AL/AO-TR-1997-0098

UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

THE RELATIONSHIP BETWEEN RELIABILITY AND VALIDITY IN A BAYESIAN WORLD

Paul D. Retzlaff
William G. Jackson

AEROSPACE MEDICINE DIRECTORATE
CLINICAL SCIENCES DIVISION
NEUROPSYCHIATRY BRANCH
2507 Kennedy Circle
Brooks Air Force Base, TX 78235-5117

July 1997

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this technical report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

William G. Jackson

WILLIAM G. JACKSON
Project Scientist

Daniel L Van Syoc

DANIEL L. VAN SYOC, Lt Col, USAF, MC, CFS
Acting Chief, Clinical Sciences Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 1997	Interim Technical Report, Aug 96 - Jul 97

4. TITLE AND SUBTITLE
The Relationship Between Reliability and Validity in a Bayesian World

5. FUNDING NUMBERS

6. AUTHOR(S)
Paul D. Retzlaff
William G. Jackson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armstrong Laboratory
Aerospace Medicine Directorate
Clinical Sciences Division, Neuropsychiatry Branch
2507 Kennedy Circle
Brooks Air Force Base, TX 78235-5117

8. PERFORMING ORGANIZATION REPORT NUMBER
AL/AO-TR-1997-0098

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

While the relationship between internal consistency and validity in traditional tests is well established, little is available for tests that predict dichotomous diagnostic outcomes. An understanding of such a relationship is important in cases such as the prediction of clinician diagnoses by psychological tests. Tests such as the MCMI are validated against such clinical diagnoses. The limiting effects on positive predictive powers given levels of Kappa are developed. The current work provides the relationship between Kappa and Positive Predictive Power for use with tests and applies it specifically to a test of psychopathology as an example. This may be applied to any situation where judgments are predicted by tests such as in mental health, medicine, or selection and training.

14. SUBJECT TERMS
Epidemiology
Psychological testing
Psychological assessment
Psychiatric diagnosis

15. NUMBER OF PAGES
22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. Z39-18
298-102

CONTENTS

PREFACE

This project was made possible by support from the Air Force Medical Operating Agency, Armstrong Laboratory, and the AFOSR Summer Faculty Research Program.

Summary

While the relationship between internal consistency and validity in traditional tests is well established, little is available for tests that predict dichotomous diagnostic outcomes. An understanding of such a relationship is important in cases such as the prediction of clinician diagnoses by psychological tests. Tests such as the MCMI are validated against such clinical diagnoses. The limiting effects on positive predictive powers given levels of Kappa are developed. The current work provides the relationship between Kappa and Positive Predictive Power for use with tests and applies it specifically to a test of psychopathology as an example. This may be applied to any situation where judgments are predicted by tests such as in mental health, medicine, or selection and training.

The relationship between reliability and validity
in a Bayesian world

Introduction

Background: The relationship between reliability and validity in Fisherian statistics is well established (Nunnally, 1978; Suen, 1990). Traditional tests calculate reliability through Cronbach alpha and validity, usually, through a correlation coefficient. With traditional norm referenced and continuous metric tests such as the MMPI-II (Butcher, Dahlstrom, Graham, Tellegen, and Kaemmer, 1991) these statistics are appropriate.

Some of the newer tests such as the MCMI-III (Millon, 1994), however, provide diagnostic hit rate data which is dichotomous and Bayesian in nature (Retzlaff, 1995; Craig, 1993). For example, while the Base Rate scores of the MCMI-III can vary from 0 to 115, the fundamental interpretation is whether the score is 85 or greater. This test was built to optimally predict membership in a diagnostic group (American Psychiatric Association, 1987; 1994) and 85 is the cut score. This test and its earlier versions (Millon, 1977; 1987) have often been used in the military (e. g., King, 1994, Retzlaff and Gibertini, 1987; 1988).

Validity of tests such as these is calculated through operating characteristics (see Table 1). These characteristics (Gibertini, Brandenburg, and Retzlaff, 1986; Williams, 1982) include diagnostic prevalence, test positives, sensitivity, specificity, positive predictive power, and negative predictive power. Positive predictive power is the most important statistic for clinicians. It is the proportion of cases who are identified as having the disorder who actually have the disorder. It answers the question, for example, "Of all patients with a high score on Antisocial, how many are actually antisocial?"

The calculation of these statistics involves a 2 by 2 hit rate matrix with the test on one marginal and clinician diagnoses on the other. The problem with this is, however, that the test is being validated against clinician diagnoses which are to some degree unreliable (e. g., Retzlaff, 1996, Retzlaff and Gibertini, 1994). Interjudge (clinician) agreement is usually established through the calculation of Kappa (see Table 2). Kappa is basically the proportion of correct agreement beyond sheer chance (Wickens, 1989). It, therefore, corrects for situations where extreme prevalence artifactually increases apparent interjudge agreement. It is calculated through a 2 by 2 matrix with each of two judges on one marginal. DSM field trials (Diagnostic and Statistical Manual; American Psychiatric Association, 1980) for example found Kappa's for the personality disorders (Millon, 1981; 1990) in the 0.26 to 0.76 range. Clinicians are not very reliable. Part of this is due to the relatively low prevalence of clinical disorders which usually are in the 0.05 to 0.15 range.

Purpose: The purpose of the current work was to calculate the ceiling effect of Kappa on positive predictive power.

Method and Results

As both operating characteristics and Kappa are calculated from 2 by 2 matrices, common cell frequencies could be used to calculate both. In effect, the summary statistics are algebraically solved through the common cell frequencies. In many ways, just as reliability is a special case of validity (the validity of a test against itself), Kappa is just another way of looking at operating characteristics.

Table 1

The calculation of operating characteristics

	Judge +	Judge -	
Test +	a	b	a + b
Test -	c	d	c + d
	a + c	b + d	1.00

positive predictive power = a/(a+b)
negative predictive power = d/(c+d)
sensitivity = a/(a+c)
specificity = d/(b+d)
prevalence = a+c
test positives = a+b

Table 2

The calculation of Kappa

	Judge A +	Judge A -	
Judge B +	a	b	a + b
Judge B -	c	d	c + d
	a + c	b + d	1.00

po = a+d
pc = ((a+b)*(a+c))+((c+d)*(b+d))
Kappa = (po-pc)/(1-pc)
prevalences = a+b and a+c

Table 3 combines the operating characteristic and Kappa tables into, in essence, a three way table including both judges and the test. This is done to allow for an integrated approach to the problem. The eight cells include all possible combinations of judgment agreement and test prediction.

Table 3

The combination of operating characteristics and Kappa tables

	$J_{1+}J_{2+}$	$J_{1+}J_{2-}$	$J_{1-}J_{2+}$	$J_{1-}J_{2-}$	
Test +	a_{test+}	b_{test+}	c_{test+}	d_{test+}	test+
Test -	a_{test-}	b_{test-}	c_{test-}	d_{test-}	test-
	a	b	c	d	1

In order to set Kappa and operating characteristics equal, a number of assumptions are necessary. The purpose of these assumptions is to limit and constrain the models in such a manner as to allow for a solution. It is impractical to attempt to solve such a problem with too many "degrees of freedom". The first assumption is that the two judges will have equal prevalence rates. In effect, each judge will diagnose the same proportion of cases as "having the disorder". In practice such is not always the case. The degree to which prevalences are different, however, impacts the reliability of the judgments and Kappa. The more the prevalences are different, the lower Kappa will be. By setting the two prevalences equal, this source of error is eliminated and allows for the desired estimation of maximal PPP given "pure disagreement". Included in this assumption is that the test positive rate will equal the clinician prevalence rates. Here again, test positive rates may be different from the underlying clinician prevalence rates but doing so will usually exact a cost in terms of a lowered PPP.

The second assumption further defines the model in asserting that the sensitivity (and given equal prevalences, the specificity) of the test to each judge is the same. This constraint is necessary to eliminate situations where the test is "better" at modeling the decisions of one judge over the other. Indeed, without this assumption, there is nothing to prevent PPP with respect to one of the judges from reaching 1.00.

Assumption #1: Prevalence of disorder is identical for Judge 1 and Judge 2. The test positives (prevalence for test) is also identical to the judges.

In the case of the Kappa table,
therefore, $a+b = a+c$,
which means $b = c$.

In the case of the Operating Characteristics table and the table above,
test positives = $a_{test+} + b_{test+} + c_{test+} + d_{test+}$,
which means test positives = a + b.

Assumption #2: Sensitivity of the test relative to each judge is the same.

So, $(a_{test+} + b_{test+}) / (a + b) = (a_{test+} + c_{test+}) / (a + c)$,
which means $b_{test+} = c_{test+}$ and $b_{test-} = c_{test-}$.

Placing these constraints on the problem and Table 3 gives Table 4. The two assumptions make the off diagonals of the original 2 by 2 matrices equal. As such, b equals c and c may be replaced by b, simplifying the matrix and cellular structure.

Incorporating Assumptions #1 and #2 gives the table below:

Table 4

Combined tables with first assumptions

	$J_{1+}J_{2+}$	$J_{1+}J_{2-}$	$J_{1-}J_{2+}$	$J_{1-}J_{2-}$	
Test +	a_{test+}	b_{test+}	b_{test+}	d_{test+}	test+
Test -	a_{test-}	b_{test-}	b_{test-}	d_{test-}	test-
	a	b	b	d	1

In our quest for maximal PPP given a specific Kappa, the maximal case will occur when the test at least always agrees when the judges agree. As such, when the two judges agree on either the diagnosis being present or being absent, the test would also agree (Assumption 3a). This suggests that the test is perfectly reliable and valid. No test is, but this constant allows for the calculation of a truly maximal PPP. Assumption 3b later will provide an alternative.

Additionally, imbedded in this assumption is the "correcting" of Kappa for the reduction in judges from two to one. The logic goes that "it takes two to disagree". One could assume that half of the disagreement is attributable to one judge and the other half to the other judge. In essence, the off diagonals are error and half the error is attributable to each of the two judges. If it is necessary to develop a model where a test is used to predict the diagnoses of a single judge, then it is necessary to correct for the "double error" and attribute only half the error to the single judge. In effect, this is a "single judge corrected Kappa". In and of itself it is an important conceptual development. However, with some algebra, maximal PPP may be defined in terms of b and prevalence (see Equation 1).

Assumption 3a: The maximum possible value of PPP will occur when the test is perfect relative to the agreements of the judges and $a_{test-} = 0$ and $d_{test+} = 0$, meaning $a_{test+} = a$ and $b_{test+} = b/2$.

Therefore, $PPP_{max} = (a + b/2) / (a + b)$
$PPP_{max} = (a + b - b/2) / (a + b)$
$PPP_{max} = 1 - (b/ (2(a+b)))$

Since prevalence = a + b,

$PPP_{max} = 1 - (b/ (2(prev))$ **Equation 1.**

With an ultimate goal of defining PPP in terms of Kappa and prevalence, b must be defined in terms of Kappa and prevalence. Equation 2 defines b in terms of Kappa and pc. Equation 3 substitutes this for b in Equation 1. Equation 4 solve for pc in a manner which allows for its substitution into Equation 3.

Now, $K = (po - pc) / (1 - pc)$,
where $po = a + d = 1 - 2b$
and $pc = (a+c)(a+b) + (b+d)(c+d)$
$= (a+b)^2 + (b+d)^2$
$= (prev)^2 + (1-prev)^2$

Therefore, $K = ((1-2b) - pc) / (1 - pc)$
$K = ((1-pc) - 2b) / (1 - pc)$
$K (1-pc) = (1-pc) - 2b$
$2b = (1-pc) - K(1-pc)$
$2b = (1-pc) (1-K)$

$b = (1-pc)(1-K) / 2$ **Equation 2.**

Substituting equation 2 for b in equation 1,

$PPP_{max} = 1 - (((1-pc)(1-K))/2) / 2(prev)$
$PPP_{max} = 1 - ((1-pc)(1-K)) /4(prev)$

$PPP_{max} = 1 - ((1-K)/4) ((1-pc)/prev)$ **Equation 3.**

Next, consider what (1-pc)/prev might be,

Recall $pc = (prev)^2 + (1- prev)^2$,

so $(1\text{-pc})/\text{prev} = (1 - (\text{prev})^2 - (1\text{-prev})^2) / \text{prev}$

$(1\text{-pc})/\text{prev} = (1 - (\text{prev})^2 - 1 + 2(\text{prev}) - (\text{prev})^2) / \text{prev}$

$(1\text{-pc})/\text{prev} = (2(\text{prev}) - 2(\text{prev})^2) / \text{prev}$

$(1\text{-pc})/\text{prev} = (2(\text{prev})(1\text{-prev})) / \text{prev}$

$(1\text{-pc})/\text{prev} = 2(1\text{-prev})$ **Equation 4.**

Substituting equation 4 for (1-pc)/prev in equation 3,

$$PPP_{max} = 1 - ((1\text{-K})/4)\ (2(1\text{-prev}))$$

Simplifying the above, we are left with the relationship between PPP and Kappa given a specific prevalence level. This, however, does include the above three assumptions. The first two are relatively appropriate. The third, though, assumes a test which is perfectly reliable and valid.

So PPPmax under the current assumptions is related to kappa in the following manner:

$PPP_{max} = 1 - ((1\text{-K})(1\text{-prev})) / 2$ **Equation 5.**

Appendix A provides these figures for a range of Kappa's at .05, .10, and .15 prevalence levels. It should be noted that this figure is truly a maximal PPP and as such will probably never be attained by a test of any sort. Note also the probably unrealistic elements at the extreme lower end of Kappa's. In the case of a .05 prevalence, the off diagonal correction for the single judge allows for more correction than reality. In effect, at a Kappa of -.05, which is below chance, the test supposedly could have a PPP_{max} of .50. This is highly unlikely and purely the result of the correction from two judges to one.

This, however, assumes that the test is perfectly reliable and valid. An additional assumption could be proposed to better reflect most tests in the "real world". Tests are less than perfectly reliable and certainly less than perfectly valid. Some estimation of the degree of imperfection is necessary. An argument could be made that the quality of the test is fairly directly related to the quality of clinician judgments. If two judges can't seem to agree on a diagnosis (and have a low Kappa), it is likely that a test of that particular diagnosis would also be relatively poor. The connection is probably even more direct when one considers that fact that it is the clinicians who write and choose items for psychological tests. One could set the imperfection of the test equal to the imperfection of the clinicians.

Assumption #3b: The level of agreement between the test and either judge is the same as the agreement between the judges.

Therefore, $a_{test+} + b_{test+} = a$.

Incorporating that assumption, the table would become:

Table 5

Combined tables with final assumption

	$J_{1+}J_{2+}$	$J_{1+}J_{2-}$	$J_{1-}J_{2+}$	$J_{1-}J_{2-}$	
Test +	a - x	x	x	b - x	a + b
Test -	x	b - x	b - x	d - (b - x)	b + d
	a	b	b	d	1

If PPP = a / (a + b) generally
then PPP_{rw} = ((a - x) + x) / (a + b), where rw stands for "real world"
then PPP_{rw} = (a + b - b) / (a + b)
PPP_{rw} = ((a + b)/(a + b)) - (b/(a + b))

PPP_{rw} = 1 - b/prev **Equation 6.**

Equation 6 is very similar to Equation 1 except that the prevalence element has been reduced. As such, more is subtracted from 1 and a more conservative and "real world" PPP is developed.

Substituting equation 2 for b in equation 6 gives,

PPP_{rw} = 1 - (((1-pc)(1-K))/2) / prev
PPP_{rw} = 1 - ((1-K)/2) ((1-pc)/prev)

Substituting equation 4 for (1-pc)/prev gives,

PPP_{rw} = 1 - ((1-K)/2) (2(1-prev))
PPP_{rw} = 1 - (1-K) (1 - prev)
PPP_{rw} = 1 - 1 + K + prev - K(prev)

So PPP is related to Kappa with the third "real world" assumption added in the following way:

PPP_{rw} = prev + K(1 - prev) **Equation 7.**

Appendix A provides these estimates along with the maximal PPP's developed earlier. Across varying prevalence rates, three things are discovered about real world PPP. 1) At perfect agreement, Kappa is 1.00 and positive predictive power is 1.00. 2) At chance agreement, Kappa is 0.00 and positive predictive power is equal to prevalence. And 3) positive predictive power of 0.00 is only possible with Kappa's below chance.

This formula presents the limits of validity given reliability in a Bayesian world. Positive predictive values for psychological tests can never be perfect given the imperfection of the clinician standards.

Discussion

Two estimations of positive predictive power have been developed given levels of Kappa and prevalence. The first is a truly maximal estimation and is heavily constrained by the assumption of perfect test psychometrics. No test would ever be able to match these numbers. This figure may be of use in a ratio approach to the validity of tests. If indeed this number is the best possible figure given Kappa and a prevalence, then perhaps it should serve as the denominator in a ratio statistic describing the ability of a test. For example, if the prevalence rate is 0.05 for a particular study with an underlying Kappa of 0.40, the maximal PPP would be 0.72 (Appendix A). If the test has a PPP of 0.36, the ratio of obtained PPP to maximal PPP would be 0.50. This ratio indicates that the test has achieved 50% of the possible and available positive predictive power. Depending upon the situation this may be considered adequate.

In developing this formula, however, an interesting development occurred. It was necessary to partial the error variance in Kappa to the two judges. As such, a single judge Kappa was developed. This concept may be a more realistic estimate of a single judge's ability to diagnose conditions. Original Kappa essentially models the error in an agreement situation involving two judges and in so doing does something of a disservice to each individual judge.

The second is an estimation which is achievable by many psychological tests. This second estimation should be considered the goal of tests which attempt to make dichotomous predictions of judgments. By way of example using psychiatric diagnoses and the MCMI, if the Kappa for Compulsive personality disorder is 0.25 in the DSM, a scale such as the MCMI-III Compulsive Personality disorder will probably only achieve a positive predictive power of about 0.29 at a disorder prevalence of 0.05. In other words, the scale will only accurately identify 29% of those who score above the cut score. This is due to the unreliability of the clinician judgments plus the probable level of test psychometrics. This estimation is important in that it allows for appropriate expectation levels of psychological tests. Tests of this type are doing "well" if they achieve that 0.29 PPP. Positive predictive powers of 0.80 or 0.90 are unlikely and should not be expected. Indeed, positive predictive powers well above these estimates should be suspect.

While the current work has focused on the use of psychological tests to predict psychiatric diagnoses, the current formulae may be used in any situation where tests are used to predict judgments. In the military, tests are used to predict who will make a good officer, who should be selected for training, who should be dropped from training, who should be promoted, who should be retained, and a large number of other situations. While some of these judgments may seem very objective such as who fails out of training, the fact of the matter is that it is always a judgment who has failed and who

should continue. Tests and check rides may seem to offer objective measures of ability but the grades are judgments. As such, most military decisions are more like the diagnosis of a psychiatric disturbance than most realize.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders.* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders.* (3rd rev. ed.). Washington, DC: Author.

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

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1991). *MMPI-2 manual for administration and scoring.* Minneapolis, MN: University of Minnesota Press.

Craig, R. J. (1993). *Millon Clinical Multiaxial Inventory: A clinical research information synthesis.* Hillsdale, New Jersey: Lawrence Erlbaum & Associates.

Gibertini, M., Brandenburg, N. & Retzlaff, P. (1986). The operating characteristics of the Millon Clinical Multiaxial Inventory. *Journal of Personality Assessment, 50,* 554-567.

King, R. E. (1994). Assessing aviators for personality pathology with the Millon Clinical Multiaxial Inventory (MCMI). *Aviation, Space, and Environmental Medicine, 65,* 227-231.

Millon, T. (1977). *Millon Clinical Multiaxial Inventory.* Minneapolis: National Computer systems.

Millon, T. (1981). *Disorders of personality.* New York: Wiley.

Millon, T. (1987). *Millon Clinical Multiaxial Inventory-II: Manual for the MCMI-II.* Minneapolis: National Computer Systems.

Millon, T. (1990). *Toward a new personology.* New York: Wiley.

Millon, T. (1994). *Manual for the MCMI-III.* Minneapolis: National Computer Systems.

Nunnally, J. C. (1978). *Psychometric Theory.* New York: McGraw-Hill.

Retzlaff, P. (1996). MCMI-III Validity: Bad test or bad validity study. *Journal of Personality Assessment, 66,* 431-437.

Retzlaff, P. (ed.) (1995). *Tactical psychotherapy of the personality disorders: An MCMI-III based approach.* Allyn & Bacon: Needham Heights, MA.

Retzlaff, P. & Gibertini, M. (1994). Neuropsychometric issues and problems. In Vanderploeg, R. (ed.), *Clinician's Guide to Neuropsychological Assessment.* Hillsdale, NJ: Erlbaum.

Retzlaff, P. D., & Gibertini, M. (1987). Air Force pilot personality: Hard data on the "right stuff." *Multivariate Behavioral Research, 22,* 383-399.

Retzlaff, P. D. & Gibertini, M. (1988). The objective psychological testing of Air Force officers in pilot training. *Aviation, Space, and Environmental Medicine, 59,* 661-663.

Suen, H. K. (1990). *Test Theories.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Williams, B. T. (1982). *Computer aids to clinical decisions.* Boca Raton, FL: CRC Press.

Appendix

Maximum and "real world" positive predictive powers at various prevalences and kappa's.

PREVALENCE:	.05		.10		.15	
KAPPA	PPP_{max}	PPP_{rw}	PPP_{max}	PPP_{rw}	PPP_{max}	PPP_{rw}
1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.95	0.98	0.96	0.98	0.95	0.98	0.96
0.90	0.95	0.91	0.95	0.91	0.96	0.91
0.85	0.93	0.86	0.93	0.86	0.94	0.87
0.80	0.91	0.81	0.91	0.82	0.91	0.83
0.75	0.88	0.77	0.89	0.77	0.89	0.79
0.70	0.86	0.72	0.86	0.73	0.87	0.74
0.65	0.84	0.67	0.84	0.68	0.85	0.70
0.60	0.81	0.62	0.82	0.64	0.83	0.66
0.55	0.79	0.58	0.80	0.59	0.81	0.62
0.50	0.76	0.53	0.77	0.55	0.79	0.57
0.45	0.74	0.48	0.75	0.50	0.77	0.53
0.40	0.72	0.43	0.73	0.46	0.74	0.49
0.35	0.69	0.39	0.71	0.41	0.72	0.45
0.30	0.67	0.34	0.68	0.37	0.70	0.40
0.25	0.65	0.29	0.66	0.32	0.68	0.36
0.20	0.62	0.24	0.64	0.28	0.66	0.32
0.15	0.60	0.20	0.62	0.23	0.64	0.28
0.10	0.57	0.15	0.59	0.19	0.62	0.23
0.05	0.55	0.10	0.57	0.14	0.60	0.19
-.00	0.53	0.05	0.55	0.10	0.57	0.15
-.05	0.50	0.00	0.53	0.05	0.55	0.10
-.10	.	.	0.50	0.01	0.53	0.06
-.15	.	.	.	.	0.51	0.02