

Misinterpretation of p -values

Important misinterpretation 1
Non-significant p -values
are not evidence
for the null hypothesis

Non-significant p -values are not evidence:

- Of no/zero effect
- Of Equivalence
- For the null hypothesis

“Absence of evidence does not equal evidence of absence”

“Abwesenheit von Evidenz für einen Effekt ist nicht gleichbedeutend mit Evidenz für die Abwesenheit eines Effekts”



The Effect of Right-Turn-On-Red on
Pedestrian and Bicyclist Accidents

D. F. Preusser
W. A. Leaf
K. B. DeBartolo
R. D. Blomberg

Dunlap and Associates, Inc.
One Parkland Drive
Darien, Connecticut 06820

Contract No. DOT-HS-6-01411
Contract Amount \$146,727

This document is available to the U.S. public through the National Technical Information
Service, Springfield, Virginia 22161

Pedestrians in New York Upstate Urban,
Signalized Location with Vehicle Turning Right.

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Avg.
1974	6	4	5	3	3	4	0	3	2	7	6	3	3.833
1975	8	3	9	4	3	3	2	1	4	5	5	6	4.417
1976	2	3	2	2	6	1	2	2	3	6	4	7	3.333
1977	4	2	9	10	5	3	0	3	3	3	12	4	4.833
1978	4	5	5	7	8	3	3	4	4	11	9	10	6.083
Avg.	4.8	3.4	6.0	5.2	5.0	2.8	1.4	2.6	3.2	6.4	7.2	6.0	4.500

Analysis of Variance				Time Series Analysis				
Source	Mean Square	d.f.	F	Model	SE residual	Q	d.f.	p
Year	13.292	4	2.959	None	2.690	44.53	25	.011
Month	16.018	11	3.566	(1-B ¹²)	3.306	68.36	25	.000
Yr x Mon	4.492	44		Pre-RTOR	1.890	20.31	22	n.s.
				Intervention	2.398	23.30	22	.39
				Hypothesis	2.267	27.55	22	.20

Descriptive Models

Pre-RTOR: (1 - .485B¹²) (Y_t - 3.758) = (1 - .383B⁴)a_t
(36 months)

Intervention: (1 - .240*B¹²) (Y_t - 3.854) = 1.467X_t + (1 - .433B⁴)a_t

Hypothesis Model (Intervention)

$$Y_t = 1.567X_t + \frac{(1 - .212*B^4)(1 - .898B^{12})}{(1 - B^{12})}a_t$$

*parameter not significant



Right turn on red
29 extra deaths (+9.4%)

Pedestrians in New York Upstate Urban,
Signalized Location with Vehicle Turning Right.

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Avg.
1974	6	4	5	3	3	4	0	3	2	7	6	3	3.833
1975	8	3	9	4	3	3	2	1	4	5	5	6	4.417
1976	2	3	2	2	6	1	2	2	3	6	4	7	3.333
1977	4	2	9	10	5	3	0	3	3	3	12	4	4.833
1978	4	5	5	7	8	3	3	4	4	11	9	10	6.083
Avg.	4.8	3.4	6.0	5.2	5.0	2.8	1.4	2.6	3.2	6.4	7.2	6.0	4.500

Analysis of Variance				Time Series Analysis				
Source	Mean Square	d.f.	F	Model	SE residual	Q	d.f.	p
Year	13.292	4	2.959	None	2.690	44.53	25	.011
Month	16.018	11	3.566	(1-B ¹²)	3.306	68.36	25	.000
Yr x Mon	4.492	44		Pre-RTOR	1.890	20.31	22	n.s.
				Intervention	2.398	23.30	22	.39
				Hypothesis	2.267	27.55	22	.20

Descriptive Models

Pre-RTOR: $(1 - .485B^{12}) (Y_t - 3.758) = (1 - .383B^4)a_t$
(36 months)

Intervention: $(1 - .240*B^{12}) (Y_t - 3.854) = 1.467X_t + (1 - .433B^4)a_t$

Hypothesis Model (Intervention)

$$Y_t = 1.567X_t + \frac{(1 - .212*B^4) (1 - .898B^{12})}{(1 - B^{12})}a_t$$

*parameter not significant

70%

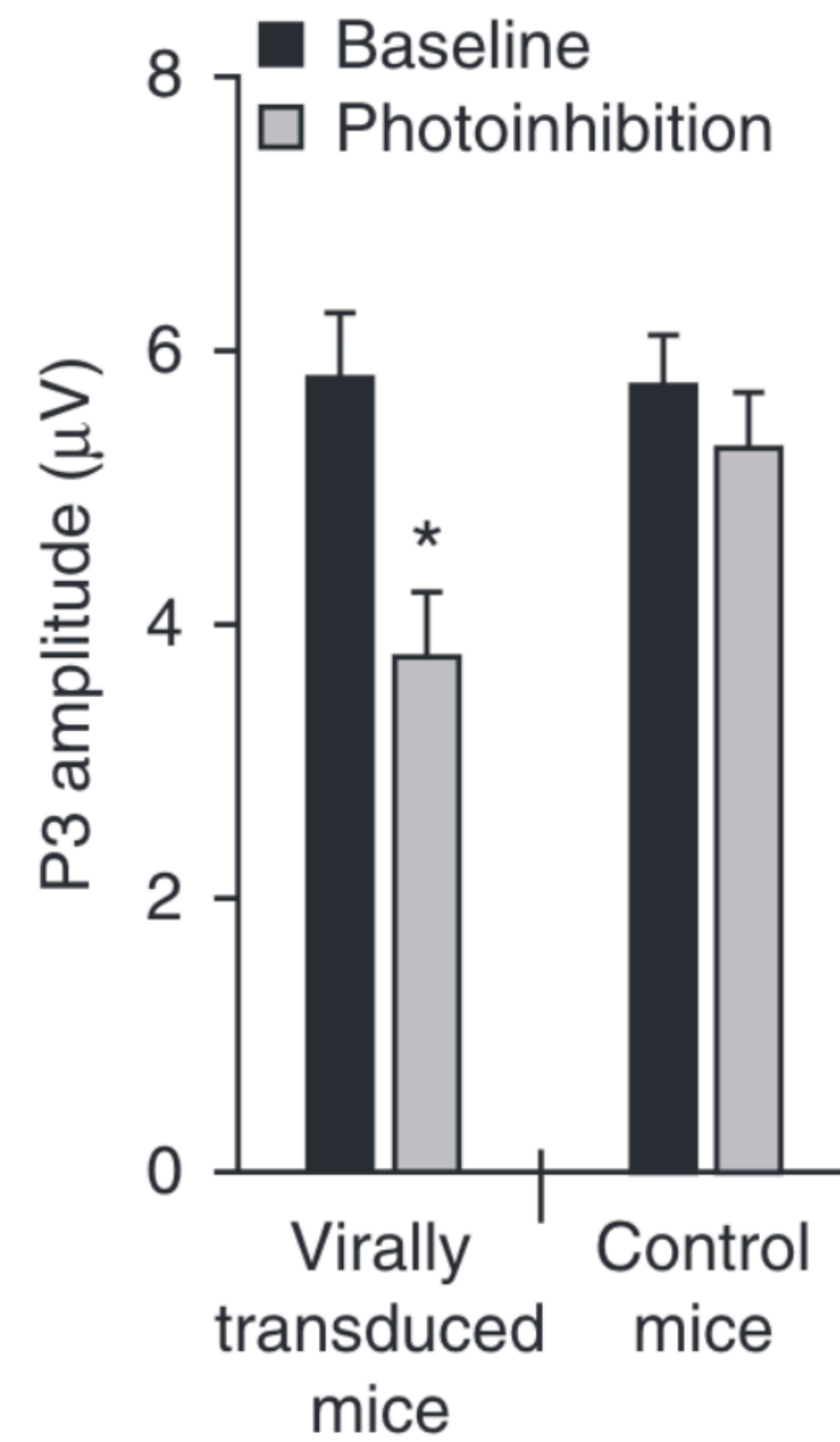
Of psychology articles with relevant comparisons
commit this error

(Aczel et al., 2018)

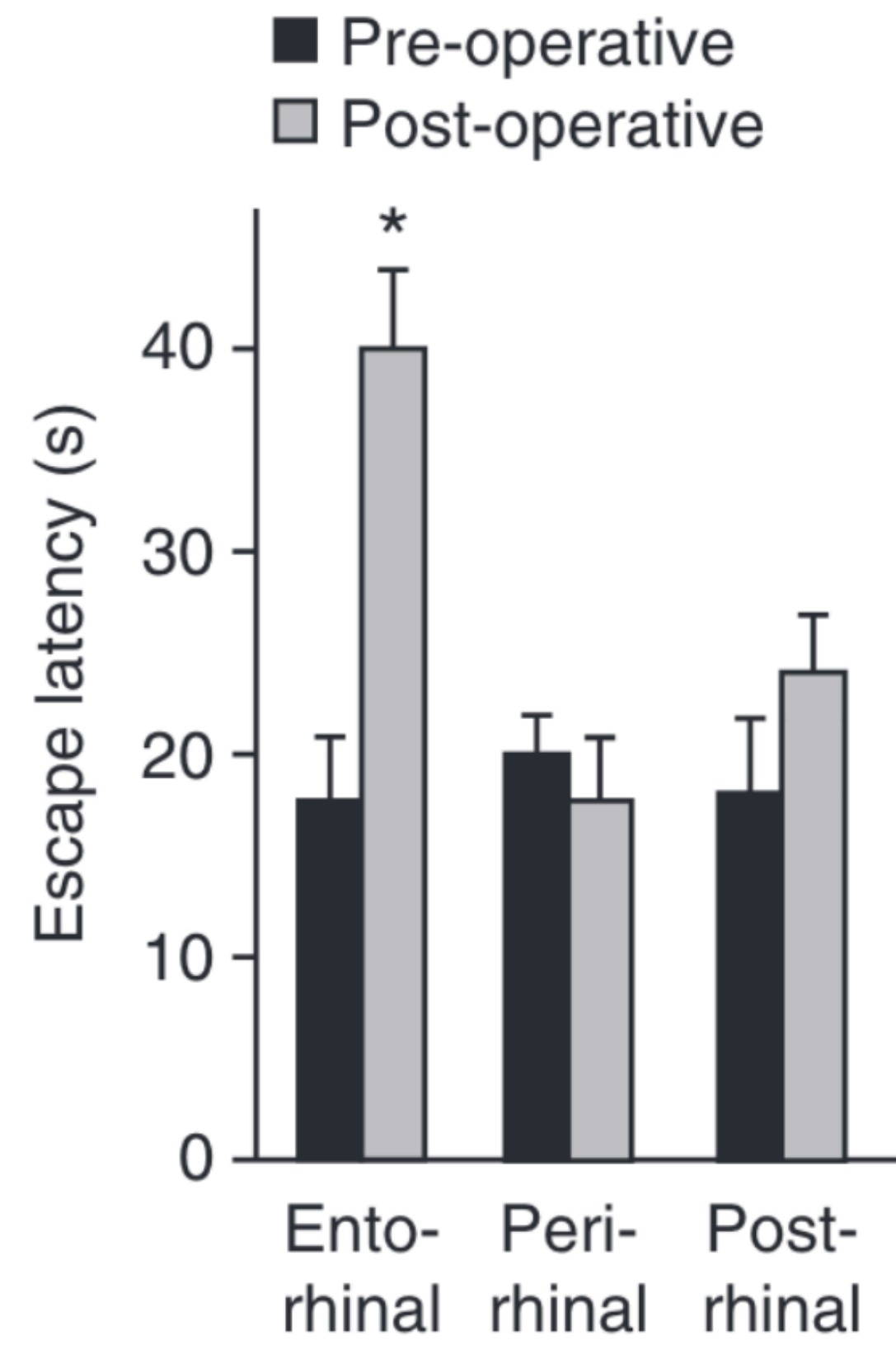
Important misinterpretation 2
p-values can't be compared
to infer differences

“The difference between significant and non-significant is not itself significant” German & Stern (2006)

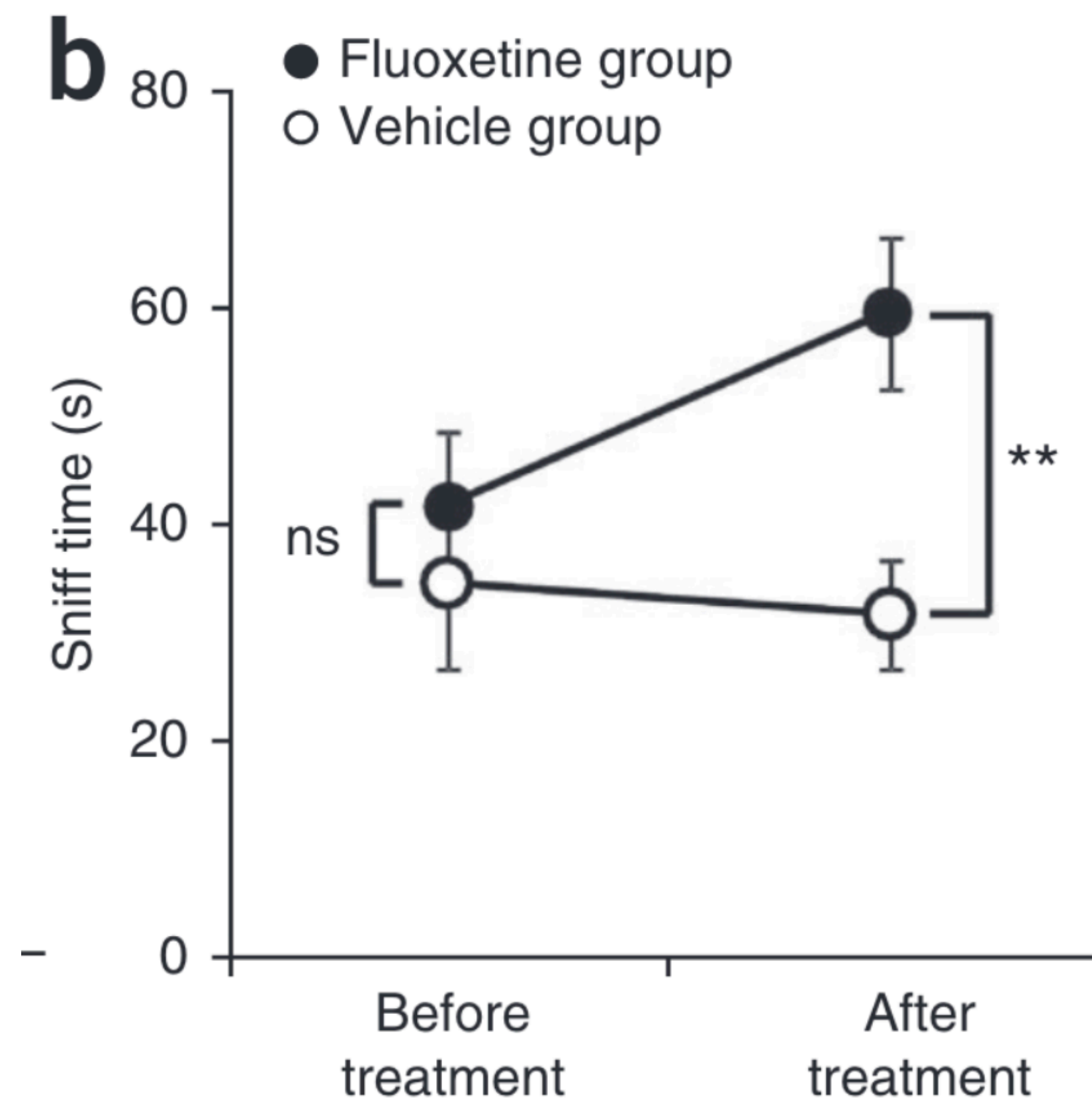
Der Unterschied zwischen „statistisch signifikant“ und „statistisch nicht signifikant“ ist selbst nicht statistisch signifikant.



Nieuwenhuis et al. (2011)



Nieuwenhuis et al. (2011)



Nieuwenhuis et al. (2011)

50%

Of neuroscience articles with relevant comparisons
commit this error

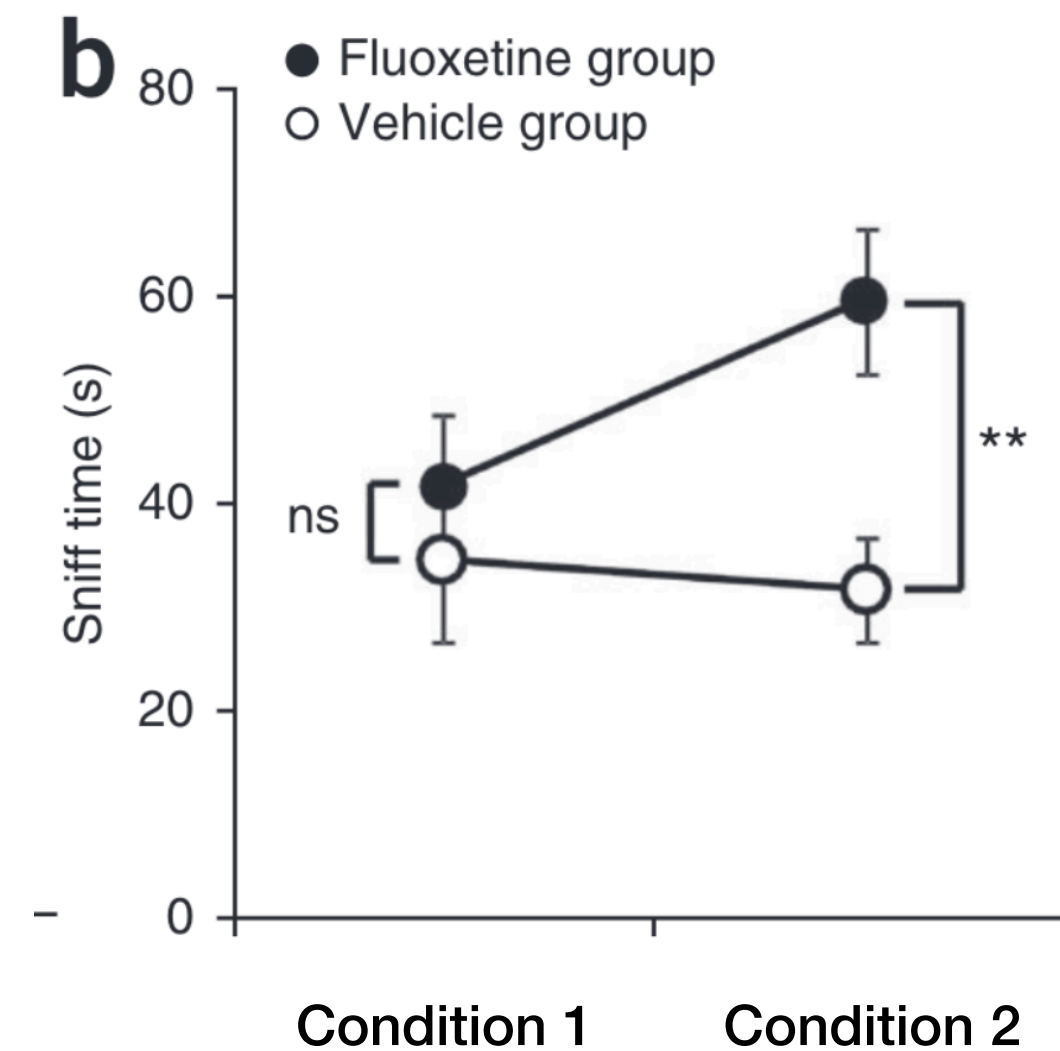
(Nieuwenhuis et al., 2011)

~25%

Of psychology articles with relevant comparisons
commit this error

If you're interested in differences, calculate a p-value for the difference

Usually, you're asking about moderation, so test for moderation. E.g.:



Wrong

t-test in condition 1

t-test in condition 2

Right

Within-between RM-ANOVA

ANCOVA controlling for baseline