Extracting data from figures using software

Prof. Livia Puljak
University of Split School of Medicine
Cochrane Croatia
Data presented only as figures

• Data extraction important for systematic reviews (SRs)
• Published randomized controlled trials (RCTs) sometimes contain numerical data presented only in figures
• Authors of SRs may request those data from study authors, which seldom yields results
Cochrane Handbook

- No guidance about data extraction from figures
- So, what do the Cochrane authors do when confronted with data presented only in figures?
Our first study

- Vucic et al. Survey of Cochrane protocols found methods for data extraction from figures not mentioned or unclear. Journal of Clinical Epidemiology. 2015;68(10):1161-1164.

- Analyzed 589 protocols
- 33 (5.6%) mentioned data extraction from figures in Methods section
- One protocol specified that computer software will be used for data extraction from figures
Manual vs software extraction: which is better?

• Our second study
• Jelicic Kadic et al. Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. Journal of Clinical Epidemiology. 2016;74:119-23
Our methods

- Data points from graphs/figures published in RCTs
- Extracted by two authors independently
- Two methods: manual estimation and software extraction with the Plot Digitizer, open source software
- Corresponding authors of each RCT were contacted up to four times via e-mail to obtain exact numbers that were used to create graphs
- Accuracy of each method was compared against the source data from which the original graphs were produced
Our results

• Software data extraction was significantly faster, reducing time for extraction for 47%

• Percent agreement between the two raters was 51% for manual and 53.5% for software data extraction

• Percent agreement between the raters and original data was 66% vs. 75% for the first rater and 69% vs. 73% for the second rater, for manual and software extraction, respectively
Where to find Plot Digitizer

• An open source software
• Available at: 
  http://plotdigitizer.sourceforge.net/
• Works with X-Y scatter or line plots
• It is free 😊
How it works

1. Install Plot Digitizer
2. Digitize your photos
   – Software will recognize GIF, JPEG, or PNG format
3. Import a figure into software
4. Calibrate a figure – „show” the software where is X and Y axis and how long are they
5. Click on a data point and software will provide numerical values for X and Y axis
How to digitize the photos?

• You can download them directly from Internet, if possible and available online in a useful format
• Or save on a computer with a snipping tool
• Scan from the paper
An example: figure
What you need to extract

• For data points that you need
• Extract the number for the value and error bar
How to install the program?

• Follow instructions from the website
• Click on the download
• Choose a folder where to save it
• Zip file will be saved to your computer
• Unpack the zip file
Contents of the folder

• To use the program:
  – Clik on **PlotDigitizer.exe**
This screen will pop up
Next step: import figure

• Click on „File”
• Then click on „Open” from the drop-down menu
• Choose the figure from your computer
After importing figure, screen looks like this:
Next step: calibrate X axis

- Choose most negative end of X axis line (click on it)
- New window will open up
- Enter the value, in this case 0
- Click „Okay“
Then enter value for Y min
Choose most positive end of X axis

- Click on the last value of X axis line (here: 72)
- New window will open up
- Enter the value, in this case 72
- Click "Okay"
Choose most positive end of Y axis

- Click on the last value of Y axis line (here: 10)
- New window will open up
- Enter the value, in this case 10
- Click „Okay“
A note

• Even though in this case you need only values from Y axis
• You still need to calibrate also values from the X axis
• The program requires it, even though irrelevant for you
Next: enter axis names and values

• After you entered value for the most positive end of Y axis
• New pop-up window appears
• You need to enter X axis name
• Choose a name, or just leave X (written on default)
• The same goes for Y axis
Now your plot is digitized

• Please note red line and number (X axis), and blue line and number (Y axis)
To obtain a value for points you need

• Now click on all points for which you need values
• You can choose multiple points
• When you are finished choosing, click on Done
• Your values will appear in a new pop-up window
• For the purpose of this demonstration, I clicked on three dots showed with yellow lines here on the right
Digitized points show up – I need Y axis here; X axis is for me here irrelevant
You need to calculate values of error bars yourself

- The program gives me value for a certain data point on this graph
- If you need a value for error bar, you will have to calculate it
- Here:
  - 6.54 main effect
  - 6.99 first error bar
  - 6.00 second error bar
  - Subtract
You can save digitized points

• Go to window with numbers (digitized points)
• Go to File
• Choose Save as
• Save in a desired location
Thank you for your attention

livia@mefst.hr