Sampling Extreme Groups Invalidates Research on the Paraphilias: Implications for DSM-5 and Sex Offender Risk Assessments

Corresponding Author: Richard Wollert

Washington State University Vancouver

P.O. Box 61849

Vancouver, WA 98666-1849

e-mail: rwwollert@aol.com

phone: 360.737.7712

fax: 360.993.0106

Elliot Cramer

University of North Carolina at Chapel Hill

Forthcoming, *Behavioral Sciences and the Law, 29;4* (July-August 2011)

Behavioral Sciences and the Law is a publication of Wiley-Blackwell.

Sampling Extreme Groups Invalidates Research on the Paraphilias: Implications for DSM-5 and Sex Offender Risk Assessments

Psychiatrist and DSM-IV Text Editor Michael First has criticized the addition of victim counts to criteria proposed by the Paraphilia Sub-Workgroup for inclusion in DSM-5 because they will increase False Positive diagnoses. Psychologist and Chair of the DSM-5 Paraphilia Sub-Workgroup Ray Blanchard responded by publishing a study of Pedohebephiles and Teleiophiles which seemed to show that victim counts could accurately identify Pedohebephiles who were selected per self-report and phallometric testing. His analysis was flawed because it did not conform to conventional clinical practice and because he sampled groups at opposite ends of the clinical spectrum. In an analysis of his full sample we found the False Positive rate for Pedohebephilia at the recommended victim count selection points was indeed very large. Why? Because data analyses that eliminate intermediate data points will generate inflated estimates of correlation coefficients, base rates, and the discriminative capacity of predictor variables. This principle is also relevant for understanding the flaws in previous research that led Hanson and Bussiere to conclude that sexual recidivism was correlated with "sexual interest in children as measured by phallometric assessment." The credibility of mental health professionals rests on the reliability of their research. Conducting, publishing, and citing research that reflects adequate sampling and cautious diagnostic theorizing is critical for preserving this credibility.

Key words: Pedohebephilia, DSM-5, selective sampling, Bayes's Theorem, sex offender assessment, child pornography.

Sampling Extreme Groups Invalidates Research on the Paraphilias: Implications for DSM-5 and Sex Offender Risk Assessments

By including the diagnosis of Pedophilia among its "mental disorders" the DSM-IV (American Psychiatric Association, 1994) and the DSM-IV-TR (American Psychiatric Association, 2000) enable clinical and forensic mental health professionals to diagnose patients who have an erotic preference for prepubescent children (defined on page 572 of DMS-IV-TR as those "generally age 13 or younger"). Psychologist Ray Blanchard, chair of the DSM-5 Paraphilia Sub-Workgroup, has proposed replacing Pedophilia in the DSM-5 with another diagnosis he refers to as "Pedohebephilia." Under this alternative, according to Blanchard and his colleagues from Canada's Kurt Freund Laboratory of the Centre for Addiction and Mental Health (Blanchard, Lykins, Wherrett et al., 2009), patients with an erotic preference for prepubescent children (those younger than 11 years old) would be one type of Pedohebephilia, patients preferring pubescent children (those ranging in age from 11 to 14) would be another, and patients attracted to both age groups would be still another.

The Sexual and Gender Disorders Work Group of the DSM-5 Task Force recently decided to recommend that Pedohebephilia be substituted for Pedophilia in DSM-5 (http://www.dsm5.org). The diagnostic criteria under consideration are presented in Table 1. As Table 1 indicates, Pedohebephilia would be diagnosed not only on the traditional paraphilic grounds of deviant fantasies and urges [A. and A.(1)] and distress or impairment [B.(1)], but also on the less traditional grounds of behavioral misconduct

[A.], victim counts [B.(2)(a) and B.(2)(b)], use of child pornography [B.(3)], and relatively strong sexual arousal to children [A.(2)] as assessed by self-report, laboratory testing like the penile plethysmograph (PPG), or behavioral analysis (Blanchard, 2009, October).

Insert Table 1 about here

This recommendation has elicited a number of criticisms from both psychologists and psychiatrists. Psychologist Karen Franklin (2010), in particular, has inventoried these criticisms and concluded that "expanding the definition of pedophilia – a diagnosis with already poor interrater reliability – into a broader definition of pedohebephilia has the potential to dramatically increase the scope and power of the sex offender civil commitment industry" in a way that "will invite … expert witnesses (to) purposely distort their testimony to achieve desired ends" (pp. 764-765). Psychiatrist Richard Green, instrumental in removing Homosexuality from the DSM (Green, 1972), has pointed out that "the age of legal consent is 14" in 11 European countries (Green, 2010a, p. 585) and that "decreeing … that a 19 year-old who prefers sex with a 14 year-old has a mental disorder will not enhance psychiatry's credibility" as a science (Green, 2010b, p. 8).

Responding to such criticisms, advocates for adding Pedohebephilia to the DSM have argued that it would foster important research on the paraphilias (Blanchard, 2009, October) while expanding the DSM's coverage to include "a sizable proportion of those men whose strongest sexual feelings are for physically immature persons" (Blanchard et al., 2009, p. 335).

In spite of these current differences, those on both sides of this argument would probably agree that Pedohebephilia should be considered for inclusion in DSM-5 if the criteria for diagnosing it included the cognitive-emotional elements of a true mental disorder and could be reliably used by clinicians.

Psychiatrist and DSM-IV Text Editor Michael First has challenged the inclusion of victim counts among the criteria for Pedohebephilia on the grounds they will increase False Positive diagnoses "by allowing a paraphilia diagnosis (to be made) simply by exceeding an arbitrary number of sexual offenses" (p. 1239). Specifically, he asserted (p. 1241) that "no empirical data were cited on the website" of the DSM-5 Paraphilia Sub-Workgroup "to explain how or why … specific thresholds" such as "three known offenses" against minors were recommended as a criterion for the proposed diagnosis. Furthermore, he observed that only "a single study by Blanchard, Klassen, Dickey, Kuban, & Blak (2001)" served "as the justification for adopting" victim counts. This study was inadequate, according to First, because it did not include an "ROC analysis" indicating that "three or more victims yielded the best balance of false negatives v. false positives based on some gold standard for a diagnosis of Pedophilia, such as the individual admitting to a preferential sexual attraction to children." [1]

[1] In the simplest diagnostic model a test, indicator, criterion, or sign that is thought to predict a condition of interest will either be present or absent for a given patient (Quinsey, Rice, Harris, & Cormier, 1998, p. 50). The patient will also either have the condition or won't. Four outcomes are possible in this model. One is that the patient shows the sign and has the condition. Such an outcome is called a "True Positive." Another possible outcome is that the patient shows the sign but doesn't have the

Blanchard (2010) responded to First's criticisms by offering "an empirical analysis along the lines he suggests" (p. 1246). He selected 181 "Pedohebephiles" (sexually-troubled patients who showed a sexual preference for minors on both self-report and phallometric measures) and 817 "Teleiophiles" (sexually-troubled patients who showed a preference for adults on both measures) from a total of 2,715 men seen at the Kurt Freund Laboratory who were agreeable to having their phallometric data used for research. Then he compiled three frequency tables and completed three ROC analyses based on these tables. In his first analysis he calculated the extent to which the total number of minor victims (e.g., 0, 1, 2, 3, 4, 5, >5) correctly differentiated the target

---

condition. This is a "False Positive." Still another is that the patient doesn't show the sign and doesn't have the condition. This is a "True Negative." Finally, a patient who doesn't show the sign may have the condition. This is a "False Negative."

An ROC analysis reflects the sign's capacity to accurately identify both True Positives and True Negatives and can range from 0 to 1. An ROC analysis for a sign that yields a True Positive or True Negative result close to 1 means, as First states in this passage, that misclassifications in the form of False Negatives and False Positives are minimized.

More complex diagnostic models may be construed. One conceptualizes a condition along a continuum of severity in which patients have high dysfunctionality (H), moderate dysfunctionality (M), or low dysfunctionality (L). If H is the target diagnostic condition, it subsumes all True Positive and False Negative outcomes. All False Positive and True Negative outcomes in this model are calculated by combining data for the M and L groups.

group of Pedohebephiles from Teleiophiles. In his second he carried out a similar analysis but first removed 120 patients who had been charged with, or admitted to, the use of child pornography. In his third analysis he "created a new variable, the *victim equivalency count*, by adding two 'victims' to the total score" for any patient who admitted to using child pornography. Blanchard also reported Sensitivity and Specificity when patients were classified as Pedohebephiles on the basis of having three or more victims. [2]

---

[2]Diagnostic studies typically categorize many patients subject to error, so there will be a number of patients in the True Positive and True Negative categories and a number of patients in the False Negative and False Positive categories. Sensitivity (True Positive rate) is the result of dividing the number of True Positives by the sum of True Positives and False Negatives. A large Sensitivity rating indicates that the presence of a criterion successfully identifies the presence of a disorder. Specificity (True Negative rate) is the result of dividing the number of True Negatives by the sum of True Negatives and False Positives. A large Specificity rating indicates that the absence of a criterion successfully identifies the absence of a disorder. The complement of Specificity (1-Specificity) indicates the extent to which a sign misidentifies the presence of the target condition. An ROC analysis of a dichotomous diagnostic model (present or absent) compares a single Sensitivity rating with the complement of a single Specificity rating (Biggerstaff, 2000).

Diagnostic models that include several levels of signs are frequently used, e.g., tests that diagnose the chances of violent recidivism (Quinsey et al., 1997). An ROC

ROC areas for Blanchard's analyses were .72, .79, and .83. Sensitivities were .54, .66, and .66. Specificities were .91, .91, and .89. Base rates for Pedophebephilia, not reported but apparent from his tabled data, were .18, .12, and .18. Blanchard's tables also did not report the total number of patients in each victim count category, the classification error rates for individual categories, or the classification error rates for a selection criterion of 3 or more victims. We have included this information and the frequency counts presented in Blanchard's first and third tables in Tables 2 and 3.

Insert Tables 2 and 3 about here

On the basis of his ROC analyses and values of Sensitivity and Specificity Blanchard concluded that either victim counts or victim equivalency counts could accurately identify Pedohebephiles. He also claimed that this conclusion applied to "other paraphilic disorders." Although he did not identify these disorders, he presumably meant the specific disorders currently included in the DSM (Exhibitionism, Frotteurism, Voyeurism, Sexual Sadism, and Pedophilia) and two other disorders (Pedohebephilia and Paraphilic Coercive Disorder) that the Paraphilia Sub-Workgroup has recommended for inclusion in the DSM.

Clinical and Statistical Flaws With Sampling Extreme Groups

Assertions

---

analysis for one of these models compares several Sensitivity ratings with the complements of several Specificity ratings.

Some readers might agree with Blanchard's conclusions, particularly after considering his ROC results in light of the somewhat optimistic error rate data presented here in Tables 2 and 3.

This would be a mistake for clinical and statistical reasons.

Clinically, it would be unusual during a diagnostic interview in either an outpatient or forensic setting for a clinician to consider whether a patient with a given set of symptoms might fit a particular diagnostic category by considering the patient's status on an alternative diagnostic category with a much different set of symptoms. More likely, the diagnostician would first implicitly consider all of the diagnostic categories that might be encountered in the population served by the clinic. Unlikely options would then be eliminated on the basis of case information until a point was reached where it was possible to make a differential diagnosis.

Blanchard used a different approach in the selection of his sample so that data were analyzed for subjects from different ends of the sexual dysfunction continuum. Data for 1,717 of his patients in the mid-range of this continuum were therefore not analyzed. The difference between typical diagnostic procedures and Blanchard's research procedures would therefore seem to limit the extent to which his results might be generalized to either outpatient or forensic settings.

Statistically, sampling extreme ends of a clinical distribution will:

1. Overestimate the "base rate" P;
2. Leave Sensitivity, the number of True Positives divided by the sum of True Positives and False Negatives, unchanged;

3. Overestimate Specificity, the number of True Negatives divided by the sum of True Negatives and False Positives ;

4. Overestimate the Likelihood Ratio $\left(\mathrm{LR} = \frac{Sensitivity}{1-Specificity}\right)$;

5. Overestimate diagnostic efficiency as reflected in $P(C|S)$, the probability that condition C is present given that diagnostic sign S is present.

Proofs

The following proofs confirm these assertions.

Assume a sexual dysfunction continuum is divided into three groups – High (H), Medium (M) and Low (L) – where the numbers of patients in each group are *h*, *m*, and *l*. Further assume that some of the patients in each group are positive for sign *S* that is thought to identify those in group H and that some patients show only the absence of *S*.

If H is the target diagnostic group, the base rate is $\frac{h}{h+m+l}$. If the M group is removed, the base rate is $\frac{h}{h+l}$.

1: The base rate $\mathrm{P}= \frac{h}{h+l}$ will always be greater than the base rate $\frac{h}{h+m+l}$.

Sensitivity is defined as $\frac{number\ of\ True\ Positives}{number\ of\ True\ Positives+number\ of\ False\ Negatives}$.

2: Sensitivity is unchanged by removing the M group since both the True Positives and False Negatives are in the one H group (also see the first paragraph of footnote 1).

Specificity is defined as $\frac{number\ of\ True\ Negatives}{number\ of\ True\ Negatives+number\ of\ False\ Positives}$.

Unlike Sensitivity, the value of Specificity will vary, depending upon the definition of the non-target group (see the third paragraph of footnote 1). For a well behaved scale, the proportion of True Negatives will be larger in the L group than in the M group, that is, $\frac{l_t}{l_t+l_f} > \frac{m_t}{m_t+m_f}$ where l_t and m_t are the numbers of True Negatives while l_f and m_f are the numbers of False Positives for the two non-target groups. These proportions – $\frac{l_t}{l_t+l_f}$ or $\frac{m_t}{m_t+m_f}$ – are the Specificities if either L or M is the non-target group. If M and L together represents the non-target group, the Specificity of this combined group is $\frac{m_t+l_t}{m_t+l_t+m_f+l_f}$. We can then show (see the Appendix) that the Specificity increases when the M group is removed, that is,

3: $\frac{m_t+l_t}{m_t+l_t+m_f+l_f} < \frac{l_t}{l_t+l_f}$. If, in a population with L, M, and H subgroups, we estimate the Specificity from a sample with the M group removed, we will then underestimate the Specificity. This is exactly what Blanchard has done.

The Likelihood Ratio is defined as $\mathrm{LR} = \frac{Sensitivity}{1-Specificity}$.

4: Since the Sensitivity is unchanged and the Specificity is increased when the M group is removed (see proofs 1 and 2), the Likelihood Ratio is increased.

The efficiency of a diagnostic sign reflects the probability that a condition such as Pedohebephilia will be present when a patient meets a criterion such as “three or more

victims." This is expressed by Bayes's Theorem[3] as a function of the Likelihood ratio LR and the base rate P.

P(diagnostic condition given a sign) = P(C|S) = $\frac{LR \times \frac{P}{1-P}}{1+(LR \times \frac{P}{1-P})}$

Since both LR and P are overestimated when the M group is removed, the numerator is overestimated, implying that P(C|S) is overestimated. To show this, note that P(C|S) is of the form $\frac{x}{1+x}$. Its reciprocal is therefore

$$\frac{1}{P(C|S)} = \frac{1+x}{x} = 1 + \frac{1}{x}.$$

If x is overestimated, $\frac{1}{x}$ is underestimated implying that $\frac{1}{P(C|S)}$ is underestimated as well.

It follows then that

5: $P(C|S)$, the discriminative capacity of the diagnostic sign, is overestimated.

---

[3] According to Wollert (2007, p. 176), "Bayes's Theorem (Bayes, 1764) is a tool for assessing the probability that a theory (e.g.. that a person with heart disease will die in 5 years) is true when considered in light of the diagnostic accuracy (i.e., LR) of some piece of evidence such as a disease criterion or test score, and what is known about the overall, or base rate, probability of the focal outcome P(D)."

An application of the "odds version" of Bayes's Theorem, discussed in connection with the fifth proof, is included in Wollert, Cramer, Waggoner, Skelton, and Vess (2010).

Example

An example for a small sample of patients may provide an intuitive understanding of these problems. Suppose that an exhaustive clinical cohort of 30 patients includes 5 Pedohebephiles and 25 Non-Pedohebephiles. Three of the Pedohebephiles have high victim counts and 2 have low counts. Eleven of the Non-Pedohebephiles have high counts while 14 have low counts. The full sample base rate for Pedohebephilia would be 17% (5/30=17%), Sensitivity would be .60 (3/5=.60), Specificity would be .56 (14/25=.56), the phi coefficient of association would be .12, and the misdiagnosis rate for Pedohebephilia would be 79% [11/(3+11)=79%].

These are dismal results. Suppose, however, that 5 Teleiophilic patients were selected from the 25 Non-Pedohebephiles because they were the only Non-Pedohebephiles who seemingly preferred adults on self-report and phallometric measures. Further suppose that 1 of the Teleiophiles had a high victim count and 4 had low counts. Now an analysis limited to only the non-randomly selected data for the Teleiophilic and Pedohebephilic groups would yield a base rate for Pedohebephilia of 50% (5/10=50%), Sensitivity would remain the same, Specificity would climb to .80 (4/5=.80), the phi coefficient would be .41, and the misdiagnosis rate for Pedohebephilia would fall to 25% [1/(1+3)=25%].

The second set of results is much more attractive than the first for publication purposes. They may also, unfortunately, be useful in a consequential (Sreenivasan, Frances, & Weinberger, 2010) or pretextual sense (Franklin, 2010) for buttressing inaccurate sex offender risk assessment opinions.

They are nonetheless misleading and should not be relied upon because they are merely artifacts of an overly selective approach to sampling.

A Failed Attempt to Replicate Blanchard's Results with a Full Sample

To evaluate whether Blanchard's results might reflect this problem we asked him for his recidivism data on the patients left out of his analyses. We then compiled frequency tables like Blanchard's except that the data for his eliminated subjects were combined with the data for his Teleiophiles to form a Non-Pedohebephilic group. After this we calculated base rates, Sensitivities, Specificities, error rates, and ROC areas based on the full sample tables.

ROC areas for the full sample analyses were .66, .74, and .77. Sensitivities were .54, .66, and .66. Specificities were .83, .83, and .80. Base rates for Pedohebephilia were .07, .04, and .07. Error rates for misdiagnosing Non-Pedophebephiles as Pedohebephiles when a count of 3 or more victims was used for the purpose of diagnostic selection were 82%, 85%, and 81%.

Tables 4 and 5 present the error rates for the categories considered by Blanchard in his first and third analyses. It is clear from comparing them with Tables 2 and 3 that both Specificity and the base rate for Pedohebephilia decreased when mid-range data that were not previously considered were taken into account. Most importantly, diagnostic error rates are discouragingly large.

Insert Tables 4 and 5 about here

We also compiled a new table for Blanchard's second analysis and obtained results that paralleled the foregoing results. This analysis, which seemed less important

to us than the other two because it eliminated molesters who had also used child pornography, will not be considered further for the sake of brevity.

Discussion

Blanchard's (2010) data initially appeared to refute First's criticism that victim counts will increase the rate with which sex offenders are erroneously classified as Pedohebephiles. The present reanalyses show that First was correct.

Why did this turn out to be the case? The answer rests on two statistical principles. First, as we have shown here, the formulas for the terms in Bayes's Theorem (Donaldson & Wollert, 2008; Waggoner, Wollert, & Cramer, 2008) indicate that eliminating intermediate groups in a Bayesian analysis will render overly optimistic estimates of both the base rate for the criterion variable and the discriminative capacity of the predictor variables. Second, the formula for the correlation coefficient indicates that calculating a single correlation coefficient on data from two groups at extreme ends of the same data set will always increase the size of a correlation coefficient or a phi coefficient over the result that would have been obtained with representative sampling. Specifically, as McCall (1975, pp. 121-132) has explained,

> *r* becomes large when there are many subjects whose *X* and *Y* scores both deviate markedly from their respective means. Consequently, by selecting extreme groups the subjects whose scores would be near the means (which would be located between the extreme groups) are systematically eliminated.

These principles are not only relevant for understanding the flaws in Blanchard's (2010) analysis but also for understanding the flaws in previous diagnostic research. As noted in the introduction, for example, the Sexual and Gender Disorders Work Group of

the DSM-5 Task Force believes that PPG testing is useful for identifying Pedohebephiles while others (e.g., Quinsey et al., 1997) believe it is also useful for identifying sexual recidivists. The DSM-IV-TR, however, describes the PPG as a procedure for which "the reliability and validity … in clinical assessment have not been well-established" (p. 567). One plausible reason for the perceived value of the PPG is that, in a meta-analysis of the correlations between sexual recidivism and PPG patterns based on 7 studies (Barbaree & Marshall, 1988; Gretton, McBride, Hare, O'Shaugnessy, & Kumka, 2001; Malcolm, Andrews, & Quinsey, 1993; Maltezky, 1993; Marshall & Barbaree, 1988; Proulx, Pellerin, Paradis, McKibben, Aubut, & Quimet, 1997; Rice, Quinsey, & Harris, 1991) reportedly including 4,853 patients, Hanson and Bussiere (1998) concluded that "sexual interest in children as measured by phallometric assessment was the single strongest predictor … (.32)" (p. 351) when compared to many other risk factors for sexual recidivism.

Upon reviewing these 7 studies we found that the data set described by Marshall and Barbaree (1988) was the same as that in Barbaree and Marshall (1988). We also found that Maletzky's (1993) article on 4,381 sex offenders did not focus specifically on sexual recidivism, but reported the rates of those who were considered to be treatment failures on the overly broad grounds that they "did not complete all active treatment sessions … reported any instance of overt or covert deviant sexual behavior at the end of active treatment or at any follow-up session … had deviant arousal greater than 20% on the plethysmograph at the end of treatment or at any follow-up session … (or were) recharged with any sexual crime within the duration of the study period" (p. 248). Four of the 5 germane studies (Barbaree & Marshall, 1988; Malcolm et al., 1993; Proulx et al.,

1997; Rice et al., 1991) sampled inadequately small bands of the very heterogeneous population of sex offenders. In contrast, the study that adopted the most comprehensive sampling approach (Gretton et al., 2001) reported that the correlation between PPG arousal and sexual recidivism was nonsignificant [r (184) = .10].

Overall, reading the original research which was the foundation of Hanson and Bussiere's meta-analysis led us to conclude that data from only 507 sex offenders were actually relevant to their meta-analysis and that only 7% to 35% of the entire sex offender population was sampled in any of the relevant cohorts except Gretton's. These findings, in turn, cast grave doubt on the value of using PPG testing to predict sexual recidivism or diagnose any DSM condition.

Hospitals, mental health providers, attorneys and the courts are looking to psychology and psychiatry to provide conceptualizations and risk assessment information that will reassure them in disposing of the criminal and civil commitment challenges that sex offenders represent. It will be self-defeating for individual evaluators and the nationally organized mental health professions to inject "illusions of certainty" (Wollert, 2007; Wollert & Waggoner, 2009) into this atmosphere when what is needed is reliable evidence. Conducting, publishing, and citing only research that reflects adequate sampling is essential for limiting confusion, expanding the range of what we can legitimately claim to know, and maintaining the credibility of the mental health professions.

Consequently, the field studies that are currently being conducted by the American Psychiatric Association to evaluate the reliability of paraphilic diagnoses proposed for inclusion in DSM-5 must reflect representative sampling of sex offenders in

general to be regarded as valid. We fear this will not happen for Pedohebephilia or any of the other proposed paraphilias because http://www.dsm-5.org currently states that "a limited number of standardardized and methodologically strong study designs" are underway at "academic or other large clinical settings with established research infrastructures" and indicates that 16 adult disorders (including Major Depressive Disorder, Schizophrenia, Personality Disorder, Post Traumatic Stress Disorder, and Mild Traumatic Brain Injury) are being studied at these sites. None of these well-established research sites are studying the paraphilias. This raises the possibility that whatever data are collected on the paraphilias may be misleading because of selective sampling or other validity threats. To avoid this we recommend that arrangements be made to rigorously study the paraphilias at a large number of well-established research sites.

Regarding the viability of approving and then applying such proposed diagnoses as Pedohebephilia and Paraphilic Coercive Disorder to real patients, we also suggest on the basis of our experience with the present analysis that psychologists and psychiatrists forego asking themselves "what are the chances I'll be right"? Instead, following the Hippocratic Oath, we recommend they ask "what are the chances I'll be wrong" and "what kinds of harm will occur if I am"?

Empirical research, social policy considerations, and the future of the integrity of the mental health professions all indicate that the most enduring and meaningful diagnostic framework, in the long run, will proceed from pessimism rather than optimism.

Appendix

Proof that removing the M group from a three-group population increases the Specificity.

We must prove that

$$\frac{m_t+l_t}{m_t+l_t+m_f+l_f} < \frac{l_t}{l_t+l_f} \text{ given that}$$

(1) $$\frac{l_t}{l_t+l_f} > \frac{m_t}{m_t+m_f}$$

We can choose positive constants κ and ε relating l_t and l_f to m_t and m_f

so that

$m_t = \kappa(l_t-\varepsilon)$ and $m_t+m_f = \kappa(l_t+l_f)$.

Then

$$\frac{m_t}{m_t+m_f} = \frac{\kappa(l_t-\varepsilon)}{\kappa(l_t+l_f)} = \frac{l_t-\varepsilon}{l_t+l_f} < \frac{l_t}{l_t+l_f} \text{ as required by (1) and}$$

$$\frac{m_t+l_t}{m_t+l_t+m_f+l_f} = \frac{l_t+\kappa(l_t-\varepsilon)}{l_t+l_f+\kappa(l_t+l_f)} = \frac{(1+\kappa)l_t-\kappa\varepsilon}{(1+\kappa)(l_t+l_f)} < \frac{l_t}{l_t+l_f}.$$

Since the term on the left is the Specificity when the combined M and L groups is the non-target group while the term on the right is the Specificity when the L group alone is the non-target group, the Specificity increases when the M group is removed.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*. Washington, D.C.: American Psychiatric Association.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (fourth edition-TR)*. Washington, D.C.: American Psychiatric Association.

Barbaree, H. E. & Marshall, W. L. (1988). Deviant sexual arousal, offense history, and demographic variables as predictors of reoffense among child molesters. *Behavioral Sciences & the Law, 6*, 267-280.

Bayes, T. (1764). An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370-418.

Biggerstaff, B. J. (2000). Comparing diagnostic tests: A simple graphic using likelihood ratios. Statistics in Medicine, 19, 649-663.

Blanchard, R. (2009, October). *DSM-V Options: Paraphilias and paraphilic disorders, Pedohebephilic Disorder, and Transvestic Disorder*. Paper presented at the meeting of the Association for the Treatment of Sexual Abusers, Dallas, Texas).

Blanchard, R., Lykins, A., Wherrett, D., Kuban, M., Cantor, J., Blak, T., Dickey, R., & Klassen, P. (2009). Pedophilia, Hebephilia, and the DSM-V. *Archives of Sexual Behavior*, 38, 335-350.

Blanchard, R. (2010). The specificity of victim count as a diagnostic indicator of Pedohebephilia [Letter to the Editor]. *Archives of Sexual Behavior, 39,* 1245-1252.

Blanchard, R., Klassen, P., Dickey, R., Kuban, M., & Blak, T. (2001). Sensitivity and specificity of the phallometric test for pedophilia in nonadmitting sex offenders. *Psychological Assessment, 13,* 118-126.

Donaldson, T. & Wollert, R. (2008). A mathematical proof and example that Bayes's Theorem is fundamental to actuarial estimates of sexual recidivism risk. *Sexual Abuse: A Journal of Research and Treatment*, *20*, 206-217.

First, M. B. (2010). DSM-5 proposal for paraphilias: Suggestions for reducing false positives related to use of behavioral manifestations [Letter to the Editor]. *Archives of Sexual Behavior,* 39, 1239-1244.

Franklin, K. (2010). Hebephilia: Quintessence of diagnostic pretextuality. *Behavioral Sciences and the Law*, *28*, 751-768.

Green, R. (1972). Homosexuality as a mental illness. *International Journal of Psychiatry*, *10*, 77-98.

Green, R. (2010a). Sexual preference for 14-year-olds as a mental disorder: You can't be serious!! [Letter to the Editor]. *Archives of Sexual Behavior*, *39*, 585-586.

Green, R. (2010b). Hebephilia is a mental disorder? Sexual Offender Treatment, 5 (1), 1-10. Retrieved on May 21, 2011 from http://www.sexual-offender.org/2-2010_01.html.

Gretton, H. M., McBride, M., Hare, R., O'Shaugnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior, 28*, 427-449.

Hanson, R. K. & Bussiere, M. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348-362.

Hanson, R. K. & Bussiere, M. (1996). *Predictors of sexual offender recidivism: A meta-analysis*. Ottawa, Canada: Ministry of the Solicitor General of Canada (No. 1996-04).

Malcolm, P. B., Andrews, D., & Quinsey, V. (1993). Discriminant and predictive validity of phallometrically measured sexual age and gender preference. *Journal of Interpersonal Violence, 8*, 486-501.

Maletzky, B. (1993). Factors associated with success and failure in the behavioral and cognitive treatment of sexual offenders. *Annals of Sex Research, 6,* 241-258.

Marshall, W. L. & Barbaree, H. (1988). The long-term evaluation of a behavioral treatment program for child molesters. *Behavior Research and Therapy, 26*, 499-511.

McCall, R. B. (1975). *Fundamental statistics for psychology.* New YorK: Harcourt Brace Jovanovich.

Proulx, J., Pellerin, B., Paradis, Y., McKibben, A., Aubut, J., & Quimet, M. (1997). Static and dynamic predictors of recidivism in sexual aggressors. *Sexual Abuse: A Journal of Research and Treatment, 9,* 7-27.

Quinsey, V.L., Harris, G., Rice, M., & Cormier, C. (1998). Violent offenders. Washington, D.C.: American Psychological Association.

Rice, M. E., Quinsey, V., & Harris, G. (1991). Sexual recidivism among child molesters released from a maximum security psychiatric institution. *Journal of Consulting and Clinical Psychology, 59,* 381-386.

Sreenivasan, S., Frances, A., & Weinberger, L. (2010). Normative versus consequential ethics in sexually violent predator laws: An ethics conundrum. *The Journal of the American Academy of Psychiatry and the Law, 38*, 386-391.

Waggoner, J., Wollert, R., & Cramer, E. (2008). A re-specification of Hanson's 2006 Static-99 experience table that controls for the effects of age in a sample of 552 young sex offenders. *Law, Probability, and Risk*, *7*, 305-312.

Wollert, R. (2007). Poor diagnostic reliability, the Null-Bayes Logic Model, and their implications for sexually violent predator evaluations. *Psychology, Public Policy, and Law*, *13*, 167-203.

Wollert, R. & Waggoner, J. (2009). Bayesian computations protect sexually violent predator evaluations from the degrading effects of confirmatory bias and illusions of certainty: A reply to Doren and Levenson (2009). *Sexual Offender Treatment*, *4* (1), 1-26.

Wollert, R., Cramer, E., Waggoner, J., Skelton, J., & Vess, J. (2010). Recent research (N=9,305) underscores the importance of using age-stratified actuarial tables in sex offender risk assessments. *Sexual Abuse, 22,* 471-490.

Authors Notes

Dr. Wollert is Research Professor of Psychology at Washington State University Vancouver and a member of the Mental Health, Law, and Policy Institute at Simon Fraser University. Dr. Cramer is Professor Emeritus of Psychology at the University of North Carolina at Chapel Hill. The authors are indebted to Ray Blanchard for his helpfulness in sharing his data and providing feedback on portions of the manuscript, and to Brian Abbott, Ted Donaldson, Bill Marshall, and Jacqueline Waggoner for reviewing earlier versions of the present article. Correspondence should be sent to the first author at P.O. Box 61849, Vancouver, WA 98666-1849 or at rwwollert@aol.com.

Table 1. Criteria proposed for Pedohebephilic Disorder by the DSM-5 Task Force

A. Over a period of at least six months, one or both of the following as manifested by fantasies, urges, or behaviors:

   (1) recurrent and intense sexual arousal from prepubescent or pubescent children

   (2) equal or greater arousal from such children than from physically mature individuals

B. One or more of the following signs or symptoms:

   (1) the person has clinically significant distress or impairment in important areas of functioning from sexual attraction to children;

   (2) the person has sought sexual stimulation, on separate occasions, from either of the following:

      (a) two or more different children, if both are prepubescent

      (b) three or more different children, if one or more are pubescent

   (3) repeated use of, and greater arousal from, pornography depicting prepubescent or pubescent children than from pornography depicting physically mature persons, for a period of six months or longer

C. The person is at least age 18 years and at least five years older than the children in Criterion A or Criterion B

Specify type:

Pedophilic Type—Sexually Attracted to Prepubescent Children (Generally Younger than 11)

Hebephilic Type—Sexually Attracted to Pubescent Children (Generally Age 11 through 14)

Pedohebephilic Type—Sexually Attracted to Both

Specify type:

Sexually Attracted to Males

Sexually Attracted to Females

Sexually Attracted to Both

Specify if:

In Remission (No Distress, Impairment, or Recurring Behavior and in an Uncontrolled Environment): State duration of remission in months: _____

In a Controlled Environment

Table 2. Victim counts for Pedohebephiles and Teleiophiles from Table 1 of Blanchard (August 2010).

Count Categories	Teleiophiles (T)	Pedohebephiles (P)	Sum (T+P)	Error Rate (T/Sum)
0	335[a]	43	378	.89
1	289[a]	32	321	.90
2	121[a]	9	130	.93
3	42[c]	25[b]	67	.63
4	14[c]	10[b]	24	.58
5	9[c]	9[b]	18	.50
>5	7[c]	53[b]	60	.12
Total	817	181[d]	998[d]	

Note. The horizontal line in the middle of the table represents the selection cut-off. Those above this line were classified as Teleiophiles, those below it as Pedohebophiles.

[a] Specificity (.91) is the sum of these numbers (745) divided by the column total (817).
[b] Sensitivity (.54) is the sum of these numbers (97) divided by the column total (181).
[c] The overall estimated misdiagnosis rate for Pedohebephilia (.43) is the sum of these numbers (72) divided by the sum of their counterparts in the "Sum" column (169).
[d] The base rate for Pedohebephilia (.18) is the first number (181) divided by the second (998).

Table 3. Victim equivalency counts for Pedohebephiles and Teleiophiles from Table 3 of Blanchard (August 2010).

Count Categories	Teleiophiles (T)	Pedohebephiles (P)	Sum (T+P)	Error Rate (T/Sum)
0	306[a]	18	324	.94
1	280[a]	13	293	.96
2	145[a]	30	175	.83
3	49[c]	39[b]	88	.46
4	19[c]	11[b]	30	.63
5	11[c]	9[b]	20	.55
6	2[c]	13[b]	15	.13
7	1[c]	7[b]	8	.12
>7	4[c]	41[b]	45	.09
Total	817	181	998	

[a] Specificity = .89; [b] Sensitivity = .66; [c] estimated misdiagnosis rate for Pedohebephilia = .42.

Table 4. Victim counts for Pedohebephiles and Non-Pedohebephiles Based on All Patients in Blanchard's Sample.

Count Categories	Non-Pedohebephiles (O)	Pedohebephiles (P)	Sum (O+P)	Error Rate (O/Sum)
0	815[a]	43	858	.95
1	867[a]	32	899	.96
2	418[a]	9	427	.98
3	174[c]	25[b]	199	.87
4	83[c]	10[b]	93	.89
5	60[c]	9[b]	69	.87
>5	117[c]	53[b]	170	.69
Total	2,534	181[d]	2,715[d]	

[a] Specificity=.83; [b] Sensitivity=.54; [c] estimated misdiagnosis rate for Pedohebephilia=.82; [d] base rate=.07.

Table 5. Victim equivalency counts for Pedohebephiles and Teleiophiles Based on all patients in Blanchard's (August 2010) sample.

Count Categories	Non-Pedohebephiles (O)	Pedohebephiles (P)	Sum (O+P)	Error Rate (O/Sum)
0	672[a]	18	690	.97
1	819[a]	13	832	.98
2	530[a]	30	560	.95
3	211[c]	39[b]	250	.84
4	109[c]	11[b]	120	.91
5	65[c]	9[b]	74	.88
6	32[c]	13[b]	45	.71
7	25[c]	7[b]	32	.78
>7	71[c]	41[b]	112	.63
Total	2,534	181	2,715	

[a] Specificity = .80; [b] Sensitivity = .66; [c] estimated misdiagnosis rate for Pedohebephilia = .81.