For ... Presentation Only

MASTER

CONF-751179--1

Observer Performance in Detecting Multiple Radiographic Signals: Prediction and Analysis Using a Generalized ROC Approach[1]

Charles E. Metz, Ph.D.,

Stuart J. Starr, B.S.,

and

Lee B. Lusted, M.D.

Presented at the 61st Scientific Assembly and Annual Meeting of the Radiological Society of North America in joint sponsorship with the American Association of Physicists in Medicine, Chicago, Illinois, 30 Nov - 5 Dec, 1975.

[1] From the Center for Radiologic Image Research, Department of Radiology, The University of Chicago, and the Franklin McLean Memorial Research Institute (operated by the University of Chicago for the U.S. Energy Research and Development Administration), Chicago, Illinois 60637.

This work was supported in part by Grant GM-18940 from the USPHS.

Observer Performance in Detecting Multiple Radiographic Signals: Prediction and Analysis Using a Generalized ROC Approach[1]

Charles E. Metz, Ph.D.,
Stuart J. Starr, B.S.,
& Lee B. Lusted, M.D.

ABSTRACT

The theories of decision processes and signal detection provide a framework for the evaluation of observer performance. Some radiologic procedures involve a secrch for multiple similar lesions, as in gallstone or pneumoconiosis examinations. A model is presented which attempts to predict, from the conventional receiver operating characteristic (ROC) curve describing the detectability of a single visual signal in a radiograph, observer performance in an experiment requiring detection of more than one such signal. An experiment is described which tests the validity of this model for the case of detecting the presence of zero, one, or two low-contrast radiographic images of a two-mm.-diameter lucite bead embedded in radiographic mottle. Results from six observers, including three radiologists, confirm the validity of the model and suggest that human observer performance for relatively complex detection tasks can be predicted from the results of simpler experiments.

---

[1] From the Center for Radiologic Image Research (USPHS Grant GM-18940) and the Department of Radiology, The University of Chicago (C.E.M., S.J.S., L.B.L.) and The Franklin McLean Memorial Research Institute (operated by the University of Chicago for the U.S. Energy Research and Development Administration) (C.E.M.), Chicago, Illinois.

A comprehensive analysis of medical imaging systems must include an evaluation of human perceptual capabilities for detection of the image features, or signals, which are relevant to diagnosis. Recently, decision theory and signal detection theory have been applied to the quantitative assessment of observer performance in radiology. The early work of our group was directed primarily toward establishing the applicability of conventional Receiver Operating Characteristic (ROC) analysis to medical imaging. We are currently exploring ways in which the classical ROC approach can be generalized to take into account clinically relevant detection tasks somewhat more complex than those usually employed in psychophysical experiments. For example, we recently reported an extension of ROC analysis to the combined task of detection and localization of a radiographic signal. In the study which I will describe today, we have derived and confirmed an ROC approach to assessing observer performance in a detection experiment in which more than one radiographic signal may be present.

In the classic signal detection experiment, an observer views a series of samples, some of which contain a signal-plus-noise and the others of which contain noise alone. The observer decides for each trial whether or not the signal is present on the basis of whether or not his confidence about the presence of the signal exceeds a certain level. Thus, the frequencies of true-positive and false-positive responses depend upon the confidence threshold adopted by the observer. The underlying detectability of the signal does not, however, depend upon the confidence threshold. The conventional ROC curve, an example of which is shown in the first slide, is a plot of the conditional true-positive decision frequency (or probability) versus the conditional false-positive decision frequency (or probability) as confidence threshold is varied.

We wanted to develop a model which would relate the conventional ROC curve for the detectability of a single radiographic signal to observer performance in an experiment requiring detection of more than one such signal. Our derivation proceeds in the following manner. Consider the radiograph to be divided into a large number of contiguous, congruent subregions. From previous theory and experiment, we know the relationship between the ROC curves of radiographs of different sizes, all other things being the same. Let us call this result the "ROC area effect". Thus, if we know the conventional ROC curve measured for the full field-of-view of the radiograph, we can compute the conventional ROC curve for a field-of-view which is the size of a subregion. Assume that the observer examines each subregion independently. Next, we argue that the observer's response concerning the full field of the radiograph is determined by the combination of decisions which he makes regarding the possible presence of a signal in each of the subregions. For example, consider the case of a radiograph containing two signals. There are a number of combinations of decisions which would lead the observer to respond that he has detected two signals. For instance, if the observer detects the signals in the two subregions which actually contain the signals <u>and</u> if he does not falsely detect a signal in any of the other subregions which in fact contain noise alone, he would respond that he had found two signals. He would make the same response if, for example, he fails to detect a signal in either of the two subregions which actually contain a signal and he falsely detects a signal in two of the subregions which in fact contain noise alone. And so on. In this manner, we formulate a given conditional response probability by summing all of the joint conditional subregion response probabilities which result in the given overall response. ((Slide 2)) By applying a change of variables according to the ROC area effect and after appropriate algebraic manipulations, we obtain the expression shown in the slide for the probability that the observer detects k signals, given that l signals are actually present. P(N|n) and P(N|s) are respectively the conditional true-negative and false-negative decision probabilities in the one-signal detection experiment. These two parameters are simply related to the axes of conventional ROC space.

We proceeded to test our model for the relatively simple case in which the observer knows that zero, one, or two signals may be present in the radiographs. Since there are three possible stimuli and three possible responses, there are three-by-three, or nine, types of correct and incorrect decisions. The predictions for the nine possible conditional response probabilities are shown in the next slide. ((Slide 3))

The notation used, for example, in Equation (6), "P sub 012 of 2 given 1", indicates the probability of the observer responding that two signals are present, given that one signal is actually present in an experiment in which either zero, one, or two signals are known to be present. P(N|n) and P(N|s) ((Slide 4)) are simply related to the conventional ROC axes as shown here in the top two equations. Since the probabilities of the three possible responses, given a particular number of signals present, must add to unity, which is stated in the bottom equation, we have three additional equations which reduce the number of degrees-of-freedom from nine to six. Therefore, the "generalized ROC curve" for the "012" experiment ((Slide 5)) should be a line in six-dimensional space.

Although a six-dimensional curve cannot be drawn, the generalized ROC curve can be represented by a set of two-dimensional projections. In presenting our experimental data, which will be shown momentarily, we have chosen to plot eight two-dimensional graphs in which eight of the conditional response probabilities, those in Equations (2) through (9), are plotted individually against the remaining one, $P_{012}(0|0)$. This approach is partially redundant, but will emphasize the fact that nine kinds of decisions are possible.

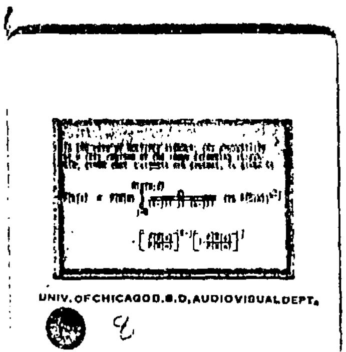
With attention to proper psychophysical methods and controls, we designed and conducted an experiment to test our predictions. We made a number of radiographs with images of 2-mm-diameter Lucite beads on RP/R film, using a Par Speed screen in a vacuum cassette. 40 samples contained two bead images, 40 contained a single image, and another 40 had radiographic mottle alone. We obtained data from six observers, consisting of three radiologists and three medical physicists. Two tasks were performed by each observer.

In the first task, we measured the conventional ROC curve of each observer by using the eighty radiographs which contained either one bead image or noise alone. ((Slide 6)) The slide shows the results of this task for the six observers. A standard five-category rating method was used. The task was repeated at least twice by every observer, the different data symbols representing separate sessions. A smooth curve was drawn through the data points in each case. Note that the data are quite reproducible.

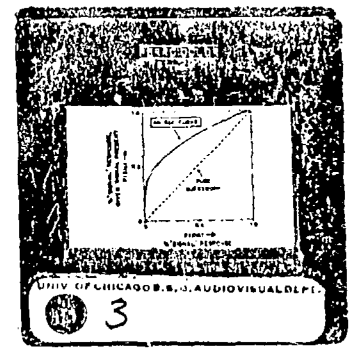
The second task performed by these same observers involved the use of all of the radiographs; that is to say, those containing zero, one, or two bead images. The approach which we have developed to measure the generalized ROC curve in a multiple-signal detection task represents a generalization of the rating method which is used to measure an ROC curve in simple detection tasks, as in the first part of the experiment. By requiring the observer to state two ratings for each radiograph, and by unfolding these responses, we obtain empirical measurements of the six degrees-of-freedom of the generalized ROC curve which describes this task.

((Slide 7)) The next slide shows the results for observer B, which are typical of all the observers. These are the eight two-dimensional projections of the generalized ROC curve, as explained earlier. Each observer performed the task at least twice, the different data symbols indicating separate sessions. The curves are predictions which are generated by substituting points on this observer's conventional ROC curve, which was shown on the previous slide, into our theoretical expressions. Note that the curves are not fit to the data points but are predictions. ((Slide8)) The final slide shows the results of observer F who repeated the task six times in order to provide a sense of the reproducibility of the data.

Although we do not know of a statistical test for goodness-of-fit which is applicable to the problem of comparing predicted with observed detection performance in the multiple-signal detection task, examination of the graphs indicates that the curves based on our theoretical model certainly predict the trends of the experimental results and indeed suggest that the quantitative relationships are predicted quite well. We believe that this demonstrated relationship between observer performance in the simple detection task and in the multiple-signal detection task is clinically relevant since some diagnostic medical imaging situations may require the radiologist to count the number of lesions present, such as the search for gallstones in cholecystography, metastases in brain scintigraphy, or pneumokoniosis lesions in chest radiography. The results of this study and the results of our previous study on the combined task of detection and localization suggest that human observer performance in relatively complex detection tasks can be predicted from the results of simpler experiments. The conventional ROC curve thus appears to provide a description of observer performance which is meaningful in clinical detection situations more complex than that in which it is measured. We also feel that this study represents a contribution to the fundamental theory of signal detection with quite general applicability.



SLIDE 2



SLIDE 1

$$P_{012}(0|0) = P_{01}(N|n) \qquad (1)$$

$$P_{012}(1|0) = -P_{01}(N|n)\, ln\, P_{01}(N|n) \qquad (2)$$

$$P_{012}(2|0) = 1 - P_{01}(N|n)[1 - ln\, P_{01}(N|n)] \qquad (3)$$

$$P_{012}(0|1) = P_{01}(N|s) \qquad (4)$$

$$P_{012}(1|1) = P_{01}(N|n) - P_{01}(N|s)[1 + ln\, P_{01}(N|n)] \qquad (5)$$

$$P_{012}(2|1) = 1 - P_{01}(N|n) + P_{01}(N|s)\, ln\, P_{01}(N|n) \qquad (6)$$

$$P_{012}(0|2) = [P_{01}(N|s)]^2 / P_{01}(N|n) \qquad (7)$$

$$P_{012}(1|2) = 2\,P_{01}(N|s) - \{[P_{01}(N|s)]^2 / P_{01}(N|n)\}\{2 + ln\, P_{01}(N|n)\} \qquad (8)$$

$$P_{012}(2|1) = 1 - 2\,P_{01}(N|s) + \{[P_{01}(N|s)]^2 / P(N|n)\}\{1 - ln\, P_{01}(N|n)\} \qquad (9)$$

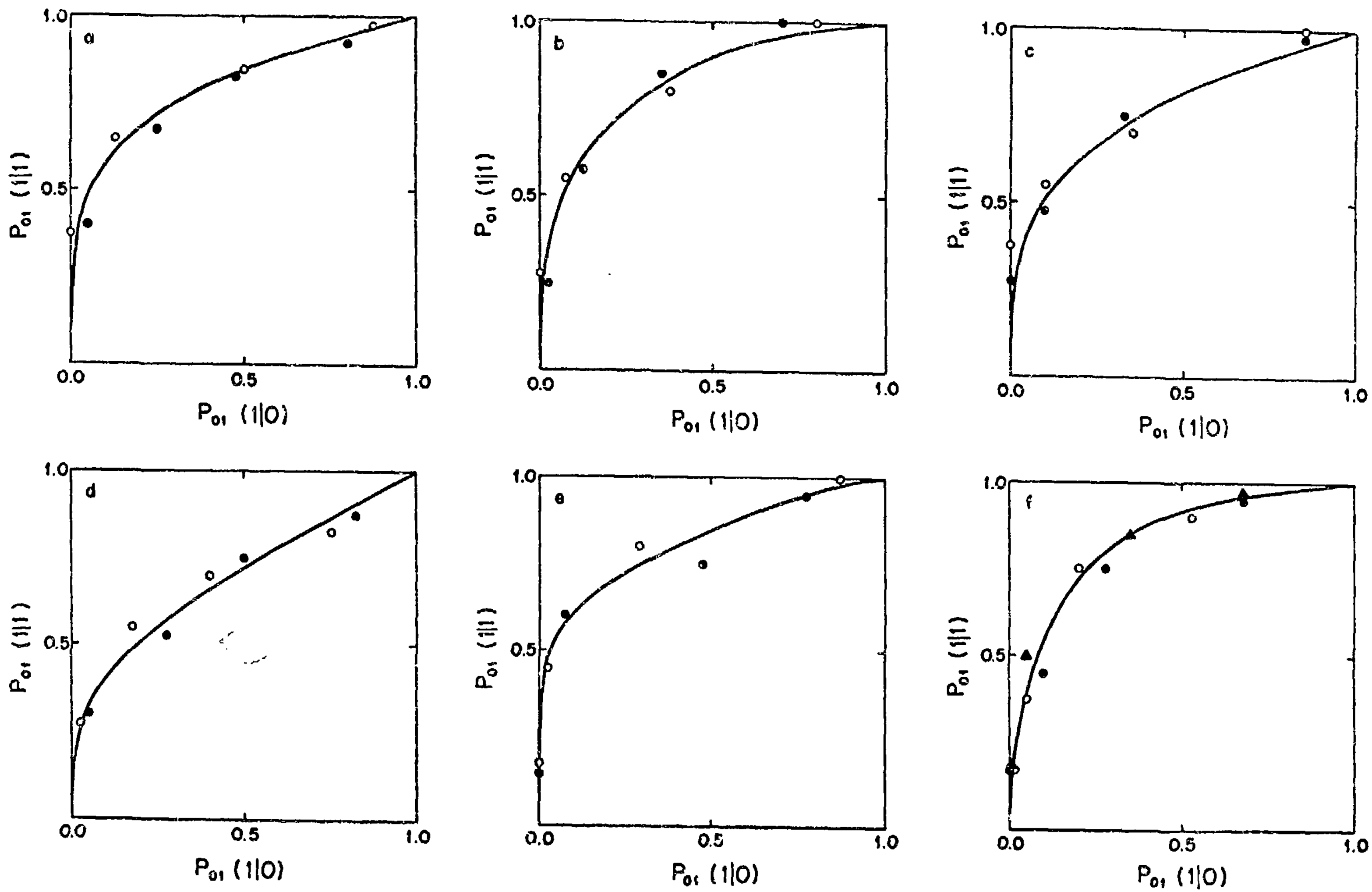
SLIDE 3 & SLIDE 5

$$P_{01}(N|n) = 1 - P_{01}(S|n)$$

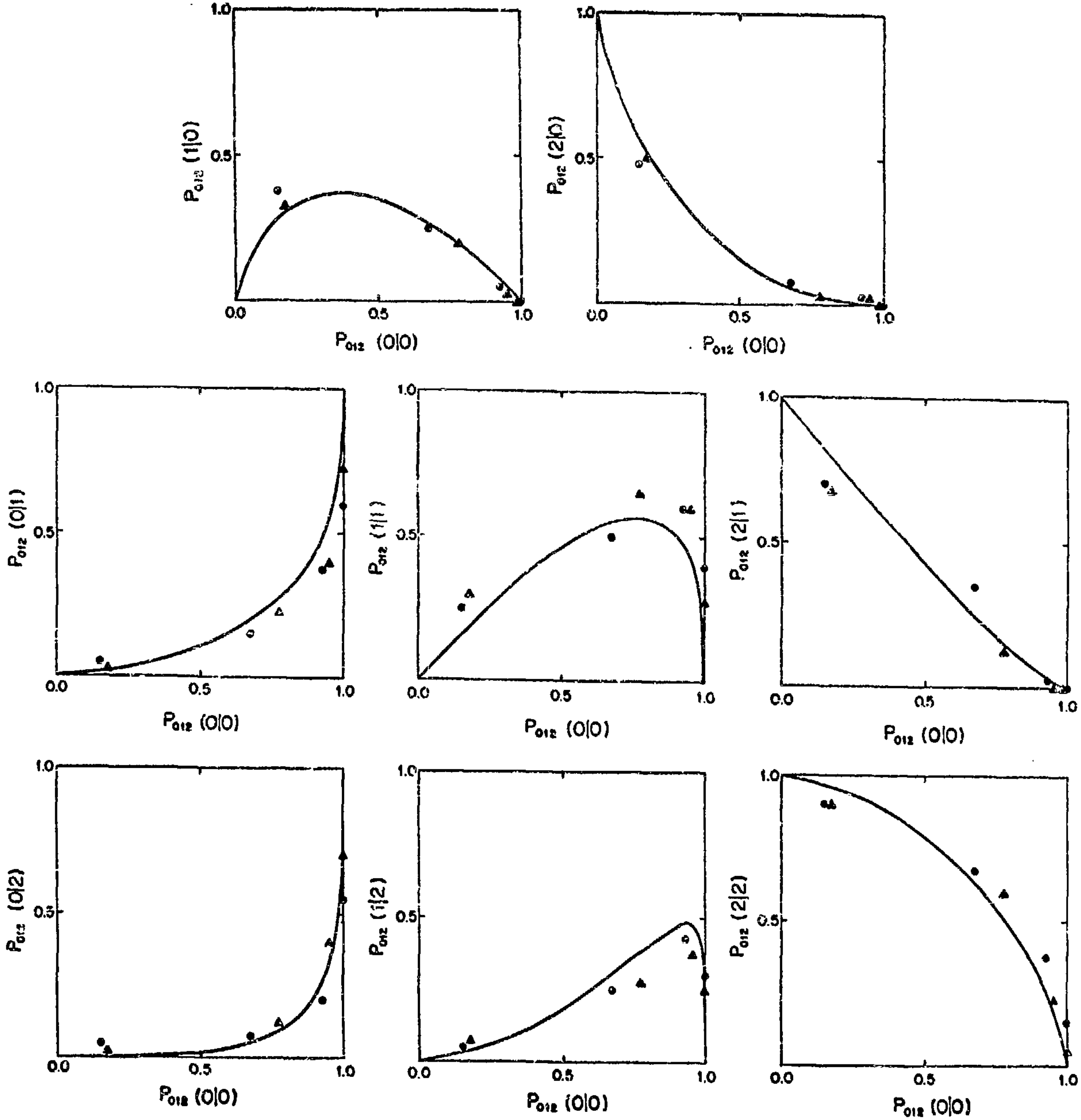
$$P_{01}(N|s) = 1 - P_{01}(S|s)$$

$$\sum_{k=0}^{2} P_{012}(k|l) = 1 \quad \text{for } l = 0,1,2$$

SLIDE 4

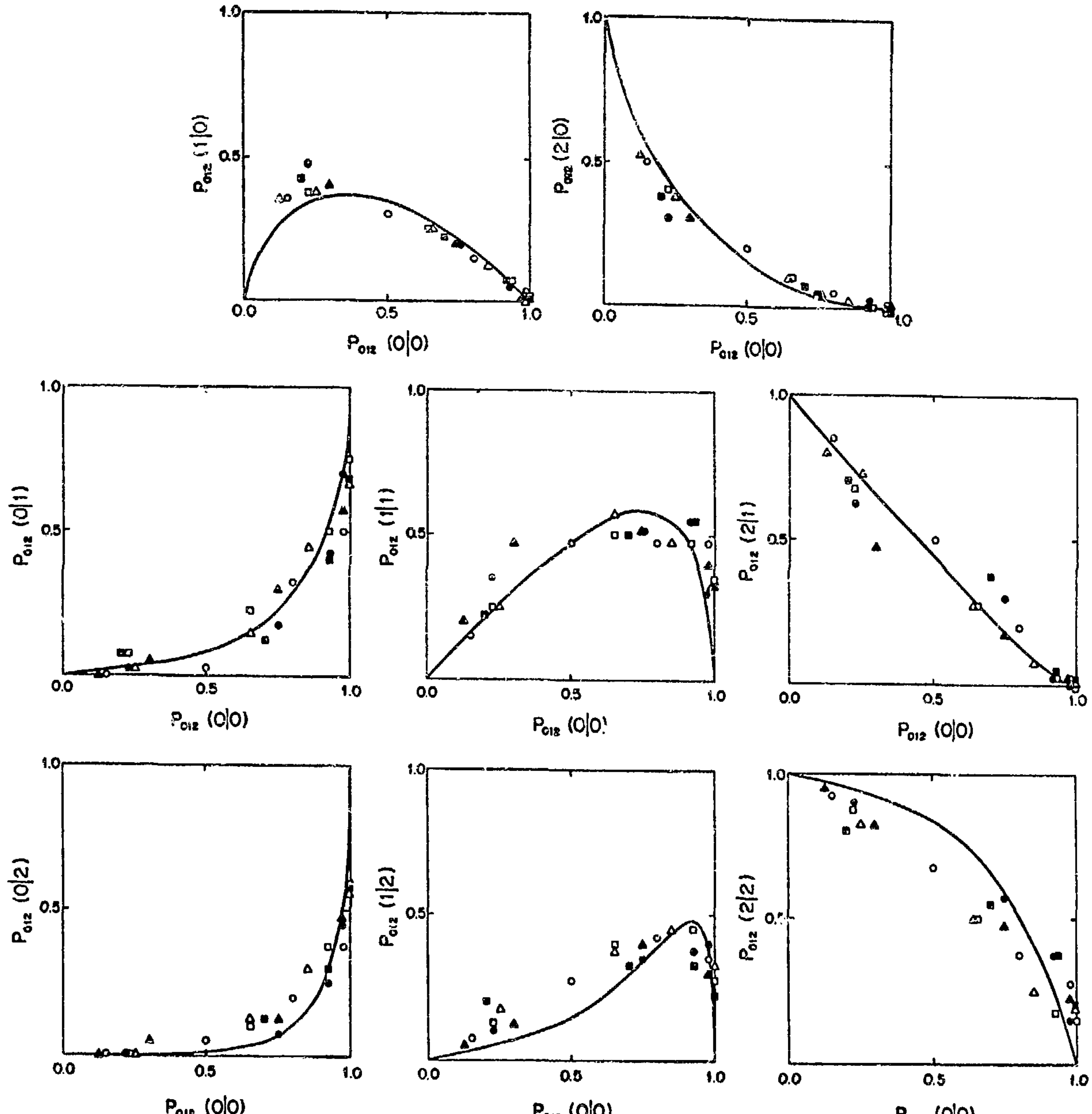


SLIDE 6



P_{012} (1|0)
P_{012} (2|0)
P_{012} (0|1)
P_{012} (1|1)
P_{012} (2|1)
P_{012} (0|2)
P_{012} (1|2)
P_{012} (2|2)
P_{012} (0|0)
0.0
0.5
1.0

SLIDE 7



P012 (1|0)
P012 (2|0)
P012 (0|1)
P012 (1|1)
P012 (2|1)
P012 (0|2)
P012 (1|2)
P012 (2|2)
P012 (0|0)
0.0
0.5
1.0

SLIDE 8