META-ANALYSIS: EFFECT OF THE DRUG FOSAMAX ON OSTEOPOROSIS

META-ANALYSIS: THE EFFECT OF THE DRUG FOSAMAX ON BONE MINERAL DENSITY IN MULTI-DOSE AND MULTI-YEAR OSTEOPOROSIS CLINICAL STUDIES

By

FRANK CIHON, B.Sc.

A Project

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (1999) **McMASTER UNIVERSITY**
(Statistics) **Hamilton, Ontario**

TITLE: **Meta-Analysis: The Effect of the Drug Fosamax on Bone Mineral Density in Multi-Dose and Multi-Year Osteoporosis Clinical Studies**

AUTHOR: **Frank Cihon, B.Sc. (McMaster University)**

SUPERVISOR: **Dr. Andrew R. Willan**

NUMBER OF PAGES: xii, 62

ABSTRACT

Meta-analysis of clinical studies has reached a stage of general acceptance in both the statistical and medical fields. In the last two decades, meta-analysis has become an increasingly popular statistical procedure designed to integrate the findings of several independent studies that address a related set of research questions. The explosion of this relatively new scientific method for evaluating clinical medicine provides insight regarding the effect of a treatment so that inferences can be made that extend far beyond the findings of primary studies. Two models for study-to-study variation may be used in a meta-analysis: the fixed-effects model assumes that the population treatment effect is a single fixed value, and the random-effects model assumes that the population treatment effect is a randomly distributed variable with its own mean and variance. A fixed-effects approach exists for multi-dose and multi-time clinical studies whereby the correlations between treatment effect estimates resulting from the common control group between doses and years are taken into account.

The goal of this research was to conduct a meta-analysis to evaluate the effect of the drug FOSAMAX on bone mineral density in postmenopausal women with established osteoporosis. In particular, the fixed-effects approach was used to determine whether a dose-effect and/or time-effect existed across nine independent studies for the 4 anatomical sites: Lumbar Spine, Femoral Neck, Total Body and Trochanter. Statistical inferences about the

dose- and/or time-effect were guided by fitting and testing regression models for the estimated treatment effects. Lastly, the summary treatment effects and 95% confidence intervals were calculated for each dose and year for all 4 sites via the fixed-effects model.

The results of the meta-analysis indicate that a dose- and time-effect exists across the studies observed. By fitting and testing numerous regression models for the estimated treatment effects, it is concluded that the "full" model (a model with a parameter for each dose) is required to model the data adequately when tested with various models collapsed with respect to either dose or year of follow-up. By testing a dose-by-time interaction model with the full model for each of the 4 sites, the results indicate that terms in the interaction model are required to explain the data for the Lumbar Spine site, and thus, dose-effect depends on year.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Andrew R. Willan, my supervisor, for his patience, guidance and wisdom during the completion of this research. His help is greatly appreciated.

In addition, recognition must be given to the other members of my committee, Dr. Peter Macdonald and Dr. Roman Viveros-Aguilera. Special thanks for their insightful suggestions.

Finally, I would like to express my heartfelt thanks to Dr. Cheryl Pause for her patience, constant support and endurance throughout the completion of my graduate studies.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONT'D)

TABLE OF CONTENTS (CONT'D)

TABLE OF CONTENTS (CONT'D)

LIST OF TABLES

LIST OF TABLES (CONT'D)

LIST OF FIGURES

CHAPTER 1 : General Introduction to the Principles of Meta-Analysis in Clinical Trials

1.1 Introduction

Scientists acknowledge that any hope of combining empirical research for the purpose of constructing generalizations requires them to build upon previous work through integration, replication or reconceptualization. Statisticians are recognized as the first scientists to support alternative procedures for combining the results of individual studies. These procedures were labeled "meta-analysis" by Gene Glass in 1976 to refer to "the statistical analysis of a large collection of analysis results for the purpose of integrating the findings" (Cooper and Hedges, 1994a).

In recent years, meta-analysis, also referred to as quantitative literature review, research synthesis or research integration, has been increasingly used in the fields of medicine, social sciences and education. Statisticians have been using these procedures for about 100 years; from Fisher (1932), Pearson (1933) and Tippett (1931) who first tested statistical significance of combined results across studies to Birge (1932), Cochran and Yates (Cochran, 1937; Yates & Cochran, 1938) and Pearson (1904) who were among the first to suggest procedures for estimating the magnitude of treatment effects across studies (Cooper and Hedges, 1994a). A carefully planned meta-analysis allows for an objective appraisal of the data providing a more precise treatment effect estimate by finding, summarizing and describing integrated results of research. In addition, comparing results of research with

additional analyses may explain the heterogeneity between the findings of individual studies and advance the theoretical issues of causation regarding a specific research question. Consequently, through combination and comparison of already existing studies, a meta-analyst can extend his/her knowledge to make inferences that go far beyond the primary results.

1.2 Statistical Considerations for Conducting a Meta-Analytic Review

Statistical considerations have an immense influence at many stages in a meta-analytic review. Some of these considerations include issues on general conceptual content concerning research strategy or interpretation while others embrace statistical practices such as appropriate tests for specific hypotheses. In particular, the formulation of a problem for meta-analytic research has significant implications for suitable statistical procedures and for proper interpretation of the findings. One important aspect of problem formulation is to specify a representative population to which the generalizations will be made. A second intrinsic issue in problem formulation relates to the nature of the treatment effect parameter to be estimated. The treatment effect may be systematically affected by bias or some aspect of experimental procedure such as restriction of range or measurement error in the response variable.

The meta-analyst must distinguish between the two types of research questions that will be investigated in a meta-analytic review. The first concerns a hypothesis specifically stated in advance whereas the second refers to a hypothesis that is not specified clearly, followed by unplanned post hoc comparisons. In many research studies, numerous hypotheses are proposed a priori, and ambiguity arises as the meta-analyst attempts to distinguish between hypotheses selected by searching the data informally and proposing a hypothesis a

posteriori. A sensible solution is to treat all hypotheses as if they were post hoc (Hedges, 1994b). Thus, the hypotheses that are proposed a priori are more powerful than post hoc test procedures for determining specific relationships of interest. However, the latter can uncover relationships that were not suspected beforehand.

Data collection is a sampling activity designed to produce studies that are representative of the intended population and that would support any inferences made to that population. Some studies in the population with a full range of attributes may not have been conducted, thus, restricting the likelihood of creating generalizations.

Much of the scientific literature on meta-analysis provides statistical procedures for the analysis of a particular measure of treatment effect. The literature presents procedures for the analysis for combining estimates of odds ratios, relative risks, risk differences and standardized mean differences as being different. However, a set of underlying statistical theories exists and provides a common theoretical justification for analyses of the treatment effects mentioned. Analyses have commonly used statistical procedures that rely on two assumptions. The first being that the treatment effect estimate is distributed normally in large samples, and the second that the standard error of the treatment effect estimate is a continuous function of the within-study sample sizes, the treatment effect and any other parameters that are estimated from within-study data (Hedges, 1994b). The statistical procedures for different measures of treatment effect appear to be dissimilar as a result of particular formulas used in the calculation of treatment effect estimates and standard errors. The implementation of a generic treatment effect and standard error, T and S(T) respectively, and a corresponding treatment effect parameter θ allows statistical procedures to be applied

to any type of treatment effect estimate by substituting the appropriate formulas.

1.3 Objectives, Capabilities and Advancements in Medical Research

The application of meta-analysis in medical research has increased sharply since 1980. Numerous medical studies with comparable treatment protocols provide valuable data on efficacy of a particular treatment. However, the data from individual studies are limited in scope to make generalizations. By combining the results across studies, the evidence regarding treatment efficacy is strengthened. The objective of a meta-analysis includes increasing power to detect a more precise summary treatment effect and to explain heterogeneity between the findings of individual studies. A meta-analysis is capable of identifying study attributes associated with effective treatments, estimating the degree of benefit associated with a study treatment and assessing the amount of study-to-study variation. It also has the ability to provide preliminary data for hypotheses to be tested and for calculation of the sample size for a definitive large trial. Furthermore, meta-analysis is capable of recommending the most reliable treatment in the absence of a definitive trial and presents the assembly of valuable study evidence in a comprehensive format. Lastly, meta-analysis increases the power of detecting a true difference (Borzak and Ridker, 1995).

Efforts to combine results from separate studies has had a historical role in medical research that can be traced as far back as half a century ago. The first meta-analysis to evaluate the effect of a therapeutic intervention (interestingly, the treatment was placebo) was published in 1955 (Egger and Smith, 1997). More recently, the advancements of meta-analysis has made it a popular statistical technique in randomized clinical trial research, particularly in the fields of cardiovascular disease (Yusuf et al., 1985), perinatal care

(Chalmers et al., 1989) and oncology (Collaborative Group, 1988) to name a few. Thus, the future of meta-analysis carries considerable promise as a tool to assess whether an overall study result varies among subgroups - for instance, among older and younger patients, men and women or patients with varying degrees of severity of disease.

1.4 Special Statistical Complications Induced by Meta-Analysis

Although it has achieved a level of general acceptance in statistical and medical literature, the implementation of meta-analyses has given rise to numerous problems with regards to establishing its validity. Common problems from combining results of independent studies are rooted from the diverse nature of the studies, in terms of design and methods applied (DerSimonian and Laird, 1986). Other concerns include problems with regards to statistical inference which stem from the application of the technique and problems which arise by the need to update meta-analyses as new research is published. Problems such as carelessness in abstracting appropriate articles, bias on behalf of the meta-analyst, failure to investigate important covariates and exaggerations regarding the reliability and precision of the findings must be overcome to improve the validity of a meta-analysis (Bailar, 1997). Also contributing to the problems of meta-analysis are those randomized controlled studies that are better controlled than others and differing sample sizes and patient populations which yield a different level of sampling error (DerSimonian and Laird, 1986).

Meta-analysis relies heavily on observational (previously published) data rather than experimental data. Dependence of only published reports results in publication bias (Normand, 1999). It is documented that there is a tendency to selectively publish statistically significant results rather than those that are nonsignificant. Pooled results do not adjust for

biases present in primary studies, and furthermore, they may introduce new biases. New sources of bias may result from small studies or prematurely terminated studies which favor extreme findings. Finally, the numerous available statistical techniques for meta-analysis may yield varying summary estimates even when integrating the findings from the exact same studies (Chalmers, 1991).

1.5 Project Objectives and Layout

It is the goal of this research to examine the effect of the drug Fosamax (alendronate sodium) on bone mineral density in patients with the bone diminishing disease, osteoporosis. A meta-analysis will be conducted to analyze 9 independent studies with data measured at 4 different anatomical sites for the purpose of detecting whether a dose or duration of treatment (time) effect exists. Fixed- and random-effects models are considered, however, complications arise as a result of dealing with multi-dose and multi-year studies which yield correlated treatment effects.

Chapter 1 examines the general principles of meta-analysis in clinical trials. Chapter 2 consists of an overview of the standard statistical procedures for meta-analysis, elaborating on appropriate models, tests of significance and statistical techniques. Chapter 3 examines the statistical procedures for analyzing stochastically dependent treatment effects, focusing on correlated treatment effects from multi-dose and multi-time studies. Chapter 4 provides a description of the data analyzed and reports the meta-analytic results. Chapter 5 summarizes the results and provides direction for future development.

CHAPTER 2 : Overview of Techniques and Statistical Procedures for Meta-Analysis

2.1 Introduction

A treatment effect refers to the magnitude of an effect or the size of the relation between two variables (Cooper and Hedges, 1994b). Treatment effect estimates are best combined when they are homogeneous or comparable in magnitude. A heterogeneity test exists and it has the capability to detect whether treatment effect estimates are too dissimilar to combine by fixed-effects models. Variation in effects from study-to-study can be treated as fixed or random by applying either the fixed- or random-effects model.

2.2 Treatment Effect Estimates for Continuous Response Variables

For study-to-study comparisons to be possible, individual study results should be recorded in (or at least, transformable to) a standardized format. The standardized difference between two means (known as the effect size) is applied when the original studies compare the responses of two independent samples of patients to two treatments (for example, experimental (treatment) versus control) where the size of the difference is influenced by the underlying population value. For example, the effect size is defined as:

$$d_i = \frac{\overline{X}_i^t - \overline{X}_i^c}{s_i} \qquad (2\text{-}1)$$

where $\overline{X}_i^t$ is the average of the treatment group in the ith study, $\overline{X}_i^c$ is the average of the

control group in the ith study and s_i is the sample standard deviation of the control group (Shadish and Haddock, 1994).

2.3 Treatment Effect Estimates for Categorical Response Variables

When comparing studies with a categorical response variable (for example, disease versus no disease) the most widely used measures of treatment effect are the relative risk and the odds ratio. In clinical studies, relative risk is associated with the proportionate decrease in the event rate of the response variable as a result of the treatment. Within the ith study, let p_{ic} and p_{it} correspond to the observed rates of the undesirable response in the control and treatment groups, respectively. The relative risk is denoted by

$$r_i = \frac{p_{ic}}{p_{it}}. \tag{2-2}$$

The odds ratio is another commonly used measure of effect for describing the results of studies that employ categorical responses. Also referred to as the cross-product ratio, the odds ratio corresponds to the ratio of the odds between two groups. The odds ratio can assume values between 0 and ∞ and an odds ratio of 1 implies no difference between the two groups. An odds ratio of greater than 1 implies a larger rate of occurrence in the group represented in the numerator, whereas, an odds ratio of less than 1 indicates a lower rate of occurrence in the group represented in the numerator. For example, the odds ratio of the proportion of subjects in each condition of a study who have the disease can be calculated by the formula,

$$o_i = \frac{p_{ic}(1-p_{it})}{p_{it}(1-p_{ic})} \tag{2-3}$$

(Shadish and Haddock, 1994).

Despite the rate difference being the most elementary parameter for describing studies that report the proportion of the treatment group with a condition (a disease, treatment success, status as case or control) and the proportion of the comparison group with that condition, it is seldom used. For example, the rate difference, also referred to as the difference between proportions is given by the formula,

$$D_i = p_{ic} - p_{it}. \qquad (2\text{-}4)$$

A reason for its infrequent use is that the range of variation is limited by the magnitudes of p_{ic} and p_{it} for study *i* (Shadish and Haddock, 1994).

2.4 Standard Approaches to Missing Treatment Effect Estimates

One of the greatest concerns the practicing meta-analyst has in conducting a meta-analysis is that of missing data. Specifically, missing treatment effects introduce a serious dilemma since a statistical measure for the findings of the study is the heart of the procedures for a meta-analytic review. Data required for the calculation of the treatment effects may be missing (for example, means, standard deviations and statistical tests) because the author of the primary results did not provide the information considered important by the meta-analyst. The generalizability of the meta-analytic results are threatened by studies with missing treatment effects because they are not included in the summary treatment effect across studies. Vote-counting procedures can be applied to studies that supply information regarding the direction or statistical significance but lack the information needed to calculate the treatment effect (Pigott, 1994). However, it is capable of making a decision concerning a null

hypothesis for a series of studies based only on counts of significant and nonsignificant study results.

Alternative techniques for handling missing treatment effects exist. However, they have serious problems that restrict their applicability. Common approaches include setting the missing treatment effects equal to zero, omitting from the analysis studies with missing treatment effects, setting the missing treatment effects equal to the mean calculated from studies with treatment effects and making use of the information presented in a research report to determine a lower limit for the treatment effect. A combined procedure of sample treatment effects and vote counts appears to be the procedure of choice for estimating the population treatment effect. (Pigott, 1994). Once again, this procedure is only proposed when studies provide the information about the direction or statistical significance of the results and lack the necessary information to calculate the treatment effect.

2.5 Fixed-Effects Models in Meta-Analysis

The fixed-effects model in meta-analysis is defined as a statistical model that specifies that the studies under analysis are the entire population of studies. The uncertainty of the meta-analytic results is only influenced by the within-study sampling error (Cooper and Hedges, 1994b).

2.5.1 Appropriate Circumstances, Assumptions and Generalizations

A fixed-effects model is used in a meta-analysis if the variation in treatment effects between studies is deemed to be the result exclusively from sampling error. That is, the lone source of variation treated as random is the within-studies sampling variation. When the variation in treatment effects between studies is significantly greater than what you would

anticipate by sampling error alone, the excess variation may be accounted for by entering known study attributes in a regression model. If the excess variation in effects can be accounted for by a small number of simple study attributes, the fixed-effects model is used (Hedges, 1994a). The fixed-effects model is based on the assumption that the population treatment effect is a single fixed value. A fixed-effects model has the capability to make generalizations to a universe of studies with comparable study attributes. When comparing the fixed-effects model to the random-effects model, the generalizability is much less for the former model; however, it has greater statistical power. As a result, confidence intervals for summary treatment effects are narrower for fixed-effects models (Rosenthal, 1995). This model is also referred to as the conditional model because it holds fixed the study attributes possibly related to the treatment effect parameter.

2.5.2 The Fixed-Effects Model

Assume that the data to be combined originate from a series of K independent studies, where the ith study yields an observed treatment effect T_i, with population treatment effect θ_i and variance v_i. Accordingly, the combined data consists of K treatment effects $T_1,\ldots,T_K$ of parameters $\theta_1,\ldots,\theta_K$, and variances $v_1,\ldots,v_K$. A common treatment effect, $\theta_1 = \ldots = \theta_K = \theta$, is assumed under the fixed-effects model. Thus, the summary treatment effect over these studies is denoted by the formula,

$$\overline{T}_{\bullet} = \frac{\sum_{i=1}^{K} w_i T_i}{\sum_{i=1}^{K} w_i}, \qquad (2\text{-}5)$$

where w_i designates the weight of the ith study that minimizes the variance of $\overline{T}_{\bullet}$. The weights are inversely proportional to the conditional variance in each study,

$$w_i = \frac{1}{v_i}, \qquad (2\text{-}6)$$

and the conditional variance of $\overline{T}_{\bullet}$ is itself a function of the conditional variances of each treatment effect combined,

$$v_{\bullet} = \frac{1}{\sum_{i=1}^{K} (1/v_i)} \qquad (2\text{-}7)$$

(Shadish and Haddock, 1994).

2.5.3 Combining Treatment Effect Estimates

Formulas for conditional variances, v_i, are different for various measures of effect, however, they are all inversely proportional to within-study sample size (a smaller variance obtained from a larger sample results in a more accurate estimate of treatment effect). Consequently, treatment effects from studies with larger sample sizes are allocated larger weights. When the treatment effect to be combined is the standardized mean difference d_i in (2-1) (for which the relevant population parameter is δ) and the data are assumed normally distributed, the summary treatment effect is denoted by $\overline{d}_{\bullet}$ and the conditional

variance of d_i is estimated as

$$v_i = \frac{n_i^t + n_i^c}{n_i^t n_i^c} + \frac{d_i^2}{2(n_i^t + n_i^c)}. \qquad (2\text{-}8)$$

In the above equation, n_i^t represents the sample size in the treatment group of the ith study, n_i^c represents the sample size in the comparison group of the ith study, and d_i estimates the population parameter δ.

When combining differences between proportions as in (2-4) the summary treatment effect is denoted by $\overline{D}_\bullet$ and the conditional variance of D_i is estimated as

$$v_i = \frac{p_{ic}(1-p_{ic})}{n_i^c} + \frac{p_{it}(1-p_{it})}{n_i^t}. \qquad (2\text{-}9)$$

All terms in the above equation are defined as in sections 2.3 and 2.5.3.

When combining odds ratios the proper measure of association to be combined is the logs odds ratio l_i. The summary treatment effect is denoted by $\bar{l}_\bullet$ and the conditional variance of l_i is

$$v_i = \frac{1}{n_i^c p_{ic}(1-p_{ic})} + \frac{1}{n_i^t p_{it}(1-p_{it})} \qquad (2\text{-}10)$$

(Shadish and Haddock, 1994).

For combining relative risks as in (2-2), the summary treatment effect of r_i is denoted by $\bar{r}_\bullet$. The conditional variance of the log relative risk is estimated as

$$v_i = \frac{1-p_{ic}}{n_i^c p_{ic}} + \frac{1-p_{it}}{n_i^t p_{it}} \qquad (2\text{-}11)$$

(Fleiss, 1993).

2.6 Random-Effects Models in Meta-Analysis

The random-effects model in meta-analysis is defined as a statistical model in which both between-studies variation and within-study sampling error are involved in the evaluation of the uncertainty of the meta-analytic results (Cooper and Hedges, 1994b).

2.6.1 Appropriate Circumstances, Assumptions and Generalizations

A random-effects model is used if the variation in treatment effects between studies is believed to be the result of both the sampling error and the between-study variability in the population treatment effect. If the excess variation (variation in treatment effects between studies which is greater than what you would expect by sampling error alone) is too complicated to be accounted for by a small number of study attributes, the meta-analyst should use the random-effects model. The random-effects model is based on the assumption that the population treatment effect is a randomly distributed variable with its own mean and variance. The study sample is assumed to represent a hypothetical collection (or population) of studies. A random-effects model has the capability to make generalizations to a universe of diverse studies (a population of studies from which a study is drawn) (Hedges, 1994a). This model is also referred to as the unconditional model because it does not hold fixed the study attributes possibly related to the treatment effect parameter.

2.6.2 The Random-Effects Model

The population treatment effect, θ_i, under a random-effects model is random (as opposed to being fixed) and it has its own distribution. The summary treatment effect for the random-effects model is

$$\overline{T}_{\bullet} = \frac{\sum_{i=1}^{K} w_i^* T_i}{\sum_{i=1}^{K} w_i^*}. \qquad (2\text{-}12)$$

The weights are inversely proportional to the total variability of an observed treatment effect v_i^* which reflects the conditional variation of that treatment effect from each population θ_i and the random variation, σ_θ^2, of the individual θ_i from the mean population treatment effect:

$$w_i^* = \frac{1}{v_i^*} = \frac{1}{\sigma_\theta^2 + v_i}. \qquad (2\text{-}13)$$

In the above equation, σ_θ^2 refers to the between-studies variance, v_i refers to the within-study variance (or conditional variance of T_i) and v_i^* is the unconditional variance of an observed treatment effect T_i (Shadish and Haddock, 1994). The estimation of σ_θ^2 is discussed in the following section.

2.6.3 Estimation of the Variance (σ_θ^2)

If the meta-analyst proceeds with the random-effects model the initial undertaking must be to test whether the variance component is zero, and if it differs significantly from zero the magnitude of the variance component needs to be estimated. When a fixed-effects analysis has already been performed the significance of the variance component is realized. If the heterogeneity test statistic,

$$Q = \sum_{i=1}^{K} \left[(T_i - \overline{T}_\bullet)^2 / v_i \right], \qquad (2\text{-}14)$$

is rejected, the variance component differs significantly from zero and the final task is to estimate its magnitude. However, if the fixed-effects analysis was not performed, Q must be determined besides estimating the magnitude of the variance component. The heterogeneity test statistic in equation (2-14) is only introduced because it is used in the second method for estimating the variance component, but the heterogeneity test is explained in detail later in this chapter.

The variance component can be estimated by two methods. The first method starts with the unweighted sample estimate of the variance of the treatment effects, $T_1,....,T_K$, calculated by the formula

$$s^2(T) = \left[\sum_{i=1}^{K} T_i^2 - \left(\sum_{i=1}^{K} T_i \right)^2 / K \right] / (K-1). \qquad (2\text{-}15)$$

The unconditional variance expected to be associated with any particular measure of effect is denoted by the expected value of the equation in (2-15) and is

$$E[s^2(T)] = \sigma_\theta^2 + (1/K) \sum_{i=1}^{K} \sigma^2(T_i|\theta_i). \qquad (2\text{-}16)$$

By definition, the observed variance from $s^2(T)$ above is an unbiased estimate of $E[s^2(T)]$. Subject to the measure of effect being combined, the v_i from equations (2-8), (2-9), (2-10) or (2-11) are used to estimate $\sigma^2(T_i|\theta_i)$. Thus, the equation in (2-16) can be

solved with these estimates and the variance component estimate is obtained:

$$\hat{\sigma}_\theta^2 = s^2(T) - (1/K)\sum_{i=1}^{K} v_i \,. \qquad (2\text{-}17)$$

The second method used to obtain the variance component estimate originates with Q in equation (2-14) which is considered an estimate of the weighted sample estimate of the unconditional variance $\sigma^2(T_i)$. The expected value of Q is

$$E[Q] = c(\sigma_\theta^2) + (K-1) \qquad (2\text{-}18)$$

and c is denoted by

$$c = \sum_{i=1}^{K} w_i - \left[\sum_{i=1}^{K} w_i^2 / \sum_{i=1}^{K} w_i\right]. \qquad (2\text{-}19)$$

A second variance component estimator is obtained by substituting Q for its expectation and solving for σ_θ^2:

$$\hat{\sigma}_\theta^2 = [Q - (K-1)]/c \qquad (2\text{-}20)$$

(Shadish and Haddock, 1994).

2.6.4 Combining Treatment Effect Estimates

The conditional variance of d_i (the standardized mean difference) estimated by equation (2-8) allows the variance component $\hat{\sigma}_\theta^2$ to be estimated by solving equations (2-14) to (2-20). If Q in (2-14) is rejected the variance component differs significantly from zero and the summary treatment effect of the random effects model is calculated from

equation (2-12). The weights in (2-12) are calculated in a similar fashion to the weights of the fixed-effects model but the unconditional variance v_i^* is substituted in for v_i as

$$w_i^* = \frac{1}{v_i^*} = \frac{1}{\hat{\sigma}_\theta^2 + v_i} \qquad (2\text{-}21)$$

where v_i is defined as (2-8).

Similarly, when combining differences between proportions, log odds ratios and log relative risks, the computations follow (2-5), (2-6) and (2-7) except that the same unconditional variance estimate , $v_i^* = \hat{\sigma}_\theta^2 + v_i$, is used in (2-6) and (2-7) in place of the conditional variances outlined in the fixed-effects model. The variance component $\hat{\sigma}_\theta^2$ is yielded by (2-17) or (2-20) and v_i is defined as (2-9), (2-10) and (2-11), as appropriate. (Shadish and Haddock, 1994).

2.7 Regression Analysis for Testing the Moderating Effects of Study Attributes in Fixed- and Random-Effects Models

When heterogeneity in treatment effects exists, regression analysis may be appropriate to determine the relation between study attributes (covariates) and treatment effects and explain the excess variation. A few examples of study attributes include average age of patients, proportion of males, average length of exposure to a treatment and year of publication. Then the association between the estimated and true treatment effect is

$$T_i = \theta_i + e_i, \qquad (2\text{-}22)$$

where e_i (assumed to be statistically independent with mean zero and variance estimate, v_i) is the error with which T_i estimates θ_i. Suppose also that there are p known covariates considered to be related to the effect, then the fixed-effects regression model is given by

$$\theta_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} \qquad (2\text{-}23)$$

where θ_i is substituted into equation (2-22) (Hedges, 1994a). In the above equation, $X_{i1}, \ldots, X_{ip}$ are the p known covariates for the *i*th study hypothesized to predict the study treatment effect θ_i and $\beta_0, \beta_1, \ldots, \beta_p$ are the unknown regression coefficients.

The random-effects regression model is given by

$$\theta_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + u_i \qquad (2\text{-}24)$$

substituted into equation (2-22) where all terms are defined as above and u_i is the random effect of study *i*, the deviation of study *i*'s true treatment effect from the value predicted by the model. The regression model for the random-effects model defined by equations (2-12) and (2-13) in matrix form is

$$T = X\beta + u + e \qquad (2\text{-}25)$$

where X is the matrix containing the covariate data. The two components, $u + e$, in equation (2-25) constitute the error term where the vector u corresponds to the random effect across studies and is assumed to be $N(0, \sigma_\theta^2 I)$, whereas, the vector e represents the sampling error. When the random effects variance is null ($\sigma_\theta^2 = 0$) the results will

duplicate those obtained from the fixed-effects regression approach defined in equation (2-23). Otherwise, if $\sigma_\theta^2 > 0$, equation (2-25) corresponds to the random-effects regression model defined in (2-24) (Raudenbush, 1994). Accordingly, the variance of T_i is

$$v_i^* = Var(u_i + e_i) = \sigma_\theta^2 + v_i \qquad (2\text{-}26)$$

or in matrix form,

$$Var(u + e) = \sigma_\theta^2 I + V \qquad (2\text{-}27)$$

when controlling for the $X's$.

2.8 Tests on Heterogeneity of Treatment Effects

The goal of a meta-analysis is to compare and possibly combine treatment effects across associated studies. The heterogeneity test examines statistically the homogeneity of treatment effects and is the primary measure in determining whether the fixed- or random-effects model is more appropriate (Egger et al., 1997). To test the homogeneity of the treatment effects across studies, the appropriate hypothesis is

$$H_0: \theta_1 = \theta_2 = \ldots = \theta_K = \theta \quad \text{versus}$$

H_1: At least one θ_i is different.

Under H_0 for large samples, the test statistic is

$$Q = \sum_{i=1}^{K} [(T_i - \overline{T}_\bullet)^2 / (v_i)] \sim \chi_{K-1}^2 \qquad (2\text{-}28)$$

where all terms are defined in section 2.6.2. In general, if Q exceeds the 100(1-α)

percentile of χ^2_{K-1}, then H_0 is rejected and the observed variance in treatment effect is significantly greater than what would be expected by chance if the K studies did share a common treatment effect. Hence, in this case the random-effects model is advocated (Schmid et al., 1991). If the test produces homogeneous results and H_0 cannot be rejected then the variation in study treatment effects is assumed to be a result of sampling error, and the fixed-effects model is appropriate.

2.9 Significance Tests and Confidence Intervals

Under the fixed-effects model, the summary estimate of treatment effect from a series of K independent studies is given by

$$\overline{T}_\bullet = \frac{\sum_{i=1}^{K} w_i T_i}{\sum_{i=1}^{K} w_i} \qquad (2\text{-}29)$$

where w_i is defined in equation (2-6) and the conditional variance of $\overline{T}_\bullet$, $v_\bullet$,is defined in equation (2-7). The hypothesis that the population treatment effect θ is zero may be tested with a two-tailed test by computing the Z -statistic. So, to test

$$H_0: \theta = 0 \qquad \text{versus}$$

$$H_1: \theta \neq 0$$

the test statistic is calculated by

$$Z = \left|\overline{T}_\bullet\right| / (v_\bullet)^{1/2}. \qquad (2\text{-}30)$$

If the Z -statistic is greater than the appropriate two-tailed critical value of the standard normal distribution, then the null hypothesis is rejected and it is concluded that the population

treatment effect differs from zero. The $100(1-\alpha)$ percent confidence interval (θ_L, θ_U) for θ is given by

$$\theta_L = \overline{T}_\bullet - z_{\alpha/2}(v_\bullet)^{1/2}, \qquad \theta_U = \overline{T}_\bullet + z_{\alpha/2}(v_\bullet)^{1/2} \tag{2-31}$$

where $z_{\alpha/2}$ denotes the $100(1-\alpha/2)$th percentile of the standard normal distribution.

Under the random-effects model, the estimate of the summary treatment effect is given by

$$\overline{T}_\bullet = \frac{\sum_{i=1}^{K} w_i^* T_i}{\sum_{i=1}^{K} w_i^*} \tag{2-32}$$

where w_i^* is defined in equation (2-13) (Shadish and Haddock, 1994). The calculations of the Z-statistic and $100(1-\alpha)$ percent confidence interval for θ are comparable to that of the fixed-effects model, by substituting $v_\bullet^*$ for $v_\bullet$ where

$$v_\bullet^* = \frac{1}{\sum_{i=1}^{K} w_i^*}. \tag{2-33}$$

CHAPTER 3 : Statistical Approaches to Stochastically Dependent Treatment Effects via Meta-Analysis

3.1 Introduction

Most often, a meta-analysis deals with analyzing treatment effects from K independent studies where a single dose (treatment) is measured against a control (or a standard treatment). The analysis is considered to be valid due to the independence of the studies. However, sometimes studies are much more complicated, and the meta-analyst may want to measure more than one dose against a common control. Consequently, more than one dose versus control treatment effect will be estimated from each study and the observed treatment effects will be correlated because of the common control group. These types of studies are referred to as multi-dose studies.

For the analysis of multi-dose studies, additional information needs to be obtained from the studies to properly account for possible correlations between the observed treatment effects. The dependency between estimated treatment effects in multi-dose studies is a result of the use of a common control. In most cases, this additional information is all that is needed to calculate the correlations. However, when the studies themselves do not provide any extra information on the measures used, the correlations can be imputed from either published research or test manuals depending on the measure. (Gleser and Olkin, 1994).

3.2 Multi-Dose Studies

Multi-dose studies randomly allocate independent groups of subjects to one of p

doses or to a control group. However, not all studies report results for the same set of doses. It is more likely that an investigator has different research goals, so results for different subsets of doses are usually reported. For the *i*th study, let y_{i0m} denote an observation for the *m*th subject of the control group and let y_{ijm} denote an observation from the *m*th subject of the *j*th dose for $j \in P_i$, where P_i is the set of indices of doses in study i; $P = \bigcup_{i=1}^{K} P_i$, where *K* is the number of studies. Let the cardinality of P_i equal f_i and let $f = \sum_{i=1}^{K} f_i$. Let the cardinality of P be g. Therefore, there are g unique doses in the *K* studies. Assume that y_{ijm} is normally distributed with mean μ_{ij} and standard deviation σ_{ij}. The formula for the effect size for the *j*th dose is

$$\delta_{ij} = \frac{\mu_{ij} - \mu_{i0}}{\sigma_{i0}}, \; j = 1,\ldots,p \qquad (3\text{-}1)$$

and the observed effect size is

$$d_{ij} = \frac{\bar{y}_{ij} - \bar{y}_{i0}}{s_{i0}}, \; j = 1,\ldots,p \qquad (3\text{-}2)$$

where $\bar{y}_{ij}$ is the sample mean for the *j*th dose, $\bar{y}_{i0}$ is the sample mean for the control group and s_{i0} is the sample standard deviation for the control group. The observed effect sizes d_{ij} have $\bar{y}_{i0}$ and s_{i0} in common, and this leads to the correlation between any two effect sizes within a study. Note that for each study, n_{i0} and n_{ij} denote the sample size of the control

group and the jth dose, respectively. Under the assumption of homogeneity of variance, $\sigma_{i0}^2 = \sigma_{i1}^2 = \ldots = \sigma_{ip}^2$, the large-sample variance of d_{ij} is given by

$$\Psi_{ijj} = \frac{1}{n_{ij}} + \frac{1+\frac{1}{2}\delta_{ij}^2}{n_{i0}},\ j = 1,\ldots,p \qquad (3\text{-}3)$$

whereas, the large-sample covariance between the effect sizes d_{ij} and d_{ij^*} $(j \neq j^*)$ is

$$\Psi_{ijj^*} = \frac{1+\frac{1}{2}\delta_{ij}\delta_{ij^*}}{n_{i0}},\ j \neq j^*,\ j, j^* \in P_i. \qquad (3\text{-}4)$$

The variances and covariances are estimated by

$$\hat{\Psi}_{ijj} = \frac{1}{n_{ij}} + \frac{1+\frac{1}{2}d_{ij}^2}{n_{i0}},\ \hat{\Psi}_{ijj^*} = \frac{1+\frac{1}{2}d_{ij}d_{ij^*}}{n_{i0}},\ j \neq j^*,\ j, j^* \in P_i. \qquad (3\text{-}5)$$

Let $\hat{\Psi}_i = \left\{\hat{\Psi}_{ijj^*}\right\}$ be the variance-covariance matrix of vector $\left\{\mathrm{d}_j\right\}$ (Gleser and Olkin, 1994).

3.2.1 Regression Model

Let δ and d be the vectors $\left\{\delta_{ij}\right\}$ and $\left\{d_{ij}\right\}$, respectively. The dimension of δ and d is f. Statistical inferences about δ are performed by fitting linear regression models for $\mathrm{d} = \left\{d_{ij}\right\}$. Let $\hat{\psi}$ be the estimated variance-covariance matrix of d. $\hat{\psi}$ is block diagonal,

of dimension $f \times f$, where the blocks in $\hat{\psi}$ are the $\hat{\psi}_i$.

The form of the most general model for the vector $\mathbf{d}$ of treatment effect estimates is

$$\mathbf{d} = \delta + \mathbf{e} \tag{3-6}$$

where the vector $\mathbf{e}$ denotes the error with mean vector 0 and covariance matrix $\hat{\psi}$. Models of the regression form

$$\mathbf{d} = X\beta + \mathbf{e} \tag{3-7}$$

linearly relate the elements of δ to each other, or to covariates measured for the studies. Let the dimensions of X and β be $f \times q$ and q, respectively. The estimator for the vector β,

$$\hat{\beta} = (X' \hat{\psi}^{-1} X)^{-1} X' \hat{\psi}^{-1} \mathbf{d}, \tag{3-8}$$

is obtained by applying the method of least squares to the model in equation (3-7). Lastly, the estimated large-sample covariance matrix of the estimator in (3-8) is

$$\mathrm{Cov}(\hat{\beta}) = (X' \hat{\psi}^{-1} X)^{-1} \tag{3-9}$$

(Gleser and Olkin, 1994).

3.2.2 Significance Tests and Confidence Intervals

Since $\hat{\beta}$ is approximately normally distributed, a $100(1-\alpha)$ percent confidence interval for any linear combination $a'\beta = a_1\beta_1 + a_2\beta_2 + \ldots + a_q\beta_q$ of the components of the

vector β is

$$a'\hat{\beta} \pm z_{\alpha/2}(a'\mathrm{Cov}(\hat{\beta})a)^{1/2}. \qquad (3\text{-}10)$$

In equation (3-10), $z_{\alpha/2}$ denotes the $100(1-\alpha/2)$th percentile of the standard normal distribution. To obtain simultaneous $100(1-\alpha)$ percent confidence intervals for all linear combinations $a'\beta$ of β, $(\chi^2_q(\alpha))^{1/2}$ is substituted in for $z_{\alpha/2}$ where $\chi^2_q(\alpha)$ is the $100(1-\alpha)$th percentile of the chi-square distribution with q degrees of freedom, the dimension of vector β.

A level α test of goodness-of-fit for the regression model in equation (3-7) versus the general model in equation (3-6), is based on the test statistic

$$Q = \mathrm{d}'\hat{\psi}^{-1}\mathrm{d} - \hat{\beta}'(X'\hat{\psi}^{-1}X)\hat{\beta} \qquad (3\text{-}11)$$

and rejects the model in equation (3-7) when $Q \geq \chi^2_{f-q}(\alpha)$.

Let X_i be a $f_i * g$ matrix whose *k*th-*j*th element is 1 if the *k*th non-placebo arm in the *i*th study received dose j, 0 otherwise. Let $X_F = X_1 // X_2 // \ldots // X_K$, where $X_1 // X_2 = \begin{bmatrix} X_1 \\ \cdots \\ X_2 \end{bmatrix}$. Then the *j*th element of

$$\hat{\beta}_F = (X_F'\hat{\psi}^{-1}X_F)^{-1}X_F'\hat{\psi}^{-1}\mathrm{d} \qquad (3\text{-}12)$$

is the summary (over study) estimate of the treatment effect for the *j*th dose. Then

$$Q_F = \mathrm{d}'\hat{\psi}^{-1}\mathrm{d} - \hat{\beta}_F'(X_F'\hat{\psi}^{-1}X_F)\hat{\beta}_F, \qquad (3\text{-}13)$$

the test statistic of the full model, compared to $\chi^2_{f-g}(\alpha)$ provides a test of the hypothesis $H_0: \delta_{1j} = \delta_{2j} = \ldots = \delta_{Kj}$ for all j (H_0 is the hypothesis that the treatment effect for dose j is common across studies). Rejection of H_0 means that study-level factors need to be added to the regression or that a random-effects model is more appropriate.

Let X_N be the vector of 1's of dimension f. Then

$$\hat{\beta}_N = (X_N'\hat{\psi}^{-1}X_N)^{-1}X_N'\hat{\psi}^{-1}\mathrm{d} \qquad (3\text{-}14)$$

is the summary (over study and dose) of the treatment effect. Let

$$Q_N = \mathrm{d}'\hat{\psi}^{-1}\mathrm{d} - \hat{\beta}_N'(X_N'\hat{\psi}^{-1}X_N)\hat{\beta}_N \qquad (3\text{-}15)$$

be the test statistic of the null model. Then if $Q_\mathrm{d} = Q_N - Q_F$ exceeds $\chi^2_{g-1}(\alpha)$, there is evidence of a dose-effect.

3.2.3 Multi-Time Studies

In addition to measuring more than one dose against a common control, outcomes may have been measured at more than one time in a particular study. These types of studies are referred to as multi-dose and multi-time studies.

3.2.3.a Modifications Based on Correlated Treatment Effects

The formula for the calculation of the standardized mean difference for multi-dose and multi-time studies is slightly more complicated than equation (3-2). To account for the correlations between treatment effect estimates resulting from the common control group between doses and times, the large-sample variance and covariance matrix must be modified.

Let y_{ijtm} denote the observation of the mth subject on the jth arm ($j = 0$ for control group) for study i at time t. Then the formula for the sample estimator in equation (3-2) is

$$d_{ijt} = \frac{\bar{y}_{ijt} - \bar{y}_{i0t}}{s_{0t}} \tag{3-16}$$

where

$$\bar{y}_{ijt} = \frac{1}{n_{ijt}} \sum_{m=1}^{n_{ijt}} y_{ijtm}, \tag{3-17}$$

and n_{ijt} is the number of patients of the jth arm measured at time t and

$$s_{0t} = \frac{1}{n_{i0t} - 1} \sum_{m=1}^{n_{i0t}} (y_{i0tm} - y_{i0t})^2. \tag{3-18}$$

Let $c_{ijj^*tt^*}$ be the estimator of $E\left\{\left[d_{ijt} - E(d_{ijt})\right]\left[d_{ij^*t^*} - E(d_{ij^*t^*})\right]\right\}$. Then

$$c_{ijj^*tt^*} = \left\{ \frac{2\mathrm{I}(j = j^*)}{n_{ijt} + n_{ij^*t^*}} + \frac{2}{n_{i0t} + n_{i0t^*}} + \frac{d_{ijt} d_{ij^*t^*} \rho_{tt^*}}{n_{i0t} + n_{i0t^*}} \right\} \rho_{tt^*} \tag{3-19}$$

where I is the indicator function and ρ_{tt^*} is the correlation between y_{ijtm} and y_{ijt^*m}, assumed independent of study. If $j = j^*$ and $t = t^*$, then $c_{ijj^*tt^*}$ is the variance of d_{ijt}, otherwise, it is one of the covariances (Willan, 1999; Gleser and Olkin, 1994).

CHAPTER 4 : Results of Meta-Analysis

4.1 Introduction

A meta-analysis was conducted to evaluate the effect of the drug FOSAMAX on bone mineral density in postmenopausal women with the disease osteoporosis. In particular, it was of interest to determine whether a dose-effect and/or time-effect existed across the nine independent studies that were included in the meta-analysis. Subjects were treated with either placebo or a particular dose of FOSAMAX and bone mineral density measurements were taken at 4 different anatomical sites. Fixed- and random-effects models were considered, however, complications arose as a result of dealing with multi-dose and multi-year studies which yielded correlated treatment effects. Since no published random-effects approach for dealing with correlated treatment effects from multi-dose and multi-year studies exists, the meta-analysis was based on the fixed-effects approach. In addition to calculating summary treatment effects and 95% confidence intervals, regression models were constructed in which the independent variables in the models were the dose and year and the dependent variable was the treatment effect (standardized mean difference). Statistical inferences about the dose- and/or time-effect were guided by fitting and testing such regression models.

4.1.1 The Epidemiology of Osteoporosis

The skeleton provides support for muscles and organs, but it also serves as a depot for the body's calcium and other minerals. The skeleton holds 99% of the body's calcium and

the remaining 1% of calcium circulates in the blood and is necessary for critical bodily functions ranging from muscle contraction to nerve function to blood clotting. Approximately 60% to 80% of the strength of bone is accounted for by bone mineral density. Bone tissue is steadily broken down and reformed to allow for growth, repair of minor damage caused from everyday stress, and for the maintenance of a properly functioning body. This breakdown or resorption of the skeleton is done by cells called osteoclasts which dig holes into the bone, releasing calcium into the bloodstream necessary for crucial bodily functions. Cells known as osteoblasts then rebuild the skeleton. Osteoblasts fill the holes of the bone with collagen by laying down crystals of calcium and phosphorus. About 10% to 30% of the adult skeleton is remodeled in this way every year (National Osteoporosis Foundation, 1999).

Osteoporosis is a progressive disease of the skeleton in which the amount of calcium present in the bones slowly decreases to the point where the bone becomes brittle, In other words, bone resorption outpaces bone formation and the bone loses density. With aging, the decrease in vitamin D3 levels and estrogen deficiency aggravate the problem of calcium absorption that, in turn, stimulate bone resorption. As bones become brittle there is an increased risk of fracture. Typical fractures include those of the spine, hip, humerus and wrist. The disease is diagnosed by the discovery of low bone density, evidence of fracture on x-ray, a history of osteoporotic fractures, height loss or kyphosis, signifying vertebral (spinal) fracture. Risk factors that increase the likelihood of developing osteoporosis include sex, age, a thin, small framed body, early menopause, lack of calcium, race, lack of physical activity, heredity, cigarette smoking, alcohol and caffeine. Osteoporosis occurs in both males and females, but it primarily affects women as bone turnover increases due to the hormonal

changes of menopause. Osteoporotic fractures occur about four times as often in women than men. About one third of all women who have this bone-weakening condition suffer from more fractures which can lead to disability and the loss of independence for postmenopausal females (Pharmaceutical Information Associates, Ltd., 1995).

4.1.2 The Drug FOSAMAX (Alendronate Sodium)

The drug FOSAMAX (Alendronate Sodium) is manufactured by the pharmaceutical company Merck & Co. and is the first non-hormonal drug to treat osteoporosis in postmenopausal women. FOSAMAX is a member of a class of drugs known as bisphosphonates and takes effect as a specific inhibitor of osteoclast-mediated bone resorption without interfering with bone formation. Tablets are administered orally on a continuous basis in an extremely specific manner. Patients must take the pill on an empty stomach with a glass of water, cease eating or drinking for at least a half hour afterwards upon which they must consume an adequate amount of calcium and remain upright due to the possibility of gastroesophageal irritation. Patients are instructed to accurately follow the treatment regimen since the absorption of FOSAMAX is extremely limited (Merck & Co., 1997). Common drug-related side effects include abdominal and musculoskeletal pain with occasional reports of mild digestive disturbances such as nausea, heartburn and irritation or pain of the esophagus.

4.1.3 Effect of FOSAMAX on Bone Mineral Density

FOSAMAX reduces the activity of osteoclasts (cells that cause bone loss) and helps build healthy bone. At the cellular level, FOSAMAX prefers to localize directly on the osteoclasts, the sites of bone resorption. The osteoclasts adhere normally to the surface of the

bone but lack the ruffled edges that signify resorption. FOSAMAX inhibits bone resorption without interrupting osteoclast recruitment or attachment. Normal bone forms on top of the FOSAMAX which becomes pharmacologically inactive when embodied inside the bone matrix. The drug is administered on a continuous basis to halt osteoclast activity on new resorption surfaces. Thus, bone formation surpasses bone resorption at these remodelling sites, resulting in continuous gains in bone density (Merck & Co., 1997). FOSAMAX has been shown to build healthy bone at the spine, hip and other sites by approximately 10%, reduce the risk of spinal fractures by approximately 50% and reduce overall height loss (Lieberman, 1995).

4.2 Nature of the Data

The data set for the meta-analysis was created from publications and clinical study reports provided by Merck & Co. that have been submitted to the Food and Drug Administration. The studies include nine placebo-controlled, double-blind, randomized trials where bone mineral density data were measured for at least one year in postmenopausal women with osteoporosis. The trials provided data regarding study design (duration of trial), patient characteristics, anatomical site, treatment duration, dosage, mean change and standard deviations for bone mineral density and study sample sizes at baseline and years 1,2 and 3. The duration of all trials included in the meta-analysis was between one and three years. Subjects were treated with either a placebo or a particular dose of FOSAMAX (dose j, $j = 1,\ldots,6$; corresponding to 1 mg, 2.5 mg, 5 mg, 10 mg, 20 mg or 40 mg, respectively). Bone mineral density was measured at the following anatomical sites: Lumbar Spine (the part

of the back between the thorax and the pelvis), Femoral Neck (pertaining to the upper part of the femur, or the thigh), Total Body and Trochanter (either of the two processes, greater trochanter or lesser trochanter, below the neck of the femur where the former is a broad, flat process at the upper end of the lateral surface of the femur, to which several muscles are attached and the latter is a short conical process projecting medially from the lower part of the posterior border of the base of the neck of the femur). All trials were double-blind. The study characteristics of the trials are contained in Table 1.

TABLE 1 : Study Characteristics of FOSAMAX Trials Measuring Bone Mineral Density

Trial No.	Study Reference	Intervention	Study Duration	Sites Measured
1	**Adami,** 1995	FOSAMAX 10, 20 mg versus placebo	2 years	lumbar spine, femoral neck, total body, trochanter
2	**Bone,** 1997	FOSAMAX 1, 2.5, 5 mg or placebo	2 years	lumbar spine, femoral neck, total body, trochanter
3	**Chesnut,** 1995	FOSAMAX 5, 10, 20, 40 mg versus placebo	2 years	lumbar spine, femoral neck, total body, trochanter
5	**Hosking (EPIC),** 1998	FOSAMAX 2.5, 5 mg versus placebo	2 years	lumber spine, femoral neck, total body, trochanter
6.1	**Liberman,** 1995	FOSAMAX 5, 10, 20 mg versus placebo	3 years	lumber spine, femoral neck, total body, trochanter
6.2	**Liberman,** 1995	FOSAMAX 5, 10, 20 mg versus placebo	3 years	lumber spine, femoral neck, total body, trochanter
7	**McClung,** 1998	FOSAMAX 1, 5, 10, 20 mg versus placebo	3 years	lumber spine, femoral neck, total body, trochanter
8	**Pols (FOSIT),** 1998	FOSAMAX 10 mg versus placebo	1 year	lumber spine, femoral neck, trochanter
9	**Rosen,** Unpublished	FOSAMAX 10 mg versus placebo	2 years	lumber spine, femoral neck, trochanter

4.3 Meta-Analytic Results by Site (Lumbar Spine, Femoral Neck, Total Body and Trochanter)

Individual analyses were conducted for each anatomical site (Lumbar Spine, Femoral Neck, Total Body and Trochanter) using the percentage change in bone mineral density at years 1, 2 and 3. The meta-analytic results are presented below.

4.3.1 Fixed-Effects Model Results

As mentioned in section 2.5.1, the fixed-effects model is based on the assumption that the population treatment effect is a single fixed value. In the case of multi-dose and multi-year studies, equations (3-8) and (3-9) of section 3.2.1 were used to calculate the summary treatment effects over studies for a particular dose and year.

4.3.1.a Summary Treatment Effects and 95% Confidence Intervals according to Dose and Year

The summary treatment effects and 95% Confidence Intervals over studies for a particular dose and year for each of the 4 sites were calculated. The results are presented by site in tables 2, 3, 4 and 5.

TABLE 2 : Summary Treatment Effects and 95% Confidence Intervals according to Dose and Year for Lumbar Spine

	Dose (mg)	Summary Treatment Effects	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Year 1	1	0.35190	0.16985	0.53395
	2.5	0.97656	0.85581	1.09731
	5	1.24521	1.14174	1.34869
	10	1.29770	1.21416	1.38124
	20	1.39615	1.25988	1.53241
	40	1.21509	0.73430	1.69588
Year 2	1	0.19736	-0.05821	0.45294
	2.5	1.00453	0.88142	1.12764
	5	1.26949	1.16136	1.37763
	10	1.30435	1.17736	1.43135
	20	1.55994	1.40640	1.71348
	40	.	.	.
Year 3	1	0.47331	0.20871	0.73791
	2.5	.	.	.
	5	1.34738	1.16746	1.52731
	10	1.87272	1.67153	2.07390
	20	.	.	.
	40	.	.	.

TABLE 3 : Summary Treatment Effects and 95% Confidence Intervals according to Dose and Year for Femoral Neck

	Dose (mg)	Summary Treatment Effects	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Year 1	1	0.04036	-0.14381	0.22454
	2.5	0.31059	0.19632	0.42485
	5	0.41493	0.32137	0.50849
	10	0.54503	0.47013	0.61992
	20	0.52985	0.39984	0.65987
	40	0.62953	0.14670	1.11236
Year 2	1	0.28555	0.10164	0.46946
	2.5	0.54626	0.42972	0.66280
	5	0.66738	0.56995	0.76480
	10	0.77482	0.65495	0.89469
	20	0.84397	0.70183	0.98611
	40	.	.	.
Year 3	1	0.33503	0.06667	0.60339
	2.5	.	.	.
	5	0.82221	0.65266	0.99176
	10	1.14918	0.96826	1.33009
	20	.	.	.
	40	.	.	.

TABLE 4 : Summary Treatment Effects and 95% Confidence Intervals according to Dose and Year for Total Body

	Dose (mg)	Summary Treatment Effects	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Year 1	1	0.30273	0.07684	0.52862
	2.5	0.47565	0.35461	0.59669
	5	0.73862	0.62984	0.84739
	10	0.93953	0.76096	1.11810
	20	1.05321	0.87028	1.23614
	40	1.34264	0.69229	1.99299
Year 2	1	0.29574	0.06040	0.53108
	2.5	0.65100	0.52695	0.77504
	5	0.88091	0.76831	0.99351
	10	0.94182	0.76396	1.11968
	20	1.04731	0.86054	1.23408
	40	.	.	.
Year 3	1	0.31569	-0.01799	0.64937
	2.5	.	.	.
	5	0.71610	0.51750	0.91469
	10	1.05911	0.85043	1.26779
	20	.	.	.
	40	.	.	.

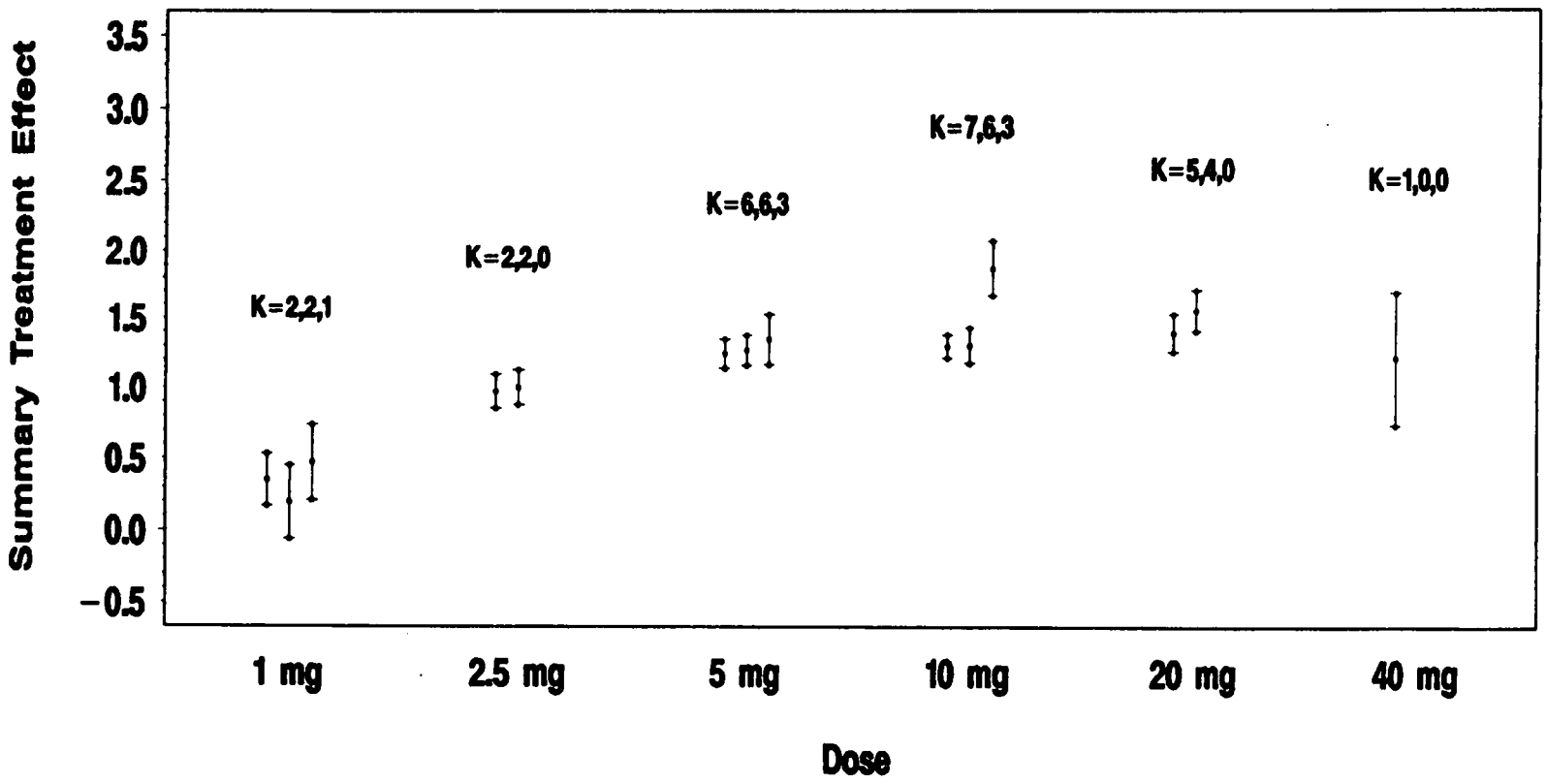
TABLE 5 : Summary Treatment Effects and 95% Confidence Intervals according to Dose and Year for Trochanter

	Dose (mg)	Summary Treatment Effects	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Year 1	1	0.08436	-0.10174	0.27047
	2.5	0.48192	0.36665	0.59719
	5	0.80182	0.70404	0.89959
	10	0.75814	0.68075	0.83553
	20	1.01806	0.87970	1.15642
	40	0.84835	0.36827	1.32842
Year 2	1	0.23296	0.04681	0.41911
	2.5	0.63111	0.51370	0.74853
	5	0.94070	0.83849	1.04291
	10	0.92119	0.79715	1.04523
	20	1.12458	0.97512	1.27404
	40	.	.	.
Year 3	1	0.45047	0.18121	0.71973
	2.5	.	.	.
	5	0.96255	0.78933	1.13578
	10	1.33789	1.15027	1.52551
	20	.	.	.
	40	.	.	.

The summary treatment effects and corresponding 95% confidence intervals for all 4 sites were plotted in Figures 1, 2, 3 and 4. All figures provide evidence of a dose- and time-effect. In Figure 1, it is evident that dose-effect depends on year for the Lumbar Spine site, particularly for the dose of 10 mg. For Figures 2, 3 and 4 corresponding to the Femoral Neck, Total Body and Trochanter sites, respectively, dose-effect does not depend on year and the effect of dose and time are additive.

Figure 1: LUMBAR SPINE, Years 1, 2 and 3

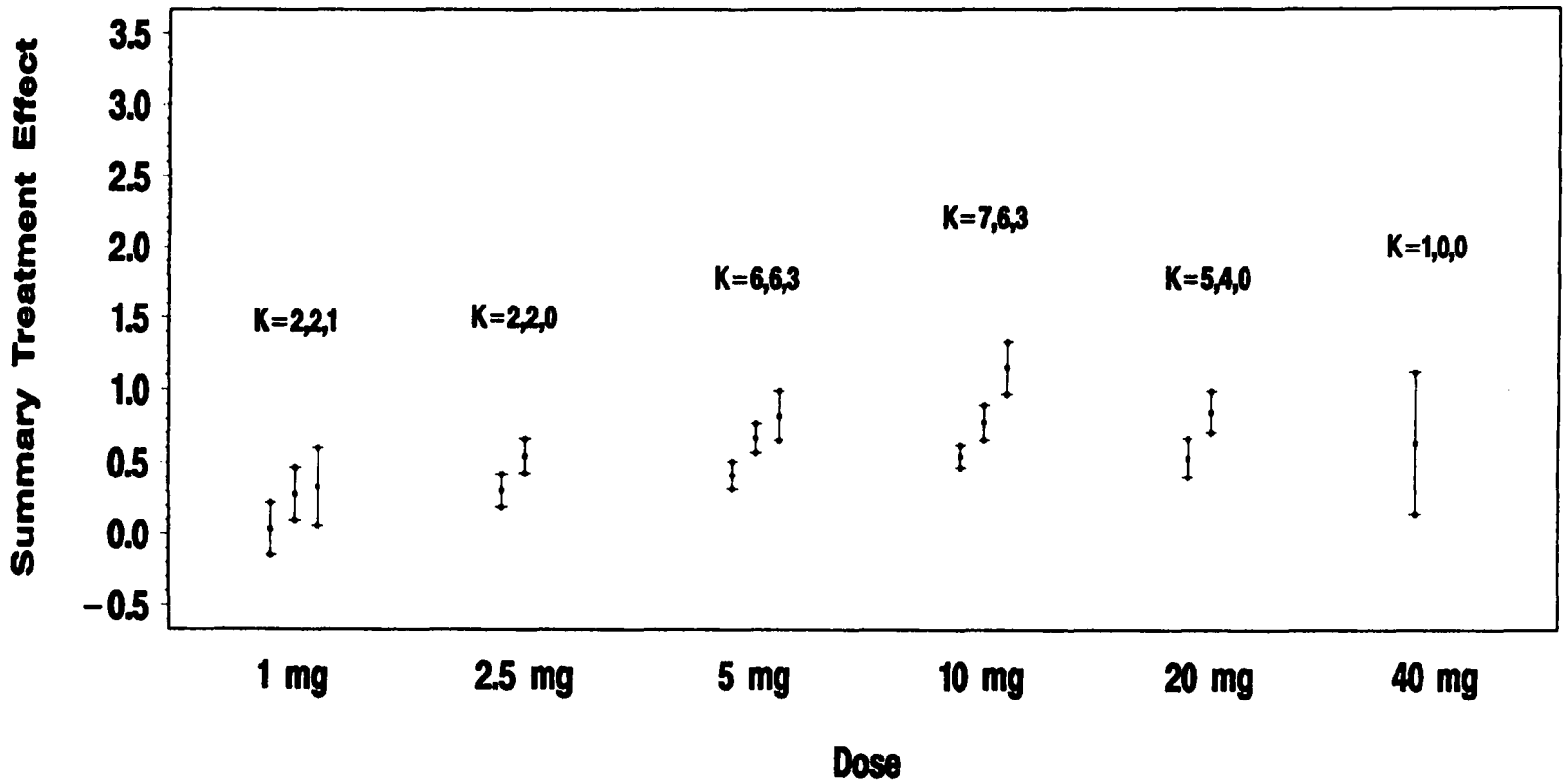
Fixed-Effects Model: Summary Treatment Effects and 95% CI



K=#1,#2,#3 #1-No. of studies in jth dose, year 1
#2-No. of studies in jth dose, year 2
#3-No. of studies in jth dose, year 3

Figure 2: FEMORAL NECK, Years 1, 2 and 3

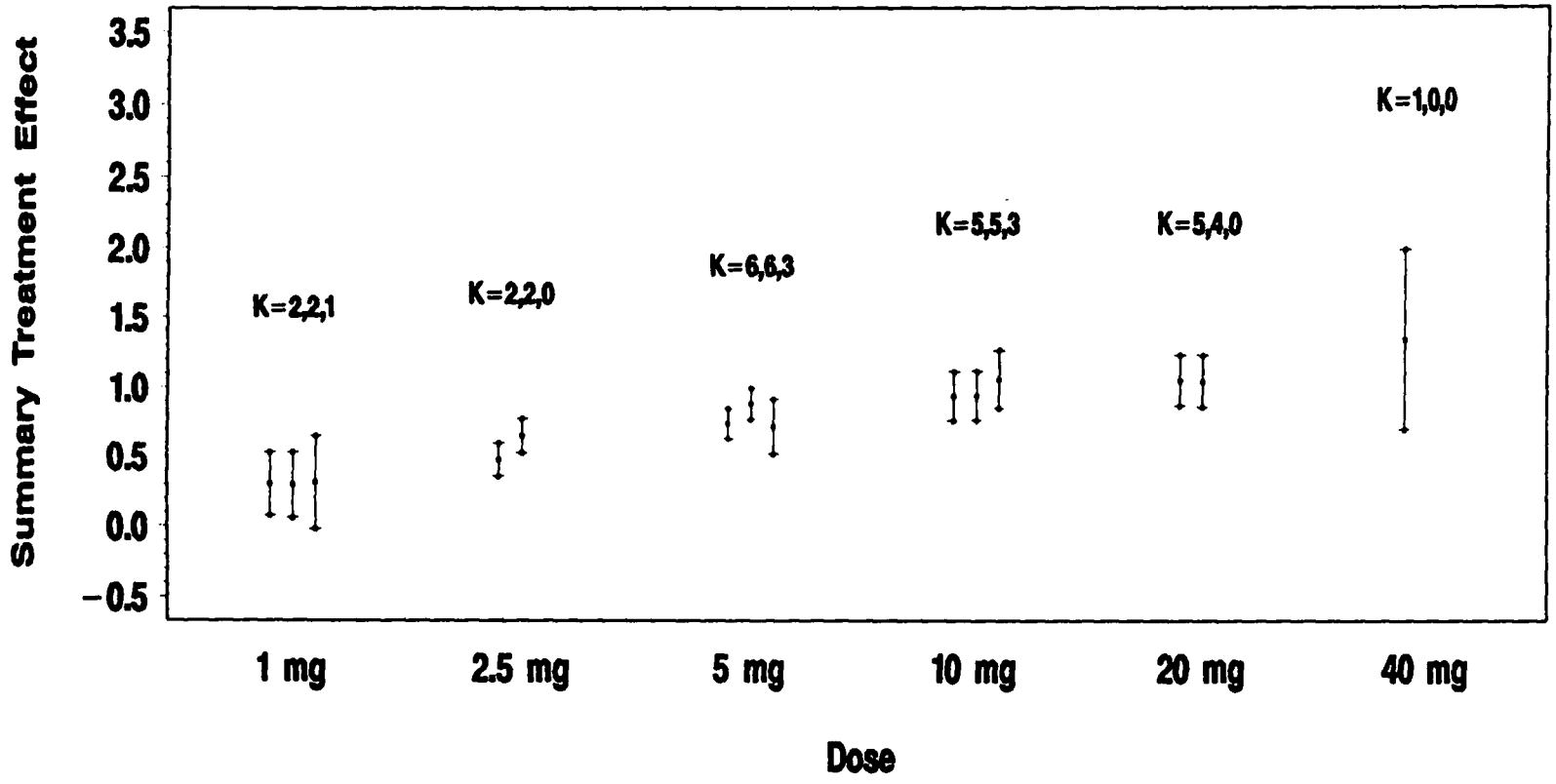
Fixed-Effects Model: Summary Treatment Effects and 95% CI



K=#1,#2,#3 #1-No. of studies in jth dose, year 1
#2-No. of studies in jth dose, year 2
#3-No. of studies in jth dose, year 3

Figure 3: TOTAL BODY, Years 1, 2 and 3

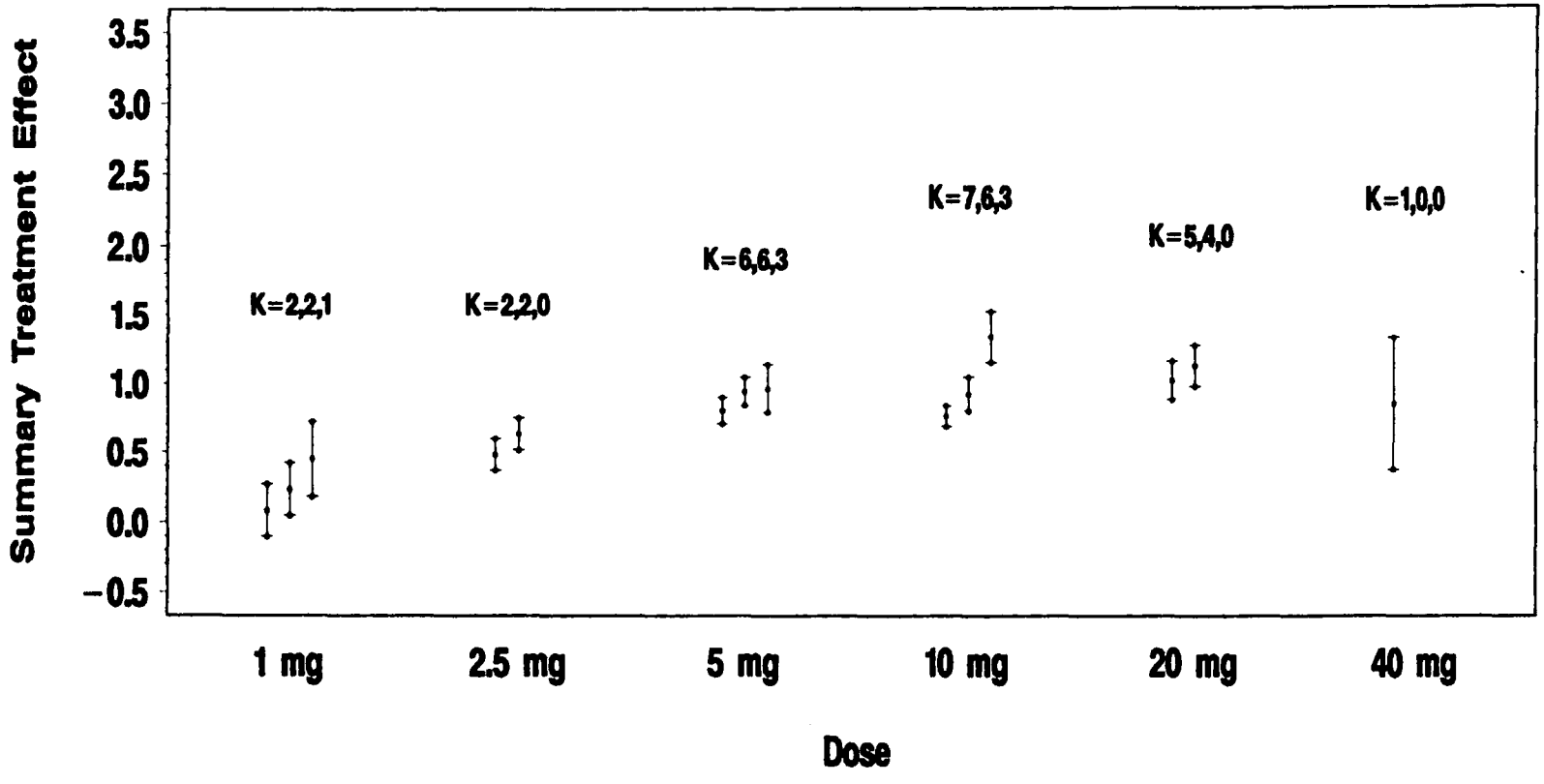
Fixed-Effects Model: Summary Treatment Effects and 95% CI



K=#1,#2,#3 #1-No. of studies in jth dose, year 1
#2-No. of studies in jth dose, year 2
#3-No. of studies in jth dose, year 3

Figure 4: TROCHANTER, Years 1, 2 and 3

Fixed-Effects Model: Summary Treatment Effects and 95% CI



K=#1,#2,#3 #1-No. of studies in jth dose, year 1
#2-No. of studies in jth dose, year 2
#3-No. of studies in jth dose, year 3

4.3.2 Dose-Effect (Years 1, 2 and 3 Individually and Combined)

Measuring more than one dose against a common control results in more than one dose versus control treatment effect being calculated for each study. To test for a dose-effect, $Q_N - Q_F$ needs to be calculated and compared to $\chi^2_{g-1}(\alpha)$ as discussed in section 3.2.2.

The quantities $Q_d = Q_N - Q_F$ are contained in table 6 for each anatomical site, each year and for years 1, 2 and 3 combined in the same model.

TABLE 6 : Dose-Effect (Years 1, 2 and 3 Individually and Combined) for all 4 Sites

	Doses Observed (mg)	Lumbar Spine	Femoral Neck	Total Body	Trochanter
Year 1	1, 2.5, 5, 10, 20, 40	Q_d =119.307 d.f.=5 P-value= 0	Q_d =32.2877 d.f.=5 P-value= 5.21E-6	Q_d =52.7686 d.f.= 5 P-value= 3.75E-10	Q_d =95.1827 d.f.= 5 P-value= 0
Year 2	1, 2.5, 5, 10, 20	Q_d =97.6639 d.f.=4 P-value= 0	Q_d =31.6493 d.f.=4 P-value= 2.26E-6	Q_d =40.4517 d.f.=4 P-value= 3.49E-8	Q_d =86.6941 d.f.=4 P-value= 0
Year 3	1, 5, 10	Q_d =98.3228 d.f.=2 P-value= 0	Q_d =35.4103 d.f.=2 P-value= 2.05E-8	Q_d =21.3268 d.f.=2 P-value= 0.00002	Q_d =42.4558 d.f.=2 P-value= 6.04E-10
Years 1, 2 and 3	1, 2.5, 5, 10, 20, 40	Q_d =187.096 d.f.=5 P-value= 0	Q_d =48.5526 d.f.=5 P-value= 2.74E-9	Q_d =54.7580 d.f.=5 P-value= 1.46E-10	Q_d =112.183 d.f.=5 P-value= 0

Since the Q_d is highly significant ($p < 0.05$) for each year and each site, there is evidence of a dose-effect. Q_d is also highly significant ($p < 0.05$) for years 1, 2 and 3 combined for all 4 sites. In the formulation of $\hat{\psi}$ from equation (3-19), from previous data, $\rho_{12} = \rho_{23} = 0.7$ and $\rho_{13} = 0.49$.

4.3.3 Low/High-Dose-Effect by Year (Years 1, 2 and 3 Individually)

Once a dose-effect was detected in section 4.3.2, it was of interest to investigate whether a low-dose, high-dose dichotomy would explain the dose-effect. The X_i would then be $f_i * 2$ where $\{X_i\}_{j1} = 1$ if the jth dose was 'low', 0 otherwise; and $\{X_i\}_{j2} = 1$ if the jth dose was 'high', 0 otherwise. Let Q_C be the corresponding lack-of-fit statistic of the collapsed model and let $Q_d = Q_C - Q_F$. The collapsed models consisted of different combinations of doses that were classified as either a low-dose or high-dose. The different combinations of low/high-dose groups are found in table 7.

TABLE 7 : Low/High-Dose Group Combinations

	Combinations of Low/High-Dose Groups				
	Comb. #1	**Comb. #2**	**Comb. #3**	**Comb. #4**	**Comb. #5**
Low	1 mg	1 and 2.5 mg	1, 2.5 and 5 mg	1, 2.5, 5 and 10 mg	1, 2.5, 5, 10 and 20 mg
High	2.5, 5, 10, 20 and 40 mg	5, 10, 20 and 40 mg	10, 20 and 40 mg	20 and 40 mg	40 mg

The quantities $Q_d = Q_C - Q_F$ are contained in table 8 for each year only for the Lumbar Spine site.

TABLE 8 : Low/High-Dose-Effect (Collapsed Model) Versus Dose-Effect (Full Model) by Year (Years 1, 2 and 3 Individually) for Lumbar Spine

	Collapsed Model (Comb. #1) vs. Full Model	**Collapsed Model (Comb. #2) vs. Full Model**	**Collapsed Model (Comb. #3) vs. Full Model**	**Collapsed Model (Comb. #4) vs. Full Model**	**Collapsed Model (Comb. #5) vs. Full Model**
Year 1	Q_d =31.031 d.f.=4 P-value= 3.02E-6	Q_d =44.835 d.f.=4 P-value= 4.30E-9	Q_d =93.126 d.f.=4 P-value= 0	Q_d =112.11 d.f.=4 P-value= 0	Q_d =119.20 d.f.=4 P-value= 0
Year 2	Q_d =40.175 d.f.=3 P-value= 9.78E-9	Q_d =52.928 d.f.=3 P-value= 1.90E-11	Q_d =83.852 d.f.=3 P-value= 0	Q_d =78.062 d.f.=3 P-value= 1.11E-16	Q_d =97.664 d.f.=4 P-value= 0
Year 3	Q_d =34.629 d.f.=1 P-value= 3.99E-9	Q_d =34.629 d.f.=1 P-value= 3.99E-9	Q_d =38.370 d.f.=1 P-value= 5.85E-10	Q_d =98.323 d.f.=2 P-value= 0	Q_d =98.323 d.f.=2 P-value= 0

Since all the Q_d for the Lumbar Spine site are highly significant ($p < 0.05$), it is concluded that the dose-effect cannot be explained by a low-dose, high-dose-effect.

4.3.4 Low/Medium/High-Dose-Effect by Year (Years 1, 2 and 3 Individually)

It was also of interest to investigate whether the dose-effect could be explained as a low/medium/high-dose-effect. The X_i would then be $f_i * 3$ where $\{X_i\}_{j1} = 1$ if the jth

dose was 'low', 0 otherwise; $\{X_i\}_{j2} = 1$ if the jth dose was 'medium', 0 otherwise; and $\{X_i\}_{j3} = 1$ if the jth dose was 'high', 0 otherwise. These collapsed models consisted of different combinations of doses that were classified as either low, medium or high. The different combinations of low/medium/high-dose groups are found in table 9.

TABLE 9 : Low/Medium/High-Dose Group Combinations

	Combinations of Low/Medium/High-Dose Groups		
	Comb. #1	**Comb. #2**	**Comb. #3**
Low	1 mg	1 and 2.5 mg	1, 2.5 and 5 mg
Medium	2.5, 5 and 10 mg	5 and 10 mg	10 mg
High	20 and 40 mg	20 and 40 mg	20 and 40 mg

The quantities $Q_d = Q_C - Q_F$ are contained in tables 10, 11, 12 and 13 for each year for all 4 anatomical sites.

TABLE 10 : Low/Medium/High-Dose-Effect (Collapsed Model) Versus Dose-Effect (Full Model) by Year (Years 1, 2 and 3 Individually) for Lumbar Spine

	Collapsed Model (Comb. #1) vs. Full Model	**Collapsed Model (Comb. #2) vs. Full Model**	**Collapsed Model (Comb. #3) vs. Full Model**
Year 1	Q_d =27.1335 d.f.=3 P-value= 5.52E-6	Q_d =40.6949 d.f.=3 P-value= 7.59E-9	Q_d =92.2917 d.f.=3 P-value=0
Year 2	Q_d =22.5817 d.f.=2 P-value=0.00001	Q_d =36.9676 d.f.=2 P-value=9.39E-9	Q_d =73.7311 d.f.=2 P-value=1.11E-16
Year 3	Q_d =34.6291 d.f.=1 P-value=3.99E-9	Q_d =34.6291 d.f.=1 P-value=3.99E-9	Q_d =38.3701 d.f.=1 P-value=5.85E-10

TABLE 11 : Low/Medium/High-Dose-Effect (Collapsed Model) Versus Dose-Effect (Full Model) by Year (Years 1, 2 and 3 Individually) for Femoral Neck

	Collapsed Model (Comb. #1) vs. Full Model	**Collapsed Model (Comb. #2) vs. Full Model**	**Collapsed Model (Comb. #3) vs. Full Model**
Year 1	Q_d =12.2970 d.f.=3 P-value= 0.00643	Q_d =13.2268 d.f.=3 P-value= 0.00417	Q_d =16.4076 d.f.=3 P-value=0.00094
Year 2	Q_d =8.4977 d.f.=2 P-value=0.01428	Q_d =9.8503 d.f.=2 P-value=0.00726	Q_d =17.7276 d.f.=2 P-value=0.00014
Year 3	Q_d =12.8452 d.f.=1 P-value=0.00034	Q_d =12.8452 d.f.=1 P-value=0.00034	Q_d =12.1454 d.f.=1 P-value=0.00049

TABLE 12 : Low/Medium/High-Dose-Effect (Collapsed Model) Versus Dose-Effect (Full Model) by Year (Years 1, 2 and 3 Individually) for Total Body

	Collapsed Model (Comb. #1) vs. Full Model	**Collapsed Model (Comb. #2) vs. Full Model**	**Collapsed Model (Comb. #3) vs. Full Model**
Year 1	Q_d =27.2557 d.f.=3 P-value= 5.20E-6	Q_d =8.1817 d.f.=3 P-value= 0.04240	Q_d =28.1515 d.f.=3 P-value=3.38E-6
Year 2	Q_d =15.6792 d.f.=2 P-value=0.00039	Q_d =8.7839 d.f.=2 P-value=0.01238	Q_d =31.6719 d.f.=2 P-value=1.33E-7
Year 3	Q_d =10.0732 d.f.=1 P-value=0.00150	Q_d =10.0732 d.f.=1 P-value=0.00150	Q_d =5.3468 d.f.=1 P-value=0.02076

TABLE 13 : Low/Medium/High-Dose-Effect (Collapsed Model) Versus Dose-Effect (Full Model) by Year (Years 1, 2 and 3 Individually) for Trochanter

	Collapsed Model (Comb. #1) vs. Full Model	**Collapsed Model (Comb. #2) vs. Full Model**	**Collapsed Model (Comb. #3) vs. Full Model**
Year 1	Q_d =31.3976 d.f.=3 P-value= 7.01E-7	Q_d =15.6233 d.f.=3 P-value= 0.00136	Q_d =70.0884 d.f.=3 P-value=4.11E-15
Year 2	Q_d =28.5922 d.f.=2 P-value=6.18E-7	Q_d =15.0233 d.f.=2 P-value=0.00055	Q_d =67.3570 d.f.=2 P-value=2.33E-15
Year 3	Q_d =16.7433 d.f.=1 P-value=0.00004	Q_d =16.7433 d.f.=1 P-value=0.00004	Q_d =13.4793 d.f.=1 P-value=0.00024

Since all the Q_d are highly significant ($p < 0.05$), it is concluded that the dose-effect cannot be explained by a low/medium/high-dose-effect.

4.3.5 Low/Medium/High-Dose-Effect for Years 1, 2 and 3 Combined

To determine if the dose-effect could be explained as low/medium/high when years 1, 2 and 3 are included, modifications were made to the variances and covariances as discussed in section 3.2.3.a. These collapsed models consisted of parameters for different combinations of grouped doses classified as either a low, medium or high-dose group. The design matrix is specified as

$\{X\}_{ij} = 1$ if the treatment effect associated with the *i*th row comes from the *j*th dose grouping, $j = 1, 2, 3$;

$= 1$ if the treatment effect associated with the *i*th row comes from year $j - 3$, $j = 4, 5$;

$= 0$ otherwise

where the *j*th dose grouping, $j = 1, 2, 3$, corresponds to the different combinations of low/medium/high-dose shown in table 9 of section 4.3.4. The lack-of-fit statistic, Q_C, was computed for models of different combinations of low/medium/high-dose groups at each site. Again, testing the collapsed models with the full model, the quantities $Q_d = Q_C - Q_F$ are contained in table 14 for each anatomical site.

TABLE 14 : Low/Medium/High-Dose-Effect (Collapsed Model) Versus Dose-Effect (Full Model) for Years 1, 2 and 3 Combined for all 4 Sites

	Lumbar Spine	**Femoral Neck**	**Total Body**	**Trochanter**
Collapsed Model (Comb. #1) vs. Full Model	Q_d =37.8046 d.f.=3 P-value= 3.11E-8	Q_d =16.1109 d.f.=3 P-value= 0.00108	Q_d =24.1491 d.f.=3 P-value= 0.00002	Q_d =35.5448 d.f.=3 P-value= 9.35E-8
Collapsed Model (Comb. #2) vs. Full Model	Q_d =70.6314 d.f.=3 P-value= 3.11E-15	Q_d =19.9486 d.f.=3 P-value= 0.00017	Q_d =10.1250 d.f.=3 P-value= 0.01753	Q_d =18.9294 d.f.=3 P-value= 0.00028
Collapsed Model (Comb. #3) vs. Full Model	Q_d =131.865 d.f.=3 P-value= 0	Q_d =25.1767 d.f.=3 P-value= 0.00001	Q_d =33.6735 d.f.=3 P-value= 2.32E-7	Q_d =81.1843 d.f.=3 P-value= 0

All the Q_d are highly significant ($p < 0.05$) and it is concluded that none of the collapsed models (low/medium/high-dose-effects) for any of the 4 sites explain the data as well as the full model. Hence, the dose-effect cannot be explained as low/medium/high when years 1, 2 and 3 are included.

4.3.6 Time-Effect for Years 1, 2 and 3

To investigate the time-effect, a similar regression approach from the above sections was applied in order to find the most parsimonious model by collapsing with respect to year of follow-up. In doing so, years 1 and 2 were combined and tested against year 3 and,

similarly, year 1 was tested against years 2 and 3 combined. The vector $\mathbf{d}$ and the variances and covariances of the collapsed models for different years grouped are defined identically as in the full model (for years 1, 2 and 3 combined in the same model) from section 4.3.2. The collapsed model consisted of a parameter for each dose as in the full model, but, in this case, only one parameter to indicate time. The design matrix is defined as

$\{X\}_{ij} = 1$ if the treatment effect associated with the *i*th row comes from the *j*th dose, $j = 1,\ldots,6$;

$= 1$ if the treatment effect associated with the *i*th row comes from year $j-6$, $j = 7$;

$= 0$ otherwise

where year $j-6$ corresponds to the yearly grouping combinations in table 15.

TABLE 15 : Yearly Grouping Combinations

Yearly Grouping Combinations	
Grouping #1	**Grouping #2**
Years 1 and 2 vs. Year 3	Year 1 vs. Years 2 and 3

The lack-of-fit statistic, Q_C, was computed for models of yearly grouping combinations for all 4 sites. Table 16 contains the quantities $Q_d = Q_C - Q_F$ for each anatomical site.

TABLE 16 : Dose-Effect for Different Combinations of Years Grouped (Collapsed Model) Versus Dose-Effect (Full Model) for all 4 Sites

	Lumbar Spine	**Femoral Neck**	**Total Body**	**Trochanter**
Collapsed Model (Grouping #1) vs. Full Model	Q_d =4.1931 d.f.=1 P-value= 0.04059	Q_d =68.0022 d.f.=1 P-value= 1.11E-16	Q_d =8.0120 d.f.=1 P-value= 0.00465	Q_d =24.9400 d.f.=1 P-value= 5.91E-7
Collapsed Model (Grouping #2) vs. Full Model	Q_d =19.4147 d.f.=1 P-value= 0.00001	Q_d =12.9175 d.f.=1 P-value= 0.00033	Q_d =7.2814 d.f.=1 P-value= 0.00697	Q_d =4.7686 d.f.=1 P-value= 0.02898

Hence, all of the collapsed models are rejected at the .05 level of significance. It is concluded that a time-effect exists and neither of the collapsed models (for yearly grouping combinations) explain the variability in the data as well as the full model.

4.3.7 Dose-by-Time Interaction Effect for Years 1, 2 and 3 Combined

Furthermore, it was of interest to determine if the dose-effect depended on the year (ie. was there a dose-by-time interaction). The vector $\mathbf{d}$ and the variances and covariances of this model are defined precisely as in the full model (for years 1, 2 and 3 combined in the same model) from section 4.3.2. The dose-by-time interaction model consisted of parameters corresponding to different dose-by-time combinations. The design matrix to test for a dose-by-time interaction is specified as

$$\{X\}_{ij} = 1 \quad \text{for interaction } j$$
$$= 0 \quad \text{otherwise}$$

where interaction j corresponds to the different dose-by-time combinations in table 17.

TABLE 17 : Dose-by-Time Combinations

j	**Dose-by-Time Combinations**	
	Dose in mg	**Time in Years**
1	1	1
2	2.5	1
3	5	1
4	10	1
5	20	1
6	40	1
7	1	2
8	2.5	2
9	5	2
10	10	2
11	20	2
12	1	3
13	5	3
14	10	3

Let Q_I be the corresponding lack-of-fit statistic and let $Q_d = Q_F - Q_I$. The test in this section was between the dose-by-time interaction model and the full model. The null hypothesis for this test was that there was no interaction between dose and time. Table 18

contains the quantities Q_d for each of the 4 sites.

TABLE 18 : Dose-Effect (Full Model) Versus Dose-by-Time Interaction Model for all 4 Sites

	Lumbar Spine	**Femoral Neck**	**Total Body**	**Trochanter**
Dose-Effect (Full Model) vs. Dose-by-Time Interaction Model	Q_d =34.2878 d.f.=6 P-value= 5.92E-6	Q_d =8.3032 d.f.=6 P-value= 0.21672	Q_d =9.8180 d.f.=6 P-value= 0.13253	Q_d =12.1036 d.f.=6 P-value= 0.05970

Thus, the null hypothesis is rejected at the .05 level of significance only for the Lumbar Spine site. However, there is marginal evidence of an interaction for the Trochanter site. It is concluded that there is no interaction for the sites, Femoral Neck, Total Body and Trochanter and this is supported by Figures 2, 3 and 4. The terms of the dose-by-time interaction model are required to explain the data for the Lumbar Spine site (refer to Figure 1). In other words, dose-effect depends on the year for the Lumbar Spine site.

CHAPTER 5 : Conclusion and Opportunities for Future Development

5.1 Summary of Meta-Analytic Results

The meta-analysis examining the effect of the drug FOSAMAX on bone mineral density in postmenopausal women with established osteoporosis provided additional information which was utilized to make inferences that extended far beyond the primary results. The meta-analysis determined that statistically significant heterogeneity ($p < 0.05$) between the treatment effects across doses and between years existed for all 4 sites. Once a dose-effect and a time-effect were detected, it was of interest to combine similar doses and years to evaluate whether collapsed models adequately explained the heterogeneity between treatment effects across doses and between years. In an attempt to arrive at the most parsimonious model, the full model (a model with a parameter for each dose) was tested with various models collapsed with respect to either dose or year of follow-up. By analyzing the results of these tests, it was concluded that the full model was needed to explain the data for all 4 sites (that is, the collapsed models did not explain the variability in the data as well as the full model).

Lastly, it was also of interest to test a dose-by-time interaction model with the full model from above for each of the 4 sites. This test determined whether the dose-effect depended on year. The results of these tests indicate that the terms of the dose-by-time interaction model are required to explain the data and that dose-effect depends on year only

for the Lumbar Spine site.

5.2 Statistical Issues for Future Research

Although meta-analysis has established itself as a popular systematic and important technique for making an objective appraisal regarding the effect of a treatment, problems and weaknesses have arisen with regards to its validity and these need to be dealt with both by discussion and by documentation of future research. Some of the shortcomings of meta-analysis that are of concern include study design differences that may exist or may not be recognized, heterogeneity between studies that is not adequately addressed and unknown sampling probabilities associated with the set of patients, settings, treatments, outcomes and times entering a meta-analysis.

Other weaknesses of meta-analysis that must be improved in future research include publication bias (the favoring of positive findings), selection of studies that influence the results, small studies and prematurely terminated studies that tend toward extreme findings and the ability to recognize risk or harm of treatment less uniformly assessed than primary endpoints (Smith and Egger, 1998). Furthermore, future documentation relating to techniques for determining the reliability and for estimating missing treatment effects of primary studies can provide support for the validity of this relatively new and maturing scientific method. Specifically, for multi-dose and multi-time studies, the development of a random-effects approach would allow both between-studies variation and within-study sampling error to be involved in the evaluation of the uncertainty of the meta-analytic findings. Therefore, meta-analysis may still be improved, by a combination of both experience and theory, to the point at which its generalizations about the effect of a treatment are considered reliable when no

other analysis or confirmation is available.

REFERENCES

Bailar, J. C. 'The Promise and Problems of Meta-Analysis', [editorial] *The New England Journal of Medicine*, **337**, 559-561 (1997).

Birge, R. T. 'The Calculation of Errors by the Method of Least Squares', *Physical Review*, **40**, 207-227 (1932).

Borzak, S., Ridker, P. M. 'Discordance Between Meta-Analyses and Large-Scale Randomized, Controlled Trials', *Annals of Internal Medicine*, **123**, 873-877 (1995).

Chalmers, T. C. 'Problems Induced by Meta-Analysis', *Statistics in Medicine*, **10**, 971-980 (1991).

Chalmers, I., Enkin, M., Keirse, M. *Effective Care during Pregnancy and Childbirth*, Oxford University Press, Oxford (1989).

Cochran, W. G. 'Problems Arising in the Analysis of a Series of Similar Experiments', *Journal of the Royal Statistical Society*, Supplement **4(1)**, 102-118 (1937).

Cooper, H., Hedges, L. V. 'Research Synthesis as a Scientific Enterprise', in *The Handbook of Research Synthesis*, Cooper, H., Hedges, L. V. Russell Sage Foundation, New York, 3-14 (1994a).

Cooper, H., Hedges, L. V. *The Handbook of Research Synthesis*, Russell Sage Foundation, New York (1994b).

DerSimonian, R., Laird, N. 'Meta-Analysis in Clinical Trials', *Controlled Clinical Trials*, **7**, 177-188 (1986).

Early Breast Cancer Trialists' Collaborative Group. 'Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women', *New England Journal of Medicine*, **319**, 1681-1692 (1988).

Egger, M., Smith, G. D. 'Meta-Analysis: Potentials and Promise', *BMJ*, **315 (7119)**, 1371-1374 (1997).

REFERENCES (CONT'D)

Egger, M., Smith, G. D., Phillips, A. N. 'Meta-Analysis: Principles and Procedures', *BMJ*, **315 (7121)**, 1533-1537 (1997).

Fisher, R. A. *Statistical Methods for Research Workers* (4th ed.), Oliver & Boyd, London (1932).

Fleiss, J.L. 'The Statistical Basis of Meta-Analysis', *Statistical Methods in Medical Research*, **2**, 121-145 (1993).

Gleser, L. J., Olkin, I. 'Stochastically Dependent Effect Sizes', in *The Handbook of Research Synthesis*, Cooper, H., Hedges, L. V. Russell Sage Foundation, New York, 339-355 (1994).

Hedges, L. V. 'Fixed Effects Models', in *The Handbook of Research Synthesis*, Cooper, H., Hedges, L. V. Russell Sage Foundation, New York, 285-299 (1994a).

Hedges, L. V. 'Statistical Considerations', in *The Handbook of Research Synthesis*, Cooper, H., Hedges, L. V. Russell Sage Foundation, New York, 29-38 (1994b).

Lieberman, W. Sr. et al. 'Effect of Oral Alendronate on Bone Mineral Density and the Incidence of Fractures in Post-Menopausal Osteoporosis', New England Journal of Medicine, **333**, 1437-1443 (1995).

Merck & Co. 'Fosamax (Alendronate Sodium Tablets)', Internet Document, 1-20 (1997).

National Osteoporosis Foundation 'Fighting Osteoporosis & Promoting Bone Health', Internet Document, (1999).

Normand, S. T. 'Tutorial in Biostatistics, Meta-Analysis: Formulating, Evaluating, Combining, and Reporting', *Statistics in Medicine*, **18**, 321-359 (1999).

Pearson, K. 'Report on Certain Enteric Fever Inoculation Statistics', *British Medical Journal*, **2**, 1243-1246 (1904).

Pearson, K. 'On a Method of Determining Whether a Sample of Size n Supposed to have been Drawn from a Parent Population having a known Probability Integral has Probably been Drawn at Random', *Biometrika*, **25**, 379-410 (1933).

Pharmaceutical Information Associates, Ltd. 'FDA Advisory Committee Recommends Alendronate for Osteoporosis', Medical Sciences Bulletin (Internet Document), 1-3 (1995).

REFERENCES (CONT'D)

Pigott, T. D. 'Methods for Handling Missing Data in Research Synthesis', *in The Handbook of Research Synthesis*, Cooper, H., Hedges, L. V. Russell Sage Foundation, New York, 163-175 (1994).

Raudenbush, S. W. 'Random Effects Models', in *The Handbook of Research Synthesis*, Cooper, H., Hedges, L. V. Russell Sage Foundation, New York, 301-321 (1994).

Rosenthal, R. 'Writing Meta-Analytic Reviews', *Psychological Bulletin*, **118**, 183-192 (1995).

Schmid, J. E., Koch, G. G., LaVange, L. M. 'An Overview of Statistical Issues and Methods of Meta-Analysis', *Journal of Biopharmaceutical Statistics*, **1(1)**, 103-120 (1991).

Shadish, W. R., Haddock, C. K. 'Combining Estimates of Effect Size', *in The Handbook of Research Synthesis*, Cooper, H., Hedges, L. V. Russell Sage Foundation, New York, 261-281 (1994).

Smith, G. D., Egger, M. 'Meta-Analysis: Unresolved Issues and Future Developments', *BMJ*, **316**, 221-225 (1998).

Tippett, L. H. C. *The Methods of Statistics*, Williams and Norgate, London (1931).

Willan, A. R. 'Personal Communication', Department of Clinical Epidemiology & Biostatistics, McMaster University (1999).

Yates, F., Cochran, W. G. 'The Analysis of Groups of Experiments', *Journal of Agricultural Science*, **28**, 556-580 (1938).

Yusuf, S., Peto, R., Lewis, J., Collins, R., Sleight, P. 'Beta Blockade During and After Myocardial Infarction: An Overview of the Randomized Trials', *Progress in Cardiovascular Diseases*, **27**, 335-371 (1985).