Ensuring Integrity in Empirical Research: Design, Analysis and Reporting

Andrew R. Willan, PhD

Senior Scientist
Program in Child Health Evaluative Sciences
SickKids Research Institute, Toronto

Professor of Biostatistics
Dalla Lana School of Public Health
University of Toronto

www.andywillan.com/talks

andy@andywillan.com

Sample Size

Best advice: seek professional statistical collaboration.

Avoid using a sample size just because it had positive results in a "similar" study.

Need to consider elements relevant to the current study, such as variances and differences that are import to detect.

If sample size is too small, the study will have low probability of achieving statistical significance if when important differences exist.

Avoid uncontrolled adaptive designs,
e.g. continue sampling until statistical significance is achieved.
Recipe for runaway Type I error.

Statistical Analysis and Inference

Best advice: seek professional statistical collaboration.

Articulate objectives and null and alternative (working) hypotheses.

Explore and understand data structures, particularly lack of independence. Are there observation more closely related to each other than to others (e.g. siblings within a family)?

A Common error is to analyze paired data as independent samples. (Inflates probability of a Type II error)

Control	Treatment
142 134 149 138 145 139 142 139 139 159 140 146 149 149 148	150 143 151 146 152 141 143 148 147 158 145 143 170 152 147
n = 15; mean = 144.3; SD = 6.353	n = 15; mean = 149.0; SD = 7.170
Difference of means = 4.69; SE = 2.47; t = 1.90; p-value = 0.0685.	

Subject	Control	Treatment	Difference
1	142	150	7.71
2	134	143	8.82
3	149	151	1.81
4	138	146	7.25
5	145	152	6.32
6	139	141	2.08
7	142	143	1.38
8	139	148	8.43
9	139	147	8.48
10	159	158	-2.14
11	140	145	3.88
12	146	143	-3.82
13	149	170	19.7
14	149	152	1.86
15	148	147	-1.48

n = 15; mean of differences = 4.69; SE = 1.51; t = 3.09; p-value = 0.0045.

Control	Treatment
38.4	40.2
38.0	39.6
41.6	41.4
37.8	40.3
37.2	41.5
44.3	44.8
42.9	57.7
46.7	56.6
45.6	55.6
45.6	58.8
40.3	56.7
.	.
.	.
.	.
.	.
.	.
.	.
52.2	53.6
55.8	46.2
n = 25; mean = 45.2; SD = 6.69	n = 25; mean = 50.3; SD = 6.48
Difference of means = 5.13; SE = 1.86; t = 2.76; p-value = 0.0083.	

Control	Treatment
38.4 (Litter 1)	40.2 (Litter 6)
38.0 (Litter 1)	39.6 (Litter 6)
41.6 (Litter 1)	41.4 (Litter 6)
37.8 (Litter 1)	40.3 (Litter 6)
37.2 (Litter 1)	41.5 (Litter 6)
44.3 (Litter 2)	44.8 (Litter 7)
42.9 (Litter 2)	57.7 (Litter 7)
46.7 (Litter 2)	56.6 (Litter 7)
45.6 (Litter 2)	55.6 (Litter 7)
45.6 (Litter 2)	58.8 (Litter 7)
40.3 (Litter 3)	56.7 (Litter 8)
. (Litter 3)	. (Litter 8)
. (Litter 3)	. (Litter 8)
.	.
.	.
. (Litter 5)	. (Litter 10)
. (Litter 5)	. (Litter 10)
. (Litter 5)	. (Litter 10)
52.2 (Litter 5)	53.6 (Litter 10)
55.8 (Litter 5)	46.2 (Litter 10)
Adjusted difference of means = 5.13; SE = 3.51; t = 1.46; p-value = 0.141.	

	Pain Scores; mean (SD)		Difference Between Means (SE)	P-value
	Placebo (n = 27)	Active (n= 25)		
Pre Treatment	53.9 (14)	60.4 (12.3)	6.5	
Post Treatment	62.3 (17.9)	79.6 (17.1)	17.3 (5.00)	0.0008
Change Score	8.4 (14.6)	19.2 (16.1)	10.8 (4.36)	0.014
ANCOVA			12.7 (4.39)	0.005

Kleinhenz *et al*. *Pain* 1999; **83**: 235-241.

Vickers and Altman. *BMJ* 2001; **323**: 1123-1124.

Statistical Analysis and Inference

"All models are wrong, some are useful" George Box.

Seek useful, not perfect, models.

The assumption of normally distributed data is very robust against some skewing and "fat tails", but can lead to problems in the presence of "outliers".

Using statistical procedures to identify "outliers" is fraught with danger, and unless there is documented experimental failure, seeking statistical advice is recommended.

Statistical Analysis and Inference

Common errors:

Reversing the independent and dependent variables in a regression analysis

Having highly correlated independent variables in a regression
(inflates probability of Type II error)

Collapsing levels of an independent variable to maximize statistical significance
(inflates probability of Type I error)

Statistical Analysis and Inference

Common errors:

Treating missing data as "no" or zeros

Confusing different types of regression models:

- *Ordinary least squares* for continuous data
- *Logistic* for binary data
- *Proportional hazards* for time-to-event data

Statistical Analysis and Inference

Avoid blindly using statistical methods used by supervisor
("sins of the father/mother").

Avoid excessive number of statistical tests of hypotheses. Inflates probability of Type I error.

Don't over-analyze the data (subset analyses, vast number of outcomes); data are like people, if you torture them enough they'll tell you anything.

Reporting

Distinguish between standard errors (of the mean) and standard deviation (of observations).
SE's for comparing means
SD's for comparing distributions

Report precise *p*-values: *i.e.* $p = 0.034$, rather than $p < 0.05$.

Report confidence intervals where appropriate.

Reporting

"The truth, the whole truth, and nothing but the truth."

Report all results, not just those supporting working hypotheses.

Resist temptation to bury "inconvenient" results.

Masked analysis with dual report.