2009-07-08

Receiver Operating Characteristic (ROC) Curve Analysis of Affinity Profiles

Benjamin C. Caswell
Brigham Young University - Provo

Receiver Operating Characteristic (ROC) Curve Analysis of Affinity Profiles

by

Benjamin C. Caswell

A thesis submitted to the faculty of

Brigham Young University

Department of Counseling Psychology and Special Education

Brigham Young University

April 2009

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Benjamin C. Caswell

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

June 1, 2009	
Date	Lane Fischer, Chair
Date	Barbara Culatta, Member
Date	Aaron Jackson, Member

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Benjamin C. Caswell in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

June 30, 2009
Date

Lane Fischer
Chair, Graduate Committee

Accepted for the Department

Ellie Young
Graduate Coordinator

Accepted for the College

Barbara Culatta
Associate Dean, David O McKay School of Education

Abstract

Receiver Operating Characteristic (ROC)

Curve Analysis of Affinity Profiles

Benjamin C. Caswell

Department of Counseling Psychology and Special Education

Education Specialist

This thesis tests the relative efficiency of Fischer's approach of Chi-square on Affinity viewing time profiles to the more commonly used deviance differential approach. Through his use of a Chi-square approach, Fischer has attempted to make a norm-referenced, comparison of ipsative scores generated using Affinity. His goal has been to create an ethically acceptable approach to identifying sexual interest without losing the efficiency generated through the use of the commonly used deviance differential. A Receiver Operating Characteristic (ROC) curve is used to generate the efficiency of each approach and provide results for comparison. These results lead to a discussion of their implications for Fischer's approach of Chi-square, the deviance differential approach, and to the general use of data generated by Affinity.

Acknowledgments

This thesis is the result of great dedication on the part of several people who are committed to my success and the success of this research. Lane Fischer, my committee chair, has devoted a portion of his life to this work and to my success, and I will always be grateful for him. Without my father, Robert Caswell, much of the work of this thesis would have been an impossible undertaking—thanks dad. I am blessed to have a brother, Dale Caswell, who is a literary genius and who spent several late nights editing and supporting my efforts; his help has been invaluable. I am grateful for Barbara Culatta, a committee member and personal friend, who has not only been a great asset to my paper, but a support to me and my family. I would also like to thank Aaron Jackson for accepting to be on another committee overseeing this work and for the comments and suggestions that have improved this paper. Lastly, I am so grateful for my wife, Megan Caswell, without her support, understanding and devotion this paper would not have been possible. Thank you all!

Table of Contents

List of Tables

List of Figures

Introduction

Sex offenses against children can be devastating, and they are distressing at any prevalence rate. Children affected by a sex crime are likely to experience lasting negative effects in physical, behavioral, emotional, and psychological domains (Dailey, 1999). An important way to protect our children from sex crimes is by identifying and treating the source. A key part in the identification process is to predict the likelihood of a future offense occurring, either relapse or an initial offense. Once a sex offender is caught and convicted, the strongest predictor of relapse is the persistence of deviant sexual interest (Hanson and Morton-Bourgon 2005). Assessment of deviant sexual interests is accomplished through several techniques including clinical interviews, self report measures, image-viewing tasks administered in conjunction with data reception strategies, and image viewing time (VT).

Although assessment of deviant sexual interest has been proven to be the best of the currently used predictors of relapse in known offenders, it is also commonly used in inpatient and outpatient settings to screen and diagnose people accused of sex crimes. While it is true that screening and diagnosis may be a rational extension of the meta-analytic prediction of relapse (assessment of sexual interest after the offense has been committed), there is a high need to develop procedures for screening and diagnosis *before* an offense has been committed and adjudicated. The effectiveness of screening procedures can be tested using a Receiver Operator Characteristic (ROC) procedure, which compares data obtained from two populations and identifies any overlap between them. An effective screening tool would result in little or no overlap between the two

population (convicted pedophiles and non-offenders) curves generated using a ROC analysis.

Fischer (2004, 2006, 2007) has identified problems with current scoring and assessment procedures in the domain of identifying sexual interest. Data reception and viewing time strategies both utilize ipsative scoring procedures, which Fischer argues are problematic. Ipsative scores make comparisons using one's own data—without the use of a norm. Fischer has attempted to resolve these problems by developing an approach using Chi-square procedures for screening and diagnosis. A Chi-square approach would allow scores, or a pattern of scores, to be normed; providing a source for direct comparison and potentially alleviating the problems associated with ipsative scores. Fischer's analysis of Chi-square (CS) procedure differs from the common procedure, which is to calculate a deviance differential (DD). A DD is created by identifying the discrepancy between two scores generated by an individual. CS scoring procedure results in a pattern of VT, which can then be standardized and norm referenced. This may result in a more favorable base for comparison and inference, as opposed to previous approaches using ipsative scores.

The purpose of this study is to test two scoring procedures (CS versus DD) using Affinity VT data via ROC analysis to determine which is the more effective at screening non-offenders from known pedophiles. Given Fischer's untested approach using a Chi-square procedure for screening non-offenders from pedophiles, it's unclear how its effectiveness will compare to a more traditional deviance differential.

Review of Literature

Crimes of sexual nature can have a devastating and long lasting affect on individuals, families, communities, and society. According to an article by Crime Magazine (O'Connor, 1998), a study conducted in 1995 by the U.S. Department of Justice estimated the number of rape/sexual assault victims in the U.S. to be 355,000, with 73.5% of those victims knowing their attacker(s). The study included data indicating that 48% of those arrested for rape were convicted, although only 2% were acquitted of the crime. Child victims of sex crimes have experienced numerous negative consequences—including fears, posttraumatic stress disorder, behavior problems, sexualized behaviors, poor self-esteem, anger, shame, guilt, interpersonal difficulties, prostitution, promiscuous sex, substance abuse, and attempted and/or completed suicide (Dailey, 1999).

Hanson & Harris (2000) note the significance of deviant sexual interest as it relates to recidivism, and according to their work sexual interest has been repeatedly found to be the strongest predictor of criminal re-offense. "There is now a general consensus that sexual recidivism is associated with at least two broad factors: (a) deviant sexual interests and (b) antisocial orientation/lifestyle instability" (Hanson & Morton-Bourgon, p. 1155). Once we determine whether or not an individual possesses a deviant sexual interest, we can then better focus our attention on factors that determine an individual's likelihood of re-offense, or maybe even an innocent individual's likelihood of an initial offense.

Prediction of Relapse

The assessment of sexual interest has become an important component in the study of sex offense, and particularly in pedophilic sexual offenses. Deviant sexual interest has been identified as the leading factor associated with sexual recidivism (Hanson & Morton-Bourgon, 2005). Hanson and Morton-Bourgon further note that given an average follow-up time of 5-6 years, the likelihood of recidivism of a sex crime is 13.7%. Although the likelihood of relapse of a sex crime is relatively low in comparison to general crime recidivism rates, it is a devastating statistic, nonetheless.

For some time now, the treatment of deviant sexual behavior, like that of drug addiction, has focused on relapse prevention. In order for this treatment to be successful we must first determine the specifics of an individual's sexual interests. Once this is established, we may then determine the high-risk situations that are likely to lead to relapse, and then help the individual in learning to avoid these situations (Fischer, 2000). The development of an assessment tool with adequate reliability and validity in identifying sexual interest could improve the way in which we screen, identify, and treat both convicted and potential sex offenders.

Commonly Used Assessments of Deviant Sexual Interest

In the hopes of identifying potentially problematic behavior, several tools have been developed for assessing an individual's sexual interest. These instruments can fit within four general categories, including clinical interviews, self-report measures, penile plethysmography, and viewing time. Each category has its own share of strengths and weaknesses, and in the field there is ongoing development for each. A review of the

strengths and weaknesses for each category will provide a greater understanding of the very problem proposed by this thesis.

Clinical interviews. Although clinical interviews may have the greatest potential for a detailed analysis of sexual interest, they are also widely criticized because of the difficulty in identifying false statements by participants (Marshall, 1996). Because such interviews involve questions and responses to highly sensitive subject matter, participants may withhold and/or provide false information in an effort to appear *normal* (Quinsey et al.). The potential for such false statements greatly affects the validity of such interviews, resulting in information that undermines a true identification of the respondent's sexual interests.

Self-report measures. Self-report commonly involves the use of a simple rating technique, such as a Likert scale, to identify a participant's own physical or sexual attraction to an identified picture or individual. Some of the proven strengths of self-report are its ability in assessing even very young children, and its stability across cultures (Crosby, 2007). Despite these strengths, self-report also contains several weaknesses. One such is the subjectivity of the participant's perceptions. As a result of this subjectivity, the participant may misunderstand instructions to identify physical attraction—leading to their questioning whether they should base ratings on known societal norms, their own attractions, or possibly a combination of the two (Quinsey, et al.). Another potential weakness of self-report is false ratings. As with clinical interviews, the sensitive nature of sexual interests may incline respondents to present inaccurate ratings in an effort to appear normal.

Penile Plethysmography. Penile Plethysmography (PPG) bases its assessment on the idea that changes in penis volume in response to visual sexual stimuli has a direct correlation with the participant's sexual attraction to those stimuli (Freund & Costell, 1970). PPG is the most scientifically accepted method in the assessment of sexual interest of males (Quinsey & Chaplin, 1988). Further strengthening it's validity, PPG has shown great success in correctly identifying male heterosexuality, homosexuality, and in distinguishing individuals with a history of child molestation from non-offenders (Freund, et al., 1973).

Although PPG has several apparent strengths in comparison to other measures of sexual interest, its use of invasive procedures and pornographic material presents an underlying ethical violation, particularly with its use on adolescents and populations opposed to such methodology (Marshall, 1996). Other identified areas of concern with PPG involve lack of standardization and temporal stability, weak criterion validity and internal consistency, and problems with data formats (Marshal & Fernandez, 2000).

Viewing time. Viewing time (VT) measurements offer a fourth assessment in the identification of deviant sexual behavior. Glasgow (2003) reports that "when people encounter others (and images of others), these are analyzed for the extent to which criteria related to sexual attractiveness are met. If this is to be an efficient process, highly salient features such as gender and age are likely to be analyzed first and if these criteria are met, then more time might be invested in further processing, perhaps in a cascading, 'gated' process unless or until a proscriptive cognition is encountered" (p. 26). Quinsey et al. (1996) further notes that "viewing time may be a measure of sexual interest because it reflects the initial stages of courtship, locating and evaluating an appropriate partner" (p.

342). And as far back as 1942, Rosenzweig (1942) found that viewing time "yields results that are sufficiently consistent (or reliable) and valid to warrant it's use" in assessing sexual interest (p. 158).

Instruments Used with Viewing Time in Assessing Sexual Interest

For the purposes of assessing deviant sexual interests, there are essentially 2 VT instruments being used today. First there's the *Abel Screen* (Abel), which has been marketed under several names, has been used since 1998 in court cases and treatment programs. The second instrument is the *Affinity*, which is relatively new. An analysis of these instruments will bring to light some of the fundamental and critical differences that exist between these two instruments.

Abel. The Abel "uses self report of attraction to 22 categories of possible sexual attraction, as well as a surreptitious measure of sustained attraction to photographs of the stimuli" (Quinsey, p. 352). According to Gress (2005), the viewing time of the Abel has been examined and "proven to be as reliable (alpha coefficient for PPG ranged from 0.66 to 0.97, alpha coefficient for viewing time ranged from 0.86 to 0.90) and valid (a consistent classification of 78% for PPG and 81 % for viewing time) as PPG" (p. 119). Gress also points out that Abel's measurement of viewing time has "the added advantages of being non-intrusive and brief, and which does not use nude slides of children" (p. 119). Furthermore, the viewing time measured by the Abel was able to accurately predict both the age and the gender preferred by the participants (Gress, p. 123).

While Fischer (2000) strongly favors the Abel Screen in its absence of the ethical baggage characteristic of the PPG, he notes two major problems with Abel's analysis of

viewing time. First, "the Abel Screen does not report the underlying raw scores, means, or standard deviations for each subject [and] there is no raw score baseline or interval with which to ground the interpretation of the scores." Second, the "z-scores are not commensurate across clients," two clients may receive the exact same z-score but have very different raw scores in viewing time (Fischer, 2000). Both problems stem from the use of Abel's ipsative scores. Ipsative scores make comparisons using an individual's own data without the use of a norm for comparison. In order to correct this problem standardized scores would have to be calculated based on the tests of large samples. Individuals could then be compared to a larger population and deviation from that population could then be measured.

Affinity. Affinity is a computer program, similar to the Abel, which uses both a self-report measurement as well as a covert measurement of viewing time in order to assess sexual interest (Glasgow, 2003). This self-report measure has the participants rate their sexual preference for both males and females and for individuals in four age groups: adult, juvenile, pre-juvenile and small child. The covert measure of sexual interest is obtained by recording viewing time as the participant rates his/her sexual attraction to images of individuals from the gender and age categories presented in the self-report measurement. Like the Abel screen, Affinity also obtains ipsative scores from the raw data collected for each participant.

Initially, the Affinity program was developed for use with male participants from a mildly mentally retarded population. Currently, Affinity 2.5 is being used to assess sexual interest with adult male sex offenders, and it may be used for research purposes

with adult male non-offenders, juvenile males, and adult and juvenile female populations (Glasgow, 2003).

Weakness of Ipsative Scoring

Affinity data is typically analyzed using ipsative scores. The term *ipsative* refers to data that is the result of using an individual's own information as the basis by which a measurement is made (Encarta, n.d.). Ipsative data represent intra-individual variance only. Johnson et al. (1988) explains that ipsative scores for an individual will always add up to the same constant. An example of this is to convert raw time into percentage of viewing time. All participant data will add up to the same total (100%). This artifact results in what is potentially its greatest weakness. The argument has been presented that "individual differences in ipsative measurements have little meaning because there is not a single scale for all individuals" (Gilford, p. 528). In the case of viewing time percentages, although all scores add up to a constant, this data is the result of raw scores that may look very different from one participant to the next. Some participants may take their time evaluating images, while other may evaluate them very quickly. In the absence of a relevant norm for comparison we are forced to make comparisons within an individual's own pattern of scores- a procedure that can be considered questionable. Unfortunately, this is the nature of VT, there is no way around it. The common approach to VT data uses a deviance differential that fails to establish a norm. Fischer (2006) proposes a second approach that allows ipsative scores to be normed, possibly overcoming their inherent weakness. This approach institutes the use of a Chi-square goodness-of-fit procedure. Further analysis and comparison of these two approaches (deviance differential and Chi-square) is necessary.

Approaches to Deviance Assessment

Viewing time of Affinity profiles have been assessed for deviance using two methods. The first approach uses a deviance differential. This approach is not only commonly used with Affinity, but with other instruments that assess sexual interest such as Penile Plethysmography. Fischer's use of a Chi-square procedure is a second approach that is distinct as a result of its identification and use of a norm referenced profile. The use of Chi-square procedures is new to both its use with Affinity and assessing sexual interest in general.

Deviance differential (DD). Deviance Differential is a simple calculation used to determine the difference between two given scores within an individual. It may be used to maximize differences in the presence of multiple points of data. DD has been used to assess standardized data of sexual interest obtained from PPG data (Harris, et al., 1992). Harris et al. converted raw data to percentages for standardization purposes and deviance has been calculated using only the most outlying scores. Although DD has been used in the determination of sexual interest its relative efficiency compared to other deviance assessments remains to be seen.

Fischer's approach of Chi-square on ipsative scores (CS). Fischer (2004, 2006, 2007) has proposed that the Chi-square goodness-of-fit test may be the most appropriate approach to an assessment of sexual deviance when applied to ipsative scores. Fischer's approach of Chi-square would allow an observed pattern of ipsative scores obtained for an individual to be compared with the expected pattern obtained from an identified non-offender sample. This would resolve the problems that Fischer believes are inherent in misinterpreting ipsative scores, inferring deviance using one's own scores, by comparing

the deviation of an observed ipsative pattern from an expected pattern. This approach has yet to be evaluated against other deviance assessments such as deviance differential.

Receiver Operating Characteristic Curve to Compare DD and CS

An evaluation and comparison of the effectiveness of both DD and CS may be done using a Receiver Operating Characteristic (ROC) curve analysis. ROC is a procedure for discriminating between examples of two categories or populations based on inputted data. An ROC curve would be generated for each, identifying their distinct efficiency, and a direct comparison is then made between the two results.

The ROC curve was first used during World War II for the analysis of radar signals. A good example of its initial need comes from the devastation of the Pearl Harbor attack and the beginning of US involvement in WWII (Green, 1966). The devastation of this attack could have been avoided if the radar could have distinguished between a formation of United States aircraft and a squadron of Japanese fighter planes. ROC is a statistical analysis that was initially designed to answer slight radar discrepancies between different aircraft, accurately detecting Japanese aircraft from radar signals (Hopley & Schalkwyk, 2007). Since its invention for military purposes ROC has been widely used in medicine. By inputting the number and severity of symptoms, medical researchers have used it to determine the likelihood of a certain diagnosis (Hopley & Schalkwyk, 2007). Receiver Operating Characteristic curves are gaining widespread use in other sciences as well.

In the field of risk assessment and specifically recidivism, ROC has been used to meet the need of improved "accuracy of decisions by reporting effect size in terms of area under the receiver operating characteristic" (Rice & Harris, p. 615). In other words,

The ROC makes a judgment, based of assessment data, whether an individual is more likely to be a member of the population that re-offends or the population that does not offend again. According to Hanley & McNeil (1982) "ROC curves are being used to judge the discrimination ability of various statistical methods," (p. 29) such as Chi-square and deviance differential. Hopley & Schalkwyk (2007) simply state that "all an ROC curve is, is an exploration of what happens to [true positives] and [false positives] as we vary the position of our arbitrary test threshold." Once we identify the most efficient ratio of true positives to false positives we have our cut score.

Method

Participants

Participants served as subjects in a prior study, and data for this study was obtained from that existing data set (Glasgow & Croxen, 2003). A total of 53 adult males with a minimum age of 18 and IQ scores that fell within the normal range participated in this data set. All participants were located in Great Britain and varied in their ethnic backgrounds. Twenty-seven of the 53 participants were convicted pedophiles that were incarcerated and were receiving treatment at the time in a mental institution. The other 26 participants were non-offenders with reportedly non-pedophilic sexual interests.

Measures

Affinity is a computer program developed to assess sexual interest. The purpose of the test is to discover the participants' sexual attraction in the categories of gender (male or female) and approximate age (adult, juvenile, pre-juvenile, and small child).

The Affinity assessment is comprised of two primary tasks. The first task is for the individual to rank order their sexual attraction to each of the 8 categories that result from combining all possibilities of age and gender. To begin the ranking task the participants are presented with a screen that has eight images on it. The images are simple stencils representing individuals of both sexes at four different developmental stages. The four stages represented are adult, juvenile, pre-juvenile, and small child. The participant is instructed to select the one image of the eight that represents the age and gender of people they are most sexually attracted to. Once an image has been selected it is removed from the screen and they again, in the same manner select from the remaining images. When the participant states they are no longer sexually attracted to any of the remaining

images they are instructed to indicate that none of the prototypes are sexually attractive. They are then instructed to select the image that is most representative of the age and gender of people that they are least sexually attracted to. They do this until all images have been selected.

The second task is a rating task. The rating task begins with a practice image that exemplifies the instructions and familiarizes the participant with the rating procedures. The images in this part of the test are actual photographs of individuals that represent one of the two genders and one of the four ages. The photographs are simple, lacking detail and other components that may distract the participant and affect the viewing time. Images are presented one at a time and they are displayed in a random order. Under each image is a scale entitled 'Sexually attractive to you.' The scale has 15 possible ratings that range from -7 to 7 with a neutral position in the middle, although the point system is not identified on the screen. A score of -7 represents *extremely sexually un-attractive* while a score of 7 represents *extremely sexually attractive*. A neutral score would mean that the rater is not sexually attracted or un-attracted to the individual in the image.

The Affinity yields both an overt and a covert measure. Performing the rating for each image is the overt measurement of the Affinity; the participants know that data is being collected as they rate each image to the 15-point scale. During this rating task the participants are also having their viewing time recorded. This is a covert measure, the participants are unknowingly being timed as they identify how sexually attractive each image is to them. Although the differences in viewing time are often fractions of a second, these differences have been found to be reliable and valid. This covert measure

can be used to confirm or falsify the participants self reported sexual interest. The data collected with viewing time also has the potential for a variety of statistical analyses.

Procedure

Viewing time conversion to ipsative scores. Since viewing time raw scores obtained from Affinity are unique to each individual and vary according to the speed with which each individual is able to visually and mentally evaluate the images, the scores must be ipsatized. This occurs when data that is unique to each individual is forced into a score that is the same for all individuals. In order to do this with Affinity data the raw VT data will be converted into a percentage of time spent rating each category with total VT equal to 1.0. No matter how slowly or quickly each participant rated Affinity's images, their times will be based on the same scale, a percentage.

Calculation of deviance differential. For the purposes of identifying a typical deviance differential (DD), the eight categories of Affinity will be separated into 4 *normal* and 4 *deviant* categories. Both heterosexuality and homosexuality are considered normal and previous studies of sexual interest have identified both adult and juvenile categories as normal areas for adult sexual interest (Crosby, 2007). The 4 normal categories include *adult female, juvenile female, adult male*, and *juvenile male.* The 4 deviant categories include *pre-juvenile female, small child female, pre-juvenile male,* and *small child male*. A DD is calculated for each participant by subtracting the percentage of viewing time in the single highest deviant category from the percentage of viewing time in the single highest normal category.

Calculation of Fischer's approach of Chi-square. Fischer's approach of Chi-square (CS) analysis of VT data must also first convert raw data into percentages for

comparison. To overcome potential problems associated with the ipsative scores from Affinity, Fischer uses the Chi-square formula $\Sigma(O-E)^2 / E$ to compare each participant's observed pattern to an empirically derived expected pattern. The resulting scores represent the participant's deviation in pattern of sexual interest from the expected pattern of sexual interests.

Data Analysis

A separate ROC analysis has been performed on both the DD scores and the CS scores. For purposes of the ROC a constant was added to all DD scores to eliminate any negative numbers, otherwise, both approaches to the ROC process are identical. Once a range of scores was identified, threshold ranges or *bins* were created in order to categorize scores and separate offenders and non-offenders into two independent population curves regardless of scores. The bins represent structured ranges of scores that will be placeholders for each Chi-square or deviance differential score. An example of this would be to divide a town's population by age, creating bins such as 0-10 years old, 11-20 years old, 21-30 years old, etc. and then tallying the number of boys and the number of girls in each range to create two distinct population curves. Any overlap in the two curves resulted in Type-I and Type-II errors.

Graphing the non-offender rate and the convicted pedophile rate on opposing axes created the ROC curve. The resulting slopes are positive and the area under the curve (AUC) represents the distinctiveness of the two curves. Simply put, the closer the AUC is to a straight line (0.5) the more overlap and less distinguishing the curves are from each other. The closer the AUC is to 1.0 the less overlap and more distinguishing the curves are from each other. Thus, the AUC of the ROC has determined the effectiveness of the

two approaches to identifying deviance (deviance differential vs. Chi-square). The approach with the greater AUC is more effective at correctly identifying participants from the known offender sample and participants from the known non-offender sample.

Results

Fischer's Approach of Chi-square

The following results were obtained using a Receiver Operating Characteristic (ROC) curve to analyze participants' data via Fischer's approach of Chi-square (CS). Chi-square scores were divided into 14 bins with the threshold ranges identified in Table 1. These threshold ranges resulted in the distribution curves found in Figure 1, with the vertical axis representing the number of participants, and the horizontal axis representing the bins (threshold ranges) of Chi-square scores. The blue curve represents non-offender participants, while the purple curve represents convicted pedophile participants. The cut score (where the two curves overlapped at their single greatest point) is located at the point where bin 3 and bin 4 meet and is identified as 17.067 on the distribution curves. Participants with Chi-square scores lower than this cut score were identified as non-offenders and separated into two groups, true non-offenders and convicted pedophiles falsely identified. Likewise, participants with Chi-square scores higher than the cut score were identified as convicted pedophiles and separated into two groups, true convicted pedophiles and falsely identified non-offenders. Table 2 distinguishes these four groups.

This ROC analysis of CS on Affinity profiles resulted in four Type I and four Type II errors. Figure 2 represents the ROC curve with the non-offender rate represented on the vertical axis and the convicted pedophile rate represented on the horizontal axis, resulting in diminished AUC as the two groups overlap. The area under the curve (AUC) is calculated by adding up all of the discrete trapezoids that make up the ROC area under the curve, resulting in an AUC of 0.876781.

Table 1. ROC threshold ranges generated by the variance of Chi-square scores using Fischer's approach.

Threshold Number	Range
1	0 - <5.067
2	5.067 - <11.067
3	11.067 - <17.067
4	17.067 - <23.067
5	23.067 - <29.067
6	29.067 - <35.068
7	35.067 - <41.069
8	41.067 - <47.070
9	47.067 - <53.071
10	53.067 - <59.072
11	59.067 - <65.073
12	65.067 - <71.074
13	71.067 - <77.075
14	77.067 - <83.076

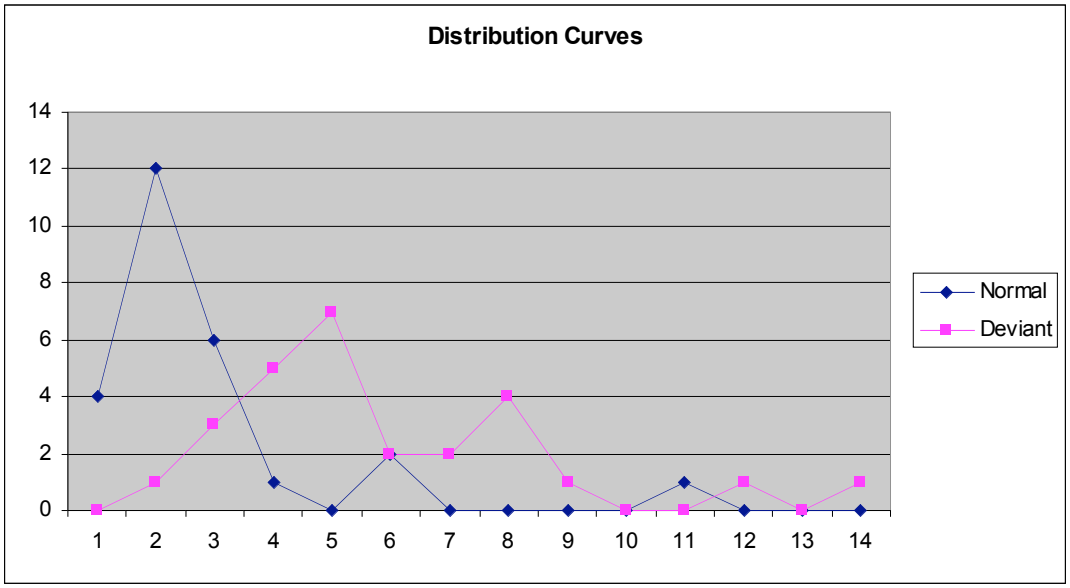

Figure 1. Distribution curves representing both the convicted pedophile (pink) and non offender (blue) populations using Chi-square data.

Table 2. This table represents the number of correctly identified individuals as well as Type I and Type II errors generated using Fischer's approach of Chi-square.

		Observed Non-offender	Observed Convicted Pedophile
Actual	Non-offender	22	4
Actual	Convicted Pedophile	4	23

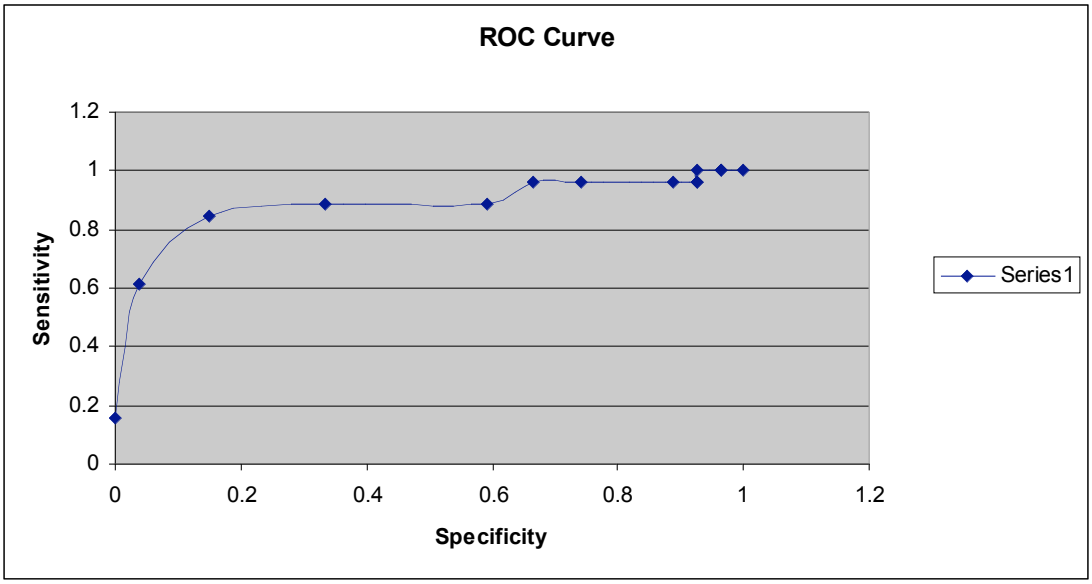

Figure 2. AUC generated using data from Fischer's approach.

Deviance Differential

An analysis using a Receiver Operating Characteristic (ROC) curve of a deviance differential (DD) on participant profiles from our data set was executed with beneficial results. DD scores were divided into 10 bins with threshold ranges identified in Table 3. These threshold ranges resulted in the distribution curves found in Figure 3, again, with the vertical axis representing the number of participants, and the horizontal axis representing the bins (threshold ranges) of deviance differential scores. The blue curve represents non-offender participants while the purple curve represents convicted pedophile participants. The cut score, obtained where the two curves overlap at their single greatest point, is located at the point where bins 5 and 6 meet; this is identified as .099 on the distribution curves. Participants with DD scores lower than this cut score were identified as convicted pedophiles and separated into two groups, *true convicted*

pedophiles and *falsely identified non-offenders*. Likewise, participants with DD scores higher than the cut score on the distribution curves were identified as non-offenders and separated into two groups, *true non-offenders* and *falsely identified pedophile convicts*. Table 4 distinguishes these four groups.

Table 3. ROC threshold ranges generated by the variance of deviance differential scores using Fischer's approach.

Threshold Number	Range
1	-0.15 - -0.100001
2	-0.1 - -0.050001
3	-0.05 - <-9.99999999987122E-07
4	1.10995197408204E-17 - 0.049999
5	0.05 - 0.099999
6	0.1 - 0.149999
7	0.15 - 0.199999
8	0.2 - 0.249999
9	0.25 - 0.299999
10	0.3 - 0.349999

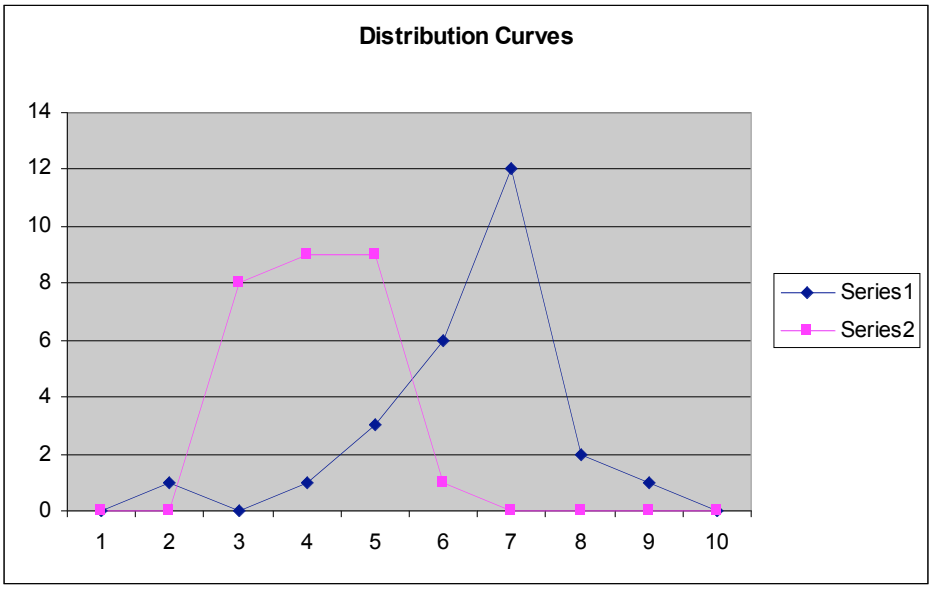

Figure 3. Distribution curves representing both the convicted pedophile (pink) and non-offender (blue) populations using deviance differential data.

Table 4. This table represents the number of correctly identified individuals as well as Type I and Type II errors generated using the deviance differential approach.

		Observed Convicted Pedophile	Observed Non-offender
Actual	Non-offender	5	21
Actual	Convicted Pedophile	26	1

This ROC analysis of DD procedures on Affinity profiles resulted in five Type I errors and one Type II error. The following graph represents the ROC curve, with the vertical axis representing the convicted pedophile rate and the horizontal axis representing the non-offender rate, resulting in a diminished AUC where the two groups overlap. The area under the curve (AUC) is calculated by adding together all of the

discrete trapezoids that make up the ROC area under the curve, resulting in an AUC of 0.913105.

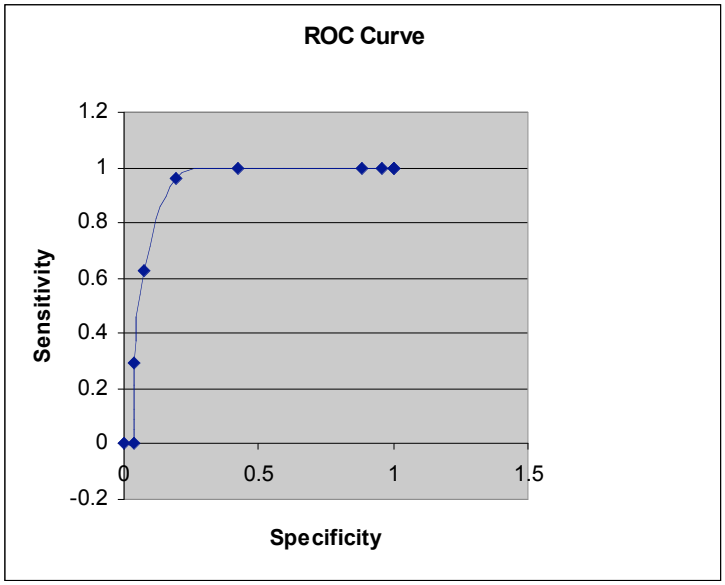

Figure 4. AUC generated using data from the deviance differential approach.

Summary

A Receiver Operating Characteristic (ROC) curve was used to test two scoring procedures using Affinity viewing time data. The results of this ROC analysis were gathered to determine the more effective of two approaches for screening non-offenders from known pedophiles: Fischer's approach of Chi-square (CS) or the commonly used deviance differential (DD).

The area under the curve (AUC) generated by the ROC analysis identifies the efficiency by which each method, CS and DD, accurately predicted and separated non-offenders and convicted pedophiles into two distinct groups. The higher the AUC, the

more efficient the distinction. The CS approach resulted in an AUC of 0.876781, while the DD approach resulted in an AUC of 0.913105.

Based on this ROC analysis, both procedures result in very similar efficiencies with the commonly used DD producing slightly greater efficiency at distinguishing between non-offenders and convicted pedophiles using Affinity 1.0 viewing time data. The difference between the two approaches AUC's comes out to .036324 and may not equate to a significant difference. Although CS shows a near equally high degree of efficiency to DD, it also results in a combined total of eight Type I and Type II errors. The DD method resulted in a combined total of six Type I and Type II errors (2 fewer than CS), producing a larger AUC and a slightly greater efficient rate of differentiation.

Discussion

The premise of this study has been to compare Fischer's approach of Chi-square (CS) to the more traditional deviance differential (DD) for the assessment of sexual interest. Others have previously identified strengths and weaknesses of deviance differential, so a majority of this discussion will focus on this studies implications for CS procedure.

Fischer's Application of Chi-square to Affinity Data Analysis

This study has identified both problems and potential strengths surrounding the use of CS procedure with sexual interest viewing time data. Further research is needed in order to substantiate or rule out these implications.

Problems. One possible problem with the CS analysis of Affinity profiles is that it includes data from each of the eight gender/age categories in its analysis of Affinity viewing time (VT). VT data collected from categories that are not significant (that do not manifest deviance) within the participant's own observed pattern, nor within a comparison to the expected pattern, may weaken the single score generated using the CS formula. An example of this would be that a pedophile interested in small child boys may spend lower then normal time viewing slides from other child categories; these scores are not significant because they are not where his attraction lies but they will weaken the overall Chi-square score. VT scores from these categories would seem to be unnecessary. Significant data may be identified by categories with high rates of VT (regardless of their deviation from the expected pattern) and by categories with high deviation from the expected pattern. Data that may not be significant, which is eliminated using the deviance differential approach, would come from those categories that have recorded VT data that

is not significant in identifying there sexual interests exist or where expected sexual interest does not exist. Further analysis of Chi-square's use of the data from each category may help to strengthen Affinity VT data, and may even lead to new norm-referenced approaches to this technology.

A second problem with Fischer's approach may stem from unfounded assumptions. When using a norm-referenced standard to make a comparison it is assumed that there is far greater deviation between groups then there is within groups. This would result when individuals within the norm referenced group are very similar to each other and to the norm, while individuals from other groups are very dissimilar to this same norm. Is this true for Affinity VT patterns? The assumption made in Fischer's approach is that there is very little variance among scores obtained from the non-offender population, while scores obtained from the convicted pedophile population would have strong variance from the entire non-offender population. However, participants from the non-offender population could potentially deviate from the normal pattern with their own hyper-normal pattern; resulting in a Chi-square score that would identify them with the other population. This presents a potential problem to norm referencing VT data as it is currently collected. Further analysis may help discern the similarity and dissimilarity within and between groups categorized by their sexual attraction.

Potential strengths. Fischer's approach of Chi-square resulted in greater equity between the generated ROC distributions. In other words, Type I and Type II errors were consistent between the populations, with each resulting in 4 errors. Although the DD resulted in fewer total errors, these errors were not evenly distributed between the two populations. The DD approach seems to have a tendency to identify non-offenders as

pedophiles. US courts may view this as a significant flaw, one that threatens fairness and equity. Such a flaw is contrary to the ethical aims of the legal system, and may threaten its acceptance as an impartial and judicial instrument. This flaw does not seem to infect the data generated by CS, resulting in a seemingly fair, equitable and impartial approach to the analysis of Affinity data. Given the relatively small sample size of this study, this potential strength is tentative and will require further examination.

A second strength of Fischer's approach also relies on a comparison of its errors with those generated using the deviance differential approach. Although CS resulted in more total errors in comparison to the DD approach, only 4 of its 8 total errors were Type I (those identifying innocent people of wrong doing). The DD approach produced 5 Type I errors, identifying more non-offenders as being part of the convicted pedophile population. This aspect of CS would seem to put it in the graces of legal ethics, whereas a DD approach would naturally elicit red flags. In the US, society confirms legal ethics by tending to react more strongly and negatively to Type 1 errors. Although Fischer's approach resulted in a greater number of total errors, these results may be more acceptable to US or similarly-minded societies due to this slight reduction in Type I errors. On the other hand, the DD approach misidentified only one individual among a total of 27 convicted pedophiles; a statistic that should not be overlooked.

Thirdly, CS analysis of Affinity profiles has the potential of becoming a more proficient diagnostic instrument. Because data is collected in each category and compared to the norm that is expected in that category, the raw data can be used to identify specific categories of interest and possible diagnosis. The DD approach fails in this respect; although it may have strength as a screening tool it lacks the data needed for diagnosis.

However, identifying CS analysis of Affinity profiles as an efficient diagnostic instrument remains unproven and will require further studies and data collection.

Utility of ROC with Affinity Data

Through this study certain implications surrounding the use of ROC on both Affinity data and Chi-square generated profiles have been identified.

Challenges in applying ROC to Chi-square analysis. A Receiver Operating Characteristic (ROC) curve may be a good approach for discerning the overall efficiency of CS method and for comparing this method to others; however, it fails at evaluating other possible strengths to this approach. Possible strengths include its ability to identify specific categories of deviance as well as its use of a norm referenced pattern for identifying an individual's deviation in each category. Further analysis using multiple methods may be beneficial in validating these and other possible strengths.

Implications for further use of Affinity. This ROC analysis does result in one implication for Affinity as an assessment of pedophilic interest. Because CS and the DD both resulted in high efficiency (high AUC), this ROC analysis has validated that VT data obtained from Affinity profiles is strongly connected to an individual's sexual interests. The results of this ROC appear to have helped to identify Affinity as a valid instrument in identifying an individual's true sexual preference to the eight categories measured. That is, Affinity seems to very accurately measure what it is reporting to measure.

References

Crosby, D. (2007). *Non-pedophilic heterosexual male response to the Affinity 2.0.* Unpublished doctoral dissertation, Brigham Young University, Provo, UT.

Dailey, T. J. (1999). The Effects of Pedophilia on Children. Retrieved April 7, 2009, from http://www.section21.m6.net/res-effect.php

Encarta dictionary. (n.d.). Retrieved May 23, 2008, from http://encarta.msn.com/encnet/refpages/search.aspx?q=ipsative.

Fawcett, T. (2003). *ROC Graphs: Notes and practical considerations for data mining research.* Retrieved April 17, 2007, from http://home.comcast.net/~tom.fawcett/public_html/papers/ROC101.pdf

Fischer, L. (2000). The Able Screen: A non-intrusive alternative? In R. D. Laws, S. M. Hudson & T. Wards (Eds.), *Remarking relapse prevention with sex offenders: A sourcebook* (pp. 303-318). Thousand Oaks, CA: Sage Publications.

Fischer, L., & Morgan, D. (2006, September). Norm-referenced clinical decision-making with Affinity viewing-time results. In L. Fischer (Chair), Current Research and Clinical Application of Affinity Viewing-Time Scores. Symposium conducted at the 25th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Chicago, IL.

Fischer, L., Byrne, P. & Glasgow, D. (2007, October). Real world decision-making using both ipsative and norm-referenced Monarch PPG and Affinity results. Paper presented at the 26th Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Fischer, L. (2004, October). Psychometric issues in viewing-time research, clinical, and forensic applications. In L. Fischer (Chair), Accessible, standardized, viewing-time measurement. Symposium conducted at the 23rd Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Albuquerque, NM.

Freund, K., & Costell, R., (1970). The structure of erotic preference in the nondeviant male. *Behavior Research and Therapy*, *8*, 15-20.

Freund, K., Langevin, R., Cibiri, S., & Zajac, Y. (1973). Heterosexual aversion in homosexual males. *British Journal of Psychiatry*. *122* (567), 163-169.

Gilford, J.P. (1954). *Psychometric Methods*, (2nd ed.). New York: McGraw-Hill.

Glasgow, D. V., & Croxen, J. D. (2003). *Viewing time as a measure of paedophile sexual interest: empirical evidence, the potential of social cognition as a theoretical model, and some possible implications for assessment and intervention.* Manuscript submitted for Publication.

Glasgow, D. V., Osborn, A., & Croxen, J. (2003). An assessment tool for investigating paedophile sexual interest using viewing time: an application of single case methodology. *British Journal of Learning Disabilities*, *31*, 96-102.

Green, D.M., & Swets, J.M. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons Inc.. ISBN 0-471-32420-5.

Gress, C. L. Z. (2005). Viewing time measures and sexual interest: Another piece of the puzzle. *Journal of Sexual Aggression, 11* (2), 117-125.

Hanley, J.A., & McNeil, B.J., The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29-36

Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sex offense recidivism. *Criminal Justice and Behavior, 27*, 6-35.

Hanson, R., & Morton-Bourgon, K. (2005), The Characteristics of Persistent Sexual Offenders: A Meta-Analysis of Recidivism Studies. *Journal of Counseling and Clinical Psychology, 73* (6), 1154-1163.

Harris, G.T., Rice, M. E., Quinsey, V.L., Chaplin, T. C., & Earls, C., (1992). Maximizing the discriminant validity of phallometric assessment data. *Psychological Assessment*, *4* (4), 502-511.

Harris, G.T., Rice, M. E., Quinsey, V.L., & Chaplin, T. C. (1996). Viewing time as a measure of sexual interest among child molesters and normal heterosexual men. *Behavior Research and Therapy, 34* (4), 389-394.

Hopley, L., & Schalkwyk, J. (2007). The magnificent ROC. Retrieved March 24, 2009, from http://www.anaesthetist.com/mnm/stats/roc/Findex.htm

Johnson, C.E., Wood, R., & Blinkhorn, S.F., (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology*, *61*, 153-162.

Laws, D. R. (1989). *Relapse prevention with sex offenders.* New York: Gilford.

Marshall, W.L. (1996). Assessments, treatments, and theorizing about sex offenders: Developments during the past twenty years and future directions. *Criminal Justice and Behavior*, *23*, 162-199.

Marshall, W.L., & Fernandez, Y. (2000). Phallometric testing with sexual offenders: Limits to its value. *Clinical Psychology Review*, *20*, 807-822.

O'Connor, P., (1998). Sex Crimes. Retrieved on April 7, 2009, from http://www.crimemagazine.com/sex_crimes.htm

Quinsey, V., & Chaplin, T. (1988). Penile responses of child molesters and normals to descriptions of encounters with children involving sex and violence. *Journal of Interpersonal Violence*, *3* (3), 259-274.

Quinsey, V.L., Rice, M.E., Grant, T.H., & Reid, K.S. (1993). The phylogenetic and ontogenetic development of sexual age preferences in males: Conceptual and measurement issues. In H.E. Barbaree, W.L. Marshal, & S.M. Hudson (Eds.), *The Juvenile Sex Offender.* New York: Guilford.

Quinsey, V.L., Ketsetzis, M., Earls, C., & Karamanoukian, A., (1996). Viewing time as a measure of sexual interest. *Ethology and Sociobiology, 17*, 341-354.

Rice, M.E., & Harris, G.T., (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior*, *29* (5), 615-620.

Rosenzweig, S., (1942). The photoscope as an object device for evaluating sexual interest. *Psychosomatic Medicine, 4*, 150-158.

Seto, M. C., Harris, G. T., Rice, M. E., Barbaree, H. E., (2003). The screening scale for pedophilic interests predicts recidivism among adult sex offenders with children victims. *Archives of Sexual Behavior*, *31* (5), 455-466.

Zamansky, H. S., (1956). A technique for assessing homosexual tendencies. *Journal of Personality*, *24*, 436-448.