Creating Effective Graphs

Sunita Patterson
What we will cover

- When to use a graph
- Types of graphs
- General principles for effective graphs
What we will cover

- When to use a graph
- Types of graphs
- General principles for effective graphs
When to use text

- Only a few data points
- No variation in the data
At diagnosis, the mean tumor volumes were 4.5 cm$^3$ for group A and 2.6 cm$^3$ for group B. At the end of treatment, mean volumes had not significantly changed (4.7 cm$^3$ for group A and 2.7 cm$^3$ for group B).
When to use a table

- To present more than a few precise numeric values
- To show detailed item-to-item comparisons or complex relationships between data
When to use a table

Table 1. Adverse events observed in a randomized phase III clinical trial of UFT vs. S-1

<table>
<thead>
<tr>
<th></th>
<th>UFT (n = 254)</th>
<th>S-1 (n = 251)</th>
<th>p-value†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All grade</td>
<td>Grade 3+4</td>
<td>All grade</td>
</tr>
<tr>
<td></td>
<td>n (% )</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Leukopenia</td>
<td>40 (15.7)</td>
<td>2 (0.8)</td>
<td>91 (36.3)</td>
</tr>
<tr>
<td>Neutropenia</td>
<td>14 (5.5)</td>
<td>0 (0.0)</td>
<td>58 (23.1)</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>17 (6.7)</td>
<td>0 (0.0)</td>
<td>32 (12.7)</td>
</tr>
<tr>
<td>Anemia</td>
<td>33 (13.0)</td>
<td>1 (0.4)</td>
<td>70 (27.9)</td>
</tr>
<tr>
<td>Total bilirubin increase</td>
<td>17 (6.7)</td>
<td>1 (0.4)</td>
<td>45 (17.9)</td>
</tr>
<tr>
<td>AST increase</td>
<td>23 (9.1)</td>
<td>3 (1.2)</td>
<td>25 (10.0)</td>
</tr>
<tr>
<td>ALT increase</td>
<td>14 (5.5)</td>
<td>2 (0.8)</td>
<td>20 (8.0)</td>
</tr>
<tr>
<td>Fatigue</td>
<td>34 (13.4)</td>
<td>3 (1.2)</td>
<td>79 (31.5)</td>
</tr>
<tr>
<td>Anorexia</td>
<td>34 (13.4)</td>
<td>3 (1.2)</td>
<td>74 (29.5)</td>
</tr>
<tr>
<td>Weight loss</td>
<td>14 (5.5)</td>
<td>1 (0.4)</td>
<td>28 (11.2)</td>
</tr>
<tr>
<td>Rash/desquamation</td>
<td>6 (2.4)</td>
<td>1 (0.4)</td>
<td>40 (15.9)</td>
</tr>
<tr>
<td>Hyperpigmentation</td>
<td>3 (1.2)</td>
<td>0 (0.0)</td>
<td>48 (19.1)</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>7 (2.8)</td>
<td>1 (0.4)</td>
<td>24 (9.6)</td>
</tr>
<tr>
<td>Mucositis/Stomatitis</td>
<td>5 (2.0)</td>
<td>0 (0.0)</td>
<td>32 (12.7)</td>
</tr>
<tr>
<td>Nausea</td>
<td>8 (3.1)</td>
<td>0 (0.0)</td>
<td>27 (10.8)</td>
</tr>
<tr>
<td>Vomiting</td>
<td>1 (0.4)</td>
<td>0 (0.0)</td>
<td>10 (4.0)</td>
</tr>
</tbody>
</table>

† Fisher’s exact test
* Number of grade 2 events are indicated, and differences in the incidences of grade 2 are tested.
When to use a graph

- To show trends in data
- To show relationships between data
When to use a graph

![Graph showing factor X concentration over weeks with error bars.](image)
When to use a graph
What we will cover

- When to use a graph
- **Types of graphs**
- General principles for effective graphs
Figure 4. Percentage distribution of five leading causes of death in the U.S. in 2011 by age group.
Figure 5. Mean body weights of weanling guinea pigs fed protein-adequate (n=10) and protein-deficient (n=10) diets for 10 months. Error bars, standard error of the mean.
Figure 6.
Sample Kaplan-Meier survival plot for three patient groups.
Figure 7. Validation of RNA-seq expression data via quantitative real-time RT-PCR for 20 genes. Average RNA-seq values were plotted against average RT-PCR values and fit into a linear regression.
Figure 8. Variations by physician specialty in performance of tests during typical office visits.
Horizonal bar graph

- **Screening for Depression**
  - General Surgery
  - Internal Medicine
  - Pediatrics

- **Auscultation**
  - General Surgery
  - Internal Medicine
  - Pediatrics

- **Blood Pressure**
  - General Surgery
  - Internal Medicine
  - Pediatrics

- **Temperature**
  - General Surgery
  - Internal Medicine
  - Pediatrics

Graph shows the percentage of physicians performing each test across different specialties.
3-D bar graph (not recommended)
Figure 11. Discharge status, by age group: United States, 2010.
Figure 12. Platelet counts for patients with primary or secondary infections.
Figure 13. Shannon diversity index for diarrheal and nondiarrheal samples, stratified by age group. ... The upper whisker extends from the 75th percentile to ... The lower whisker extends from ...
Figure 14. Persons infected with the outbreak strain of Salmonella Heidelberg in six states, 2011.
What we will cover

- When to use a graph
- Types of graphs
- General principles for effective graphs
General principles for effective graphs

- Design for clarity
  - Arrange components logically
  - Label clearly
  - Make data easily distinguishable
- Avoid clutter
- Use color wisely
- Include an effective legend
- Be consistent
Figure 16. Validation of expression of innate immune genes during viral persistence.
Label clearly

- Make sure labels are large enough to read (no smaller than 8 pt at printed size)
- Use a consistent capitalization style
  - Body weight, g
  - Body Weight (g)
- Indicate units
Make data easily distinguishable

- Data points: open circles, triangles, squares; closed circles, triangles, squares

- Lines
  - A few lines: black solid, dashed, dotted
  - A lot of lines: may need color

- Bars
  - A few bars: black, white, gray (avoid patterned fills)
  - A lot of bars: may need color
Avoid clutter
Use color wisely

- *Ann Oncol:* No charge for color figures in online article; **$600**/figure for color figures in print article

- *Cancer:* No charge for color

- *Cancer Cell:* **$1,000** for first color figure; **$275**/figure for others

- *Cancer Res, Clin Cancer Res:* **$675**/figure for color figures

- *J Clin Oncol:* No charge for color
Use color wisely

- Black, red, and blue reproduce well
- Avoid yellow
- Avoid red and green together
- Avoid “loud” colors
- Keep it simple
Use color wisely
Use color wisely
Use color wisely
Include an effective legend

- Overall subject or main message of graph
- Number of patients or samples
- Definitions of abbreviations and symbols
- Method
- Statistical test
Figure 12. Platelet counts for patients with primary infections (n=39) or secondary infections (n=57). Counts were measured at the time of admission. The median values are indicated by red lines. ** $P < 0.005$, Mann-Whitney test.
Be consistent

- Use the same software
- Use a consistent order of categories
- Use the same font and symbols
- Use consistent nomenclature, spelling, and capitalization
For more information


- Day RA, Gastel B. *How to Write and Publish a Scientific Paper*. 8th ed. Santa Barbara, CA: Greenwood; 2016:

Graph sources

Line graph (Fig. 3): CDC, https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6053a1.htm [public domain]

Pie chart (Fig. 4): Minino AM, https://www.cdc.gov/nchs/products/databriefs/db115.htm [public domain]

Survival curve (Fig. 6): Gross et al, https://doi.org/10.1371/journal.pone.0038960 [CC BY 4.0]

Scatterplot (Fig. 7): Chen et al, https://doi.org/10.1371/journal.pone.0144927 [CC BY 4.0]

Stacked bar graph (Fig. 11): CDC, https://www.cdc.gov/nchs/products/databriefs/db182.htm [public domain]

Dot plot (Fig. 12): Singla et al, https://doi.org/10.1371/journal.pntd.0004497 [CC BY 4.0]
Graph sources

Box-and-whiskers plot (Fig. 13): Lindsay et al, https://wwwnc.cdc.gov/eid/article/21/2/14-0795_article [public domain]

Histogram (Fig. 14): CDC, https://www.cdc.gov/salmonella/heidelberg-chickenlivers/011112/epi.html [public domain]

Line graph (Fig. 15): CDC, https://www.cdc.gov/nchs/data/hus/hus16.pdf [public domain]

Bar graph (Fig. 16): Nayak et al, https://doi.org/10.1371/journal.ppat.1003395 [CC0]

Line graph (Fig. 18): Zhao et al, https://doi.org/10.1371/journal.pone.0169962 [CC BY 4.0]

Stacked bar chart (Fig. 19): Numico et al, https://doi.org/10.1371/journal.pone.0101170 [CC BY 4.0]

Box-and-whiskers plot (Fig. 20): Zhao et al, https://doi.org/10.1371/journal.pone.0152356 [CC BY 4.0]
Scientific Publications

- [Link](http://inside.mdanderson.org/departments/scipub/index.html)
- Tel: 713-792-3305
- Email: [scientificpublications@mdanderson.org](mailto:scientificpublications@mdanderson.org)
- Location: Mid-Campus Building, 16th floor