individual elements.

Neyeloff, Fuchs and Moreira (2012) published a step-by-step guide for building simple forest plots in Excel¹. Their tutorial provides a guide for creating the standard format produced by most meta-analytic programs. Example worksheets accompany the tutorial and are available online². Bax, Yu, Ikeda and Moons (2007) have published a comprehensive review of the utility of various software for meta-analysis, including discussion of their graphing capabilities³. For those interested in the history of the forest plot, Lewis & Clarke (2001) provide a concise but detailed review⁴.

The present article explores increasingly advanced uses of the forest plot with the intent of highlighting the flexibility of Excel in generating both simple and complex forest plots. A step by step tutorial is included with specific directions for generating a stratified forest plot and general suggestions for modifying the forest plot to meet the user's specific needs. Excel provides an alternative for generating and

¹ Available for free at: www.biomedcentral.com/1756-0500/5/52

² Available for free at: www.biomedcentral.com/1756-0500/5/52/additional

³ Available for free at: www.biomedcentral.com/1471-2288/7/40/

⁴ Available for free at: www.ncbi.nlm.nih.gov/pmc/articles/PMC1120528/pdf/1479.pdf

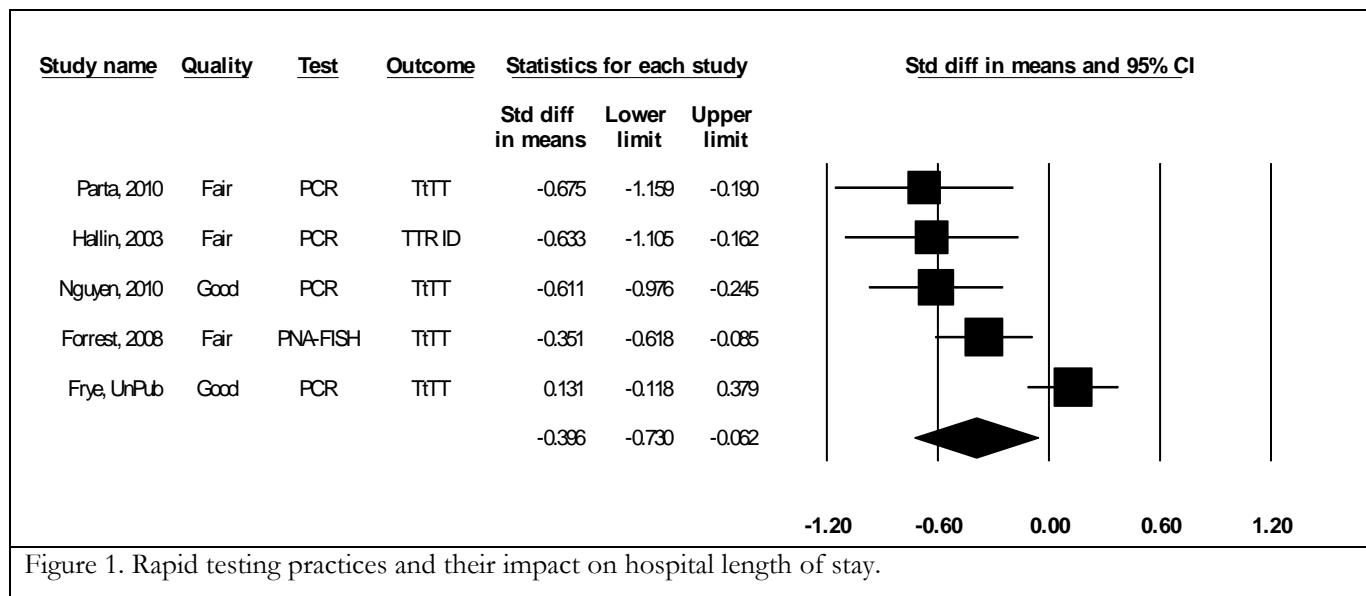
manipulating all of the above mentioned forms of information in the forest plot. Until recently, Excel did not offer a strong alternative to statistical programs, but the recent versions of Excel are much more flexible in creating complex graphs and Excel is ubiquitous in professional and academic environments.

Beyond a single forest

The standard forest plot displays a series of estimates of central tendency and their associated confidence intervals. Often included with the series is an estimate of the series' grand mean. For example, Figure 1 below displays a subset of data examining the impact on length of stay from laboratory practices to increase timeliness of providing targeted therapy for inpatients with bloodstream infections. This figure was generated using Comprehensive Meta-analysis 2.2.064, a commonly used program for completing meta-analyses.⁵

comparator. A confidence interval that includes the zero line is generally not significantly different from null. Finally, the grand mean is described by the vertical axis of the diamond at the bottom of the graph with its associated 95 percent confidence interval described by the horizontal axis of the diamond.

To the right of the forest plot, there are several columns of text and numbers. These columns report the data sources (first author and year of publication) with additional columns describing study characteristics and displaying standardized values (effect sizes) of practice impact and precision. Since interval data contributed to this example, the standardized difference in means (Cohen's *d*) is used to summarize study findings. The upper and lower limits represent the 95 percent confidence interval associated with the precision of each value.



In Figure 1, each box represents the measured impact of the practice while the 95 percent confidence intervals show the precision of each estimate. The size of the boxes corresponds to each study's weight, while the random effect estimate at the bottom of the graph displays the grand mean and 95 percent confidence intervals of the included studies. Negative findings to the left of the 0.00 line indicate the tested practice reduced length of stay relative to its comparator, while findings to the right favor the

There are many scenarios where it is useful to display several aggregate means in addition to all of their constituent estimates. This need often arises in the case of sensitivity analyses or when describing the grand mean and estimates of several related constructs. While it is possible to present the individual aggregate estimates in a single standard plot, the visual information within group variance is lost. Likewise, each grand mean and its constituent estimates can be presented in separate standard forest plots (As in Figure 1). However, this results in multiple plots, making the differences between groups difficult to discern.

Figure 2 displays the results of a sensitivity analysis between several related constructs. In this study, over 70 school

⁵ An introductory description of meta-analysis and forest plot interpretation is provided by the Cochrane Collaborative at: http://130.226.106.152/openlearning/PDF/Module_3.pdf

systems received Safe Schools/Healthy Student grants to implement drug and violence prevention programs (data from Derzon, et al, 2012). Each district was afforded the freedom to allocate funds and choose intervention programs according to their local agenda. This figure is the result of several analyses of the combined impact of the grant on three grantee outcomes. As with the standard forest plot, the grand mean of each construct is presented along with the constituent estimates. In this scenario, the grand means are labeled 'Grantee' and negative results indicate reductions in substance use. This graph allowed the client to simultaneously view the overall effect of the program and whether results differed by three school types. By retaining all three constructs on the same forest plot, the viewer can instantly compare the sample sizes and differences in variance and central tendency across each of the constructs.

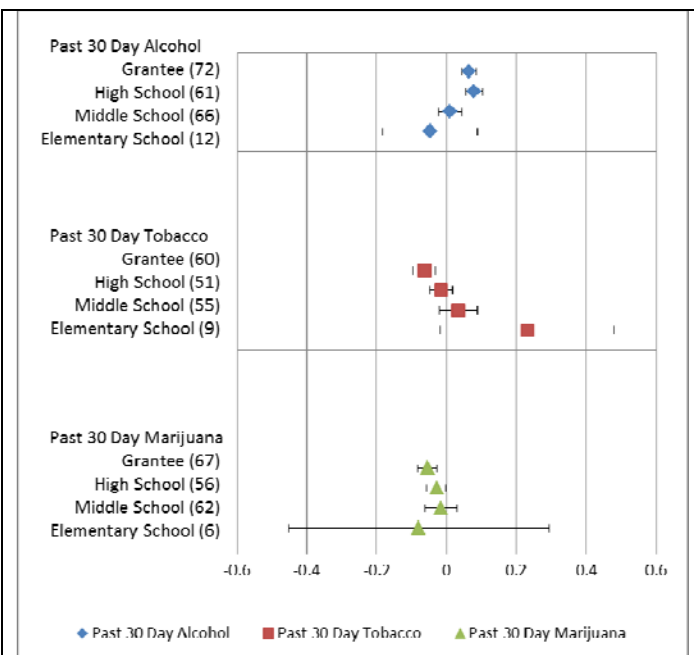


Figure 2. Forest plot of data from 72 SS/HS grantees showing pre- and post-test change in three outcomes broken out by school type.

Adding an integrated descriptive construct and ordering by strength of effect

Ordering estimates and manipulating the key provide several useful means of adding additional information to a forest plot. Figure 3 displays both modifications. This figure displays the results of many separate meta-analyses of the impact of personal and family characteristics on violent behavior (data from Derzon, 2010). Each construct on the y-axis represents the summary estimate from a single meta-

analysis. The symbols representing the effect sizes differ to show that age is a mediator of these relationships. The inclusion of a key in the graph itself helps the reader decipher the meaning of the symbols more quickly.

This graph was designed to address the commonly held belief that family life has a direct and deterministic impact on violent and antisocial behavior. As can be seen in the graph, family life is moderated by age and clearly not more strongly related to violence behavior than are individual-level predictors. The correlations show that nearly all the commonly cited precursors of violent and antisocial behaviors are, in fact, significantly and positively predictive of these outcomes, but perhaps not to the degree many believe.

Paired plots

At times it is useful to compare two estimates next to each other in order to reinforce their differences or similarities. This approach is particularly powerful when you are attempting to display how this difference operates over many pairs as in Figure 4 (data from Derzon, 2007). In this plot, two mean correlations are presented for each predictor labeled on the y-axis. The box symbol is used to show each predictor's relationship with marijuana use while the diamond indicates its relationship with alcohol use. The symbol denoting each estimate uniquely identifies the two groups being compared. Displaying both estimates for each construct on the y-axis serves to highlight those pairs that are very similar or different. For example, the individual characteristic 'sensation seeking' has a small impact on marijuana use and a significantly stronger impact on alcohol use. Individual characteristics tend to have statistically similar relations with later alcohol and marijuana use, while characteristics of the family show significantly different relations of these outcomes for three of the six predictors summarized. The paired display of these estimates makes observing these differences immediately and visually apparent.

This approach is by no means limited to pairs. Unfortunately, as the number of estimates presented per unit of axis increase, the visual information quickly becomes overwhelming to the casual reader. While they may not be optimal as conference slides, these noisy plots are incredibly powerful for visualizing and analyzing large amounts of information. The large bulk of information can be useful for reinforcing how similar various estimates are within construct (Figure 5) or for exploring linear trends.

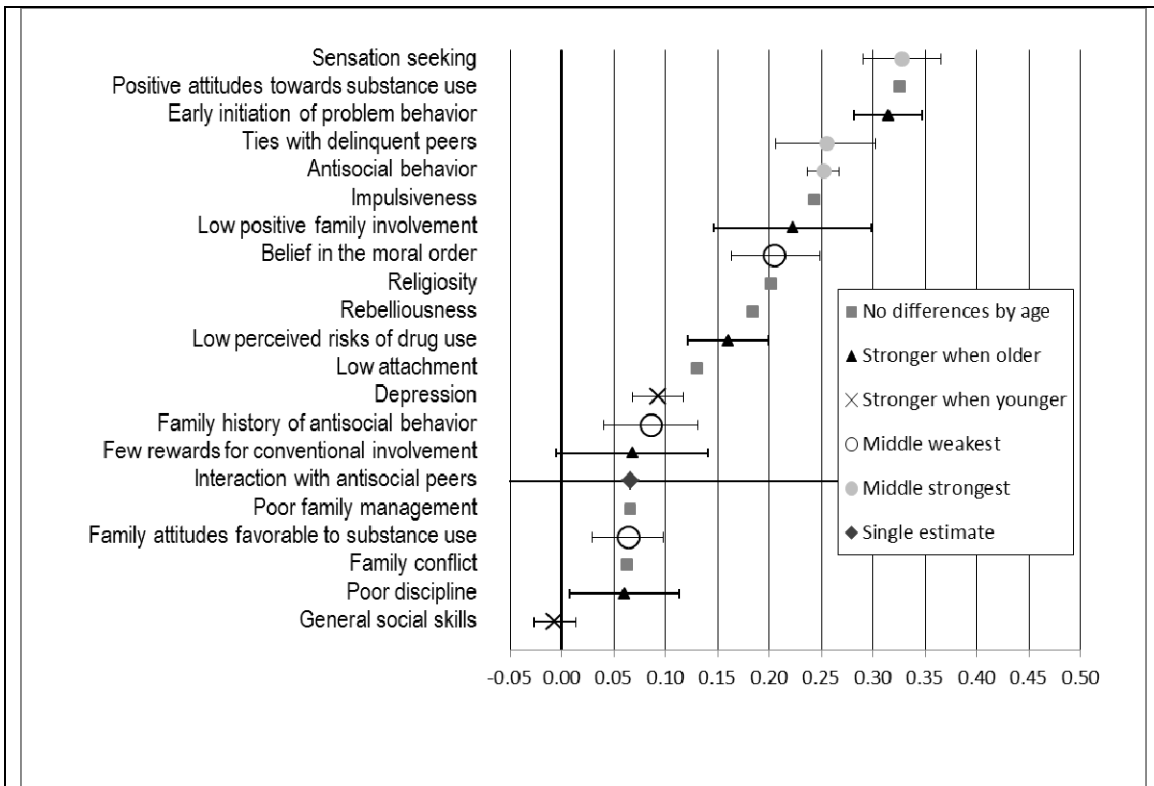


Figure 3. The impact of age of measurement of individual and family predictors on later violent behavior rank ordered by strength of the estimate.

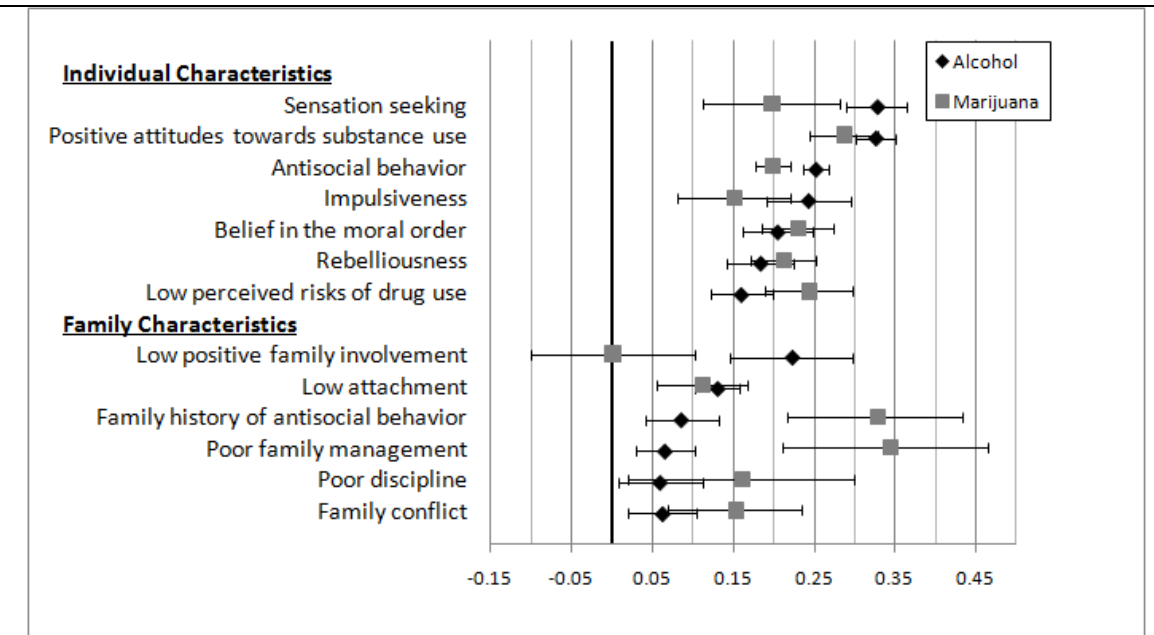


Figure 4. A comparison of individual and family characteristics in predicting later alcohol and marijuana use.

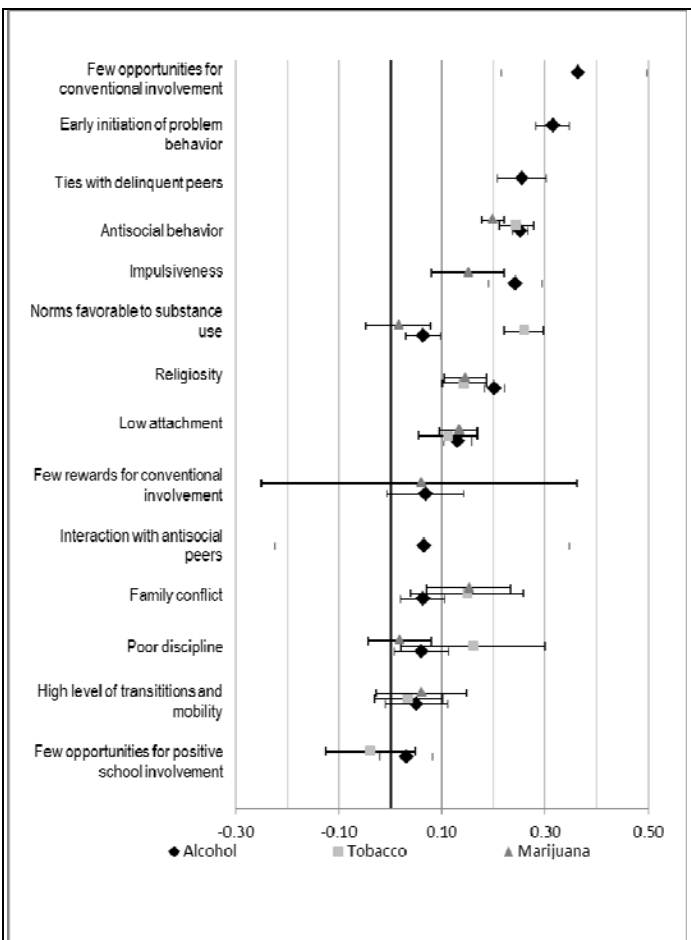


Figure 5. Consistent predictors of later alcohol, tobacco and marijuana use.

Organizing your data

The initial task is to construct your data in a format that can be used to create a forest plot. Table 1 displays example data from Figure 5 (see Supplemental Materials for a completed example with table and graph in Excel). The necessary columns are:

1. Study Title (or element to be graphed): Provides the name used for each element on the y-axis, including hierarchy titles.
2. Order (position on graph): Organizes the data in the order that you prefer. This is especially useful if you wish to explore several different orders. This variable allows the user to quickly resort the data into alternate configurations.

3. Group: Contains a numeric indicator for grouping and sorting. This column is only necessary if the graph contains stratified estimates.
4. Group name: Contains the group name to be displayed in the legend. This column is only necessary if the graph contains stratified estimates.
5. Sample size: Contains the sample size for each individual estimate, as well as any aggregate estimates (group or grand mean estimates). This column is not integral to the graphing procedure. However, it is useful to maintain this data in the tables in case the user wishes to include this information visually (see the 'Other Options' section below).
6. Initial y-axis: Provides the initial spacing for the y-axis. In general, this should be numbered 1-x to reflect the number of labels (x) on the y-axis, including strata names for stratified graphs. The next column will be used to re-scale this axis estimate more finely.
7. Final y-axis: Use this column to quickly and uniformly rescale the estimates in the previous column. The easiest way to do this is to cut and paste a calculation into this field that divides every value in 'Initial y-axis' by the same number. In general, aim for the values in this column to range between 0 and 3.
8. Effect size: This column contains the estimates of interest for display. This column contains both individual and aggregate estimates (in the case of stratified plots).
9. Lower confidence interval: Lower confidence interval for individual and aggregate estimates.
10. Upper confidence interval: Upper confidence interval for individual and aggregate estimates.

Constructing Forest Plots in Excel

The following instructions provide a general 'how-to' for constructing simple and complex forest plots in Excel⁶. For the sake of illustration, these instructions follow the process used to build Figure 5 and reference the data in Table 1.

⁶ These directions were generated for MS Excel 2007, but the menus remain essentially unchanged for later versions.

Table 1. Example data from Figure 5.

Element Title	Order	Grp	Group Name	Sample size	Initial Y-axis	Final Y-axis	EFFECT SIZE	Lower confidence interval	Upper confidence interval
Few opportunities for conventional involvement	1	1	Alcohol	1	14	2.8	0.37	0.15	0.13
Early initiation of problem behavior	2	1	Alcohol	7	13	2.6	0.32	0.03	0.03
Ties with delinquent peers	3	1	Alcohol	5	12	2.4	0.26	0.05	0.05
Antisocial behavior	4	1	Alcohol	49	11	2.2	0.25	0.02	0.02
Impulsiveness	5	1	Alcohol	4	10	2	0.24	0.05	0.05
Norms favorable to substance use	6	1	Alcohol	9	9	1.8	0.06	0.03	0.03
Religiosity	7	1	Alcohol	23	8	1.6	0.20	0.02	0.02
Low attachment	8	1	Alcohol	20	7	1.4	0.13	0.03	0.03
Few rewards for conventional involvement	9	1	Alcohol	4	6	1.2	0.07	0.07	0.07
Interaction with antisocial peers	10	1	Alcohol	1	5	1	0.07	0.29	0.28
Family conflict	11	1	Alcohol	9	4	0.8	0.06	0.04	0.04
Poor discipline	12	1	Alcohol	6	3	0.6	0.06	0.05	0.05
High level of transitions and mobility	13	1	Alcohol	3	2	0.4	0.05	0.06	0.06
Few opportunities for positive school involvement	14	1	Alcohol	4	1	0.2	0.03	0.05	0.05
Antisocial behavior	4	2	Tobacco	11	11.1	2.22	0.24	0.03	0.03
Norms favorable to substance use	6	2	Tobacco	7	9.1	1.82	0.26	0.04	0.04
Religiosity	7	2	Tobacco	4	8.1	1.62	0.14	0.04	0.04
Low attachment	8	2	Tobacco	5	7.1	1.42	0.11	0.06	0.06
Family conflict	11	2	Tobacco	1	4.1	0.82	0.15	0.11	0.11
Poor discipline	12	2	Tobacco	2	3.1	0.62	0.16	0.14	0.14
High level of transitions and mobility	13	2	Tobacco	3	2.1	0.42	0.04	0.07	0.07
Few opportunities for positive school involvement	14	2	Tobacco	1	1.1	0.22	-0.04	0.09	0.09
Antisocial behavior	4	3	Marijuana	22	11.2	2.24	0.20	0.02	0.02
Impulsiveness	5	3	Marijuana	4	10.2	2.04	0.15	0.07	0.07
Norms favorable to substance use	6	3	Marijuana	4	9.2	1.84	0.02	0.06	0.06
Religiosity	7	3	Marijuana	6	8.2	1.64	0.15	0.04	0.04
Low attachment	8	3	Marijuana	7	7.2	1.44	0.13	0.04	0.04
Few rewards for conventional involvement	9	3	Marijuana	1	6.2	1.24	0.06	0.31	0.30
Family conflict	11	3	Marijuana	2	4.2	0.84	0.15	0.08	0.08
Poor discipline	12	3	Marijuana	3	3.2	0.64	0.02	0.06	0.06
High level of transitions and mobility	13	3	Marijuana	1	2.2	0.44	0.06	0.09	0.09

1. Insert a scatter plot. Use the 'Scatter with only markers' option in the 'Insert' menu.
2. Right-click on the new chart and click the 'Select Data' option.
3. Using the resulting dialog box, create the first series to be plotted by clicking 'Edit'. This brings up the 'Edit Series' dialog box. In this example, there are three series. Each corresponds to the groups indicated in the columns 'Group' and 'Group Name.' Name the series by selecting only one of the cells in the column 'Group Name.' For Figure 5 and Table 1, the first series is 'Alcohol'. Select the x-axis data (effect size) for the first group. Do not include date for the other groups in this selection. Select the y-axis data from the 'Final y-axis' column. Do not include date for the other groups in this selection. Select OK. (NOTE: If you are creating a non-stratified plot with only one estimate per construct, you will only have one series. This is true even if you are including a grand mean estimate).
4. If you have multiple strata or want multiple nested values per y-axis label, add the second series following the steps outlined above.
5. To add confidence intervals, select 'Layout' from menu bar.
 - a. Select 'Error Bars' (far right side of menu bar)
 - b. Select 'More Error Bar Options'
 - c. Select 'Custom and Specify Value' by clicking on the chart icon and highlighting the Graph L95-CI and the Graph U95-CI
 - i. A quirk of Excel: sometimes it will put in both vertical and horizontal error bars and the horizontal error bar (the one you want displayed) is rarely the data you specified.
 1. The easiest way to fix this is to click on the error bar you want to modify, right click and select 'Format Error Bars,' select 'Custom,' select 'Specify Value' and highlight the Graph L95-CI and the Graph U95-CI again.
 - ii. You can delete the unwanted error bar by selecting it and hitting 'Delete.' Alternatively, you can click in the box labeled 'Chart Area' in the menu bar (in 'Layout'), select the y-error bars (this will select the y-error bars in the graph). They can then be deleted by clicking back into the graph and hitting 'Delete.'
6. Nesting multiple estimates within a single series can be done by editing the column 'Final y-axis data.'
7. Once all the data are graphed, this is a good time to make any adjustments to the graph itself.
 - a. The y-and x-axes can be edited (e.g., changing the range displayed) by selecting the axis right clicking, selecting 'Format Axis' and changing the options to create a visually appealing image.
 - b. The y-axis (numbers) can be deleted by selecting the axis and hitting 'Delete.'
 - i. The axis can also be deleted or added by selecting 'Layout,' 'Axis' and the option desired.
 - c. Grid lines can be added or deleted by selecting 'Layout,' 'Gridlines' and the desired option. Once inserted, they can be edited by selecting a 'Major' or 'Minor Gridline,' right clicking and selecting the options desired.
 - d. A reference line (e.g., at zero) can be added by selecting 'Insert,' 'Shape,' selecting the 'Line' option and drawing the line where desired. It is often helpful to zoom the graphic (either by selecting 'View,' 'Zoom' or by holding the control key and using the mouse scroll wheel).
 - e. Symbols and error bars can be edited by selecting either the symbol or error bar, right clicking, selecting format data series and choosing the options 'Fill,' 'Marker Color,' etc. to create a visually appealing image.
8. The easiest way to add labels is select the graph and resize it by pulling in the left-hand side. Then the labels (e.g., study names) can be added by inserting them in text boxes. Although they can be pasted en masse, it's generally easier to add them separately as they are than to work with alignment, font size, etc.
 - a. Create a 'Template' text box by creating a single text box (e.g., Study 1) and resize the box so that it is the same size as the y-axis major grid lines. This makes rough alignment of the boxes to the data easy.
 - i. Click in the label and select 'Format Text Effects.' The vertical alignment of the label can then be centered by selecting 'Text

Box, Vertical Alignment.’ Other text options can be selected as well.

- b. Copy, place and edit the remaining text boxes.
 - i. Note that selecting the text box is a bit tricky. Select the box by clicking near the edge of the text box so the edge becomes a solid line. When you get the cross, something is selected to be moved. A label saying chart area or a text box will appear. You will want to move the item that is the text box. It is often easiest to do this from a side rather than the top.
 - c. Text boxes can be aligned and evenly distributed by selecting all (Shift-select each box) and choosing the options desired in ‘Format, Align.’
9. Pasting the final image into a Microsoft Word or PowerPoint file sometimes creates surprises. Selecting ‘Paste Special’ and one of the alternative formats may produce better results. Bitmap will reproduce the graph exactly as shown in Excel, but is the most memory intensive. An alternative is to use the Visual Basic for Applications (VBA) option to create a macro ‘export’ (under “Developer,” Peltier, 2008b). To create a .PNG image of the graph (.PNG is recommended for most publications), create the folder ‘My Charts’ on the C: drive, return to your Excel file and select the image you wish to export. Under ‘Developer’ select ‘Macro,’ create a macro titled ‘Export’ and paste the code below into the macro. Be sure to size the chart to its final dimensions prior to running the macro and exporting the image. More information and advanced macros for exporting charts can be found in Peltier (2008a):

```
Sub Export()
ActiveChart.Export "C:\My
Charts\SpecialChart.png"
End Sub
```

Other Options

It is possible to show relative sample size or absolute sample size in Excel forest plots. To display absolute sample sizes, record the sample size for each y-axis label in the label names (e.g. “Recidivism (N=12)”).

To display relative size, the size of the point estimates can be scaled to reflect the sample size of each estimate. To do this, select a single data point in the graph, starting with the data

point that corresponds to the record with the largest sample. Then right-click on the data point. In the resulting menu, select the option ‘Format Data Point.’ Then select ‘Marker Options.’ In this menu, you can select the size of the marker, allowing you to scale markers relative to one another. For the record with the largest sample size, select the largest symbol size that you wish to include in your graph. Scale the remaining symbols relative to that symbol.

Summary

The flexibility and availability of Excel makes it an attractive option for generating standard and complex forest plots for displaying meta-analytic and other research results. Other programs are available to meet this need, but many are specialized, may require a license purchase or be more restrictive in the types of data they manage. Excel, on the other hand, can be used for multiple common activities beyond graphing forest plots, resides on most modern computers and provides multiple data analysis and visualization options beyond creating forest plots.

References

- Bax, L., Yu, L.M., Ikeda, N., Moons, K.G.M. (2007). A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology* 7(1):40. (Available for free at: <http://www.biomedcentral.com/1471-2288/7/40/>)
- Derzon, J. H. (2007). Using correlational evidence to select youth for prevention programming. *J Prim Prev*, 28(5), 421-447. doi: 10.1007/s10935-007-0107-7
- Derzon, J. H. (2010). The correspondence of family features with problem, aggressive, criminal, and violent behavior: a meta-analysis. *Journal of Experimental Criminology*, 6(3), 263-292. doi: 10.1007/s11292-010-9098-0
- Derzon, J. H., Yu, P., Ellis, B., Xiong, S., Arroyo, C., Mannix, D., . . . Rollison, J. (2012). A national evaluation of Safe Schools/Healthy Students: outcomes and influences. *Evaluation and Program Planning*, 35(2), 293-302. doi: 10.1016/j.evalprogplan.2011.11.005
- Freiman JA, Chalmers TC, Smith HJ, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials. *N Engl J Med* 1978; 299:690-4.
- Hedges, L. V., Olkin, I., Statistiker, M., Olkin, I., & Olkin, I. (1985). *Statistical methods for meta-analysis*.
- Lewis, S., & Clarke, M. (2001). Forest plots: trying to see the wood and the trees. *Bmj*, 322(7300), 1479-1480.

Lewis, J. A., & Ellis, S. H. (1982). A statistical appraisal of post-infarction beta-blocker trials. *Prim Cardiol*, (suppl 1), 31-37.

Neyeloff, J. L., Fuchs, S. C., & Moreira, L. B. (2012). Meta-analyses and Forest plots using a microsoft excel spreadsheet: step-by-step guide focusing on descriptive data analysis. *BMC Research Notes*, 5(1), 52. (Available for free at: <http://www.biomedcentral.com/1756-0500/5/52>; Supplemental materials available for free at:

<http://www.biomedcentral.com/1756-0500/5/52/additional>).

Peltier, J. (2008a, June 9, 2008). Enhanced Export Chart Procedure Retrieved 11/28/2012, from <http://peltiertech.com/WordPress/enhanced-export-chart-procedure/>

Peltier, J. (2008b, June 5, 2008). Export Chart as Image File Retrieved 11/28/2012, from <http://peltiertech.com/WordPress/export-chart-as-image-file/>

Note:

The Excel worksheet used in this example is available at <http://pareonline.net/sup/v18n7/forestplot.xls>.

Citation:

Derzon, James H. & Alford, Aaron A. (2013). Forest plots in Excel: Moving beyond a clump of trees to a forest of visual information. *Practical Assessment, Research & Evaluation*, 18(7). Available online: <http://pareonline.net/getvn.asp?v=18&n=7>

Corresponding Author:

James H. Derzon
Battelle Centers for Public Health Research and Evaluation,
2111 Wilson Blvd. #900
Arlington, VA, 22201
Email: DerzonJ [at] battelle.org