

Research Statistics 101 for EBP

Salvador Bondoc, OTD, OTR/L, BCPR, CHT
For OT 604: Directed Study on Evidence-Based Practice

Objectives

- ◎ At the end of this presentation, the student should be able to:
 - Demonstrate familiarity with basic statistical terms used in research literature
 - Provide basic interpretation of the meaning of common statistical terms found in occupational therapy research literature

Basic Terms

- “n” or “N”
 - Refers to the size of the sample (n) or the population (N) in which the sample was drawn from
- “p” level
 - Refers to the probability value
 - Most *p-values* are in reference to the likelihood of that a result could have occurred by chance. The lower the *p-value*, the lower the probability that the result occurred by chance.
 - An acceptable *p-value* is .05 or less, which suggests that the probability that a result occurred by chance is 5% or less.
- Confidence interval (CI)
 - A confidence interval is an interval in which a measurement or trial falls within or corresponds with a given probability

Research Variables

Generally, variables are factors or attributes that are manipulated and measured in research. There are two types: independent and dependent variables.

◎ **Independent variable**

- the variable representing the value being manipulated by the researcher
- **Example:**
 - In a true experiment or randomized controlled trial (RCT) of fall prevention, the participant is randomly assigned to one of two independent variable conditions – either experimental (fall prevention program) or control condition (usual care)

Research Variables

Continued

◎ **Dependent Variable**

- the event studied and expected to change when the independent variable is manipulated
- The dependent variable reflects the output resulting from manipulation of the independent variable
- **Example:**
 - In the RCT on fall prevention, the fall prevention program (experimental versus control) is the independent variable, and the number of falls at follow-up is the dependent variable

Types of Variables (based on measurement)

Variables differ in how well they can be measured, and the type of measurement scale. Variables may be classified as either parametric or nonparametric.

◉ Parametric Variables

• Interval variables

- allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them.
- Examples: *age, weight, muscle power (in lbs), blood pressure, etc.*

• Ratio variables

- similar to interval variables; in addition to all the properties of interval variables, they feature an identifiable absolute zero point.
- most statistical data analysis procedures do not distinguish between the interval and ratio variables.

Types of Variables (based on measurement)

continued

◎ Non-parametric Variables:

- **Nominal variables**

- allow for only qualitative classification.
- Examples: *gender, race, color, city, etc.*

- **Ordinal variables**

- allow us to rank order the items we measure in terms of which has more or less of the quality represented by the variable but not exactly by how much.
- Examples: *socioeconomic status, level of agreement, developmental level, etc*

Two General Categories of Statistics

Data is represented and analyzed using statistical tools which may be broadly classified as follows:

◎ Descriptive Statistics

- **Used to provide a description of a given variable – in what is referred to as univariate analysis. Includes:**
 - averages, ranges and degree of variation

◎ Inferential Statistics

- **Used to draw conclusions about a population from a sample by**
 - Estimation : to determine the true value of the parameter from a sample
 - Hypothesis testing: to determine whether a difference exists in a parameter between groups of data.

Descriptive Statistics: Measures of Central Tendency

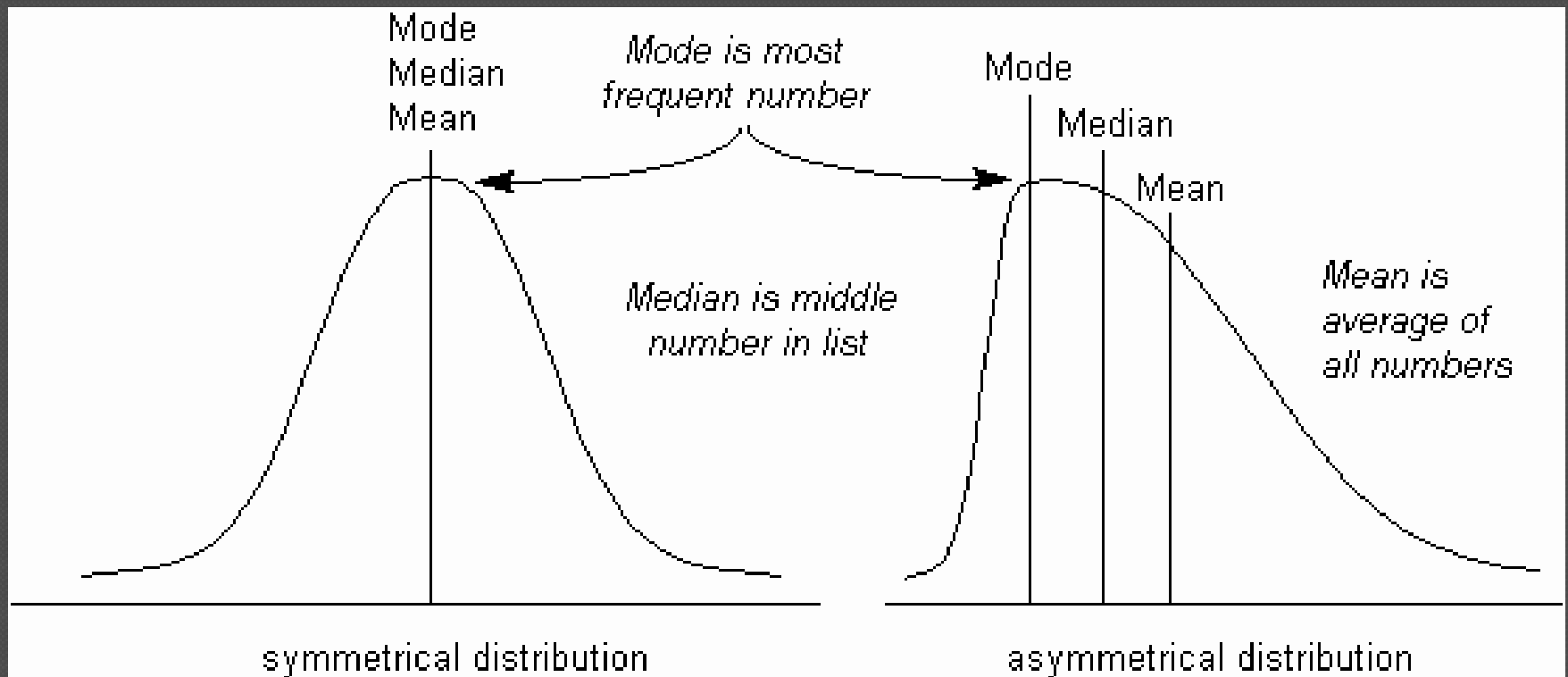
◎ Central tendency

- Refers to what is typical or average
- These measures are reported as descriptive statistics

◎ 3 measures of central tendency

- **Mean = mathematical average**
 - Example: the average of {1, 3, 5, 7, 9} is 5
 - Typically expressed as **M** or **X** in research papers
- **Median = mathematical half point**
 - Example: the median of {1, 2, 4, 7, 9} is 4
- **Mode = most frequently occurring value**
 - Example: the mode of {2, 3, 3, 5, 9} is 3

Mean – Median - Mode



Descriptive Statistics: Measures of Variability

◎ Measures of variability

- Reflects the degree of spread or dispersion that characterizes a group of values

◎ Terms related to measures of variability:

- **Standard deviation**

- Amount that a score varies from the mean or average of scores
- Typically expressed as *sd* or **SD** in the literature

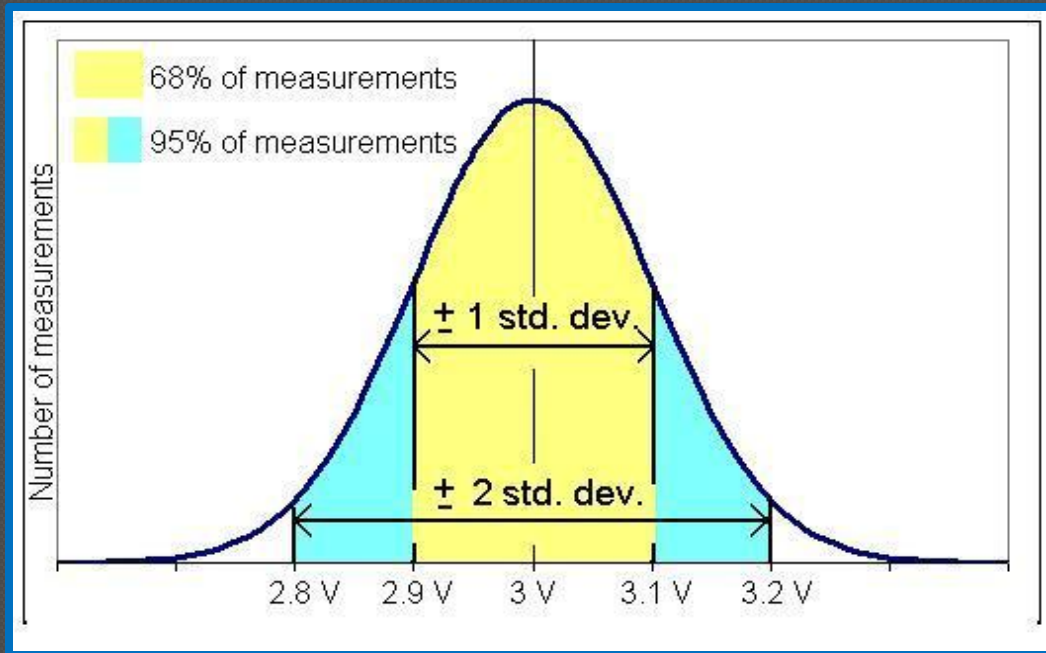
- **Variance**

- Represents the square of the standard deviation or s^2
- Measure of how far a set of numbers are spread out from each other

- **Range**

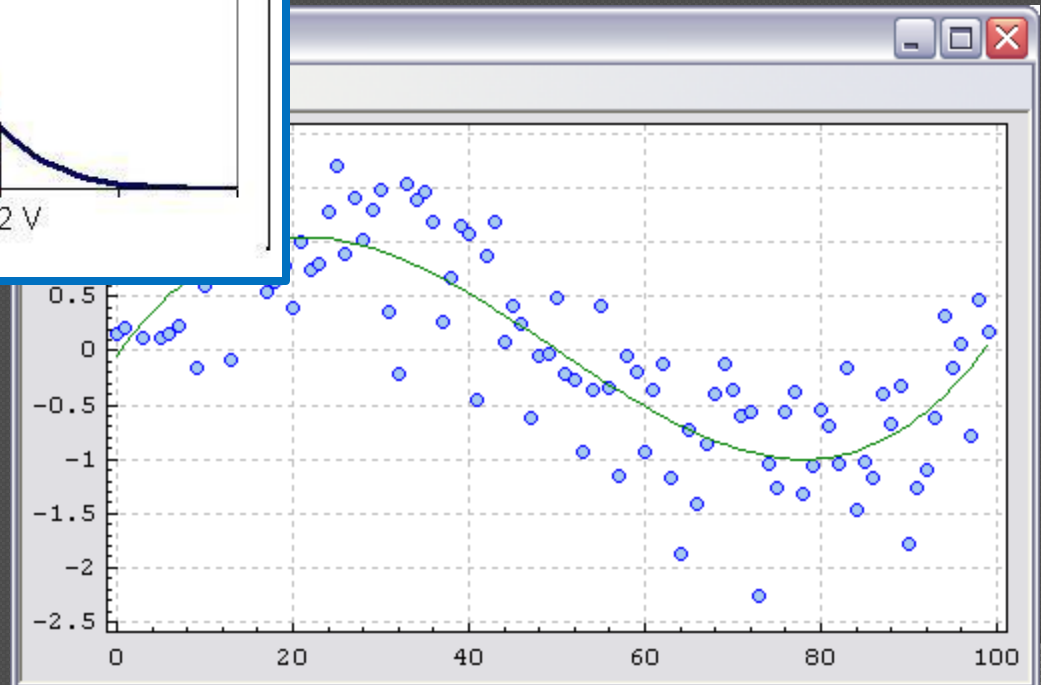
- Represents the difference between the highest and lowest score in a distribution

Standard Deviation and Variance



Standard Deviation

Variance



This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

**Enhancing Quality of Life in Older Adults: A Comparison of Muscular Strength
and Power Training**

Health and Quality of Life Outcomes 2008, **6**:45 doi:10.1186/1477-7525-6-45

Jeffrey A. Katula (katulaj@wfu.edu)
W. JACK Rejeski (rejeski@wfu.edu)
Anthony P. Marsh (marshap@wfu.edu)

Table 4 – Changes in Satisfaction with Life Across Treatment Conditions*

Treatment	Baseline: M ±SD	Post-Test: M ±SD	Δ Scores: Raw (Adjusted M ± SE)
Control	22.93 ± 6.46	21.46 ± 6.06	-1.47 (-2.17 ± 1.46) ¹
ST	22.28 ± 6.79	25.27 ± 6.03	2.99 (2.38 ± 1.61) ^{1, 2}
PT	25.78 ± 7.47	29.25 ± 6.38	3.47 (4.27 ± 1.62) ^{2, a}

ST = Strength Training, PT = Power Training

* = Baseline adjusted means in the last column not sharing a common superscript differ at the $p < .05$ level using a Bonferroni post hoc adjustment.

^a = significant within group change at $p < .05$ (paired t-test)

Common Inferential Statistics

◎ T-test

- Often used to compare the means of 2 groups of measurements including:
 - Independent samples
 - Comparison of outcomes of control vs. experimental groups in experimental designs
 - Paired samples
 - Comparison between pretest and posttest scores in a single group before-and-after design (a.k.a. paired samples t-test)
- Reported as a $t = \#.##, p < .##$ in the literature
 - Example: The difference in the mean Falls Efficacy Scores between control and experimental group was statistically significant ($t=3.86, p<.05$).
 - It is important to determine what the scores are to be able to interpret with confidence which group did statistically better.

Table 2. Success Rates and Improvement in Primary Outcomes After 1, 3, 6, 12, and 18 Months (Intention-to-Treat Analyses)*

Primary Outcome by Month	Surgery	Splinting	Difference†	P Value
Success rate, No./total (%)				
1	23/80 (29)	37/88 (42)	-13 (-28 to 1)	.07
3	62/78 (80)	46/86 (54)	26 (12 to 40)	<.001
6	72/77 (94)	57/84 (68)	26 (14 to 37)	<.001
12	67/73 (92)	60/83 (72)	20 (8 to 31)	.002
18	61/68 (90)	59/79 (75)	15 (3 to 27)	.02

JAMA®

Online article and related content
current as of June 1, 2010.

Splinting vs Surgery in the Treatment of Carpal Tunnel Syndrome: A Randomized Controlled Trial

Annette A. M. Gerritsen; Henrica C. W. de Vet; Rob J. P. M. Scholten; et al.

JAMA. 2002;288(10):1245-1251 (doi:10.1001/jama.288.10.1245)

<http://jama.ama-assn.org/cgi/content/full/288/10/1245>

mean (SD)				
1	1.6 (2.9)	2.1 (2.2)	-0.5 (-1.3 to 0.3)	.22
3	5.1 (3.3)	3.2 (2.7)	1.9 (1.0 to 2.8)	<.001
6	6.6 (2.4)	4.4 (3.2)	2.2 (1.4 to 3.1)	<.001
12	6.4 (2.7)	5.1 (3.1)	1.3 (0.4 to 2.2)	.005
18	6.2 (2.8)	5.0 (3.3)	1.2 (0.2 to 2.3)	.02
Paresthesia during the day, mean (SD)				
1	1.5 (3.0)	1.4 (2.1)	0.1 (-0.6 to 1.0)	.66
3	4.8 (3.2)	2.2 (3.2)	2.6 (1.6 to 3.6)	<.001
6	5.5 (2.9)	3.7 (3.2)	1.8 (0.8 to 2.8)	<.001
12	5.5 (2.9)	4.0 (3.4)	1.5 (0.5 to 2.5)	.004
18	5.3 (3.0)	4.0 (3.6)	1.3 (0.3 to 2.5)	.01
Paresthesia at night, mean (SD)				
1	1.3 (3.1)	2.5 (3.0)	-1.2 (-2.1 to -0.2)	.02
3	4.6 (3.8)	3.5 (3.3)	1.1 (0 to 2.2)	.046
6	5.4 (3.5)	4.1 (3.7)	1.3 (0.2 to 2.4)	.02
12	5.2 (3.6)	4.5 (3.4)	0.7 (-0.4 to 1.8)	.20
18	5.0 (3.6)	4.4 (3.6)	0.6 (-0.6 to 1.7)	.35

*Patient cohorts are presented in Figure 1. The values are expressed as the mean (SD) improvements from baseline unless otherwise indicated.

†Indicates the differences (surgery minus splint) in success rates and in the mean improvements from baseline (95% confidence interval).

Gutman, S. A., Kerner, R., Zombek, I., Dulek, J., & Ramsey, C. A. (2009). Supported education for adults with psychiatric disabilities: Effectiveness of an occupational therapy program. *American Journal of Occupational Therapy, 63*, 245–254.

A Mann–Whitney U test found no statistically significant differences between experimental and control group pretest scores on the Participant Comfort With the Student Role Scale, TSS, ISS, and SBS ($p > .05$). We found a statistically significant difference, however, between experimental and control group posttest scores on the Participant Comfort With the Student Role Scale, TSS, ISS, and SBS ($p \leq .000$). Paired t tests found statistically significant differences between pre- and posttest scores for all 12 modules taught in the Bridge Program (see Table 1).

Table 1. Statistically Significant Differences Between Pre- and Posttest Scores for Each of the 12 Modules

Module	p
1. Exploration of training programs, degrees, and work options	$\leq .000$
2. Study skills for school or work	$\leq .000$
3. Time management skills for school or work	$\leq .000$
4. Effective reading skills for school and job training	$\leq .000$
5. Basic writing skills for school or work	$\leq .000$
6. Basic computer skills	$< .01$
7. Introduction to Internet skills	$< .01$
8. Basic math skills for school and job placement tests	$\leq .000$
9. Use of library resources	$\leq .000$
10. Public speaking strategies for school or work	$< .05$
11. Professional behaviors and social skills	$< .05$
12. Stress management skills for school or work	$\leq .000$

Inferential Statistical Procedures

◎ *Mann-Whitney U test*

- is the nonparametric equivalent of **independent samples t-test** that can be used to analyze ordinal data from a two-group independent groups design when measurement is at least ordinal.
- It analyses the degree of separation (or the amount of overlap) between groups.
- It is represented by a “U” or “U”

◎ *Wilcoxon signed-ranks test*

- is another nonparametric test that is the nonparametric version of **paired samples t-test** when measurement is at least ordinal.
- It is represented by a “T”

Inferential Statistical Procedures

◎ Analysis of Variance (ANOVA)

- more complex version of a t-test because it allows comparison of 3 or more groups when there are at least 2 outcomes being analyzed, or in repeated measures comparison.
- One way ANOVA is used if there is only one outcome or independent variable being analyzed; Two-way ANOVA for more than 1 independent variable. However, when we are dealing with more than 1 dependent variable, the term changes to Multivariate ANOVA or MANOVA.
- Reported as: $F(df, df) = \#.##, p < .##$
 - Example: Differences in community participation and health related quality of life were found between With Treatment, Alternative Treatment, and No Treatment Groups. The differences were significant ($F(df, df) = 9.82, p < .01$)

Miller, L. J., Coll, J. R., & Schoen, S. A. (2007). A randomized controlled pilot study of the effectiveness of occupational therapy for children with sensory modulation disorder. *American Journal of Occupational Therapy, 61*, 228–238.

Table 1 displays demographic characteristics for participants. Using Fisher's exact test for categorical variables and a one-way analysis of variance (ANOVA) for age, no significant group differences were found on age, gender, mother's education, or ethnicity.

Table 1. Demographic Characteristics of Children Participating in Randomized Controlled Pilot Study

	OT <i>N</i> = 7		AP <i>N</i> = 10		NT <i>N</i> = 7		<i>p</i> value
	<i>N</i>	(%)	<i>N</i>	(%)	<i>N</i>	(%)	
Gender							0.85
Female	1	(14.3)	3	(30.0)	2	(28.6)	
Male	6	(85.7)	7	(70.0)	5	(71.4)	
Ethnicity							0.84
Caucasian	6	(85.7)	9	(90.0)	7	(100.0)	
Hispanic	1	(14.3)					
Other			1	(10.0)			
Mother's education							0.85
High school	1	(14.3)	3	(30.0)	2	(28.6)	
College	6	(85.7)	7	(70.0)	5	(71.4)	
Age mean (<i>SD</i>)	6.09	(1.53)	6.88	(1.35)	6.67	(2.31)	0.65

Note. OT = occupational therapy group; AP = activity protocol group; NT = no treatment group.

Inferential Statistical Procedures

◎ The Kruskal-Wallis H-test

- Is similar to the ANOVA but is used for nonparametric (ordinal) measures when 3 or more independent groups are involved
- Just as how the ANOVA extends the independent samples t-test, the H-test extends the Mann-Whitney U-test

◎ Friedman ANOVA test

- is also similar to the repeat measures ANOVA (i.e., comparing multiple points of ordinal data from the same sample)
- As such it may be compared to the Wilcoxon signed rank test but for a comparison of 3 or more variables

Inferential Statistical Procedures

◎ Analysis of Covariance (ANCOVA)

- In factorial designs, there is a potential for co-effect among variables. ANCOVA is used as a statistical procedure comparing two or more groups while controlling for the effect of covariates. When we are dealing with more than one dependent variable, the term used is Multivariate ANCOVA or MANCOVA.
- Covariates are continuous explanatory variables that are not of primary interest, but may affect the outcome
- By using ANCOVA, the dependent variable means are adjusted to what they would be if all groups were equal on covariates

Inferential Statistical Procedures

Table 2. Summary of Analysis of Covariance (ANCOVA) and Multivariate Analysis of Covariance (MANCOVA) Results for Intervention Effects on Measures of Sensory Impairment, Behavior, and Developmental Skills

Variable	Group Main Intervention Effect		
	F^a	p	Partial η^{2b}
Univariate ANCOVA analyses			
Autism Behavior Checklist (0–167)	1.116	.309	.074
Parent Stress Index	15.141	.0003	.219
Multivariate analyses: Parent PDDBI			
Post hoc ANCOVA	2.407	.066	.202
Sensory score	4.906	.032	.107
Maladaptive Behavior (AWP/C)	10.025	.003	.196
Language/Social Abilities (REXCSA)	1.194	.281	.028
Autism Composite score	9.116	.004	.182
Multivariate Sensory and Self-Regulation			
Post hoc ANCOVA	5.618	.002	.256
Sense Checklist	12.164	.001	.193
Self-Regulation Checklist	15.626	.00002	.235

Note. Pretreatment scores were used as covariates to control for initial difference. PDDBI = Pervasive Developmental Disorders Behavior Inventory; AWP/C = Approach/Withdrawal Problems Composite; REXCSA = Receptive/Expressive Social Communication Abilities.

^aMANCOVA F s are the Pillai's Trace.

^bPartial η^2 as an effect size estimate can be categorized as *small* (.01–.06), *medium* (.06–.14), or *large* (>.14).

Early Intervention for Autism With a Parent-Delivered Qigong Massage Program: A Randomized Controlled Trial

Louisa M. T. Silva, Mark Schallock, Kristen Gabrielsen

A recent randomized controlled trial (RCT) of a dual parent and trainer-delivered qigong massage intervention for young children with autism resulted in improvement of measures of autism as well as improvement of abnormal sensory responses and self-regulation. The RCT evaluated the effects of the parent-delivered component of the intervention. Forty-seven children were randomly assigned to treatment and wait-list control groups. Treatment group children received the parent-delivered program for 4 mo. Trained therapists provided parent training and support. Improvement was evaluated in two settings—preschool and home—by teachers (blind to group) and parents. Results showed that the parent-delivered program was effective in improving measures of autism (medium effect size) and sensory and self-regulatory responses (large effect size). Teacher data on measures of autism were confirmed by parent data. Results indicate that the parent-delivered component of the program provided effective early intervention for autism that was suitable for delivery at home.

Silva, L. M. T., Schallock, M., & Gabrielsen, K. (2011). Early intervention for autism with a parent-delivered qigong massage program: A randomized controlled trial. *American Journal of Occupational Therapy, 65*, 550–559. doi: 10.5014/ajot.2011.000661

Inferential Statistical Procedures

◎ Chi-Square

- **Used to analyze tallies or frequencies of nominal or categorical data. It may be used to:**
 - Analyze the “goodness of fit” of a single group sample
 - Example: *Given a sample of students, is there a specific preference for teaching styles*
 - Test the degree of independence of 2 or more groups (also known as Pearson’s Chi Square)
 - Example: *Do audiences who watch more TV per week more obese?*
- **Reported as X^2 (df) = ###.#, $p < .##$**
 - Example: In a sample of 1000 QU students, there are more students who prefer flip-flops over traditional close-toed leather shoes as formal wear (X^2 (999) = 189.6, $p < .05$)

Kielhofner, G., Braveman, B., Fogg, L., & Levin, M. (2008). A controlled study of services to enhance productive participation among people with HIV/AIDS. *American Journal of Occupational Therapy, 61*, 36–45.

To compare differences in productivity status at 3, 6, and 9 months, we performed a chi-square analysis on the dichotomous variable (productively engaged vs. not productively engaged). We also calculated effect sizes (odds ratios). Effect sizes are independent of sample size and are especially relevant for determining an intervention's effectiveness (Wilkinson & Task Force on Statistical Inference, 1999). Frequencies for different types of productive engagement (e.g., working, volunteering, being a student) observed at 9 months also were compared across groups.

Table 4. Productive Outcomes Among Participants Who Completed the Enabling Self-Determination (ESD) and Standard Care Programs at 3, 6, and 9 Months Postintervention

Outcome	ESD Program Participants With Productive Participation <i>f/n (%)</i>	Standard Care Participants With Productive Participation <i>f/n (%)</i>	Chi-Square (<i>df</i> = 1)	<i>p</i> ^a	Odds Ratio
3 months	20/28 (71.4)	6/17 (35.3)	5.66	.019	4.58
6 months	21/27 (77.8)	5/18 (27.8)	11.07	.001	9.10
9 months	18/25 (72.0)	5/14 (35.7)	4.88	.031	5.66

Note. *f* = The number of people who had productive outcomes at each time point; *n* = the total number of people available at each time point.

^aOne-tailed Fisher's Exact Test.

Correlational Statistical Procedures

◎ Correlation

- **Used to measure the extent to which two variables are associated (i.e., when X changes, Y tends to change also)**
 - Correlations should NOT be interpreted as a cause and effect relationship, i.e., X does not cause Y to change or vice versa.
- **Reported as $r = .##$, $p < .##$. The “r” is a number that falls between 1.0 and -1.0.**
- **The positive value indicates “positive correlations” while the negative value indicates “negative correlations”**
- **The closer the r is to 1.0 or -1.0 the stronger the association is.**
 - **Example: In a sample of 86 clients who were discharged from an inpatient rehab facility, independence in bathing strongly correlated with community reintegration within 3 months post-discharge ($r = .72$, $p < .05$)**

Gillen, G. (2006). Coping during inpatient stroke rehabilitation: An exploratory study. *American Journal of Occupational Therapy, 60*, 136–145.

When data were further analyzed for frequency of use patterns between the individual strategies, significant positive correlations were found between the coping strategy of humor and the strategies of denial ($r = .734, p < .001$) and behavioral disengagement ($r = .573, p < .05$). In addition, significant positive correlations were found between the strategy of seeking emotional support and the strategies of acceptance ($r = .739, p < .001$) and planning ($r = .579, p < .019$). The strategy of active coping was significantly and positively correlated with the strategy of positive reframing ($r = .571, p < .05$), whereas the coping strategy of planning was significantly and positively correlated with the coping strategy of acceptance ($r = .709, p < .002$).

Correlational Statistical Procedures

◎ **Pearson's r**

- Also known as *Pearson's Product Moment Correlation Coefficient* is used for parametric data.
- Represents quantitatively the extent to which scores on two variables occupy the same relative position.

◎ **Spearman's Rho (r_s)**

- This correlation coefficient is appropriate when either of the following two conditions are met:
 - One variable is an ordinal scale and the other is an ordinal scale or higher.
 - One of the distributions is markedly skewed.
 - In either case, both scales must be converted to ranks

Correlational Statistical Procedures

◎ Regression

- Is a type of analysis designed to predict the levels of another variable.
- A common use for this statistical procedure is to identify risk factors as a predictors of a certain condition.
- Reported as $R^2 = .##$ for each independent variable that is significantly predictive.
- In other words, only those variables that could “account for” as predictive of a condition are included.
 - Example: In one study (Halfon et al, 2001), length of stay, age, and morbidity predisposition were identified as predictive factors in falls within a hospital setting. Gender and type of surgery were minimally correlated but were not considered predictive of falls in the hospital setting.

Neurorehabilitation and Neural Repair

<http://nnr.sagepub.com>

Assessing the Motor Status Score: A Scale for the Evaluation of Upper Limb Motor Outcomes in Patients after Stroke

Mark Ferraro, Jennifer Hogan Demaio, Jennifer Krol, Chris Trudell, Keren Rannekleiv, Lisa Edelstein, Paul Christos, Mindy Aisen, Jill England, Susan Fasoli, Hi Krebs, N Hogan and Bruce T. Volpe
Neurorehabil Neural Repair 2002; 16: 283
DOI: 10.1177/154596830201600306

Figure 1 shows the correlation between the FM score and the MSS score for the upper extremity in the 12 patients who had had evaluations by different examiners. The mean MSS score for each patient was used in this calculation. The correlation was significant ($Y = 3.5 + 0.81 X$, $r^2 = 0.96$, $P < 0.0001$).

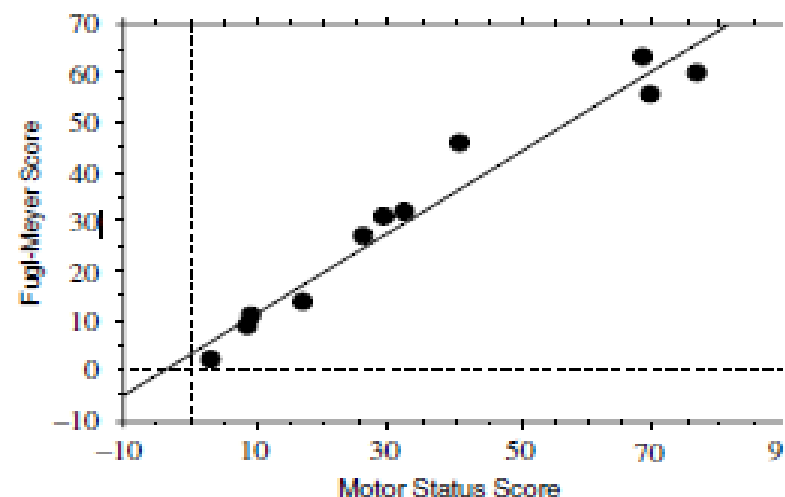


Figure 1. The regression plot of Fugl-Meyer (maximum for the proximal and distal = 66) and Motor Status Scale (maximum for the shoulder/elbow, and the wrist, hand, and fingers = 82) scores for the upper extremity in the 12 patients with stroke. The regression is significant ($Y = 3.52 + 0.82 X$, $r^2 = 0.96$, $P < 0.0001$).

Logistic Regression

○ Logistic regression

- a type of regression analysis that is used to predict the outcome of categorical variables (dependent variables).
- A common use of logistic regression is when the outcome is dichotomous (e.g., fell, did not fall). This procedure is termed as binomial logistic regression.
 - Predicts the odds associated with the presence or absence of the dependent variable/outcome
 - Results presented as odds ratio (OR) – how much more or less likely it is for a participant to belong to the target group than the reference group
- When there are more than two categorical possibilities (e.g., better, no change, worse), the procedure is known as multinomial logistic regression.

Close, J., Ellis, M., Hooper, R., Glucksman, E., Jackson, S., & Swift, C. (1999).

Prevention of falls in the elderly trial (PROFET: A randomized controlled trial. *The Lancet*, 353, 93-97.

Statistical analysis

Based on an average of two falls per year with an SD of 1.5 and a 25% rate of attrition, a sample size of 352 would have 90% power to detect a 30% reduction in the rate of falls from 2.0 to 1.4 in the intervention group with a probability of $p < 0.05$.

At 12-month follow-up, patients were classified according to whether they had died, moved to institutional care, were otherwise lost to follow-up, or remained in the study. This status at follow-up was compared in the two study groups by means of a χ^2 test. Participants who remained in the study at 12 months were included in subsequent analyses.

We compared the two groups in terms of number of falls and serious injuries (fracture or joint dislocation) by Mann-Whitney tests, and reported ability to go out alone by χ^2 test. Differences between groups were compared after adjustment for baseline differences in Barthel index and AMT scores, by multiple logistic regression. For the analyses of risk of falling and risk of

Logistic regression analysis showed that the risk of falling in the 12-month follow-up period was lower in the intervention group than in the control group after adjustment for differences in Barthel and AMT scores at baseline and number of falls in the 12 months before the index fall (odds ratio 0.39 [95% CI 0.23–0.6]). Similarly, there was a significantly lower risk of recurrent falling in the intervention group than in the control group after adjustment for the same baseline variables (0.33 [0.16–0.68]). The odds of at least one hospital admission (as recorded from hospital admission data) were lower in the intervention group than in the control group after adjustment for baseline Barthel and AMT scores (0.61 [0.35–1.05]). The proportions who were able to go out alone at 12 months did not differ significantly after adjustment for Barthel and AMT scores and ability to go out alone at baseline ($p = 0.27$).

Level of Measurement	Sample Characteristics					Correlation
	1-Sample	2-Sample		K-Sample (i.e., >2)		
		Independent Samples	Paired Samples	Independent Samples	Paired Samples	
Categorical or Nominal	χ^2 or Binomial	χ^2	Macnamar's χ^2	χ^2	Cochran's Q	
Rank or Ordinal		Mann-Whitney U	Wilcoxin Matched Pairs Signed Ranks	Kruskal Wallis H	Friedman's ANOVA	Spearman's rho
Parametric (Interval & Ratio)	z test or t test	t test between groups	t test within groups	1-way ANOVA between groups	1-way ANOVA (within or repeated measure)	Pearson's r
		Factorial (2-way) ANOVA For more than 1 Independent Variable (Outcome)				

Adapted from <http://www.uwsp.edu/psych/stat/indexTests.htm>

Measurement Statistics

- ◎ Most correlational statistics are used!
- ◎ Cronbach alpha coefficient
 - is a coefficient of reliability or internal consistency; a measure of how closely related a set of items are as a group.
- ◎ Intraclass correlation coefficient (ICC)
 - is a measure of the reliability of measurements or ratings; often used to determine inter-rater reliability

Research article

Open Access

A modified QuickDASH-9 provides a valid outcome instrument for upper limb function

C Philip Gabel*^{1,2}, Michael Yelland², Markus Melloh³ and Brendan Burkett¹

Stage	Reliability R_{xx} (ICC)	Internal consistency Alpha
Calibration	(n = 46)	(n = 139)
QuickDASH-9	0.94	0.89
QuickDASH	0.94	0.92
DASH	0.98	0.96
ULFI	0.96	0.89
Validation	(n = 22)	(n = 67)
QuickDASH-9	0.94	0.925
QuickDASH	0.91	0.92
ULFI	0.98	0.92

Useful Web References

MedCalc

- www.medcalc.org/index.php

Mathworld

- mathworld.wolfram.com/

Dimension Research

- www.dimensionresearch.com/index.html

Research Methods Knowledge Base

- www.socialresearchmethods.net/kb/index.php