

# Journal of Interpersonal Violence

## Measuring Domestic Violence Offender Change: A 'Real World' strategy for the Practicing Clinician

Journal:	<i>Journal of Interpersonal Violence</i>
Manuscript ID:	Draft
Manuscript Type:	Original Research
Keywords:	Domestic Violence, Assessment < Domestic Violence, Intervention/treatment < Domestic Violence

SCHOLARONE™  
Manuscripts

Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Running Head: MEASURING OFFENDER CHANGE

Measuring Domestic Violence Offender Change: A ‘Real World’ Strategy for the Practicing  
Clinician

For Peer Review

### Abstract

Written specifically for practicing clinicians who are interested in measuring client outcomes and treatment effectiveness but lack the statistical background or confidence to analyze client data, this article provides an explanation and directions for generating a Reliability Change Index (RCI) to measure clinically significant and statistically reliable change after treatment. The study examined pre/posttest data from 1, 228 domestic violence offenders from across the United States and revealed that 12% or fewer offenders made clinically significant improvements as measured by the Domestic Violence Inventory- Pre Post (DVI-PP).

*Keywords: Reliability Change Index (RCI), domestic violence, treatment outcomes, clinical significance*

## Measuring Domestic Violence Offender Change: A 'Real World' Strategy for the Practicing Clinician

This article is written specifically for practicing clinicians who are interested in measuring client outcomes and treatment effectiveness but do not have the statistical background or lack confidence in their ability to analyze and interpret data. A review of key research concepts are provided and clinicians will learn about a specific assessment tool for measuring domestic violence treatment, as well as a formula and directions for calculating client change. Outcomes can be calculated using a spreadsheet software program, widely available in most software packages (e.g., Microsoft EXCEL, Open Office, Google Docs). The focus on this paper is domestic violence (DV) and treatment outcomes, however the principles can be applied to other client outcome data.

Domestic violence and batterer treatment programs have undergone extensive study (list references) and several predictors of treatment completion (Buttelle & Pike, 2003; Dalton, 2001; Gondolf, 2000; Muftic and Bouffard, 2007) and recidivism (Daly & Pelowski, 2000; Gondolf & Wernik, 2009; Tollefson & Gross, 2006) have been identified. Buttelle and Pike (2003) used pre/post design and results revealed changes in the desired direction for the offenders after treatment; however, approximately 20% reoffended within 12-months of completing treatment. Herman and colleagues (2014) also used a pre/posttest design and administered several questionnaires prior to and after completing domestic violence treatment. Comparisons between pre/post test scores found statistically significant changes in self-reported attitudes and behaviors of treatment completers but approximately 30% reoffended during the 9-year follow-up period.

Pre/posttest research designs are commonly used to measure changes in self-reported attitudes, beliefs and behaviors (Dimitrov & Rumrill, 2003) and can be useful for measuring

1  
2  
3 overall treatment effectiveness (group level) but are not appropriate to address questions about  
4  
5 individual improvement or decline after treatment. In the current era of accountability, clinicians  
6  
7 need to provide empirically, as well as clinically supported, treatment-related decisions at the  
8  
9 individual level and pre/posttest comparisons do not provide sufficient information at the person  
10  
11 level or necessary detail about the magnitude of change a person may have experienced  
12  
13 (Chelune, Naugle, Luders, Sedalk, & Awad, 1993). A Reliability Change Index (RCI) offers  
14  
15 clinicians a relatively simple approach for identifying clinically significant improvement at the  
16  
17 individual level.  
18  
19  
20  
21

### 22 **Reliability Change Index**

23  
24 Reliability Change Indices (RCI) have a long history of use in neuropsychology measuring  
25  
26 change across test battery administrations (Woods, Childers, Ellis, Guaman, Grant & Heaton,  
27  
28 2006) in Alzheimer's patients, (Malek-Ahmadi, Chen, Davis, Belden, Powell, Jacobson, &  
29  
30 Sabbagh, 2015), in post-concussive athletes (Parsons, Notebaert, Shields, & Guskiewicz, 2009)  
31  
32 and stroke patients (Middel & van Sonderen, 2010). Moreover, RCI has been recommended for  
33  
34 use in rehabilitation counseling (Johnson, Dow, Lynch & Hermann, 2006) mental health  
35  
36 treatment (Evans, Margison, & Barkham, 1998) and for measuring psychotherapy outcomes  
37  
38 (Margison, Barkham, Evans, McGrath, Cark, Audin, & Connell, 2000; Wise, 2004).  
39  
40  
41  
42

43 Reliability Change Index (RCI) measures clinical effectiveness in standardized units,  
44  
45 summarizes the direction of change and indicates whether the change is reliable (Zahra & Hedge,  
46  
47 2010). An RCI score provides clinicians with two key pieces of information. The first is whether  
48  
49 the posttest score demonstrates clinical improvement that is statistically reliable (i.e., not the  
50  
51 result of chance or error). For example, using a .05 (two-tailed) alpha represents  $\pm 1.64$ ; this  
52  
53 means that an RCI value exceeding  $\pm 1.64$  would occur by chance 10% of the time. We could  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 conclude that a score which exceeded  $\pm 1.64$  was the result of the treatment and not because of  
4 chance or measurement error. The second use of an RCI is its role in creating a confidence  
5 interval (Chelune et al., 1993; Johnson, et al., 2006) sometimes referred to as a prediction  
6 interval (Temkin, Heaton, Grant, & Dimen, 1999) to establish critical values. The confidence  
7 interval uses these critical values to establish cutoff scores based on a predetermined alpha level  
8 (e.g.,  $\pm 1.64$ ). A change score (posttest-pretest) must exceed the cutoff score to be considered  
9 clinically significant. This will be discussed in greater detail with examples in subsequent  
10 paragraphs.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

22 A review of the various forms of significance are provided to clarify and distinguish between  
23 the various types of significance that a clinician encounters when administering assessments and  
24 interpreting scores. Readers are directed to Thompson (2002) for additional information and a  
25 historical perspective.  
26  
27  
28  
29  
30

31 **Statistical significance** for pre/posttest designs, indicates whether the sample scores are  
32 different from the null hypotheses (i.e., there is no difference between pre/posttest  
33 scores). Conventionally, .05 (5/100), .01 (1/100), and .001 (1/1000) are used as  
34 thresholds for establishing whether an event is random or related to the treatment. Note  
35 that statistical significance does not describe the impact or importance of the results.  
36  
37  
38  
39  
40  
41  
42

43 **Practical significance** is commonly referred to as effect size and represents the  
44 magnitude of the treatment effect. In other words, how large were the differences in  
45 pretest and posttest scores (Ferguson, 2009; Thompson, 2002). In general, large effect  
46 sizes are more likely to be clinically significant than small effect sizes but not always.  
47  
48  
49  
50

51 **Clinical significance** refers to the value or the importance of the treatment effect; the  
52 impact on everyday functioning, interactions and relationships.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 It is important to remember that an improvement in everyday functioning may not be  
4 statistically significant and a result that is statistically significant may have no noticeable impact  
5 on everyday functioning. This paradox supports the value of using an RCI to assess treatment  
6 effectiveness.  
7  
8  
9  
10

11  
12 Jacobson and Truax are often credited with establishing the first RCI that examined clinical  
13 significance, while simultaneously measuring score consistency across the pretest and posttest  
14 administrations (Johnson et al., 2006). The statistic used in this study is a modification of the  
15 RCI reported by Jacobsen and Truax and controls measurement error and practice effects.  
16  
17 Practice effects, as readers may recall, refers to score changes that are not related to the treatment  
18 but are associated with exposure to the test materials or testing procedures (Duff, 2012). The  
19 RCE formula used in this study was:  
20  
21  
22  
23  
24  
25  
26  
27

$$28 \quad RCI + P = \frac{29 \quad ((X_2 - X_1) - (M_2 - M_1))}{30 \quad SDD}$$

31  
32  
33 This formula was initially reported and used by Chelune and colleagues (1993) to examine  
34 cognitive performance of individuals after surgery for treatment of epileptic seizures. Parsons,  
35 Notebaert, Shields and Guskiewicz (2009) adapted it to evaluate post-concussion improvement  
36 and decline in college athletes and its use in this study is considered appropriate.  
37  
38  
39  
40  
41  
42

43  
44 If we put the formula into words,  $X_1$  is the offender's pretest score,  $X_2$  is the offender's  
45 posttest score,  $M_1$  is the group mean pretest score,  $M_2$  is the group mean posttest score. The  $M_2 -$   
46  $M_1$  difference represents practice effects; thus the same amount is subtracted from each score to  
47 correct for systematic bias. SDD is the standard deviation of the difference and is generated  
48 using pre/posttest difference scores (change scores) from the sample. Readers may recognize this  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 as a modification of the  $z$ - score formula, which transforms scores into standardized units for  
4  
5 comparison.  
6

7  
8 The paragraphs above have reviewed the essential basics of calculating an RCI+P, which we  
9  
10 noted earlier provides clinicians with information about offender change that it clinically  
11  
12 significant and statistically reliable. The remainder of this article describes the analysis process  
13  
14 and results using pre/posttest data from a sample of domestic violence offenders from across the  
15  
16 United States. The aim of this paper is to illustrate how calculating an RCI can inform clinical  
17  
18 decision making. The formula, interpretation and clinical application are reviewed using  
19  
20 examples.  
21  
22  
23

## 24 **Methods**

### 25 **Procedures**

26  
27  
28 Data were extracted from the test developer's database. There were 13, 309 DVI Pre-Post  
29  
30 (DVI-PP) completed from April 1, 2011 – December 31, 2014; 11, 890 pretests were  
31  
32 administered and 1, 419 posttests were submitted. Pre and posttest records were matched,  
33  
34 duplicate and incomplete records were removed ( $N = 123$ ) leaving 1, 296 offenders who had  
35  
36 completed a pretest and posttest. A second data cleaning process removed observations with  $<6$   
37  
38 days and  $>701$  treatment days, which represented data at the 99<sup>th</sup> percentile ( $N = 68$ ). After this  
39  
40 process, there were 1, 228 matched pretest and posttest administrations.  
41  
42  
43  
44

### 45 **Participants**

46  
47  
48 Demographic information and criminal history were reported by offenders on their  
49  
50 answer sheets. In general, the overwhelming majority of offenders were male (85%), Caucasian  
51  
52 (62%) with a high school education or higher (75%). Approximately 44% of offenders reported  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 they were single. The average age of offenders was 35 and the average age at time of first  
4  
5 conviction was 26.  
6  
7

8 Eighty-eight percent of offenders reported one or more arrests, 80% reported one or more  
9  
10 domestic violence arrests, 37% reported one or more alcohol-related arrests, 22% reported one or  
11  
12 more drug-related arrests, and 29% reported one or more arrests for assault.  
13  
14

15 Using *pretest* risk range classifications, approximately 70% of offenders scored in the  
16  
17 Low Risk range on the Alcohol Scale, Control Scale, and Drug Scale. Approximately, 60% of  
18  
19 offenders scored in the Low Risk range on the Violence Scale. *Posttest* risk range classifications  
20  
21 revealed that over 80% of offenders scored in the Low Risk range on the Alcohol Scale, Control  
22  
23 Scale, Drug Scale and Violence Scale. Figure 1 summarizes group changes in risk classification.  
24  
25  
26

27 The average number of treatment days (number of days between pretest and posttest  
28  
29 administration) was 263 days; range 7 – 700 days of treatment. No additional information about  
30  
31 treatment methods or approaches were available for review or analysis.  
32  
33

### 34 **Instrument**

35  
36 The Domestic Violence Inventory Pre-Post (DVI-PP) uses 147 true/false, multiple choice  
37  
38 items that comprise seven domains associated with domestic violence. Scale descriptions,  
39  
40 provided by the test developer are summarized here for reader convenience. Additional  
41  
42 information on the DVI Pre-Post (DVI-PP) can be found at <http://www.dvi-pre-post.com>.  
43  
44  
45

46 *Truthfulness Scale* serves as validity measure and uses 21 true/false items. All interview  
47  
48 and self-report information is subject to the dangers of untrue answers due to defensiveness,  
49  
50 guardedness or deliberate falsification. The Truthfulness Scale identifies these self-protective,  
51  
52 recalcitrant and guarded people who minimize or even conceal information.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

*Alcohol and Drug Scales:* Frequently, domestic violence offenders are under the influence of alcohol or drugs at the time of the incident and substance abuse was associated with recidivism and failure to complete treatment. The Alcohol Scale and Drug Scale, individually measure frequency of use and the degree of severity of alcohol and drug problems including misuse and abuse. The Alcohol Scale uses 21 true/false and multiple choice items and the Drug Scale uses 22 true/false and multiple choice items to assess problem severity.

*Control Scale* measures control of others and control of self uses 25 true/false and multiple choice items. Controlling behaviors vary from swearing and intimidation to battering. Control is often synonymous with power. Controlling behaviors can represent subtle acts of manipulation, influence and persuasion to gain power over others, or these behaviors can escalate to anger and aggression. In its extreme form, control can become an obsession and personal power may be found through the control of others.

*Violence Scale* uses 32 true/false and multiple choice items to measure the use of physical force to injure, damage, or destroy. Its purpose is to identify individuals who are dangerous to themselves and others or demonstrate a propensity for violence. A person's aggression (e.g., acting out potential) may be related to substance abuse, overall adjustment, emotional problems, and traits such as aggressiveness or risk-taking.

*Stress Coping Abilities Scale* asks offenders to rate themselves on a 4-point Likert-type scale that is comprised of 32 items. This scale is the only scale that measures protective and prosocial factors. It is a measure of experienced stress, as well as buffers of stress and a person's coping abilities.

## Analysis

### Reliability

1  
2  
3 Cronbach's Alpha, a measure of reliability, was used to measure the internal consistency of  
4 the items in each of the DVI-PP scales for pretest and posttest administrations. In addition, test-  
5 retest reliability scores were calculated for each scale.  
6  
7  
8  
9

### 10 **t-test**

11  
12 This study also included a pre/posttest design. Scores were used to measure change in  
13 attitudes and behaviors after completing domestic violence treatment for the sample of  
14 offenders. Bonferroni correction was applied to control for experimentwise error ( $p = .007$ ).  
15  
16  
17  
18

### 19 **Reliability Change Index (RCI+P)**

20  
21 On the DVI-PP, a decrease in posttest scores indicated a decrease in problem severity, as  
22 measured by DVI-PP scales; whereas, an increase in posttest scores indicated increased  
23 problem severity after treatment. At the individual level, reliable improvement occurred  
24 when values exceeded +1.64 and reliable decline occurred when values exceeded -1.64.  
25  
26  
27  
28  
29  
30  
31  
32 RCI+P was also used to identify cutoff scores to identify clinically significant improvement.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Cutoff scores were established using the formula ( $SDD * 1.64$ ). Cutoff scores were generated  
for each DVI-PP scale. Offenders were then grouped into three categories, Improvement,  
Decline, and No Change based on whether they exceeded the cutoff score values.

For clinicians interested in analyzing their own clients' data the steps used in this study  
are provided or clinicians can use the standard deviation of the difference (SDD) found in  
Table 3 to calculate RCI+P using their own pretest and posttest data.

### **Initial Steps:**

1. Administer pretest instrument
2. Administer posttest instrument
3. Determine alpha

**In a spreadsheet program:**

4. Enter pre/posttest scores
5. Calculate difference between pre/posttest scores ( $X_2 - X_1$ )
6. Calculate mean and standard deviation of pretest, posttest and difference scores
7. Calculate cutoff scores ( $\pm 1.64 \times SDD$ )
8. Calculate the RCI (use earlier formula) for each offender
9. Compare difference scores to RCI cutoff scores
10. Complete steps 4 – 6 for each DVI-PP scale.

**Results**

Table 1 displays pretest and posttest reliability coefficients for the scales scores. Reliability coefficients were  $>.85$  for all DVI-PP scales. Perfect reliability is 1.00 and the professionally accepted standard of reliability that is often reported is  $.75$  (Nunally, 1978). Some researchers have suggested that reliability coefficients between  $.60-.90$  may be appropriate depending on the nature of the instrument and the construct being measured (Murphy & Davidshofer, 2001). Test-retest reliability results were significantly lower than internal consistency measures, as would be expected for a test that measures treatment effectiveness-- scores are expected to change. Test-retest coefficients were statistically significant for all scales ( $p = .01$ ).

Results of the pre/posttest results showed that, *as a group*, the sample made statistically significant improvement across all domains, as measured by the DVI-PP ( $p < .001$ ). Table 2 summarizes results including Cohen's  $d$  a measure of practical significance. Effect sizes ranged from small to medium, respectively.

Reliability Change Index (RCI+P) results provided additional insight into the offender sample and results are presented in Table 3. Table 4 summarizes the percentage of offenders who

1  
2  
3 improved, decline and those who experienced no clinically significant change. On the Alcohol  
4 Scale, 6% of offenders who scored 18 points lower than their pretest score made clinical gains;  
5  
6 6% of offenders who scored 18 points higher than their pretest score endorsed more problems  
7  
8 after treatment and approximately 88% did not meet the threshold for clinically or statistically  
9  
10 significant change. On the Control Scale, 5% showed clinical improvement by scoring 20 points  
11  
12 lower than their pretest score, 1% scored worse in this domain after treatment, and 94% did not  
13  
14 experience clinical or statistical change as measured by the Control Scale. There were 5% of  
15  
16 offenders who showed improved on the Drug Scale by scoring 17 points lower after treatment,  
17  
18 2% reported more drug-related problems after treatment, and 93% reported no change with drug  
19  
20 problems. A clinically significant decrease in propensity for violence was achieved with a 26  
21  
22 point decrease on the Violence Scale. Ten percent of offenders of offenders made clinical  
23  
24 improvement in the area of violence, 1% of offenders reported more violent behaviors and  
25  
26 attitudes after treatment and 89% indicated no change. Eleven percent of offenders made clinical  
27  
28 improvement in coping and stress management; 3% reported a decrease in coping skills and 86%  
29  
30 reported no clinically or statistically significant change in stress management and coping.  
31  
32  
33  
34  
35  
36  
37  
38

### 39 Discussion

40  
41 The purpose of this article was to explain a relatively simple method for measuring treatment  
42 effectiveness, which did not require extensive statistical experience. This study used data  
43 submitted by domestic violence offenders from across the United States who had completed the  
44 DVI-PP. Data were extracted from the test developer's research database. Interested clinicians  
45  
46 can use the group means and standard deviations presented in this study, or use the directions  
47  
48 provided to generate their own statistics.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 As indicated by the risk classification comparison, a majority of offenders were considered  
4 low risk on many of the DVI-PP scales at the time of initial testing. Pre/posttest score analyses  
5 revealed statistically significance improvements on all DVI-PP scales for the sample. At the  
6 individual level, RCI criteria were relatively stringent (i.e., 5% of scores at the clinically  
7 significant level). With these parameters, approximately 6% of offenders showed changes in  
8 problem minimization and alcohol misuse. Approximately 5% of offenders made improvements  
9 in the area of personal control and drug misuse. A larger percentage of offenders reduced  
10 violence (10%) and improved coping skills (11%).

11  
12 The relatively high percentage of offenders classified as Low Risk on many of the DVI-PP  
13 scales may account for why so few offenders made any significant clinical improvement. This is  
14 referred to a *floor effects*. Simply stated, scores that start low (i.e., Low Risk) often cannot go  
15 lower (Duff, 2012). This factor should be considered when interpreting results.

### 31 **Limitations**

32  
33 Additional considerations and limitations include sample characteristics, methodological  
34 considerations, and practical implications. With regard to the sample, test data were submitted  
35 from across the country with no information on treatment modality, frequency or type of service  
36 provided. Without more specific treatment information it is difficult to know whether a specific  
37 approach or treatment intensity was more effective than another. Future research should include  
38 treatment information to improve generalization and best-practice approaches in the field.

39  
40 Including practice effects added precision to the analysis (Temkin et al., 1999), however  
41 the same quantity was applied across all offenders in the sample. As Duff (2012) rightly points  
42 out, it is a one size fits all approach that does not account for individual differences. Regression  
43 approaches are more individualized and more precise but calculations and interpretation are often  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 beyond the statistical training and interest of most clinicians (Parsons et al., 2009). Despite the  
4  
5 limited variability, RCI +P produces results comparable to more sophisticated regression  
6  
7 approaches (Heaton et al., 2001; Temkin, et al., 1999).  
8  
9

10 The retrospective nature of this study also limits its practical applications. No data  
11  
12 describing recidivism post-treatment were available for this study; using a prospective or  
13  
14 longitudinal methodology could, more specifically, examine long term impact of treatment on  
15  
16 recidivism. In addition, the absence of a control group limits the generalization of the  
17  
18 information. A control group comprised of individuals who have not been referred for or  
19  
20 participated in domestic violence treatment would strengthen the efficacy and score  
21  
22 interpretation.  
23  
24  
25

26  
27 Another issue, salient for clinicians, is the instance when an offender demonstrates higher  
28  
29 risk after attending treatment but whose score does not exceed the predetermined threshold.  
30  
31 These offenders do not seem to have benefited from domestic violence treatment but do meet the  
32  
33 threshold of increased risk and problem severity. RCI+P is only one factor that clinicians can use  
34  
35 when making discharge decisions and recommendations. Additional aftercare programming or  
36  
37 complimentary treatment (e.g., substance abuse) may also be necessary. It is important to note  
38  
39 that the use of 10% (i.e.,  $\pm 1.64$ ) was selected by this researcher (using previous research as a  
40  
41 guide). Clinicians interested in adapting this approach may opt for less stringent criteria or more  
42  
43 stringent criteria. Modifications should be grounded in a well-reasoned rationale that supports  
44  
45 clinical decision-making.  
46  
47  
48  
49

50 The RCI+P is a useful method that clinicians can implement, using simple calculations,  
51  
52 to demonstrate whether treatment was effective. This study used pre/posttest data from domestic  
53  
54 violence offenders but the methods can be applied to other outcome assessments. RCI+P can  
55  
56  
57  
58  
59  
60

1  
2  
3 provide clinicians with another tool to aid in decision-making. Future areas of research include  
4  
5 the use of a control group, a longitudinal methodology and gathering more specific treatment  
6  
7 related information to determine whether some interventions are more effective than others at  
8  
9 reducing recidivism.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review



## References

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- Buttell, F. P., & Pike, C. K. (2003). Investigating the Differential Effectiveness of a Batterer Treatment Program on Outcomes for African American and Caucasian Batterers. *Research on Social Work Practice, 13*, (6) 675-692. doi: 10.1177/1049731503254055
- Chelune, G. J., Naugle, R. I., Luder, H., Sedlak, J., & Awad, I. A. (1993). Individual change after epilepsy surgery: Practice effects and base-rate information. *Neuropsychology, 7* (1), 41-52.
- Dalton, B. (2001). Batterer characteristics and treatment completion. *Journal of Interpersonal Violence, 16* (12), 1223- 1237.
- Daly J., & Pelowski, S. (2000). Predictors of dropout among men who batter: A review of studies with implications for research and practice. *Violence and Victims, 15*, 137-160.
- Dimitrov, M. D., & Rumrill, P. D. Jr.(2003). Pretest-Posttest designs and measurement of change. *Work, 20*159-165.
- Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. *Archives of Clinical Neuropsychology, 27*, (3) 248-261
- Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health, 1*, 70-72. DOI: 10.1136/ebhm.1.3.70.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40* (5) 532-538.
- Gondolf, E. W. (2000). Mandatory court review and batterer program completion. *Journal of Interpersonal Violence, 15*, 428-437.

- 1  
2  
3 Gondolf, E. W., & Wernik, H. (2009). Clinician ratings of batterer treatment behaviors in  
4 predicting reassault. *Journal of Interpersonal Violence*, 29 (11), 1792-1815.  
5  
6  
7  
8 Herman, K. Rotunda, R., Williamson, G., & Vodanovich, S. (2014). Outcomes from a Duluth  
9 model batterer intervention program at completion and long term follow-up. *Journal of*  
10 *Offender Rehabilitation*, 53, 1-18. DOI: 10.1080/10509674.2013.861316  
11  
12  
13  
14  
15 Johnson, E. K., Dow, C., Lynch, R. T., & Hermann, B. P. (2006). Measuring clinical significance  
16 in rehabilitation research. *Rehabilitation Counseling Bulletin*, 50 (1), 35-45.  
17  
18  
19  
20 Malek-Ahmadi, M., Chen, K., Davis, K., Belden, C. M., Powell, J., Jacobson, S. A., & Sabbagh,  
21 M. N. (2015). Sensitivity to change and prediction of global change for the Alzheimer's  
22 Questionnaire. *Alzheimer's Research and Therapy*, 7 (1) 1-9. DOI: 10/1186/s13195-014-  
23 0092-z  
24  
25  
26  
27  
28  
29 Middel, B. & van Sonderen, E. (2010). Responsiveness and validity of three outcome measures  
30 of motor function after stroke rehabilitation. *Stroke*, 41, DOI:  
31  
32 10:1161/Strokeha.110.580365. Retrieved <http://stroke.ahajournals.org/content/41/7/e463>  
33  
34  
35  
36  
37 Margison, F. R., Barkham, M., Evans, C., McGrath, G., Clark, J., Audin, K., & Connell, J.  
38 (2000). *British Journal of Psychiatry*, 177, 123-130.  
39  
40  
41 Muftic, L. R., & Bouffard, J. A. (2007). An evaluation of gender differences in the  
42 implementation and impact of a comprehensive approach to domestic violence. *Violence*  
43 *Against Women*, 13, 1 46-69.  
44  
45  
46  
47  
48 Murphy, K. R. & Davidshofer, C. O. (2001). (5<sup>th</sup> ed.). *Psychological Testing: Principles and*  
49 *Applications*. Upper Saddle River, NJ: Prentice Hall.  
50  
51  
52  
53 Nunally, J. C. (1978). *Tests and measurements*. New York: McGraw Hill.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- Parsons, T. D., Notebaert, A. J., Shields, E. W., & Guskiewicz, K. M. (2009). Application of reliable change indices to computerized neuropsychological measure of concussion. *International Journal of Neuroscience*, *119*, 492-507. DOI:10.1080/00207450802330876
- Temkin, N. R., Heaton, R. K., Grant, I., & Dikmen, S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neurological Society*, *5*, 357-369.
- Thompson, B. (2002). "Statistical", "Practical" and "Clinical": How many kinds of significance do clinicians need to consider? *Journal of Counseling and Development*, *80* (1), 64-71.
- Tollefson, D. R., & Gross, E. R. (2006). Predicting recidivism following participation in a treatment program for batterers. *Journal of Social Service Research*, *32* (4), 39-61. DOI:10.1300/J079v32n04\_03
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessments*, *82* (1) 50-59.
- Woods, S. P., Childers, M., Ellis, R. J., Guaman, S., Grant, T., & Heaton, R.K. (2006). A battery approach for measuring neuropsychological change. *Archives of Clinical Neuropsychology*, *21*, 83-89. DOI:10.1016/j.acn.2005.07.008.
- Zahra, D. & Hedge, C. (2010). The reliability change index: Why isn't it more popular in academic psychology? *PsyPag Quarterly*, *76*, 14-19.

Table 1.

## DVI-PP Reliability Coefficients

<u>Scale</u>	<u>Pre-Test</u>	<u>Posttest</u>	<u>ICC</u>
Truthfulness	.88	.89	.25*
Alcohol	.90	.86	.21*
Control	.89	.84	.18*
Drug	.85	.83	.29*
Violence	.89	.84	.17*
Stress Coping Abilities	.93	.92	.27*

N = 1, 228

ICC = test-retest coefficient

\*p &lt; .01

Table 2.

Pre and Posttest Score *t*-test Results

<u>Scales</u>	<u>Pretest</u>		<u>Posttest</u>		<i>t</i>	<i>p</i>	<i>d</i>
	Mean	SD	Mean	SD			
Truthfulness	8.10	5.32	8.85	5.59	3.91	<.001	.13
Alcohol	7.68	8.82	5.51	6.48	7.77	<.001	.28
Control	8.36	7.96	5.49	5.66	11.32	<.001	.42
Drug	6.46	7.84	4.65	5.97	7.62	<.001	.26
Violence	16.77	12.12	9.07	7.65	20.52	<.001	.76
Stress Coping	111.34	43.64	130.57	44.40	12.69	<.001	.43

N=1, 228; SD = standard deviation

Table 3.

Pre/Posttest Means and Standard Deviations for the DVI-PP

Scales	<u>Pre-Test</u>		<u>Post-test</u>		<u>Difference</u>		<u>Standard Error</u>		<u>Cut score</u>
	Mean	SD	Mean	SD	$M_2 - M_1$	SDD	<u>SEM</u>	$SE_{diff}$	Points
Truthfulness	8.10	5.32	8.85	5.59	.79	6.64	4.61	6.52	±11
Alcohol	7.68	8.82	5.51	6.48	-2.23	9.93	7.84	11.09	±18
Control	8.36	7.96	5.49	5.66	-2.72	8.95	7.21	10.19	±20
Drug	6.46	7.84	4.65	5.97	-1.82	8.57	6.61	9.34	±17
Violence	16.77	12.12	9.07	7.65	-7.57	13.17	11.04	15.62	±26
Stress Coping	111.34	43.64	130.57	44.40	19.19	53.18	37.29	52.73	±86

N = 1, 228

Cut scores were established using formula (SDD x ±1.64)

Table 4.

Percentage of Offenders with Statistically Reliable Changes on the DVI-PP

Scale	IMP % (n)	DEC % (n)	NC % (n)
Truthfulness	6% 74	6% 75	88% 1079
Alcohol	6% 79	3% 32	91% 1117
Control	5% 55	1% 15	94% 1158
Drug	5% 64	2% 26	93% 1138
Violence	10% 119	1% 12	89% 1097
Stress Coping	11% 133	3% 39	86% 1056

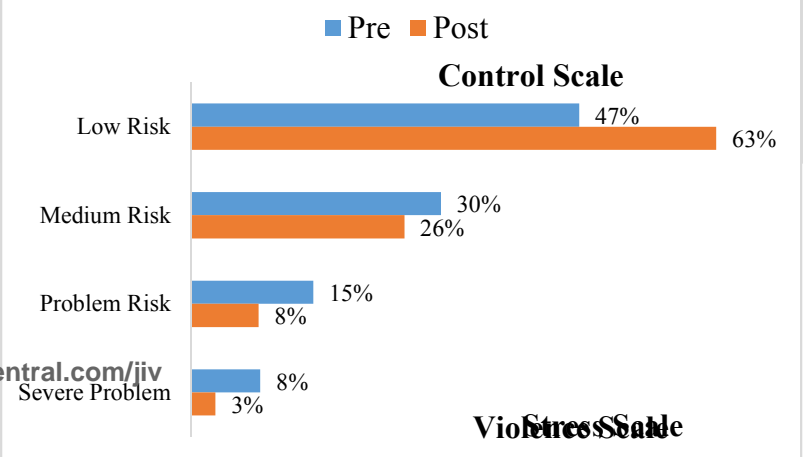
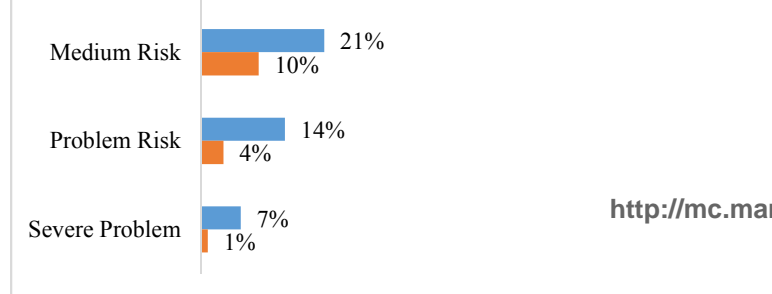
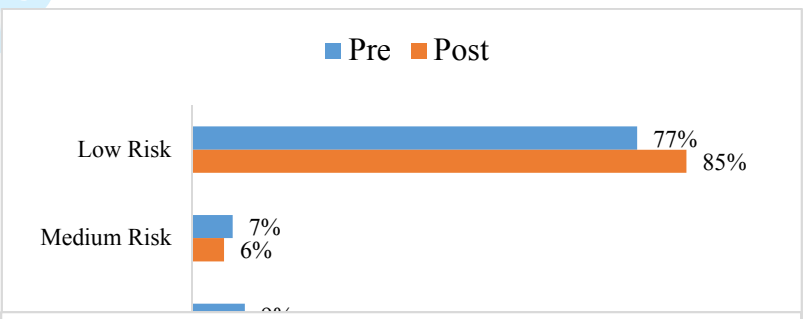
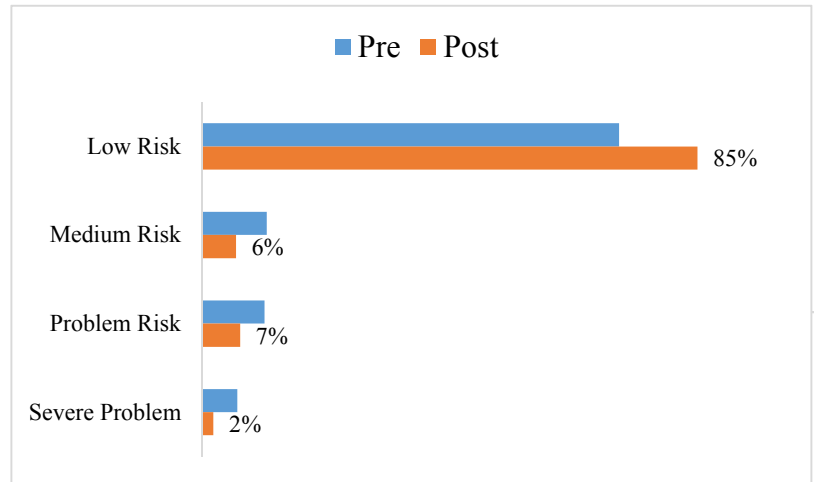
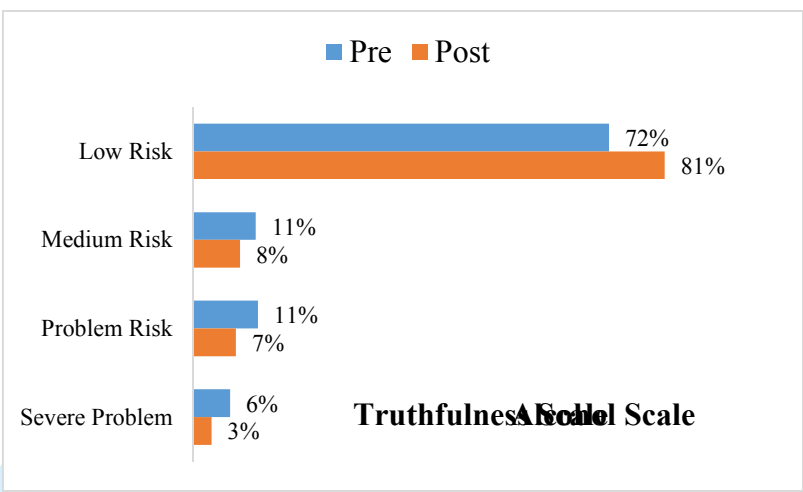
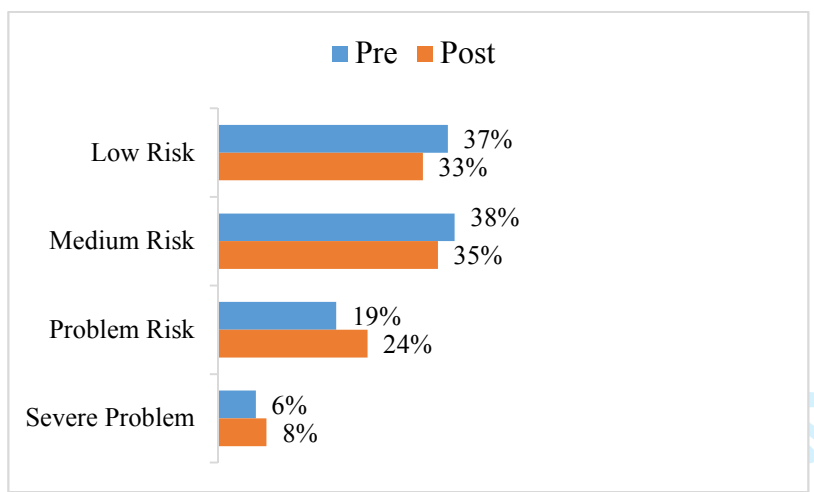
IMP = Improvement

DEC = Decline

NC = No Change

Figure 1

Risk Range Comparisons by Scale





For Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review