Behavioral Sciences and the Law
Behav. Sci. Law **31**: 141–153 (2013)
Published online 16 January 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/bsl.2047

Classification Accuracy of Actuarial Risk Assessment Instruments[‡]

Daniel J. Neller, Psy.D., ABPP (Forensic)* and
Richard I. Frederick, Ph.D., ABPP (Forensic)[†]

Users of commonly employed actuarial risk assessment instruments (ARAIs) hope to generate numerical probability statements about risk; however, ARAI manuals often do not explicitly report data that are essential for understanding the classification accuracy of the instruments. In addition, ARAI manuals often contain data that have the potential for misinterpretation. The authors of the present article address the accurate generation of probability statements. First, they illustrate how the reporting of numerical probability statements based on proportions rather than predictive values can mislead users of ARAIs. Next, they report essential test characteristics that, to date, have gone largely unreported in ARAI manuals. Then they discuss a graphing method that can enhance the practice of clinicians who communicate risk via numerical probability statements. After the authors review several strategies for selecting optimal cut-off scores, they show how the graphing method can be used to estimate positive predictive values for each cut-off score of commonly used ARAIs, across all possible base rates. They also show how the graphing method can be used to estimate base rates of violent recidivism in local samples. Copyright © 2013 John Wiley & Sons, Ltd.

Classification decisions based on actuarial instruments are generally more accurate than those based on unaided professional judgments (Dawes, Faust, & Meehl, 1989; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000). This superiority extends to assessments of violence risk (Ægisdotter et al., 2006; Hanson & Morton-Bourgon, 2009; Mossman, 1994; Singh, Grann, & Fazel, 2011; cf. Litwack, 2001). Among the most widely used (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Jackson & Hess, 2007; Viljoen, McLachlan, & Vincent, 2010), generally accepted (Lally, 2003), and extensively validated (Campbell, French, & Gendreau, 2009; Hanson & Morton-Bourgon, 2009; Singh et al., 2011) actuarial risk assessment instruments (ARAIs) are the Violence Risk Appraisal Guide (VRAG; Quinsey, Harris, Rice, & Cormier, 1998, 2006), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 1998, 2006), the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR; Hanson, 1997), and the Static-99 (Hanson & Thornton, 2000).

Some users of ARAIs report numerical probability statements to convey recidivism risk. They state, for example, “Based on this ARAI score, Mr. Smith’s risk for violent recidivism over the next 10 years is x%.” This communication format has been sharply

*Correspondence to: Daniel J. Neller, Independent Practice, P.O. Box 1194, Wynne, AR 72396, U.S.A. E-mail: danieljneller@gmail.com
[†]Independent Practice, Springfield, MO.
[‡]Opinions expressed in this article do not necessarily represent the views of any department, agency, or institution with which the authors are affiliated.

criticized (e.g., Cooke & Michie, 2010, 2011; Hart, Michie, & Cooke, 2007; Scurich & John, 2012; cf. Hanson & Howard, 2010; Harris, Rice, & Quinsey, 2008; Mossman & Sellke, 2007). Nevertheless, this method of risk communication is sensible for at least three reasons. First, clinicians use ARAIs in contexts in which probabilities are relevant to decision-making. For example, ARAIs might be used in circumstances in which a probability threshold must be crossed before a decision is made to deprive an examinee of liberty (Janus & Meehl, 1997).

Second, some research indicates that clinicians do not reach high levels of agreement about the meaning of categorical risk statements (e.g., "low," "moderate," or "high" risk; McNiel & Binder, 1998). For example, Douglas and Reeves (2010) reported reliability estimates of summary risk ratings for what is perhaps the most widely used and empirically validated structured professional judgment guide designed to assess violence risk, the HCR-20 (Webster, Eaves, Douglas, & Hart, 1997). The median inter-rater reliability coefficients were 0.66 and 0.41 in forensic and correctional samples, respectively. Figures such as these might be unacceptably low for use in legal settings (Heilbrun, 1992).

Third, some evidence indicates decision-makers might make more accurate decisions when presented with numerical probability statements than with other risk communication formats (Hilton, Carter, Harris, & Sharpe, 2008). As a result, some authors (e.g., Kwartner, Lyons, & Boccaccini, 2006) encourage clinicians to add numerical probability values to categorical descriptions, when appropriate, to ensure risk communication is properly understood (also see Monahan & Steadman, 1996).

Although the communication of risk via numerical probability statements may be sensible, the manuals of many commonly used ARAIs do not make explicit all of the information necessary for making such statements. This is because ARAI manuals offer risk estimates in the form of proportions rather than predictive values; do not report such essential test characteristics as true- and false-positive rates for individual cut-off scores; and do not provide clear, practical guidance to clinicians whose local base rates of violent recidivism differ significantly from those studied in the ARAI calibration samples.

The information and analyses presented in this article are intended improve the accuracy of classification decisions made by ARAI users, particularly those who incorporate numerical probability statements into their violence risk assessments. The article shows how the reporting of numerical probability statements based on proportions rather than predictive values can mislead users of ARAIs. Next, it offers an analysis of calibration sample data of commonly used ARAIs, enabling users to consider essential test characteristics that have, to date, gone largely unreported in ARAI manuals. These essential test characteristics are then applied to a graphing procedure that enables users to estimate positive predictive values for all cut-off scores across all possible base rates. Finally, the article explains how the graphing method can be used to estimate local base rates of violent recidivism.

PROPORTIONS VERSUS PREDICTIVE VALUES

Clinicians who use numerical probability statements in their communication of risk estimates bear the responsibility to understand the meaning and limits of such statements. A clinician who, based on an ARAI score, reports a numerical estimate of the

DOI: 10.1002/bsl

likelihood that an examinee will recidivate violently is reporting the positive predictive value (PPV); however, manuals of commonly used ARAIs do not report the PPV. Instead, they typically report the proportion of people at each score or group of scores who recidivated violently. This is not the PPV.

The distinction between reporting the PPV and proportions cannot be overstated. Making a predictive statement about behavior based on a score or group of scores is typically taken to mean that the behavior is predicted in reference to all scores (i.e., with respect to the entire distribution of individuals). For instance, if a test has possible scores ranging from 0 to 100, and the chosen cut-off score is 65 (with 65 or higher predicting the condition of interest, such as violent recidivism), then the interpretation of a score of 65 is taken to mean "The risk of violent recidivism for persons scoring 65 *or higher* is x%." To make a predictive statement from "test score = 65" means considering the entire distribution of test scores.

In Figure 1, we contrast proportion-based predictions with PPV-based predictions. Panel A illustrates the method followed by the authors of the VRAG, the most extensively validated contemporaneous ARAI designed for assessing violence risk (Singh et al., 2011; Yang, Wong, & Coid, 2010). First, raw scores are gathered into groups

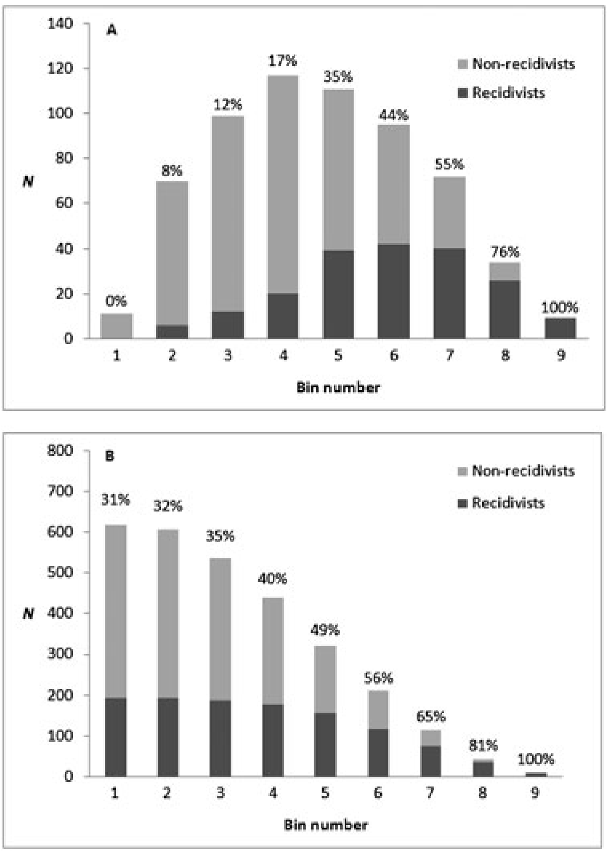

Figure 1. Distinguishing proportion (A) from positive predictive values (B). In panel A, each bar reflects the proportion of recidivists in the bin. In panel B, each bar reflects the probability that an individual will recidivate if he or she scores in a particular bin or in a higher numbered bin.

DOI: 10.1002/bsl

referred to as "bins." Next, the proportion of recidivists within each bin is calculated. As shown, the proportion of recidivists within each bin increases as the test score increases; this sort of representation is useful to show how a test scoring method captures more individuals with the condition of interest as test scores increase. But, as demonstrated in the following, this method does not correspond to highly accurate statements about violence risk.

Predictive statements are made possible when a validation sample closely resembles the population characteristics about which clinicians wish to make predictive statements (e.g., Singh et al., 2011). Because a single score or bin does not closely resemble the population they wish to know about, clinicians must consider the test's performance from bin 1 through bin 9 in total. Clinicians can begin this sort of analysis with panel B in Figure 1. In panel B, bin 1 refers to all participants because all of them obtained a score contained in bin 1 or higher. Bin 2 refers to all participants who obtained a score contained in bin 2 or higher, and so on.

Figure 1 clearly illustrates how two clinicians, using the same information from the VRAG, could arrive at different predictive statements depending on the type of probability they report. Presume, for instance, an examinee is assigned a VRAG score contained in bin 1. If the clinician relies upon the proportions reported by the VRAG authors, he or she might state that the examinee's likelihood of violent recidivism is 0%. If the clinician utilizes the PPV, he or she might state the examinee's likelihood of violent recidivism is 31%, the base rate (BR) of the study. Additional discrepancies are readily observed upon inspection of Figure 1.

The differences between these two forms of prediction are further elucidated when comparing the equation used to calculate mere proportion against the equation used to calculate PPV. To calculate the mere proportion of offenders who violently recidivated at each score or in each bin, the authors of commonly employed ARAIs used the following equation:

$$\text{Proportion} = \frac{n \text{ offenders in bin or at score who recidivated violently}}{n \text{ offenders in bin or at score}} \tag{1}$$

This method does not consider the BR of violent recidivism. Moreover, the equation does not contain the true-positive rate (TPR) and false-positive rate (FPR) of the respective ARAI cut-off score.

By contrast, to calculate PPV – the numerical probability statement in which ARAI users may be most interested – a clinician or researcher must consider the values for the recidivism BR and the ARAI's TPR and FPR at a particular cut-off score:

$$\text{PPV} = \frac{\text{TPR (BR)}}{[(\text{TPR} \times \text{BR}) + (\text{FPR}\ (1 - \text{BR}))]} \tag{2}$$

Equation 2 shows that estimates of the BR, TPR, and FPR are necessary to calculate PPV. The developers of the VRAG, SORAG, RRASOR, and Static-99 provide recidivism BRs for their respective calibration samples; however, they do not report estimates of TPR and FPR for each cut-off score. Based on data published on the

DOI: 10.1002/bsl

calibration samples (Hanson, 1997; Hanson & Thornton, 2000; Quinsey et al., 1998, 2006; for updated Static-99 estimates, see www.static99.org), these essential test characteristics for each potential cut-off score are presented in Table 1.

Test authors sometimes report PPV and negative predictive value (NPV, the likelihood of non-recidivism for offenders who score below identified cut-off scores) for each cut-off score. This can be highly misleading because, as Equation 2 shows, predictive values change as BR estimates change; that is, estimates of PPV and NPV are not static for any single cut-off score. The fluctuation of BRs, PPV, and NPV creates a quandary for clinicians who report numerical probability statements.

THE TEST VALIDATION SUMMARY

The Test Validation Summary (TVS; Frederick & Bowden, 2009a, 2009b) can assist clinicians who want to report predictive values in their risk assessments. The TVS is a graphing method that displays TPR, FPR, BR, PPV, NPV, and the proportion of positive scores in one figure for any particular cut-off score on any particular test.[1] An example of a TVS graph is presented in Figure 2. The BR varies along the *x*-axis. The proportion of positive test scores varies along the y-axis. The positive proportion line (PPL) shows the range of expected positive test scores at any particular BR by considering the FPR when the BR is 0% (point A), the TPR when the BR is 100% (point B), and all points between. As shown, the proportion of positive test scores varies linearly with the BR, whereas the PPV and NPV vary curvilinearly with the BR. TVS-generated PPV estimates for the VRAG, SORAG, RRASOR, and Static-99 across all possible BRs and cut-off scores are provided in Figure 3.

Use of the TVS to Estimate Local Base Rates

For any cut-off score with estimates of TPR and FPR, clinicians can observe the proportion of positive scores on the *y*-axis and find that value on the PPL to then determine their local BR on the *x*-axis (see Figure 4). (This determination is, of course, subject to the standard errors of proportion for the TPR and FPR estimates and the reliability restrictions of the procedure.) Once the BR is estimated, clinicians can discern the PPV and NPV of the cut-off score in their local sample.

To use the TVS to estimate a local BR, clinicians first identify a single cut-off score for a chosen ARAI. Far from an entirely objective process, the choice of a cut-off score depends on multiple factors, such as the purpose of the assessment, the benefits of correct decisions, and the costs of incorrect decisions (Haynes, Smith, & Hunsley, 2011). Therefore, a clinician might consider several strategies (e.g., Swets, 1992). One strategy is to consider the point at which the PPV would be expected to exceed 0.50 at a justifiable base rate, and where the proportion of correct decisions (i.e., true-positives and true-negatives) would be expected to be maximized. Ordinarily these occur at the intersection of the recidivist

[1] The TVS is distinguished from other graphical depictions of classification accuracy. For example, the predictive receiver operating characteristic curve (Shiu & Gastonis, 2008) enables users to estimate PPV and NPV for a given cut-off score; however, it does not enable users to estimate PPV and NPV across all possible cut-off scores and BRs.

DOI: 10.1002/bsl

Table 1. Essential test characteristics of commonly used actuarial risk assessment instruments (ARAIs)

ARAI: Follow-up	Cut point	*N*	Percentile	Rr (CI)	Rn	NRn	TPR	FPR
VRAG: 7-year								
	Bin 1	11	0.02	0.00 (0.00–0.26)	0	11	1.00	1.00
	Bin 2	70	0.13	0.08 (0.04–0.17)	6	64	1.00	0.97
	Bin 3	99	0.29	0.12 (0.07–0.20)	12	87	0.97	0.82
	Bin 4	117	0.48	0.17 (0.11–0.25)	20	97	0.91	0.62
	Bin 5	111	0.66	0.35 (0.27–0.44)	39	72	0.80	0.39
	Bin 6	95	0.81	0.44 (0.35–0.54)	42	53	0.60	0.22
	Bin 7	72	0.93	0.55 (0.44–0.66)	40	32	0.39	0.09
	Bin 8	34	0.99	0.76 (0.60–0.88)	26	8	0.18	0.02
	Bin 9	9	1.00	1.00 (0.70–1.00)	9	0	0.05	0.00
SORAG: 10-year								
	Bin 1	11.52	0.04	0.09 (0.02–0.38)	1.04	10.48	1.00	1.00
	Bin 2	25.92	0.13	0.12 (0.04–0.29)	3.11	22.81	1.00	0.91
	Bin 3	40.32	0.27	0.39 (0.26–0.54)	15.72	24.60	0.98	0.73
	Bin 4	57.60	0.47	0.59 (0.46–0.70)	33.98	23.62	0.88	0.53
	Bin 5	51.84	0.65	0.59 (0.46–0.72)	30.59	21.25	0.68	0.33
	Bin 6	46.08	0.81	0.76 (0.62–0.86)	35.02	11.06	0.49	0.16
	Bin 7	31.68	0.92	0.80 (0.64–0.91)	25.34	6.34	0.28	0.07
	Bin 8	17.28	0.98	0.89 (0.66–0.97)	15.38	1.90	0.13	0.02
	Bin 9	5.76	0.99	1.00 (0.61–1.00)	5.76	0.00	0.03	0.00
RRASOR: 10-year								
	Score 0	527	0.20	0.07 (0.05–0.09)	34	493	1.00	1.00
	Score 1	806	0.51	0.11 (0.09–0.14)	90	716	0.93	0.76
	Score 2	742	0.80	0.21 (0.18–0.24)	157	585	0.75	0.40
	Score 3	326	0.93	0.37 (0.32–0.42)	120	206	0.45	0.14
	Score 4	139	0.98	0.49 (0.41–0.57)	68	71	0.21	0.04
	Score 5	52	1.00	0.73 (0.60–0.83)	38	14	0.07	0.01
Static-99: 10-year (old data)								
	Score 0	107	0.10	0.11 (0.07–0.19)	12	95	1.00	1.00
	Score 1	150	0.24	0.07 (0.04–0.13)	11	139	0.95	0.89
	Score 2	204	0.42	0.13 (0.09–0.19)	27	177	0.90	0.77
	Score 3	206	0.61	0.14 (0.10–0.19)	29	177	0.79	0.52
	Score 4	190	0.79	0.31 (0.25–0.38)	59	131	0.66	0.31
	Score 5	100	0.88	0.38 (0.29–0.48)	38	62	0.41	0.16
	Score 6+	129	1.00	0.45 (0.37–0.54)	58	71	0.25	0.08
Static-99: 10-year (new data)								
	Score 0	175	0.11	5.7 (0.03–0.10)	10	165	1.00	1.00
	Score 1	196	0.23	4.6 (0.02–0.09)	9	187	0.97	0.87
	Score 2	226	0.37	6.6 (0.04–0.11)	15	211	0.94	0.73
	Score 3	229	0.51	17.5 (0.13–0.23)	40	189	0.89	0.57
	Score 4	272	0.68	18.0 (0.14–0.23)	49	223	0.76	0.43
	Score 5	178	0.79	28.7 (0.23–0.36)	51	127	0.60	0.26
	Score 6	150	0.88	32.7 (0.26–0.41)	49	101	0.44	0.16
	Score 7	105	0.94	44.8 (0.36–0.54)	47	58	0.28	0.08
	Score 8	60	0.98	38.9 (0.26–0.49)	22	38	0.12	0.04
	Score 9	16	0.99	46.4 (0.33–0.77)	9	7	0.05	0.01
	Score 10	14	1.00	54.0 (0.27–0.73)	7	7	0.02	0.01
Static-99R: 5-year								
	Score −3	40	0.017	0.00 (0.00–0.09)	0	40	1.00	1.00
	Score −2	65	0.044	0.00 (0.00–0.06)	0	65	1.00	0.98
	Score −1	260	0.152	0.03 (0.01–0.05)	7	253	1.00	0.95
	Score 0	294	0.274	0.03 (0.01–0.05)	8	286	0.95	0.84
	Score 1	350	0.419	0.03 (0.02–0.05)	10	340	0.90	0.72
	Score 2	350	0.565	0.04 (0.02–0.07)	14	336	0.83	0.56
	Score 3	343	0.707	0.06 (0.04–0.09)	20	323	0.73	0.42
	Score 4	277	0.823	0.06 (0.04–0.10)	17	260	0.59	0.27
	Score 5	193	0.903	0.15 (0.10–0.20)	28	165	0.48	0.16

(*Continues*)

Table 1. (Continued)

ARAI: Follow-up	Cut point	N	Percentile	Rr (CI)	Rn	NRn	TPR	FPR
	Score 6	110	0.948	0.13 (0.08–0.20)	14	96	0.28	0.09
	Score 7	75	0.980	0.16 (0.09–0.26)	12	63	0.19	0.04
	Score 8	28	0.991	0.29 (0.15–0.47)	8	20	0.10	0.02
	Score 9	13	0.997	0.38 (0.18–0.64)	5	8	0.05	0.01
	Score 10	7	0.999	0.29 (0.08–0.64)	2	5	0.01	0.00
	Score 11	1	1.000	0.00 (0.00–0.79)	0	1	0.00	0.00

Note. Violent recidivism is the outcome variable for the VRAG and SORAG. Sexual-specific recidivism is the outcome variable for the RRASOR and Static-99. *N*, number of offenders at the identified score; Rr (CI), rate of recidivism and associated confidence interval; Rn, number of offenders at the cut-off score who recidivated; NRn, number of offenders at the cut-off score who did not recidivate; VRAG, Violence Risk Appraisal Guide; SORAG, Sex Offender Risk Appraisal Guide; RRASOR, Rapid Risk Assessment for Sexual Offense Recidivism. *Note 2*. Correction added on 28 January 2013 after initial online publication on 16 January 2013. The VRAG data for Bins 8 and 9 were originally omitted from the table. This error has been corrected in this version of the article.

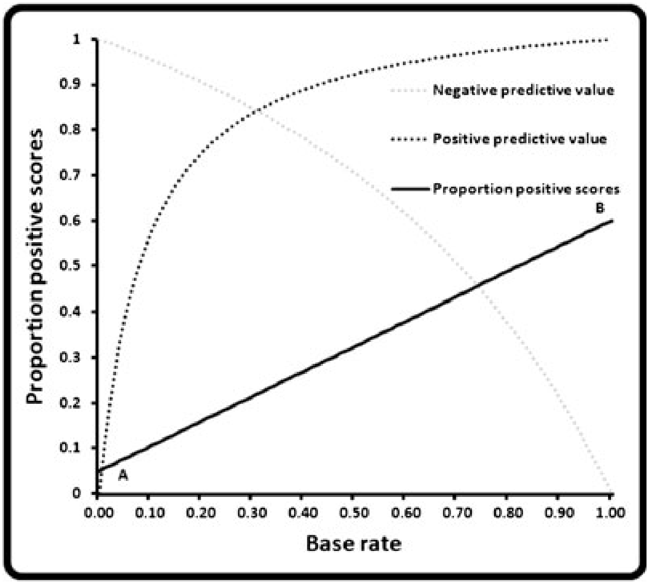

Figure 2. Example of a Test Validation Summary graph. The positive proportion line (PPL) extends from A, the false-positive rate (FPR, the rate of positives when base rate = 0) to B, the true-positive rate (TPR, the rate of positives when base rate = 1). The rate of positive outcomes varies between A and B, as the base rate changes. The curved lines, positive predictive value (dark dotted line) and negative predictive value (light dotted line), are influenced by base rate, FPR, and TPR.

and non-recidivist distributions (Cureton, 1957; Meehl & Rosen, 1955; Rorer & Dawes, 1982; Rorer, Hoffman, LaForge, & Hsieh, 1966). Based on the actuarial tables and normative data published on the calibration samples, recidivist and non-recidivist distributions overlap at VRAG bin 7, SORAG bin 4, RRASOR score 4, Static-99 score 9 (new data), and Static99R score 9 (see Figure 5).[2]

[2] Older data published for the Static-99 (Hanson & Thornton, 2000) show that the distribution of recidivists remains beneath the distribution of non-recidivists across all cut-off scores offered by the authors; that is, the distributions never cross, even at the highest cut-off score offered by the Static-99 authors (6+).

DOI: 10.1002/bsl

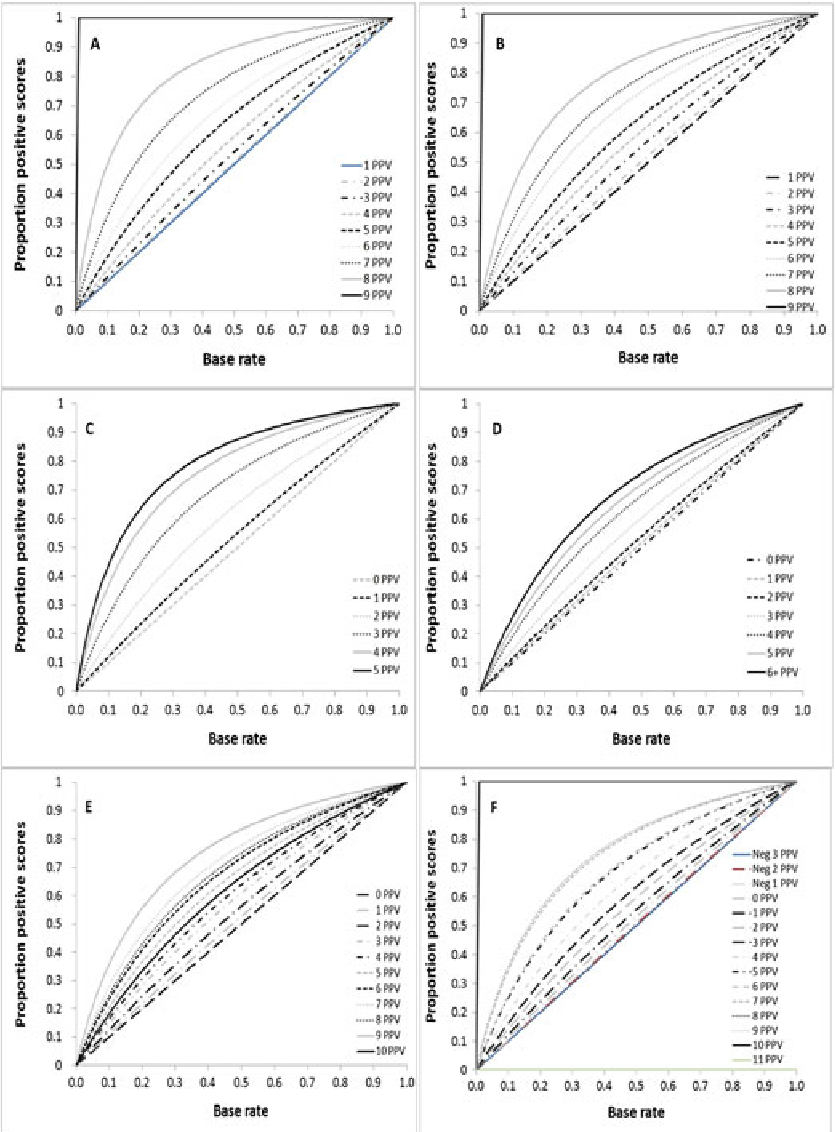

Figure 3. Positive predictive values of various cut-off scores for: (A) the Violence Risk Appraisal Guide (VRAG, 7-year); (B) the Sex Offender Risk Appraisal Guide (SORAG, 10-year); (C) the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR, 10-year); (D) the Static-99 (10-year, old data); (E) the Static-99 (10-year, new data); and (F) the Static-99R (5-year).

An alternate strategy for selecting a cut-off score is to inspect a representative receiver operating characteristic (ROC) curve of an ARAI (Swets, Dawes, & Monahan, 2000). Depending upon the goals of the assessment, the clinician might seek to determine the point at which the instrument's TPR (i.e., sensitivity) and true-negative rate (TNR, or specificity, or 1 – FPR) approximate each other (Swets, 1992). This is accomplished by drawing

DOI: 10.1002/bsl

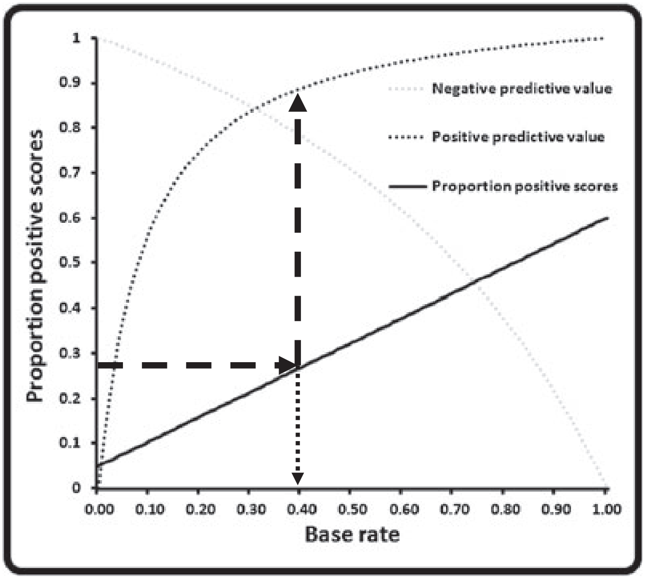

Figure 4. Using the Test Validation Summary (TVS) to derive local base rate (BR). Plot the false-positive rate (FPR) at BR = 0 and the true-positive rate (TPR) at BR = 1, and draw the positive proportion line (PPL) between those two points. Observe the rate of positive scores in the sample and plot as the *y*-coordinate on the PPL. The *x*-coordinate at that point is the BR estimate for the sample. The estimates of positive predictive value (PPV) and negative predictive value (NPV) are observed as the *y*-values of the curved lines at the BR estimate.

a diagonal from the upper left corner of the figure (0, 1) to the lower right corner of the figure (1, 0). Sensitivity and specificity typically approximate each other at the point where the diagonal intersects the ROC curve. As shown in Figure 6, sensitivity and specificity most closely balance at the following cut-off points: VRAG bin 6, SORAG bin 5, RRASOR score 2, Static-99 score 4 (old data), Static-99 score 5 (new data), and Static-99R score 4.

Another strategy for selecting an ARAI cut-off score is to inspect its TVS graph. If test users value the cut-off score that jointly maximizes PPV and NPV at a given BR, they can easily observe the point at which PPV and NPV are jointly maximized – ordinarily the point at which these parameters intersect in the TVS graph. Although other cut-off scores may produce higher values for either PPV or NPV, usually no other combination of PPV and NPV will have a higher value.

As depicted in Figure 4, after clinicians select an optimal cut-off score, they are prepared to take the next steps toward using the TVS to estimate their local BR. Clinicians review a sample of violence risk assessments they previously conducted; calculate the proportion of cases in which examinees scored above the optimal cut-off score; locate that value on the *y*-axis; and identify that same value on the PPL. The local BR is the corresponding *x*-value.[3]

[3] The TVS graphing spreadsheet is available at www.richardfrederick.com.

DOI: 10.1002/bsl

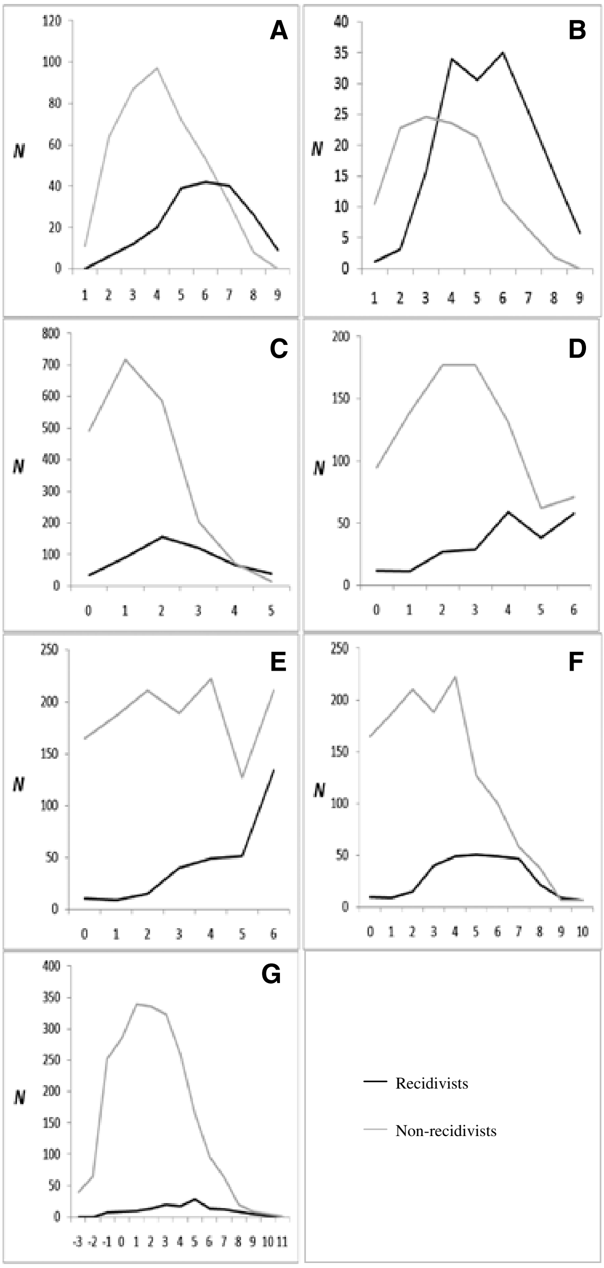

Figure 5. Distributions of recidivists and non-recidivists are plotted for: (A) Violence Risk Appraisal Guide (VRAG, 7-year) bins; (B) Sex Offender Risk Appraisal Guide (SORAG, 10-year) bins; (C) Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR, 10-year) scores; (D) Static-99 (10-year, old data) scores; (E) Static-99 (10-year, 0 to 6+, new data) scores; (F) Static-99 (10-year, 0 to 10, new data) scores; and (G) Static-99R (5-year) scores. Ordinarily, classification accuracy is expected to be highest at points at which recidivist and non-recidivist distributions overlap.

DOI: 10.1002/bsl

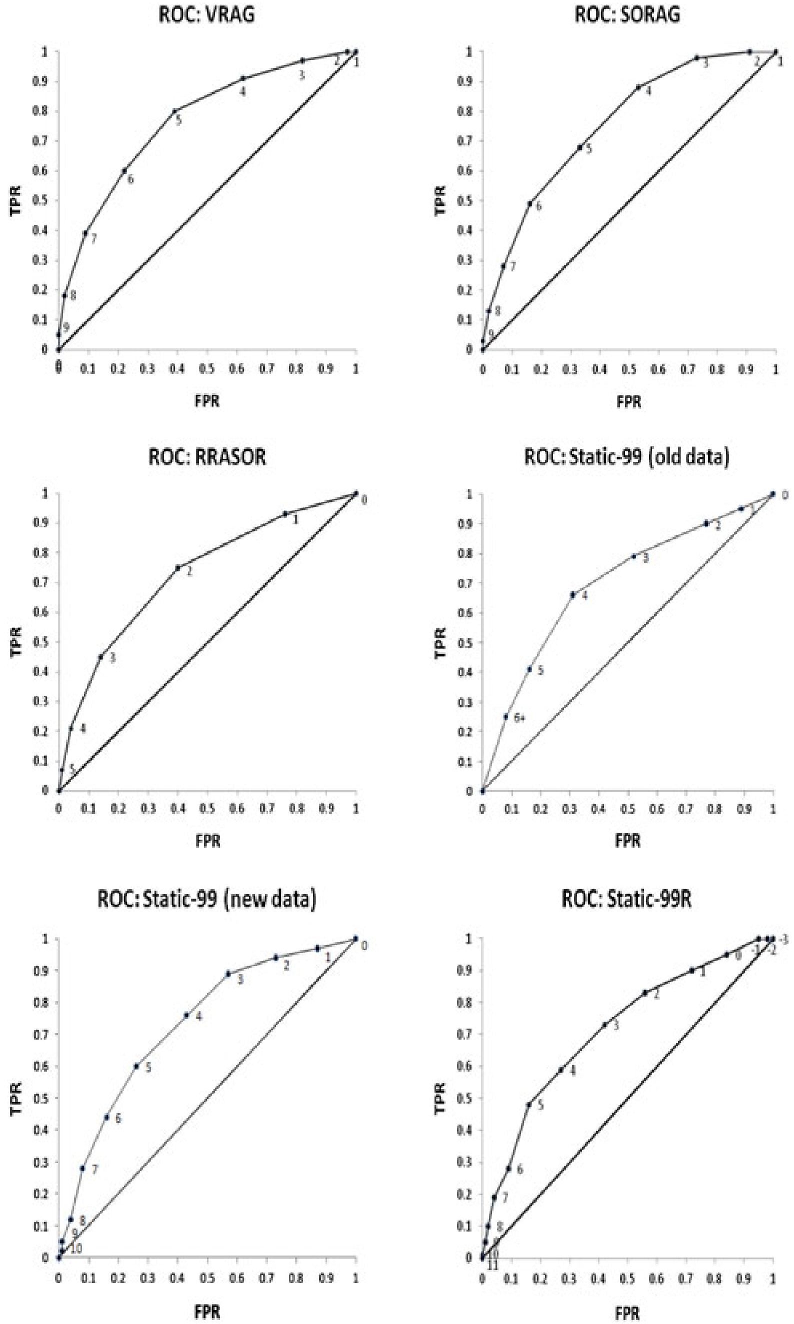

Figure 6. Receiver operating characteristic (ROC) curves with cut-off scores for the Violence Risk Appraisal Guide (VRAG, 7-year), the Sex Offender Risk Appraisal Guide (SORAG, 10-year), the Rapid Risk Assessment for Sexual Offense Recidivism (RRASOR, 10-year), the Static-99 (10-year, old data), the Static-99 (10-year, new data), and the Static-99R (5-year). FPR, false-positive rate; TPR, true-positive rate.

DOI: 10.1002/bsl

CONCLUSION

Accurate numerical probability statements are made possible when certain information is available. Manuals of commonly used ARAIs contain useful information; however, they do not make explicit the information necessary for generating accurate numerical probability statements. Without justifiable cut-off scores, local BR data, and clear estimates of essential test characteristics, in most circumstances clinicians will lack sufficient data upon which to make meaningful predictive statements about violence risk. By contrast, clinicians who are equipped with justifiable cut-off scores, local BR data, and reliable estimates of test characteristics have important information that can contribute to meaningful probability statements about violence risk. The material presented in this article will hopefully assist clinicians in their efforts to improve the accuracy of their classification decisions and enhance the precision of their risk communication.

REFERENCES

Ægisdotter, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Rush, J. D. (2006). The meta- analysis of clinical judgment project: Fifty six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*, 341–382. DOI: 10.1177/0011000005285875

Archer, R. A., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, *87*(1), 84–94. DOI: 10.1207/s15327752jpa8701_07

Campbell, M., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior*, *36*, 567–590. DOI: 10.1177/0093854809333610

Cooke, D., & Michie, C. (2010). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior*, *34*, 259–274. DOI: 10.1007/s10979-009-9176-x

Cooke, D., & Michie, C. (2011). Violence risk assessment: Challenging the illusion of certainty. In McSherry B. & Keyzer P. (Eds.), *Dangerous people: Policy, prediction, and practice* (pp. 147–161). New York, NY: Routledge. ISBN: 978-0-415-88495-2

Cureton, E. E. (1957). Recipe for a cookbook. *Psychological Bulletin*, *54*(6), 494–497. DOI: 10.1037/h0041406

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668–1674. DOI: 10.1126/science.2648573

Douglas, K. S., & Reeves, K. A. (2010). Historical-Clinical-Risk Management – 20 (HCR-20) Violence Risk Assessment Scheme: Rationale, application, and empirical overview. In Otto R. K., & Douglas K. S. (Eds.), *Handbook of violence risk assessment* (pp. 147–185). New York: Taylor & Francis Group. ISBN: 978-0-415-96214-8.

Frederick, R. I., & Bowden, S. (2009a). The test validation summary. *Assessment*, *16*, 215–236. DOI: 10.1177/1073191108325005

Frederick, R. I., & Bowden, S. (2009b). Evaluating constructs represented by symptom validity tests in neuropsychological assessment of traumatic brain injury. *The Journal of Head Trauma Rehabilitation*, *24*, 105–122. DOI: 10.1097/HTR.0b013e31819b1210

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293–323. DOI: 10.1037/1076-8971.2.2.293

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*(1), 19–30. DOI: 10.1037/1040-3590.12.1.19

Hanson, R. K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. Ottawa, Ontario: Department of the Solicitor General of Canada.

Hanson, R. K., & Howard, P. D. (2010). Individual confidence intervals to not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior*, *34*(4), 275–281. DOI: 10.1007/s10979-010-9227-3

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, *21*(1), 1–21. DOI: 10.1037/a0014421

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, *24*(1), 119–136. DOI: 10.1023/A:1005482921333

Harris, G. T., Rice, M. E., & Quinsey, V. L. (2008). Shall evidence-based risk assessment be abandoned? *The British Journal of Psychiatry*, *192*, 154. DOI: 10.1192/bjp.192.2.154

Hart, S. D., Michie, C., & Cooke, D. J. (2007). Precision of actuarial risk assessment instruments: Evaluating the 'margins of error' of group v. individual predictions of violence. *The British Journal of Psychiatry*, *190*, s60–s65. DOI: 10.1192/bjp.190.5.s60

Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2011). *Scientific foundations of clinical assessment*. New York: Routledge. ISBN: 978-0-415-87651-3.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, *16*(3), 257–272. DOI: 10.1007/BF01044769.

Hilton, Z., Carter, A., Harris, G., & Sharpe, A. (2008). Does using nonnumerical terms to describe risk aid violence risk communication? Clinician agreement and decision making. *Journal of Interpersonal Violence*, *23*(2), 171–188. DOI: 10.1177/0886260507309337

Jackson, R. L., & Hess, D. T. (2007). Evaluation for civil commitment of sex offenders: A survey of experts. *Sexual Abuse: A Journal of Research and Treatment*, *19*, 425–448. DOI: 10.1177/107906320701900407

Janus, E., & Meehl, P. E. (1997). Assessing the legal standard for predictions of dangerousness in sex offender commitment proceedings. *Psychology, Public Policy, and Law*, *3*(1), 33–64. DOI: 10.1037/1076-8971.3.1.33.

Kwartner, P. P., Lyons, P. M., & Boccaccini, M. T. (2006). Judges' risk communication preference in risk for future violence cases. *International Journal of Forensic Mental Health*, *5*(2), 185–194. DOI: 10.1080/14999013.2006.10471242

Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research & Practice*, *34*(5), 491–498. DOI: 10.1037/0735-7028.34.5.491

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law*, *7*(2), 409–443. DOI: 10.1037/1076-8971.7.2.409

McNiel, D. & Binder, R. (1998, March). Comparison of categorical and probabilistic approaches to communication about psychiatric patients' risk of violence in clinical practice. Paper presented at the biennial Conference of the American Psychology-Law Society/Division 41 of the American Psychological Association, Redondo Beach, CA.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. *Psychological Bulletin*, *52*, 194–216. DOI: 10.1037/h0048070

Monahan, J., & Steadman, H. J. (1996). Violent storms and violent people: How meteorology can inform risk communication in mental health law. *American Psychologist*, *51*(9), 931–938. DOI: 10.1037/0003-066X.51.9.931

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology*, *62*(4), 783–792. DOI: 10.1037/0022-006x.62.4.783

Mossman, D., & Sellke, T. (2007). Avoiding errors about 'margins of error.' *The British Journal of Psychiatry*, *191*, 561. DOI: 10.1192/bjp.191.6.561

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association. ISBN: 1-55798-495-6.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk*, second edn. Washington, DC: American Psychological Association. ISBN: 1-59147-343-8.

Rorer, L. G., & Dawes, R. M. (1982). A base-rate bootstrap. *Journal of Consulting and Clinical Psychology*, *50*(3), 419–425. DOI: 10.1037/0022-006x.50.3.419

Rorer, L. G., Hoffman, P. L., LaForge, G. E., & Hsieh, K. (1966). Optimum cutting scores to discriminate groups of unequal size and variance. *Journal of Applied Psychology*, *50*(2), 153–164. DOI: 10.1037/h0022944

Scurich, N., & John, R. S. (2012). A Bayesian approach to the group versus individual prediction controversy in actuarial risk assessment. *Law and Human Behavior*, *36*(3), 237–246. DOI: 10.1037/H0093973

Shiu, S., & Gastonis, C. (2008). The predictive receiver operating characteristic curve for the joint assessment of the positive and negative predictive values. *Philosophical Transactions of the Royal Society*, *366*(1874), 2313–2333. DOI: 10.1098,rsta.2008.0043

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, *31*(3), 499–513. DOI: 10.1016/j.cpr.2010.11.009

Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, *47*(4), 522–532. DOI: 10.1037/0003-06x.47.4.522

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*(1), 1–26. DOI: 10.1111/1529-1006.001

Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, *17*(3), 377–395. DOI: 10.1177/1073191109359587

Webster, C. D., Eaves, D., Douglas, K. S., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence, version 2*. Burnaby, BC, Canada: Simon Fraser University. ISBN: 978-0864911575.

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, *136*(5), 740–767. DOI: 10.1037/a0020473

DOI: 10.1002/bsl