University of Massachusetts Medical School

#### eScholarship@UMMS

Office of Institutional Research, Evaluation, and Assessment Publications and Presentations

Office of Educational Affairs

2004-11

## The Use of "Effect Size" in Augmenting the Results of Significance Testing: A Comparison of Pre/Post Data from a Geriatric Interclerkship

Urip Purwono University of Massachusetts Medical School

Et al.

### Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/res\_eval

Part of the Educational Assessment, Evaluation, and Research Commons, and the Medicine and Health Sciences Commons

#### **Repository Citation**

Purwono U, Barrett SV, Carlin MM, Zanetti ML, McGee SM. (2004). The Use of "Effect Size" in Augmenting the Results of Significance Testing: A Comparison of Pre/Post Data from a Geriatric Interclerkship. Office of Institutional Research, Evaluation, and Assessment Publications and Presentations. Retrieved from https://escholarship.umassmed.edu/res\_eval/14

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Office of Institutional Research, Evaluation, and Assessment Publications and Presentations by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.



# The Use of "Effect Size" in Augmenting the Results of Significance Testing: A Comparison of Pre/Post Data from a Geriatric Interclerkship

Urip Purwono, Susan V. Barrett, Michele M. Carlin, Mary L. Zanetti, Sarah McGee University of Massachusetts Medical School

### BACKGROUND

Most research in medical education, when examining the impact of an intervention, report findings based primarily on significance testing despite the controversy of its appropriate use. Moreover, the p-value used to determine rejection or acceptance of the null hypothesis tells nothing about the magnitude of the significance. Using a pre/post assessment of a Geriatric Interclerkship as a case study, this study examines the utility of "effect size" measures in augmenting significance testing results.

Effect size (ES) is a name given to a family of indices that measure the magnitude of a treatment effect. Unlike significance tests, these indices are independent of sample size. The formula used to calculate ES in this study follows:

$$d = \mathbf{M}_1 - \mathbf{M}_2 / \sigma_{\text{pooled}}$$
$$\sigma_{\text{pooled}} = \sqrt{\left[\left(\sigma_1^2 + \sigma_2^2\right) / 2\right]}$$

Where  $M_1$  and  $M_2$  are the means of the pre- and post- groups, respectively, and the pooled standard deviation is the square root of the average of the squared standard deviations (Cohen, 1988).

### **METHOD**

A pre/post evaluation consisting of knowledge and attitude items was constructed and administered to students participating in the Geriatric Interclerkship during academic year 2003-2004. While percent of student change on attitude and knowledge items from pre- to post- was measured, matched pre/post data was examined in terms of significance testing (paired t-test) and one measure of "effect size" (Cohen's d).

## **RESULTS**

Tables 1 and 2 highlight the percent of student pre/post change on attitude and knowledge items. Reliability of the instruments was .63 (knowledge test) and .82 (attitude scale). Ninety and ninety-two matched responses were collected to analyze clusters of attitude and knowledge items, respectively. Table 3 indicates the pre-post difference was significant for both the knowledge and attitude domains ( $t_{91} = 17.60$  and  $t_{89} = 5.82$ , respectively; p <. 01). Table 4 shows the effect size for the knowledge domain was very high (d = 1.84) indicating a substantial change in student pre to post knowledge, thus complementing the result of the significance testing. On the contrary, the effect size for the attitude domain was small (d = .34), indicating a minimal change in student attitude.

#### Table 1: Percent of Students Changing onAttitude Questions From PRE to POST

	Neutral/Postive to Negative	Positive to Neutral	Negative to Negative	Neutral/Negative to Neutral	Positive to Positive	Neutral/Negative to Positive
Item 1		3%		11%	73%	12%
Item 2r	2%	1%		18%	66%	13%
Item 3r	6%	4%	34%	28%	18%	10%
Item 4		6%	1%	8%	80%	6%
Item 5r	2%	6%	4%	18%	63%	7%
Item 6r	13%	4%	21%	18%	26%	18%
Item 7	6%	8%	7%	32%	33%	14%
Item 8r	8%	2%	28%	28%	22%	12%
Item 9	13%	4%	28%	30%	19%	6%
Item 10r	1%	6%		4%	78%	11%
Item 11r		2%	2%	11%	73%	11%
Item 12r		3%		11%	78%	8%
Item 13r		4%	3%	7%	71%	14%
Item 14	2%	1%	1%	6%	79%	11%

Negative = Disagree, Strongly Disagree; Positive = Agree, Strongly Agree. Items with reverse coding (2,3,5,6,8,10-13) reflect reverse scale.

 

 Table 2: Please rate your level of competence in the following content areas:

 
 Image: Section of the sectio

	Right-Wrong	Wrong-Wrong	Right-Right	Wrong-Right
Item 1	1%	5%	89%	4%
Item 2		9%	14%	77%
Item 3	4%	49%	5%	41%
Item 4	11%	25%	29%	35%
Item 5		48%	8%	45%
Item 6	1%	9%	32%	59%
Item 7	10%	18%	51%	21%
Item 8	3%	32%	25%	40%
Item 9		38%	15%	47%
Item 10	8%	25%	49%	18%
Item 11	5%	72%	5%	17%
Item 12	9%	14%	58%	20%
Item 13	10%	32%	18%	40%
Item 14	2%	11%	70%	17%
Item 15	2%	3%	86%	9%
Item 16			99%	1%
Item 17	1%	3%	79%	16%
Item 18	1%		89%	10%
Item 19			99%	1%
Item 20	10%	8%	60%	23%
Item 21	2%	2%	88%	8%
Item 22	1%	1%	92%	5%
Item 23			93%	7%
Item 24	10%	14%	58%	18%
Item 25	15%	40%	29%	15%

#### Table 3: Paired Samples Test

		Mean (SD)	N	Std. Error Mean	Paired Differences						
					Mean (SD)	Std. Error Mean	95% tl Diffe	CI of te rence	t	df	Sig. (2-tailed)
Pair 1	Post Knowledge	19.36 (2.72)	92	.284	4.88 (2.66)	.277	4.33	5.43	17.60	91	.000
	Pre Knowledge	14.48 (2.58)	92	.269							
Pair 2	Post Attitude	3.69 (.431)	90	.045	.16 (.25)	022	.10	.21	5.82	89	.000
	Pre Attitude	3.54 (.453)	90	.048		.027					

#### Table 4: Effect Size: Cohen's d

Domain	d	95% CI of	Effect Size	Cohen's Standard
Knowledge	1.84	1.49	2.18	Large
Attitude	.34	0.04	0.63	Small



# CONCLUSION

The "effect size" provides additional practical information to the significance testing. The routine use of this analysis is recommended to enhance the quality of research and evaluation in medical education. More specifically, the results from this Geriatric Interclerkship could eventually be compared to the change in student knowledge and attitude in another Interclerkship experience. Therefore, the most important benefit of reporting effect sizes is that this information provides the researcher with a more standard tool that allows for meta-analysis across studies.



#### **Evaluation Tools**

	You got a call from a patient's daughter. You have known this patient for many years. The daughter reports that her 83 year oil mether has failen >2.1 times over the past year, has difficulty going up and downstaire, and needs some help with huthing and dressing. She is continent and alert. The best course of action for you as a primary care provide is:	
Ē	Secure that the datablety coups is and talk to your fighter about these issues. Becommends a nursing hours of the gatemet, increasing in a state of the state of	rreks.
	Mr K is an 82 year old freefighter with a high school education living alone in edder housing. He comes to your office complications of memory problems and his langther reports he is having difficulty remembering in take his medications and has neglected to pay his hills recently. You perform a MMRI and his score is 2.4. The next spin your evaluation is:	
=	Have Mr & evaluated for nursing home placement, Perform a more pathemice vehiculation of his neuropy problem. Advise the daughter to pursue guardiandop.	
	Indicate True or False for each of the following statements on the psychosocial aspects of aging. True	144
=	The elderly commonly experience loss.	
Ξ	Personality style changes dramatically with aging	
=	Religion contributes importantly to the emotional well-being of the elderly	
	Maintaining involvement with others and activities leads to successful aging.	
=	Living alone increases an elderly person's risk of institutionalization.	
	Please continue with these True False items.	
=	Defirium occurs rarely in hospitalized elders.	
Ξ	Falls are typically multi-factorial in their ofology.	
Ξ	Functional assessment includes ADLs, IADLs, gait, cognitive status, mood, and psychosocial issues.	
=	An abnormal score on the Geriatric Depression Scale confirms the diagnosis of depression.	
=	Less medication is always better in managing elderly patients.	

-									
	ATTITUDE SCALE								
	Please use the scale to indicate the degree to which you agree or disagree with each statement. There are no right or wrong answers. The best response is the one that truly reflects your personal opioion.								
	Sirveg) Dispare	e Disarre	Neurod	Arrest	Stringly Autor				
=	Most old people are pleasant to be with,								
Ξ	The federal government should reallocate money from Medicare to research on AIDS or pediatric diseases.								
=	If I have the choice. I would rather see younger patients rather than elderly ones.								
=	It is society's responsibility to provide care for its elderly persons								
-	Medical care for old people uses up too much human and material resources.								
=	As people grow older, they become less organized and more confused.								
=	Elderly patients tend to be more appreciative of the medical care I provide than are younger patients.								
=	Taking a medical history from elderly patients is frequently an ordeal.								
-	I tend to pay more attention and have more sympathy towards my elderly patients than my younget patients.								
=	Old people in general do not contribute much to society								
=	Treatment of chronically ill old patients is hopeless.								
-	Old penions don't contribute their fair share towards paying for their health care.								
=	In general, old people act too slow for modern society								
-	It is interesting listening to old people's accounts of their past experiences.								

**Geriatric Experience** 

