BIVARIATE RANDOM EFFECTS AND HIERARCHICAL META-ANALYSIS OF SUMMARY RECEIVER OPERATING CHARACTERISTIC CURVE ON FINE NEEDLE ASPIRATION CYTOLOGY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS
OF
THE MIDDLE EAST TECHN0ICAL UNIVERSITY

BY

İDİL ERTE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF MEDICAL INFORMATICS

SEPTEMBER 2011

Approval of the Graduate School of Informatics

_____

Prof. Dr. Nazife BAYKAL
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

_____

Assist. Prof. Dr. Didem GÖKÇAY
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

_____ _____

Assoc. Prof. Dr. Mehtap AKÇİL
Co-Supervisor

Prof. Dr. Nazife BAYKAL
Supervisor

Examining Committee Members

Prof. Dr. Ergun KARAAĞAOĞLU (Hacettepe, TIP)_____

Prof. Dr. Nazife BAYKAL (METU, II)_____

Assoc. Prof. Dr. Mehtap AKÇİL (Başkent , FEF)_____

Assist. Prof. Dr. Yeşim AYDIN SON (METU, II)_____

Assist.Prof.Dr.Vilda PURUTÇUOĞLU (METU, FBE)_____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name and Surname : İdil ERTE

Signature :

ABSTRACT

BIVARIATE RANDOM EFFECTS AND HIERARCHICAL META-ANALYSIS OF SUMMARY RECEIVER OPERATING CHARACTERISTIC CURVE ON FINE NEEDLE ASPIRATION CYTOLOGY

ERTE, İdil

M.Sc., Department of Medical Informatics
Supervisor: Prof. Dr. Nazife BAYKAL
Co-Supervisor: Assoc. Prof. Dr. Mehtap AKÇİL

September 2011, 73 pages

In this study, meta-analysis of diagnostic tests, Summary Receiver Operating Characteristic (SROC) curve, bivariate random effects and Hierarchical Summary Receiver Operating Characteristic (HSROC) curve theories have been discussed and

accuracy in literature of Fine Needle Aspiration (FNA) biopsy that is used in the diagnosis of masses in breast cancer (malignant or benign) has been analyzed. FNA Cytological (FNAC) examination in breast tumor is, easy, effective, effortless, and does not require special training for clinicians. Because of the uncertainty related to FNAC's accurate usage in publications, 25 FNAC studies have been gathered in the meta-analysis. In the plotting of the summary ROC curve, the logit difference and sums of the true positive rates and the false positive rates included in the meta-analysis's codes have been generated by SAS. The formula of the bivariate random effects model and hierarchical summary ROC curve is presented in context with the literature. Then bivariate random effects implementation with the new SAS PROC GLIMMIX is generated. Moreover, HSROC implementation is generated by SAS PROC HSROC NLMIXED. Curves are plotted with RevMan Version 5 (2008). It has been stated that the meta-analytic results of bivariate random effects are nearly identical to the results from the HSROC approach. The results achieved through both random effects meta-analytic methods prove that FNA Cytology is a diagnostic test with a high level of distinguish over breast tumor.

Keywords: Meta-Analysis, Summary Receiver Operating Characteristic Curve, Diagnostic Tests, Fine Needle Aspiration Cytology, Breast Cancer

ÖZ

İNCE İĞNE ASPİRASYON SİSTOLOJİ'SİNİN İKİ DEĞİŞKENLİ RASGELE ETKİ MODELİNE GÖRE META-ANALİZİ'NİN ÖZET İŞLEM KARAKTERİSTİĞİ EĞRİSİ VE HİYERARŞİK ÖZET İŞLEM KARAKTERİSTİĞİ EĞRİSİ

ERTE, İdil

Yüksek Lisans, Tıp Bilişimi
Tez Yöneticisi: Prof. Dr. Nazife BAYKAL
Ortak Tez Yöneticisi: Doç. Dr. Mehtap AKÇİL

Eylül 2011, 73 sayfa

Bu çalışmada, tanı testlerinin meta-analizi, Özet İşlem Karakteristiği Eğrisi (SROC), İki değişkenli rasgele etki modeli ve Hiyerarşik Özet İşlem Karakteristiği Eğrisi (HSROC) teorisi anlatılmış ve meme kitlelerinin (malign ya da belign), meme kanserinin tanısında kullanılan, İnce İğne Aspirasyon (FNA) biyopsisinin

literatürdeki doğruluğu incelenmiştir. Meme tümöründe, İnce İğne Aspirasyon Sitolojik (FNAC) muayenesi, kolay, elverişli, etkili, zahmetsiz ve klinisyenlerin eğitilmesini gerektirmemektedir. İnce İğne Aspirasyon Sitoliji'sinin (FNAC) yayınlardaki doğruluğuna ilişkin belirsizlik nedeniyle, 25 FNAC çalışması meta-analizine dahil edilmiştir. Özet işlem karakteristiğinin oluşturulmasında; meta-analizine dahil edilen çalışmaların doğru pozitif oranları ve yanlış pozitif oranlarının lojit fark ve toplamları SAS programıyla kodları yazılmıştır. İki değişkenli rasgele etki modeli ve Hiyerarşik Özet İşlem Karakteristiği Eğrisi (HSROC) yöntemleri tanıtılmış ve bu modellerin parametre tahminleri SAS PROC GLIMMIX ve HSROC NLMIXED ile hesaplanmıştır. SROC Eğrileri, RevMan Version 5 (2008) ile çizdirilmiştir. Sonuç olarak, iki değişkenli rasgele etki modeli ve Hiyerarşik Özet İşlem Karakteristiği Eğrisi (HSROC) sonuçları yaklaşık olarak aynı bulunmuştur. İki farklı meta-analitik yöntemden elde edilen sonuçlar gösteriyor ki FNA Sitoliji'si yüksek ayırıcılık gücüne sahip bir tanı testidir.

Anahtar Kelimeler: Meta-Analizi, Özet İşlem Karakteristiği Eğrisi, Tanı Testi, İnce İğne Aspirasyon Biyopsisi, Meme Kanseri

To My Mother Kudret & My Father Erhan Erte

ACKNOWLEDGMENTS

I am greatly appreciative to *my Supervisor Prof. Nazife BAYKAL* for her support throughout my study and the grand influence on me which has helped me to decide that this sector, being an academician in a university, is my first preference for my career.

I am also deeply thankful to *my Co-Supervisor Assoc. Prof. Mehtap AKÇİL for* her guidance throughout my study and the first-hand opportunity to work with a great academician. They were all great at sharing their knowledge and experiences.

It is also to *my mother Kudret and my father Erhan, my cousins especially to Gökçe,* addition to *my whole family.*

I would like to express my deepest gratitude to *my friend*s; *Özlem, Barbaros, Bilge, Burcu, Tuğba, Bengi, Begüm*, *Zeynep* and *Pınar*. Also, I would like to express my deepest gratitude to *my teachers; Umut, Özlem, İsmail and Timur*. They were always with me and ready to support me all time.

I am thankful to the *Staff of Informatics Institute* for their helps in every stage of the bureaucratic tasks. I am also grateful to my thesis committee for their suggestions and valuable comments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ROC: Receiver Operating Characteristic

FPR: False Positive Rate

TPR: True Positive Rate

SROC: Summary Receiver Operating Characteristic

FNA: Fine Needle Aspiration

FNAC: Fine Needle Aspiration Cytology

HSROC: Hierarchical Summary Receiver Operating Characteristic

DOR: Diagnostic Odds Ratio

LOR: Logarithmic Odds Ratio

GLMMIX: Generalized Linear Mixed Model

NLMIXED: Nonlinear Mixed Model

CHAPTER 1

INTRODUCTION

This study is on bivariate random effects meta-analysis of Receiver Operating Characteristic curve and Hierarchical Summary Receiver Operating Characteristic curve on Fine Needle Aspiration Cytology (FNAC). In this study, these terms will be explained and an outline of the thesis will be given.

In medical research, generalizing the results of a sampling study to population is usually impossible due to lack of time, money, staff and patients (Normand, 1999).

Moreover, Normand (1999) stated that most of the studies on the same subject display inconsistency and incompatibility among the results due to biological variation.

In order to address the aforementioned problems, meta-analysis method was developed in 1976 (Glass, 1976) and its usage has increased sharply since then. Meta-analysis differs from the traditional review by including both medical and statistical approaches in the method (Yach, 1990).

Meta-analysis of Receiver Operating Characteristic (ROC) curve data is usually plotted with fixed-effects models, which have drawbacks. To present a method that addresses the shortcomings of the fixed effects summary ROC (SROC) method, Littenberg and Moses (1993), proposed random-effects model to execute a meta-analysis of ROC curve data.

In this study, sensitivities and specificities are analyzed using a bivariate random-effects model and Hierarchical Summary Receiver Operating Characteristic curve. The analyses are carried out by developing code in the software package SAS (PROC NLMIXED and PROC GLIMMIX).

1.1 Design of the study

A meta-analysis study starts with a well-structured problem or an organized planning. After that, literature should be researched through all relevant databases (MEDLINE, PUBMED, etc.). These studies are selected through the inclusion or exclusion criteria that the researchers provide. All the studies' related parameters and variables included in the meta-analysis study should be demonstrated in a table. One of the many meta-analysis methods can be selected and used in the study. In this study, bivariate random effects meta-analysis and Hierarchical Summary Receiver Operating Characteristic curve is used.

1.2 Purpose of the Study

Clinical and epidemiologic studies are usually done on a limited sample of population due to deficiency of practitioners, money and time. Meta-analysis aims to address such drawbacks (Glass, 1976).

In addition to meta-analysis, other methods were developed on combining results of several studies for parameter estimation which are dependent to the kind of studies and types of findings (Hasselblad, & Hedges, 1995). Combining probabilities, effectiveness indexes, correlations, and accuracy of diagnostic measurements is some of the methods used in parameter estimation (Kardaun, & Kardaun, 1990).

The purpose of this study is to assess the diagnostic characteristics of a diagnostic test with Fine Needle Aspiration Cytology (FNAC) by using meta-analysis. FNAC is a quick, and reliable technique involves inserting a very small needle into the lesion in question to aspirate cells. FNAC is safe than open surgery, in which a lesion in a variety of sites (thyroid, breast, skin etc.) can be observed (Temel, 2000). FNAC's

purpose is to distinguish patients having a certain breast cancer with the final diagnosis.

Consistency of the FNAC on the breast palpable will be observed in the literature by Summary Receiver Operating Characteristic, bivariate random effects and Hierarchical Summary Receiver Operating Characteristic curve with SAS programming software. Both model codes are generated in SAS PROC NLMIXED and PROC GLIMMIX.

Furthermore, meta-analysis can provide consortium decision upon subjects such as FNAC (Borenstein, Hedges, Higgins, & Rothstein, 2009).

1.3 Significance of the Study

Clinicians should decide whether to use the diagnostic test or how to interpret the results. Literature review can be applied in order to support these decisions. Unfortunately, diagnostic test is usually compared with the same reference test, and results are often inconsistent (Temel, 2000).

In literature, there are studies on FNAC on breast palpable inconsistence with each other (Arends, Hamza, Van Houwelingen, Heijenbrok-Kal, Hunink, & Stijnen, 2008). These inconsistencies can be confusing for a clinician deciding whether to use FNAC on a health related subject or not. In order to assess and support such decisions, meta-analysis of the FNAC data are carried out.

Meta-analysis that assesses the reliability, accuracy, and impact of diagnostic tests are vital to guide test selection and the interpretation of test results (Arends et al., 2008).

Arends et al. (2008) note that bivariate random-effects approach not only extends the SROC approach but also provides a similar framework for other approaches.

An alternative approach for fitting HSROC curves has been proposed by Rutter and Gatsonis (2001). It has been used to fit an ROC curve when data are available at

multiple thresholds in a single study. The models allows for asymmetry in the ROC throughout inclusion of the scale parameter which determines the shape of the ROC. Rutter and Gatsonis (2001) proposed that this model can be used in the estimation of Summary ROC curves.

Meta-analysis can be done with handy and user-friendly software packages like Meta-DiSc (Meta-DiSc, 2006) which has some drawbacks and limitations on how the estimates generates. In order to generate needed estimates, SAS code was generated.

In this study the bivariate GLIMMIX model was compared with the HSROC NLMIXED model by using the published data of 25 meta-analysis about diagnostic test accuracy.

CHAPTER 2

LITERATURE REVIEW

2.1 Diagnostic Tests

Chappell, Raab, and Wardlaw (2009) stated that a diagnostic test is any kind of medical test performed to assist the diagnosis or detection of disease so as to determine appropriate treatment, screen for disease or monitor substances such as drugs. Diagnostic test includes; diagnosing diseases, measuring the progress or convalescence and confirming that a person is free from disease (Broemeling, 2007).

Diagnostic accuracy is the capability of distinguishing the patient and the healthy individual (Broemeling, 2007). Diagnostic accuracy is evaluated with specificity and sensitivity. Diagnostic test emerged from the idea that reference test/ gold standard is hard to apply (Walter, 2002). In order to determine the test's accuracy parameters are forecasted (Table 1, Sutton, Abrams, Jones, Sheldon, & Song, 2000).

Table 1 Distribution of Reference and Diagnostic Tests.

	Disease +	***Disease -***	***TOTAL***
TEST **(positive)**	TP(A)	FP(B)	TP+FP
TEST **(negative)**	FN(C)	TN(D)	FN+TN
TOTAL	TP+FN (n_1)	FP+TN(n_2)	TP+FP+FN+TN

Table 1 shows the distribution of reference and diagnostic tests (Sutton et al., 2000);

- True Positive (TP): Diseased people correctly diagnosed as diseased.
- False Positive (FP): Healthy people incorrectly identified as diseased.
- True Negative (TN): Healthy people correctly identified as healthy.
- False Negative (FN): Diseased people incorrectly identified as healthy.

In addition to that, TPR represents the number of patients who have disease, and support this by having a TEST (positive) (whatever cutoff level is selected). FPR represents false positives (the test has tricked us, and told us that non-diseased patients are definitely diseased). Similarly, true negatives are represented by TNR, and false negatives by FNR (Broemeling, 2007).

Table 2 Conditional Probability of Reference and Diagnostic Tests.

Statement	**The name of the parameter**
P(T+ | D+)	sensitivity, True Positive Rate
P(T- | D-)	specificity, True Negative Rate
P(T+ | D-)	False positive Rate
P(T- | D+)	False Negative Rate

The sensitivity is how good the test is at discriminating patients with disease. It is simply the True Positive Rate (Littenberg, & Moses, 1993).

TPR=Sensitivity= TP/ (TP + FN) (EQUATION 1)

Specificity is the ability of the test in distinguishing patients not suffering from any disease. This is the synonymous with the True Negative Rate. (Littenberg, & Moses, 1993).

TNR=Specificity= TN/ (FP + TN) (EQUATION 2)

2.2 Receiver Operating Characteristic Curve

Receiver Operating Characteristic (ROC) Curve is plotted for displaying accuracy of diagnostic test (Krzanowski, & Hand, 2009). Chappell et al. (2009) explained that TPR and FPR are corresponded to cumulative probabilities of two related normal distributions is assumed for the models used for ROC analysis.

After the disease status of each subject is determined, TPR and FPR can be estimated at each level of this cut point and the data then can be plotted as a ROC curve. Chappell et al. (2009) explained that true values of TPR and FPR arrive from cumulative probabilities that assumed to have two normal distributions. Only two parameters are required in order to describe ROC curves due to the independence of ROC curve from the scale or location of the data.

An ideal ROC curve in Figure 1 that passes through the upper corner almost with 100% sensitivity and 100% specificity. In other words, the closer the ROC curve is to the left, the higher the overall accuracy of the test (Zweig, & Campbell, 1993).

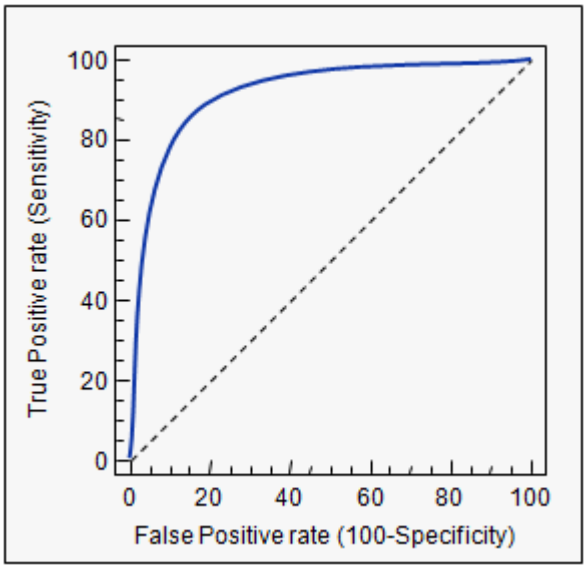


Figure 1 Ideal ROC Curve with High Accuracy

2.3 Meta-Analysis

Meta-analysis emerged from the idea that, when pooling the results of the studies for estimation of outcomes has started to fall behind (Van Houwelingen, Arends, & Stijnen, 2002).

Meta-analysis is the quantitative review and synthesis of the results of related but also independent studies (Normand, 1999). According to Arends et al. (2008), these independent studies are mostly the published ones. Meta-analysis is a tool for summarizing the results in the literature in a quantitative manner and resolve uncertainty in clinical trials. Also, it is capable of exploring the heterogeneity among study results (Arends et al., 2008).

Besides the increase in the number of fields of application, also usage of meta-analysis has been amplified throughout recent years. Nevertheless, Medical field is in the lead among other areas that the meta-analysis is used (Sutton et al., 2000).

The aim of meta-analysis is to ensure precision in parameter estimating, to contribute to the future research and to analyze the overall and common problems of all studies

included in the meta-analysis rather than individual studies' own problems (Borenstein, Hedges, Higgins, & Rothstein, 2009).

(Brand, & Kragt, 1992; Thompson, 1994; DerSimonian, & Laird, 1986) pointed out that, an analysis which ignores the heterogeneity in treatment outcome can be clinically misleading.

Along with the amplification in the usage of meta-analysis new methods also were developed (Kardaun, & Kardaun, 1990);

- Integration of probabilities
- Integration of effect size
- Combining correlations

However, Meta-analysis of diagnostic tests is a newly developed method (Hasselblad, & Hedges, 1995).

2.4 Fixed and Random Effects Meta-Analysis

Arends et al. (2008) explained that, sampling error differences in studies causes the point estimates of the effect size to differ in almost every meta-analysis. The true underlying effects are determined homogeneous after finding effect sizes differ from the sampling error alone. As a result the differences between the estimates are not systematic because of the random variation (Arends, Hoes, Lubsen, Grobbee, & Stijnen, 2000).

In the earlier years, homogeneity in true effect is assumed to be accepted through all studies and the modeling in meta-analysis is under that assumption with fixed effects model (Sutton et al., 2000).

In reality, true effects for each study differ because the variability in the effect size exceeds what expected from sampling error alone (Broemeling, 2007). DerSimonian and Laird (1986) stated that there is heterogeneity between the treatment effects in the different studies.

In this study, a model is used, called ‘random effects model’; where true effects are assumed to have a distribution, including parameters as mean and the standard deviation estimated from the data. Thompson (1994) discussed that mean parameter is the average effect. On the other hand, standard deviation parameter identifies the heterogeneity between the true effects. Meta-analysis with the univariate effect measure is started to use this model for simple cases (Normand, 1999). Although the fixed effects method is the widely used one, it ignores the potential between-trial component which can lead to misinterpretation (Thompson, & Pocock, 1991).

In such situations, random effects model is much more effective since, the trials in meta-analysis and the quantitative results are clinically heterogeneous (DerSimonian, & Laird, 1986).

2.5 Meta Regression

To combine all studies despite the heterogeneity between studies is a challenge although it enables both clinically and scientifically results to define the effects of heterogeneity overall treatment effects (DerSimonian, & Laird, 1986). The dependence of the treatment effect on characteristics as mean of gender and age can be observed by meta-regression. Since, TPR and FPR show same trends, linear regression is decided to be used with logistic transformations (Littenberg, & Moses, 1993).

— with (EQUATION 3)

p is the probability that the event occurs,

p/(1-p) is the odds ratio and

ln[p/(1-p)] is the log odds ratio

The logistic regression model is a non-linear transformation of the linear regression. Its distribution is an S-shaped distribution function which is close to the normal distribution. Also probit regression model can be used but it is easier to work with

logistic regression in health field since clinicians can also calculate the probabilities easily (Temel, 2000).

In meta-regression, dependent variable is the estimated trial treatment effect where as the covariates are determined as the trial or patient characteristics. Covariates should be specified to reduce risk of false positive results which may be the reason for the use of fixed effects regression model (Armitage, & Colton, 1998).

Meta regression is used for explaining the between-trial component of the variance with the covariates. Random effects model can be changed to fixed effects model if all the variance is explained by the covariates.

2.6 Summary ROC

The most commonly used statistical method for summarizing results of the independent studies is the Summary ROC (SROC) method submitted by Littenberg, & Moses, (1993); Moses, Shapiro, & Littenberg, (1993). The difference versus the sum of the logit (true positive rate) and logit (false positive rate) from each study is plotted. After that, regression line to these points is plotted. Lastly, the line is transformed to ROC space.

When generating a Summary ROC curve, the studies included in the meta-analysis's TPR and FPR are calculated (Littenberg, & Moses, 1993). These TPR and FPR's logit differences and sums are calculated. Linear regression slope is formed with dependent variable taken as logit difference and independent variable taken as logit sum (Sutton, 2000).

SROC allows evaluation of diagnostic test accuracy by using all possible cut points of TPR and FPR rate rather than merely one cut point (Littenberg, & Moses, 1993). By the calculated several cut points in SROC, it is decided that cut points do affect the results. If the difference is not generated by the different cut points, other analyses should be considered (Arends et al., 2008).

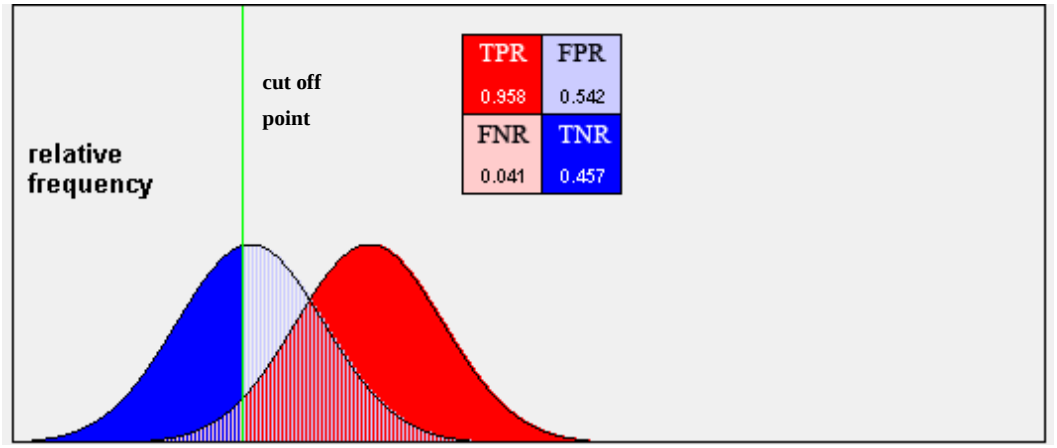


Figure 2 Graph with High TPR Value

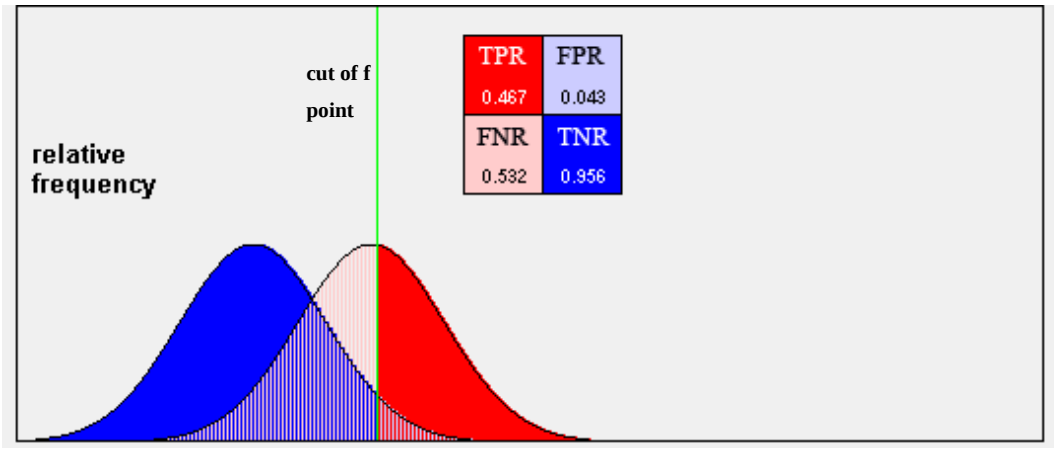


Figure 3 Graph with High TNR Value

When a disease is detected, it is important to decide on the cut point, because as the TPR increases the TNR decreases (Sutton et al., 2000). As seen in Figure 2 and Figure 3 changes in the cut point directly affect the TPR and TNR. Ideally, it should be more appropriate to assign a cut point that allows higher specificity (Sutton et al., 2000). Moses et al. (1993) stated that TPR and TNR should be closed to 1. On the other hand, FPR and FNR expected to be closer to 0.

The area under the ROC, points out the accuracy of distinguishing between healthy and diseased people. (Littenberg, & Moses, 1993)

2.7 Fine Needle Aspiration Cytology (FNAC)

“Breast cancer” refers to a malignant tumor that has developed from cells in the breast area (Breast Cancer Organization, 2011). Breast cancer is one of the dangerous diseases among all kind of diseases. The American Cancer Society's statistics for breast cancer in the United States are for 2010:

- About 210,000 new cases of breast cancer will be diagnosed in women.
- About 40,000 women will die from breast cancer

In most cases, breast cancer starts in the cells of the lobules, which are the milk-producing glands in the breast. Sometimes symptoms occur in the ducts, the passages that drain milk from the lobules to the nipple. Less commonly, breast cancer can develop in the stromal tissues, which have the fatty and fibrous connective tissues of the breast (Breast Cancer Organization, 2011).

Biopsy of breast cancer could be made via open surgery which costs time and money. On the other hand Fine Needle Aspiration is a fast, easy and a low cost operation. Moreover, patient doesn’t need anesthesia during this operation (Phui-Ly, et al. 2011). Unfortunately, the accuracy of the FNAC is needed to analyze over open surgery.

In this study, proving the accuracy of the FNAC was tried to be achieved with bivariate random effects and hierarchical meta-analysis of Summary Receiver Operating Characteristic Curve (SROC).

2.8 Data Example of FNAC

2.8.1 FNAC of the Breast

An example of bivariate random effects meta-analysis of 29 studies on FNAC (Arends et al., 2008) was reviewed in literature for comparison with bivariate random effects meta-analysis of our 25 studies.

For the examination of the breast to assess inclusive or exclusive of breast cancer, both sensitivity and specificity of FNAC were calculated for 29 study (Giard, & Hermans, 1992). The probability of shortage of abnormal cells in the patients without breast cancer is determined for specificity and the probability of a malignancy in patients with breast cancer is determined for sensitivity (Giard, & Hermans, 1992). The diagnosis of benign or malignant breast cancer frequencies of the FNAC is represented in Table 3.

Arends et al. (2008) fitted the bivariate model on the data of the 29 studies of FNAC of the meta-analysis of Giard and Hermans (1992) in Table 3. Choices of the ROC curves were presented in Table 4.

Table 3 Patients who Underwent a Fine Needle Aspiration Cytological Examination (FNAC)

	FNAC Results for Patients with Benign Disease			FNAC Results for Patients with Malignant Disease		
Study	**False Pos**	**True Neg**	**Total (n_0)**	**True Pos**	**False Neg**	**Total (n_1)**
Linsk	70	939	1009	979	89	1068
Furnival	3	163	166	51	22	73
Zajdela	55	394	949	1569	152	1721
Wilson	25	259	284	35	15	50
Thomas	4	121	125	59	12	71
Duguid	18	216	234	56	4	60
Klini	602	3117	3719	329	39	368
Gardecki	10	213	223	125	17	142
Strawbridge	88	499	587	211	63	274
Shabot	0	31	31	49	1	50
Azzarelli	26	643	669	336	178	514
Bell	147	746	893	210	42	252
Norton*	5	25	30	16	3	19
Dixon	16	356	372	258	53	311
Aretz	9	107	116	56	18	74
Ulanow	16	112	128	162	28	190
Wanebo	6	112	118	116	13	129
Wollenberg	99	145	244	65	12	77
Somers	5	78	83	94	10	104
Lannin*	0	70	70	26	4	30
Eisenberg	28	136	164	1318	249	1567
Barrows	55	539	594	569	120	689
Watson*	1	287	288	46	16	62
Hammond	13	76	89	64	6	70
Dundas	1	104	105	39	4	43
Smith*	16	426	442	132	20	152
Palombini	17	161	178	470	22	492
Langmuir*	25	200	225	28	4	32
Wilkinson*	43	22	65	42	3	45

*These studies are also included in our study of 25 meta-analysis of FNAC.

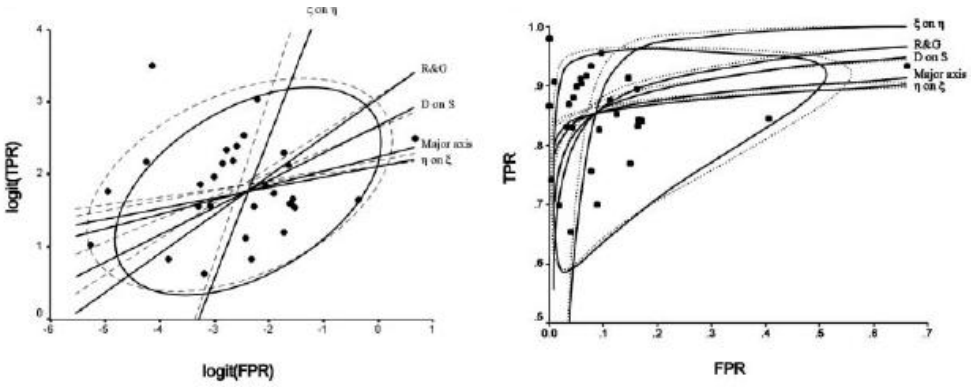
In Figures 4, the different choices of ROC curves were plotted in the logit-logit space and ROC space. Addition to that, 95% coverage regions were given in ellipse shape.

Arends et al. (2008) stated that α (cut-point) and β (threshold) calculations as seen in Table 4 and Figures 4.

Table 4 Data Example: FNAC of the Breast

	Approximate Likelihood	
Type of Summary ROC	**α (se)**	**β (se)**
1. η on ξ	2.126 (0.32)	0.148 (0.13)
2. ξ on η	6.431 (3.95)	1.954 (1.65)
3. *D* on *S*	2.680 (0.37)	0.380 (0.15)
4. Rutter and Gatsonis	3.054 (0.31)	0.537 (0.12)
5. Major axis	2.249 (0.42)	0.199 (0.17)

The five different SROC curve was chosen by Arends et al. (2008) including "η on ξ", "ξ on η", Rutter and Gatsonis, "D on S" and major axis (Figures 4).



Figures 4 Summary Receiver Operating Characteristic (SROC) Curves for the 5 Different Choices of the SROC Curve, as a Graphical Illustration of Table 4.

From the 5 types of SROC, "D on S" (logit differences of TPR and FPR on sum of their logits) was found to be the most comparable to the standard SROC by Littenberg and Moses (1993).

Also it has been declared that the ROC curves differ from each other Table 4 and Figures 4. Especially difference is obvious for "η on ξ" and "ξ on η" in Figures 4.

However, Rutter and Gatsonis, "D on S" and major axis SROC curves lied between "η on ξ" and "ξ on η" SROC curves. Arends et. al (2008) stated that Rutter and Gatsonis, "D on S" and major axis SROC curves were closer to the "η on ξ" although which is not true usually.

It has been stated that Rutter and Gatsonis is the geometric mean of the two regression lines and does not use the correlation between Sensitivity and Specificity. Although Rutter and Gatsonis curve has been stated as an alternative approach to Littenberg and Moses (1993) curve, it has not been easy to generate it.

To decide which one is preferable, Hamza, Van Houwelingen, Heijenbrok-kal, Stijnen, (2009) has stated that D on S is most popular with more than 300 times in literature, Rutter and Gatsonis almost used 20 times in literature (Hamza et al., 2009). Chu and Guo (2009) proposed that "η on ξ" is the correct one which described the median sensitivity of studies with a fixed value of the specificity. Also "ξ on η" can be chosen for describing the median specificity with a fixed specificity (Hamza, Arends, Van Houwelingen, Stijnen, 2009). Chapell and Raab (2009) argued that the major axis is the best.

In our bivariate random effects meta-analysis of 25 studies, "ξ on η" was chosen for plotting the SROC. Because variance in bivariate model explains the variation between true sensitivities on the studies having same specificities.

CHAPTER 3

METHODOLOGY

3.1 Data Collection and Meta-Analysis

Arends et al. (2008) stated that the designs of test accuracy evaluations differ from the designs of studies that calculate the effectiveness of treatments. Test accuracy proves that different criteria must be developed when evaluating quality and bias in the study. Additionally, often each evaluation of diagnostic tests reports summary statistics (sensitivity and specificity) rather than a single statistic. It requires alternative statistical methods for pooling study results (Egger, Ebrahim, & Smith, 2002).

In the past decade, several methods for meta-analysis of diagnostic tests have been developed (Littenberg, & Moses, 1993; Moses, et al., 1993; Hasselblad, & Hedges, 1995; Hellmich, Abrams, & Sutton, 1999; Walter, Irwig, Macaskill, Glasziou, & Fahey, 1995; Walter, 2002). The proposed methods depend on the type of data available. (Hellmich, et al., 1999) and (Kester, & Buntinx, 2000) used methods for individual patient data of the studies. Some methods can be used when estimated of the area under the ROC curve is calculated (McClish, 1992). Some methods are applicable to the situation when only one estimated pair of sensitivity and specificity (different diagnostic thresholds is corresponded) per study is available (Arends et al., 2008).

3.2 Summary ROC (SROC) Method

One of the methods at summarizing the results of the different studies results for observing the same diagnostic test is the plotting the Summary Receiver Operating Characteristic curve. ROC curves goal is to define the pattern of sensitivities and specificities observed when the performance of the test is assessed at different thresholds.

Meta-analysis of ROC curve data aims to provide firstly information on a continuous diagnostic marker or variable M on a number of studies. The data calculated by each study are the number of patients with a positive test result (y_1) and the total number of patients (n_1) in the group with the disease. On the other hand, the number of patients with a positive test result is (y_0) and the total number of patients is (n_0) in the group without the disease. Estimating the overall ROC curve of the diagnostic marker M based on the provided data from the different studies is the main goal (Arends et al., 2008).

It has been noted that a ROC curve plots the sensitivity as a function of false positivity a test at different smooth curve through these points (Reitsma, Glas, Rutjes, Scholten, Bossuyt, & Zwinderman, 2005).

The common method used in diagnostic thresholds but in SROC analysis, first the sensitivity and false positive is plotted for each study then a SROC curve is plotted to fit all these points.

Littenberg and Moses (1993) stated that transformed test, X, follows a logistic distribution both in the population without the disease and in the population with the disease assuming there is the transformation of the continuous diagnostic variable M. Transformation is done assuming that large values of X correspond with the diseased population.

The cumulative distribution of X in the healthy and the diseased populations for $\alpha \geq 0$ and $\beta>0$;

—— and (EQUATION 4)

——— (EQUATION 5)

The difference between the mean value with the disease and without the disease is α/ β in the population. Here, α is the accuracy parameter along with β being the scale parameter (Rutter and Gatsonis, 2001). The ratio between the standard deviation of the population of the diseased and the healthy is 1/ β. Arends et al., 2008 showed that in Figure 6, 0< β <1 corresponds with a higher variance in the population with the disease and β>1 with a smaller variance.

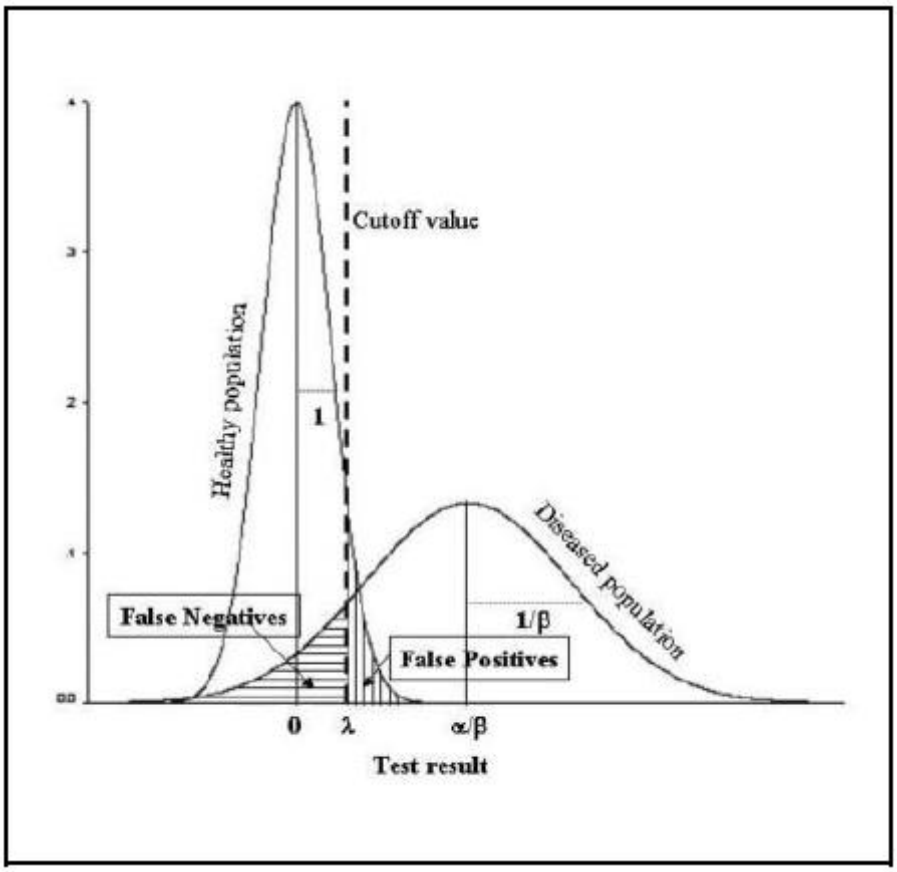


Figure 5 Illustration of α and β.

λ signify the threshold X value for the test being positive, then according to cumulative distribution of X the probability of a false positive result is 1−e λ =(1+e λ) (Littenberg, & Moses, 1993).

(EQUATION 6)

(EQUATION 7)

Linear relationship is;

(EQUATION 8)

The estimation of the ξ and η are (Table 1);

and (EQUATION 9)

(EQUATION 10)

In the SROC approach of Littenberg and Moses (1993), the relation is written as

(EQUATION 11)

Accepting that $=2\alpha/(\beta+1)$ (with ≥ 0) and $=(\beta-1)/(\beta+1)$ (with $-1<\beta\ 0<1$). Cox (1970) stated that if D and S are the estimated values of $\eta-\xi$ and $\eta+\xi$ from a study (avoid division by 0, 0.5 is added), then it is (Table 1);

and (EQUATION 12)

and (EQUATION 13)

and gives the regression; (EQUATION 14)

(EQUATION 15)

Where D is the difference between logit(TPR) and logit(FPR). Also S is the sum of logit(TPR) and logit(FPR).

Following that the values of and are estimated by a simple weighted or unweighted linear regression (Littenberg, & Moses, 1993). The log odds ratio of a α positive test result for diseased population relative to healthy population is D, and is often called the diagnostic odds ratio (Moses, et al., 1993). Briefly its estimated variance is;

(EQUATION 16)

The summary ROC curve is plotted by transforming the estimate of + , to the ROC space and the relation between p and q is calculated as (Littenberg, & Moses, 1993);

(EQUATION 17)

Though, the SROC method assumes that the values of α and β do not vary across studies which means it is a fixed effects method (Sutton, 2000). Usually, variation is according to the threshold effect and within-study sampling variability. Also, D and S are correlated within a study, positively or negatively, depending on the study which is ignored in fixed effects SROC model. On the other hand, in many practical cases it is likely that there is between-study variation beyond those sources in which fixed effect model can give biased estimates and misinterpret the standard errors (DerSimonian, & Laird, 1986; Normand, 1999; Hardy, & Thompson, 1998; Thompson, 1994; Berkey, Hoaglin, Antczak-Bouckoms, Mosteller, & Colditz, 1998). Chappell et al. (2009) argue that fixed effect method of (Moses, & Littenberg, 1993) can be criticized since it does not allow for the non-linear transformation, for the binomial variance of individual studies or variables does not provide valid estimates of precision because of error.

Intercept (α) and the slope (β) of the linear regression model declaration is not straightforward. The diagnostic odds ratio does not depend on the threshold S the intercept would show a summary estimate of the Diagnostic Odds Ratio (DOR).

However, when the DOR vary with S, the coefficient of slope (β) has no clarification directly, but has a great effect on the shape of the SROC (Walter, 2002).

Moreover, new meta-analytical methods consider variation across studies by introducing random effects (Van Houwelingen, 1995; DerSimonian, & Laird, 1986; Normand, 1999; Hardy, & Thompson, 1998; Thompson, 1994).

As a result, regression to the mean (Senn, 1994) and attenuation due to measurement errors (Carroll, Ruppert, & Stefanski, 1995) could seriously bias the slope of the regression line (Van Houwelingen, et al., 2002) if the variable S in regressions measurement error is regarded (bias in β□ and α□ , β, σ) (Thompson, 1994).

As discussed earlier ROC shows the TPR and FPR changes through the cut points changes according to only one study. However, SROC summarize the studies without noticing the variables differences from study to study (Moses, et al., 1993).

3.3 Effectiveness Index

Effect size (δ) in meta-analysis is the one–dimensional version of distance is simply the standardized mean difference (Hasselblad, & Hedges, 1995).

(EQUATION 18)

The estimated effect size, d, is calculated by estimated , and σ, where s is the pooled standard deviation for the continuously valued screening tests. The distribution of the d is normal when sample sizes are not small with a mean of δ, a variance of v and , showing sizes of the samples from the two sub-populations (Hasselblad, & Hedges, 1995).

(EQUATION 19)

(EQUATION 20)

Hasselblad, & Hedges (1995) explained that log odds ratio (sum of the logits of sensitivity and specificity) is a constant multiplied by the standardized difference between means. Hence index of effectiveness (δ) for the continuous case, an estimate d of δ, and an estimate of the variance of d, for the binary valued screening tests can be calculated respectively;

(EQUATION 21)

(EQUATION 22)

(EQUATION 23)

A,B,C and D are components of two-by-two contingency table in Table 1 showing the Distribution of Reference and Diagnostic Tests.

3.4 Random Effects Meta-Analysis

As discussed earlier, fixed effect and random-effect methods can be used for integrating effectiveness index (meta-analyzing) just as any other effect size estimates. Thus estimates can come from different valued effect size estimates as binary and continuously (Hasselblad, & Hedges, 1995).

Fixed effects model can be used in case of homogeneity. In this case, random effect analysis is used for FNAC due to heterogeneity.

As mentioned before, if there is heterogeneity, random effect model is more appropriate (DerSimonian, & Laird, 1986). Heterogeneity can be calculated with I^2 or Cochran Q.

where (EQUATION 24)

(EQUATION 25)

with is the estimation between studies variance

Θ is TPR or TNR or DOR

(EQUATION 26)

Moreover, random effect models included the measure v of the variation between studies in the calculation of the total uncertainty (Hasselblad, & Hedges, 1995).

The (variant component) is used as a measure of the amount of between-studies heterogeneity in effectiveness. The random effects weighted mean is *, the variance of the weighted mean in random effect model is calculated as;

(EQUATION 27)

(EQUATION 28)

(EQUATION 29)

Where and (m-1) is the DF of χ^2 distribution (EQUATION 30)

Hasselblad and Hedges (1995) assumed that the effect size (δ) measure, is valued without error, populations are normal distributed with same variances and the cutoff level can be fixed.

Distinctively, d can be transformed to ROC. Sensitivity (S_n) can be calculated through given d and specificity (S_p) and ROC curve can be plotted S_n versus 1-S_p (Hasselblad, & Hedges, 1995).

(EQUATION 31)

3.5 The Bivariate Model

SROC converts each pair of sensitivity and specificity into a single measure of accuracy (DOR), which does not distinguish between ability of detecting sensitivity and specificity (Reitsma et al. 2005). On the other hand, bivariate model has an advantage of displaying the two-dimensional nature of the underlying data throughout the analysis. Integrating effective, DerSimonian and Laird (1986) proposed random effects model this is the most used model for meta-analyzing FPR of a diagnostic test:

with — —— (EQUATION 32)

and ξ are the observed and true logit(FPR) of study i, respectively. The parameter shows the overall mean logit FPR. defines the between-study variance in true logit FPR. Also, TPR are analyzed using this model:

with — —— (EQUATION 33)

Assuming a bivariate normal model for the pair ξ_i and η_i (Reitsma et al., 2005):

(EQUATION 34)

Arends et al. (2008) noted that above model (Equation 34) describes the univariate random-effects meta-analysis model for the ξ_i and η_i separately but now assure that ξ_i and η_i are correlated. One way to characterize the overall accuracy of the diagnostic test would be to get the estimated and and transform them to the ROC space also characterizing the bivariate normal distribution by a line and then transform that line to the ROC space is an another way. Chappell, et al. (2009) added that logit FPR and logit TPR are modeled directly as a bivariate normal distribution with mean $(\mu_1, \mu_2)^T$ and variance- covariance matrix Ω , which can be showed as and also covariance or the correlation ρ.

Random effect approach is the sensitivities from individual studies within a meta-analysis are approximately normally distributed around a mean value. Also, there is variability around this mean. The combination of the logit transformed sensitivities and specificities and the correlation between them indicates the bivariate normal distribution (Reitsma et al., 2005).

The bivariate model can be seen as an improvement to SROC approach (Reitsma et al., 2005);

1. The bivariate model will estimate the amount of between-study variation in sensitivity and specificity separately also the correlation of between them which is important for heterogeneity of result between studies.
2. The bivariate model calculates summary estimates of sensitivity and specificity and 95% confidence interval.
3. The parameters of bivariate model can be used for plotting a SROC curve.
4. DOR and likelihood ratios (LR) can be calculated (Reitsma et al., 2005);

——————— and (EQUATION 35)

——————— (EQUATION 36)

with (EQUATION 37)

5. Covariates can be added to the bivariate model.

3.6 HSROC (Hierarchical SROC)

The SROC model is simple for summarizing paired estimates of Sensitivity and (1-Specificity) across studies. However, the SROC have some important limitations.

For instance, the model cannot distinguish between within study and between study variability. In other words, it assigns equal weight to all pairs of (Sensitivity, 1-

Specificity) even though potentially large differences between studies exists with respect to sample sizes. The SROC model can be fitted using weighted analysis. Unfortunately Walter (2002) proposed that, it produces biased estimates. Thus, a hierarchical SROC model (HSROC) would be more preferred.

The HSROC model can be considered as having two levels as within and between studies (Rutter, & Gatsonis, 2001). Macaskill (2004) showed that for the *i*th study, the number of test positives for the diseased (I) and healthy (N) test subjects, t_{ij}, j = I, N, respectively, are assumed to have binomial distributions;

(EQUATION 38)

(EQUATION 39)

(EQUATION 40)

(EQUATION 41)

where p_{ij} is the probability of a positive test result for the *i*th study and *j*th diseased status; n_{ij} is the sample size for the *i*th study and *j*th diseased status; d_{ij} is the true disease status for the *i*th study and *j*th diseased status (coded as 0.5 for j = I and -0.5 for j = N); and the random effects (and) are assumed to be independent and normally distributed.

Toft and Nielsen (2009) stated that each study has its own certain threshold , estimating the average log odds of a positive test result for the diseased and healthy groups, and diagnostic accuracy , estimating the expected diagnostic log odds ratio (LOR). The scale parameter allows the accuracy to vary with certain threshold, thereby letting asymmetry in the SROC. The must be modelled as a fixed effect, since each study only contributes one point to the SROC curve.

Hence the association between threshold and accuracy must be derived from the studies considered jointly. Essentially, and in Equation (39) can be interpreted

analogously to and in Equation (15). Thus using and the HSROC can be derived as (Macaskill, 2004):

(EQUATION 42)

by varying (1-Specificity) across the relevant range.

Because each study presents only one point in ROC space to the analysis, a single study does not provide information on the shape of the SROC.

One of the advantage of the HSROC models is the observed sensitivity and 1-specificity for each study, taking account the correlation between them (Macaskill, 2004).

On the other hand, in the Moses method the D and S in Equation 15 are computed before modeling which causes the information lost on sensitivity and specificity (Toft, & Nielsen, 2009).

Thus, HSROC allows the sampling variability in the sensitivity and specificity to be taken into account in the modeling addition to that it also allows the summary estimates of sensitivity and specificity to be included as a function of the model parameters (Macaskill, 2004).

CHAPTER 4

RESULTS

4.1 FNAC Data Profile

Studies comparing FNAC results of specimens from palpable breast masses with the histological diagnosis of each mass are identified (Akçil et al., 2008). Thus, the meta-analysis data's consisted of 25 FNAC studies.

The summaries of the studies analyzed are presented in Table 5. The years of publication of the 25 FNAC studies ranged from 1984 to 2007. The number of patients included in the meta-analysis is 10455.

4.2 Sensitivity and Specificity Calculations

Figure 6 displays the sensitivities and specificities with a forest plot for FNAC in 25 studies which the sensitivities changes from 78% to 100%, specificities changes from 76% to 100%.

Norton et al. (1984) displays the lowest sensitivity and specificity pair in Figure 6. Also Somers et al. (1985), Watson et al. (1987) and Wilkinson et al. (1989) display lower sensitivities (Figure 6). Forest plots and elements of Figure 6 are plotted by RevMan Version 5 (2008).

Table 5 Main Elements of the 25 Studies Assessed for Meta-Analysis

STUDY	AUTHOR	YEAR	N	TP	TN	FP	FN
1	Vetrani et al.	1992	256	136	108	7	5
2	Watson et al.	1987	260	46	200	1	13
3	Sheikh et al.	1987	2263	293	2290	40	0
4	Griffith et al.	1986	236	110	95	15	16
5	Atamded et al.	1993	51	32	17	1	1
6	Ciatto et al.	1989	563	489	60	1	13
7	Gelabert et al.	1990	107	90	12	0	5
8	Horgon et al.	1991	1742	222	1471	11	38
9	Lannin et al.	1986	93	26	65	0	2
10	Norton et al.	1984	37	10	19	6	2
11	Collaço et al.	1999	260	175	69	1	15
12	Arıkan et al.	1992	134	63	69	0	2
13	Langmuir et al.	1989	101	24	65	11	1
14	Smith et al.	1988	317	113	181	15	8
15	Dominguez et al.	1997	427	158	247	11	11
16	Feichter et al.	1997	323	153	145	1	24
17	Wilkinson et al.	1989	240	27	206	0	7
18	Silverman et al.	1987	93	33	47	2	6
19	Wang et al.	1998	165	114	45	3	3
20	Zardawi et al.	1998	437	100	329	3	5
21	Somers et al.	1985	185	80	82	0	23
22	Kaufman et al.	1994	234	102	120	4	8
23	Chaiwun et al.	2002	424	194	193	1	36
24	Mizuno et al.	2005	94	72	14	1	7
25	Ariga et al.	2002	1058	814	222	3	19

Heterogeneity is generated, at least because of sampling error, among the 25 FNAC studies in the meta-analysis. It is evaluated by the Cochran Q-test and inconsistency (I^2) statistics that are provided by the free meta-analysis program Meta-DiSc (Meta-DiSc, 2006; Higgins, Thompson, Deeks, & Altman, 2003).

Akçil et al. (2008), pointed out that I^2 for sensitivity, specificity and DOR were 88.8%, 85.1%, and 74.1%, respectively which are considered to be large deviations for 25 FNAC studies (Equation 26). Thus, the calculations were based on the random effects model.

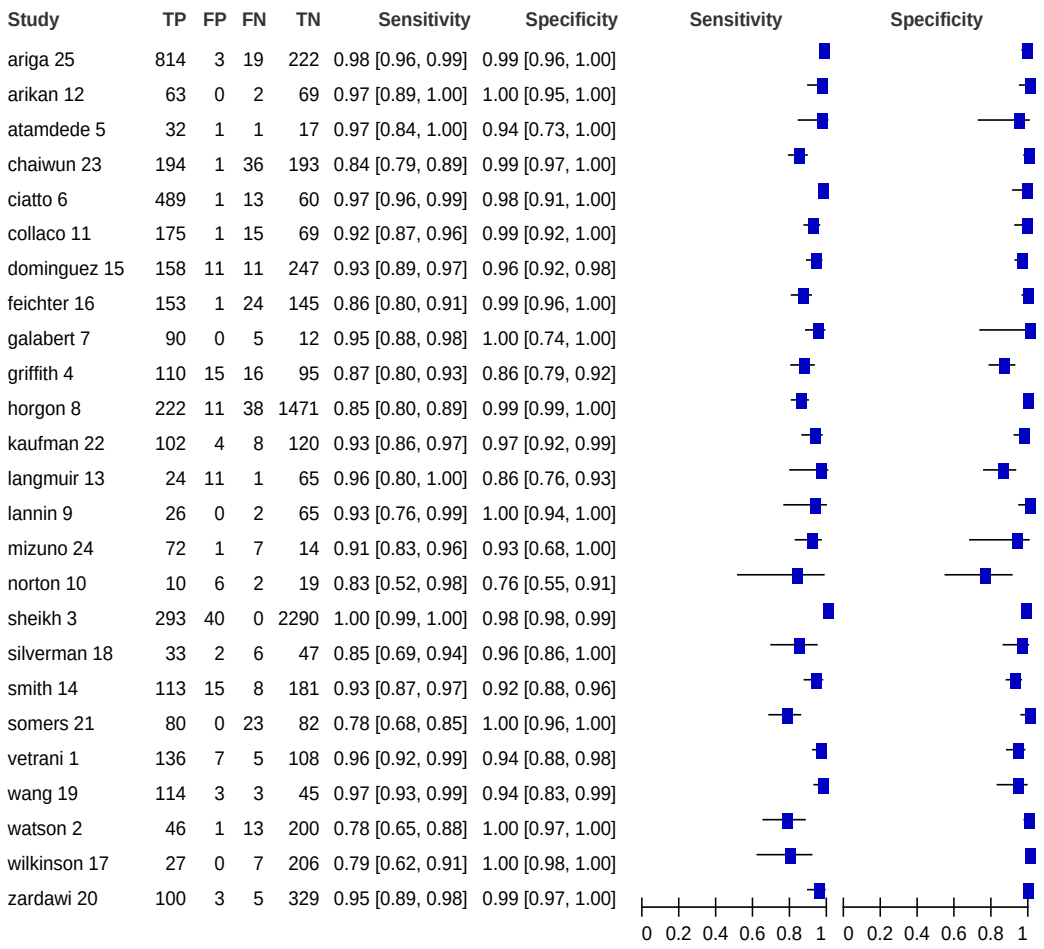


Figure 6 Forest Plot of Specificities and Sensitivities of 25 FNAC Studies

The estimated pooled sensitivity, specificity, and DOR are 0.9316, 0.9751 and 628.55, respectively in Table 6. Estimations developed in The NLMIXED Procedure in SAS for SROC nonlinear fixed-effects model (Appendix A, Appendix B).

Table 6 Estimations of Specificity, Sensitivity and DOR of 25 Studies in NLMIXED

Estimate	Value	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Sens	0.9316	0.004019	24	231.78	<.0001	0.05	0.9233	0.9399
Spec	0.9788	0.001786	24	548.18	<.0001	0.05	0.9751	0.9825
DOR	628.55	67.0492	24	9.37	<.0001	0.05	490.17	766.93

In addition to that, Table 6 states that the diagnostics measures are heterogeneous for 25 studies in the meta-analysis ($p<0.001$ for homogeneity tests).

4.3 Bivariate Random Effects

In Figure 7, plots of the standard SROC curve, based on the Littlenberg and Moses (1993) linear regression model is presented with RevMan Version 5 (2008).

However, SROC curves, average operating points including 95% confidence intervals and 95% prediction regions is plotted in RevMan Version 5 (2008) using estimates from more valid statistical models: the bivariate model as in Figure 8 and the hierarchical SROC (HSROC) model as in Figure 9.

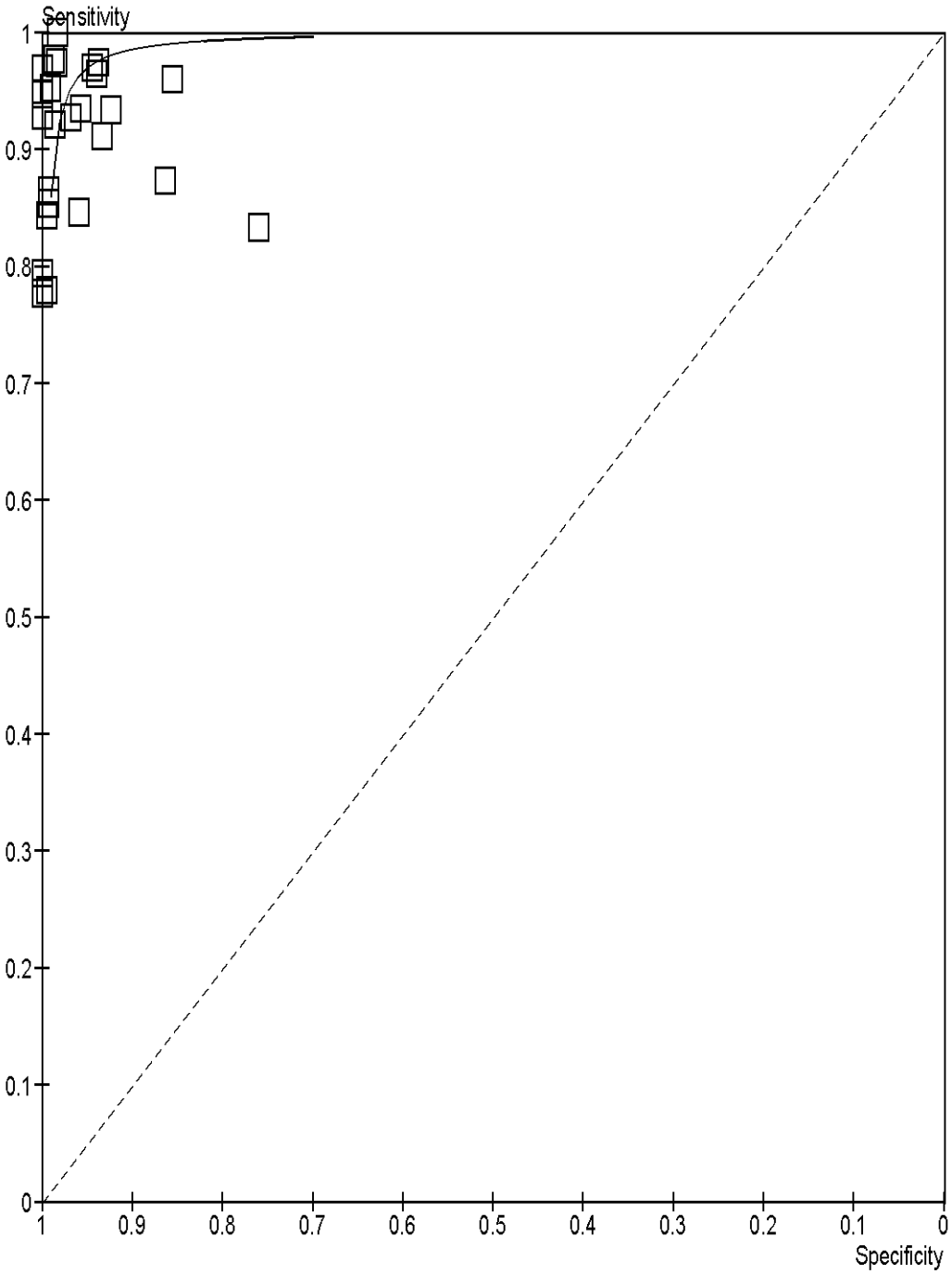


Figure 7 Standard SROC of 25 FNAC Studies

The standard way of meta-analyzing in literature is the random-effects model of DerSimonian and Laird, (1986) as discussed in Equation 34. The straightforward meta-analytic approach is to generate a bivariate random-effects model for pooling sensitivity and specificity.

In Figure 8, bivariate random-effects meta-analysis of sensitivity and specificity SROC is plotted. SROC is represented with the solid line. Observed bivariate pairs of sensitivity and specificity of the 25 FNAC studies are represented by rectangles.

Bivariate summary point is represented with the central point. Bivariate boundary of the 95% confidence region for the summary point (central point) is represented with the ellipse. The bivariate random effects meta-analysis of SROC plot is generated by RevMan Version 5 (2008).

To plot the bivariate random effects SROC, the model parameters codes are generated in SAS PROC GLIMMIX (Appendix A). Its estimations are calculated in SAS as below;

```
model true/total = status / noint s cl corrb covb ddfm=bw;
random status / subject=study S type=chol G;
estimate 'logit_sensitivity' status 1 0 / cl ilink;
estimate 'logit_specificity' status 0 1 / cl ilink;
estimate 'LOR' status 1 1 / cl exp;
```

Table 7 Estimations of Bivariate Random Effects in SAS PROC GLIMMIX

Estimate	Value
E(logitSe)	2.6304
E(logitSp)	3.9010
Var(logitSe)	0.7886
Var(logitSp)	1.6773
Cov(logits)	-0.3065
SE(E(logitSe))	0.2001
SE(E(logitSp))	0.3149
Cov(Es)	-0.01235
Studies	25

For plotting the SROC, the parameters are obtained from SAS PROC GLIMMIX (Table 7, Appendix B).

In SAS PROC GLIMMIX, expected mean value of logit transformed sensitivity and logit transformed specificity were calculated as, 2.6304 and 3.9010. Between-study variance of logit transformed specificity found as 0.7886 and logit transformed specificity 1.6773, respectively.

Also, covariance between logit transformed sensitivity and specificity was -0.3065. For creating confidence and prediction regions, standard error of the expected mean value of logit transformed sensitivity and logit transformed of specificity was calculated as 0.2001 and 0.3149. Finally, covariance between expected mean logit sensitivity and specificity was -0.01235 (Table 7).

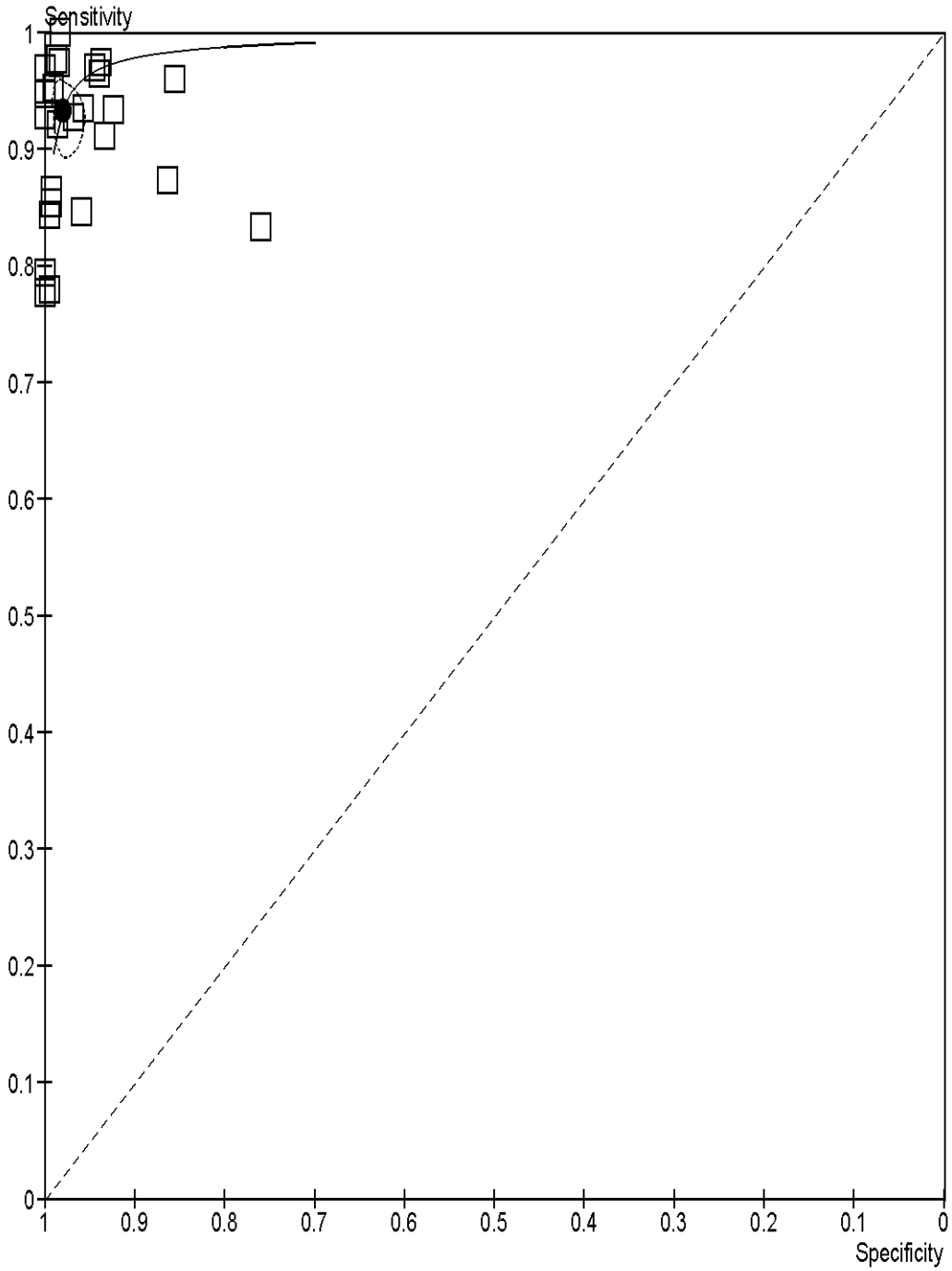


Figure 8 Bivariate Random Effects Meta-Analysis of SROC of 25 FNAC Studies

4.4 Hierarchical SROC (HSROC)

In Figure 9, hierarchical SROC is plotted. HSROC is represented with the solid line. Observed bivariate pairs of sensitivity and specificity of the 25 FNAC studies are represented by rectangles.

Summary point is represented with the central point. Boundary of the 95% confidence region for the summary point (central point) is represented with the ellipse. The HSROC plot is generated by RevMan Version 5 (2008).

To plot the bivariate random effects SROC, the model parameters codes are generated in SAS PROC GLIMMIX (Appendix A). Its estimations are calculated in SAS as below;

```
estimate 'var_usen' var_usen; /*bivariate model paramaters estimated below/*
estimate 'var_uspec' var_uspec;
estimate 'cov_usenspec' cov_usenspec;
estimate 'logit_sensitivity' mu_sen;
estimate 'logit_specificity' mu_spec;
estimate 'LOR' mu_sen + mu_spec;
estimate 'sensitivity' exp(mu_sen)/(1+exp(mu_sen));
estimate 'specificity' exp(mu_spec)/(1+exp(mu_spec));
estimate 'DOR' exp(mu_sen + mu_spec);
```

For plotting the HSROC, the parameters obtained with the parameters obtained from SAS PROC NLMIXED (Table 8, Appendix B).

Table 8 Estimations of HSROC in SAS PROC NLMIXED

Estimate	Value
Alpha	6.4112
Theta	0.01907
Beta	0.3829
Var(accuracy)	0.7247
Var(threshold)	1.6239
SE(E(logitse))	0.1993
SE(E(logitsp))	0.3142
Cov(Es)	-0.0190725
Studies	25

In SAS PROC NLMIXED, Alpha (accuracy parameter), Theta (threshold parameter) and Beta (shape parameter) were calculated as 6.4112, 0.01907 and 0.3829, respectively.

Also, variance of accuracy parameter was 0.7247 and variance of threshold parameter was 1.6239. For creating confidence and prediction regions, standard error of the expected mean value of logit transformed sensitivity and logit transformed of specificity was calculated as 0.1993 and 0.3142.

Finally, covariance between expected mean logit sensitivity and specificity was -0.0190725.

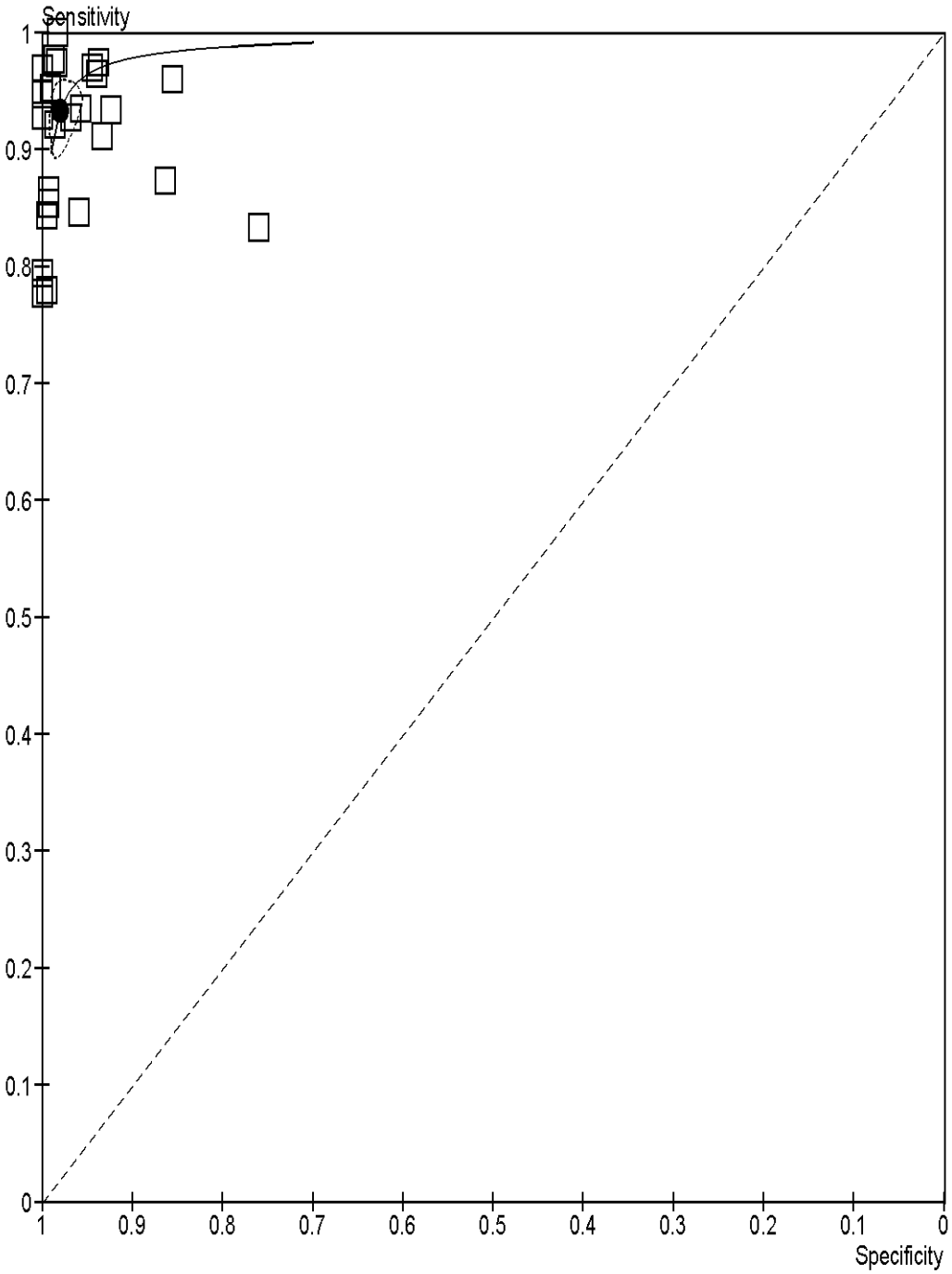


Figure 9 HSROC of 25 FNAC Studies

4.5 Comparison of Bivariate Random Effects Model and HSROC

Table 9 gives the mean and %95 confidence interval of sensitivity and specificity for the fixed effects meta-analysis with the GLIMMIX and NLMIXED approaches.

Table 9 Fixed Effects Meta-Analysis of FNAC Data

Approach	SAS Proc	Pooled sensitivity	Pooled specificity
Bivariate	GLIMMIX	93.2% (92.3-93.9)	97.9% (97.5-98.2)
HSROC	NLMIXED	93.2% (92.3-94.0)	97.9% (97.5-98.3)

Pooled sensitivities were calculated as same in both models as well as pooled specificity. However, their confidence intervals changes slightly from each other (Table 9).

It is assumed that ξ_i and η_i of the estimated true sensitivity and true specificity in each study are assumed to have a bivariate normal distribution across the studies (Menke, 2010). Also it allows possible correlation between the true sensitivity and true specificity.

Reitsma et al. (2005) proposed the bivariate random effects model and its covariance structure as in Equation 34. This covariance matrix includes the random effects between study variances and of the studies' sensitivities and specificities, and their covariance . Equation 34 is written as with U being the random effect for sensitivity and specificity of study i;

with (EQUATION 43)

(EQUATION 44)

Equation 44 is the bivariate generalized linear mixed model by Chu and Cole, 2006. and are normally distributed random effects estimations that are correlated

by covariance . However, fixed effect model is structured by omitting and .

Table 10 Random Effects Meta-Analysis of FNAC Data

Approach	SAS Proc	Pooled sensitivity	Pooled specificity
Bivariate	GLMMIX	93.3% (90.2-95.5)	98.0% (96.3-99.0)
HSROC	NLMIXED	93.3% (90.7-95.9)	98.0% (96.8-99.3)

Table 10 gives the mean and %95 confidence interval of sensitivity and specificity for the random effects meta-analysis with the GLIMMIX and NLMIXED approaches.

Pooled sensitivities were calculated as same in both models with 93.3% as well as pooled specificities with 98.0% (Table 10). However, their confidence intervals vary slightly from each other. Pooled sensitivity intervals were (90.2-95.5) in GLMMIX and (90.7-95.9) in NLMIXED (Table 10). Moreover, Pooled specificity intervals were (96.3-99.0) in GLMMIX and (96.8-99.3) in NLMIXED (Table 10).

GLIMMIX and NLMIXED codes are generated from literature (Peng, 2009; Littell, Stroup, & Freund, 2002; Kleinman, & Horton, 2010; Marasinghe, & Kennedy, 2008; Geoff, & Gueritt, 2002).

CHAPTER 5

DISCUSSION

Diagnostic tests are important parts of the clinical diagnosis (Menke, 2010). Diagnostic test supports diagnosis and allows evaluating the disease. On the other hand, they do not completely show gold standards without knowing the accuracy of the test.

Diagnostic test accuracy is needed when assessing the true result of a test. The diagnostic test accuracy is achieved by several studies including meta-analysis (Sutton et al., 2000). Meta-analysis aim is to quantitatively summarize studies to achieve pooled test accuracy estimates that are globally effective results than the results of a one study (Chappell et al., 2009).

If there is little variation between trials then I^2 will be low and a fixed effects model might be more appropriate. Since I^2 is high, an alternative approach, 'random effects' is considered in this study due to heterogeneity. Random effects allow the study outcomes to vary in a normal distribution between studies (DerSimonian, & Laird, 1986).

I^2 was calculated as 88.8%, 85.1%, and 74.1%, for sensitivity, specificity and DOR, respectively. This was discussed as large deviations for 25 FNAC studies (Akçil et al., 2008).Thus, Standard SROC cannot distinguish between and within study variability (Figure 8).

Almost every meta-analysis of diagnostic tests studies reports their results as bivariate pairs of sensitivity and specificity. Sensitivity and specificity are based on binomial distribution which also approaches the Gaussian normal distribution only with large numbers. Therefore, PROC GLIMMIX is implemented with SAS since it considers the binomial distribution (Appendix A.3).

The HSROC approach replaces random effects model in most studies. Macaskill, (2004) proved that, the empirical Bayesian estimates for sensitivity and specificity pair from HSROC NLMIXED (Appendix A.5) accounts for full Bayesian analysis. The HSROC approach models the diagnostic accuracy (alpha) and threshold (theta) and the scale parameter (beta) to ensure for asymmetry in the SROC curve by allowing accuracy to vary with the theta (Macaskill, 2004).

This study has shown that the bivariate GLIMMIX approach perform an alternative to the HSROC approach. The evaluation of 25 meta-analysis showed that, both bivariate GLIMMIX and HSROC NLMIXED meta-analytic approaches are almost the same, despite being calculated by different models (Figure 9, Figure 10).

Addition to the result proposed above, Menke (2010) stated that bivariate GLIMMIX and HSROC NLMIXED meta-analytic approaches were almost identical in 50 meta-analysis for the lymphangiography (LAG) data of Scheidler, Hricak, Yu, Subak, Segal (1997). Pooled sensitivities were calculated as same in both models for 50 meta-analysis for the LAG data with 67.4% as well as pooled specificities with 83.7%. However, pooled sensitivity intervals were (59.8-74.2) in GLMMIX and (60.2-74.6) in NLMIXED with slightly varies in calculations. Addition to that, Pooled specificity intervals were (75.1-89.8) in GLMMIX and (76.5-91.0) in NLMIXED (Menke, 2010).

Therefore the bivariate GLIMMIX approach has the same accuracy as the HSROC approach. Direct modeling of sensitivity and specificity can be considered an advantage of the bivariate GLIMMIX approach (Menke, 2010).

In our study, pooled sensitivities were calculated as same in both models with 93.3% as well as pooled specificities with 98.0% (Table 10). However, pooled sensitivity intervals were (90.2-95.5) in GLMMIX and (90.7-95.9) in NLMIXED with slightly varies in calculations for 25 FNAC data (Table 10).

Addition to that, pooled specificity intervals were (96.3-99.0) in GLMMIX and (96.8-99.3) in NLMIXED (Table 10).

Furthermore, bivariate GLIMMIX approach codes are generated much more easily which is more comprehensible than the HSROC NLMIXED approach (Appendix A.2, Appendix A.5).

In SAS PROC GLIMMIX, standard error of the expected mean value of logit transformed sensitivity and logit transformed of specificity was calculated as 0.2001 and 0.3149 (Table 7).

On the other hand, standard error of the expected mean value of logit transformed sensitivity and logit transformed of specificity was calculated as 0.1993 and 0.3142 in SAS PROC NLMIXED (Table 8). This value is close to the result of the SAS PROC GLIMMIX as stated above when comparing both models.

Finally, when comparing the results of GLIMMIX and NLMIXED for both models, covariance between expected mean logit sensitivity and specificity was calculated as -0.01235 for bivariate model (Table 7).

Similarly, covariance between expected mean logit sensitivity and specificity was calculated as -0.0190725 for HSROC model.

This study has shown the implemented SAS PROC GLIMMIX and SAS PROC GLIMMIX for FNAC of 25 studies.

The codes are generated for fixed effects meta-analysis as shown in the Appendix A.2 and Appendix A.4 which are not included in the Results part (Chapter 4).

My contribution was to generate both SAS PROC GLIMMIX and HSROC NLMIXED codes and compare the bivariate random effects model SROC and hierarchical SROC curve for 25 FNAC studies.

CHAPTER 6

CONCLUSION

Breast cancer is one of the mortal diseases especially among women. According to estimation of Breast Cancer organization in the US (2011), About 1 in 8 women (approximately 12%) will develop breast cancer over the course of her lifetime (Breast Cancer Organization, 2011).

The symptom of breast cancer is a tumor that can be examined by palpation. Physician should evaluate the mass whether it is malignant or benign. For the identification of breast cancer, patient must undergo an operation of excision biopsy which is very painful, risky and expensive (Akçil, Karaagaoğlu, & Demirhan, 2008).

On the other hand, patients those who undergo a Fine Needle Aspiration Cytology do not require an anesthetic. Furthermore, operation is practical and inexpensive (Giard, & Hermans, 1992).

Today, meta-analysis is used almost in every area in literature. Medical field is one of the largest areas of usage. Moreover, meta-analysis of the diagnostic test is the newly developed method which is very popular nowadays.

Necessarily, meta-analysis of FNAC has been conducted in literature several times in Literature. However, in the literature, there is no example of comparing the study of bivariate random effects SROC and HSROC on FNAC. In this manner, this study will contribute a lot to the literature.

Sutton et al. (2000) stated that, if several studies are conducted then their results usually differ. In other words, meta-analysis provides consortium upon subjects. Thus meta-analysis of the 25 FNAC studies is conducted for showing the effectiveness of FNAC.

The main goal of the study was to calculate pooled summary estimates for sensitivity and specificity that were more reliable than a result of a single study.

A generalized linear random effects model with PROC GLIMMIX offers a straightforward method for bivariate meta-analysis of sensitivity and specificity, and is thus an alternative to the HSROC approach (Menke, 2010).

It has been showed the HSROC model and the bivariate random effects model for meta-analysis of diagnostic studies are closely related, and in this study almost identical (Figure 8 and Figure 9).

The bivariate model allows addition of covariates that impose both sensitivity and specificity, while the HSROC model allows covariates impose both accuracy and threshold parameters.

Comparing two FNAC studies, bivariate random effects meta-analysis of 29 FNAC studies found that mean logit transformed sensitivity was 1.774 and mean logit transformed specificity was -2.384 in Table 4 (Arends et al., 2008).

In our FNAC study of 25 studies, expected mean value of logit transformed sensitivity and logit transformed specificity were calculated as, 2.6304 and 3.9010 in SAS PROC GLIMMIX for bivariate model (Table 7).

Addition to that, between-study variance of logit transformed specificity found as 0.7886 and logit transformed specificity as 1.6773, respectively for our 25 FNAC study (Table 7).

On the other hand, between-study variance of logit transformed specificity found as 0.286 and logit transformed specificity as 0.990 for the 29 FNAC studies by Arends et al. (2008).

Furthermore, covariance between logit transformed sensitivity and specificity was -0.3065 in 25 FNAC studies (Table 7).

However, covariance between logit transformed sensitivity and specificity was calculated as 0.146 for 29 FNAC studies by Arends et al. (2008).

Lastly, meta-analysis of 25 studies stated that FNAC analysis of palpable breast masses is effective at differentiating tumors with the final diagnosis (malignant or benign disease). This shows that the FNAC can be used and as diagnostic modality which has cost-effectiveness and reliability and safeness.

6.1 Limitations

This study has limitations of comparing the bivariate GLIMMIX approach and the HSROC NLMIXED approach which are mathematically similar approaches for the bivariate meta-analysis of sensitivity and specificity.

A further approach should be the applications of the bivariate meta-regression (Chapter 2.5).

Plotting the bivariate meta-analysis SROC and HSROC in SAS are not available (SAS Institute Inc, 2004). So, the SROC plotting was established in the RevMan Version 5.

6.2 Future Works

The PROC MIXED codes of Van Houwelingen et al., (2002) and Reitsma (2005) will be adapted to the PROC GLIMMIX syntax for bivariate meta-regression.

By adding the characteristics of the study to the model as aspiration number and study design, multivariate SROC and HSROC will be plotted.

Although, the plotting of the SROC is tried in the SAS coding unfortunately it couldn't be achieved. Package of the plotting in SAS programming will be developed in order to achieve coding integrity.

REFERENCES

Akçil, M., Karaagaoğlu, E., & Demirhan, B. (2008). Diagnostic accuracy of fine-needle aspiration cytology of palpable breast masses: An SROC curve with fixed and random effects linear meta-regression models. *Diagnostic Cytopathology, 36*, 303-10.

Arends, L. R., Hamza, T. H., Van Houwelingen, J. C., Heijenbrok-Kal, M. H., Hunink, M. G. M., & Stijnen T. (2008). Bivariate random-effects meta-analysis of ROC curves. *Medical Decision Making, 28* , 621.

Arends, L. R., Hoes, A. W., Lubsen, J., Grobbee, D. E., & Stijnen, T. (2000). Baseline risk as predictor of treatment benefit: three clinical meta-reanalyses. *Statistics in Medicine, 19*, 3497–518.

Ariga, R., Bloom, K., Reddy, V. B., Kluskens, L., Francescatti, D., Dowlat, K., Siziopikou, P., & Gattuso, P. (2002). Fine-needle aspiration of clinically suspicious palpable breast masses with histopathologic correlation. *Am J Surg*, *184*(5), 410-413.

Arikan, P., Pak, I., Demir, A., Ozgen, K., & Kutun, S. (1992). The diagnostic value of fine-needle aspiration in breast masses. *Ankara Pathology Bulletin, 9*, 13-17.

Armitage, P., & Colton, T. (1998). *Encyclopedia of Biostatistics (1st ed., Vol. 4).* John Wiley and Sons, Inc,: New York

Atamdede, F. I., & Isaacs, J. H. (1993). The role of fine-needle aspiration in the diagnosis of breast lesions. *Gynecologic Oncology*, *50*(2), 159-163.

Berkey, C. S., Hoaglin, D. C., Antczak-Bouckoms, A., Mosteller, F., & Colditz, G. A. (1998). Meta-analysis of multiple outcomes by regression with random effects. *Statistics in Medicine*, *17*, 2537–50.

Borenstein, M., Hedges, L. V., Higgins, J. P. T, & Rothstein, H. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley & Sons, Ltd.

Brand, R., & Kragt, H. (1992). Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statistics in Medicine, 11*(16), 2077- 2082.

Breast Cancer Organization. (2011). *U.S. Breast Cancer Statistics.* Retrieved from: http://www.breastcancer.org/symptoms/understand_bc/statistics.jsp.

Broemeling, L. D. (2007). Bayesian biostatistics and diagnostic medicine. *Boca Raton : Taylor & Francis.*

Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models.* London: Chapman & Hall.

Chaiwun, B., Settakorn, J., Ya-In, C., Wisedmongkol, W., Rangdaeng, S., & Thorner, P. (2002). Effectiveness of fine-needle aspiration cytology of breast: analysis of 2,375 cases from northern Thailand. *Diagn Cytopathol 26*(3), 201-5

Chappell, F. M., Raab, G. M., & Wardlaw, J. M. (2009). When are summary ROC curves appropriate for diagnostic meta-analyses? *Statististics in Medicine, 28*, 2653–2668 DOI: 10.1002/sim.3631

Chu, H., & Cole, S. R. (2006). Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol, 59*(12), 1331-2.

Chu, H., & Guo, H. (2009). Letter to the Editor. *Biostatistics, 10*(1), 201–203.

Ciatto, S., Cecchini, S., Grazzini, G., Iossa, A., Bartoli, D., Cariaggi, M. P., & Bulgaresi, P. (1989). Positive predictive value of fine needle aspiration cytology of breast lesions. *Acta Cytological, 33*(6), 894-898.

Collaço, L. M., De Lima, R. S., Werner, B., & Torres, L. F. (1999). Value of fine needle aspiration in the diagnosis of breast lesions. *Acta Cytological, 43*(4), 587-592.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trial, 7*(3), 177-188.

Dixon, J. M., Clarke, P. J., Crucioli, V., Dehn, T. C. B., Lee, E. C. G, & Greenal, M. J. (1987). Reduction of the surgical excision rate in benign breast disease using fine needle aspiration cytology with immediate reporting. *Br J Surg, 74*, 1014–16.

Domínguez, F., Riera, J. R., Tojo, S., & Junco, P. (1997). Fine needle aspiration of breast masses. An analysis of 1,398 patients in a community hospital. *Acta Cytological, 41*(2), 341-347.

Egger, M., Ebrahim, S., & Smith, G. D. (2002). Editorial: Where now for meta-analysis? *International Journal of Epidemiology, 31*(1), 1-5.

Feichter, G. E., Haberthür, F., Gobat, S., & Dalquen, P. (1997). Breast cytology. Statistical analysis and cytohistologic correlations. *Acta Cytological, 41*(2), 327-332.

Gelabert, H. A., Hsiu, J. G., Mullen, J. T., Jaffe, A. H., & D'Amato, N. A. (1990). Prospective evaluation of the role of fine-needle aspiration biopsy in the diagnosis and management of patients with palpable solid breast lesions. *Am Surg, 56*(4), 263-267.

Geoff, D., & Everitt, B. S. (2002). *A handbook of statistical analyses using SAS.* Boca Raton, Chapman & Hall/CRC.

Giard, R. W. M., & Hermans, J. (1992). The value of aspiration cytologic examination of the breast—a statistical review of the medical literature. *Cancer, 69*, 2104–10.

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Griffith, C. N., Kern, W. H., & Mikkelsen, W. P. (1986). Needle aspiration cytologic examination in the management of suspicious lesions of the breast. *Surg Gynecol Obstet, 162*(2), 142-144.

Hamza, T.H., Arends, L.R., Van Houwelingen J,C., & Stijnen, T. (2009). Multivariate random effects meta-analysis for diagnostic tests with multiple thresholds. *BMC Med Res Methodol*, 9, 73.

Hamza, T.H., Van Houwelingen, H.C., Heijenbrok-kal, M.H., & Stijnen, T. (2009). Associating explanatory variables with summary receiver operating characterstics curves in diagnostic meta-analysis. *Journal of Clinical Epidemiology, 62*(12), 1284-1291.

Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine, 17*, 841–56.

Hasselblad, V., & Hedges, L. V. (1995). Meta-analysis of screening and diagnostic tests. *Psychol Bull, 117*, 167–78.

Hellmich, M., Abrams, K. R, & Sutton, A. J. (1999) Bayesian approaches to meta-analysis of ROC curves. *Medical Decision Making, 19*, 252–64.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in metaanalyses. *BMJ 32*, 557–560.

Horgan, P. G., Waldron, D., Mooney, E., O'Brien, D., McGuire, M., & Given, H. F. (1991). The role of aspiration cytologic examination in the diagnosis of carcinoma of the breast. *Surg Gynecol Obstet, 172*(4), 290-292.

Irwig, L., Macaskill, P., Glasziou, P., & Fahey, M. (1995). Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol, 48*, 119–30; discussion 131–2.

Kardaun, J. W. P. F., & Kardaun, O. J. W. F. (1990). Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. *Methods of Information in Medicine, 29*, 12-22

Kaufman, Z., Shpitz, B., Shapiro, M., Rona, R., Lew, S., & Dinbar, A. (1994). Triple approach in the diagnosis of dominant breast masses: combined physical examination, mammography, and fine-needle aspiration. *J Surg Onco, 56*(4), 254-257.

Kester, A. D., & Buntinx, F., Meta-analysis of ROC curves. (2000). *Medical Decision Making, 20*, 430–9.

Kleinman, K., & Horton, N. J. (2010). *SAS and R, data management, statistical analysis, and graphics.* Boca Raton, FA, CRC Press.

Krzanowski, W. J., & Hand, D. J. (2009). *ROC curves for continuous data*. CRC Press/Taylor & Francis Group, Boca Raton, London, New York.

Langmuir, V. K., Cramer, S. F., & Hood, M. E. (1989). Fine needle aspiration cytology in the management of palpable benign and malignant breast disease. Correlation with clinical and mammographic findings. *Acta Cytological,. 33*(1), 93-98.

Lannin, D. R., Silverman, J. F., Walker, C., & Pories, W. J. (1986). Cost-effectiveness of fine needle biopsy of the breast. *Ann Surg, 203*(5), 474-480.

Littell, R. C., Stroup, W., W., & Freund, R. J. *SAS for linear models.* (2002). Cary, N.C., SAS Institute, New York, Wiley.

Littenberg, B., & Moses, L. E. (1993). Estimating diagnostic-accuracy from multiple conflicting reports—a new meta-analytic method. *Medical Decision Making, 13*,313–21.

Macaskill, P. (2004). Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. *J Clin Epidemiol, 57*, 925–932.

Marasinghe, M. G., & Kennedy, W. J. (2008). *SAS for data analysis, intermediate statistical methods*. New York , London, Springer.

McClish, D. K. (1992). Combining and comparing area estimates across studies or strata. *Medical Decision Making, 12*, 274–9.

Menke, J. (2010). Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods of Information in Medicine*, *49*(1), 54-62, 62-64.

Meta-DiSc version 1.4. (2006). http://www.hrc.es/investigacion/metadisc_en.htm. Accessed June 10,2011.

Mizuno, S., Isaji, S., Ogawa, T., Tabata, M., Yamagiwa, K., Yokoi, H., & Uemoto, S. (2005). Approach to fine-needle aspiration cytology-negative cases of breast cancer. *Asian J Surg, 28*(1), 13-17.

Moses, L. E., Shapiro, D., & Littenberg, B. (1993). Combining independent studies of a diagnostic test into a summary ROC curve: dataanalytic approaches and some additional considerations. *Statistics in Medicine, 12*, 1293–316.

Mushlin, A. I. (1985). Diagnostic tests in breast cancer: clinical strategies based on diagnostic probabilities. *Ann Intern Med, 103*, 79–85.

Normand, S. L. (1999). Meta-analysis: formulating, evaluating, combining and reporting. *Statistics in Medicine, 18*, 321-359.

Norton, L. W., Davis, J. R., Wiens, J. L., Trego, D. C., & Dunnington, G. L. (1984) Accuracy of aspiration cytology in detecting breast cancer. *Surgery, 96*(4), 806-814.

Peng, C. Y. J. (2009). *Data analysis using SAS*. Los Angeles, SAGE

Phui-Ly, L., Tse-Jia, L., Mao-Chih, H., Hsiu-Pen, L., Ching-Fang, L., Min-Szu, Y., & Chi-Long, C. (2011). Rapid staining and immediate interpretation of fine-needle aspiration cytology for palpable breast lesions: diagnostic accuracy, mammographic, ultrasonographic and histopathologic correlations. *Acta cytological, 55*(1), 30-7.

Reitsma, J. B., Glas, A. S., Rutjes, A.W., Scholten, R. J., Bossuyt, P. M., & Zwinderman, A.H. (2005). Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol, 58*, 982–990

Review Manager (RevMan) version 5.0.16. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008. http://www.ccims.net/ RevMan/RevMan5. Accessed June 1, 2011.

Rutter, C. M., & Gatsonis, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Statistics in Medicine, 20, 2865–84.

SAS Institute Inc: SAS/STAT(r) 9.1 User's Guide. Cary, NC: SAS Institute Inc; 2004.

Scheidler, J., Hricak, H., Yu, K. K., Subak, L., & Segal, M.R. (1997). Radiological evaluation of lymph node metastases in patients with cervical cancer. A meta-analysis. *JAMA, 278,* 1096–1101.

Senn, S. (1994). Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials (letter). *Statistics in Medicine, 13,* 293–6.

Sheikh, F. A., Tinkoff, G. H., Kline, T. S., & Neal, H. S. (1987). Final diagnosis by fine-needle aspiration biopsy for definitive operation in breast cancer. *Am J Surg, 154*(5), 470-474.

Silverman, J. F., Lannin, D. R., O'Brien, K., & Norris, H. T. (1987). The triage role of fine needle aspiration biopsy of palpable breast masses. Diagnostic accuracy and cost-effectiveness. *Acta Cytological, 31*(6), 731-736.

Smith, C., Butler. J., Cobb, C., & State, D. (1988). Fine-needle aspiration cytology in the diagnosis of primary breast cancer. *Surgery, 103*(2), 178-183.

Somers, D. (1985). Fine-needle aspiration biopsy in the management of solid breast tumors. *Arch Surg, 120*(6), 673-677.

Sutton, A.J. Abrams, K. R., Jones, D. R., Sheldon, T. A., & Song, F. (2000). Methods for meta-analysis in medical research. Chichester, New York, Wiley.

Somers, R.G., Young, G.P., Kaplan, M.J., Bernhard, V.M., Rosenberg, M., & Sutton, A., Abrams, K., Jones, D., Sheldon, T., & Song, F. (2000). *Methods for meta-analysis in medical research.* Wiley: Chichester.

Temel, M. A. (2000). *Tanı testlerinin meta-analizi: özet işlem karakteristiği (SROC) ve bir uygulama* (PhD's thesis). Available from National Thesis Center of the Turkish Council of Higher Education Dissertations and Theses database. (UMI No. 99229)

Thompson, S. G. (1994). Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal, 309*(6965), 1351-1355.

Thompson, S. G., & Pocock, S. (1991). Can meta-analysis be trusted? *Lancet, 338*, 1127-1130.

Toft, N., & Nielsen, S. S. (2009). Summary receiver operating characteristics (SROC) and hierarchical SROC models for analysis of diagnostic test evaluations of antibody ELISAs for paratuberculosis. *Preventive Veterinary Medicine, 92*(3), 249-255.

Van Houwelingen, J. C. (1995). Meta-analysis; Methods, limitations and applications. *Biocybernetics and Biomedical Engineering, 15*(1-2), 53-61.

Van Houwelingen, J. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine, 21*, 589–624.

Vetrani, A., Fulciniti, F., Di Benedetto, G., Zeppa, P., Troncone, G., Boscaino, A., de Rosa, G., & Palombini, L. (1992). Fine-needle aspiration biopsies of breast masses. An additional experience with 1153 cases (1985 to 1988) and a meta-analysis. *Cancer, 69*(3), 736-740.

Walter, S. D. (2002). Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Statistics in Medicine, 21*, 1237–56.

Walter, S. D., Irwig, L., & Glasziou, P. P. (1999). Meta-analysis of diagnostic tests with imperfect reference standards. *J Clin Epidemiol, 52*, 943–51.

Wang, H. H., & Ducatman, B. S. (1998). Fine needle aspiration of the breast. A probabilistic approach to diagnosis of carcinoma. *Acta Cytological, 42*(2), 285-289.

Wang, M. C., & Bushman B. J. (1999). Integrating results through meta-analytic review using SAS software. Cary, NC: SAS Institute Inc.

Watson, D. P., McGuire, M., Nicholson, F., & Given, H. F. (1987). Aspiration cytology and its relevance to the diagnosis of solid tumors of the breast. *Surg Gynecol Obstet, 165*(5), 435-441.

Wilkinson, E. J., Schuettke, C. M., Ferrier, C. M., Franzini, D. A., & Bland, K. I. (1989). Fine needle aspiration of breast masses. An analysis of 276 aspirates. *Acta Cytological, 33*(5), 613-619.

Yach, D. (1990). Meta-analysis in Epidemiology. *S-Afr-Med-J, 78*, 94-97

Zardawi, I. M. (1998). Fine needle aspiration cytology in a rural setting. *Acta Cytological, 42*(4), 899-906.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem, 39*,561-77.

APPENDICES

APPENDIX A: SAS CODES

```
/*****************************************************************/
/* 1. Data Files */
/*****************************************************************/
%let nStudies = 25;
%let n_minus_1 = 24;

DATA FNAC;
length author $30;
input author $ number year N TP TN FP FN;
RP=TP+FN;
RN=TN+FP;
study = _N_;
datalines;
vetrani et al. 1 1992 256 136 108 7 5
watson et al. 2 1987 260 46 200 1 13
sheikh et al. 3 1987 2263 293 2290 40 0
griffith et al. 4 1986 236 110 95 15 16
atamdede et al. 5 1993 51 32 17 1 1
ciatto et al. 6 1989 563 489 60 1 13
gelabert et al. 7 1990 107 90 12 0 5
horgon et al. 8 1991 1742 222 1471 11 38
lannin et al. 9 1986 93 26 65 0 2
norton et al. 10 1984 37 10 19 6 2
collaço et al. 11 1999 260 175 69 1 15
arıkan et al. 12 1992 134 63 69 0 2
langmuir et al. 13 1989 101 24 65 11 1
smith et al. 14 1988 317 113 181 15 8
dominguez et al. 15 1997 427 158 247 11 11
feichter et al. 16 1997 323 153 145 1 24
wilkinson et al. 17 1989 240 27 206 0 7
silverman et al. 18 1987 93 33 47 2 6
wang et al. 19 1998 165 114 45 3 3
zardawi et al. 20 1998 437 100 329 3 5
```

```
somers et al. 21 1985 185 80 82 0 23
kaufman et al. 22 1994 234 102 120 4 8
chaiwun et al. 23 2002 424 194 193 1 36
mizuno et al. 24 2005 94 72 14 1 7
ariga et al. 25 2002 1058 814 222 3 19
;
run;
PROC PRINT data=FNAC;
run;

DATA bivariate_FNAC; set FNAC;
status='SEN'; true=TP; pos=TP; total=RP; output;
status='SPEC'; true=TN; pos=FP; total=RN; output;
keep study status true pos total; run;

/*****************************************************************/
/* 2. Bivariate generalized linear fixed-effects model */
/*****************************************************************/
PROC GLIMMIX data=bivariate_FNAC method=quad;
title 'Bivariate generalized linear fixed-effects model GLMMIX';
class study status;
model true/total=status/noint s cl corrb covb df=&n_minus1;
estimate 'logit_sensitivity' status 1 0 / cl ilink;
estimate 'logit_specificity' status 0 1 / cl ilink;
estimate 'LOR' status 1 1 / cl exp;
run;

/*****************************************************************/
/* 3. Bivariate generalized linear random-effects model */
/*****************************************************************/
PROC GLIMMIX data=bivariate_FNAC method=quad;
title 'Bivariate generalized linear random-effects model GLMMIX';
class study status;
model true/total = status / noint s cl corrb covb ddfm=bw;
random status / subject=study S type=chol G;
estimate 'logit_sensitivity' status 1 0 / cl ilink;
estimate 'logit_specificity' status 0 1 / cl ilink;
estimate 'LOR' status 1 1 / cl exp;
run;

/*****************************************************************/
/* 4. SROC nonlinear fixed-effects model */
/*****************************************************************/
PROC NLMIXED data=bivariate_FNAC tech=quanew df=&n_minus1;
title 'SROC nonlinear fixed-effects model NLMIXED';
parms Theta=0 Alpha=0;
if (status='SEN') then eta = Theta + Alpha/2;
```

```
if (status='SPEC') then eta = Theta - Alpha/2;
pi = exp(eta)/(1+exp(eta));
model pos ~ binomial(total,pi);

mu_sen = Alpha/2 + Theta;
mu_spec = Alpha/2 - Theta;

estimate 'logit_sens' mu_sen;
estimate 'logit_spec' mu_spec;
estimate 'LOR' mu_sen + mu_spec;
estimate 'sens' exp(mu_sen)/(1+exp(mu_sen));
estimate 'spec' exp(mu_spec)/(1+exp(mu_spec));
estimate 'DOR' exp(mu_sen + mu_spec);
run;

/*****************************************************************/
/* 5. HSROC nonlinear random-effects model */
/*****************************************************************/
PROC NLMIXED data=bivariate_FNAC tech=quanew;
title 'HSROC nonlinear random-effects model NLMIXED';
parms Theta=0 Alpha=0 Beta=0 s2ut=0 s2ua=0;
bounds s2ut >= 0, s2ua >= 0;
if (status=' SEN') then eta = exp(-Beta/2)*(Theta+ut +(Alpha+ua)/2);
if (status='SPEC') then eta = exp( Beta/2)*(Theta+ut -(Alpha+ua)/2);
pi = exp(eta)/(1+exp(eta));
model pos ~ binomial(total,pi);
random ut ua ~ normal([0,0],[s2ut,0,s2ua]) subject=study;

mu_sen = exp(-Beta/2) * (Alpha/2 + Theta);
mu_spec = exp( Beta/2) * (Alpha/2 - Theta);
var_usen = exp(-Beta) * (s2ut + s2ua/4);
var_uspec = exp( Beta) * (s2ut + s2ua/4);
cov_usenspec= s2ua - s2ut/4;

estimate 'var_usen' var_usen;
estimate 'var_uspec' var_uspec;
estimate 'cov_usenspec' cov_usenspec;
estimate 'logit_sensitivity' mu_sen;
estimate 'logit_specificity' mu_spec;
estimate 'LOR' mu_sen + mu_spec;
estimate 'sensitivity' exp(mu_sen)/(1+exp(mu_sen));
estimate 'specificity' exp(mu_spec)/(1+exp(mu_spec));
estimate 'DOR' exp(mu_sen + mu_spec);
run;
```

APPENDIX B: SAS RESULTS

Obs	author	number	year	n	TP	TN	FP	FN	RP	RN	study
1	Vetrani et al.	1	1992	256	136	108	7	5	141	115	1
2	Watson et al.	2	1987	260	46	200	1	13	59	201	2
3	Sheikh et al.	3	1987	2263	293	2290	40	0	293	2330	3
4	Griffith et al.	4	1986	236	110	95	15	16	126	110	4
5	Atamdede et al.	5	1993	51	32	17	1	1	33	18	5
6	Ciatto et al.	6	1989	563	489	60	1	13	502	61	6
7	Gelabert et al.	7	1990	107	90	12	0	5	95	12	7
8	Horgon et al.	8	1991	1742	222	1471	11	38	260	1482	8
9	Lannin et al.	9	1986	93	26	65	0	2	28	65	9
10	Norton et al.	10	1984	37	10	19	6	2	12	25	10
11	Collaço et al.	11	1999	260	175	69	1	15	190	70	11
12	Arıkan et al.	12	1992	134	63	69	0	2	65	69	12
13	Langmuir et al.	13	19 9	10	24	65	11	1	25	76	13
14	Smith et al.	14	1988	317	113	181	15	8	121	196	14
15	Dominguez et al.	15	1997	427	158	247	11	11	1	25	15
16	Feichter et al.	16	1997	323	153	145	1	24	177	14	16
17	Wilkinson et al.	17	1989	240	27	206	0	7	34	206	17
18	Silverman et al.	18	1987	93	33	47	2	6	39	49	18
19	Wang et al.	19	1998	165	114	45	3	3	117	48	19
20	Zardawi et al.	20	1998	437	100	329	3	5	105	332	20
21	Somers et al.	21	1985	185	80	82	0	23	103	82	21
22	Kaufman et al.	22	1994	234	102	120	4	8	110	124	22
23	Chaiwun et al.	23	2002	424	194	193	1	36	230	194	23
24	Mizuno et al.	24	2005	94	72	14	1	7	79	15	24
25	Ariga et al.	25	2002	1058	814	222	3	19	833	225	25

Bivariate generalized linear fixed-effects model

The GLIMMIX Procedure

Model Information	
Data Set	WORK.BIVARIATE_FNAC
Response Variable (Events)	true
Response Variable (Trials)	total
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix	Diagonal
Estimation Technique	Maximum Likelihood
Degrees of Freedom Method	Residual

Class Level Information		
Class	**Levels**	**Values**
study	25	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
status	2	SEN SPEC

Number of Observations Read	50
Number of Observations Used	50
Number of Events	10047
Number of Trials	10455

Dimensions	
Columns in X	2
Columns in Z	0
Subjects (Blocks in V)	1
Max Obs per Subject	50

Optimization Information	
Optimization Technique	Newton-Raphson
Parameters in Optimization	2
Lower Boundaries	0
Upper Boundaries	0
Fixed Effects	Not Profiled

Iteration History					
Iteration	**Restarts**	**Evaluations**	**Objective Function**	**Change**	**Max Gradient**
0	**0**	4	299.49231296	.	102.9168
1	**0**	3	265.15617378	34.33613918	13.12135
2	**0**	3	264.28077188	0.87540191	0.541436
3	**0**	3	264.27958337	0.00118851	0.001034
4	**0**	3	264.27958337	0.00000000	3.788E-9

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
status	2	24	1849.33	<.0001

Parameter Estimates									
Effect	status	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
status	SEN	2.6112	0.06305	24	41.41	<.0001	0.05	2.4810	2.7413
status	SPEC	3.8323	0.08604	24	44.54	<.0001	0.05	3.6547	4.0098

Covariance Matrix for Fixed Effects				
Effect	status	Row	Col1	Col2
status	SEN	1	0.003976	
status	SPEC	2		0.007403

Correlation Matrix for Fixed Effects				
Effect	status	Row	Col1	Col2
status	SEN	1	1.0000	
status	SPEC	2		1.0000

Estimates															
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Mean	Standard Error Mean	Lower Mean	Upper Mean	Exponentiated Estimate	Exponentiated Lower	Exponentiated Upper
logit_sens	2.6112	0.06305	24	41.41	<.0001	0.05	2.4810	2.7413	0.9316	0.004019	0.9228	0.9394	.	.	.
logit_spec	3.8323	0.08604	24	44.54	<.0001	0.05	3.6547	4.0098	0.9788	0.001786	0.9748	0.9822	.	.	.
LOR	6.4434	0.1067	24	60.40	<.0001	0.05	6.2233	6.6636					628.55	504.34	783.35

Bivariate generalized linear random-effects model

The GLIMMIX Procedure

Model Information	
Data Set	WORK.BIVARIATE_FNAC
Response Variable (Events)	true
Response Variable (Trials)	total
Response Distribution	Binomial
Link Function	Logit
Variance Function	Default
Variance Matrix Blocked By	study
Estimation Technique	Maximum Likelihood
Likelihood Approximation	Gauss-Hermite Quadrature
Degrees of Freedom Method	Between-Within

Class Level Information		
Class	**Levels**	**Values**
study	25	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
status	2	SEN SPEC

Number of Observations Read	50
Number of Observations Used	50
Number of Events	10047
Number of Trials	10455

Dimensions	
G-side Cov. Parameters	3
Columns in X	2
Columns in Z per Subject	2
Subjects (Blocks in V)	25
Max Obs per Subject	2

Optimization Information	
Optimization Technique	Dual Quasi-Newton
Parameters in Optimization	5
Lower Boundaries	2
Upper Boundaries	0
Fixed Effects	Not Profiled
Starting From	GLM estimates
Quadrature Points	5

Iteration History					
Iteration	**Restarts**	**Evaluations**	**Objective Function**	**Change**	**Max Gradient**
0	**0**	4	313.85951729	.	91.73475
1	**0**	2	293.26769221	20.59182508	9.449572
2	**0**	2	292.88399358	0.38369864	6.577754
3	**0**	2	292.3046356	0.57935797	4.633307

4	**0**	3	292.1237926	0.18084300	3.169826
5	**0**	2	291.98278248	0.14101012	1.21244
6	**0**	3	291.95254377	0.03023871	0.228667
7	**0**	3	291.95158976	0.00095401	0.012346
8	**0**	3	291.95158556	0.00000420	0.003001
9	**0**	3	291.95158544	0.00000012	0.000027

Convergence criterion (GCONV=1E-8) satisfied.

Estimated G Matrix

Effect	status	Row	Col1	Col2
status	SEN	**1**	0.7786	-0.3065
status	SPEC	**2**	-0.3065	1.6773

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error
CHOL(1,1)	study	0.8824	0.1619
CHOL(2,1)	study	-0.3473	0.3456
CHOL(2,2)	study	1.2477	0.2537

Solutions for Fixed Effects

Effect	status	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
status	SEN	2.6304	0.2001	23	13.14	<.0001	0.05	2.2164	3.0443
status	SPEC	3.9010	0.3149	23	12.39	<.0001	0.05	3.2496	4.5525

Type III Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
status	2	23	202.82	<.0001

Solution for Random Effects

Effect	status	Subject	Estimate	Std Err Pred	DF	t Value	Pr > |t|
status	SEN	study 1	0.5706	0.4197	48	1.36	0.1803
status	SPEC	study 1	-1.0782	0.4881	48	-2.21	0.0320
status	SEN	study 2	-1.2208	0.3604	48	-3.39	0.0014
status	SPEC	study 2	1.0729	0.7501	48	1.43	0.1591
status	SEN	study 3	2.0068	0.5668	48	3.54	0.0009
status	SPEC	study 3	0.1315	0.3477	48	0.38	0.7070
status	SEN	study 4	-0.6039	0.3250	48	-1.86	0.0693
status	SPEC	study 4	-1.9425	0.4214	48	-4.61	<.0001
status	SEN	study 5	0.4485	0.6191	48	0.72	0.4723
status	SPEC	study 5	-0.6690	0.9167	48	-0.73	0.4691
status	SEN	study 6	0.9029	0.3150	48	2.87	0.0061
status	SPEC	study 6	-0.01050	0.7786	48	-0.01	0.9893
status	SEN	study 7	0.1936	0.4282	48	0.45	0.6532
status	SPEC	study 7	0.2216	1.1117	48	0.20	0.8428
status	SEN	study 8	-0.8368	0.2604	48	-3.21	0.0023
status	SPEC	study 8	0.9584	0.4129	48	2.32	0.0246
status	SEN	study 9	-0.1059	0.5720	48	-0.19	0.8539
status	SPEC	study 9	0.8818	0.9528	48	0.93	0.3593

status	SEN	study 10	-0.2676	0.7295	48	-0.37	0.7154
status	SPEC	study 10	-2.3615	0.5906	48	-4.00	0.0002
status	SEN	study 11	-0.1616	0.3167	48	-0.51	0.6121
status	SPEC	study 11	0.2300	0.7868	48	0.29	0.7713
status	SEN	study 12	0.4577	0.5225	48	0.88	0.3854
status	SPEC	study 12	0.7934	0.9276	48	0.86	0.3966
status	SEN	study 13	0.4416	0.6745	48	0.65	0.5157
status	SPEC	study 13	-1.9943	0.4544	48	-4.39	<.0001
status	SEN	study 14	0.05376	0.3820	48	0.14	0.8887
status	SPEC	study 14	-1.3474	0.4095	48	-3.29	0.0019
status	SEN	study 15	0.04643	0.3436	48	0.14	0.8931
status	SPEC	study 15	-0.7443	0.4293	48	-1.73	0.0894
status	SEN	study 16	-0.7379	0.2879	48	-2.56	0.0136
status	SPEC	study 16	0.7935	0.7506	48	1.06	0.2958
status	SEN	study 17	-1.0757	0.4434	48	-2.43	0.0191
status	SPEC	study 17	1.6562	0.8956	48	1.85	0.0706
status	SEN	study 18	-0.6952	0.4561	48	-1.52	0.1341
status	SPEC	study 18	-0.4637	0.7382	48	-0.63	0.5329
status	SEN	study 19	0.7638	0.4725	48	1.62	0.1125
status	SPEC	study 19	-1.0164	0.6416	48	-1.58	0.1197
status	SEN	study 20	0.2605	0.4227	48	0.62	0.5407
status	SPEC	study 20	0.6463	0.5604	48	1.15	0.2545
status	SEN	study 21	-1.2989	0.3029	48	-4.29	<.0001
status	SPEC	study 21	1.2480	1.0125	48	1.23	0.2238
status	SEN	study 22	-0.05913	0.3808	48	-0.16	0.8773
status	SPEC	study 22	-0.4231	0.5595	48	-0.76	0.4533
status	SEN	study 23	-0.9125	0.2643	48	-3.45	0.0012
status	SPEC	study 23	1.0101	0.7379	48	1.37	0.1774
status	SEN	study 24	-0.2241	0.4062	48	-0.55	0.5837
status	SPEC	study 24	-0.6214	0.9611	48	-0.65	0.5210
status	SEN	study 25	1.0484	0.2864	48	3.66	0.0006
status	SPEC	study 25	0.2655	0.5688	48	0.47	0.6428

Covariance Matrix for Fixed Effects

Effect	status	Row	Col1	Col2
status	SEN	**1**	0.04004	-0.01235
status	SPEC	**2**	-0.01235	0.09918

Correlation Matrix for Fixed Effects

Effect	status	Row	Col1	Col2
status	SEN	**1**	1.0000	-0.1960
status	SPEC	**2**	-0.1960	1.0000

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Mean	Standard Error Mean	Lower Mean	Upper Mean	Exponentiated Estimate	Exponentiated Lower	Exponentiated Upper
logit_sens	2.6304	0.2001	23	13.14	<.0001	0.05	2.2164	3.0443	0.9328	0.01254	0.9017	0.9545	.	.	.
logit_spec	3.9010	0.3149	23	12.39	<.0001	0.05	3.2496	4.5525	0.9802	0.006118	0.9627	0.9896	.	.	.
LOR	6.5314	0.3384	23	19.30	<.0001	0.05	5.8314	7.2314					686.37	340.83	1382.22

SROC nonlinear fixed-effects model

The NLMIXED Procedure

Specifications	
Data Set	WORK.BIVARIATE_FNAC
Dependent Variable	pos
Distribution for Dependent Variable	Binomial
Optimization Technique	Dual Quasi-Newton
Integration Method	None

Dimensions	
Observations Used	50
Observations Not Used	0
Total Observations	50
Parameters	2

Parameters		
Theta	Alpha	NegLogLike
0	0	5858.10359

Iteration History					
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	2	2542.5988	3315.505	1603.68	-78048.8
2	3	516.691116	2025.908	387.3721	-41949.3
3	5	339.959859	176.7313	231.9965	-108.138
4	6	285.814665	54.14519	126.5521	-246.815
5	7	267.771924	18.04274	46.34373	-27.2065
6	9	265.58942	2.182504	14.56456	-4.22412
7	11	264.284226	1.305193	1.886044	-0.58884
8	13	264.279585	0.004641	0.017266	-0.00893
9	15	264.279583	1.62E-6	0.000072	-3.11E-6
10	17	264.279583	6.59E-12	4.558E-9	-133E-13

NOTE: GCONV convergence criterion satisfied.

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
Theta	-0.6105	0.05334	24	-11.45	<.0001	0.05	-0.7206	-0.5005	-4.56E-9
Alpha	6.4434	0.1067	24	60.40	<.0001	0.05	6.2233	6.6636	-297E-12

Additional Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
logit_sens	2.6112	0.06305	24	41.41	<.0001	0.05	2.4810	2.7413
logit_spec	3.8323	0.08604	24	44.54	<.0001	0.05	3.6547	4.0098
LOR	6.4434	0.1067	24	60.40	<.0001	0.05	6.2233	6.6636
sens	0.9316	0.004019	24	231.78	<.0001	0.05	0.9233	0.9399
spec	0.9788	0.001786	24	548.18	<.0001	0.05	0.9751	0.9825
DOR	628.55	67.0492	24	9.37	<.0001	0.05	490.17	766.93

HSROC nonlinear random-effects model

The NLMIXED Procedure

Specifications	
Data Set	WORK.BIVARIATE_FNAC
Dependent Variable	pos
Distribution for Dependent Variable	Binomial
Random Effects	ut ua
Distribution for Random Effects	Normal
Subject Variable	study
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions	
Observations Used	50
Observations Not Used	0
Total Observations	50
Subjects	25
Max Obs Per Subject	2
Parameters	5
Quadrature Points	1

Parameters					
Theta	Alpha	Beta	s2ut	s2ua	NegLogLike
0	0	0	0	0	5858.10359

Iteration History					
Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	18	4358.25008	1499.854	117307.7	-5.83E9
2	19	2657.26939	1700.981	43160.87	-1.22E10
3	35	1791.59488	865.6745	9178.802	-2.041E9
4	36	408.815063	1382.78	1301.263	-2828.72
5	38	242.098415	166.7166	490.8461	-735.57
6	40	191.050923	51.04749	198.3244	-111.725
7	42	189.87713	1.173793	199.5491	-2.29045
8	44	188.212645	1.664485	199.7302	-0.13978
9	45	185.66206	2.550585	178.1634	-2.28702
10	46	181.780953	3.881108	145.0935	-1.74165
11	48	179.656029	2.124924	126.9116	-2.51794
12	50	178.832769	0.82326	113.4607	-1.24976
13	51	177.703337	1.129432	99.38527	-0.43219
14	53	172.708709	4.994627	70.28812	-1.34913
15	55	163.895016	8.813693	37.24912	-4.44868
16	56	158.574185	5.320832	36.37896	-27.2365
17	57	156.562106	2.012079	20.26779	-3.56816
18	58	153.479769	3.082337	19.28687	-7.90541
19	60	152.491977	0.987792	17.23299	-1.04871
20	62	152.057851	0.434126	14.23893	-0.34587

21	64	151.814103	0.243747	10.96705	-0.16936
22	65	151.651423	0.16268	9.025893	-0.07403
23	67	151.55523	0.096193	8.913537	-0.15341
24	69	151.543275	0.011955	8.518309	-0.01061
25	70	151.538022	0.005252	8.704062	-0.0045
26	71	151.534083	0.003939	8.706584	-0.00905
27	73	151.533191	0.000892	8.750883	-0.00168
28	75	151.533014	0.000177	8.842063	-0.00007
29	78	151.520694	0.01232	8.680679	-0.00016
30	79	151.507125	0.013569	8.873015	-0.02341
31	81	151.501501	0.005624	8.782307	-0.00236
32	87	149.811326	1.690176	10.36507	-0.00892
33	89	149.433214	0.378112	5.922718	-1.75637
34	91	149.204005	0.229209	7.393754	-0.25731
35	93	149.086757	0.117248	8.276728	-0.14608
36	95	148.73054	0.356217	8.24024	-0.29055
37	97	147.584285	1.146254	4.339049	-1.29564
38	98	146.828959	0.755327	8.866983	-1.01934
39	99	146.157827	0.671131	2.706495	-1.86982
40	101	146.087918	0.069909	0.255826	-0.10802
41	103	146.077304	0.010614	0.196091	-0.01462
42	105	146.074785	0.00252	0.062335	-0.0033
43	107	146.074745	0.000039	0.004382	-0.00007
44	109	146.074744	1.263E-6	0.000543	-1.56E-6
45	111	146.074744	2.532E-8	0.000062	-4.24E-8

NOTE: GCONV convergence criterion satisfied.

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
Theta	-0.01907	0.4366	23	-0.04	0.9655	0.05	-0.9223	0.8842	0.000019
Alpha	6.4112	0.3155	23	20.32	<.0001	0.05	5.7585	7.0639	0.000028
Beta	0.3829	0.2674	23	1.43	0.1657	0.05	-0.1704	0.9361	0.000019
s2ut	0.7247	0.2681	23	2.70	0.0127	0.05	0.1701	1.2794	0.000062
s2ua	1.6236	0.6490	23	2.50	0.0199	0.05	0.2811	2.9661	-4.95E-6

Additional Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
var_usen	0.7710	0.2816	23	2.74	0.0117	0.05	0.1885	1.3534
var_uspec	1.6581	0.6723	23	2.47	0.0215	0.05	0.2674	3.0488
cov_usenspec	1.4424	0.6519	23	2.21	0.0371	0.05	0.09382	2.7910
logit_sens	2.6313	0.1993	23	13.20	<.0001	0.05	2.2190	3.0437
logit_spec	3.9051	0.3142	23	12.43	<.0001	0.05	3.2552	4.5550
LOR	6.5364	0.3358	23	19.47	<.0001	0.05	5.8419	7.2310
sens	0.9329	0.01249	23	74.71	<.0001	0.05	0.9070	0.9587
spec	0.9803	0.006080	23	161.24	<.0001	0.05	0.9677	0.9928
DOR	689.81	231.61	23	2.98	0.0067	0.05	210.70	1168.93