INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI®
800-521-0600

Modeling outcome estimates in meta-analysis using fixed and mixed effects linear models

presented

by

Asmaâ Mansour

Department of Mathematics and Statistics

Faculty of Arts and Sciences

McGill University, Montreal

July, 1998

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements of the degree of

Master's of Science (M.Sc.) in Statistics

1998

McGill University

Graduate faculty

This thesis is entitled:

Modeling outcome estimates in meta-analysis using fixed and mixed effects linear model

presented

by

Asmaâ Mansour

has been evaluated by a jury made of the following:

Dr. François Bellavance
(thesis supervisor)

Dr. Robert Platt
(external examiner)

Thesis accepted:

August, 1998

SUMMARY

The main objective of this thesis is to present a quantitative method for modeling data collected from different studies on a same research topic. This quantitative method is called meta-analysis.

The first step of a meta-analysis is the literature search, conducted using computerized and manual search strategies to identify relevant studies. The second step is the data abstraction from different relevant papers. In general, at least two independent raters systematically abstract the information, and interrater reliability check is performed.

The next step is the quantitative analysis of the abstracted data. For this purpose, it is possible to use either fixed or mixed effects linear model. Under the fixed effects model, only the variability due to sampling error is considered. In contrast, under the mixed effects model, an additional random effects variance is being considered. Both, the method of moments and the method of maximum likelihood can be used to estimate the parameters of the model.

Finally, the use of the above mentioned models and methods of estimation is illustrated with a data set on the prognosis of depression in the elderly, made available by Dr. Martin Cole from the Department of Psychiatry at St. Mary's Hospital Center in Montreal.

SOMMAIRE

L'objectif de ce mémoire est de présenter une méthode pour modéliser les données provenant de différentes études sur un même sujet de recherche. Cette méthode quantitative d'analyse est appelée méta-analyse.

La première étape d'une méta-analyse est la revue de la littérature, réalisée à l'aide de stratégies de recherche informatisée ou manuelle afin d'identifier les études pertinentes. La deuxième étape consiste à extraire les données des différents articles pertinents. En général, cette tâche est faite de façon systématique par au moins deux investigateurs indépendants, et la fiabilité de leurs performances est ensuite vérifiée.

L'étape suivante est l'analyse quantitative des données préalablement extraites. Pour ce faire, il est possible de choisir entre le modèle à effets fixes ou celui à effets mixtes. En présence du modèle à effets fixes, seulement la variabilité de l'erreur échantillonale est considerée. Par contre, sous le modèle à effets mixtes, un terme supplémentaire ayant un effet aléatoire est consideré. La méthode des moments et celle du maximum de vraisemblance peuvent être utilisées pour estimer les paramètres du modèle.

Finalement, les modèles ainsi que les méthodes d'estimation mentionnés ci-haut sont illustrés à l'aide d'un jeu de donnée sur le pronostic de la dépression chez les personnes âgées, mis à notre disposition par le Dr. Martin Cole du Département de Psychiatrie du Centre Hospitalier de St. Mary's à Montréal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

First, I would like to take this opportunity to thank my thesis supervisor, Dr. François Bellavance for his invaluable advice, availability and patience.

Secondly, I would like to thank my family, and most of all my parents for their love and support, because without them I would not have come this far.

Thirdly, I would like to thank Dr. Martin Cole for providing me with the data set and for his cooporation and assistance throughout my year at St. Mary's .

Finally, I would like to thank my co-worker, Eric Belzile for his advice in computer programming.

In conclusion, special thanks goes to the staff of the Department of Epidemiology and Community Studies for making my work experience at St. Mary's Hospital a great one, I shall never forget.

INTRODUCTION

Meta-analysis is a quantitative method which allows us to model results abstracted from different studies on the same research topic. This method permits us to establish relations and to draw conclusions on many research areas dealing with the same topic of interest in medicine, education, social science, etc. . .

If we want to present a brief history on the origin and development of meta-analysis, one would go back to 1896 when the famous biometrician, Karl Pearson, had been asked to review the efficacy of the typhoid vaccine developed in that same year by Sir Almroth Wright. After being tested in different settings, this vaccine had been recommended for routine use in the British army for soldiers at risk for the disease. Karl Pearson had reviewed the empirical evidence from five studies reporting data about the relationship between inoculation status and typhoid immunity, and six studies reporting data on inoculation status and fatality among those who contracted the disease. He computed tetrachoric correlations for each of these eleven cases, and then averaged these correlations to describe average inoculation effectiveness. In an article published in 1904, Pearson concluded that the average correlations were too low to warrant adopting the vaccine, since other accepted vaccines at that time produced correlations at or far above his findings, quoting: "I think the right conclusion to draw would be not that it was desirable to inoculate the whole army, but that improvement of the serum and method of

dosing, with a view to a far higher correlation, should be attempted" (p. 1245). It was with Glass (1976) that the term "meta-analysis" had been introduced for the first time, to refer to "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3). The methodology for combining estimates across studies goes back to 1932 when many papers had been included in the physical sciences by Birge and in statistics in 1937 by Cochran and in 1938 by Yates and Cochran. Whereas the methodology for combining probabilities across studies dates at least from procedures suggested in Tippett's *Method of Statistics* (1931) and Fisher's *Statistical methods for Research Workers* (1932).

Four books had appeared in the first half of 1980's, primary devoted to quantitative methods used in meta-analysis. Glass, McGaw, and Smith were the first to introduce in 1981 analysis of variance and multiple regression approach in meta-analysis with the effect sizes as the dependent variable. Hunter, Schmidt, and Jackson introduced in 1982, procedures in meta-analysis focusing on comparing the observed variation in study outcomes with that expected by chance, and considering sources of bias for correcting observed estimates and their variances. Rosenthal presented in 1984 new techniques, in the combining of significance levels, effect size estimated, and the analysis of variation in effect sizes involving assumptions, specially made for the analysis of study outcomes. Finally in 1985, Hedges and Olkin published *Statistical Methods for Meta-analysis* which helped elevate meta-analysis to an independent specialty within statistical science.

In this thesis, we shall focus primarily on two models to analyze the results of different studies namely, the fixed and the mixed effects models using a regression approach to encounter for covariates that may explain the variability between study results. In the first part, we present a brief overview of the methods used to identify articles relevant to the topic of research namely, the identification of computerized databases, and the design of the search strategies, followed by data abstraction. In the second part, the theory of the fixed effects model is presented. Under this model, all study results differ from each other only by virtue of having used just a sample of observations from the total population so that observed studies yield results that differed from the true population parameters only by sampling error. In the third part, the theory of the mixed effects model is described. Under this latter model, one should not assume that there is one overall population parameter, but rather that a distribution of population parameters exists, generated by a distribution of possible realizations. Hence, observed results in studies differ from each other not just because of sampling error but also by the true underlying differences. Both, the method of moments and the method of maximum likelihood are used to obtain estimates of the model parameters and to further make inference about them. Finally, we apply the theory of the two models to a data set made available by Dr. Martin Cole from the Department of Psychiatry at St. Mary's Hospital Center. This data set deals with the prognosis of depression in the elderly in primary care and community based settings.

CHAPTER 1

LITERATURE REVIEW

The objective of this chapter is to briefly introduce methods for identifying articles relevant to a specific research topic, through literature search techniques usually done via computerized databases. Besides, systematic data abstraction from the relevant literature, interrater reliability, assessment of validity, and publication bias will be discussed.

1.1 LITERATURE SEARCH

A proper application of the method of meta-analysis begins with the development of a systematic and explicit procedures for identifying studies. The first step in information retrieval is almost always a search of personal files of the investigator and discussion with knowledgeable colleagues to identify materials that are already in hand, about the particular research topic of interest. This research is usually followed by a computerized search of one or more computer databases.

1.1.1 COMPUTERIZED DATABASES

There exists several computerized databases in different research areas; for example **MEDLINE & PsycINFO** in the field of health sciences, **Education resources**

information center (ERIC), in the field of education, and **Ei Compendex*Plus™ (CDEX),** in the fields of engineering and management. These databases are the most important ones available in the computerized data bank, with almost no overlap between **ERIC** and **CDEX,** about 25% overlap between **MEDLINE** and **PsycINFO** and a small overlap between **PsycINFO** and **ERIC**. In the health sciences, the computerized search virtually always includes **MEDLINE**.

Below is a brief description of each of these computerized databases.

1.1.1.1 MEDLINE

MEDLINE is a bibliographic database that is the computerized counter part of *index medicus*. It is the primary source of information on publications in the biomedical literature. It encompasses information from *Index Medicus*, *index to Dental literature*, and *International Nursing*, as well as other sources of coverage in the areas of allied health, biological and physical sciences, humanities and information science as they relate to medicine and health care. It contains 8.7 million records from more than 3600 journals, and covers *Indexes* from the period 1966 onward. **MEDLINE** is not a full-text database. That is the complete text of publication is not available in computer-stored form. Rather, for each indexed publication, **MEDLINE** contains the title, the authors, the source of publication, the journal title, the volume number, and page numbers; the abstract, if it is available (the abstracts are included for about 67% of the records), and a fair number of medical "subject headings" (MeSH) terms. The MeSH terms are chosen from a limited vocabulary, and they are assigned to each published article by a

professional indexer working under a set of highly structured rules. These MeSH terms are extremely important in developing literature search strategies to identify relevant articles for meta-analysis

1.1.1.2 PsycINFO

The **PsycINFO** database covers the professional and academic literature in psychology and related disciplines such as medicine, psychiatry, nursing, sociology, education, pharmacology, physiology, and linguistics. **PsycINFO** includes references and abstracts to over 1300 journals in more than 30 languages, and to book chapters and books in the English language. Over 50000 references are added annually covering indexes from 1967 onward.

1.1.1.3 ERIC

ERIC is presently the largest education database in the world. It contains over 700000 citations covering research documents, journal articles, technical reports, thesis, and curricular material in the field of education. This database is a key source for education information to researchers, teachers, librarians, journalists, and students. **ERIC** covers indexes from the period 1966 onward.

1.1.1.4 CDEX

CDEX is a recently indexed database which provides abstracts and full bibliographic citations for worldwide engineering and technical literature. It encompasses all engineering disciplines, as well as related fields in science and management. The references in this database are drawn from 2600 published journals, conferences, technical reports, monographs, and other materials, and it covers indexes from 1996 onward.

1.1.2 COMPUTERIZED SEARCH STRATEGY TO IDENTIFY POTENTIALLY RELEVANT STUDIES

The first step before conducting a computerized search is to identify critical "subject headings" terms which describe the topic of interest. The goal, is to use the right set of terms in the right way to accurately describe the topic at the appropriate level of specificity. When this is done well, it would result in a search of high precision with a minimum of "false positives" (irrelevant identified articles) and a small number of "false negatives" (relevant articles not identified). The search strategy should contain "subject headings" terms identified to be inclusion criteria for the meta-analysis. The reviewer should not only avoid search that is too broad and waste time by requiring review and rejection of the many false positives retrieved, but also avoid the search that is too narrow and either, results in an incomplete synthesis or requires another search to locate the many false negatives missed in the initial effort. Librarians who are specialized in electronic searching can often be helpful in developing and executing a search strategy. Librarians may lack knowledge in the subject area of a review but, they are trained to run

searches that are precise. Therefore, it is very important to work closely with a librarian in deciding which databases to search and what terms to use in each database. It is often helpful to provide one or more papers, published on the topic of interest to be able to look for "subject headings" and keywords to develop the search strategy.

A good approach in developing a comprehensive search strategy is to begin with a string of terms that describe the disease or the condition of interest, and join them with the Boolean "OR" operator. The result then can be narrowed down with the Boolean "AND" operator, by using terms that describe the outcome being evaluated, which results in pulling out only articles that address both the condition and the outcome of interest. If we consider the example dealing with the prognosis of depression in the elderly in primary care and community settings, we can clearly notice that the condition of interest is depression in elderly and the outcome of interest is the prognosis of depression. The "subject heading" terms used in this meta-analysis for the computerized literature search strategy were, "depression" **and** "aged" which restrict the search to only the depressed patients 55 years of age and over, linked to the outcome of interest by **and** to {" prognosis" **or** "course" **or** "follow-up"}.

Once the search strategy is being developed, potentially relevant records are retrieved for selection, according to inclusion/exclusion criteria set by the review group, and systematic data abstraction is performed.

1.2 SYSTEMATIC DATA ABSTRACTION

The basis for any excellent scientific research study is the collection of information that is reliable, valid and free of bias. The quality of a research synthesis also depends on these same basic methodological principles. In a meta-analysis, there are usually two levels of data collection or what we may call "data abstraction". First, data that document whether or not each identified potentially relevant paper should be included for the meta-analysis study based on the inclusion/exclusion criteria". Next, for all included studies, data on the characteristics and results of the study need to be abstracted for eventual statistical analysis. In general, data are systematically abstracted onto structured forms that have been pre-tested.

1.2.1 REVIEW OF STUDIES FOR INCLUSION IN THE META-ANALYSIS

Once the computerized databases search has identify a set of potentially relevant articles, the first step is to review each of them according to a set of inclusion/exclusion criteria set by the review group. For example, the list of inclusion/exclusion criteria for the meta-analysis on the prognosis of depression in the elderly, in primary care and community settings, consisted of the following five inclusions criteria: 1) original research; 2) published in English or French; 3) study population of community residents or primary care patients; 4) mean age of subjects in the study of 60 years or over; 5) reported affective state as an outcome. These criteria are usually checked, using what we

may call "a first level data abstraction form" to keep track on both excluded and included articles and on the reasons for exclusion. The first level data abstraction form should be designed in such a way to explicitly mention the criteria that the investigator is hoping for. The form usually starts by a unique identifier number followed by some general item identification such as the first author's name and initials, the title of the article, the journal title, the volume and page numbers, the date of publication, the screening date and, the reviewers initials. After this step, a list of inclusion/exclusion criteria should be cited using a simple and clear vocabulary. These criteria should be numbered sequentially and the layout should be simple to facilitate data entry.

To save time and energy, an initial screening is done based on the abstracts and/or titles only. For the articles in which the key information is missing in the abstract to make the final decision on including them in the meta-analysis or rather excluding them from it, their full text is retrieved, and each paper is read to determine if it meets the required criteria.

1.2.2 DATA ABSTRACTION FOR STATISTICAL ANALYSIS

Once relevant articles had been selected, according to the list of inclusion/exclusion criteria, the following step is the abstraction of data for statistical analysis. This step is done using what we may call a "second level data abstraction form". Usually, the second level data abstraction form differs from the first level one by adding the information needed for data synthesis in meta-analysis, namely, the outcomes and the characteristics

of the studies which may account completely or partially for the variation across studies regarding the outcome of interest. In the example on the prognosis of depression, the outcome was the affective state, categorized as percent well, depressed, dead, or lost to follow-up. Whereas population settings, mean age, length of follow-up were the main study characteristics considered to potentially explain, totally or in part, the variability in the outcome across studies.

1.2.3 RELIABILITY OF DATA ABSTRACTION

As mentioned previously, an extremely important element of any scientific research study is the reliability of the collected information. Often, the information needed to be abstracted from the papers, in order to perform statistical modeling, is not clearly stated. So it is always better to reduce the error that can result from gathering the bits and pieces of the desired information from different papers than to attempt to control and correct it later on in the research process. However, to free data from collection error, it is highly recommended that the data abstractor should be formally trained for this purpose, and that each item to be abstracted should be done under the assistance of the principle investigator, who would undergo the process of abstracting the data from a single paper with the abstractor observing, using a detailed abstracting instruction form. The early forms abstracted by the newly trained abstractor should be re-abstracted to assess the reliability of the abstraction process.

It may be useful to have two or more independent abstractions. Discrepancies in the judgments of the abstractors can then be adjusted by a consensus of the study abstractors meeting as a committee. Several methods to measuring interrater reliability are available among which we can state, the Kappa coefficient, (FLeiss, 1981), concordance correlation coefficient , (Lin, Biometrics, 1989), and percent agreement, (Fleiss, 1981).

1.2.4 ASSESSMENT OF VALIDITY OR QUALITY

An other important element in any scientific research is the assessment of validity or quality, where this item is considered as important as the assessment of reliability discussed in the previous section.

A measure of validity is usually done using a quality rating system which begins with a listing of elements that define poor or good quality of a research study. These elements can differ from one type of research study to the other. For example, in randomized clinical trials, the items defining quality that the investigator may be interested in considering are: 1) randomization, where this item is regarded as good if it allows each study participant to have the same chance of receiving each intervention, and the investigators could not predict which intervention was next; 2) double blinding, where this item is regarded as good if neither the person doing the assessments nor the study participants could identify the intervention being assessed; 3) withdrawals and dropouts, where this item is regarded as good if the participants who were initially included in the study and did not complete the required period or were not included for the analysis, are

clearly stated as well as the number of withdrawals and the reasons for none completion of the study are mentioned (Jadad et al,1996; Chalmers et al, 1981). On the other hand, in observational or prognosis studies, these elements defining quality can be: 1) the sample size, where this item is regarded as good if the sample of patients is representative and well-defined at a similar point in the course of the disease; 2) follow-up sufficiently long, where this item is regarded as good if the patients had been followed long enough to detect the outcomes of interest; 3) completion of follow-up, where this item is regarded as good if at least a certain percentage of the patients in the cohort had completed the study; 4) objective outcome criteria, where this items is regarded as good if the methods by which the outcomes of a study in terms of diagnostics systems and scales had been clearly stated (Laupasis et al, 1994).

1.2.5 BIAS

When abstracting data, bias can occur in many ways. The abstractor who believe that one treatment is better than the other may select the data from the article in favor of his/her position. Knowledge that a study had been published in a prestigious journal or knowledge that a study was unpublished or that it comes from technical reports or from a master's thesis may lead the abstractor to rate the paper highly in the first instance and rather poorly in the second one on measures of quality, which may lead eventually to a bias in the judgment.

here are many ways to reduce bias in data collection, blinding the abstractor to aspects of publication is one of them. The aspects of the articles that are more likely to influence the abstractors are the authors, the title of the article, the journal title, and the source of funding. To blind the abstractor, each study should be assigned a code and a black marker should be used to cross out all aspects that can make the paper identifiable. Sometimes it is very difficult to totally blind the abstractor to certain type of journals which can be easily identified by their style of writing, layout or page size. Therefore, the inability to blind experienced investigators may be a good argument for hiring an independent person to do data abstraction, when funds are available.

In this chapter, we have presented some of the procedures surrounding literature search, and data abstraction which constitute the body of a systematic review, but once the quantitative data are available, statistical methods need to be used in order the model them. For this purpose two models will be presented in the next two Chapters. First, the fixed effects model, and second the mixed effects model using two methods, namely, the method of moments and the method of maximum likelihood.

CHAPTER 2

FIXED EFFECTS MODEL

Data collected from different studies dealing with the same research topic need to be analyzed and summarized in order to draw conclusions on the particular subject of interest. Modeling the data can be done in many different ways. Among them, two general linear models are available: 1) A fixed effects model and, 2) a mixed effects model.

In this chapter, the focus will be made on describing the fixed effects model. But first, it is important to briefly review the type and format of results from research studies that are generally modeled in a meta-analysis.

2.1 TYPE OF DATA AVAILABLE FROM RESEARCH STUDIES

In research synthesis, the outcome of interest that we want to model can be of different types. For example, in a randomized control trial where subjects are randomly assigned to either a control or treatment group, the difference, at the end of the trial, in means or proportions, depending on the nature of the outcome, i.e. continuous or discrete, between the two groups, would constitute the information of interest to be modeled. In case control studies, where the relative risk or rate ratio expressing the relationship of the

characteristics or attributes, as present or absent, to the disease are usually estimated by the odds ratios. In the cohort studies, the characteristics related to the development of the disease are initially measured in the study population and the outcome of the condition of interest can than be expressed in the form of a proportion. These different types of outcome, summarizing the results of research studies, are often refer to as effect size estimates.

Besides, as mentioned in section 2 of Chapter 1, research studies included in a meta-analysis always differ from each other in many methodological and substantive ways. For example, not all the studies have the same sample size. In general, studies with larger sample sizes will give more accurate estimates of the effect size, i. e. the estimated variance of the effect size, also known as the estimation variance, will be smaller.

Table I below, presents the most important effect sizes encountered in research studies along with their corresponding estimation variance (Cooper, 1994).

Table I : Different types of effect size and their corresponding estimation variance

#	Parameter	Parameter estimate	Estimation variance (v_i)
1	Difference between means	$d_i = \bar{X}_{i1} - \bar{X}_{i2}$	$\frac{n_{i1}+n_{i2}}{n_{i1}n_{i2}}\left[{}^{\bullet}s_{pi}^2\right]$
2	Difference between proportions	$D_i = p_{i1} - p_{i2}$	$\frac{p_{i1}(1-p_{i1})}{n_{i1}} + \frac{p_{i2}(1-p_{i2})}{n_{i2}}$
3	Standardized mean difference	$d_i^{'} = \frac{\bar{X}_{i1} - \bar{X}_{i2}}{{}^{\bullet}s_{pi}}$	$\frac{n_{i1}+n_{i2}}{n_{i1}n_{i2}} + \frac{d_i^{'2}}{2(n_{i1}+n_{i2})}$
4	z-transformed correlation	$\zeta_i = .5\ln\left[\frac{(1+\rho_i)}{(1-\rho_i)}\right]$	$\frac{1}{n_i - 3}$
5	Natural logarithm of odds ratio	$l_i = \ln\left[\frac{p_{i1}(1-p_{i2})}{p_{i2}(1-p_{i1})}\right]$	$\frac{1}{n_{i1}p_{i1}(1-p_{i1})} + \frac{1}{n_{i2}p_{i2}(1-p_{i2})}$
6	Proportions	p_i	$\frac{p_i(1-p_i)}{n_i}$

$${}^{\bullet}s_{pi}^2 = \frac{\left[(n_{i1}-1)s_{i1}^2 + (n_{i2}-1)s_{i2}^2\right]}{\left[n_{i1}+n_{i2}-2\right]}.$$

In research synthesis, it is often desirable to investigate the relationship between the true effects size and the characteristics of research studies, which can usually be presented by a combination of continuous and discrete independent variables. One analytic procedure for investigating these relationships is an analogue to multiple regression analysis, or more generally a linear fixed effects model.

A fixed effects model is a model in which the universe to which generalizations are made consists of a set of studies identical to those in the study sample except for the primary sampling units that appear in the study, and the specific characteristics of the study, where the universe is the hypothetical collection of studies that could be conducted in principle and about which we wish to generalize, and where the study sample is the set of studies used in the meta-analysis. Since, the studies in the universe differ from those of the study sample only as a result of the sampling of people into the groups of the studies, the only source of sampling error is the variation resulting from the sampling of the people into studies.

In the next sections, we present methods to fitting effect sizes and to making inference about independent variables which correspond basically to the study characteristics. The fixed effects model will provide estimates of the parameters with their significance levels and confidence intervals .

2.2 MODEL AND NOTATION

Given a collection of k studies to be considered in a meta-analysis, we can denote the true population effect size for study i, where $i = 1,\ldots,k$, by θ_i, its estimate by T_i and the estimation variance of T_i by v_i. Schematically we have:

Study	True population effect size	Estimated effect size	Estimation variance
1	θ_1	T_1	v_1
⋮	⋮	⋮	⋮
k	θ_k	T_k	v_k

The vectors of the true population and estimated effect sizes are $\theta = (\theta_1, \theta_2, \ldots \theta_k)'$ and $T = (T_1, T_2, \ldots, T_k)'$ respectively, and the linkage between them is given by the relation,

$$T_i = \theta_i + \varepsilon_i \qquad \text{for the } i^{th} \text{ study}, \quad i = 1,\ldots,k, \tag{2-1}$$

where ε_i is the random or sampling error.

In the fixed effects model we assume that the true population effect size θ_i for the *ith* study depends on a vector of p fixed independent study characteristics $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})'$.

More specifically, we have,

$$\begin{aligned}
\theta_1 &= \beta_0 + \beta_1 x_{11} + \cdots + \beta_p x_{1p} \\
\theta_2 &= \beta_0 + \beta_1 x_{21} + \cdots + \beta_p x_{2p} \\
&\vdots \qquad\qquad\qquad\qquad\qquad\quad \vdots \\
\theta_k &= \beta_0 + \beta_1 x_{k1} + \cdots + \beta_p x_{kp}
\end{aligned},$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are unknown regression coefficients, and $x_{i1}, x_{i2}, \ldots, x_{ip}$ are the values of the independent or predictor variables $X_1, X_2, \ldots, X_p$ for the *ith* study.

Thus, in matrix notation, the model can be written as:

$$\underset{k\times 1}{\mathrm{T}} = \underset{k\times P}{\mathrm{X}}\underset{P\times 1}{\beta} + \underset{k\times 1}{\varepsilon}, \tag{2-2}$$

where $P = p + 1$, and

$$\mathrm{T} = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_k \end{bmatrix}, \qquad \mathrm{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \vdots \\ 1 & x_{k1} & \cdots & x_{kp} \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \text{ and, } \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}.$$

2.3 ESTIMATION

Estimating the parameters of the fixed effects model for effect sizes is a little more complicated than estimation in a standard regression model. To help understand why, let us consider the *ith* observation of equation (2-2), yielding the linear fixed effects model for the effect size estimate T_i:

$$T_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i. \qquad (2\text{-}3)$$

The variance of ε_i is

$$Var(\varepsilon_i) = Var(T_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_p X_{ip}) = Var(T_i) = v_i,$$

Furthermore, the covariance between ε_i and $\varepsilon_{i'}$, for $i \neq i'$, i and $i' = 1,\ldots,k$, is,

$$Cov(\varepsilon_i, \ \varepsilon_{i'}) = Cov(T_i, \ T_{i'}) = 0,$$

because it is natural to assume that the results of the k studies in the research synthesis are independent.

To estimate the parameters $(\beta_0, \beta_1, \ldots, \beta_p)$ using the ordinary least square method (OLS) would be clearly inappropriate, since this later assumes that every residual ε_i has the same variance. In contrast, the residual variances $v_i's$ of the model are in general unequal, i.e. heteroscedastic, since data from research synthesis will nearly always be

unbalanced giving unequal precision in the estimation of the study's effect size estimates. Therefore, a weighted least square (WLS) approach is preferred and the weights used for this purpose are the inverse of each study estimation variance, namely,

$$w_i = 1/v_i, \qquad i = 1,\ldots,k, \tag{2-4}$$

and so the calculation of w_i is straightforward. Table I in the previous section provides expressions for v_i in the case of several measures of effect size T_i.

Hence, in order to estimate the β's in equation (2-3), we use a weighted least square estimation with the variance-covariance matrix of the residuals equals to,

$$D[\varepsilon] = D[T] = diag(v_1, v_2, \ldots v_p) = diag(w_1^{-1}, w_2^{-1}, \ldots, w_p^{-1}) = \mathrm{V} = \mathrm{W}^{-1}.$$

The weighed least square estimate (WLSE) of β is well known (Seber 1977, p. 61) and is given by,

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}T = (X'WX)^{-1}X'WT. \tag{2-5}$$

The estimate $\hat{\beta}$ is unbiased for β and its variance-covariance matrix is given by,

$$
\begin{aligned}
\mathrm{D}[\hat{\beta}] &= \mathrm{D}[(X'V^{-1}X)^{-1}X'V^{-1}T] \\
&= (X'V^{-1}X)^{-1}X'V^{-1}\ \mathrm{D}[\mathrm{T}]\ ((X'V^{-1}X)^{-1}X'V^{-1})' \\
&= (X'V^{-1}X)^{-1}X'V^{-1}(V)V^{-1}X(X'V^{-1}X)^{-1} \\
&= (X'V^{-1}X)^{-1} \\
&= (X'WX)^{-1}.
\end{aligned}
$$

2.4 INFERENCE ON INDIVIDUAL PARAMETERS

The independent variables may or may not contribute in explaining some of the variability of the effect sizes among the studies included in the meta-analysis. Thus, it is of interest to make inference on the regression coefficients $\beta_0, \beta_1, \ldots, \beta_p$ that are capturing the possible associations between the study characteristics and the effect sizes.

The usual null hypothesis considered is,

$$H_0: \beta_j = 0, \qquad j = 0,1,2,\ldots,p.$$

This hypothesis may be tested by computing the ratio of the estimate to its standard error, that is,

$$t_j = \frac{\hat{\beta}_j}{S(\hat{\beta}_j)}, \qquad j = 0,1,2,\ldots,p. \tag{2-6}$$

where $S(\hat{\beta}_j)$ is the j^{th} square root element on the diagonal of the matrix $(X'WX)^{-1}$. The t ratio is approximately normally distributed, with the approximation improving as the number of studies increases. However, it is preferable to compare the t Statistic with the critical values of Student t distribution with $k-p-1$ degrees of freedom when the number of studies is small, which is often the case in systematic reviews.

All computations can be done using many standard statistical packages, with minor adjustements or hand calculations. In general, these statistical packages give the correct estimates of the regression coefficients, but the standard errors and significance levels are incorrect for the fixed effects meta-analysis model. However, the inverse of the weigthed sum of squares, $(X'WX)^{-1}$, can usually easily be printed out as an option. Alternatively, the correct standard error $S(\hat{\beta}_j)$ of the estimated coefficient estimate β_j is simply,

$$S(\hat{\beta}_j) = \frac{SE_j}{\sqrt{MS_{error}}},$$

where SE_j is the standard error of $\hat{\beta}_j$ as given by the statistical program and MS_{error} is the « error » or « residual » mean square from the analysis of variance table as given by the computer program as well.

A $100(1-\alpha)\%$ confidence interval for each β_j can be obtained by multiplying $S(\hat{\beta}_j)$ by the two-tailed critical value of the Student t distribution with $k-p-1$ degrees of freedom, and then adding and substracting this product from $\hat{\beta}_j$. More specifically, the 100(1-α)% confidence interval for β_j is given by:

$$\hat{\beta}_j - t^{\alpha/2}_{k-p-1} S(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + t^{\alpha/2}_{k-p-1} S(\hat{\beta}_j),$$

where $t^{\alpha/2}_{k-p-1}$ is the critical value such that the probability of the Student t distribution with $k-p-1$ degrees of freedom exceeding $t^{\alpha/2}_{k-p-1}$ is equal to $\alpha/2$.

One may want to look at a simultanous estimation of the β_j's to ensure that the overall type I error does not exceed α. For example, we can use the Bonferroni method leading to the following confidence interval for β_j :

$$\hat{\beta}_j - t(1-\alpha/2p; k-p-1) S(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + t(1-\alpha/2p; k-p-1) S(\hat{\beta}_j),$$

where p is the number of confidence intervals, corresponding to the number of study characteristics in the model.

CHAPTER 3

MIXED EFFECTS MODEL

As we have seen in Chapter 2, under the fixed effects model the independent or predictor variables $X_1,\ldots,X_p$ are assumed to account completely for the variation in the true effect sizes across studies. In contrast, the mixed effects model, which is the main focus of this Chapter, assumes that part of the variability in the true effect sizes remains unexplained by the model. Hence, the name ***mixed effects linear model*** was given to this model since it contains a combination of: 1) fixed effect parameters and 2) a random effect parameter.

In the mixed effects linear model, the effect size estimate T_i in any given study i, $i = 1,\ldots,k$, differs from the true effect size θ_i due to both, sampling error and prediction error. In the fixed effects linear model, the estimates of the effect size θ_i vary as a result of chance differences between study's samples. This variation is conventionally called "sample variance" or "estimation variance" and was denoted by v_i in Chapter 2. On the other hand, due to numerous unidentifiable and/or uncontrollable sources of influence in the true effect size, it is in general difficult to make exact prediction. Hence, the true effect size may itself vary. Therefore, in addition to the "estimation variance" arising from random sampling, we refer to the variance of the true effect size as a "random

effects variance", denoted by σ_θ^2. Thus, the variance of the estimated effect sizes has two independent components:

Variance of estimated effects	=	random effects variance	+	estimation variance

.

The variance described above can be written as follows:

$$Var(T_i) = v_i^* = \sigma_\theta^2 + v_i, \quad i = 1,\ldots,k.$$

Usually, v_i is available from the study, but σ_θ^2 is unknown and hence, need to be estimated from the data. One should note that, if the model characteristics explain all the variation in the true effect size, σ_θ^2 would be zero; i.e. all of the variation across studies, once those study characteristics are taken into account in the model, would be attributed to the estimation variance v_i. This latter situation brings us back to the fixed effects model discussed in Chapter 2. Therefore, the mixed effects model encompasses the fixed effects model as a special case.

3.1 MODEL AND NOTATION

Suppose that we have calculated an effect size estimate T_i of the true effect size θ_i for each of k studies, $i = 1,\ldots,k$. From equation (2-1), we recall the linkage between the true and the estimated effects size,

$$T_i = \theta_i + \varepsilon_i, \qquad i = 1,\ldots,k, \tag{3-1}$$

where the $\varepsilon_i's$ are independent random errors with mean zero and estimation variance v_i. Under the mixed effects linear model, the prediction for the true effect sizes depends on a set of independent variables or study characteristics plus an error term, that is,

$$\theta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + u_i, \qquad i = 1,\ldots,k, \tag{3-2}$$

where

β_0 is the model intercept;

$X_{i1},\ldots,X_{ip}$ are the independent variables hypothesized to predict the study effect size θ_i;

$\beta_1, \beta_2, \ldots, \beta_p$ are regression coefficients capturing the association between study characteristics and effect sizes; and

u_i is the random effect of study i, that is, the deviation of study $i's$ true effect size from the value predicted on the basis of the model. The u_i's are assumed to be independent with mean zero and variance σ_θ^2.

The model (3-1) is identical to the fixed effects model (2-2) with one exception: the addition of the random effect u_i.

3.2 ESTIMATION

Estimating the parameters in the mixed effects model is more complicated than estimation in standard regression model, and even more complicated than in the fixed effects model for a meta-analysis, as described in Chapter 2. To understand why, it is better to substitute equation (3-2) into (3-1) leading to the regression model,

$$T_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + u_i + \varepsilon_i . \qquad (3\text{-}3)$$

From equation (3-3), we can clearly note that there are two random components, u_i and ε_i, and hence, the variance of T_i becomes,

$$Var(T_i) = Var(u_i + \varepsilon_i) = Var(u_i) + 2Cov(u_i, \varepsilon_i) + Var(\varepsilon_i).$$

As mentioned above, it is usually natural to assume that u_i and ε_i are independent, so that $Cov(u_i, \varepsilon_i) = 0$ and,

$$Var(T_i) = Var(u_i) + Var(\varepsilon_i) = \sigma_\theta^2 + v_i = v_i^*.$$

Similarly to the fixed effects model, it would be inappropriate to estimate the parameters of (3-3) using the ordinary least squares estimation, because OLS assumes that every residual has equal variance (homoscedastic). On the contrary, the residual variances in model (3-3) are in general unequal (heteroscedastic) because, as seen in Chapter 2, the v_i's varies across studies, and hence the v_i^*'s as well. Therefore, the weighted least squares approach is needed, and the weights denoted, by $w_1^*, w_2^*, \ldots, w_k^*$, are the inverse of each study's variance,

$$w_i^* = \frac{1}{v_i^*} = \frac{1}{(\sigma_\theta^2 + v_i)}, \qquad i = 1, \ldots, k.$$

The weights w_i^*'s differ from the weights in the fixed effects model mentioned in equation (2-4), since they depend on the additional random effects variance σ_θ^2 which is generally unknown and must be estimated from the data. However, to estimate σ_θ^2 and hence w_i^*, an estimate of the regression coefficients β's is required.

The dilemma posed here is that the estimation of the $\beta's$ depends on the unknown value of σ_θ^2 that is part of the weights, and the estimation of σ_θ^2 depends on the unknown values of the $\beta's$. Two methods are proposed to solve this problem: a three step ***method of moments*** and ***the method of maximum likelihood.***

3.2.1 A THREE STEPS METHOD OF MOMENTS

Using the method of moments, provisional estimates of the $\beta's$ in equation (3-3) can be computed. Based on these provisional estimates, the random variance σ_θ^2 and the weights w_i^* are then estimated. Finally, these weights are employed in a weighted least squares regression to obtain new and final estimates of the $\beta's$.

There are two approaches to the three steps method of moments. One approach starts by computing ordinary least squares estimates of the fixed effects model. The second approach begins by computing the weighted least squares estimates of the same fixed effects model using the weights $w_i = 1/v_i$. The specific details of the three steps are described in the following sections.

3.2.1.1. Step1:Computing the provisional estimates of the β's

Approach I: Starting with OLS

This approach computes provisional least squares estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ using ordinary least squares. For this purpose, let us consider model (3-3) in matrix notation:

$$\underset{k\times 1}{T} = \underset{k\times P}{X}\underset{P\times 1}{\beta} + \underset{k\times 1}{u} + \underset{k\times 1}{\varepsilon}, \tag{3-4}$$

where $P = p + 1$, and

$$T = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_k \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \vdots \\ 1 & x_{k1} & \cdots & x_{kp} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix}.$$

Also,

$$Var(T) = Var(u+\varepsilon) = \sigma_\theta^2 I_k + V,$$

where I_k is the identity matrix of dimension k and V is the $k \times k$ diagnonal estimation variance matrix with the $v_i's$ on the diagnonal, $i = 1, \ldots, k$.

The ordinary least square estimate of β is well known (Seber, 1977) and is given by,

$$\hat{\beta}_{ols} = (X'X)^{-1}X'T.$$

The residual sum of squares (RSS) is well known as well and is given by,

$$RSS = (T - X\hat{\beta}_{ols})'(T - X\hat{\beta}_{ols})$$

$$= T'[I_k - X(X'X)^{-1}X']T.$$

Approach II: Starting with WLS

Recall from Chapter 2 that the weighted least square estimate of β from the fixed effects model is given by the equation (2-5)

$$\hat{\beta}_{wls} = (X'V^{-1}X)^{-1}X'V^{-1}T,$$

and the residual sum of squares is given (Seber, 1977) by,

$$RSS = (T - X\hat{\beta}_{wls})'V^{-1}(T - X\hat{\beta}_{wls})$$

$$= T'[V^{-1}(I_k - X(X'V^{-1}X)^{-1}X'V^{-1})]T. \qquad (3\text{-}5)$$

3.2.1.2 Step2: Computing the estimate of σ_θ^2

To find an estimate of σ_θ^2, we need in the second step to compute the expected value of the residual sum of squares obtained in step 1. For that purpose, the following theorem on quadratic forms will be useful (Seber, 1977, p. 13, theorem 1.7):

> **Theorem 1**: Let Y be an $n \times 1$ vector of random variables and let A be an $n \times n$ symmetric matrix. If E [Y] = θ and D[Y] = Σ then,
>
> E [Y'AY] = tr[A Σ] + θ'A θ,
>
> where tr[A Σ] is the trace of the matrix A Σ.

Approach I: Starting with OLS

The expected value of the residual sum of squares based on the provisional OLS estimates of β is:

$$
\begin{aligned}
& E(RSS) \\
&= E[T'(I_k - X(X'X)^{-1}X')T] \\
&= E[T'MT] && \text{say, where } M = (I_k - X(X'X)^{-1}X') \\
&= tr[MVar(T)] + \beta'X'MX\beta && \text{by Theorem 1} \\
&= tr[M(\sigma_\theta^2 I_k + V)] + \beta' X'MX\beta
\end{aligned}
$$

$$= tr[M\sigma_\theta^2] + tr[MV] + \beta' X' MX\beta$$

$$= \sigma_\theta^2 tr(M) + tr[MV] + \beta' X' MX\beta$$

$$= \sigma_\theta^2[tr(I_k) - tr(X(X'X)^{-1}X') + tr[(I_k - (X(X'X)^{-1}X'))V] +$$

$$\beta' X' X\beta - \beta' X'(X(X'X)^{-1}X')X\beta$$

$$= \sigma_\theta^2[k - tr[(X'X)^{-1}(X'X)] + tr[(I_k - (X(X'X)^{-1}X'))V] + \beta' X' X\beta - \beta' X' X\beta$$

$$= \sigma_\theta^2[k - tr[I_{(p+1)}] + tr[(I_k - (X(X'X)^{-1}X'))V]$$

$$= \sigma_\theta^2(k - p - 1) + tr[(I_k - (X(X'X)^{-1}X'))V].$$

Therefore, an estimate of σ_θ^2 would be,

$$\hat{\sigma}^2_{\theta(ols)} = \frac{RSS - tr[(I_k - (X(X'X)^{-1}X'))V]}{(k - p - 1)}.$$

Approach II: Starting with WLS

The expected value of the residual sum of squares based on the preliminary WLS estimates of β is:

$E(RSS)$

$= E(T'(V^{-1}[I_k - X(X'V^{-1}X)^{-1}X'V^{-1}])T)$

$= E(T'AT)$ say , where $A = (V^{-1}[I_k - X(X'V^{-1}X)^{-1}X'V^{-1}])$

$= tr[AVar(T)] + \beta X'AX\beta$ by Theorem 1

$= tr[A(\sigma_\theta^2 I_k + V)] + \beta' X'AX\beta$

$= tr[\sigma_\theta^2 A] + tr[AV] + \beta' X'AX\beta$

$= \sigma_\theta^2[tr(V^{-1}(I_k - X(X'V^{-1}X)^{-1}X'V^{-1})] + tr[I_k - V^{-1}X(X'V^{-1}X)^{-1}X'] +$
$\beta' X'V^{-1}X\beta - \beta' X'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}X\beta$

$= \sigma_\theta^2[tr(V^{-1}) - tr(V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1})] + tr[I_k] - tr[X'V^{-1}X(X'V^{-1}X)^{-1}] +$
$\beta' X'V^{-1}X\beta - \beta' X'V^{-1}X\beta$

$= \sigma_\theta^2[tr(V^{-1}) - tr(X'V^{-2}X(X'V^{-1}X)^{-1})] + tr[I_k] - tr[I_{(p+1)}]$

$= \sigma_\theta^2[tr(V^{-1}) - tr(X'V^{-2}X(X'V^{-1}X)^{-1})] + (k - p - 1).$ (3-6)

Therefore, an estimate of σ_θ^2 is given by,

$$\hat{\sigma}^2_{\theta(wls)} = \frac{RSS - (k - p - 1)}{tr(V^{-1}) - tr[X'V^{-2}X(X'V^{-1}X)^{-1}]}. \quad (3\text{-}7)$$

3.2.1.3 Step3: Computing new estimates of the β's and inference

Once the random effect variance, $\hat{\sigma}_\theta^2$, is estimated from the data, new estimate of the mixed effects model can now be computed using the weighted least square method with weights given by,

$$w_i^* = \frac{1}{(v_i + \hat{\sigma}_\theta^2)}.$$

The estimated weighted regression coefficients, $\hat{\beta}_{wls}^*$, can easily be computed using equation (2-4) for the fixed effects model, replacing $w_i = \frac{1}{v_i}$ by $w_i^* = \frac{1}{(v_i + \hat{\sigma}_\theta^2)}$. Thus, we obtain,

$$\hat{\beta}_{wls}^* = (X'W^*X)^{-1}X'W^*T, \qquad (3\text{-}8)$$

where $W^* = (\hat{\sigma}_\theta^2 I_k + V)^{-1}$.

Considering $\hat{\sigma}_\theta^2$ as fixed in w^*, $\hat{\beta}_{wls}^*$ is an unbiased estimate of β and its variance is:

$$
\begin{aligned}
Var(\hat{\beta}^{*}_{wls}) &= Var[(X'W^{*}X)^{-1}X'W^{*}T] \\
&= [(X'W^{*}X)^{-1}X'W^{*}]Var(T)[(X'W^{*}X)^{-1}X'W^{*}]' \\
&= [(X'W^{*}X)^{-1}X'W^{*}](W^{*})^{-1}[(X'W^{*}X)^{-1}X'W^{*}]' \\
&= [(X'W^{*}X)^{-1}X'W^{*}](W^{*})^{-1}[(W^{*}X(X'W^{*}X)^{-1}] \\
&= (X'W^{*}X)^{-1}X'W^{*}X(X'W^{*}X)^{-1} \\
&= (X'W^{*}X)^{-1} \\
&= [X'(\hat{\sigma}^{2}_{\theta}I_{k} + V)^{-1}X]^{-1}. \qquad (3\text{-}9)
\end{aligned}
$$

Once the $\hat{\beta}^{*}_{wls}$ parameters have been calculated, it becomes straightforward to make inference about them in terms of hypethesis testing and confidence interval. The usual null hypothesis for the fixed effect coefficients β_j is,

$$H_0: \beta_j = 0, \qquad j = 0,1,\ldots,p.$$

This hypothesis is to be tested by computing the ratio of the estimate to its standard error, that is,

$$t = \frac{\hat{\beta}^{*}_{j(wls)}}{S(\hat{\beta}^{*}_{j(wls)})}, \qquad (3\text{-}10)$$

where $S(\hat{\beta}^{*}_{j(wls)})$ is the square root of the j^{th} diagonal element of the estimated variance matrix given in (3-9).

Under the null hypothesis, the t ratio in (3-10) will follow approximately a Student t distribution with $k-p-1$ degrees of freedom. Also, a 100(1-α)% confidence interval for β_j can be written as follows,

$$\hat{\beta}^{*}_{j(wls)} - t^{\alpha/2}_{k-p-1} S(\hat{\beta}^{*}_{j(wls)}) \le \beta_j \le \hat{\beta}^{*}_{j(wls)} + t^{\alpha/2}_{k-p-1} S(\hat{\beta}^{*}_{j(wls)}),$$

where $t^{\alpha/2}_{k-p-1}$ is the critical value such that the probability of the Student t distribution with $k-p-1$ degrees of freedome exceeding $t^{\alpha/2}_{k-p-1}$ is equal to $\alpha/2$.

Furthermore, it is of interest to test the hypothesis that the random effects variance σ^2_θ is null, that is,

$$H_0: \quad \sigma^2_\theta = 0.$$

If this hypothesis is retained, we can conclude that the study characteristics in model (3-3), namely the X's, fully account for the variation in the true effect sizes. In contrast, if this hypothesis is rejected, significant variation among the random effects (the values of u_i) remain unexplained after controlling for these study characteristics.

To derive a test for this hypothesis, the last part of the following theorem on distribution theory is needed (adapted from Seber, 1977, p. 60-64 and p.54, theorem3.5):

Theorem 2: If $Y \sim N_k(X\beta, \Sigma)$, where X is $k \times P$ of rank P and Σ is nonsingular, then:

(i) $\hat{\beta}_{wls} \sim N_p(\beta, (X'\Sigma^{-1}X)^{-1})$, where $\hat{\beta}_{wls} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$

(ii) $(\hat{\beta}_{wls} - \beta)'X'\Sigma^{-1}X(\hat{\beta}_{wls} - \beta) \sim \chi^2_P$.

(iii) $\hat{\beta}_{wls}$ is independent of RSS, where $RSS = (Y - X\hat{\beta}_{wls})'\Sigma^{-1}(Y - X\hat{\beta}_{wls})$

(iv) $RSS \sim \chi^2_{k-P}$

Now, under the fixed effects model, that is under the null hypothesis of $\sigma_\theta^2 = 0$, and assuming that the vector of estimated effect sizes $T \sim N_k(X\beta, V)$, the weighted residual sum of squares given by equation (3-5) will follow a Chi-square distribution with $k - p - 1$ degrees of freedom, by Theorem 2 (iv).

Therefore, we will reject the null hypothesis, at the significance level α, if the RSS exceeds the $100(1-\alpha)$ percent point of the Chi-square distribution with $k - p - 1$ degrees of freedom. In other words, this test can be viewed as a test for greater than expected residual variation, as we can see by comparing the expected value of RSS in (3-

6) under the mixed effects model and $k-p-1$ which corresponds to the expected value of RSS when $\sigma_\theta^2 = 0$; i.e. under the fixed effects model.

3.2.2 METHOD OF MAXIMUM LIKELIHOOD

In this method, the maximum likelihood estimates of the fixed effect parameters $(\beta_0, \beta_1, \ldots, \beta_p)$ and of the random effect parameter σ_θ^2 are to be computed using a simple iterative procedure untill the estimates converge. One advantage of this method, is that the procedure naturally produces an estimate of the standard error of $\hat{\sigma}_\theta^2$ at convergence, not available in the method of moments.

If we assume that T is normally distributed with mean $X\beta$ and variance $V^* = \sigma_\theta^2 I_k + V$, then the likelihood function of the fixed effect estimates $(\beta_0, \beta_1, \ldots, \beta_p)$ and random effect estimates σ_θ^2 is,

$$L = L(\beta, \sigma_\theta^2) = (2\pi)^{-\frac{1}{2}k} |V^*|^{-\frac{1}{2}} \exp[-\frac{1}{2}(T - X\beta)'(V^*)^{-1}(T - X\beta)]$$

$$= (2\pi)^{-\frac{1}{2}k} |(\sigma_\theta^2 I_k + V)|^{-\frac{1}{2}} \exp[-\frac{1}{2}(T - X\beta)'(\sigma_\theta^2 I_k + V)^{-1}(T - X\beta)].$$

The natural logarithm of the likelihood function is then given by,

$$Ln(L) = -\frac{1}{2}k[Ln(2\pi)] - \frac{1}{2}[Ln|\ (\sigma_\theta^2 I_k + V)\ |] - \frac{1}{2}[(T - X\beta)'(\sigma_\theta^2 I_k + V)^{-1}(T - X\beta)]. \tag{3-11}$$

The maximum likelihood estimators of β and σ_θ^2 are the vector $\hat{\beta}_{ML}$ and the scalar $\hat{\sigma}_{\theta_{ML}}^2$ that maximize $Ln(L)$. For this reason, we need to solve for β in $\frac{\partial Ln(L)}{\partial \beta} = 0$ and for σ_θ^2 in $\frac{\partial Ln(L)}{\partial \sigma_\theta^2} = 0$.

First we differentiate (3-11) with respect to β, that is,

$$\begin{aligned}\frac{\partial Ln(L)}{\partial \beta} &= \frac{\partial[-\frac{1}{2}(T - X\beta)'(V^*)^{-1}(T - X\beta)]}{\partial \beta} \\ &= -\frac{1}{2}\frac{\partial[(T'(V^*)^{-1}T) - 2\beta' X'(V^*)^{-1}T + \beta' X'(V^*)^{-1}X\beta]}{\partial \beta} \\ &= X'(V^*)^{-1}T - X'(V^*)^{-1}X\beta. \end{aligned} \tag{3-12}$$

Setting (3-12) to zero and solving for β, we get,

$$\begin{aligned}\hat{\beta}_{ML} &= (X'(V^*)^{-1}X)^{-1}X'(V^*)^{-1}T \\ &= (X'(\sigma_\theta^2 I_k + V)^{-1}X)^{-1}X'(\sigma_\theta^2 I_k + V)^{-1}T.\end{aligned}$$

Hence, the maximum likelihood estimate $\hat{\beta}_{ML}$ is identical to the weighted least squares estimate and is a function of the unknown random variance σ_θ^2.

Now, to obtain $\hat{\sigma}^2_{\theta_{ML}}$, we differentiate (3-11) with respect to σ_θ^2:

$$\frac{\partial Ln(L)}{\partial \sigma_\theta^2} = \frac{\partial}{\partial \sigma_\theta^2}\left(-\frac{1}{2}Ln|V^*| - \frac{1}{2}(T-X\beta)'(V^{*-1})(T-X\beta)\right)$$

$$= -\frac{1}{2}\frac{\partial Ln|V^*|}{\partial \sigma_\theta^2} - \frac{1}{2}\frac{\partial[(T-X\beta)'(V^{*-1})(T-X\beta)]}{\partial \sigma_\theta^2}$$

$$= A + B \qquad \text{say,}$$

where,

$$A = -\frac{1}{2}\frac{\partial Ln|V^*|}{\partial \sigma_\theta^2} = -\frac{1}{2}\frac{\partial}{\partial \sigma_\theta^2} Ln|\sigma_\theta^2 I_k + V| = -\frac{1}{2}\frac{\partial}{\partial \sigma_\theta^2} Ln\left[\prod_{i=1}^{k}(\sigma_\theta^2 + v_i)\right]$$

$$= -\frac{1}{2}\frac{\partial}{\partial \sigma_\theta^2}\left[\sum_{i=1}^{k} Ln(\sigma_\theta^2 + v_i)\right]$$

$$= -\frac{1}{2}\sum_{i=1}^{k}\frac{1}{\sigma_\theta^2 + v_i}$$

$$= -\frac{1}{2}\sum_{i=1}^{k} w_i^*$$

$$= -\frac{1}{2}\sum_{i=1}^{k} w_i^*\left(\frac{v_i^*}{v_i^*}\right)$$

$$= -\frac{1}{2}\sum_{i=1}^{k} w_i^{*2} v_i^{*}, \qquad (3\text{-}13)$$

and,

$$B = -\frac{1}{2}\frac{\partial}{\partial \sigma_\theta^2}\left[(T - X\beta)'(V^{*-1})(T - X\beta)\right]$$

$$= -\frac{1}{2}\frac{\partial}{\partial \sigma_\theta^2}\left(r' V^{*-1} r\right), \qquad \text{where } r = (T - X\beta)$$

$$= -\frac{1}{2}\frac{\partial}{\partial \sigma_\theta^2}\sum_{i=1}^{k}\frac{r_i^2}{\sigma_\theta^2 + v_i}, \qquad \text{where } r_i \text{ is the } i^{th} \text{ element of } r,$$

$$= \frac{1}{2}\sum_{i=1}^{k} w_i^{*2} r_i^2 . \qquad (3\text{-}14)$$

Hence, from (3-13) and (3-14) we obtain:

$$\frac{\partial Ln(L)}{\partial \sigma_\theta^2} = -\frac{1}{2}\sum_{i=1}^{k} w_i^{*2} v_i^{*} + \frac{1}{2}\sum_{i=1}^{k} w_i^{*2} r_i^2 . \qquad (3\text{-}15)$$

Setting (3-15) to zero, we get,

$$-\sum_{i=1}^{k} w_i^{*2} v_i^{*} + \sum_{i=1}^{k} w_i^{*2} r_i^2 = 0 . \qquad (3\text{-}16)$$

Equation (3-16) is an implicit function, and to solve for σ_θ^2 we use an iteration procedure, that is to pull out a new σ_θ^2 as a function of an initial one such that ,

$$\sum_{i=1}^{k} w_i^{*2}(\sigma_\theta^2 + v_i) - \sum_{i=1}^{k} w_i^{*2} r_i^2 = 0$$

$$\sigma_\theta^2 \sum_{i=1}^{k} w_i^{*2} + \sum_{i=1}^{k} w_i^{*2} v_i - \sum_{i=1}^{k} w_i^{*2} r_i^2 = 0,$$

and solving for σ_θ^2, we get,

$$\hat{\sigma}_{\theta(new)}^2 = \frac{\sum_{i=1}^{k} w_i^{*2}(r_i^2 - v_i)}{\sum_{i=1}^{k} w_i^{*2}}.$$

The iteration procedure starts with an inital estimate of σ_θ^2 in $w_i^* = \left(\sigma_\theta^2 + v_i\right)^{-1}$ given by (3-7) and an initial estimate of r_i from (3-8). Each estimate of σ_θ^2 should yield a positive value, and the process iterates until the estimates converges. Every negative estimate of σ_θ^2 should be converted to zero. At convergence, the process produces the maximum likelihood estimate $\hat{\sigma}_{\theta_{ML}}^2$.

Since $\hat{\beta}_{ML}$ is identical to the weighted least squares estimate $\hat{\beta}_{wls}^*$ in (3-8), then its variance is given by equation (3-9).

From the properties of the maximum likelihood estimation, the variance of $\sigma^2_{\theta_{ML}}$ is given by the i^{th} element of the inverse of the information matrix, that is

$$Var(\hat{\sigma}^2_{\theta_{ML}}) = -\frac{1}{E\left(\frac{\partial^2 LnL}{\partial \sigma_\theta^4}\right)},$$

where,

$$E\left(\frac{\partial^2 LnL}{\partial \sigma_\theta^4}\right) = E\left(\frac{\partial^2}{\partial \sigma_\theta^4}\left[-\frac{1}{2}Ln\left|\sigma_\theta^2 I_K + V\right| - \frac{1}{2}(T - X\beta)'(\sigma_\theta^2 I_k + V)^{-1}(T - X\beta)\right]\right)$$

$$= -\frac{1}{2}E\left(\frac{\partial^2}{\partial \sigma_\theta^4}Ln\left|\sigma_\theta^2 I_k + V\right|\right) - \frac{1}{2}E\left(\frac{\partial^2}{\partial \sigma_\theta^4}(T - X\beta)'(\sigma_\theta^2 I_k + V)^{-1}(T - X\beta)\right)$$

$$= -\frac{1}{2}E\left(\frac{\partial}{\partial \sigma_\theta^2}\sum_{i=1}^{k}\frac{1}{(\sigma_\theta^2 + v_i)}\right) + \frac{1}{2}E\left(\frac{\partial}{\partial \sigma_\theta^2}\sum_{i=1}^{k}\frac{r_i^2}{(\sigma_\theta^2 + v_i)^2}\right) \qquad \text{by (3-13) \& (3-14)}$$

$$= \frac{1}{2}E\left(\sum_{i=1}^{k}\frac{1}{(\sigma_\theta^2 + v_i)^2}\right) - E\left(\sum_{i=1}^{k}\frac{r_i^2}{(\sigma_\theta^2 + v_i)^3}\right)$$

$$= \frac{1}{2}tr(V^{*-2}) - E\left[(T - X\beta)'V^{*-3}(T - X\beta)\right]$$

$$= \frac{1}{2}tr(V^{*-2}) - tr(V^{*-2}) \qquad \text{by Theorem 1}$$

$$= -\frac{1}{2}tr(V^{*-2})$$

$$= -\frac{1}{2}\sum_{i=1}^{k}\frac{1}{(\sigma_\theta^2 + v_i)^2}$$

$$= -\frac{1}{2}\sum_{i=1}^{k} w_i^{*2} .$$

Therefore,

$$Var(\hat{\sigma}^2_{\theta_{ML}}) = \frac{2}{\sum_{i=1}^{k} w_i^{*2}} .$$

In this section, we have used a simple iterative likelihood procedure to find the maximum likelihood estimates, also known as the full information likelihood estimation, (Longford, 1987). On the other hand, the restricted approach of the maximum likelihood estimation (Bryk et al., HLM program, 1988) for univariate or multivariate models is a bit more satisfactory, specially when the number of studies are small, because it adjusts variance estimates for the uncertainty associated with estimation of the fixed effects. Due to the lack of time, this later approach had not been tried and therefore, we do not have enough evidence to support our argument stated above.

CHAPTER 4

EXAMPLE

In this chapter, we illustrate the steps to a systematic review using the data of a recent meta-analysis on the prognosis of depression in the elderly, accepted for publication in the American Journal of Psychiatry (July 1998). We have applied the theory of both, the fixed and mixed effects model to the outcome, namely, the percent of subjects well, i.e. the percent of subjects who recovered at the end of the study period. Under the mixed effects model both, the three steps method of moment and the method of maximum likelihood are presented.

The data synthesis had been preceded by a literature review including the computerized literature search, interrater reliability of the data abstraction and assessment of the validity and quality of the research studies.

4.1.LITERATURE REVIEW

4.1.1 LITERATURE SEARCH

The selection process involved four steps. First, two computer databases, **MEDLINE** and **PsycINFO**, were searched by Dr. Martin Cole (MC) for potentially relevant articles

published from January 1980 to November 1996, and from January 1984 to November 1996, respectively, using the following search strategy: "depression" **AND** ["prognosis" **OR** "course" **OR** "follow-up"] **AND** "aged". Second, relevant articles were retrieved for more detailed evaluation. The selection process yielded 711 potentially relevant studies; most of which were not studies of prognosis. Thirdly, the bibliographies of relevant articles were searched for additional references. Twenty-seven articles were retrieved for more detailed evaluation. They were screened by MC to see if they met the following five inclusion criteria: 1) original research; 2) published in English or French; 3) study population of community residents or primary care patients; 4) subjects' mean age 60 years or over; 5) reported affective state as an outcome.

Four studies of primary care patients, involving 843 patients with depression (Callahan et al, 1994; Kennedy et al, 1991; Kukull et al, 1986; Van Marwijk, 1997), and 8 studies of community residents, involving 425 subjects with depression (Ben-Arie et al,1990; Bowling et al, 1996; Copeland et al, 1992; Forsell et al, 1994; Kivela et al, 1991; Kivela, 1995; Kua, 1993; O'Connor et al, 1990; Snowdon and Lane, 1995), met all the inclusion criteria. The other 15 studies were excluded for the following reasons: one was not original research, two had subjects' mean age lower than 60 years, nine did not report affective state as an outcome and three did not meet two or more of these inclusion criteria.

4.1.2 ASSESSMENT OF VALIDITY OR QUALITY

To determine validity, Dr. Martin Cole (MC) and and I (Asmaâ Mansour, AM), independently assessed the methods and design of each study, according to the seven criteria for prognostic studies described by the Evidence-Based Medicine Working Group (Laupasis, 1994), namely, 1) formation of an inception cohort; 2) description of referral pattern; 3) adequate length of follow-up to determine outcome; 4) completion of follow-up (i. e. determination of outcomes for at least 80% of the inception cohort); 5) objective outcome criteria; 6) unbiased outcome assessment and 7) adjustment for extraneous prognostic factors (i. e., severity of physical illness, cognitive impairment). Each study was scored with respect to meeting (+), not meeting (-) or partially meeting (+/-) each of the above criteria. Interrater agreement was calculated for each criteria as the % of studies where independent assessments of both raters were exactly the same; Interrater agreement ranged between 50% to 100% for the seven criteria. After discussion between the two raters, a consensus was reached in where disagreement was initially observed.

All studies had some methodological limitations. For the primary care studies, the limitations were related to description of referral pattern, completion of follow-up or adjustment for extraneous prognostic factors; for the community studies, the limitations were related to unbiased outcome assessment or adjustment for extraneous prognostic factors.

4.1.3 DATA ABSTRACTION

Information about the population, sample size, lower age limit at enrollment, number of subjects belonging to each gender category, diagnostic criteria, proportion of cases detected and treated by primary care physicians, length of follow-up, affective outcomes and prognostic factors was independently abstracted by two raters, MC and AM, from each relevant paper. Similarly to the validity criteria, inter-observer agreement, was calculated (see bottom of table II). The percent agreement ranged between 58% and 100%. The lower level of agreement for the outcomes variables (58%) was due mainly to the way it had been calculated: for each study, both raters had to have exactly the same percentages in each category. If the criterion of agreement was relaxed to same $\pm 3\%$ in each outcome category, the inter-observer agreement increased to 92%.

Once again, in instances of disagreement, articles were re-examined to reach a consensus about the abstracted information. Table II presents the abstracted data after consensus.

Table II : Information about study characteristics and outcomes at consensus

Study/yr	N	Lower age limit, year (mean)	Women/ Men	Population	Diagnostic criteria	% detected by G.P.*	% treated	Length of follow-up months	Outcomes (%)	
PRIMARY CARE:										
Kukull et al, 1986	78 (est)	≥60	0/78	VA general medical clinic	Zung SDS≥60	-	-	33	Well Depressed Died No follow-up	24 27 18 31
Kennedy et al, 1991	313	≥65 (75.6)	255/58	Representative sample of Medicare recipients	CES-D≥16	-	9% seen by mental health specialist	24	Well Depressed No follow-up, died	36 31 33
Callahan et al, 1994	410	60 (65.6)	328/82	University affiliated primary care practice	CES-D≥16	-	-	9	Well Depressed No follow-up	36 34 30
Van Marwijk et al, 1997	42	65	-	9 general practices	DIS/DSM 3	28	10	12	Well Depressed No follow-up	43 17 40
COMMUNITY:										
Ben-Arie et al, 1990	23	65	17/6		(PSE) CATEGO	0/6 general clinic attenders	9	42	Well Relapse Depressed Other Died No follow-up	31 4 43 9 9 4
O'Connor et al, 1990	27	≥75	-		CAMDEX (DSM 3)	-	-	12	Well Continuously ill Other Died	22 44 4 30
Kivela et al, 1991, 1995	42	60-94 (70)	29/13		DSM 3	-	-	12	Well Relapsed Continuously ill Other Died	46 12 14 14 14
								60	Well Continuously ill Other Died	12 26 17 45

(Table II continued)

Copeland et al, 1992	123	≥65	91/32	GMS-AGECAT (DSM 3)	-	4% received anti-depressants	36	Well Continuously ill Other Died No follow-up	20 27 20 20 13
Kua, 1993	35	≥65 (72.4)	-	GMS-AGECAT (DSM 3)	2/8 regular medical clinic attenders	6% received anti-depressants	60	Recovered Depressed Subcase Died Other diagnoses No follow-up	23 29 14 14 9 11
Forsell et al, 1994	34	≥75	-	DSM 3R	32	32	36	Well Depressed Dysthymic Demented Died No follow-up	6 35 15 3 38 3
Snowdon et al, 1995	12	≥65	-	DSM 3	-	16% received anti-depressants	48**	Well Depressed Demented Dead No follow-up	8 43 8 8 33
Bowling et al, 1996	129 (est)	≥65	90/39 (est)	GHQ≥6 for anxiety/ depression	-	37% received psychotropic medication	30	Well Depressed No follow-up	38 42 20
% agreement between the two raters:	92	100	100/92	100		92	100	58	

DSM-3: Diagnostic and Statistical Manual of Mental Disorders, 3rd edition, APA, 1980
CES-D: Centre for Epidemiologic Studies Depression Scale, Radloff, 1997
Zung SDS: Zyng Depression Scale, Zung, 1965
PSE: Present State Examination, Wing, 1974
CAMDEX: Cambridge Examination for Mental Disorders in the Elderly, Roth et al, 1986
DSM-3R: Diagnostic and Statistical Manual of Mental Disorders, 3rd edition revised, APA, 1987
GMS: Geriatric Mental State, Copeland et al, 1986
GHQ: General Health Questionnaire, Goldberg, 1978
*G.P. General practitioners
**One of four follow-up periods

4.1.3.1 Primary care studies

One study used DSM- 3 diagnostic criteria (APA, 1980) and 3 used cut-offs on depression symptom rating scales: in two instances, 16 or more on the Center for Epidemiologic Studies Depression Scale (Radloff, 1997) and, in the third instance, 60 or more on Zung Self-Rating Depression Scale (Zung, 1965). Samples varied from 42 to 410 patients. The patients mean ages were reported in 2 studies (65.6 and 75.6 years). One study included men only and in two others 80% or more of the patients were women. Lengths of reported follow-up varied between 9 and 33 months. One study reported the rate of detection by primary care physicians (28%). Two studies reported rates of eventual antidepressant treatment: 9% and 10%, respectively.

4.1.3.2 Community studies

One study used CATEGO (Wing et al, 1974), one used CAMDEX (Roth et al, 1986), two used DSM- 3 criteria, one used DSM- 3R criteria (APA, 1987), two used Geriatric Mental State-AGECAT (Copeland et al, 1986) and one used a score of 6 or more for anxiety-depression on the General Health Questionnaire (Goldberg, 1978). The samples ranged from 12 to 129 patients. The subjects mean ages were reported in 2 studies (70 and 72.4 years). Only 4 reported the gender distribution: most subjects were women. Lengths of reported follow-up varied between 12 and 60 months. Three studies reported rates of detection of depression by a primary care physician: rates ranged from 0 to 32%.

Six studies reported rates of eventual antidepressant treatment: rates ranged from 4 to 37%.

4.1.3.3 Prognostic factors

A variety of prognostic factors were reported in 10 studies although measurement of these factors varied from one study to the next. Older age (Kennedy et al, 1991), poor perceived health (Callahan et al,1994) and total number of life events (Kivela et al, 1991; Kivela, 1995) were associated with poor outcome in one study each. Physical illness was associated with poor outcome in 4 studies (Bowling et al, 1996; Kennedy et al, 1991; Kua, 1993; Snowdon and lane, 1995) but not in 2 others (Kivela et al, 1991; Kivela, 1995; Kukull et al, 1986); physical disability was associated with poor outcomes in 2 studies (Bowling et al, 1996; Kivela et al, 1991; Kivela, 1995) and cognitive impairment in 2 studies (O'Connor et al, 1990; Snowdon and lane, 1995). Finally, severe depression was associated with poor outcome in 2 studies (Bowling et al, 1996; Callahan et al, 1994) but not in 2 others (Kivela et al, 1991; Kivela, 1995; Snowdon and lane, 1995).

4.2 QUANTITATIVE DATA SYNTHESIS

To analyze the results of the different studies, we dichotomized the outcome categories in table II into "percent well" and "percent other" in which this later category, included "percent depressed", "percent dead", "percent lost to follow-up" and "percent in

other categories", see table II. We then used the fixed and mixed effects linear models, described in Chapters 2 and 3 respectively, to summarize the results of the outcome category, percent of subjects well across the different studies at the end of follow-up. The study population (i.e. community or primary care), length of follow-up and lower age limit at enrollment, were considered as study characteristics in the model. In a preliminary analysis, the other study characteristics reported in table II (i. e. gender, diagnostic criteria, percent detected by G.P and percent treated) were also considered but were not found to be associated with the outcome and/or were missing for many studies.

Under the mixed effects model, the parameter estimates were computed using both, the method of moments involving the two approaches: the first approach starting with an ordinary least square estimate (OLS) and the second one starting with a weighted least square estimate (WLS), and the method of maximum likelihood. Finally, we performed a test of homogeneity of the primary outcome across studies by testing that the random effects variance of the mixed effects model is null.

A raw percent of subject well at the end of follow-up and an adjusted percent of subjects well were modeled. The difference between the raw and the adjusted percent well was that, in the first instance, subjects lost to follow-up were grouped in the percent other outcome category, while in the second instance, subjects lost to follow-up were redistributed proportionately across the two outcome categories since we had no evidence on what happen to these subjects who were lost to follow-up. We were suspicious about the fact that they were probably doing well and they just interrupted the study they were

involved in, or they died and there were no relatives to contact or may be they just moved. Therefore, we thought that if we kept the percent lost as a separate category, it might influence our conclusion on the prognosis of the subjects who had completed the studies and were doing well versus all the others.

The prognostic outcomes for raw and adjusted percent well are presented in Tables III and IV respectively, in the following pages, as well as the study characteristics considered in the statistical model i.e. population settings, lower age limit at enrollment and length of follow-up.

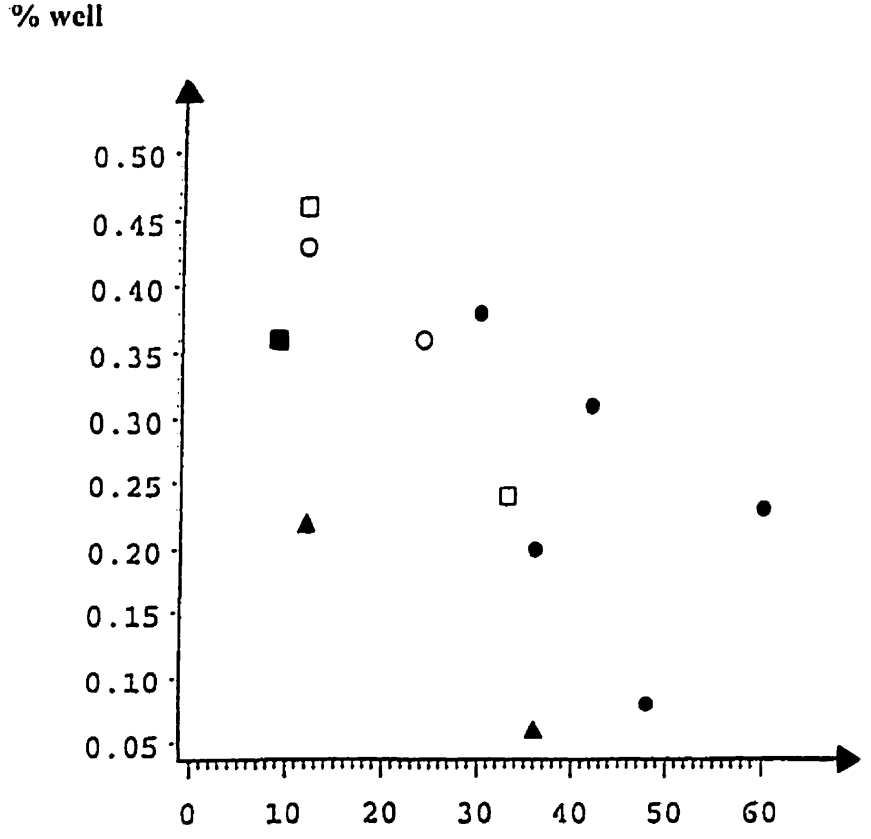
Looking at these tables, we observe heterogeneity in the outcome across studies. The length of follow-up and lower age limit at enrollment seem to be negatively associated with both the raw and adjusted percent well, as can be seen in figures 1 and 2 respectively as well. Clearly, both raw and adjusted percent of subjects well decreases with the increasing length of follow-up and lower age limit 75. There was no significant difference between lower age limit 60 and 65, therefore they were pooled into a single category in the final mixed effects linear model.

No significance difference had been noticed between estimates of the fixed and mixed effects models using first the raw pecent well (table V) then the adjusted percent well (table VI) taking into account population settings, lower age limit at enrollment, and length of follow-up as predictor variables in the model. On the other hand tables VII and VIII respectively, shows the same models mentioned above with one exception, the

Table III : Description of data used in the meta-analysis with raw % well as the outcome

Study	sample size $n_{i(Raw)}$	Raw %well $T_{i(Raw)}$	Settings (X_1)	Lower Age limit (X_2)	Length of follow-up (in months) (X_3)	Estimation variance $V_{i(Raw)}$
Kukull et al.	78	24	Primary care	60	33	0.00234
Kennedy et al.	313	36	Primary care	65	24	0.00074
Callahan et al.	410	36	Primary care	60	9	0.00056
Van Marwijk et al.	42	43	Primary care	65	12	0.00584
Ben-Arie et al.	23	31	Community	65	42	0.00930
O'Connor et al.	27	22	Community	75	12	0.00636
Kivela et al.	42	46	Community	60	12	0.00591
Copeland et al.	123	20	Community	65	36	0.00130
Kua et al.	35	23	Community	65	60	0.00506
Forsell et al.	34	6	Community	75	36	0.00166
Snowdon et al.	12	8	Community	65	48	0.00614
Bowling et al.	129	38	Community	65	30	0.00183

Figure1 : Percent well vs length of follow-up by patient population and lower age limit at enrollment



Legend

Lower age limit at enrollment	Population
□ 60	
O 65	empty = primary care
Δ 75	filled = community

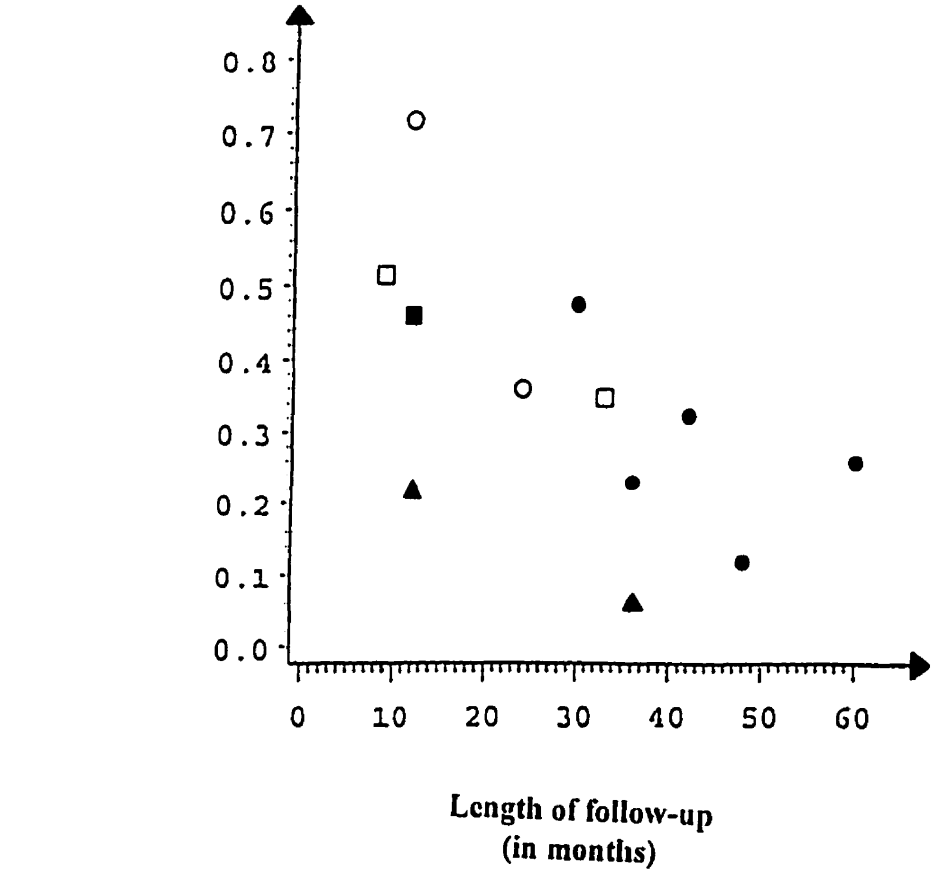
Table IV : Description of data used in the meta-analysis with adjusted % well as the outcome

Study	Adjusted sample size $n_{i\,(Adjusted)}^{a}$	Adjusted % well $T_{i\,(Adjusted)}^{b}$	Settings (X_1)	Age (X_2)	Length of follow-up (in months) (X_3)	Estimation variance $v_{i\,(Adjusted)}$
Kukull et al.	54	34.8	Primary care	60	33	0.00421
Kennedy et al.	313	36.0	Primary care	65	24	0.00074
Callahan et al.	287	51.4	Primary care	60	9	0.00087
Van Marwijk et al.	25	71.7	Primary care	65	12	0.00806
Ben-Arie et al.	22	32.3	Community	65	42	0.00990
O'Connor et al.	27	22.0	Community	75	12	0.00636
Kivela et al.	42	46.0	Community	60	12	0.00591
Copeland et al.	107	23.0	Community	65	36	0.00165
Kua et al.	31	25.9	Community	65	60	0.00615
Forsell et al.	33	6.2	Community	75	36	0.00176
Snowdon et al.	8	12.0	Community	65	48	0.01308
Bowling et al.	103	47.5	Community	65	30	0.00242

a: $n_{i(Adjusted)} = n_{i(Raw)} - (\%\ \text{lost to follow-up}) * n_{i(Raw)}$

b: $T_{i(Adjusted)} = \{ T_{i(Raw)} * n_{i(Raw)} \} / n_{i(Adjusted)}$

Figure2 : Adjusted % well vs length of follow-up by patient population and lower age limit at enrollment



Legend

Lower age limit at enrollment	Population
□ 60	
O 65	empty = primary care
Δ 75	filled = community

population settings predictor had been excluded, since it made no contribution to the models.

In the method of moments, the estimates coming from the two approachs are identical to each other to two decimal places and hence, by this fact we have shown that whether we start by the ordinary least squares estimates from the fixed effects model or from the weighted least square estimate from the same model, these approaches adjust the final estimate accordingly so that they would yield to approximatly to more of less the same final estimates.

The iteration procedure used for the method of maximum likelihood, reached convergence of both the random effect variance and the β estimates along with the corresponding confidence intervals at the 5^{th} and 8^{th} iteration , with raw and adjusted percent well as effect sizes or outcomes, respectively and with lower age limit at enrollment, length of follow-up and population settings as predictor variables in the first instance, then without population settings in the second instance. No major improvement had been noticed in terms of β estimates compared to the ones of the method of moments but on the other hand, a slight improvement in the random effect variance was made. In contrast, there was a slight difference between the mixed and fixed effects model estimates since the null hypothesis of the random effects variance had been rejected at the 0.05 level of significance, which brings us to the conclusion that the mixed effects model best suit our data (see tables V, VI, VII and VIII).

Table V : Results for the raw percent well

	Fixed effects model			Mixed effects model								
				Method of moments						Method of maximum likelihood		
				Approach I (starting with OLS)			Approach II (starting with WLS)					
Parameters	**β**	**p_value**	**95% CI**	**β**	**p-value**	**95% CI**	**β**	**p-value**	**95% CI**	**β**	**p-value**	**95% CI**
Intercept	0.4460	<0.001	(0.3286, 0.5633)	0.4881	< 0.001	(0.3242, 0.6520)	0.4932	< 0.001	(0.3139, 0.6725)	0.4802	< 0.001	(0.3318, 0.6287)
Settings	-0.0127	0.736	(-0.0970, 0.0716)	-0.0339	0.5391	(-0.1558, 0.0879)	-0.0363	0.552	(-0.1714, 0.0987)	-0.0300	0.542	(-0.1388, 0.0787)
Age 75 vs 60 or 65	-0.2029	<0.001	(-0.3025, -0.1032)	-0.2137	0.007	(-0.3542, -0.0732)	-0.2155	0.0127	(-0.3710, -0.0599)	-0.2112	0.005	(-0.3369, -0.0856)
Length of follow-up	-0.0048	0.005	(-0.0077, -0.0019)	-0.0057	0.010	(-0.0097, -0.0017)	-0.0058	0.0145	(-0.0102, -0.0015)	-0.0056	0.007	(-0.0093, -0.0020)
				$\sigma_\theta^2 = 0.0020$			$\sigma_\theta^2 = 0.0030$			$\sigma_\theta^2 = 0.0012$		
$H_0 : \sigma_\theta^2 = 0$	RSS= 18.0710, p-value =0.0200											

Table VI: Results for the adjusted percent well

	Fixed effects model			Mixed effects model								
				Method of moments						Method of maximum likelihood		
				Approach I (starting with OLS)			Approach II (starting with WLS)					
Parameters	β	p-value	95% CI	β	p-value	95% CI	β	p-value	95% CI	β	p-value	95% CI
Intercept	0.5979	<0.001	(0.4677, 0.7282)	0.5829	< 0.001	(0.3556, 0.8101)	0.5830	< 0.001	(0.3674, 0.7986)	0.5844	< 0.001	(0.4084, 0.7604)
Settings	-0.0209	0.614	(-0.1131, 0.0711)	0.0151	0.849	(-0.1627, 0.1931)	0.0128	0.865	(-0.1551, 0.1806)	0.0017	0.977	(-0.1314, 0.1348)
Age 75 vs 60 or 65	-0.2638	<0.001	(-0.3689, -0.1588)	-0.2715	0.014	(-0.4716, -0.0714)	-0.2714	0.010	(-0.4599, -0.0827)	-0.2699	0.003	(-0.4185, -0.1213)
Length of follow-up	-0.0077	<0.001	(-0.0110, -0.0044)	-0.0070	0.019	(-0.0126, -0.0015)	-0.0070	0.015	(-0.0123, -0.0017)	-0.0071	0.005	(-0.0115, -0.0027)
				$\sigma_\theta^2 = 0.0064$			$\sigma_\theta^2 = 0.0054$			$\sigma_\theta^2 = 0.0023$		
$H_0 : \sigma_\theta^2 = 0$	RSS = 22.1290, p-value = 0.0047											

Table VII : Results for raw percent well without population settings

	Fixed effects model			Mixed effects model								
				Method of moments						Method of maximum likelihood		
				Approach I (starting with OLS)			Approach II (starting with WLS)					
Parameters	β	p-value	95% CI	β	p-value	95% CI	β	p-value	95% CI	β	p-value	95% CI
Intercept	0.43086	<0.001	(0.3709, 0.4908)	0.4515	< 0.001	(0.3534, 0.5496)	0.4540	< 0.001	(0.3487, 0.5593)	0.4474	< 0.001	(0.3594, 0.5354)
Age 75 vs 60 or 65	-0.1966	<0.001	(-0.2854, -0.1078)	-0.1967	0.005	(-0.3174, -0.0760)	-0.1970	0.007	(-0.3252, -0.0688)	-0.1964	0.003	(-0.3071, -0.0857)
Length of follow-up	-0.0045	0.001	(-0.0067, -0.0023)	-0.0051	0.005	(-0.0082, -0.0019)	-0.0051	0.006	(-0.0084, -0.0018)	-0.0050	0.003	(-0.0078, -0.0021)
				σ_θ^2 = 0.0017			σ_θ^2 = 0.0023			σ_θ^2 =0.0011		
$H_0 : \sigma_\theta^2 = 0$	RSS= 18.1927, p_value =0.0330											

Table VIII : Results for the adjusted percent well without population settings

	Fixed effects model			Mixed effects model								
				Method of moments						Method of maximum likelihood		
				Approach I (starting with OLS)			Approach II (starting with WLS)					
Parameters	β	p-value	95% CI	β	p-value	95% CI	β	p-value	95% CI	β	p-value	95% CI
Intercept	0.5732	<0.001	(0.5026, 0.6438)	0.5955	< 0.001	(0.4519, 0.7391)	0.5922	< 0.001	(0.4611, 0.7233)	0.5861	< 0.001	(0.4750, 0.6971)
Age 75 vs 60 or 65	-0.2532	<0.001	(-0.3454, -0.1609)	-0.2779	0.005	(-0.4458, -0.1100)	-0.2756	0.002	(-0.4288, -0.1224)	-0.2707	0.001	(-0.4014, -0.1400)
Length of follow-up	-0.0072	<0.001	(-0.0098, -0.0047)	-0.0072	0.005	(-0.0117, -0.0028)	-0.0074	0.003	(-0.0113, -0.0032)	-0.0072	0.001	(-0.0107, -0.0036)
				$\sigma_\theta^2 = 0.0056$			$\sigma_\theta^2 = 0.0042$			$\sigma_\theta^2 = 0.0023$		
$H_0 : \sigma_\theta^2 = 0$	RSS = 22.405, p-value =0.0077											

To test the null hypothesis H_0: $\sigma_\theta^2 = 0$, the weighted residual sum of square (RSS) for the fixed effect model had been computed then compared to the Chi-square distribution. From the bottom of the above tables V, VI, VII and VIII respectively, we can clearly see that the p-value of the test is less than 0.05 and therefore we have enough evidence to reject the null hypothesis. This fact had then been confirmed by the method of maximum likelihood, where the random effect variance for example, is equal to 0.0011 and 0.0023 at convergence for both raw and adjusted percent well respectively (table VII and VIII).

Therefore we can conclude that part of the variability in our model had not been explained by its characteristics, such as lower age limit at enrollment and length of follow-up. Hence the mixed effects model is the appropriate one to fit our data.

All statistical analyses were conducted using the procedure IML in ***SAS statistical software***, version 6.12 (SAS, 1997). The SAS IML programs and outputs are provided in appendices A to D.

CONCLUSION

The first fundamental element of any systematic research synthesis is gathering the essential material to performing this task in the most reliable, valid an free of bias way possible. Once the information had been made available, data is being abstracted and therefore, a quantitative synthesis is required in order to be able to answer questions about the generality of an effect.

In the first Chapter of this thesis, the review of the literature which includes the search of the computerized databases such as **MEDLINE** and **PsycINFO** along with methods of developing a search strategy, data abstraction, reliability and validity check had been discussed in details.

Methods of modeling the data available from different studies had been presented in the second and third Chapters respectively. Both the fixed and the mixed effects models had been discussed.

Finally in the fourth and last Chapter, the theory of the above mentioned models had been applied to a data set dealing with the prognosis of depression in elderly in primary care and community based settings.

One may ask the question of when to use the fixed versus mixed effects model. Generally, if we consider the extreme case of a synthesis of two studies, the fixed effects approach seems more sensitive. The mixed effects analysis would supply a poor summary. The random effects variance would be estimated with extremely poor precision and the notion of generalizing to a larger population of studies will be ironic. On the other hand, if we consider a research synthesis with several hundred studies, the fixed effects approach would be clearly inappropriate, since the assumption that the studies are sampled from a well define population would make a little sense.

The mixed effects linear model has the advantage of including the fixed effect linear model as its special case when the random effect variance is tested to be null. The mixed effects model, as it had been discussed in Chapter 3, takes into account the variability existing between studies using study characteristics such as, age and length of follow-up, reducing with this fact the biases resulting from pooling effect size estimates by the mean of weighted averages. However, we may be confronted to the problem of "overfitting" if we include a large number of predictors some of which may seem to improve prediction strictly as a result of chance, or rather "underfitting", if some of the predictors which are related to the outcome are not included in the model and hence may result in a bias in the model estimates. Therefore, in case of a large amount of unexplained heterogeneity remains after fixed effect modeling, one should consider turning to the mixed effects synthesis in which a model is augmented by the addition of the term representing unexplained sources at between-study heterogeneity. In the simplest special case of the

fixed and mixed effects models, the estimated coefficients ($\hat{\beta}$) of the study characteristics are shifted toward zero by an amount directly dependent on their estimated variances, and therefore as if the outcomes had been pooled by the means of weighted averages. On the other hand, when the heterogeneity between studies is small relative to study's specific variance, that is the residual sum of squares ends up smaller than its degrees of freedom, essentially the same summary conclusions should be arrived at using a fixed or a mixed effects approach (Greenland and Rothman, 1998)

At this stage of the thesis, one may be able to define a meta-analysis as a method focusing on contrasting and combining results from different studies hoping to identify consistent patterns and sources of disagreement among those results. This data synthesis method had been greeted warmly in educational, social science and medical research areas, whereas major criticism had faced this method in epidemiology. Heterogeneity between study results is often a major factor resulting from poor quality of the data available in these studies. Therefore, due to the fact that in epidemiology, systematic reviews involve much smaller number of studies compared to social sciences or medicine where hundreds of studies are available, greater pressure to refine summarization techniques is been made. However, because of the rapid growth of epidemiologic research, a simple narrative (or qualitative) review is no longer reliable. Both qualitative and quantitative aspect of a meta-analysis in a systematic review can and should be complementary in order to convey a balanced picture of the material. A purely statistical analysis cannot convey explanations of results in terms of bias, while a purely qualitative

analysis will lack precision and can easily miss small but important associations or patterns in the material (Greenland and Rothman, 1998).

In recognizing the need for the meta-analysis, one should also be aware of its limitations. In particular, the causal explanation of similarities and differences among study results noted in meta-analysis is a qualitative aspect of the review and hence outside the scope of quantitative analysis. In Greenland, 1994, there was a debate on whether to favor meta-analysis or rather to ban it from publication. Shapiro (Shapiro, S., 1994) strongly disagreed on the use of meta-analysis in particular meta-analyses of observational studies. Whereas, Greenland argue this matter in a less categorical way by strongly recommending the use of random effects models in the data synthesis, and reducing the biases by making more emphasis on methods of attributing quality measures to the different studies at hand.

What had been discussed in this thesis is the frequentist or the classical approach to a systematic data synthesis. However, statistical models used in research syntheses are based on assumptions that require justification which the analyst may find little difficulty proving that there are correct. Alternatively, the bayesian approach provides a formal structure for incorporating such uncertainties. All unknown parameters in the model are treated as random variables that are governed by a joint probability distribution specified prior to viewing the data. This prior distribution is based on evidence that previously exists and is updated in the light of the given data to produce the posterior distribution

under which the statistical inferences will be made. Under the next investigation, this posterior distribution will act as the prior, and so on.

A generalized linear mixed model (GLMM), which is a generalized form of the mixed effect model described in Chapter 3 is available when the outcome estimates are discrete, that is binomial or Poisson. Using the bayesian approach, two methods can be employed for estimation: 1) the penalized quasi-likelihood (PQL) method and 2) the marginal quasi-likelihood (MQL) method. To avoid numerical integration, these previously stated methods use Gibbs sampling techniques by taking repeated samples from the posterior distribution (Breslow and Clayton, 1993).

For future work, it would be interesting to investigate meta-analysis of a multinomial random variable as an outcome where all its categories would be grouped in one single vector as opposed to analyzing each vector category separately, as was done in the work of this thesis.

REFERENCES

American Psychiatric Association (APA) (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd Edition. Washington, DC.

American Psychiatric Association (APA) (1987). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd Edition, Revised. Washington, DC.

Ben-Arie, O., Welman, M., Teggin, A. F. (1990). The depressed elderly living in the community: a follow-up study. *Br. J. Psychiatry*, **157**, 425-7.

Birge, R. T. (1932). The calculation of errors by the method of least squares. *Physical Review*, **40**, 207-227.

Breslow, N. E., Clayton, D. G. (1993). Approximate inference in generalized linear models. *Journal of the American Statistical Association*, **88**, 9-24.

Bryk, A. S., Raudenbush, S. W., Seltzer, M. and Congdon, R. T. (1988). *An introduction to HLM: Computer program and user's guide*. University of Chicago, Department of Education: Chicago.

Bowling, A., Farquhar, M., Grundy, E. (1996). Outcome of anxiety and depression at two and a half years after baseline interview: associations with changes in psychiatric morbidity among three samples of elderly people living at home. *Int. J. Geriatr. Psychiatry*, **11**, 119-29.

Callahan, C. M., Hui, S. L., Nienaber, N. A., Musick, B. S., Tierney, W. M. (1994). Longitudinal study of depression and health services use among elderly primary care patients. *J. Am. Geriatr. Soc.*, **42**, 833-8.

Chalmers, T. C., Smith, H., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., Ambroz, A. (1981). A method for assessing the quality of a randomized control trial. *Controlled clinical trials*, **2**, 31-49.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society (Supplement)*, **4**, 102-118.

Cooper, H., Hedges, L. (1994). *The handbook of research synthesis*. Russell Sage Foundation : New York.

Copeland, J. R., Dewey, M. E., Griffiths-Jones, H. M. (1986). A computerized psychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGECAT. *Psychol. Med.*, **16**, 89-99.

Copeland, J. R. M., Davidson, I. A., Dewey, M. E. et al. (1992). Alzheimer's disease, other dementia, depression and pseudodementia: prevalence, incidence and three-year outcome in Liverpool. *Br. J. Psychiatry*, **161**, 230-9.

Fisher, R. A. (1932). *Statistical methods for research workers*. Oliver & Boyd : London.

Forsell, Y., Jorm, A. F., Winblad, B. (1994). Outcome of depression in demented and non-demented elderly: observations from a three-year follow-up in a community-based study. *Int. J. Geriatr. Psychiatry*, **9**, 5-10.

Glass, G. V., McGaw, B., and Smith, M. L. (1981). *Meta- analysis in social research.* Sage : Beverly Hills.

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

Greenland, S, Rothman, K. J. (1998). *Modern epidemioloy*. Philadelphia: Lippincott-Raven.

Greenland, S. (1994). Can meta-analysis be salvaged?. *American Journal of Epidemioloy*, **140**, 783-787.

Goldberg, D. P. (1978). *Manual of the General Health Questionnaire*. NFER-Nelson, Windsor.

Hedges, L. V. and Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press : Orlando.

Hunter, J. E., Schmidt, F. L. and Jackson G. B (1982). *Meta-analysis: Cumulating research findings across studies*. Sage : Beverly Hills.

Jadad, A. R., Moore, A., Carroll, D. et al. (1996). Blind assessment of the quality of trials reports. *Controlled clinical trials*, **17**, 1-12.

Fleiss, Joseph, L. (1981). *Statistical methods for rates and proportions.* Wiley: New york.

Kennedy, G. J., Kelman, H. R., Thomas, C. (1991). Persistence and remission of depressive symptoms in late life. *Am. J. Psychiatry*, **148**,174-8.

Kivela, S. L., Pahkala, K., Laippala, P. (1991). A one-year prognosis of dysthymic disorder and major depression in old age. *Int. J. Geriatr. Psychiatry*, **6**, 81-7.

Kivela, S. L. (1995). Long term prognosis of major depression in old age: a comparison with the prognosis of dysthymia disorder. *Int. Psychogeriatr.*, **7**, 69-82.

Kua, E. H. (1993). The depressed elderly Chinese living in the community: a five-year follow-up study. *Int. J. Geriatr. Psychiatry*, **8**, 427-30.

Kukull, W. A., Koepsell, T. D., Inui, T. S. et al. (1986). Depression and physical illness among elderly general medical clinic patients. *J. Affective Disord.*, **10**, 153-62.

Laupacis, A., Wells, G., Richardson, S., Tugwell, P. (1994). Users' guides to the medical literature: V. How to use an article about prognosis. *J. A. M. A.* , **272**, 234-7.

Lin, Lawrence I-Kuei (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255-268.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, **74**, 817-27.

O'Connor, D. W., Pollitt, P. A., Roth, M. (1990). Coexisting depression and dementia in a community survey of the elderly. *Int. Psychogeriatr.*, **2**,45-53.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical journal*, **3**, 1243-1246.

Radloff, L. S. (1997). The CES-D Scale: a new self report depression scale for research in the general population. *Appl. Psychol. Measurement*, **1**, 385-401.

Rosenthal, R. (1984.). *Meta- analytic procedures for social research*. Sage: Beverly Hills.

Roth, M., Tym, E., Mountjoy, C. P., Huppert, F. A., et al. (1986). CAMDEX: A standardized instrument for the diagnosis of mental disorder in the elderly with special reference to the detection of dementia. *Br. J. Psychiatry*, **149**, 698-709.

S.A.S. Institute Inc. (1997). Version 6.12. Cary NC.

Seber, G. A. F. (1971). *Linear regression models*. Wiley: New york.

Shapiro, S. (1994). Meta-analysis/shmeta-analysis. *Am. J. Epidemiol.*, **140**, 771-778.

Snowdon, J., Lane, F. (1995). The Botany survey: a longitudinal study of depression and cognitive impairment in an elderly population. *Int. J. Geriatr. Psychiatry*, **10**, 349-58.

Tippett, L. H. C. (1931). *The methods of statistics*. William & Norgate.

Van Marwijk, H. W. J. (1995). Recovery of late life depression: A study among elderly Dutch general practice patients. In: Depression in the elderly as seen in General Practice, eds Van Marwijk, H. W. J. Leiden, The Netherlands, Procfshrift Rijkuniversiteit.

Wing, P. K., Cooper, J. E., Mann, S. A., et al. (1974). *The Measurement and Classification of Psychiatric Symptoms*. Cambridge University Press : London.

Yates, F., Cochrane, W. G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science*, **28**, 556-580.

Zung ,W. W. K. (1965). A self-rating depression scale. *Arch. Gen. Psychiatry*, **12**, 63-70.

ANNEX A

FIXED AND MIXED EFFECTS MODELS

Method of moments

Option pagesize=77 linesize=130 ;

Data meta;

infile '----------\consensus.dat';

input type study$ fic drp fle coff ooa boa aepf nb age men women dc pdetect ptreat lfoll well dep dyst dement relapse subcase lost dead other;

label fic =" formation of inception cohort" drp = "description of referral pattern" fle = "follow-up long enough" coff = "completion of follow-up" ooa = "objective outcome assessment" boa = "blind outcome assessment" aepf = "adjustment for extraneous prognostic factors" nb = "sample size" age = "age" men = "men" women = "women" dc = "diagnostic criteria" detect = "percent detected by primary physicians" ptreat = "percent treated" lfoll = "length of follow-up" well ="patients doing well" dep = "patients depressed" lost ="patients lost to follow-up" dead = "patients dead" other ="other patients" dyst = " dysthymic patients" dement = "demented patients"

run;

nb1=nb-(nb*(lost/100)); {Adjusted sample size}

well1=(nb*(well/100));
ad_well=(well1/nb1); {Adjusted percent well}

p_well=well/100; {Raw percent well}

t=p_well;
*t=ad_well;

```
v=((t)*(1-t))/nb;                    {Raw estimation variance}
*v=((t)*(1-t))/nb1;                  {Adjusted estimation variance}


title 'Meta-Analysis: Consensus Data';
proc print data=meta; run;

proc glmmod data=meta  outparm=meta1 outdesign=meta2 noprint;
class type age ; model t= type age lfoll v; run;

proc print data=meta1; run;
proc print data=meta2; run;

run;

PROC IML;

use meta2;
read all var {col1 col2 col6 col7} into x;
read all var {col8} into e;
read all var {t} into y;

b_label ={Intercept, type, age, lfoll};


n=nrow(y);
I=i(n);

/*************************************************************************

                        Mixed Effects Model
              Y = XB + u + e; E(Y)=XB, Var(u+e)=sigma2*I + V

                        & Fixed Effects Model
                        when sigma2 is set to zero

*************************************************************************/

V  = diag(e);
VI = inv(V);
M  = x*(ginv(t(x)*x))*t(x);
MV = x*(ginv(t(x)*VI*x))*t(x);
df = trace(I-M);
```

```
/*******  Method of moments starting with OLS estimates  ********/


B_ols  = (ginv(t(x)*x))*t(x)*y;
RSS_ols = t(y)*(I-M)*y;
p_value = 1 - probchi(RSS_ols, df);
sig_ols = ( RSS_ols - (trace(V - M*V)) ) / trace(I-M);
sig2_ols= max(0, sig_ols);
W_ols  = diag(j(n,1,sig2_ols)) + V;
WI_ols  = inv(W_ols);
Bw_ols  = (ginv(t(x)*WI_ols*x))*t(x)*WI_ols*y;
v_Bw_ols= (ginv(t(x)*WI_ols*x))*(t(x)*WI_ols*x)*t(ginv(t(x)*WI_ols*x));
s_Bw_ols= ( ( diag(v_Bw_ols) )*j(nrow(Bw_ols),1,1) )##.5;


t05    = tinv(.975,df);

p_ols  = (1- probt(abs(Bw_ols/s_Bw_ols), df))*2;

L95_ols = Bw_ols - (t05*s_Bw_ols);
U95_ols = Bw_ols + (t05*s_Bw_ols);

print 'Meta Analysis';
print x e y;



/*******  Method of moments starting with WLS estimates  ********/


B_wls  = (ginv(t(x)*VI*x))*t(x)*VI*y;
RSS_wls = t(y)*(VI - VI*MV*VI)*y;
p_value = 1 - probchi(RSS_wls,df);
sig_wls = ( RSS_wls - trace(I - VI*MV) )/( trace(VI-VI*MV*VI) );
sig2_wls= max(0, sig_wls);

W_wls  = diag(j(n,1,sig2_wls)) + V;
WI_wls  = inv(W_wls);
Bw_wls  = (ginv(t(x)*WI_wls*x))*t(x)*WI_wls*y;
v_Bw_wls= (ginv(t(x)*WI_wls*x))*(t(x)*WI_wls*x)*t(ginv(t(x)*WI_wls*x));
s_Bw_wls= ( ( diag(v_Bw_wls) )*j(nrow(Bw_ols),1,1) )##.5;

p_wls  = (1- probt(abs(Bw_wls/s_Bw_wls), df))*2;

L95_wls = Bw_wls - (t05*s_Bw_wls);
```

```
U95_wls = Bw_wls + (t05*s_Bw_wls);

print '/******* Method of moments starting with WLS estimates ********/';

print  sig2_wls b_label Bw_wls s_Bw_wls p_wls L95_wls U95_wls;

print '/** Test of Ho: sigma2=0  (i.e. random effects variance is null) **/';

print  RSS_wls df p_value;

QUIT;
```

ANNEX B

MIXED EFFECTS MODEL

Method of maximum likelihood

```
Option pagesize=77 linesize=130 ;

Data meta;

infile '----------\consensus.dat';

input type study$ fic drp fle coff ooa boa aepf nb age men women dc pdetect ptreat lfoll
well dep dyst dement relapse subcase lost dead other;


*label fic =" formation of inception cohort" drp = "description of referral pattern" fle =
"follow-up long enough" coff = "completion of follow-up" ooa = "objective outcome
assessment" boa= "blind outcome assessment" aepf = "adjustment for extraneous
prognostic factors"  nb = "sample size" age = "age" men = "men" women = "women" dc
= "diagnostic criteria" detect = "percent detected by primary physicians" ptreat = "
percent treated" lfoll = "length of follow-up" well ="patients doing well" dep = " patients
depressed" lost =" patients lost to follow-up" dead =" patients dead " other =" other
patients" dyst = "dysthymic patients" dement = " demented patients";

run;

*nb1=nb-(nb*(lost/100));


well1=(nb*(well/100));

ad_well=(well1/nb1);
p_well=well/100;

t=p_well;
*t=ad_well;
```

```
v=((t)*(1-t))/nb;

*v=((t)*(1-t))/nb1;

run;
```

Title '*****Meta-Analysis: Consensus Data*******';**

```
proc print data=meta; run;


proc glmmod data=meta outparm=meta1 outdesign=meta2 noprint;
class type age ; model t= age lfoll v; run;

proc print data=meta1; run;
proc print data=meta2; run;
```

PROC IML;

```
use meta2;

read all var {col1 col4 col5} into x;
read all var {col6} into e;
read all var {t} into y;

b_label={Intercept, age, lfoll};
```

/**********

Mixed Effect Model

Y = XB + u + e; E(Y)=XB, Var(u+e)=sigma2*I + V

**********/

```
n=nrow(y);
I=i(n);
V  = diag(e);
```

```
VI = inv(V);
M  = x*(ginv(t(x)*x))*t(x);
MV = x*(ginv(t(x)*VI*x))*t(x);
df = trace(I-M);

Title '/*******  Method of maximum likelihood  ********/';

RSS_wls = t(y)*(VI - VI*MV*VI)*y;
sig_wls = (RSS_wls - trace(I - VI*MV))/(trace(VI-VI*MV*VI));


Title '/**********Iteration procedure using SAS macro*************/';


%macro f(x);
%do %while (&x le 3);
%let y=%eval(&x-1);

if &x=1 then sig2_i=max(0, sig_wls);
if &x>1 then sig2_i=sig2_n&y;

W_wls  = diag(j(n,1,sig2_i)) + V;
WI_wls  = inv(W_wls);
Bw_wls=(ginv(t(x)*(WI_wls)*x))*t(x)*(WI_wls)*y;
print Bw_wls;

invL= inv(t(x)*WI_wls*(x));
P= x*(invL)*t(x)*WI_wls;
N=(I - P)*y;
R=diag(N);
sq_R=R*R;
RV=sq_R - V;
W_wls_2= W_wls* W_wls;
RVW=(W_wls_2*RV);

n_sig_2 = trace(RVW);
d_sig_2= trace(W_wls_2);
sig2_&x= trace(RVW) / trace(W_wls_2);

n=nrow(y);
sig2_n&x=max(0, sig2_&x);
print sig2_&x sig2_n&x;

W_n  = diag(j(n,1,sig2_n&x)) + V;
```

```
WI_n  = inv(W_n);
Bw_n&x  = (ginv(t(x)*(WI_n)*x))*t(x)*(WI_n)*y;
vBw_n&x= (ginv(t(x)*WI_n*x))*(t(x)*WI_n*x)*t(ginv(t(x)*WI_n*x));
sBw_n&x= ( ( diag(vBw_n&x) )*j(nrow(Bw_n&x),1,1) )##.5;
p_n&x   = (1- probt(abs(Bw_n&x/sBw_n&x), df))*2;

t05  = tinv(.975,df);
L95_n&x = Bw_n&x - (t05*sBw_n&x);
U95_n&x = Bw_n&x + (t05*sBw_n&x);

sqt_s2=sqrt(2/d_sig_2);


          print '/*******  Method of maximum likelihood ********/';


print sig2_i d_sig_2 sig2_&x sig2_n&x  b_label Bw_n&x  p_n&x
L95_n&x  U95_n&x  sqt_s2;

%let x = %eval(&x+1); %end;
%mend f;
%f(1);

QUIT;
```

ANNEX C

SAS PRINT OUT OF THE ESTIMATE OF THE MIXED EFFECTS MODEL USING THE METHOD OF MOMENT

Using the raw percent well as the effect size

OBS	_COLNUM_	EFFNAME	TYPE	AGE
1	1	INTERCEPT		
2	2	AGE		1
3	3	AGE		2
4	4	AGE		3
5	5	LFOLL		
6	6	V		

OBS	T	COL1	COL2	COL3	COL4	COL5	COL6
1	0.24	1	1	0	0	33	.0023385
2	0.36	1	0	1	0	24	.0007361
3	0.36	1	1	0	0	9	.0005620
4	0.43	1	0	1	0	12	.0058357
5	0.31	1	0	1	0	42	.0093000
6	0.22	1	0	0	1	12	.0063556
7	0.46	1	1	0	0	12	.0059143
8	0.20	1	0	1	0	36	.0013008
9	0.23	1	0	1	0	60	.0050600
10	0.06	1	0	0	1	36	.0016588
11	0.08	1	0	1	0	48	.0061333
12	0.38	1	0	1	0	30	.0018264

Meta Analysis

X			E	Y
1	0	33	0.0023385	0.24
1	0	24	0.0007361	0.36
1	0	9	0.000562	0.36
1	0	12	0.0058357	0.43
1	0	42	0.0093	0.31
1	1	12	0.0063556	0.22

1	0	12	0.0059143	0.46
1	0	36	0.0013008	0.2
1	0	60	0.00506	0.23
1	1	36	0.0016588	0.06
1	0	48	0.0061333	0.08
1	0	30	0.0018264	0.38

corresponding to table vii in chapter 4

/** Method of moments starting with OLS estimates ****/**

SIG2_OLS	B_LABEL	BW_OLS	S_BW_OLS	P_OLS	L95_OLS	U95_OLS
0.001768	INTERCEPT	0.4515693	0.0433713	2.555E-6	0.3534565	0.5496821
	AGE	-0.196741	0.0533679	0.0050245	-0.317468	-0.076015
	LFOLL	-0.005103	0.0013814	0.0049682	-0.008227	-0.001978

/** Method of moments starting with WLS estimates ****/**

SIG2_WLS	B_LABEL	BW_WLS	S_BW_WLS	P_WLS	L95_WLS	U95_WLS
0.0022952	INTERCEPT	0.4540194	0.0465379	4.3937E-6	0.3487433	0.5592954
	AGE	-0.197024	0.0566909	0.006989	-0.325267	-0.06878
	LFOLL	-0.005156	0.0014622	0.0064509	-0.008464	-0.001848

/Test of Ho: sigma2=0 (i.e.random effects variance is null)**/**

RSS_WLS	DF	P_VALUE
18.19271	9	0.0330025

Using the adjusted % well as the effect size

OBS	T	COL1	COL2	COL3	COL4	COL5	COL6
1	0.34783	1	1	0	0	33	0.004215
2	0.36000	1	0	1	0	24	0.000736
3	0.51429	1	1	0	0	9	0.000870
4	0.71667	1	0	1	0	12	0.008058
5	0.32292	1	0	1	0	42	0.009902

6	0.22000	1	0	0	1	12	0.006356
7	0.46000	1	1	0	0	12	0.005914
8	0.22989	1	0	1	0	36	0.001654
9	0.25843	1	0	1	0	60	0.006152
10	0.06186	1	0	0	1	36	0.001760
11	0.11940	1	0	1	0	48	0.013078
12	0.47500	1	0	1	0	30	0.002416

Meta Analysis

X			E	Y
1	0	33	0.0042148	0.3478261
1	0	24	0.0007361	0.36
1	0	9	0.0008704	0.5142857
1	0	12	0.0080578	0.7166667
1	0	42	0.0099022	0.3229167
1	1	12	0.0063556	0.22
1	0	12	0.0059143	0.46
1	0	36	0.0016544	0.2298851
1	0	60	0.0061522	0.258427
1	1	36	0.0017595	0.0618557
1	0	48	0.0130778	0.119403
1	0	30	0.0024164	0.475

corresponding to table viii in chapter 4

/* Method of moments starting with OLS estimates ***/**

SIG2_OLS	B_LABEL	BW_OLS	S_BW_OLS	P_OLS	L95_OLS	U95_OLS
0.0055723	INTERCEPT	0.5955402	0.0634806	6.0721E-6	0.4519372	0.7391432
	AGE	-0.277935	0.0742173	0.0045908	-0.445826	-0.110044
	LFOLL	-0.007279	0.0019509	0.004689	-0.011693	-0.002866

/* Method of moments starting with WLS estimates ***/**

SIG2_WLS	B_LABEL	BW_WLS	S_BW_WLS	P_WLS	L95_WLS	U95_WLS
0.0041652	INTERCEPT	0.5922536	0.0579336	2.9773E-6	0.4611987	0.7233085
	AGE	-0.275621	0.0677496	0.0028071	-0.428881	-0.122361
	LFOLL	-0.007231	0.001804	0.0030704	-0.011312	-0.00315

/Test of Ho: sigma2=0 (i.e.random effects variance is null)**/**

RSS_WLS	DF	P_VALUE
22.404637	9	0.0076813

ANNEX D

SAS PRINT OUT OF THE ESTIMATES OF THE MIXED EFFECTS MODEL USING THE METHOD OF MAXIMUM LIKELIHOOD

Using raw percent well as the effect size

corresponding to table vii in chapter 4

/********Iteration procedure using SAS macro **********/**

1st iteration

BW_WLS
0.4540194
-0.197024
-0.005156

SIG2_1	SIG2_N1
0.0011257	0.0011257

SIG2_I	D_SIG_2	SIG2_1	SIG2_N1	B_LABEL	BW_N1	P_N1	L95_N1	U95_N1	SQT_S2
0.0022952	561373.93	0.0011257	0.0011257	INTERCEPT	0.4473838	1.0915E-6	0.3595115	0.5352561	0.0018875
				AGE	-0.196378	0.0030371	-0.307	-0.085756	
				LFOLL	-0.005005	0.0033616	-0.007873	-0.002138	

2nd iteration

BW_WLS
0.4473838
-0.196378
-0.005005

SIG2_2	SIG2_N2
0.0011311	0.0011311

SIG2_I	D_SIG_2	SIG2_2	SIG2_N2	B_LABEL	BW_N2	P_N2	L95_N2	U95_N2	SQT_S2
0.0011257	1249511.4	0.0011311	0.0011311	INTERCEPT	0.447427	1.1007E-6	0.359459	0.535395	0.0012652
				AGE	-0.196381	0.0030518	-0.307093	-0.085669	
				LFOLL	-0.005006	0.0033742	-0.007876	-0.002136	

3rd iteration

BW_WLS
0.447427
-0.196381
-0.005006

SIG2_3	SIG2_N3
0.0011313	0.0011313

SIG2_I	D_SIG_2	SIG2_3	SIG2_N3	B_LABEL	BW_N3	P_N3	L95_N3	U95_N3	SQT_S2
0.0011311	1243508.3	0.0011313	0.0011313	INTERCEPT	0.4474292	1.1012E-6	0.3594563	0.5354402	0.0012682
				AGE	-0.196381	0.0030526	-0.307098	-0.085665	
				LFOLL	-0.005006	0.0033748	-0.007876	-0.002136	

4th iteration

BW_WLS
0.4474292
-0.196381
-0.005006

SIG2_4	SIG2_N4
0.0011314	0.0011314

SIG2_I	D_SIG_2	SIG2_4	SIG2_N4	B_LABEL	BW_N4	P_N4	L95_N4	U95_N4	SQT_S2
0.0011313	1243205.2	0.0011314	0.0011314	INTERCEPT	0.4474293	1.1012E-6	0.3594562	0.5354024	0.0012684
				AGE	-0.196381	0.0030526	-0.307098	-0.085664	
				LFOLL	-0.005006	0.0033749	-0.007876	-0.002136	

5th iteration

BW_WLS
0.4474293
-0.196381
-0.005006

SIG2_5	SIG2_N5
0.0011314	0.0011314

SIG2_I	D_SIG_2	SIG2_5	SIG2_N5	B_LABEL	BW_N5	P_N5	L95_N5	U95_N5	SQT_S2
0.0011314	1243189.9	0.0011314	0.0011314	INTERCEPT	0.4474293	1.1012E-6	0.3594562	0.5354024	0.0012684
				AGE	-0.196381	0.0030526	-0.307098	-0.085664	
				LFOLL	-0.005006	0.0033749	-0.007876	-0.002136	

Using adjusted percent well as an effect size

corresponding to table viii in chapter 4

/**********Iteration procedure using SAS macro ***********/

1th iteration

BW_WLS
0.5922536
-0.275621
-0.007231

SIG2_1	SIG2_N1
0.002548	0.002548

SIG2_I	D_SIG_2	SIG2_1	SIG2_N1	B_LABEL	BW_N1	P_N1	L95_N1	U95_N1	SQT_S2
0.0041652	219787.79	0.002548	0.002548	INTERCEPT	0.5870986	9.9331E-7	0.4730586	0.7011387	0.0030166
				AGE	-0.271547	0.0013196	-0.405543	-0.137551	
				LFOLL	-0.007174	0.0015759	-0.010813	-0.003536	

2th iteration

BW_WLS
0.5870986
-0.271547
-0.007174

SIG2_2	SIG2_N2
0.0023488	0.0023488

SIG2_I	D_SIG_2	SIG2_2	SIG2_N2	B_LABEL	BW_N2	P_N2	L95_N2	U95_N2	SQT_S2
0.002548	410472.74	0.0023488	0.0023488	INTERCEPT	0.5863185	8.3915E-7	0.4746674	0.6979696	0.0022074
				AGE	-0.27087	0.0011781	-0.402241	-0.139499	
				LFOLL	-0.007168	0.0014202	-0.010745	-0.003591	

3rd iteration

BW_WLS
0.5863185
-0.27087
-0.007168

SIG2_3	SIG2_N3
0.0023101	0.0023101

SIG2_I	D_SIG_2	SIG2_3	SIG2_N3	B_LABEL	BW_N3	P_N3	L95_N3	U95_N3	SQT_S2
0.0023488	452250.79	0.0023101	0.0023101	INTERCEPT	0.5861624	8.1115E-7	0.4749853	0.6973394	0.0021029
				AGE	-0.270732	0.0011516	-0.401585	-0.13988	
				LFOLL	-0.007167	0.0013907	-0.010732	-0.003602	

4th iteration

BW_WLS

0.5861624
-0.270732
-0.007167

SIG2_4 SIG2_N4
0.002302 0.002302

SIG2_I	D_SIG_2	SIG2_4	SIG2_N4	B_LABEL	BW_N4	P_N4	L95_N4	U95_N4	SQT_S2
0.0023101	461172.62	0.002302	0.002302	INTERCEPT	0.5861296	8.0538E-7	0.4750518	0.6972073	0.0020825
				AGE	-0.270703	0.0011461	-0.401447	-0.139959	
				LFOLL	-0.007167	0.0013846	-0.010729	-0.003604	

5th iteration

BW_WLS
0.5861296
-0.270703
-0.007167

SIG2_5 SIG2_N5
0.0023003 0.0023003

SIG2_I	D_SIG_2	SIG2_5	SIG2_N5	B_LABEL	BW_N5	P_N5	L95_N5	U95_N5	SQT_S2
0.002302	463070.42	0.0023003	0.0023003	INTERCEPT	0.5861226	8.0417E-7	0.4750659	0.6971793	0.0020782
				AGE	-0.270697	0.001145	-0.401418	-0.139976	
				LFOLL	-0.007166	0.0013833	-0.010728	-0.003605	

6th iteration

BW_WLS
0.5861226
-0.270697
-0.007166

SIG2_6 SIG2_N6
0.0023 0.0023

SIG2_I	D_SIG_2	SIG2_6	SIG2_N6	B_LABEL	BW_N6	P_N6	L95_N6	U95_N6	SQT_S2
0.0023003	463473.78	0.0023	0.0023	INTERCEPT	0.5861212	8.0391E-7	0.4750689	0.6971734	0.0020773
				AGE	-0.270696	0.0011447	-0.401412	-0.139979	
				LFOLL	-0.007166	0.001383	-0.010728	-0.003605	

7th iteration

BW_WLS
0.5861212
-0.270696
-0.007166

SIG2_7 SIG2_N7
0.0022999 0.0022999

SIG2_I	D_SIG_2	SIG2_7	SIG2_N7	B_LABEL	BW_N7	P_N7	L95_N7	U95_N7	SQT_S2
0.0023	463559.5	0.0022999	0.0022999	INTERCEPT	0.5861209	8.0385E-7	0.4750696	0.6971721	0.0020771
				AGE	-0.270695	0.0011447	-0.401411	-0.13998	
				LFOLL	-0.007166	0.0013829	-0.010728	-0.003605	

8th iteration

BW_WLS
0.5861209
-0.270695
-0.007166

SIG2_8	SIG2_N8
0.0022999	0.0022999

SIG2_I	D_SIG_2	SIG2_8	SIG2_N8	B_LABEL	BW_N8	P_N8	L95_N8	U95_N8	SQT_S2
0.0022999	463577.71	0.0022999	0.0022999	INTERCEPT	0.5861208	8.0384E-7	0.4750697	0.6971719	0.0020771
				AGE	-0.270695	0.0011447	-0.40141	-0.13998	
				LFOLL	-0.007166	0.0013829	-0.010728	-0.003605	