Paper 177-30

Using SAS® to implement proportional odds ratio model for meta-analysis of diagnostic tests

Mir S Siadaty, Jianfen Shu, University of Virginia, Charlottesville, VA

ABSTRACT

This work shows how we can use SAS® system to implement the proportional odds ratio (POR) model (Siadaty 2004). POR is used to pool results of several studies together in a meta-analysis of diagnostic tests. It generalizes comparison of competing tests when there are two or more tests available for disease and each test has been studied in one or more papers and each paper might have studied more than one test. The POR uses odds ratio (OR) as performance measure, and makes no assumptions about the shape of OR_0, a baseline function capturing the way OR changes across papers. The POR merely specifies a relationship between the ORs of two tests instead of assuming homogeneity of ORs.

In this presentation, we will briefly introduce the POR model. The focus will be on how to use SAS/STAT on Windows to fit the model for a real example, and estimate the discrimination accuracy of competing tests. We will show some useful graphics to capture the numerical results. Furthermore, we will show SAS MACRO codes for converting estimated ORs into other measures of test performance like predictive values, and post-test probabilities.

The flexibility of POR model and its generalized applicability, coupled with easy implementation in familiar software like SAS®, suits the daily practice of meta-analysis and improves clinical decision-making. If you are a SAS® user with medium level experience and with statistics modeling knowledge, then this is for you.

INTRODUCTION

For a particular disease there may be several diagnostic tests invented, where each of the tests is subject of one or more studies. One may want to combine all such studies to see how the competing tests are performing with respect to each other, and choose the best for clinical practice.

To pool several studies and estimate a summary statistic some assumptions are made. One such assumption is that differences seen between individual study results are due to chance (sampling variation). Equivalently, this means all study results are reflecting the same "true" effect (L'Abbe 1987). However, meta-analysis of studies for some diagnostic tests show that this assumption, in some cases, is not empirically supported. This difference in test performance across studies may be due to differences in study design, patient population, case difficulty, type of equipment, abilities of raters, and dependence of OR on threshold chosen (Rutter 1995). One solution is to relax the assumption that every study is pointing to the same value and to accept explicitly that different studies may correctly give "different" values for performance of the same test.

Another concern is when comparing two (or more) diagnostic tests, where each study reports results on more than one test, the performance statistics (in the study results) are correlated. Then standard errors computed by summary ROC (Toledano 1995) are invalid.

The proportional odds ratio (POR) model was proposed to relax the "OR-homogeneity" assumption. The model accommodates complex missing patterns, and accounts for correlated results. In the model, when comparing two (or more) tests, each test has its own trend of ORs across studies, while the trends of two tests are (assumed to be) proportional to each other, the "proportional odds ratio" assumption. Besides accounting better for between-study variation, we show how to use the POR model to "explain why" such variation exists. This potentially gives valuable insights and may have direct clinical applications. It may help define as to when, where, how, and on what patient population to use which test, to optimize performance. In the following section we show brief formulation of the POR model. For details, please refer the paper Siadaty 2004.

Using odds ratio (OR) as measure of diagnostic discrimination accuracy, consider the model

$$OR_i(Paper) = OR_0(Paper) * \exp(\beta_i) \quad , \; i = 1, 2, \ldots, k$$

where i is an index for the k diagnostic tests, and *Paper* is a categorical variable representing the studies included in the analysis. OR_0 is a function capturing the way OR changes across papers. Then to compare two diagnostic tests

$$OR_i(Paper) / OR_j(Paper) = \exp(\beta_i - \beta_j)$$

where the ratio of the two ORs depends only on the difference between the effect estimates of the two tests, and is independent of the underlying OR_0 and *Paper*. Thus the model makes no assumptions about the shape of OR_0 (and in particular homogeneity of ORs) but merely specifies a relationship between the ORs of the two tests.

Consider the following logistic regression model fitted to the grouped binary data of the 2-by-2 tables (the cell counts).

(1) $logit(Result_{pt}) = \beta_0 + \beta_1 * Disease_{pt} + \beta_2 * PaperID_{pt} + \beta_3 * Disease_{pt} * PaperID_{pt} + \beta_4 * X_{pt} + \beta_5 * Disease * X_{pt} + \beta_6 * TestID_{pt} + \beta_7 * Disease_{pt} * TestID_{pt}$

where Result is an indicator for positive test result (depending on software choice, for grouped binary data, usually Result is replaced by number of positive test results over the total sample size, for each group); Disease is an indicator for actual presence of disease, ascertained by the gold standard; PaperID is a categorical variable for papers included in the meta-analysis; and TestID is a categorical variable for tests included; and X represents variables one is interested to measure their effect on the performance of the tests. Regression coefficients β_2 to β_7 can be vector valued, meaning having several components, so the corresponding categorical variables should be represented by suitable number of indicator variables in the model. Indexes p and t signify paper p and test t. They define the repeated measures structure of the data. The main effect of paper, β_2, allows baseline odds to vary for each paper (otherwise the model assumes the odds for healthy people are the same across all the papers, effectively summing the four cell counts across all papers); while the interactions β_3 and β_5 allow LOR of the test to be different for each paper and dependent on covariate X. In the case X is a test-level categorical variable, using main effect of X rather than paper, may give more degrees of freedom to the baseline odds.

One can derive log-odds-ratio (LOR) of tests from the previous logistic model. Substituting Disease with 1 and 0 in two separate rounds in the model (2), one subtracts the two resulting models. Then some terms are omitted and it simplifies to

(2) $LOR_{pt} = \beta_1 + \beta_3 * PaperID_{pt} + \beta_5 * X_{pt} + \beta_7 * TestID_{pt}$

Note that the model assumes difference between LOR of two tests are the same across all papers (the POR assumption), but does not assume the OR of a test is the same for all papers (no homogeneity-of-ORs assumption). When there is no perceived referent test to which the other tests are to be compared (especially when there are more than two tests), a "deviation from means" coding is preferred for the TestID (Hosmer 1989). Using deviation coding, the $\beta_1 + \beta_3 * PaperID_{pt} + \beta_5 * X_{pt}$ would define overall average LOR across all the tests, while components of β_7 show amount of deviation of each test from this overall mean LOR, and whether the test is significantly better or worse than the overall mean. Hence, by using deviation contrast (rather than simple contrast), even if components of β_3 , β_5, or β_7 are statistically significant, β_1 is interpretable and can be reported as overall average LOR (but not the 'common' LOR). In model (1), β_2 allows different odds bases (for when Disease=0) per paper, while β_3 *allows different ORs per paper,* so relaxing the OR-homogeneity assumption. These parameters are surrogates for reasons of heterogeneity between studies. In their presence, effects of covariates X_{pt} may not be estimable.

Among the approaches for modeling repeated-measures data, we use generalized estimating equations to estimate marginal logistic regression (Diggle 2002). Software is widely available for estimation of parameters of a marginal POR model. These include SAS® (genmod procedure), R (function geese), and STATA (command xtgee).
For the implementation of POR model, we have focused on LOR. However, once the LORs have been estimated by the model, one can translate the LOR to other test performance measures, such as area-under-curve, predictive values, post-test probabilities, and likelihood ratios.

IMPLEMENTATION OF POR MODEL BY SAS®

AN EXAMPLE

We use a recent meta-analysis of diagnostic tests for deep vein thrombosis (DVT) by Heim et al. (Heim 2004). DVT is a common condition where blood clots are formed in veins that are situated deep inside legs. It has significant morbidity and mortality if not diagnosed and treated in a timely manner. The clinical signs and symptoms of DVT are nonspecific and objective testing is required for diagnosis. D-dimer is a compound in blood that is directly linked to formation of such clots. Measurement of D-dimer is the basis for several diagnostic tests of DVT. In their meta-analysis, Heim et al selected 23 papers. Each paper studied from one to 13 different tests. Table 2 shows the distribution of studied tests per paper.

This is a rather irregular and incomplete dataset. In this meta-analysis there are 23 papers and 21 tests, comprising 483 potential performance measurements, while only 66 are actually observed, thus 86% of cells are not measured. The papers overlap partially in the types of tests they studied. In other words, not every test has been studied in every paper. In addition, the number of studied tests per paper varies. There were papers that studied only one test (and tests that were studied by only one paper).

For this dataset (DATA=DVT) there are three variables to be entered in the model as covariates or confounders (X's). They are type of gold standard used by a paper (Gold) as different studies used different gold standard, clinical setting (inpatient or outpatient) in which the patients were recruited (Setting), and the observed prevalence of the disease for each paper (Prevalence). All the three are study-level variables, with the last one being continuous.

Table 1. Tests Studied in Each Paper

		Diagnostic Test																					
		Asserachrom	Auto Dimertest	BC D-Dimer	D-Dimer test	Dimertest	Dimertest EIA	Dimertest GOLD EIA	Dimertest II	Enzygnost	Fibrinostika	IL Test	Instant I.A.	Liatest	LPIA	Minutex	Nephelotex	NycoCard	SimpliRED	Tinaquant	Turbiquant	VIDAS	TOTAL
	Paper	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
1	Brenner, B. 1995 (86)						X		X										X				3
2	D'Angelo, A. 1996 (103)																					X	1
3	Elias, A. 1996 (171)	X			X					X	X		X					X				X	7
4	Escoffre-Barbe, M. 1998 (464)													X									1
5	Farrell, S. 2000 (48)																		X				1
6	Fiessinger, J. 1997 (30)																		X				1
7	Janssen, M. 1997 (132)					X					X					X			X	X		X	6
8	Legnani, C. 1997 (81)							X					X		X	X	X	X				X	7
9	Legnani, C. 1999 (99)			X				X														X	3
10	Lennox, A. 1999 (200)																		X				1
11	Leroyer, C. 1997 (448)	X											X										2
12	Scarano, L. 1997 (126)									X			X					X					3
13	van der Graaf, F. 2000 (99)	X		X						X	X	X	X	X		X		X	X	X	X	X	13
14	Wells, P. 1999 (150)																		X				1
15	Wells, P. 1995 (214)																		X				1
16	Funfsinn, N. 2001 (106)	X	X																	X		X	4
17	Harper, P. 2001 (235)											X							X				2
18	Carter, C. 1999 (199)																		X				1
19	Permpikul, C. 2000 (65)																		X				1
20	Sadouk, M. 2000 (177)																			X	X		2
21	Wijns, W. 1998 (74)	X											X									X	3
22	Kharia, HS. 1998 (79)																	X					1
23	Perrier, A. 1999 (474)																					X	1
	TOTAL	5	1	2	1	1	1	2	1	3	3	2	6	2	1	3	1	5	11	4	2	9	66

Numbers in parentheses in front of the paper names are sample sizes for the corresponding paper.

MARGINAL EFFECT MODEL

We fitted the marginal logistic regression of (1).

```
/*Using TRANSREG to compute design matrix for deviation contrasts, for main effects,
and for interactions.*/

proc transreg data=dvt design;
model class(paper gold setting test/deviations zero='1' '1' '1' '1' cprefix =1 4 7 1)
identity(prvlnc / tstandard=center name=(cprevalence));
id n result disease prvlnc cprvlnc;
output out=dvtno1;
run;

/*The above Transreg centralizes the prevalence, but it counts each paper to the
number of tests studied by that paper.  Fitting the marginal model by GEE:*/

proc genmod data=dvtno1 ;
class paper test disease gold setting;
model result/n = disease p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18
p19 p20 p21 p22 p23 disease*p2 disease*p3 disease*p4 disease*p5 disease*p6 disease*p7
```

```
disease*p8 disease*p9 disease*p10 disease*p11 disease*p12 disease*p13 disease*p14
disease*p15 disease*p16 disease*p17 disease*p18 disease*p19 disease*p20 disease*p21
disease*p22 disease*p23 gold2 disease*gold2 setting2 disease*setting2 cprevalence
disease*cprevalence t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19
t20 t21 disease*t2 disease*t3 disease*t4 disease*t5 disease*t6 disease*t7 disease*t8
disease*t9 disease*t10 disease*t11 disease*t12 disease*t13 disease*t14 disease*t15
disease*t16 disease*t17 disease*t18 disease*t19 disease*t20 disease*t21
/ dist=bin link=logit;
repeated subject=paper / within=test*disease type=exch covb orrw;
output out=dvtrsdl resraw=rawr reschi=pearsonr;
run;
```

Here shows the abbreviated output:

```
                                The GENMOD Procedure

                                  Model Information

 Data Set                          WORK.DVTNO1
 Distribution                         Binomial
 Link Function                           Logit
 Response Variable (Events)           RESULT    RESULT
 Response Variable (Trials)                N    N
 Observations Used                       132
 Number Of Events                       3125
 Number Of Trials                       9168

                              Class Level Information

Class    Levels    Values

PAPER    23        1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
                   21 22 23
TEST     21        1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
                   21
DISEASE   2        0 1
GOLD      2        1 2
SETTING   2        1 2

                              GEE Model Information

        Correlation Structure                          Exchangeable
        Within-Subject Effect              TEST*DISEASE (42 levels)
        Subject Effect                            PAPER (23 levels)
        Number of Clusters                                       23
        Correlation Matrix Dimension                             42
        Maximum Cluster Size                                     26
        Minimum Cluster Size                                      2

                  Analysis Of GEE Parameter Estimates
                   Empirical Standard Error Estimates

                                        Standard   95% Confidence
Parameter        Estimate    Error      Limits            Z Pr > |Z|

Intercept         -2.5300   0.0735  -2.6741  -2.3858  -34.40   <.0001
DISEASE        0   2.4890   0.0365   2.4175   2.5606   68.18   <.0001
DISEASE        1   0.0000   0.0000   0.0000   0.0000     .      .
P2                 1.2717   0.4254   0.4379   2.1055    2.99   0.0028
.
.
P23                0.3181   0.4254  -0.5158   1.1519    0.75   0.4547
P2*DISEASE     0  -0.6713   0.3212  -1.3009  -0.0418   -2.09   0.0366
P2*DISEASE          1   0.0000   0.0000   0.0000   0.0000     .      .
.
.
```

```
P23*DISEASE          0  -0.1385   0.3212  -0.7680   0.4911   -0.43   0.6664
P23*DISEASE          1   0.0000   0.0000   0.0000   0.0000      .       .
GOLD2                    0.0000   0.0000   0.0000   0.0000      .       .
GOLD2*DISEASE        0   0.0000   0.0000   0.0000   0.0000      .       .
GOLD2*DISEASE        1   0.0000   0.0000   0.0000   0.0000      .       .
SETTING2                 0.0000   0.0000   0.0000   0.0000      .       .
SETTING2*DISEASE     0   0.0000   0.0000   0.0000   0.0000      .       .
SETTING2*DISEASE     1   0.0000   0.0000   0.0000   0.0000      .       .
cprevalence              0.0000   0.0000   0.0000   0.0000      .       .
cprevalence*DISEASE  0   0.0000   0.0000   0.0000   0.0000      .       .
cprevalence*DISEASE  1   0.0000   0.0000   0.0000   0.0000      .       .
T2                      -1.4606   0.2775  -2.0044  -0.9168   -5.26   <.0001
.
.
T21                  -1.7863   0.4358  -2.6404  -0.9322   -4.10   <.0001
T2*DISEASE       0    0.2223   0.1882  -0.1466   0.5912    1.18   0.2376
T2*DISEASE       1   0.0000   0.0000   0.0000   0.0000      .       .
.
.
T21*DISEASE      0   1.0037   0.3259   0.3650   1.6424    3.08   0.0021
T21*DISEASE      1   0.0000   0.0000   0.0000   0.0000      .       .
```

Table 2. Parameter Estimates for Test Effects

Coefficient		Test	Deviation*	95% Confidence Limits	p value**
β_7 †	1	Asserachrom	0.524	0.2293, 0.8186	0.0005
	2	Auto Dimertest	0.222	-0.1466, 0.5912	0.2376
	3	BC D-Dimer	-0.993	-2.4195, 0.4333	0.1724
	4	D-Dimer test	0.225	0.1, 0.3494	0.0004
	5	Dimertest	-2.092	-2.3392, -1.8439	<.0001
	6	Dimertest EIA	-0.929	-1.1756, -0.6825	<.0001
	7	Dimertest GOLD EIA	-0.193	-0.4784, 0.0935	0.1871
	8	Dimertest II	-0.731	-0.9774, -0.4843	<.0001
	9	Enzygnost	0.399	0.1209, 0.6766	0.0049
	10	Fibrinostika	0.857	0.6865, 1.0266	<.0001
	11	IL Test	0.809	0.0914, 1.5256	0.0271
	12	Instant I.A.	0.558	0.216, 0.9006	0.0014
	13	Liatest	-0.143	-0.3375, 0.0511	0.1486
	14	LPIA	0.182	-0.0354, 0.3997	0.1007
	15	Minutex	-0.323	-0.8394, 0.193	0.2197
	16	Nephelotex	0.654	0.4325, 0.8745	<.0001
	17	NycoCard	-0.797	-1.0434, -0.5506	<.0001
	18	SimpliRED	0.393	0.1467, 0.6398	0.0018
	19	Tinaquant	0.703	0.0113, 1.3948	0.0464
	20	Turbiquant	-0.328	-1.6596, 1.0032	0.629
	21	VIDAS	1.004	0.365, 1.6424	0.0021
β_1		Overall LOR	2.489	2.4175, 2.5606	< .0001***

* estimate of deviation from overall LOR

** p-value for null hypothesis of Deviation = 0

*** p-value for null hypothesis of LOR=0

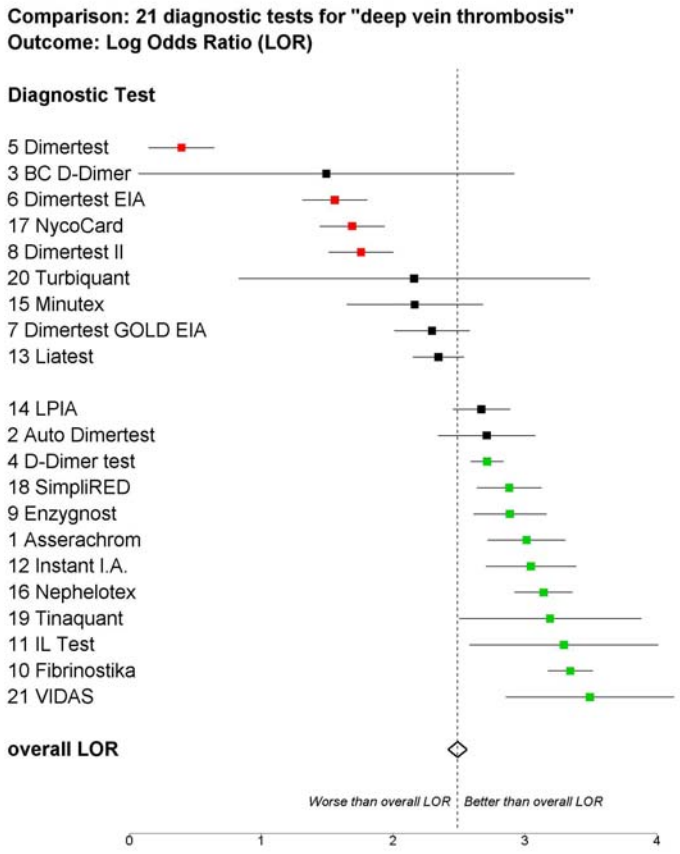
† $LOR(Result_{pt}) = \beta_1 + \boldsymbol{\beta_3}*PaperID_{pt} + \boldsymbol{\beta_5}*X_{pt} + \boldsymbol{\beta_7}*TestID_{pt}$

Since we have used deviation contrast for the variables, estimate of β_1 is the "overall mean" for the log-OR. This is similar to an ANOVA analysis where the overall mean is estimated by the model. Therefore the average OR is equal to exp(2.489) = 12.049. Components of β_7 estimate deviation of LOR of each test from the overall LOR. Software gives estimates of SEs, plus confidence intervals and p-values, so inference is straightforward.

Variables Gold and Setting were parameterized as deviation contrasts, while variable Prevalence was centered on zero. However, parameters of the three X variables, the β_5 in the model, were not estimable. This is due to over-parameterization of the model. In other words, in the design matrix the columns that represent those variables were linearly dependent on other columns, so the software omitted them from the estimation process.

A forest plot can be used to present the results of the modeling in a graphical way. This may connect better with clinically oriented audience. In Figure 4 we have sorted the 21 tests based on their LOR estimate.

Figure 4. Comparing Performance of each Diagnostic Test to the Overall LOR



The horizontal axis is log-OR, representing test performance. The dashed vertical line shows overall mean LOR. For each diagnostic test the solid square shows the LOR, while the horizontal line shows the corresponding CI. If the horizontal line does not intersect the vertical line, the test is significantly different from the overall mean LOR.

Next step, one may choose a specific test as RefCat (like the test with biggest favorable deviation from overall LOR), and use "simple" parameterization, to answer the question "for disease DVT, what tests are included in the equivalence-class of best diagnostic tests?" (However, note the design of the meta-analysis was not of an equivalence-trial.) After VIDAS, tests Fibrinostika, IL-Test, Tinaquant, Nephelotex, Instant I.A., Asserachrom, Enzygnost, SimpliRED, and Turbiquant are performing in a descending order.

SAS® code to compare tests to a good RefCat (a test with good LOR), like VIDAS.

```
proc transreg data=dvt design;
model
class(test / zero='21' cprefix = 1)
class( paper gold setting /deviations zero='1' '1' '1' cprefix = 1 4 7)
identity(prvlnc / tstandard=center name=(cprevalence));
id n result disease prvlnc cprvlnc;
output out=dvtno1;
```

```
run;

proc genmod data=dvtno1 ;
class paper test disease gold setting;
model result/n = disease t1 t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17
t18 t19 t20 disease*t1 disease*t2 disease*t3 disease*t4 disease*t5 disease*t6
disease*t7 disease*t8 disease*t9 disease*t10 disease*t11 disease*t12 disease*t13
disease*t14 disease*t15 disease*t16 disease*t17 disease*t18 disease*t19 disease*t20
p2 p3 p4 p5 p6 p7 p8 p9 p10 p11 p12 p13 p14 p15 p16 p17 p18 p19 p20 p21 p22 p23
disease*p2 disease*p3 disease*p4 disease*p5 disease*p6 disease*p7 disease*p8
disease*p9 disease*p10 disease*p11 disease*p12 disease*p13 disease*p14 disease*p15
disease*p16 disease*p17 disease*p18 disease*p19 disease*p20 disease*p21 disease*p22
disease*p23 gold2 disease*gold2 setting2 disease*setting2 cprevalence
disease*cprevalence / dist=bin link=logit;
repeated subject=paper / within=test*disease type=exch covb corrw;
output out=dvtrsdl resraw=rawr reschi=pearsonr;
run;
```

Needs to change line class (test / zero='1' cprefix = 1) such that zero='' of the category number one wants to be RefCat. Also needs to change lines t2 t3 t4 t5 t6 t7 t8 t9 t10 t11 t12 t13 t14 t15 t16 t17 t18 t19 t20 t21 disease*t2 disease*t3 disease*t4 disease*t5 disease*t6 isease*t7ch that the omitted RefCat name matches with these lines in Genmod. Then beta of Disease would be LOR of that particular test (with its CI), and beta of Test are deviation of the rest of tests from RefCat test, and whether significant.

Table 3. Parameter estimate for test effects comparing with VIDAS.

Test	Deviation*	95% Confidence Limits	p value**
5 Dimertest	-3.0952	-3.8397, -2.3507	<.0001
17 NycoCard	-1.8007	-2.3916, -1.2098	<.0001
6 Dimertest EIA	-1.9328	-2.676, -1.1895	<.0001
8 Dimertest II	-1.7345	-2.4778, -0.9913	<.0001
15 Minutex	-1.3269	-2.002, -0.6518	0.0001
2 Auto Dimertest	-0.7814	-1.2875, -0.2753	0.0025
7 Dimertest GOLD EIA	-1.1962	-1.9729, -0.4194	0.0025
13 Liatest	-1.1469	-1.9406, -0.3531	0.0046
3 BC D-Dimer	-1.9968	-3.5341, -0.4595	0.0109
4 D-Dimer test	-0.779	-1.4037, -0.1543	0.0145
14 LPIA	-0.8215	-1.6111, -0.0319	0.0414
9 Enzygnost	-0.6049	-1.3182, 0.1083	0.0965
20 Turbiquant	-1.3319	-2.9143, 0.2506	0.099
18 SimpliRED	-0.6104	-1.3537, 0.1328	0.1074
1 Asserachrom	-0.4797	-1.1662, 0.2067	0.1708
12 Instant I.A.	-0.4454	-1.306, 0.4152	0.3104
16 Nephelotex	-0.3502	-1.143, 0.4426	0.3866
19 Tinaquant	-0.3007	-1.1175, 0.5162	0.4707
10 Fibrinostika	-0.1471	-0.8139, 0.5197	0.6654
11 IL Test	-0.1952	-1.279, 0.8886	0.7241

* estimate of deviation from LOR of test VIDAS

** p-value of difference from LOR=0

One can translate LOR to other measures of test performance. There are numerous types of these measures. We provide code to convert the LOR estimated by the POR model to such measures. Note that almost all of them, unlike LOR, are in pairs. This means in order to compare two tests, one needs to use two numbers to represent each single

test. For example, sensitivity-specificity is a pair. If a test has a higher sensitivity than the other test, while having a lower specificity, it is not immediately clear which test is better. Also, note that some performance measures are independent of disease prevalence, while others depend on prevalence. This means the same test would perform differently for populations with different disease prevalence. Below is a SAS® MACRO for the conversion.

```
%macro convert (LORdat, LOR, prev, sens, spec);
data newdat;
     set &LORdat;
     **compute OR based on LOR;
     OR=exp(&LOR);
     **compute sens based on spec;
     if &sens=. then %do;
            &sens=OR*(1-&spec)/&spec;
            &sens=&sens/(1+&sens);
            %end;
     **compute spec based on sens;
     if &spec=. then %do;
            &spec=OR*(1-&sens)/&sens;
            &spec=&spec/(1+&spec);
            %end;
     ppv=1/ (1+ ((1-&prev)/&prev) * (1/ (OR-&sens*(OR-1))) );
     npv=1/ (1+ (&prev/(1-&prev)) * (1/ (OR-&spec*(OR-1))) );
     lrab= &sens/(1-&spec);
     lrnr = (1-&sens)/&spec;
     preodds = &prev/(1-&prev);
     postoddsab = preodds * lrab;
     postoddsnr = preodds * lrnr;
     postpab = postoddsab/(postoddsab+1);
     postpnr = postoddsnr/(postoddsnr+1);

     /** homogeneous AUC
      based on Walter SD. Properties of the summary receiver operating characteristic
     (SROC)curve for diagnostic test data. Stat Med. 2002 May 15; 21(9):1237-56. MID:
     12111876; */
     h.auc <- (OR*(OR-1-log(OR))) / ((OR-1)^2);*/
run;
%mend convert;

**example;
data test;
     LOR=log(20.3);
     prev=0.4;
     sens=.;
     spec=.9;
run;

%convert(test, lor, prev, sens, spec);
```

Note in order to compute some of the performance measures, one needs to assume a prevalence and sensitivity or specificity. We assumed a disease prevalence of 40%, and a specificity of 90%, for Table 6, as the tests are mainly used for ruling out the DVT.

Table 4. Other Performance Measures for the 21 Diagnostic Tests of DVT

	Diagnostic Test	Diagnostic Odds Ratio	Prevalence	Specificity	Sensitivity	AUC (assuming homogeneous OR)	Positive Predictive Value	Negative Predictive Value	Likelihood Ratio For Abnormal Test	Likelihood Ratio For Normal Test	Pre Test Odds	Post Test Odds Of Abnormal Test	Post Test Odds Of Normal Test	Post Test Probability Of Abnormal Test	Post Test Probability Of Normal Test
1	Asserachrom	20.3	0.4	0.9	0.693	0.888	0.822	0.815	6.933	0.341	0.667	4.622	0.227	0.822	0.185
2	Auto Dimertest	15.0	0.4	0.9	0.626	0.864	0.807	0.783	6.258	0.416	0.667	4.172	0.277	0.807	0.217
3	BC D-Dimer	4.5	0.4	0.9	0.332	0.732	0.688	0.669	3.315	0.743	0.667	2.210	0.495	0.688	0.331
4	D-Dimer test	15.1	0.4	0.9	0.626	0.865	0.807	0.783	6.263	0.415	0.667	4.175	0.277	0.807	0.217
5	Dimertest	1.5	0.4	0.9	0.142	0.566	0.486	0.611	1.419	0.953	0.667	0.946	0.636	0.486	0.389
6	Dimertest EIA	4.8	0.4	0.9	0.346	0.741	0.697	0.674	3.459	0.727	0.667	2.306	0.485	0.697	0.326
7	Dimertest GOLD EIA	9.9	0.4	0.9	0.525	0.826	0.778	0.740	5.248	0.528	0.667	3.499	0.352	0.778	0.260
8	Dimertest II	5.8	0.4	0.9	0.392	0.766	0.723	0.689	3.920	0.676	0.667	2.613	0.450	0.723	0.311
9	Enzygnost	18.0	0.4	0.9	0.666	0.879	0.816	0.802	6.661	0.371	0.667	4.440	0.247	0.816	0.198
10	Fibrinostika	28.4	0.4	0.9	0.759	0.910	0.835	0.849	7.592	0.268	0.667	5.061	0.178	0.835	0.151
11	IL Test	27.0	0.4	0.9	0.750	0.907	0.833	0.844	7.503	0.277	0.667	5.002	0.185	0.833	0.156
12	Instant I.A.	21.1	0.4	0.9	0.701	0.890	0.824	0.818	7.006	0.333	0.667	4.671	0.222	0.824	0.182
13	Liatest	10.4	0.4	0.9	0.537	0.831	0.782	0.745	5.371	0.514	0.667	3.581	0.343	0.782	0.255
14	LPIA	14.5	0.4	0.9	0.616	0.861	0.804	0.779	6.163	0.426	0.667	4.109	0.284	0.804	0.221
15	Minutex	8.7	0.4	0.9	0.492	0.813	0.766	0.727	4.921	0.564	0.667	3.281	0.376	0.766	0.273
16	Nephelotex	23.2	0.4	0.9	0.720	0.897	0.828	0.828	7.202	0.311	0.667	4.801	0.207	0.828	0.172
17	NycoCard	5.4	0.4	0.9	0.376	0.758	0.715	0.684	3.763	0.693	0.667	2.509	0.462	0.715	0.316
18	SimpliRED	17.9	0.4	0.9	0.665	0.878	0.816	0.801	6.648	0.372	0.667	4.432	0.248	0.816	0.199
19	Tinaquant	24.3	0.4	0.9	0.730	0.900	0.830	0.833	7.300	0.300	0.667	4.867	0.200	0.830	0.167
20	Turbiquant	8.7	0.4	0.9	0.491	0.812	0.766	0.726	4.909	0.566	0.667	3.273	0.377	0.766	0.274
21	VIDAS	32.9	0.4	0.9	0.785	0.918	0.840	0.863	7.851	0.239	0.667	5.234	0.159	0.840	0.137

RANDOM EFFECTS

One may use a non-linear mixed effects modeling approach on the cell-count data for estimation of parameters of the POR model. The Paper effect is declared as random, and interaction of the random effect with Disease is included in the model, as indicated in model (1). However, such mixed effects non-linear models are hard to converge, especially for datasets where there are many papers studying only one or a small number of the included tests (such as the dataset presented as example in this paper). If the convergence is good, it may be possible to fit a mixed model with the interaction of Disease, Test, and the Paper random effect. Such model relaxes the POR assumption, besides relaxing the assumption of OR-homogeneity. In other words, one can use the model to quantitatively test the POR assumption. **One should understand that the interpretation of LOR estimate from a marginal model is of a population-average, while that of a mixed model is a conditional-average**. Therefore there is a slight difference in their meaning. In this presentation, we won't show the random effect model.

CONCLUSION

When collecting papers from biomedical literature for a particular meta-analysis of a few diagnostic tests, it is hard to assume OR-homogeneity and hard to come up with a complete square dataset, where every paper has included all the tests of interest. Usually the dataset contains missing values, and a case-wise deletion of papers with missing tests means a lot of data is thrown away. A method of analysis that can utilize incomplete matched groups may be helpful. The POR model allows complex missing patterns in data structure. SAS® procedures TRANSREG and GENMOD make the implementation of POR model easy.

The flexibility of POR model and its generalized applicability, coupled with ease with which it can be estimated in familiar software, suits the daily practice of meta-analysis and improves clinical decision-making.

REFERENCES

Diggle P, Heagerty P, Liang KY, Zeger S: *Analysis of Longitudinal Data.* New York: Oxford University Press; 2002.

Heim SW, Schectman JM, Siadaty MS, Philbrick JT. D-dimer testing for deep venous thrombosis: a metaanalysis. *Clin Chem* 2004, 50(7):1136-47.

Hosmer DW, Lemeshow S: Applied Logistic Regression. New York: Wiley-Interscience; 1989.

Ihaka R, Gentleman RR: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 1996, 5: 299-314.

L'Abbe KA, Detsky AS, O'Rourke K: Meta-analysis in clinical research. Ann Intern Med 1987, 107:224-33.

Rutter CM, Gatsonis, CA: Regression methods for meta-analysis of diagnostic test data. *Acad Radiol* 1995, 2:S48-S56.

Siadaty MS, Shu j: Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. BMC Medical Research Methodology 2004, 4:27.

Toledano A, Gatsonis CA: Regression analysis of correlated receiver operating characteristic data. Acad Radiol 1995, 2:S30-S36.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Mir S Siadaty
University of Virginia
POBOX 800717
Charlottesville, VA 22908
Work Phone: 434-982-4436
Email: MirSiadaty@virginia.edu