Guidelines for Reporting Statistical Results: Tables and Figures

Jiangxia Wang, MS MA

Johns Hopkins Biostatistics Center

Wilmer Biostatistics Center

09/04/2019

Content counts most of all

"Analytical presentations ultimately stand or fall depending on the quality, relevance, and integrity of their content."

- Edward Tufte

Be CLEAR on the Message to Be Conveyed

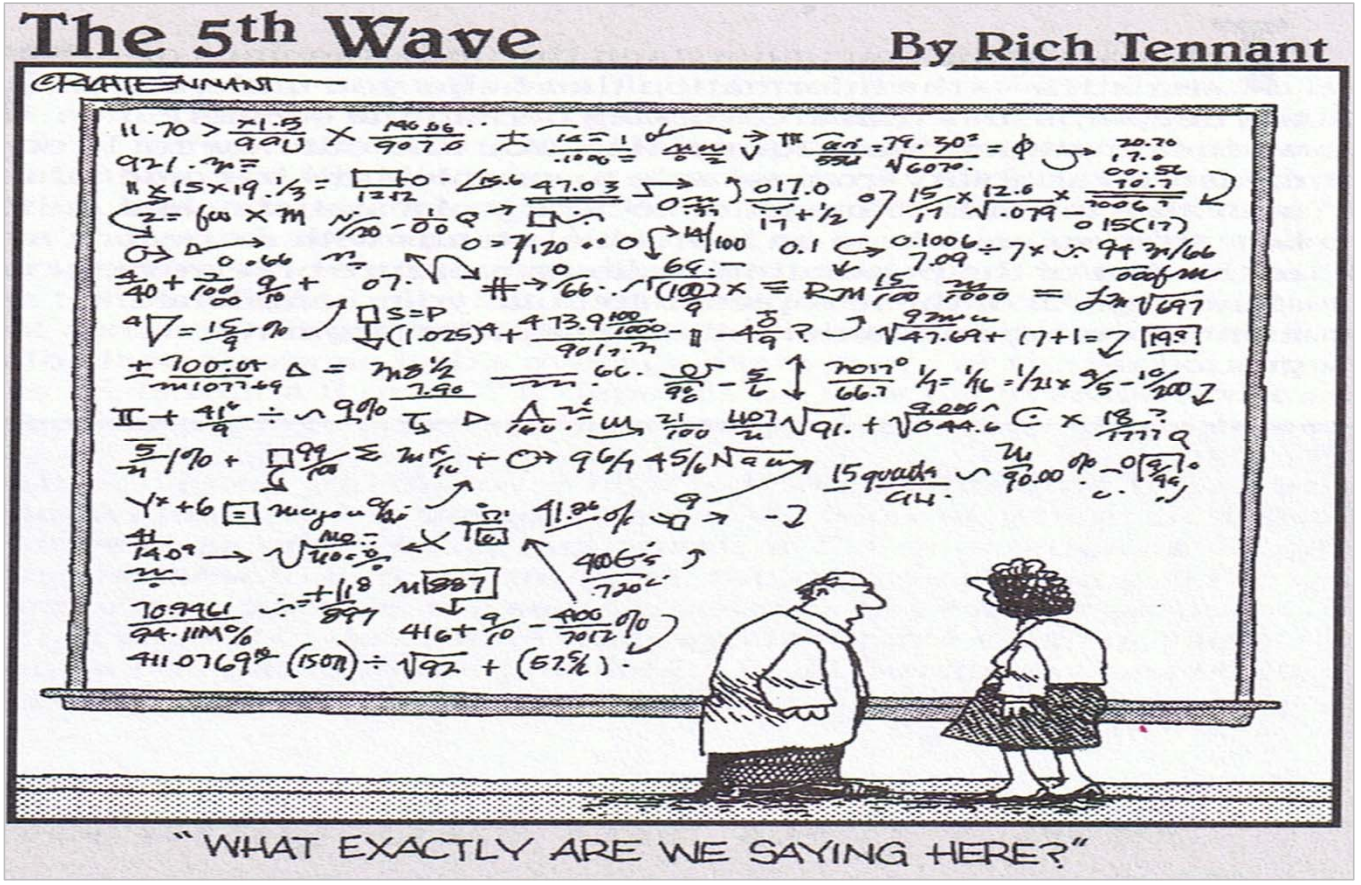


Figure Example - Before

- This example summarizes survey data collected in a convenience sample of JHU SON students
- The survey included questions about the following:
  - Psychological type: feeling, thinking, intuitive, sensitive
  - Having a Facebook account: yes, no
  - Average number of hours of sleep per night

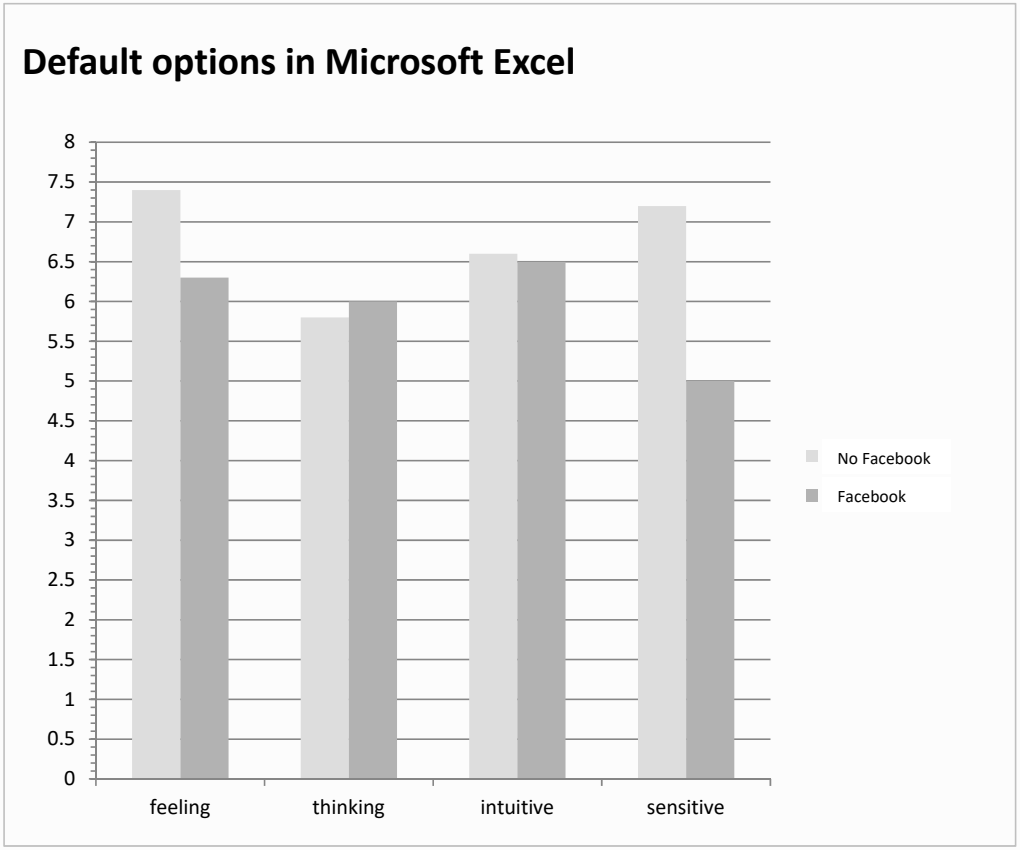
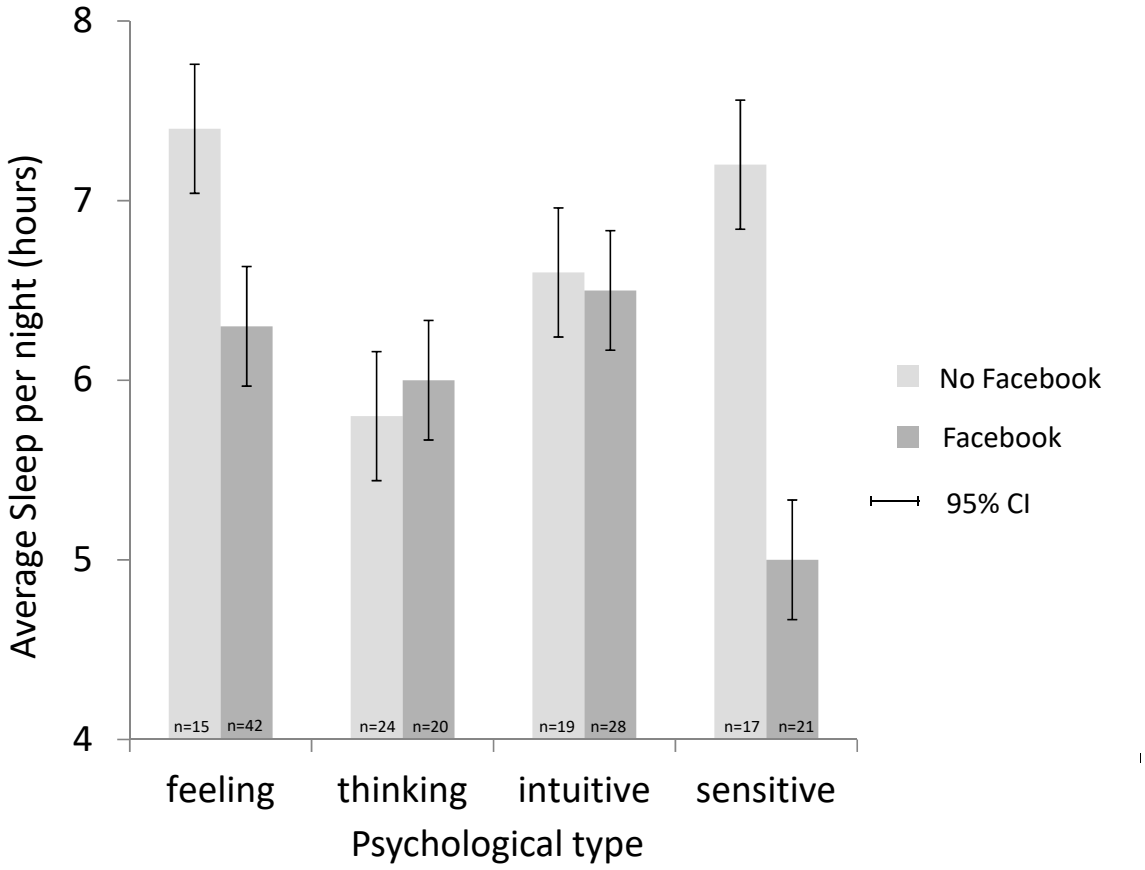


Figure Changes

- Decrease clutter: gridlines, tick marks
- Increase readability with font size
- Match y-axis more closely to range of data
- Label axes
- Add units of measure
- Provide 95% CI in addition to means
- Add information about group sample sizes

Figure Example - After



Principles of Analytical Design

- Principle 1: *Show Comparisons, contrasts, differences*
  - *Be clear about "Compared with what?"*
- Principle 2: *Causality, Mechanism, Structure, Explanation*
- Principle 3: *Multivariate Analysis - Show more than 1 or 2 variables*
- Principle 4: *Integration of Evidence - Words, numbers, images, diagrams*
- Principle 5: *Documentation – What? Who? When? Where? Assumptions?*
- Principle 6: *Content Counts Most of All – What are the content-reasoning tasks that this display is supposed to help with?*

General Reporting Principle from a Technical Editor

- Tables and figures should be able to **stand alone**
  - That is, the reader should not have to go back to the text for:
    - Acronym definitions (subject matter and statistical)
    - Details about analysis, such as, confounders used in the analysis
    - Perhaps how missing values are handled in summaries
    - Units of measurement, N's
  - Title should give full description of table or figure

Guidelines for Tables

- Limit your table to data that are relevant to the hypotheses in the study.
- Be certain that your table can stand alone without any explanation.
- Make sure that your table is supplementary to your text and does not replicate it.
- Always give units of measurement in table headings.
- Align decimal places.
- Round numbers as much as possible – re precision of measurement.
- Decide on a reasonable amount of data to be presented.
- Do not use tables if you only have two or fewer columns and rows. In such cases, a textual description is enough.
- If you have identical columns or rows of data in two or more tables, combine the tables.
- Be ***consistent*** with your tabular presentations. Use consistent table, title, and heading formats.

http://www.ncsu.edu/labwrite/res/gh/gh-tables.html#practice
http://www2.le.ac.uk/offices/ld/resources/numerical-data/numerical-data
https://www.inf.ethz.ch/personal/markusp/teaching/guides/guide-tables.pdf

Guidelines for Figures

- Figures are an efficient display of meaningful and unambiguous data
- Limit your figure to data that are relevant to the hypotheses in the study.
- Be certain that your figure can stand alone without any explanation.
- Keep it simple.
- Figures should be no more complex than the data they describe.
- Figures should not provide a distorted picture of the values they present. (Edward Tufte and the "Lie Factor" - http://www.infovis-wiki.net/index.php?title=Lie_Factor)
- Scaling should be easy to interpret.
- Include units on the axes.
- Do not use figures if you only have a small number of data points
- Be consistent with your graphical presentation, as appropriate. Use consistent title, axes and legend formats.

https://www.stat.auckland.ac.nz/~ihaka/120/Lectures/lecture03.pdf
https://www.biostat.wisc.edu/~kbroman/topten_worstgraphs/
http://www.statisticshowto.com/misleading-graphs/

Design Exercise

- In your groups, discuss how to redo the following statistical results slides to better communicate the ideas.

- Consider the following questions during your discussion:
  - ✓ Is the problem effectively communicated?
  - ✓ Are results stated so that lay people can understand them?
  - ✓ How could graphs and tables improve communications?

- Select a spokesperson from your group to describe the results of your makeover to the class, focus on:
  - What's working and not working?
  - How will you do it differently?

15 minutes

Exercise 1

- Make over the “before” table on the next page

Table Example – Before

Table 1. Summary statistics for depression score by age group			
Age: 8-12 years	n Mean SD	Males: 25 Males: 45.78136 Males: 3.30625	Females: 12 Females: 38.340 Females: 6.3715
Age: 13 – 19 years	n Mean SD	Males: 15 Males: 27.6473 Males: 8.261	Females: 52 Females: 20.04 Females: 7.1464
Age: 20 – 29 years	n Mean SD	Males: 33 Males: 11.07 Males: 4.0341	Females: 91 Females: 8.2369 Females: 5.165

Exercise 2

- Make over the “before” slide on the next page

- Note the following:
  - % change is the % increase in VTG to be detected.
  - DN: Dunn’s non-parametric comparisons to control
  - DT: Dunnett’s test
  - JT: step-down Jonckheere-terpstra test
  - MW: Mann-Whitney with Bonferroni-Holm adjustment
  - WL: Williams test

Power Simulation Results

TEST	25	50	100	110	120	130	140	150	160	170	180	190	200	250	300	350	400
DN	6	14	33	40	44	47	50	53	58	59	62	64	67	77	83	86	90
DT	8	18	46	52	58	62	65	69	74	76	80	82	85	93	97	98	99
JT	16	29	58	65	69	72	74	77	80	82	84	86	88	93	95	96	98
MW	6	12	28	33	37	41	46	47	52	54	59	61	64	76	85	88	93
WL	15	27	59	66	72	75	77	80	84	86	89	91	92	97	99	100	100

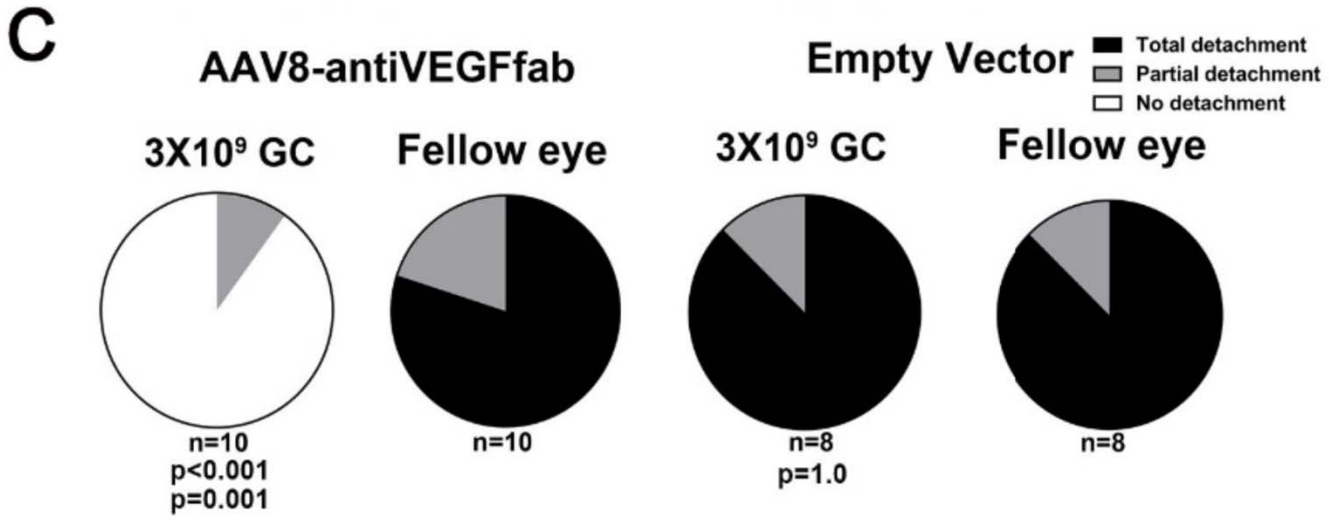
We want to detect a minimum of 10% VTG increase. Each of the statistical test is listed in the first columns and the sample sizes in the first row. The numbers in the table represent the achieved % power. The Williams' test is the most powerful in this case.

Exercise 3

- Make over the “before” slide on the next page

- Experiment description:
  Adult *Tet/opsin/VEGF* double transgenic mice had subretinal injection of $3x10^9$ GC AAV8-antiVEGFfab in one eye and no injection in the fellow eye or $3x10^9$ GC of null vector in one eye and no injection in the fellow eye. One month after injection, photos were graded for presence of total, partial, or no retinal detachment.

Subretinal injection of AAV8-AntiVEGFfab prevents retinal detachment



Nine of 10 eyes injected with AAV8-antiVEGFfab had no retinal detachment, while 8 of 10 fellow eyes had total retinal detachment ($p<0.001$ by Fisher's test). In contrast, 7 of 8 eyes injected with empty vector and fellow eyes had total retinal detachment ($p=1.0$). Compared to eyes injected with empty vector, there was significant prevention of retinal detachment in eyes injected with AAV8-antiVEGFfab ($p=0.001$ by Fisher's test).

Exercise 4

- Make over the “before” regression table on the next page
- Background information about the table:

Objective: To identify pre-residency characteristics associated with post-residency academic productivity in ophthalmology.

Design: A retrospective cohort study.

Participants: Ophthalmology residents who graduated from the Johns Hopkins Wilmer Eye Institute residency training program, Baltimore, Maryland, between 1990 and 1999.

Methods: Graduates were asked to complete an electronic survey and submit their most updated curricula vitae. Information pertaining to pre-residency scholastic performance and research experience, as well as post-residency professional activities and publication records was obtained from these sources.

Main Outcome Measures: Faculty status as of 2009. The statuses have three categories: 1) Currently on full-time faculty; 2) Never on full-time faculty; and 3) Once was on full-time faculty .

Table: Regression results for current faculty statuses

Faculty Category	Characteristics	Relative risk ratio	95% Confidence Interval	P
Full-time faculty	Years from end of residency to present	1.35	0.84, 2.19	0.2
	Age at start of residency	0.99	0.58, 1.71	NS
	Male vs. female	1.31	0.09, 19.27	NS
	With a graduate degree vs. without any graduate degree	1.26	0.04, 36.49	NS
	AOA vs. not AOA	2.24	0.16, 31.77	NS
	USMLE1 or FLEX percentile	1.05	1.01, 1.09	0.01
	Total number of first- and second-author publications	1.16	0.8, 1.6	0.38
	Pre-residency research was in ophthalmology vs. not	0.80	0.03, 20.91	NS
	Influential person is ophthalmologist vs. not	0.04	0.00, 0.96	0.05
Once was a full-time faculty	Years from end of residency to present	2.57	1.35, 4.87	0.00
	Age at start of residency	0.71	0.39, 1.32	0.3
	Male vs. female	26.94	0.90, 806.66	0.06
	With a graduate degree vs. without any graduate degree	1.84	0.06, 57.22	NS
	AOA vs. not AOA	2.75	0.20, 38.25	NS
	USMLE1 or FLEX percentile	1.04	1.00, 1.07	0.02
	Total number of first- and second-author publications	0.90	0.6, 1.3	0.59
	Pre-residency research was in ophthalmology vs. not	0.09	0.00, 2.56	0.2
	Influential person is ophthalmologist vs. not	0.07	0.00, 1.78	0.1

Exercise 1 - Table Example

Before:

Table 1. Summary statistics for depression score by age group			
Age: 8-12 years	n Mean SD	Males: 25 Males: 45.78136 Males: 3.30625	Females: 12 Females: 38.340 Females: 6.3715
Age: 13 – 19 years	n Mean SD	Males: 15 Males: 27.6473 Males: 8.261	Females: 52 Females: 20.04 Females: 7.1464
Age: 20 – 29 years	n Mean SD	Males: 33 Males: 11.07 Males: 4.0341	Females: 91 Females: 8.2369 Females: 5.165

After:

Table 1. Summary statistics for depression score by age group

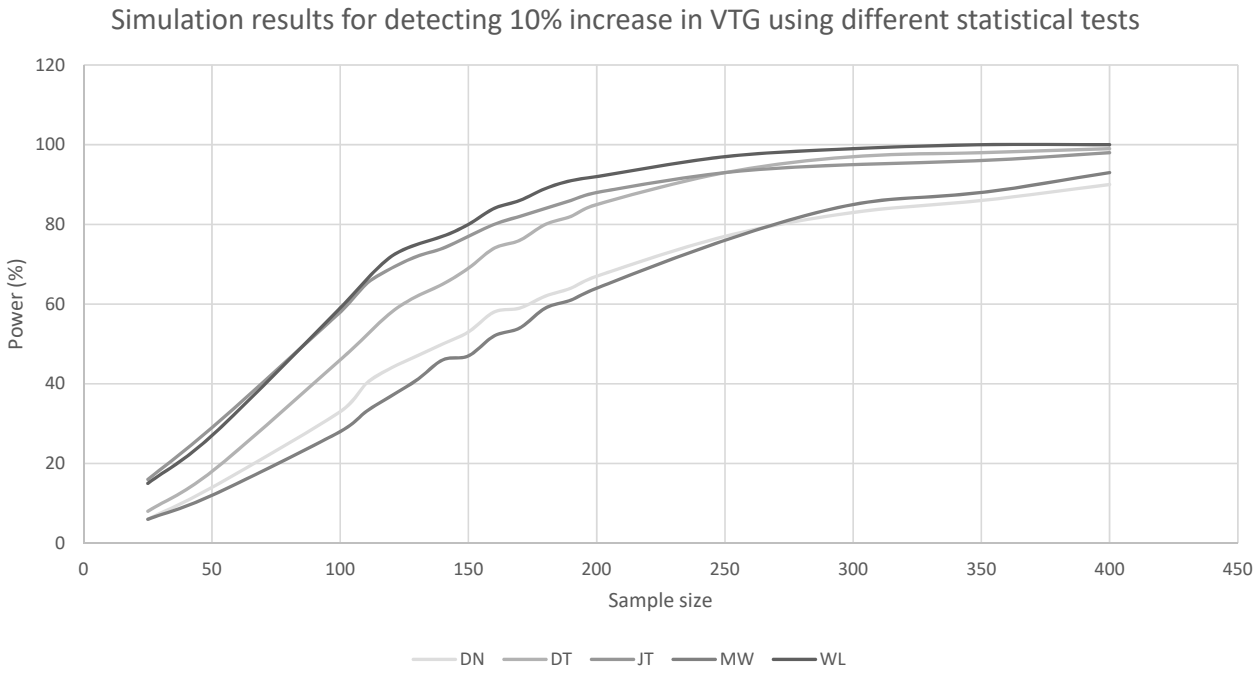
Age Range (years)	Males		Females	
	n	Mean (SD)	n	Mean (SD)
8 - 12	25	45.78 (3.31)	12	38.34 (6.37)
13 – 19	15	27.65 (8.26)	52	20.04 (7.15)
20 – 29	33	11.07 (4.03)	91	8.24 (5.17)

Exercise 2 - Before

TEST	25	50	100	110	120	130	140	150	160	170	180	190	200	250	300	350	400
DN	6	14	33	40	44	47	50	53	58	59	62	64	67	77	83	86	90
DT	8	18	46	52	58	62	65	69	74	76	80	82	85	93	97	98	99
JT	16	29	58	65	69	72	74	77	80	82	84	86	88	93	95	96	98
MW	6	12	28	33	37	41	46	47	52	54	59	61	64	76	85	88	93
WL	15	27	59	66	72	75	77	80	84	86	89	91	92	97	99	100	100

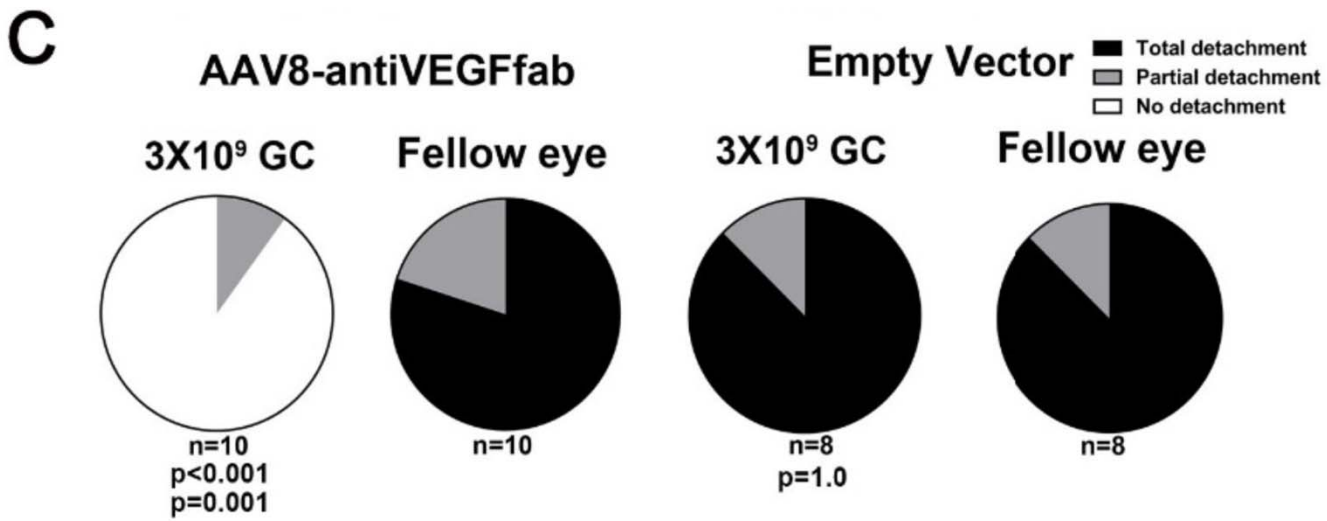
We want to detect a minimum of 10% VTG increase. Each of the statistical test is listed in the first columns and the sample sizes in the first row. The numbers in the table represent the achieved % power. The Williams' test is the most powerful in this case.

Exercise 2 - After



DN: Dunn's non-parametric test
DT: Dunnett's test
JT: step-down Jonckheere-terpstra test
MW: Mann-Whitney test with Bonferroni-Holm adjustment
WL: Williams' test

Exercise 3 - Before



Nine of 10 eyes injected with AAV8-antiVEGFfab had no retinal detachment, while 8 of 10 fellow eyes had total retinal detachment ($p<0.001$ by Fisher's test). In contrast, 7 of 8 eyes injected with empty vector and fellow eyes had total retinal detachment ($p=1.0$). Compared to eyes injected with empty vector, there was significant prevention of retinal detachment in eyes injected with AAV8-antiVEGFfab ($p=0.001$ by Fisher's test).

Exercise 3 - After

Table: Subretinal injection of AAV8-AntiVEGFfab prevents retinal detachment

Experiment	Injection	Number of eyes	No detachment	Partial detachment	Total detachment	P value*	
AAV8-antiVEGFfab	Yes	10	9	1	0	(reference)	
	No	10	0	2	8	<.001	
Empty vector	Yes	8	0	1	7	-	(reference)
	No	8	0	1	7	0.001	1.0

* P values are calculated using fisher's exact tests comparing to the eyes with the reference group.

Exercise 4

- Make over the “before” regression table on the next page
- Background information about the table:

Objective: To identify pre-residency characteristics associated with post-residency academic productivity in ophthalmology.

Design: A retrospective cohort study.

Participants: Ophthalmology residents who graduated from the Johns Hopkins Wilmer Eye Institute residency training program, Baltimore, Maryland, between 1990 and 1999.

Methods: Graduates were asked to complete an electronic survey and submit their most updated curricula vitae. Information pertaining to pre-residency scholastic performance and research experience, as well as post-residency professional activities and publication records was obtained from these sources.

Main Outcome Measures: Faculty status as of 2009. The statuses have three categories: 1) Currently on full-time faculty; 2) Never on full-time faculty; and 3) Once was on full-time faculty .

Table: Regression results for current faculty statuses

Faculty Category	Characteristics	Relative risk ratio	95% Confidence Interval	P
Full-time faculty	Years from end of residency to present	1.35	0.84, 2.19	0.2
	Age at start of residency	0.99	0.58, 1.71	NS
	Male vs. female	1.31	0.09, 19.27	NS
	With a graduate degree vs. without any graduate degree	1.26	0.04, 36.49	NS
	AOA vs. not AOA	2.24	0.16, 31.77	NS
	USMLE1 or FLEX percentile	1.05	1.01, 1.09	0.01
	Total number of first- and second-author publications	1.16	0.8, 1.6	0.38
	Pre-residency research was in ophthalmology vs. not	0.80	0.03, 20.91	NS
	Influential person is ophthalmologist vs. not	0.04	0.00, 0.96	0.05
Once was a full-time faculty	Years from end of residency to present	2.57	1.35, 4.87	0.00
	Age at start of residency	0.71	0.39, 1.32	0.3
	Male vs. female	26.94	0.90, 806.66	0.06
	With a graduate degree vs. without any graduate degree	1.84	0.06, 57.22	NS
	AOA vs. not AOA	2.75	0.20, 38.25	NS
	USMLE1 or FLEX percentile	1.04	1.00, 1.07	0.02
	Total number of first- and second-author publications	0.90	0.6, 1.3	0.59
	Pre-residency research was in ophthalmology vs. not	0.09	0.00, 2.56	0.2
	Influential person is ophthalmologist vs. not	0.07	0.00, 1.78	0.1

Exercise 4 - After

Table: Unadjusted and adjusted relative risk ratios for current faculty status by the characteristics of the surveyed residents

Faculty Category	Characteristics	Unadjusted[1]			Adjusted[2]		
		Relative risk ratio	95% Confidence Interval	P value	Relative risk ratio	95% Confidence Interval	P value
Currently on full-time faculty vs. Never on full-time faculty	Years from the end of residency to the time of survey	1.06	0.8, 1.4	0.67	0.96	0.7, 1.3	0.75
	Age at the start of residency	0.86	0.7, 1.1	0.19	0.85	0.7, 1.1	0.19
	Male vs. female	0.69	0.1, 3.3	0.64	-	-	-
	With a graduate degree vs. without any graduate degree	0.75	0.2, 3.0	0.68	-	-	-
	AOA vs. not AOA	2.25	0.5, 10.8	0.31	2.7	0.5, 13.9	0.23
	USMLE1 or FLEX above 90 percentile vs. below 90	0.53	0.2, 1.3	0.15	-	-	-
	Total number of first- and second-author publications	1.16	0.8, 1.6	0.38	-	-	-
	Pre-residency research was in ophthalmology vs. not	1.77	0.4, 7.6	0.44	-	-	-
	Had Influential person in ophthalmology or academics	0.47	0.1, 2.2	0.33	-	-	-
Once was on full-time faculty vs. Never on full-time faculty	Years from the end of residency to the time of survey	1.46	1.1, 2.0	0.01	1.45	1.0, 2.0	0.03
	Age at the start of residency	0.81	0.6, 1.0	0.10	0.82	0.6, 1.1	0.17
	Male vs. female	1.3	0.2, 7.9	0.78	-	-	-
	With a Masters degree, or PhD, or both vs. without	0.53	0.1, 2.4	0.41	-	-	-
	AOA vs. not AOA	5.0	1.0, 26.1	0.06	4.8	0.8, 30.0	0.09
	USMLE1 or FLEX above 90 percentile vs. below 90	1.05	0.4, 2.6	0.91	-	-	-
	Total number of first- and second-author publications	0.90	0.6, 1.3	0.59	-	-	-
	Pre-residency research was in ophthalmology vs. not	1.38	0.3, 6.4	0.67	-	-	-
	Had Influential person in ophthalmology or academics	0.66	0.1, 3.5	0.63	-	-	-

1. The unadjusted estimates are from simple multinomial logistic regressions that only have one predictor.
2. The adjusted estimates are from multivariate multinomial logistic regression model including years from the end of residency to the time of survey, age at the start of residence and AOA statuses as the predictors.

How to Avoid Upsetting a Stat Referee

- No significant difference does not → no difference
- Include confidence intervals
- Give exact p-values where possible
- Differentiate SD from SE when used, avoid ± notation (vs CI)
- Avoid bar charts with error bars – why?
- Check assumptions for statistical methods
- Give clear descriptions of statistical methods

-- Altman: http://www-users.york.ac.uk/~mb55/talks/upset.htm

Tables vs. Figures: When to Use

Use Tables When:

- The document you produce will be used to look up individual values.
- It will be used to compare individual values.
- Precise values are required.
- The quantitative information to be communicated involves more than one unit of measure.

Use Figures When:

- The message is contained in the shape of the values.
- The document will be used to reveal relationships among multiple values.

Stephen Few. Show Me the Numbers: Designing Tables and Graphs to Enlighten. 2004. Analytics Press.

Table with Annotations

Table 1. Associations between respondents' demographic characteristics and self-reported willingness to respond (WTR) to a pandemic flu emergency

			WTR if required			WTR if asked, but not required		
		% [a]	% Agree [b]	OR [c]	(95%CI)	% Agree	OR	(95%CI)
All [d]			82.5			72.0		
By respondent characteristics								
Gender	Female	72.7	81.6		Reference	69.9		Reference
	Male	27.3	84.9	1.27	(1.00 - 1.61)	77.1	1.45	(1.18 - 1.78)
Age (years)	<30	16.5	80.6		Reference	66.5		Reference
	30-39	21.8	79.8	0.95	(0.68 - 1.32)	65.8	0.97	(0.73 - 1.28)
	40-49	25.7	82.2	1.11	(0.80 - 1.55)	72.7	1.34	(1.02 - 1.77)
	50-59	27.0	85.1	1.38	(0.99 - 1.92)	76.3	1.63	(1.23 - 2.14)
	60+	9.0	84.3	1.29	(0.84 - 1.99)	79.0	1.90	(1.30 - 2.76)
Duration at JHH [e] (years)	<1	11.0	81.3		Reference	69.3		Reference
	1-5	32.5	82.8	1.10	(0.77 - 1.58)	72.3	1.16	(0.85 - 1.57)
	6-10	17.3	80.1	0.93	(0.63 - 1.37)	70.0	1.04	(0.74 - 1.45)
	>10	39.2	83.5	1.16	(0.81 - 1.65)	73.1	1.21	(0.90 - 1.63)

[a] Percent of respondents in category within characteristic
[b] Percent agreeing with WTR statement (positive response)
[c] OR is the odds ratio provided in the logistic regression which compares the odds between a positive WTR response and a negative WTR response with respect to a particular characteristic category compared to its reference category,,nadjusted for other demographic characteristics.
[d] Percent covers all respondents.
[e] Johns Hopkins Hospital (JHH)

Figure with Annotations

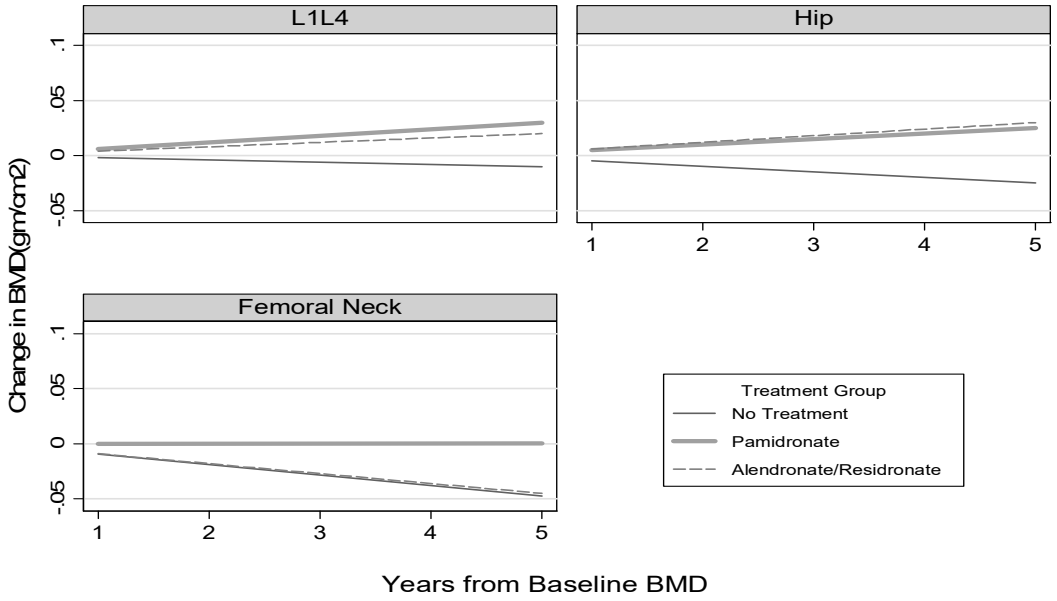


Figure 1. Projected levels of BMD change for Type I OI patients by treatment group and site starting at a baseline measurement across a span of 5 years. The projected BMD changes are based on the following [annualized linear rates of change (95% CI) adjusted for age at baseline, DXA scan and gender]: ***for L1L4*** *a) No treatment* [-0.002 (-0.009, 0.006)], *b) Pamidronate* [0.006 (0.008, 0.012)], *c) Alendronate/Residronate* [0.004 (0.0006, 0.008)]; ***for Total Hip*** *d) No Treatment* [-0.005 (-0.012, 0.003)], *e) Pamidronate* [0.005 (-0.003, 0.013)], *f) Alendronate/Residronate* [0.006 (0.002, 0.011)]; ***for Femur Neck*** *g) No Treatment* [-0.0095 (-0.018, -0.001)], *h) Pamidronate* [0.001 (-0.008, 0.008)], and *i) Alendronate/Residronatel* [-0.009 (-0.007, 0.005)].