Loyola University Chicago
## Loyola eCommons

1988

# A Meta-Analysis of Eighteen Clinical Studies Related to the Comparative Nephrotoxicity of Gentamicin and Tobramycin

J Russell Teagarden
*Loyola University Chicago*

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Part of the Education Commons

## Recommended Citation

A META-ANALYSIS OF EIGHTEEN CLINICAL STUDIES

RELATED TO THE COMPARATIVE NEPHROTOXICITY

OF GENTAMICIN AND TOBRAMYCIN

by

J. Russell Teagarden

A Thesis Submitted to the Faculty of the Graduate School

of Loyola University of Chicago in Partial Fulfillment

of the Requirements for the Degree of

Master of Arts

November

1988

## ACKNOWLEDGEMENTS

The assistance in formulating and conducting this research provided by the thesis committee director, Dr. Jack Kavanaugh, was greatly appreciated. His guidance in helping me comply with thesis requirements unrelated to the research was also invaluable. I am also grateful for the input received from the other committee members, Ms. Michelle Parker, and Dr. Ronald Morgan. I am particularly grateful for all the thesis committee members' assistance that was often requested during evening hours or on weekends.

This research was heavily dependent on various microcomputer-based applications. I am indebted to Dr. and Mrs. Richard E. Derman for providing this equipment.

Finally, completion of this research project would not have been possible without the patience, sacrifice, support, and encouragement of my wife, Celine Marie, and our two children, Alexis Leigh, and Benedict Jamison. It's good to be back home again.

VITA

The author, John Russell Teagarden, is the son of John Herbert and Lois (Heiden) Teagarden. He was born April 26, 1955, in San Francisco, California.

His elementary education was obtained in the public school systems of suburban San Francisco and Chicago. His secondary education was completed in June, 1973 at John Hersey High School in Arlington Heights, Illinois.

In September, 1973, Mr. Teagarden enrolled in Harper Community College in Palatine, Illinois, and completed the one-year pre-pharmacy curriculum in June, 1974. He then entered the University of Illinois College of Pharmacy and earned the degree of Bachelor of Science in Pharmacy in June, 1978. In July, 1978, he entered the Residency in Hospital Pharmacy program at Northwestern Memorial Hospital in Chicago and received a certificate of completion in June, 1979.

In July, 1979, Mr. Teagarden accepted a position as staff pharmacist in the intensive care areas of Northwestern Memorial Hospital in Chicago. In October, 1983, he accepted the position of Drug Information Specialist at Saint Joseph Hospital in Chicago and remained until May, 1985 when he became Technical Support Manager at Baxter Healthcare Corporation in Deerfield, Illinois. In February, 1987, Mr. Teagarden returned to Saint Joseph Hospital as Drug Information Specialist.

Mr. Teagarden is a member of the American Society of Hospital Pharmacists, Illinois Council of Hospital Pharmacists, Northern Illinois Society of Hospital Pharmacists, Drug Information Association,

and Northeastern Illinois Chapter of the American Statistical
Association. He serves as a reviewer for Hospital Formulary and Drug
Interaction Facts, and is an advisory board member of the Iowa Drug
Information Service and Curaflex Health Services.

Since 1983, Mr. Teagarden has consistently contributed to the
published medical literature. He has authored or coauthored
"Intravenous nitroglycerin for the treatment of angina at rest
unresponsive to standard nitrate therapy", published in the American
Journal of Cardiology in 1983; "Online literature retrieval", published
in the Illinois Pharmacist, in 1984; "Association of methemoglobinemia
and intravenous nitroglycerin administration", published in the
American Journal of Cardiology in 1985; "Glyburide and glipizide:
clinical profile of two new oral hypoglycemic agents", published in
Hospital Formulary in 1986; "Therapeutic Role of zidovudine in the
treatment of acquired immune deficiency syndrome", published in
The Consultant Pharmacist in 1987; and "Stability of procainamide
hydrochloride in neutralized 5% dextrose injection", published in the
American Journal of Hospital Pharmacy in 1988. He also coauthored
"Factors associated with the time to arrhythmic recurrence in survivors
of ventricular fibrillation", and "Predictors of arrhythmic recurrence
in survivors of ventricular fibrillation rendered free of inducible
ventricular tachycardia", both of which were presented at the 36th
Annual Scientific Session of the American College of Cardiology, in New
Orleans, Louisiana, in March, 1987.

Mr. Teagarden enrolled in the Graduate School at Loyola
University of Chicago in September, 1984. He received the degree of

Master of Arts in research methodology in January, 1989.   In May 1988,

he was inducted into the Jesuit Honor Society, Alpha Sigma Nu.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Aminoglycoside antibiotics are important in the treatment of serious to life-threatening bacterial infections. These antibiotics are used extensively as evidenced by the approximately 3 million doses given annually in the United States alone. However, the use of aminoglycoside antibiotics is often hampered by associated toxicities. One of the most common toxicities encountered with these antibiotics affects the kidneys (nephrotoxicity). Although aminoglycoside nephrotoxicity is usually reversible and mild in severity, it can result in some degree of morbidity and lengthen hospital stay. If gone undetected, however, aminoglycoside nephrotoxicity can progress to irreversible renal (kidney) failure and condemn a patient to dialysis (Cooper & Bennett, 1987).

Considerable research has been directed at determining the interrelationships between aminoglycoside antibiotics and nephrotoxicity. Gentamicin and tobramycin specifically, have drawn a great deal of attention, particularly after early animal data indicated that tobramycin might be associated with less nephrotoxicity than gentamicin (Kahlmeter & Dahlager, 1982). Comparative clinical trials have produced discrepant results as to whether gentamicin is associated with nephrotoxicity more often than tobramycin (Burkle, 1986). Results from the same investigators have even been discrepant (Smith, Lipsky, Laskin, Hellmann, Mellitis, Longstreth, & Lietman, 1980; Moore, Smith, Lipsky, Mellits, & Lietman, 1984).

Several attempts have been made to discern from the empirical research whether there is a difference in the incidence of

1

nephrotoxicity between gentamicin and tobramycin (Burkle 1986; Cone, 1982; Darr & Elenbaas, 1981; Hubler, 1984; Kahlmeter & Dahlager, 1982; Meyer, 1986; Smith & Lietman, 1982). However, the conclusions presented in these reviews have been as discrepant as the independent empirical (i.e. primary) findings.

The question of gentamicin and tobramycin comparative nephrotoxicity is important because if tobramycin is associated with less nephrotoxicity than gentamicin, morbidity related to gentamicin nephrotoxicity could be reduced by preferentially using tobramycin. However, since gentamicin is significantly less expensive to use than tobramycin, if there is no difference in nephrotoxicity between them, then by using gentamicin preferentially, financial resources that would have been consumed by tobramycin use could be reallocated for other purposes.

Given that which is reported above, it is apparent that the previously published reviews of the empirical research comparing gentamicin and tobramycin nephrotoxicity yielded inconsistent results. It is important to note that meta-analytical techniques were never systematically applied to the existing database. In the study discussed below, meta-analytical techniques were used in an attempt to provide a better understanding of the comparative nephrotoxicity of gentamicin and tobramycin than achieved by the previously published narrative reviews.

Meta-analysis represents a group of methodologies that are used to systematically and quantitatively combine results of individual empirical research efforts to derive conclusions that may not be

achievable otherwise. Meta-analyses are distinguished from narrative reviews by their quantitative nature. The procedure has been criticized because of the heterogeneity that may exist among the results and methods that are combined. Although there are methods to control for the possible heterogeneity across studies, meta-analytic synthesis of research findings will never take the place of a well-done, definitive study. Meta-analytic procedures are perhaps best reserved for situations where definitive studies are not logistically possible, or as an exploratory activity to determine whether such a study should be undertaken (Mintz, 1983).

Considering the importance of the comparative nephrotoxicity of gentamicin and tobramycin, and the discrepancy that presently exists in both the primary and secondary literature, a meta-analysis might provide a better overall picture. Thus, a meta-analysis of the comparative nephrotoxicity of gentamicin and tobramycin was undertaken to primarily determine in a quantitative fashion whether such a difference exists, and if so, to what degree.

A parametric meta-analytic procedure (standardized mean differences) was used to detect and quantify any differences in the comparative nephrotoxicity of gentamicin and tobramycin. However, not all comparative studies of gentamicin and tobramycin nephrotoxicity provided enough information to apply the parametric procedures. Therefore, a modified vote-counting method was used to analyze those studies that could not be analyzed by the parametric procedures. Thus, a secondary purpose of this research project was to compare these two meta-analytical techniques.

## Comparative Nephrotoxicity of Gentamicin and Tobramycin

Aminoglycoside antibiotics. Aminoglycoside antibiotics are a group of antibiotics that share similar chemical structures and properties. Many of the aminoglycoside antibiotics are commonly used in the treatment of serious to life-threatening bacterial infections. In some cases they represent the most effective or the only effective antibiotics available (Pancoast, 1988).

The first aminoglycoside antibiotic made available for general clinical use in the United States was streptomycin in 1944. The next aminoglycoside antibiotic to be approved for use was kanamycin in 1957, followed by gentamicin in 1969, tobramycin in 1975, amikacin in 1976, and netilmicin in 1983. The aminoglycosides antibiotics have seen extensive use with approximately three million doses administered annually in the United States (Pancoast, 1988).

The use of aminoglycoside antibiotics, however, is hampered by their associated toxicity. The most common toxicities encountered with the use of aminoglycoside antibiotics are ototoxicity and nephrotoxicity. Ototoxicity refers to toxicity affecting auditory function and nephrotoxicity refers to toxicity affecting kidney function (Pancoast, 1988). Nephrotoxicity, specifically, has been the subject of significant research and debate. Part of the research and debate has concerned the relative nephrotoxicity of one aminoglycoside to another, particularly gentamicin and tobramycin (Kahlmeter & Dahlager, 1984).

Aminoglycoside nephrotoxicity. Nephrotoxicity occurs in

4

approximately 10-20% of aminoglycoside courses of therapy.
Aminoglycoside antibiotics are taken up into renal tubular cells;
however, the cellular mechanism of toxicity is not known. The
clinical presentation of aminoglycoside nephrotoxicity is usually an
asymptomatic accumulation in the serum of measurable metabolic products
that are normally excreted by the kidneys (Cooper & Bennett, 1987).
For example, creatinine, which is a metabolic product of muscle, is
produced at a relatively constant rate and is excreted by the kidney.
Therefore, as renal function decreases (as occurs secondary to
nephrotoxicity), excretion of creatinine decreases correspondingly and
accumulates in the serum (Ravel, 1978). Other manifestations of
nephrotoxicity can include detection of various enzymes or proteins in
the urine (Schentag, 1983).

Aminoglycoside nephrotoxicity typically occurs within seven to 10
days after initiation of therapy and is usually reversible with
discontinuation. Aminoglycoside nephrotoxicity that goes undetected
can progress to severe degrees ultimately requiring dialysis. Risk
factors that have been associated with aminoglycoside nephrotoxicity
include age, aminoglycoside dose, duration of therapy, recent
aminoglycoside exposure, preexisting renal dysfunction, concurrent
administration of other nephrotoxins, potassium depletion, and
intravascular volume depletion (Cooper & Bennett, 1987). The degree
to which specific aminoglycoside antibiotics contribute to the risk of
nephrotoxicity has been the subject of considerable debate.

Gentamicin and tobramycin nephrotoxicity. Early animal data
suggesting that tobramycin might be less nephrotoxic than gentamicin

resulted in subsequent clinical trials (Cooper & Bennett, 1987). The importance of determining the comparative nephrotoxicity of gentamicin and tobramycin encompasses both clinical and economic considerations. From a purely clinical perspective, even the slightest suggestion that tobramycin is less nephrotoxic than gentamicin would lead many clinicians to use tobramycin to minimize any undue morbidity related to gentamicin. However, economic considerations cloud the decision because tobramycin is several times more expensive to use than gentamicin. If there is no difference in the degree of nephrotoxicity associated with gentamicin and tobramycin, then use of gentamicin would permit reallocation of the financial resources necessary for tobramycin to other uses.

Published comparative studies of gentamicin and tobramycin nephrotoxicity have produced equivocal results; some studies showing tobramycin to be less nephrotoxic than gentamicin while others showed no difference. Many authors (Burkle, 1986; Cone, 1982; Darr & Elenbaas, 1981; Hubler, 1984; Kahlmeter & Dahlager, 1982; Meyer, 1986; Smith & Lietman, 1982) have attempted to evaluate the empiric research. These evaluations were either narrative reports with subjective conclusions or analyses of pooled data; none of which used recognized meta-analytical techniques. Like the empiric research they reviewed, these evaluations produced equivocal conclusions.

Burkle (1986), Darr and Elenbaas (1981), Hubler (1984), Meyer (1986), and Smith and Lietman (1982) each reported the results of published comparisons of gentamicin and tobramycin nephrotoxicity, cited methodological and clinical considerations, and rendered

subjective conclusions. Burkle evaluated 12 comparative trials and concluded "that these 12 clinical trials failed to demonstrate any difference in nephrotoxicity between these agents" (p. 516). Hubler reached a similar conclusion after evaluating 15 comparative trials, stating " the results of controlled studies in humans suggest that there are no marked clinical differences in the nephrotoxicity of gentamicin and tobramycin" (p. 3), as did Meyer in stating that "it is still too risky to conclude definitely that one agent is significantly less nephrotoxic that another and that controversy still abounds" (p. 126). In contrast, after evaluating approximately the same published database, Darr and Elenbaas concluded "that tobramycin has less nephrotoxic potential than does gentamicin" (p. 325) and Smith and Lietman concluded "tobramycin causes nephrotoxicity less frequently than does gentamicin" (p. 507).

Cone (1982) and Kahlmeter and Dahlager (1982) attempted quantitative analyses of the comparative studies of gentamicin and tobramycin nephrotoxicity. Cone pooled the results as reported in selected comparative studies and conducted pairwise comparisons (chi-square) to test for statistical significance. The difference between gentamicin and tobramycin nephrotoxicity did not reach statistical significance. Similarly, Kahlmeter and Dahlager pooled the results as reported from selected comparative studies of gentamicin and tobramycin nephrotoxicity. However, the pooled proportions of gentamicin and tobramycin nephrotoxicity (14% versus 12.9%) were not subjected to hypothesis testing.

In summary, despite several attempts to determine the comparative

nephrotoxicity of gentamicin and tobramycin by summarizing published results, the question of comparable nephrotoxicity still remains. Applying meta-analytical techniques to this database could provide more meaningful information than the previously published reviews to help solve this important question.

Meta-Analysis

Definition and characterization. Definitions and characterizations of "meta-analysis" vary because meta-analysis as a research methodology is relatively new and is still evolving (Mintz, 1983). The beginning of meta-analysis as a distinct methodological entity has been traced to Glass in 1976 (Mintz, 1983; Thacker, 1988); however, research techniques associated with meta-analysis had been employed prior to 1976 (Glass, McGaw, & Smith, 1981; Leviton & Cook, 1981; Sacks, Berrier, Reitman, Ancona-Berk, & Chalmers, 1987).

Glass (1976) originally defined meta-analysis as "the statistical analysis of a large collection of results from individual studies for the purpose of integrating the findings" (p. 3). Later, Glass, McGaw and Smith (1981) defined meta-analysis as "the analysis of analyses (i.e., the statistical analysis of the findings of many individual analyses)" (p. 12). Other definitions of meta-analysis are similar. Mintz (1983) defined meta-analysis as "a quantitative methodology for integrating empirical research literature" (p. 71). Meta-analysis is defined by Thacker (1988) as "an attempt to improve traditional methods of narrative review by systematically aggregating information and quantifying its impact" (p. 1658), and by L'Abbe, Detsky and O'Rourke (1987) as "the process of combining study results that can be used to

draw conclusions about therapeutic effectiveness or plan new studies"
(p. 224). Thus, most authors define meta-analysis as a method or as
methods to combine empirical (i.e., primary) research for the purpose
of deriving or improving generalizations.

Glass et al. (1981) have characterized meta-analysis as a method
by which quantitative analyses of empirical research are conducted by
adopting an "attitude" of data analysis (i.e., using measurement and
statistical analysis techniques). It can be considered as a method of
summarizing an accumulated knowledge and highlighting important
aspects (Thacker, 1988). Meta-analysis also addresses research
questions that remain unresolved because (a) empirical data are in
disagreement as to the direction or magnitude of an effect, (b) sample
sizes used in the primary research were too small to detect an effect,
or (c) the large trials necessary are not logistically feasible (L'Abbe
et al. 1987). In contrast to traditional narrative reviews in which
typically there are no rules by which the reviewer assesses the
relevant primary research, meta-analysis requires systematic approaches
to aggregating empirical information and quantifying its effect to
produce more valid generalizations (Fiske, 1983; Thacker, 1988).

Common to the definitions and characterizations of meta-analysis
is the "quantitative" nature of the methods used to review empirical
research, particularly relative to the traditional "narrative"
methods. The degree to which meta-analysis "quantifies" empiric
research is variable and often limited. Mintz (1983) conceptualized
the review process on a continuum based on the degree to which
quantitative methods are used as follows:

> As the review process progresses from the descriptive narrative
> summary to the abstract heights of the multiple regression
> analysis, a series of steps is taken by the reviewer. Each step
> involves increased quantification and abstraction. (p. 71)

Thus, on one end of the continuum are narrative reviews in which

quantitative integration of empiric research is absent and

subjectivity reigns. The next step along the continuum finds

narrative reviews that include tabular or graphical presentations of

the empiric research substrate that invite summarization but do not

integrate individual findings or synthesize new information. The next

step crosses into meta-analytic methodology wherein coding schemes are

used to facilitate descriptive summaries of the empiric research. The

complexity of methodologies continue to increase along the continuum to

ultimately "the introduction of inferential statistical hypothesis

testing" (p. 72).

Although quantitative aspects are emphasized when defining or

characterizing meta-analysis, there are necessary qualitative aspects

as well. For example, qualitative judgments in a particular

meta-analysis could include the population of studies considered

relevant, the scope of the empiric research substrate to analyze, and

the methodological approaches to employ (Leviton & Cook, 1981). Thus,

"just as quantitative research presupposes qualitative judgments, so

qualitative research is impossible without quantitative estimates" (p.

232).

Nomenclature. "Meta-analysis" was the term used by Glass in 1976

to denote methods by which empirical research is integrated to

emphasize or synthesize information from large bodies of data.

However, like the methodologies of meta-analysis, the nomenclature

remains unsettled (Light, 1987).

Meta-analysis is a term used frequently in both the social (Glass et al., 1981) and medical sciences (L'Abbe et al., 1987). "Research integration" and "research synthesis" are also terms that have been used with some regularity, while other authors prefer the term "overview" (Light, 1987). Presently none of these terms refer to specific types of methods to integrate empiric research, and are therefore used interchangeably. Terms used exclusively in the physical sciences for meta-analysis are "critical review" and "critical evaluation" (Hedges, 1987).

Need for meta-analysis. One model for scientific research specifies two components. The first component is empiric research from which primary data are derived. The second component is integration and interpretation of the results of empiric research. Meta-analysis serves as one methodological approach to this second component of scientific research of integration and interpretation (Fiske, 1983).

The need for the second component of scientific research in clinical medicine relates to the variability in results that occur despite the use of controlled methods in empiric research such as the randomized controlled trial. Horwitz (1987) enunciated the problem as follows:

> Clinical medicine is awash in controversy. At every level of clinical practice today, from prevention of the chronic diseases of aging such as cancer, to the treatment of acute disorders such as myocardial infarction, the evaluation and application of medical therapies is assailed by disagreement and uncertainty. In contemplating the health hazards of such diverse entities as tampons (and the alleged risk of toxic shock syndrome) or aspirin (and the alleged risk of Reye's Syndrome), the methodologic strategies and details of the research are frequently challenged and criticized, creating controversy and dissent in the

scientific literature and the public press. (p. 91)

Reasons for the state of the medical literature as Horowitz describes it have been attributed to specific methodological errors or problems inherent in the research paradigm itself. Specific methodological errors can include experimental designs that result in bias or statistical shortcomings such as insufficient sample sizes or employment of inappropriate analytical techniques. Inherent problems in research paradigms often relate not to compliance with the componentry of the paradigm, but with the variable interpretations of their use and applications (Horowitz, 1983).

The narrative review has been the predominant method by which empiric research has been assessed in the clinical medicine literature. Prior to the acceptance and application of controlled methods of experimentation in clinical research, the form of the published medical literature was primarily reports of random observations. Thus, there was little primary research that could be integrated; and therefore, the narrative review served to describe the state of the art (Fye, 1987).

As controlled methods of experimentation in clinical research became widely applied, mostly in the form of randomized controlled trials, narrative reviews became less reliable as a means to summarize the empiric data. The burgeoning size of the medical literature, as a result of specialization and the pressures to publish, also add to the inadequacy of narrative reviews to accurately summarize primary data (Fye, 1987). From January 1, 1984 to August 1, 1986 alone, approximately 6,000 randomized clinical trials were indexed in MEDLINE

(Chalmers, Levin, Sacks, Reitman, Berrier, & Nagalingam, 1987b). Thus, as the complexity and amount of the empiric research continue to increase, the chance for misinterpretation and bias will increase accordingly (Einarson, McGhan, Bootman, & Sabers, 1985; Strube & Hartmann, 1983; Thacker, 1987).

It could be argued that what is needed are not methods to integrate empirical research, but empiric research that is conducted to definitely answer the research questions at hand. However, logistical considerations often preclude design of the definitive empiric research effort, particularly those that will likely only demonstrate small to moderate magnitudes in effect that necessarily require large sample populations difficult to assemble (Collins, Gray, Godwin, & Peto, 1987). The inherent nature of the randomized clinical trial paradigm also often produces divergent results from seemingly identical methods (Horowitz, 1987). Thus, as Fiske (1983) noted:

> In the long-range perspective, no one study makes much difference (except the rare one that falls more in the context of discovery by uncovering something previously undemonstrated). Granted that the single study may stimulate or irritate in a healthy fashion, only the distillations from the entire body of research in an area have lasting effects. (p. 65)

The meta-analysis of Yusef, Collins, Peto, Furberg, Stampfer, Goldhaber, and Hennekens (1985) assessing the effect of intravenous fibrinolytic therapy for acute myocardial infarction is an illustrative example. Prior to their meta-analysis, the place of fibrinolytic therapy for acute myocardial infarction had been uncertain despite the publication of over 20 clinical trials over a period of 25 years. Of the 24 randomized clinical trials of intravenous fibrinolytic agents for acute myocardial infarction included in the meta-analysis, only

five suggested any benefit from this therapeutic intervention in terms of mortality. However, Yusef et al. derived an overall reduction of mortality of approximately 22% using the odds ratio of mortality in the fibrinolytic groups to mortality in the control groups. The overall odds ratio was derived by weighting individual study odds ratios inversely by variance. It thus appeared that intravenous fibrinolytic therapy for acute myocardial infarction could affect a mortality benefit, but that the magnitude of effect might be moderate thereby necessitating large sample sizes for reliable detection. As a result of this meta-analysis, two large, multi-center, randomized, placebo-controlled trials were undertaken to confirm these findings. Both the Gruppo Italiano Per Lo Studio Della Streptochinasi Nell'Infarto Miocardico (GISSI) (1985) trial and the ISIS-2 (Second International Study of Infarct Survival) Collaborative Group (1988) trial that enrolled 11,806 and 17,189 patients, respectively, demonstrated a reduction in mortality associated with the use of intravenous fibrinolytic therapy (streptokinase) for acute myocardial infarction of a similar magnitude as the Yusef et al. meta-analysis.

In summary, it is rare that an individual empirical research effort can provide definitive and reproducible results. Therefore, meta-analysis can be considered as "an equally important activity of interpreting and integrating the results of the empirical studies that have been done" (Fiske, 1983, p. 65).

Meta-analysis in clinical medicine. Meta-analysis as a technique to integrate empiric research in clinical medicine has lagged behind the need. As evidence, in reviewing the first ten issues published in

1982 of the New England Journal of Medicine, Journal of the American Medical Association, British Medical Journal, and Lancet, Halvorsen (cited in DerSimonian & Laird, 1986) found only one of 589 articles that applied formal statistical methods to combine results. Mulrow (1987) evaluated 50 review articles published between June, 1985 and June, 1986 in the Annals of Internal Medicine, Archives of Internal Medicine, Journal of the American Medical Association, or New England Journal of Medicine. Although some degree of qualitative synthesis (e.g., describing differences in sample populations, intervention approaches, outcome measures) was attempted in 43 of the 50 reviews, qualitative synthesis of the empiric research covered was attempted in only three. In an assessment of review articles published in Clinical Pharmacy, Drug Intelligence and Clinical Pharmacy, Drugs, and Pharmacotherapy, Hendrickson and Amerson (1986) did not even include an analysis of the methodologies used in the reviews. Thus, it appears that not only is the primary literature lagging in the application of meta-analytical techniques, but some of the assessments of the review literature even fail to look for them.

In 1987, Sacks et al. published an evaluation of meta-analyses in clinical medicine to date. In their search for meta-analyses they discovered that although the first was published as early as 1955, only 13 others were published during the subsequent 25 years. However, they discovered an apparent new appreciation for meta-analysis beginning in 1980 by finding 69 published between 1980 and 1987.

Since meta-analysis has been only sparingly used in clinical medicine, there have been few assessments of them. Sacks et al. (1987)

evaluated 86 meta-analyses published in the clinical medicine literature meeting the inclusion requirement that at least one of the studies used in an individual meta-analysis be a randomized controlled trial. Each meta-analysis was reviewed for study design, combinability, control and measurement of potential bias, statistical analysis, sensitivity analysis, and application of results.

The most notable aspect concerning the study design of the meta-analyses evaluated by Sacks et al. (1987) was the paucity of details provided. In only seven percent of the meta-analyses was the protocol described and in only 35% was the literature search strategy detailed. Although the studies included were reported in nearly all meta-analyses, a list of the studies excluded was rarely provided. Treatment assignment (i.e., randomization) within included studies was described for most meta-analyses but few (22%) provided details concerning the ranges in patient, disease, and treatment characteristics across studies.

Sacks et al. (1987) found that less than half (45%) of the 86 meta-analyses evaluated described any differences that existed among studies included. Less common among the evaluated meta-analyses (20%) were statistical methodologies used to determine homogeneity among included studies.

Overall, Sacks et al. (1987) found adequate control and measurement of potential bias infrequently among the 86 evaluated meta-analyses. None reported details to ensure that methods and results were considered separately by the individual meta-analysts. In addition, in none were blinded data-extraction and measurement of

interobserver agreement employed in conjunction.

Adequate statistical methods for meta-analysis for the purpose of their evaluation was defined by Sacks et al. (1987) as "any recognized method of pooling except the simple addition of successes across all trials to give an overall average" (p. 452). Those that reported only simple addition of successes were considered "partial". Adequate statistical methods were used in 66% of the 86 evaluated meta-analyses. Consideration of Type I and Type II errors were acknowledged in 45% of the meta-analyses. Confidence intervals were reported in 43% and subgroup analyses were conducted in 63%.

Among the 86 meta-analyses evaluated by Sacks et al. (1987), few sensitivity analyses were applied. Assessing the quality or making adjustments for differences in quality among studies in individual meta-analyses were discovered in only 19%. The issue of quality of individual studies was acknowledged in less than half (47%). Only 16% of the meta-analyses assessed the effects of different assumptions, tests and criteria. While about 17 (20%) of the meta-analyses acknowledged the problem of publication bias, in only two were adjustments attempted.

Sacks et al. (1987) found that the implications of the meta-analyses as the authors saw them were included in 77%. However, economic considerations were only fully explored in one and addressed to a lesser degree in 17 (20%).

Overall, of the six major categories of meta-analyses assessed by Sacks et al. (1987), only 24 (28%) of the 86 addressed at least one issue in all six categories. Thirty-one (36%) addressed at least one

issue in five of the categories, 25 (29%) addressed four categories, five (6%) addressed three categories, and one (1%) addressed two categories. Therefore, although the use of meta-analysis is increasing in clinical medicine, methodologies and quality vary considerably and improvement is generally warranted.

Using the same meta-analyses as Sacks et al. (1987), Chalmers et al. (1987b) assessed the degree by which meta-analyses of smaller controlled trials agreed with larger co-operative studies. A meta-analysis involving 12 studies of intravenous beta-adrenergic receptor antagonists for acute myocardial infarction in a total of 4,408 patients produced similar results (i.e, confidence intervals of effect) as two separate large co-operative trials, one of which included 5,778 patients and the other 16,027. A meta-analysis of intravenous streptokinase for acute myocardial infarction involving 11 randomized controlled trials and a total of 5,268 patients resulted in a similar magnitude of effect as a large co-operative study involving 11,712 patients; however, the confidence interval was narrower for the large co-operative study. Of interest in this comparison were the contrasting results of a particular subgroup analysis wherein the meta-analysis indicated a favorable effect from the treatment and the large co-operative study indicated a favorable effect from the control. The other comparison of Chalmers et al. was that between a meta-analysis of the effect of phenobarbital for prevention of intracranial hemorrhage in newborn infants involving seven studies and a total of 413 patients with a co-operative study involving 280 patients. The results differed; however, the confidence intervals

overlapped.

Again using the same group of meta-analyses as Sacks et al. (1987), Chalmers, Berrier, Sacks, Levin, Reitman, and Nalgalingham (1987a) evaluated statistical and clinical agreement of meta-analyses concerning the same empiric research. To the original 86 published meta-analyses, five unpublished meta-analyses were added. Among the 91 meta-analyses, 46 represented replicate analyses of 20 different treatments (i.e., 20 cohorts). The levels of statistical agreement were (a) experimental therapy significantly better ($p < .05$), (b) trend in favor of experimental therapy ($p > .05$), (c) no apparent statistical effect, (d) trend favoring control group ($p > .05$), and (e) control group significantly better ($p < .05$). The levels of clinical agreement were gauged on the meta-analysis authors' enthusiasm and were (a) strongly favoring experimental therapy, (b) moderately favoring experimental therapy, (c) no difference of clinical interest, (d) moderately favoring control, and (e) strongly favoring control. The 20 cohorts were divided into two groups; one group in which all meta-analyses agreed within each cohort and another in which at least one meta-analysis within each cohort was in disagreement. This was done for both statistical and clinical scales.

Among the 20 cohorts there was statistical agreement in 10 and disagreement in 10. Among the 10 cohorts in which there was statistical agreement, treatment was favored in eight. In the 10 cohorts in which statistical disagreement existed, the disagreement was often between adjacent levels (e.g., $p < .05$ and $p > .05$); therefore, agreement in direction of effect often occurred despite statistical

disagreement. Clinical agreement was recorded for six of the 20

cohorts. Of the six cohorts with clinical agreement, treatment was

favored in five. As occurred with statistical disagreement, the

magnitude of clinical disagreement was typically adjacent levels. All

six cohorts in clinical agreement were also in statistical agreement.

No differences in agreement/disagreement status were observed within

selected cohorts in which inclusion and exclusion criteria differed

(e.g., meta-analyses including all published and unpublished research

versus meta-analyses including only randomized controlled trials).

Therefore, this preliminary evaluation of meta-analysis in clinical

medicine indicates that there may be differences in the results between

meta-analyses covering the same empirical research; however, the

difference is usually in magnitude and not direction. In addition,

differences are more common to authors' interpretations of the results

than statistical results. As concluded by the authors:

> Although this paper does not settle the question of whether meta-
> analyses of clinical trials as now performed have sufficient
> scientific rigor to reveal reproducible facts, the process must
> continue in the future; hopefully, disagreements will disappear as
> meta-analyses methodology becomes more rigorous. The extent of
> agreement is encouraging, and, taken with the apparent lack of
> disagreement between results of meta-analyses of small trials
> compared with large, co-operative studies, suggest that one should
> not discourage, on the basis of their anticipated size alone, well
> designed and conducted small trials. (p. 740)

The current need for meta-analysis may soon become an

expectation. The Ad Hoc Working Group for Critical Appraisal of the

Medical Literature (Mulrow, Thacker, & Pugh, 1988) recently published

guidelines that call for meta-analytical techniques to be applied to

reviews of medical literature. Einarson et al. (1985) have recommended

"that meta-analysis be used for drug reviews published in the pharmacy

literature" (p. 1962).

Meta-analysis methodology. If the analogy of Louis, Fineberg, and Mosteller (1985) that "meta-analysis is to primary a research study as a primary research study is to its study subjects" (p. 1) is accepted, then conceptual approaches to empiric research can be applied to meta-analysis. Thus, the typical steps required in conducting a meta-analysis include (a) defining a research question, (b) searching and retrieving relevant literature (i.e., subjects), (c) defining inclusion and exclusion criteria, and screening the relevant literature retrieved (i.e, screening subjects), (d) describing and analyzing the data, and (e) reporting and interpreting results (Louis et al., 1985; Thacker, 1988).

The foundation upon which any meta-analysis is built is the clearly defined research question. Concerning research questions as they relate to clinical medicine, Yusuf (1987) advised that, "the question should always be framed in the context of the supposed mechanisms of drug action and the known epidemiology of that particular disease" (p. 281). All subsequent steps are necessarily related to the research question. In addition, covariates of interest also determine subsequent methodological direction. Therefore, no other steps toward conducting a meta-analysis should be taken until the research question is clearly settled (Light, 1987; Thacker, 1988).

The validity and generalizability of a completed meta-analysis is related, in part, to the degree in which the data relevant to the research question is covered. Therefore, systematic processes to retrieve all relevant data are necessary. These data retrieval

processes include (a) electronic searches of appropriate databases (e.g., MEDLINE and Embase for clinical medicine literature), (b) manual searches through the reference sections of previously retrieved literature, and (c) contacting colleagues or other possible sources (e.g., governmental agencies, manufacturers) for unpublished information (Thacker, 1988).

None of the methods used for meta-analysis directly address the choice of inclusion and exclusion criteria to be employed. The inclusion and exclusion criteria are dependent on the research question and researcher predispositions concerning what can and cannot be legitimately pooled (Demets, 1987). At present there are no accepted rules concerning the basic parameters that must be present for a particular study to be included in a meta-analysis. This is a subject of continued debate among meta-analysts. The study parameters considered by meta-analysts acceptable for inclusion span a continuum from randomized controlled trials without confounding variables to all "relevant" studies (independent of form) including those considered flawed. Independent of the parameters by which the meta-analyst employs in selecting empiric data, it must be consistent and taken into consideration when making inferences from the results (Light, 1987).

There are many analytical methods used in meta-analyses. In general, there are two basic analytical approaches used. One is combining significance levels and the other is combining magnitudes of effect (Strube & Hartmann, 1983). The form of the outcome data of interest and the amount of information available dictate, in part, which analytical approach is employed.

Analyses that combine significance tests are generally used when little information is provided in the empiric research substrate. The basic premise of combining significance levels is "that it allows the reviewer to determine whether a set of results could have arisen by chance" (Strube & Hartmann, 1983, p. 15). There are several procedures used to combine statistical significance levels, some of which have been described by Hedges & Olkin (1985, chap. 3). These procedures are necessarily nonparametric and can be difficult to interpret. They only determine whether a difference exists and provide no information in terms of magnitude of effect (Demets, 1987).

A related approach to combining statistical significance levels known as "vote-counting" is based on the proportion of studies within a meta-analysis that reach statistical significance. A relationship between independent and dependent variables is considered significant if a "plurality" of studies reach statistical significance. Hedges and Olkin (1985, chap. 4) have criticized conventional vote-counting methods because of frequently insufficient power to detect small differences even with large sample sizes. However, they have derived methods by which the vote-counting approach can be used to more accurately estimate the magnitude of effect. Like combining statistical significance levels, the usefulness of vote-counting methods are restricted to situations where little information is supplied in the empiric research substrate.

The two most common analytical approaches used in meta-analysis, particularly when two groups are compared, are effect size estimations and odds ratios. Effect size estimations are often used when the form

of the outcome variable of interest is continuous, whereas odds ratios are useful when the outcome variable of interest is dichotomous (Demets, 1987; Strube & Hartmann, 1983).

For meta-analyses in which the outcome variable of interest is continuous, effect sizes and confidence intervals are estimated for each study by using the standardized mean difference and associated standard deviation, respectively (Hedges & Olkin, 1985, chap. 5). An overall effect size and confidence interval can then be derived by averaging across individual studies after weighting them by appropriate factors (e.g., variance, quality) (Hedges & Olkin, 1985, chap. 6). For meta-analyses in which the outcome variable of interest is dichotomous, odds ratios and associated confidence intervals are derived for each study using the proportion of "successes" in one group over the proportion of successes in the comparison group. Odds ratios different than one indicate an effect and the distance from an odds ratio of one indicates the magnitude of effect. Overall odds ratios and confidence intervals are also derived with weighting individual studies for the appropriate factors.

It is assumed that effect sizes (when estimated by using standardized mean differences) and odds ratios are the same across individual studies (fixed effect model). This assumption of homogeneity can be tested. Where heterogeneity of effect sizes or odds ratios exist, outliers can be identified and procedures can be used to cluster groups of studies with homogeneous effect sizes or odds ratios (Hedges & Olkin, 1985). As an alternative, a random effects model could be used to account for the degree of heterogeneity (Hedges &

Olkin, 1985, chap. 9; DerSimonian & Laird, 1986). The importance of homogeneity among individual studies is an issue of debate among meta-analysts and is related to the debate concerning study parameters for inclusion into meta-analyses (Light, 1987).

Independent of the analytical methods used, the risk of publication bias usually exists. Publication bias refers to the dependency of meta-analyses on published literature that is generally selective for studies with positive results. Chan, Sacks, and Chalmers (1982) surveyed 291 authors of randomized clinical trials published in medical journals and found that 41% of the 141 responders had conducted unpublished studies. Among the randomized clinical trials conducted by the authors responding to the survey, 77% of those reporting positive results were published in contrast to 42% of those reporting negative results being published. Therefore, the published literature on which a meta-analysis is based may not be representative of all the relevant empirical research (Begg, 1985).

Methods have been proposed to account for publication bias. For situations in which a positive effect has been detected by a meta-analysis, Rosenthal (1979) has derived a formula whereby the number of unpublished negative trials necessary to make the result of the meta-analysis null can be estimated. Similarly, L'Abbe et al. (1987) developed a method of quantifying publication bias by simulating either the sample size of one unpublished negative trial or the number of small negative trials (with a fixed sample size) that would be required to make the results of a positive meta-analysis negative. The estimates of Rosenthal and L'Abbe et al. are qualitative in nature in

that they provide a level of confidence in the positive results of a meta-analysis, i.e., if only a few negative unpublished studies would make the results null there would be less confidence than if hundreds of unpublished negative trials would be necessary. In contrast, Begg (1985) derived a method whereby the magnitude of publication bias for each study in a meta-analysis is estimated in units of standard deviation relative to the true mean. While this method is more quantitative than those of Rosenthal and L'Abbe et al., it requires knowledge of the incidence of a specific occurrence (e.g., disease) and the total number of subjects possible (independent of consent to participate). Methods to determine and adjust for negative publication bias have not been developed (L'Abbe, et al., 1987)

Reporting results of meta-analyses is similar to reporting results of empirical research. However, detail to the descriptive aspects of the research substrate of a meta-analysis (empiric research) may have added importance for two reasons. One is that meta-analyses naturally accumulate research methods and procedures related to a particular research front that can be easily consulted by researchers investigating future endeavors. Another reason is that reviewers can more easily determine the applicability of a meta-analysis from detailed descriptions (Strube & Hartmann, 1983). Graphical depictions, especially of effects size estimates and odds ratios with their associated confidence intervals are useful adjuncts to descriptions (Walker, Martin-Moreno, & Artalejo, 1988).

The interpretation of meta-analysis results is not a simple matter. Interpretations must take into account the general nature of

meta-analyses. Louis et al. (1985) emphasized this point by stating that:

> Although the collection of papers leading to a meta-analysis might be based on experiments, observational studies, sample surveys, or other forms of investigation, the meta-analysis itself is an observational study with the strengths and weaknesses associated with that design. (p. 2)

Another important consideration is that from a "melange of treatments and mix of patients", quantitative estimates are derived that are generally representative of average effects (Wittes, 1987, p. 275). This is generally not a problem for policy makers such as insurance carriers and governmental agencies who are usually more interested in the types of average effects generated by meta-analyses. However, the quantitative estimates with "very high degrees of power does not gainsay the annoying reality that these estimates of 'average' effects may be very difficult to apply to specific clinical problems" (p. 275).

It is difficult to resist the temptation among those in search of more specific information to dredge the data within a meta-analysis. However, they do so at the risk of finding an apparent effect by chance that is not representative of the true effect (Collins et al., 1987). The peril of post hoc subgroup analysis was demonstrated in the ISIS-1 trial (cited in Collins et al., 1987) of beta-adrenergic receptor antagonists in acute myocardial infarction. Subjects born under the astrological sign of Scorpio benefitted more from the therapeutic intervention than those born under other astrological signs. This result is more likely due to chance than any biological explanation. Peto (1987) has thus suggested that "most of the subgroup analyses from

individual trials or from overviews of randomized trials should be just
reported, but not believed" (p. 235), and Collins et al. (1987)
suggest:

> Inference about the true size of any effects in subsets may be more
> reliable if based indirectly on an overview of all randomized
> patients in all trials, rather than on direct examination of only
> those subsets. (p. 249)

For use in specific clinical situations, the information derived from
any given meta-analysis will rarely be decisive. The information
should be viewed in the context of a specific patient or a specific
therapeutic regimen (Wittes, 1987). It provides some of the
information needed for specific clinical decisions (Yusuf, 1987).

Roles of meta-analysis in clinical medicine. Meta-analysis plays
many roles in clinical medicine. One of the roles meta-analysis plays
is one of stabilization of treatment effects. If individual studies
can vary from the true treatment effect as individual subjects can
within a treatment group, meta-analyses can provide better estimates of
the true effect as do group means derived from individual subjects.
Similarly, meta-analyses can counterbalance any "overenthusiasm" that
might be related to a particular outcome (Furberg & Morgan, 1987).

Meta-analysis has been particularly useful in evaluating moderate
treatment effects. The large sample sizes necessary to detect moderate
treatment effects often result in a series of studies that leave the
research question unresolved. Properly conducted meta-analyses can
provide adequate power to substantially reduce or eliminate the
equivocation (DerSimonian & Laird, 1986). The Food and Drug
Administration used this approach in approving labelling for aspirin
specifying that it could be used to reduce the risk of death in

specifying that it could be used to reduce the risk of death in patients who had previously suffered an acute myocardial infarction. Individual studies had indicated that aspirin might confer such a benefit, but the effect was sufficiently moderate and the sample sizes sufficiently inadequate to reach statistical significance. However, a meta-analysis covering these studies subsequently confirmed the benefit of aspirin in patients previously experiencing acute myocardial infarction and the Food and Drug Administration acted on this information (Furberg & Morgan, 1987; Hennekens, Buring, & Hebert, 1987).

Similarly, meta-analysis can be used to analyze certain subgroups from an aggregate of studies not possible with individual studies. However, considering the danger in subgroup analysis as previously described, the use of meta-analysis for subgroup analysis should be reserved for those subgroups defined a priori. Where subgroups are identified in a meta-analysis by data dredging, they should only serve as topics for future research (Furberg & Morgan, 1987).

Meta-analysis can and has been used in the planning of clinical trials (Hennekens et al, 1987). Research questions can be generated from the results of meta-analyses. For example, in the meta-analysis of intravenous streptokinase for the treatment of acute myocardial infarction (Yusuf et al., 1985), a reduction of mortality was recorded for patients who received treatment within 24 hours of symptom onset. Conventional wisdom at the time suggested that only those patients treated within four to six hours of symptom onset would benefit. To resolve this discrepancy, a large clinical trial was designed that

called for treatment with streptokinase during the first 24 hours after onset of acute myocardial infarction symptoms (ISIS-2, 1988). The results confirmed the earlier meta-analysis in that benefits were recorded in all patients treated within 24 hours of symptom onset.

In planning clinical trials, effect size estimates provided by meta-analyses can assist in estimating necessary sample sizes. Meta-analyses can produce more accurate estimates of effect size than pilot studies. This was illustrated by two studies assessing the effects of intravenous beta-adrenergic receptor antagonists in acute myocardial infarction. One of the studies based sample sizes on an earlier meta-analysis that suggested a 10% reduction in mortality was possible while the other study based sample sizes on a pilot study that suggested a 36% reduction in mortality. Therefore, the sample sizes were substantially different and although each study resulted in the same magnitude of effect (13-15% reduction in mortality), only the results of the study based on the meta-analysis reached statistical significance (Hennekens et al., 1987).

Another role of meta-analysis is permitting a view of "the forest through the trees" such that details or patterns that may not have been discernable in any individual study can be highlighted (Furberg & Morgan, 1987). Similarly, meta-analyses can identify "gaps" in current knowledge, thereby exposing "weaknesses in the empirical assessment of a given theory" (Strube & Hartmann, 1983, p. 23).

Criticisms of meta-analysis. The recent introduction of meta-analysis as a formal method of research synthesis has not been universally embraced. [Eysenck (1978) has referred to it as

"mega-silliness".] The criticisms can be divided into those that are of non-technical (i.e., emotional) origins and those of more technical (i.e., methodological) origins.

Some of the non-technical objections to meta-analysis are rooted in investigators' ownership of research findings and methodologies they used to derive them. It is difficult for some investigators to accept the fact that rarely do individual studies affect the long-range perspective of any particular paradigm. Meta-analyses serve to emphasize this principle as well as to question individual methodologies and underlying assumptions (Fiske, 1983; Glass & Reinhold, 1983). Therefore, investigators unable to dispassionately view meta-analyses that include their work will likely reject them as a legitimate undertaking.

Among clinical medicine researchers, non-technical objections to meta-analysis have been raised in the context of its effects on future research. Where a consensus has arisen with regard to a particular mode of therapy, a reluctance to submit subjects to investigations of alternatives can emerge. An example that has been cited (Yusuf, 1987) is the reluctance of some clinicians to enter post-menopausal women with Stage II breast cancer to chemotherapy regimens because of an existing consensus that tamoxifen (a non-chemotherapeutic agent) is effective even though these clinicians may be uncertain as to the best approach.

Most of the objections and criticisms directed at meta-analysis are on methodological grounds and these primarily relate to the appropriateness of combining study populations, methodologies, and

results, as well as using empiric research of varying quality (Glass &
Kliegl, 1983; L'Abbe et al., 1987). Critics have referred
meta-analysis as comparing apples with oranges; however, the degree to
which one considers this a significant problem depends on whether one
is viewing the meta-analysis as one pertaining to apples, oranges, or
fruit (Mintz, 1983).

Integrating studies of different degrees of quality, and
especially studies considered low in quality, has generated debate as
to the usefulness of meta-analysis. Eysenck (1978), in referring to
the use of low quality studies in meta-analyses, evoked an axiom used
in the computer sciences, "garbage in - garbage out". However, the
quality of studies can be taken into account by either specifying
methodological requirements in the inclusion and exclusion criteria or
by using quality as a covariate (L'Abbe et al., 1987). Chalmers,
Smith, Blackburn, Silverman, Schroeder, Reitman, & Ambroz (1981) have
developed a method by which the quality of a study can be quantified
and weighted accordingly. The seriousness of study design flaws could
then be assessed by the degree in which they correlate with effect
size. Glass and Kliegl (1983) have thus countered Eysenck's contention
by suggesting that differences in study quality handled appropriately
can result in "garbage in - information out".

Pooling results of different studies using different
methodologies, involving different subject types, and done at different
times has long been debated among statisticians. This debate has been
appropriately extended by critics to meta-analysis. [Proponents of
meta-analysis have suggested that narrative reviews suffer the same

problems (Strube & Hartmann, 1983; Thacker, 1988).] However,
techniques such as sensitivity analysis and weighted regression have
been applied to meta-analyses to partially take the heterogeneity of
methods, subjects, and time into account (L'Abbe et al., 1987).
Meta-analysis is still evolving and methods to improve methodological
approaches that address some of the current limitations are under study
(Thacker, 1988).

Summary. Meta-analysis refers to a group of methodologies that
can be used to combine related empiric research to arrive at
conclusions not possible by reviewing individual studies, or improving
generalizations of individual studies. Meta-analysis is distinguished
from the traditional narrative review in that statistical methodologies
are applied to derive "objective" conclusions whereas narrative reviews
are more subjective. However, meta-analysis is not a substitute for a
definitive study in which conclusions are usually based on a more
homogeneous sample than possible with a meta-analysis. Therefore, a
major role of meta-analysis is where the appropriate definitive study
is not logistically feasible or where there is uncertainty as to
whether such a study is warranted.

Meta-analysis is not without its critics; it is perhaps best
described as an evolving entity. To assist the evolutionary process,
L'Abbe et al., (1987) have suggested that a consensus conference be
convened to develop standard protocols. They, among others (Stube &
Hartmann, 1983), have also suggested that central registries of ongoing
trials specific to well-defined content areas (such as the National
Institutes of Health or World Health Organization for clinical

medicine) be established as a means to reduce the sampling bias known to plaque meta-analyses. In addition, brief summaries might be made available to assist investigators in assessing whether certain studies are relevant. Calls have also been issued for continued investigation into statistical methods that will address the shortcomings of meta-analysis, development of methods to assess their quality, and better reporting of research reports (Strube & Hartmann, 1983; Thacker, 1988). Strube & Hartmann have gone one step further in proposing "a generative function for meta-analysis that is an extension of the predictive function" (p. 24)

Although many important and useful meta-analyses related to clinical medicine have been conducted, its acceptance in clinical medicine has been slow in coming. However, the number of meta-analyses published related to clinical medicine is steadily increasing and there is evidence that meta-analytical techniques will eventually be required as part of all literature reviews.

## General Approach

The method by which the comparative nephrotoxicity of gentamicin and tobramycin was assessed using previously completed studies centered on the meta-analytical concept of "effect size". Effect size generally represents the magnitudes of difference between pairs of treatment conditions. Thus, an effect size of zero suggests that there is no difference between a pair of treatment conditions (i.e., gentamicin and tobramycin nephrotoxicity) while an effect size of either less than or greater than zero suggests that a difference exists. The greater the effect size (independent of sign), the greater the magnitude of difference between treatment pairs (Glass et al., 1981).

There are several methods by which to estimate effect size. In the meta-analysis reported here, where the comparative nephrotoxicity of gentamicin and tobramycin could be evaluated with a continuous variable (i.e., degree of nephrotoxicity), effect sizes were estimated by directly calculating standardized mean differences (referred to in this paper as the parametric analysis). Where the comparative nephrotoxicity could only be evaluated with a dichotomous variable (i.e., nephrotoxicity occurred or not), a modified vote-counting procedure was used to estimate effect sizes (Hedges & Olkin, 1985). [Although Hedges and Olkin refer to the modified vote-counting method as "partially parametric" (p. 47), in the meta-analytic procedure reported here, it is referred to as the nonparametric analysis.]

Overall, the meta-analysis of the comparative nephrotoxicity of gentamicin and tobramycin involved three distinct procedures. First,

searching and retrieving the relevant literature; second, screening

the retrieved literature for inclusion and exclusion criteria; and

third, analyzing the data in the literature meeting the screening

criteria.

Literature Search

Both electronic and manual searches of the medical literature

were conducted to locate and retrieve published and unpublished studies

related to the comparative nephrotoxicity of gentamicin and

tobramycin. In addition, the manufacturers of gentamicin (Schering,

Inc.) and tobramycin (Eli Lilly, Inc.) were contacted in order to

retrieve any related unpublished information they might have had on

file.

Electronic literature search. MEDLINE [MEDLARS (Medical

Literature Automated Retrieval System) online], International

Pharmaceutical Abstracts (IPA), Embase, and Dissertation Abstracts

Online were searched electronically. With the exception of

Dissertation Abstracts Online, the published controlled vocabularies

for each of the databases searched electronically were used to find the

most appropriate terms for the search strategy. All searches were

limited to human studies published in the English language.

MEDLINE is an electronic database of predominantly clinical

medicine literature produced by the National Library of Medicine. It

is derived from approximately 3,000 biomedical journals published

worldwide beginning in 1966 (Kruse, 1983). The terms used to search

MEDLINE were, "kidney failure, acute" or "kidney tubular necrosis,

acute" with both "gentamicin" and "tobramycin" (National Library of

Medicine, 1987).

Embase is an electronic database of predominantly clinical medicine literature produced by Elsevier Science Publishers. It is derived from approximately 4,000 biomedical journals published worldwide since 1975 (Kruse, 1983). The terms used to search Embase were, "acute renal failure" with both "gentamicin" and "tobramycin" (Excerpta Medica, 1984).

IPA is produced by the American Society of Hospital Pharmacists and is an electronic database derived from over 600 journals primarily related to pharmacy practice published worldwide since 1970 (Kruse, 1983). The terms used to search IPA were, "kidney failure" with both "gentamicin" and "tobramycin" (Tousignaut, 1987).

Dissertation Abstracts Online is produced by Dissertation Abstracts International and is an electronic database comprised of nearly every doctoral dissertation dating back to 1860 (Perry, 1986). It was searched to determine whether relevant information regarding the comparative nephrotoxicity of gentamicin and tobramycin has been subject of a doctoral dissertation that had not been otherwise published. The terms used for a free-text search of Dissertation Abstracts Online were, "gentamicin", "tobramycin", and "nephrotoxicity".

Manual literature search. The manual literature search consisted primarily of scanning the reference lists of the studies retrieved from the electronic search. In addition, bibliographies provided by the drug manufacturers contacted were scanned for appropriate citations.

Inclusion and Exclusion Criteria.

        The inclusion criteria for entry of individual studies into the

meta-analysis were different depending on the method used to estimate

effect size.   The exclusion criteria were not similarly affected.

        Parametric analysis.  The criteria for the inclusion of studies

that evaluated the comparative nephrotoxicity of gentamicin and

tobramycin using a continuous variable were (a) methods and results

were in the English language; (b) investigations were limited to human

subjects; (c) there were at least two independent groups in each study,

one of which received gentamicin and the other tobramycin; (d) renal

function was measured by either serum creatinine concentrations or

creatinine clearances; and (e) means and measures of variance (i.e.,

standard deviation, standard error, variance, or range) of either

continuous measure of renal function were reported.

        Serum creatinine concentrations and creatinine clearances were

the continuous variables selected as the basis for estimation of effect

sizes because these have been the laboratory values most often used to

measure nephrotoxicity in the comparative studies involving gentamicin

and tobramycin (Schentag, 1983).  Creatinine is a metabolic product

produced in muscle that is released at a relatively constant rate.   In

the absence of renal failure, excretion of creatinine through the

kidneys occurs at a rate (creatinine clearance) approximately that of

blood filtered by the kidneys [glomerular filtration rate (GFR)].

Thus, renal function or changes in renal function can be measured by

creatinine clearance.  Logistical considerations, however, frequently

prohibit accurate measurement of creatinine clearance directly;

therefore, serum creatinine concentration is often used. Since creatinine is produced at a relatively constant rate and is eliminated almost entirely by excretion through the kidneys, a change in renal function can be approximated by corresponding changes in serum creatinine (Ravel, 1978). Serum creatinine concentration and creatinine clearance are therefore necessarily related, and in fact, serum creatinine concentrations may be more sensitive to changes in renal function than creatinine clearance (Morgan, & Will, 1983). Both are considered late markers of aminoglycoside nephrotoxicity (Schentag, 1983).

The exclusion criteria included (a) studies that included data reported in another study, and (b) studies not obtainable either directly or by available intra-library loan programs.

Nonparametric analysis. The inclusion criteria for studies that compared the nephrotoxicity of gentamicin and tobramycin with a dichotomous variable were same as those for studies in the parametric analysis except that the incidence of nephrotoxicity for both the gentamicin groups and tobramycin groups had to be reported instead of a continuous measure of renal function. In addition, the definition of nephrotoxicity used had to be specified. The exclusion criteria were the same as those for studies in the parametric analysis.

Data Collection

From the lists of references available as a result of the literature searches, studies were identified that appeared related to the comparative nephrotoxicity of gentamicin and tobramycin. These references were obtained and screened according to the inclusion and

exclusion criteria. Those studies that met the screening criteria were entered into the meta-analysis and separated into (a) studies to be used for the parametric analyses, (b) studies to be used for the nonparametric analyses, and (c) studies that could be used for both the parametric and nonparametric analyses.

Data Analysis

Parametric analysis. The estimated effect size for each study that met the screening criteria for studies that compared gentamicin and tobramycin nephrotoxicity using a continuous measure of renal function was derived by the standardized mean difference,

$$(\underline{Y}^G - \underline{Y}^T)/\underline{s}, \qquad (1)$$

where $\underline{Y}^G$ and $\underline{Y}^T$ are the mean continuous measures of renal function (serum creatinine concentration or creatinine clearance) for the gentamicin and tobramycin groups, respectively, and $\underline{s}$ is the pooled sample standard deviation as derived by,

$$\underline{s} = \sqrt{\frac{(\underline{n}^G - 1)(\underline{s}^G)^2 + (\underline{n}^T - 1)(\underline{s}^T)^2}{\underline{n}^G + \underline{n}^T - 2}}, \qquad (2)$$

where $\underline{n}^G$ and $\underline{n}^T$ are the gentamicin and tobramycin group sample sizes, respectively, and $\underline{s}^G$ and $\underline{s}^T$ are the standard deviations of the continuous measures of renal functions for the gentamicin and tobramycin groups, respectively. Pooled estimates of sample standard deviations were used because equal population variances for the gentamicin and tobramycin groups could be assumed (Hedges & Olkin,

1985, pp. 78-79). Where the range of measurements of renal function were provided in place of standard deviations, the standard deviation was approximated by $\underline{s}$ = range/$\sqrt{\underline{n}}$. Where the standard error of the renal function measurements were provided in place of the standard deviation, the standard deviation was derived by $\underline{s}$ = standard error $\cdot\sqrt{\underline{n}}$ (Littenberg, 1988).

The estimate of effect size as derived by Equation 1 is associated with a small sample bias approximated by $3\delta/(4\underline{N} - 9)$; thus, as the sample size increases, the bias is reduced. To adjust for this bias, the effect size derived by Equation 1 was multiplied by the correction factor $J(\underline{m}) = [1 - (3/\{4\underline{m} - 1\})]$, where $\underline{m} = [(\underline{n}^T + \underline{n}^G) - 2]$ (Hedges & Olkin, 1985, pp. 79-80). Thus, the effect size ($\underline{d}$) for each study was estimated by

$$\underline{d} = (J\underline{m}) \; \frac{\underline{y}^G - \underline{y}^T}{\underline{s}}. \tag{3}$$

The large sample distribution of Equation 3 approximates normality if $\underline{n}^G$ and $\underline{n}^T$ increase at the same rate. The estimated variance of $\underline{d}$ is thus,

$$\hat{\sigma}^2(\underline{d}) = \frac{\underline{n}^G + \underline{n}^T}{\underline{n}^G\underline{n}^T} + \frac{\underline{d}^2}{2(\underline{n}^G + \underline{n}^T)}, \tag{4}$$

and the 95% confidence intervals are then,

$$\delta_L = \underline{d} - C_{\alpha/2}\hat{\sigma}(\underline{d}), \qquad \delta_U = \underline{d} + C_{\alpha/2}\hat{\sigma}(\underline{d}), \tag{5}$$

where $\delta$ is the population effect size estimated by $\underline{d}$ and $C_{\alpha/2}$ is the two-tailed critical value of the standard normal distribution (Hedges &

Olkin, 1985, pp. 85-88).

Therefore, according to Equation 3, an estimated effect size greater than zero suggests that gentamicin is associated with nephrotoxicity to a greater degree than tobramycin. Conversely, an estimated effect size less than zero suggests that tobramycin is associated with nephrotoxicity to a greater degree than gentamicin. Otherwise an estimated effect size approximating zero suggests that there is no significant difference in the degree of nephrotoxicity between the two agents.

Confidence intervals can also be used for interpretation. A 95% confidence interval as derived by Equation 5 that is comprised of values only greater than zero suggests that gentamicin is associated with nephrotoxicity to a greater degree than tobramycin. Conversely, a 95% confidence interval comprised of only values less than zero suggests that tobramycin is associated with nephrotoxicity to a greater degree. Other-wise, 95% confidence intervals that include zero suggest that there is no significant difference in the degree of nephrotoxicity between the two agents.

To derive an estimated effect size for the series of comparative trials, a weighted linear combination of the individual effect sizes was used ($d_w = w_1 d_1 + \ldots + w_k d_k$, where $w_1 \ldots w_k$ are nonnegative weights summing to unity). Weights were assigned based on the inverse of the effect size variances,

$$w_i = \frac{1}{\hat{\sigma}^2(d_i)} \Big/ \sum_{j=1}^{k} \frac{1}{\hat{\sigma}^2(d_j)}. \tag{6}$$

The weighted estimate of $\delta_+$ ($d_+$) based on the sample estimate of $\delta$ ($d$) to derive the weights for each study was

$$d_+ = \sum_{i=1}^{k} \frac{d_i}{\hat{\sigma}^2(d_i)} \bigg/ \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_i)} \tag{7}$$

(Hedges & Olkin, 1985, pp. 109-111).

Like the effect size estimates for individual studies, the weighted effect size estimate for the series of studies ($d_+$) approximates normality; thus, confidence intervals for $\delta_+$ can be derived using $d+$, assuming $n^T$ and $n^G$ increase in size at the same rate. The confidence intervals for the estimated effect size for the series of studies were then,

$$\delta_L = d_+ - c_{\alpha/2}\, \hat{\sigma}(d_+), \qquad \delta_U = d_+ + c_{\alpha/2}\, \hat{\sigma}(d_+), \tag{8}$$

where $c_{\alpha/2}$ is the two-tailed critical value of the standard normal distribution and $\hat{\sigma}^2(d_+)$ is derived by

$$\hat{\sigma}^2(d+) = \left( \sum_{i=1}^{k} \frac{1}{\hat{\sigma}^2(d_i)} \right)^{-1} \tag{9}$$

(Hedges & Olkin, 1985, pp. 112-113).

Therefore, like estimations of effect size for individual studies, an estimated effect size for the series of studies that is greater than zero suggests that gentamicin is associated with nephrotoxicity to a greater degree than tobramycin, and conversely, an estimated effect size less than zero suggests that tobramycin is associated with nephrotoxicity to a greater degree than gentamicin. Otherwise, estimated effect sizes approximating zero suggest there is

no significant difference in the degree of nephrotoxicity between the two agents.

Confidence intervals can be used to make similar interpretations. A 95% confidence interval comprised of only values greater than zero suggests that gentamicin is associated with nephrotoxicity to a greater extent than tobramycin and an estimated effect size 95% confidence interval comprised of only values less than zero suggests that tobramycin is associated with nephrotoxicity to a greater extent than gentamicin. Otherwise, 95% confidence intervals that include zero suggest no significant difference in the degree of nephrotoxicity between the two agents.

In order to make inferences from the aggregate effect size estimate, $(\underline{d}_+)$, the assumption of homogeneity of effect sizes among the population effect sizes must be met (i.e., $\delta_1 = \delta_2 = \ldots = \delta_k$). Homogeneity of effect sizes was tested by using the Q statistic,

$$Q = \sum_{i=1}^{k} \frac{(\underline{d}_i - \underline{d}_+)^2}{\hat{\sigma}^2(\underline{d}_i)} \quad , \tag{10}$$

wherein Q has an asymptotic chi-square distribution with k-1 degrees of freedom when there is homogeneity of population effect sizes in the series of k studies. Therefore, if Q exceeds the .05 percent critical value of the chi-square distribution with k-1 degrees of freedom, the null hypothesis of homogeneous population effect sizes is rejected (Hedges & Olkin, 1985, pp. 122-123). A group of homogeneous studies was identified by withdrawing studies until the homogeneity assumption was met as defined by the Q statistic.

Sensitivity analyses were conducted to determine the relationship of certain variables with estimated effect size. The effect of homogeneity on the overall effect size estimate was determined by with-drawing individual studies until homogeneity was satisfied by the $Q$ statistic. Relationships of individual study characteristics with estimated effect size were investigated using simple linear regression techniques (Godfrey, 1985) with Systat (Wilkinson, 1985). Independent variables selected were those in which differences of clinical significance existed between the gentamicin and tobramycin groups and included mean age, mean durations of therapy, initial renal function, and incremental changes in renal function. In addition, whether the studies were blinded or randomized was investigated as well. The regression analyses were restricted to the homogeneous group of studies.

Nonparametric analysis. The dichotomous variable used in studies comparing the nephrotoxicity of gentamicin and tobramycin was whether nephrotoxicity occurred or not according to arbitrary criteria. To estimate the effect size of the difference in nephrotoxicity of gentamicin and tobramycin in this series of studies, the modified vote-counting method of Hedges and Olkin (1985, chap. 4) was employed. This method is based on the proportion of studies within a series of k studies in which the difference between groups reach statistical significance.

To estimate the effect size of a series of k studies using the vote-counting method of Hedges and Olkin (1985, chap. 4), each group within each study and between studies must be equal in size, (i.e, the

number of patients in the gentamicin group must be equal to the number of patients in the tobramycin group and all studies must have the same numbers of patients in each treatment group). However, an average value can be derived using the square mean root (SMR),

$$n_{SMR} = \left( \frac{\sqrt{n_1} + \sqrt{n_2} + \dots + \sqrt{n_{k'}}}{k} \right)^2 \tag{11}$$

where $\underline{n}$ is the equivalent sample size in each treatment group and k is the number of studies in the series (Hedges & Olkin, 1985, pp. 67-69).

The modified vote-counting method to estimate effect size also requires that differences between groups were tested statistically by using the $\underline{t}$ distribution. To satisfy this criteria, the differences in proportions of patients considered nephrotoxic in the gentamicin group versus the tobramycin group in each study were tested statistically using the Relative Deviate Test which produced $\underline{z}$ scores (O'Brien & Shampo, 1981). Since the average sample size ($n_{SMR}$) exceeded 30, the $\underline{z}$ distribution approximates the t distribution. A difference reaching a critical $\underline{z}$ score of 1.96 ($\underline{p} = .05$, two-tailed) was considered statistically significant.

The estimated effect size of the difference in gentamicin and tobramycin nephrotoxicity based on the proportions of patients considered nephrotoxic was derived by

$$P_{.05}(\hat{\theta}) = U/k, \tag{12}$$

where U/k is the proportion of studies in which the difference between the gentamicin and tobramaycin groups reached statistical significance at the .05 level (Hedges & Olkin, 1985, pp. 52-53). A table giving P.05($\hat{\theta}$) as a function of effect size and the common sample size ($n_{SMR}$) was used to derive the estimated effect size for the series of studies (pp. 60-61).

Confidence intervals for the proportion of studies in which the difference between the two treatment groups reach statistical significance at the .05 level were computed by,

$$P_L = \hat{p} - C_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{k}} \; ; \qquad P_U = \hat{p} + C_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{k}}, \qquad (13)$$

where p is the proportion of studies in which the difference in proportion of gentamicin and tobramycin patients considered nephrotoxic reached statistical significance, $C_{\alpha/2}$ is the two-tailed critical value for the standard normal distribution, and k is the number of studies in the series (Hedges & Olkin, 1985, p. 54) From these values, the same table used to derive the mean estimated effect size is used to derive the estimated effect size confidence interval.

Because with this method an effect size can be estimated only for a series of studies, sensitivity analyses involving relationships of individual study characteristics with estimated effect size cannot be investigated. However, the effect of specific studies on the study series estimated effect size was investigated by removing individual studies or grouping others.

## Literature Searches

The literature searches identified 36 studies in which the nephrotoxicity associated with gentamicin and tobramycin was compared. After retrieval and screening each study according to the inclusion and exclusion criteria, 18 (50%) were eligible for analysis. All of the studies were published between 1976 and 1985 (Table 1).

Among the clinical trials that were not included in the analysis, three were excluded because they were not comparative, four because renal function was not assessed, one because neither a continuous measure of renal function nor the proportion of patients considered nephrotoxic were reported, and eight because the same data were reported in other studies included in the analysis. In addition, one study identified was not obtainable and one was not published in the English language.

## Descriptive Data

Of the 18 studies included in the analyses (Table 1), 11 provided documentation of renal function sufficient for the parametric analyses (i.e., 10 reporting serum creatinine concentration and one creatinine clearance). Two of these studies were derived from one published article (Matzke, Lucarotti, & Shapiro, 1983). Two separate independent investigations were conducted in this study; and therefore, represent two of the 18 studies included in the meta-analysis (Study 10 and Study 11). Seven of the studies provided only enough information for the nonparametric analyses. However, six of the studies included in the

Table 1

Clinical Trials Included in Analysis

| Study Number | First Author | Year Published | Source |
|---|---|---|---|
| | | | |

Clinical Trials Included Only in Parametric Analysis

| Study Number | First Author | Year Published | Source |
|---|---|---|---|
| 1 | Madsen | 1976 | Journal of Infectious Disease |
| 2 | Kahlmeter | 1978 | Journal of Antimicrobial Chemotherapy |
| 3 | Goodwin | 1979 | Proceedings of 11th International Congress of Chemotherapy |
| 4 | Itsarayoungyuen | 1982 | Pediatric Pharmacology |
| 5 | Donta | 1985 | Antimicrobial Agents and Chemotherapy |

Table continued

Table 1 (continued)

Clinical Trials Included in Analysis

| Study Number | First Author | Year Published | Source |
|---|---|---|---|
| | | | |
| Clinical Trials Included in Both Parametic and Nonparametic Analyses | | | |
| 6 | Smith | 1980 | New England Journal of Medicine |
| 7 | Fong | 1981 | Journal of Antimicrobial Chemotherapy |
| 8 | Schentag | 1981 | Antimicrobial Agents and Chemotherapy |
| 9 | Feig | 1982 | Journal of Antimicrobial Chemotherapy |
| 10 | Matzke (A)[a] | 1983 | American Journal of Nephrology |
| 11 | Matzke (B)[a] | 1983 | American Journal of Nephrology |

Table continued

Table 1 (continued)

Clinical Trials Included in Analysis

| Study Number | First Author | Year Published | Source |
|---|---|---|---|
| | | | Clinical Trials Included in Only Nonparametric Analysis |
| 12 | Walker | 1976 | Journal of Infectious Disease |
| 13 | Wade | 1978 | Lancet |
| 14 | Kunin | 1980 | Journal of the American Medical Association |
| 15 | Keys | 1981 | Mayo Clinic Proceedings |
| 16 | Brown | 1982 | Antimicrobial Agents and Chemotherapy |
| 17 | Pancorbo | 1982 | Biopharmaceutics and Drug Distribution |
| 18 | Fee | 1983 | Review of Infectious Disease |

[a] The Matzke study included two separate analyses and are thus labeled as Matzke (A) and Matzke (B), respectively.

para-metric analysis also provided the proportions of patients considered nephrotoxic in each treatment group so these studies were also included in the nonparametric analysis. Therefore, a total of 13 clinical trials were included in the nonparametric analyses.

The 18 studies included in the meta-analysis involved a total of 967 treatment courses of gentamicin and 876 treatment courses of tobramycin. Of the 11 studies used in the parametric analyses, there were 525 courses of gentamicin and 523 courses of tobramycin. Of the seven clinical trials that could only be used for the nonparametric analyses, there were 442 courses of gentamicin and 353 courses of tobramycin. When the clinical trials that could be used for both analyses were combined with those that could only be used for the nonparametric analyses, there were 862 courses of gentamicin and 758 courses of tobramycin.

As shown in Table 2, the sample sizes were generally equivalent in each treatment group for nearly all the clinical trials included in the analyses. The two exceptions were Study 4 (Itsarayoungyuen, Riff, Schauf, Hamilton et al., 1982), and Study 17 (Pancorbo, Compty, & Heissler, 1984). In Study 4, 20 patients received gentamicin and 30 patients received tobramycin due to a randomization scheme designed to assign patients to gentamicin or tobramycin in a 2:3 ratio. The basis for this randomization scheme was that because gentamicin had been previously studied more extensively in the patient population randomized (neonates), it was desirable to randomize more patients into the tobramycin group. In Study 17, 125 patients received gentamicin and 39 patients received tobramycin. In this study,

Table 2

Descriptive Data of Clinical Trials Included in Analyses

| Study Number | Number of Patients G[a] | T[b] | Patient Age[c] G | T | Total Dose[d] G | T | Duration of Therapy[e] G | T | Serum Conc.[f] G | T | Creatinine[g] G | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Clinical Trials Included Only in Parametric Analysis | | | | | | | |
| 1 | 36 | 36 | NR[h] | NR | NR | NR | 7 | 7 | NR | NR | 106 | 106 |
| 2 | 18 | 19 | 59 | 67 | 3.68 | 2.57 | 21 | 15 | NR | NR | 79 | 103 |
| 3 | 22 | 22 | NR | NR | 2.32 | 2.88 | 15 | 15 | 2.1 | 2.2 | 149 | 157 |
| 4 | 20 | 30 | 0.004 | 0.004 | 0.67 | 0.68 | 8 | 8 | NR | NR | 97 | 88 |
| 5 | 9 | 11 | 55 | 59 | 2.05 | 2.60 | 9 | 9 | NR | NR | 177 | 108 |

Table continued

Table 2 (continued)

Descriptive Data of Clinical Trials Included in Analyses

| Study Number | Number of Patients G[a] | T[b] | Patient Age[c] G | T | Total Dose[d] G | T | Duration of Therapy[e] G | T | Serum Conc.[f] G | T | Creatinine[g] G | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Clinical Trials Included in Both Parametic and Nonparametic Analyses | | | | | | | | | |
| 6 | 72 | 74 | 58 | 59 | 1.90 | 1.92 | 6 | 6 | 2.8 | 2.5 | 150 | 159 |
| 7 | 102 | 103 | 51 | 55 | 1.99 | 1.83 | 8 | 8 | 1.6 | 1.3 | 88 | 97 |
| 8 | 137 | 121 | 67 | 66 | 1.70 | 1.70 | 10 | 11 | 1.6 | 1.4 | 51* | 51 |
| 9 | 25 | 29 | 57 | 57 | NR | NR | 8 | 9 | 1.1 | 1.4 | 85 | 78 |
| 10 | 49 | 49 | 61 | 64 | 1.73 | 1.81 | 10 | 11 | 1.4 | 1.7 | 95 | 104 |
| 11 | 50 | 48 | 62 | 62 | 1.99 | 1.98 | 10 | 11 | 1.4 | 1.5 | 109 | 103 |

Table continued

Table 2 (continued)

Descriptive Data of Clinical Trials Included in Analysis

| Study Number | Number of Patients G[a] T[b] | | Patient Age[c] G T | | Total Dose[d] G T | | Duration of Therapy[e] G T | | Serum Conc.[f] G T | | Creatinine[g] G T | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clinical Trials Included in Only Nonparametric Analysis | | | | | | | | | | | | |
| 12 | 40 | 40 | NR/NS[i] | NR/NS | NR | NR | 7 | 7 | NR | NR | NR | NR |
| 13 | 43 | 47 | NR | NR | NR | NR | NR | NR | NR | NR | NR | NR |
| 14 | 29 | 33 | 55 | 57 | 2.1 | 2.2 | 10 | 11 | NR | NR | NR | NR |
| 15 | 15 | 12 | 51 | 45 | NR | NR | 19 | 14 | 1.6 | 1.9 | NR | NR |
| 16 | 103 | 96 | NR | NR | NR | NR | NR | NR | NR | NR | NR | NR |
| 17 | 125 | 39 | 59 | 59 | 1.7 | 1.8 | 9 | 10 | 1.4 | 1.2 | 128 | 131 |
| 18 | 87 | 86 | 44 | 46 | 1.9 | 2.4 | 9 | 11 | 1.0 | 0.9 | NR | NR |

[a]G = gentamicin. [b]T = tobramycin. [c]Mean age in years. [d]Mean total dose in grams.
[e]Mean duration in days. [f]Mean trough serum concentrations in milligrams/liter.
[g]Mean serum creatinine (milligrams/liter) except for Study 8 which is creatinine
clearance (milliliters/minute). [h]NR = data not reported. [i]NR/NS = data not reported but
difference described as not statistically significant.

patients were not assigned to receive gentamicin or tobramycin by random allocation. The drug prescribed was determined by the primary care physician and at the time of this study, there was an apparent preference for gentamicin at the institution where this study was conducted.

The mean patient age was reported for both the gentamicin and tobramycin groups in 13 studies (72%). Although the mean patient age for both groups was not reported in Study 12 (Walker & Gentry, 1976), it was noted that the difference in ages did not reach statistical significance. With exception of Study 4 (Itsarayoungyuen et al., 1982), which involved neonates, all the studies involved mostly adults. The mean patient ages for both treatment groups generally occurred in the fifth to seventh decades. Within the studies in which mean patient ages were reported, mean ages were always similar for the gentamicin and tobramycin groups (Table 2).

The mean total amount of gentamicin and tobramycin used was recorded in 12 studies (67%). Excluding Study 4 (Itsarayoungyuen et al., 1982), which involved neonates, the mean total amounts of gentamicin and tobramycin used ranged from 1.69 to 3.68 grams and 1.70 to 2.88 grams, respectively. Within the studies reporting the mean amount of gentamicin and tobramycin used, the amounts were very similar for each group with the exception of two studies. In Study 2 (Kahlmeter, Hallberg, & Kamme, 1978), the mean total dose of gentamicin was 3.68 grams compared to 2.57 grams of tobramycin. In Study 18 (Fee, Vierra, & Lathrop, 1978), the mean total dose of gentamicin was 1.9 grams compared to 2.4 grams of tobramycin (Table 2).

The mean duration of gentamicin and tobramycin therapy was reported in 16 studies (89%). The mean duration of therapy ranged from six to 21 days and six to 15 days for gentamicin and tobramycin, respectively. With the exception of two studies, the mean duration of therapy for gentamicin and tobramycin were nearly identical. For both exceptions, gentamicin was used for a longer period of time than tobramycin. In Study 2 (Kahlmeter, Hallberg, & Kamme, 1978), mean gentamicin use duration was 21 days as compared to a mean of 15 days for tobramycin use. In Study 15 (Keys, Kurtz, Jones, & Muller, 1981), mean gentamicin use duration was 19 days compared to a mean of 14 days for tobramycin use. In Study 2, the longer mean duration of therapy for gentamicin correlated with the larger mean total dose reported. Mean total dose for Study 15 was not reported (Table 2).

Mean trough gentamicin and tobramycin serum concentrations were reported in 10 studies (56%). Mean trough serum concentrations in the studies in which they were reported ranged from 1.0 to 2.8 milligrams/liter and 0.9 to 2.5 milligrams/liter for gentamicin and tobramycin, respectively. Within each reporting study, the gentamicin and tobramycin trough serum concentrations were very similar (Table 2).

The mean serum creatinine concentrations prior to initiation of therapy were reported for both the gentamicin and tobramycin groups in 11 studies (61%) and the mean creatinine clearance prior to initiation of therapy was reported in one (6%). The mean initial serum creatinine concentration ranged from 79 to 177 micromoles/liter and 78 to 159 micromoles/liter in the gentamicin and tobramycin groups, respectively. Within each of the reporting studies, initial serum creatinine

concentrations were similar for both gentamicin and tobramycin groups with the exception of Study 5 (Donta & Lembke, 1985) in which the initial serum creatinine concentration for the gentamicin group was 177 micromoles/ liter compared to 108 micromoles/liter for the tobramycin group. Serum creatinine concentrations considered to be indicative of normal renal function range from 71 to 177 micromoles/liter (Ravel, 1978). Only the gentamicin group in Study 5 reached the upper limit.

In Study 8 (Schentag, Plaut, & Cerra, 1981) creatinine clearance was used instead of serum creatinine concentrations. The initial creatinine clearance for both the gentamicin and tobramycin groups was 51 milliliters/minute (Table 2). Creatinine clearances indicative of normal renal function are between 90 and 120 milliliters/minute; however, normal values decrease with age (Ravel, 1978). Thus, Study 8 differed from the others in that both groups had compromised renal function at the initiation of therapy. This is consistent with the greater severity of illness among the patients in Study 8 than the other studies.

The incidence of nephrotoxicity for each treatment group was recorded in each of the 13 studies included in the nonparametric analysis according to the definition of nephrotoxicity established by the investigators of each study. The definitions of nephrotoxicity used in each of these studies are listed in Table 3. Although the definitions varied among studies, they were generally similar.

Among the 13 studies included in the nonparametric analysis (Table 4), the nephrotoxicity incidence ranged from 4 to 55% and 2.5 to 58% for the gentamicin and tobramycin groups, respectively. The

Table 3

Definitions of Nephrotoxicity in Nonparametric Analysis

| Study Number | Definition of Nephrotoxicity |
|---|---|
| **Studies Included in Both Parametic and Nonparametic Analyses** | |
| 6 | $SCR^a$ increase $\geq$ 44 if initial SCR < 265 or, SCR increase $\geq$ 88 if initial SCR $\geq$ 265 |
| 7 | SCR increase $\geq$ 44 if initial SCR < 265 or, SCR increase $\geq$ 88 if initial SCR $\geq$ 265 |
| 8 | SCR increase $\geq$ 44 |
| 9 | SCR increase $\geq$ 27 |
| 10 | SCR increase $\geq$ 44 if initial SCR < 177 or, SCR increase $\geq$ 30% if initial SCR $\geq$ 177 |
| 11 | SCR increase $\geq$ 44 if initial SCR < 177 or, SCR increase $\geq$ 30% if initial SCR $\geq$ 177 |
| **Studies Included in Only Nonparametric Analysis** | |
| 12 | SCR increase |
| 13 | SCR increase $\geq$ 35 if initial SCR < 265 or, SCR increase $\geq$ 80 if initial SCR $\geq$ 265 |
| 14 | SCR increase $\geq$ 33% |
| 15 | Iothalamate decrease to < 14% of initial |
| 16 | SCR increase > 35 |
| 17 | SCR increase $\geq$ 30% |
| 18 | Final SCR > 133 with decrease in CRCL > 33% or, SCR increase $\geq$ 88 if initial "abnormal" |

$^a$SCR = serum creatinine concentration in micromoles per liter.
$^b$CRCL = creatinine clearance in milliliters per minute.

Table 4

Components of Nonparametric Effect Size Estimations

| Study Number | Proportion Nephrotoxic[a] $G^b$ | $T^c$ | Z-Score | Difference Statistically Significant |
|---|---|---|---|---|
| Studies Included in Both Parametic and Nonparametic Analyses | | | | |
| 6 | 19/72 | 9/74 | 2.8 | Yes |
| 7 | 8/102 | 7/103 | 0.3 | No |
| 8 | 51/137 | 27/121 | 2.5 | Yes |
| 9 | 10/25 | 8/29 | 1.0 | No. |
| 10 | 5/49 | 9/49 | -1.1 | No |
| 11 | 4/50 | 8/48 | -1.3 | No |
| Studies Included in Only Nonparametric Analyses | | | | |
| 12 | 7/40 | 2/40 | 1.8 | No |
| 13 | 13/43 | 11/47 | 0.8 | No |
| 14 | 16/29 | 5/33 | 3.3 | Yes |
| 15 | 6/13 | 7/12 | -0.9 | No |
| 16 | 5/103 | 2/96 | 1.1 | No |
| 17 | 5/125 | 1/39 | 0.4 | No |
| 18 | 22/87 | 13/86 | 1.8 | No |

[a] Number of patients nephrotoxic/total number of patients.
[b] G = gentamicin group.   [c] T = tobramycin group.

incidence of nephrotoxicity was lower for the gentamicin group in three

of the studies and higherin the other 10. The difference in the

incidence of nephrotoxicity between groups reached statistical

significance in three studies, in all of which a lower incidence was

recorded for the tobramycin groups. The greatest differential between

the two groups in any one study was recorded in Study 14 (Kumin, 1980)

in which the incidence of nephrotoxicity for the tobramycin group was

15% compared to 55% for the gentamicin group.

The ranges of nephrotoxicity for both groups remained the same

when the seven studies that could only be used in the nonparametric

analysis were considered. In only one study was the incidence of

nephrotoxicity lower for gentamicin than tobramycin. Also in only one

study did the difference in the incidence of nephrotoxicity between the

treatment groups reach statistical significance (Table 4).

Parametric Analyses

Effect size estimations. The components used to estimate effect

sizes for each study based on a continuous measure of renal function

are listed in Table 5. The estimated effect sizes derived, and their

respective variance terms (standard deviations and 95% confidence

intervals), are listed in Table 6. Figure 1 is a plot of the estimated

effect sizes and associated 95% confidence intervals for each study.

The estimated effect sizes ranged from -0.887 to 1.666. Four of

the 11 studies were associated with estimated effect sizes of negative

values (suggesting tobramycin is associated with nephrotoxicity to a

greater degree than gentamicin) and the remaining seven studies were

associated with estimated effect sizes of positive values (suggesting

Table 5

Components of Effect Size Calculations in Parametric Analysis

| Study Number | Creatinine[a] G[b] | T[c] | Standard Deviation G | T | Pooled Standard Deviation | Correction Factor [J(m)] |
|---|---|---|---|---|---|---|
| 1 | 114 | 111 | 26.5 | 26.5 | 25.5 | 0.989 |
| 2 | 100 | 109 | 15.2 | 31.0 | 24.6 | 0.978 |
| 3 | 323 | 157 | 133.3 | 38.8 | 98.1 | 0.982 |
| 4 | 124 | 88 | 39.5 | 48.4 | 45.1 | 0.984 |
| 5 | 207 | 126 | 275.7 | 59.7 | 189.1 | 0.958 |
| 6 | 186 | 168 | 150.2 | 152.3 | 151.2 | 0.995 |
| 7 | 203 | 230 | 28.0 | 31.3 | 29.8 | 0.996 |
| 8[d] | 41 | 50 | 27.0 | 34.0 | 30.5 | 0.997 |
| 9 | 100 | 78 | 43.3 | 23.9 | 35.0 | 0.957 |
| 10 | 125 | 134 | 40.4 | 35.4 | 38.0 | 0.992 |
| 11 | 148 | 149 | 30.0 | 122.5 | 88.3 | 0.992 |

[a] Initial serum creatinine concentrations in micromoles/liter.
[b] G = gentamicin group.  [c] T = tobramycin group.  [d] Initial creatinine clearance in milliliters/minute.

Table 6

Estimated Effect Sizes in Parametric Analysis

| Study Number | Estimated Effect Size | Standard Deviation | 95% Confidence Interval |
|:---:|:---:|:---:|:---:|
| 1 | 0.131 | 0.236 | −0.332 − 0.593 |
| 2 | −0.334 | 0.331 | −0.983 − 0.316 |
| 3 | 1.666 | 0.350 | 0.980 − 2.352 |
| 4 | 0.772 | 0.300 | 0.187 − 1.358 |
| 5 | 0.420 | 0.454 | −0.480 − 1.300 |
| 6 | 0.116 | 0.166 | −0.201 − 0.440 |
| 7 | −0.887 | 0.146 | −1.160 − −0.332 |
| 8 | 0.294 | 0.125 | 0.049 − 0.540 |
| 9 | 0.605 | 0.457 | −0.291 − 1.502 |
| 10 | −0.233 | 0.203 | −0.630 − 0.165 |
| 11 | −0.020 | 0.202 | −0.416 − 0.376 |
| Mean[a] | 0.007 | 0.063 | −0.116 − 0.131 |

[a]Weighted by inverse of estimated effect size variance.

Figure 1.  Effect Size Estimates in Parametric Analysis.
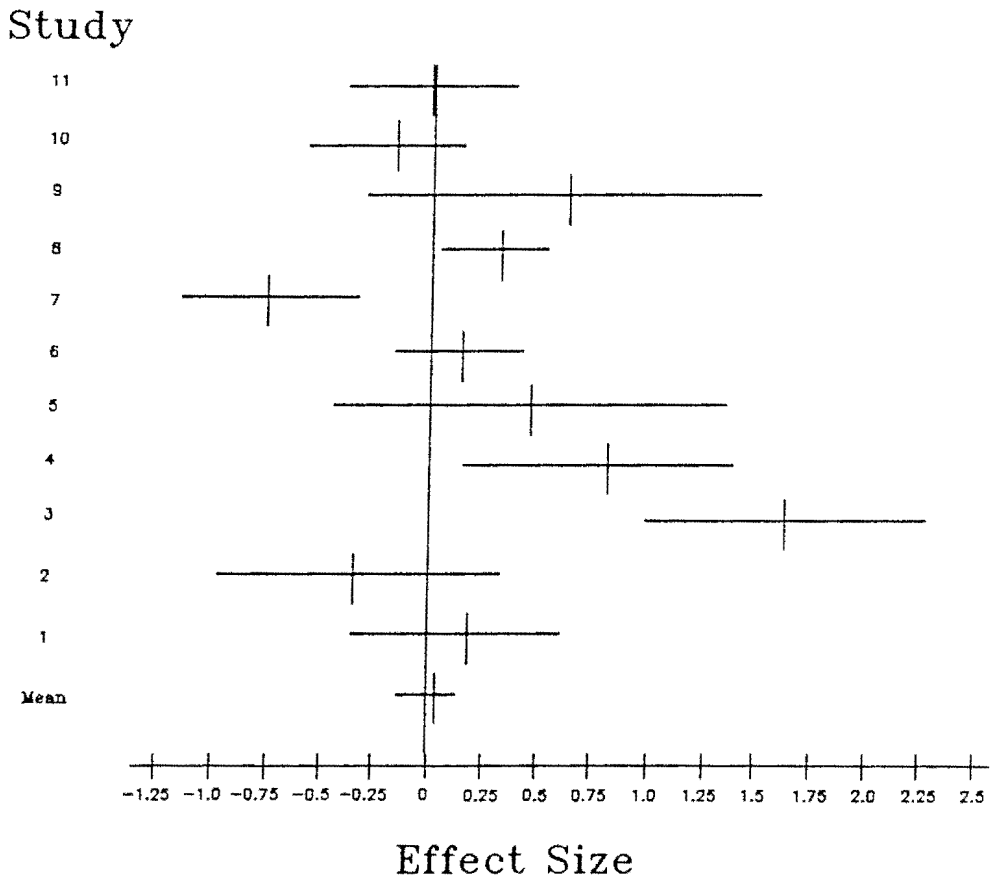
gentamicin is associated with nephrotoxicity to a greater degree than tobramycin).

The standard deviations of the estimated effect sizes ranged from 0.125 to 0.457. The widest 95% confidence interval was 1.81 and the narrowest was 0.26 (Figure 1). The 95% confidence intervals included zero for seven studies (suggesting no difference in the degree of nephrotoxicity associated with gentamicin and tobramycin). Of the four studies in which the 95% confidence intervals did not include zero, three encompassed only positive values (suggesting gentamicin is associated with nephrotoxicity to a greater degree than tobramycin) and one encompassed only negative values (suggesting tobramycin is associated with nephrotoxicity to a greater degree than tobramycin).

The estimated effect size for this series of 11 studies was derived from the data shown in Table 7. The estimated effect size for this series was 0.007 with an associated standard deviation of 0.063 and a 95% confidence interval of -0.116 to 0.131 (Figure 1). Since the aggregate effect sized estimate approximates zero and the associated confidence interval encompasses zero (Table 6), it would appear from these results that there is no significant difference in the degree of nephrotoxicity between gentamicin and tobramycin. Technically, the interpretation of these results is that after treatment with either gentamicin or tobramycin, the serum creatinine concentrations or creatinine clearances would not be different. Whether one chooses to interpret this to mean the degree of nephrotoxicity does not differ between the two drugs depends on whether serum creatinine concentrations and creatinine clearances are accepted as representative

Table 7

Components of Estimated Effect Size Calculation For

Parametric Study Series

| Study Number | $\underline{d}^a$ | $\hat{\sigma}^2(\underline{d})^b$ | $\underline{d}/\hat{\sigma}^2(\underline{d})$ | $1/\hat{\sigma}^2(\underline{d})$ | $\underline{d}^2/\hat{\sigma}^2(\underline{d})$ |
|---|---|---|---|---|---|
| 1 | 0.131 | 0.056 | 2.345 | 17.692 | 0.306 |
| 2 | −0.334 | 0.110 | − 3.042 | 9.116 | 1.016 |
| 3 | 1.666 | 0.122 | 13.606 | 8.166 | 22.670 |
| 4 | 0.772 | 0.772 | 8.650 | 11.198 | 6.681 |
| 5 | 0.420 | 0.206 | 1.989 | 4.849 | 0.816 |
| 6 | 0.116 | 0.027 | 4.217 | 36.432 | 0.488 |
| 7 | −0.887 | 0.021 | −41.404 | 46.656 | 36.742 |
| 8 | 0.294 | 0.016 | 18.713 | 63.566 | 5.509 |
| 9 | 0.605 | 0.209 | 2.893 | 4.781 | 1.751 |
| 10 | −0.233 | 0.041 | − 5.659 | 24.335 | 1.316 |
| 11 | −0.020 | 0.041 | − 0.495 | 24.489 | 0.010 |
| Sum | | | 1.813 | 251.551 | 77.306 |

[a]Estimated effect size.  [b]Estimated effect size variance.

of renal function.

Homogeneity assumption. The criterion for homogeneity among the 11 studies was not met [$Q$ (10, $N$ = 11) = 77.3, $p$ < .05]. Inspection of Table 6 and Figure 1 suggested that Study 3 (Goodwin, 1979) and Study 7 (Fong, Fenton, & Bird, 1981) may have contributed most to the heterogeneity among the studies. When these two studies were excluded from the analysis, the remaining nine studies met the criteria for homogeneity [$Q$ (8, $N$ = 9) = 13.4, $p$ > .05].

The range of the estimated effect sizes among the nine homogeneous studies was from -0.334 to 0.772. Three of the estimated effect sizes were negative and six were positive values. The estimated effect size standard deviations ranged from 0.125 to 0.457. The widest 95% confidence interval was 1.81 and the narrowest 0.49. The estimated effect size for the series of nine homogeneous studies was 0.15 with an associated standard deviation of 0.005 and a 95% confidence interval of 0.01 to 0.29.

The interpretation of comparative nephrotoxicity changes when the effect size estimate of only the nine homogeneous studies are considered below. Assuming a normal distribution, the average serum creatinine concentration or creatinine clearance in patients treated with gentamicin will exceed those of approximately 55% (standard normal deviate of 0.15) of the patients treated with tobramycin. Considering the 95% confidence interval, the average serum creatinine concentration or creatinine clearance in patients treated with gentamicin could exceed those of as many as approximately 61% (standard normal deviate of 0.29) or as few as 50% (standard deviate of 0.01) of

the patients treated with tobramycin. Again, the degree to which this represents a difference in the comparative nephrotoxicity of gentamicin and tobramycin will be dependent on the degree to which these variables are accepted as representative of renal function. If these variables are accepted as representative, the clinical significance of the difference in comparative nephrotoxicity will be determined by counter-balancing the excess risk of nephrotoxicity with the economic advantages associated with gentamicin use.

Sensitivity analyses. Sensitivity analyses were conducted to determine whether any relationships existed between estimated effect sizes for individual studies and certain study characteristics. The sensitivity analyses were restricted to the nine clinical trials that were homogeneous.

With the exception of one study, the sample populations involved mostly adult patients (Table 2). In Study 4 (Itsarayoungyuen et al., 1982), the sample population included only neonates. The mean age was 1.5 days. By excluding this patient population, the estimated effect size for the remaining series of eight studies was 0.113 with an associated standard deviation of 0.073 and 95% confidence interval of -0.031 to 0.257. The assumption of homogeneity remained after exclusion of Study 4 $[Q (7, \underline{N} = 8) = 8.84 \underline{p} > .05]$. Thus, by including only clinical trials involving adult patients, the estimated effect size for the series of homogeneous studies involving mostly adult patients does not change appreciably from when the study involving neonates is included.

The basis for the effect size estimate in all but one of the

studies was serum creatinine concentrations.  In Study 8 (Schentag et al., 1981), the basis for the effect size estimate was creatinine clearance.  By excluding this study from the analysis, the estimated effect size for the remaining eight studies was 0.082 with an associated standard deviation of 0.087 and 95% confidence interval of -0.088 to 0.252.  Excluding Study 8 also reduced heterogeneity among the remaining eight studies [$Q$ (7, $N$ = 8) = 11.49, $p$ > .05].  Thus, by excluding Study 8 so that only homogeneous clinical trials employing serum creatinine concentrations as the endpoint are included in the analysis, the change in the aggregate estimated effect size is not of clinical significance.

As shown in Table 2, for the studies reporting patient ages, the mean ages were always very similar within each study; however, the mean ages differed between studies.  Therefore, mean age for both treatment groups were pooled within each study and regressed on estimated effect size.  Age accounted for about 32% of the variability in estimated effect size; however, this relationship did not reach statistical significance [$r^2$ (6, $N$ = 7) = .318, $p$ = .085].  Of note was the direction of the relationship (regression coefficient of -1.012).  As age increased, estimated effect size decreased.  This may have been due to the inclusion of Study 4 (Itsarayoungyuen et al., 1982) which included only neonates and was associated with a relatively high effect size estimate.  When this study was taken out of the analysis, the variability in estimated effect size associated with age decreased to approximately 9%; however, this relationship also did not reach statistical significance [$r^2$ (5, $N$ = 6) = 0.089, $p$ = .264].  Therefore,

it does not appear that any differences in nephrotoxicity associated with gentamicin and tobramycin are linearly related to patient age (Table 8).

In the studies reporting the duration of gentamicin and tobramycin use, the durations of use were similar for both treatment groups for nearly all of the studies (Table 2). Therefore, the durations of gentamicin and tobramycin use were pooled within studies and regressed on effect size estimates to determine whether duration of use affected effect size estimates (Table 8). Duration of aminoglycoside use accounted for approximately 21% of the variability in estimated effect size; however, this relationship did not reach statistical significance [$r^2$ (7, $N$ = 8) = .209, $p$ = .121]. Thus, it does not appear that any differences in nephrotoxicity associated with gentamicin and tobramycin are linearly related to the duration of use.

Differences in initial renal function between treatment groups, as measured by initial serum creatinine concentrations or creatinine clearances, existed among some of the clinical trials (Table 2); however, not all were clinically significant. The differences between initial serum creatinine concentrations or creatinine clearances between treatment groups were regressed on effect size estimates (Table 8). The differences in initial serum creatinine concentrations and creatinine clearances accounted for approximately 16% of the variability in estimated effect size; however, this relationship did not reach statistical significance [$r^2$ (7, $N$ = 8) = .156, $p$ = .159].

Variation existed among the studies in the differences between the treatment groups in the incremental changes during therapy in serum

Table 8

Sensitivity Analysis - Linear Regression Analyses

| Variable | Coefficient | Adjusted $R^2$ | D.F.[a] | p[b] |
|---|---|---|---|---|
| Age(1)[c] | -1.012 | .318 | 6 | .09 |
| Age(2)[d] | -0.045 | .089 | 5 | .26 |
| Duration[e] | -0.061 | .209 | 7 | .12 |
| Initial CR[f] | 0.007 | .156 | 7 | .16 |
| Increase CR[g] | -0.002 | .000 | 7 | .88 |
| Random[h] | 0.105 | .000 | 7 | .71 |
| Blind[i] | 0.457 | .299 | 7 | .07 |

[a]D.F. = degrees of freedom. [b]P = two-tailed probability of coefficient not equaling zero. [c]Age(1) = pooled ages for all homogeneous studies. [d]Age(2) = pooled ages for all homogeneous studies excluding the study with neonates (Study 9). [e]Pooled duration of gentamicin and tobramycin use (days). [f]Difference in initial serum creatinine concentration or creatinine clearance between gentamicin and tobramycin groups. [g]Difference in incremental increase in serum creatinine concentration or creatinine clearance between gentamicin and tobramycin groups. [h]Whether patients were randomized to treatment groups. [i]Whether study was blinded.

creatinine concentrations or creatinine clearances (Tables 2 and 5).
The difference in incremental change between thetreatment groups were
regressed on the effect size estimates (Table 8). The incremental
change in serum creatinine concentration or creatinine clearance did
not account for any variability in the estimated effect size
($r^2$ =.000); therefore, it does not appear that any differences in the
incremental changes in measurements of renal function between
gentamicin and tobramycin are associated with a difference in
nephrotoxicity.

Only three studies were blinded [Study 9 (Feig et al., 1982),
Study 6 (Smith et al., 1980), and Study 4 (Itsarayoungyuen et al.,
1982)] (Table 6). Blinding status did not appreciably affect effect
size estimates and did not reach statistical significance
[$r^2$ (7, $N$ = 8) = .105, $p$ = .71].

Three studies did not randomize patients to either treatment
group [Study 2 (Kahlmeter et al., 1978), Study 5 (Donta & Lembke,
1985), and Study 8 (Schentag et al., 1981)] (Table 8). Randomization
status accounted for approximately 30% of the variability in effect
size estimates; however, this relationship did not reach statistical
significance [$r^2$ (7, $N$ = 9) = .299, $p$ = .07). Of note was the
direction of the relationship (regression coefficient of 0.105)
suggesting that higher effect size estimates (i.e., differences in the
degree of nephrotoxicity between gentamicin and tobramycin) may be
expected more often in randomized studies.

Nonparametric Analyses

Effect size estimation. The nonparametric estimations of effect

size were based on the differences in the proportions of studies in which the nephrotoxicity incidence between the gentamicin groups and tobramycin groups reached statistical significance. The proportion of studies in which the difference in the incidence of nephrotoxicity between gentamicin and tobramycin reached statistical significance by the Relative Deviate Test was 0.231 (3/13) with an associated 95% confidence interval of 0.002 to 0.437. The estimated effect size based on this proportion was 0.117 with an associated 95% confidence interval of 0 to 0.226 (Table 9).

Sensitivity Analysis. When the seven studies that could only be used in the nonparametric analyses were considered (Studies 12-18), the proportion in which the difference in nephrotoxicity incidence between gentamicin and tobramycin reached statistical significance was 0.14 (1/7) with an associated 95% confidence interval of -0.116 to 0.402. The estimated effect size based on this proportion was 0.080 with an associated 95% confidence interval of -0.60 to 0.200.

In the nonparametric analysis that included the studies that were also in the parametric analysis, one of the studies included [Study 7 (Fong et al., 1981)] was one that was excluded in the homogeneous parametric analysis. When this study was eliminated from the nonparametric analysis, the proportion of studies in which the difference in nephrotoxicity incidence between gentamicin and tobramycin reached statistical significance was 0.25 (3/12) with an associated 95% confidence interval of 0.005 to 0.495. The estimated effect size was 0.134 with an associated 95% confidence interval of 0 to 0.232.

Table 9

Estimated Effect Sizes for Nonparametric Analyses

| Study Group[a] | Proportion of Differences Significant[b] | 95% Confidence Interval of Proportion | Estimated Effect Size | 95% Confidence Interval of Effect Size |
|---|---|---|---|---|
| 1 | 3/13 (23%) | -0.008 - 0.437 | 0.117 | 0.000 - 0.226 |
| 2 | 3/12 (25%) | -0.001 - 0.469 | 0.134 | 0.000 - 0.232 |
| 3 | 1/7  (14%) | -0.116 - 0.402 | 0.080 | -0.060 - 0.200 |

[a]Study groupings:
  1 = All studies used in nonparametric analysis (Studies 6-18).
  2 = Studies used in nonparametric and homogeneous parametric analyses (Studies 6, 8-18).
  3 = Studies only used in nonparametric analysis (Studies 12-18).

[b]Number of studies in which differences in proportions nephrotoxic significant/total number of studies.

The nonparametric estimations of effect size were remarkably similar to those resulting from the parametric analysis. If one accepts the arbitrary definitions of nephrotoxicity in the studies used in the nonparametric analyses and serum creatinine concentrations or creatinine clearance as markersof nephrotoxicity, the interpretations of the nonparametric analysis results would parallel those of the parametric analysis. Using the same assumptions, it could be suggested that the modified vote-counting method of Hedges and Olkin (1985) may be a reliable alternative when the empirical research substrate provides only limited information (Table 10 and Figure 2).

Summary

Eighteen clinical studies related to the comparative nephrotoxicity of gentamicin and tobramycin met the criteria for this meta-analysis. Analysis by two different methods indicated that if there is a difference in nephrotoxicity between the two drugs, it is not of a great magnitude. In addition, none of the selected covariates affected the difference in nephrotoxicity between gentamicin and tobramycin to an extent that reached statistical significance.

Secondarily, the modified vote-counting method produced very similar results as the parametric analysis (Table 10 and Figure 2). Thus, despite that conventional vote-counting methods are often dismissed as not being useful, the modified vote-counting method of Hedges and Olkin (1985, chap. 4) may indeed have a role in situations where the empirical research under investigation does not provide enough information to apply the more parametric procedures.
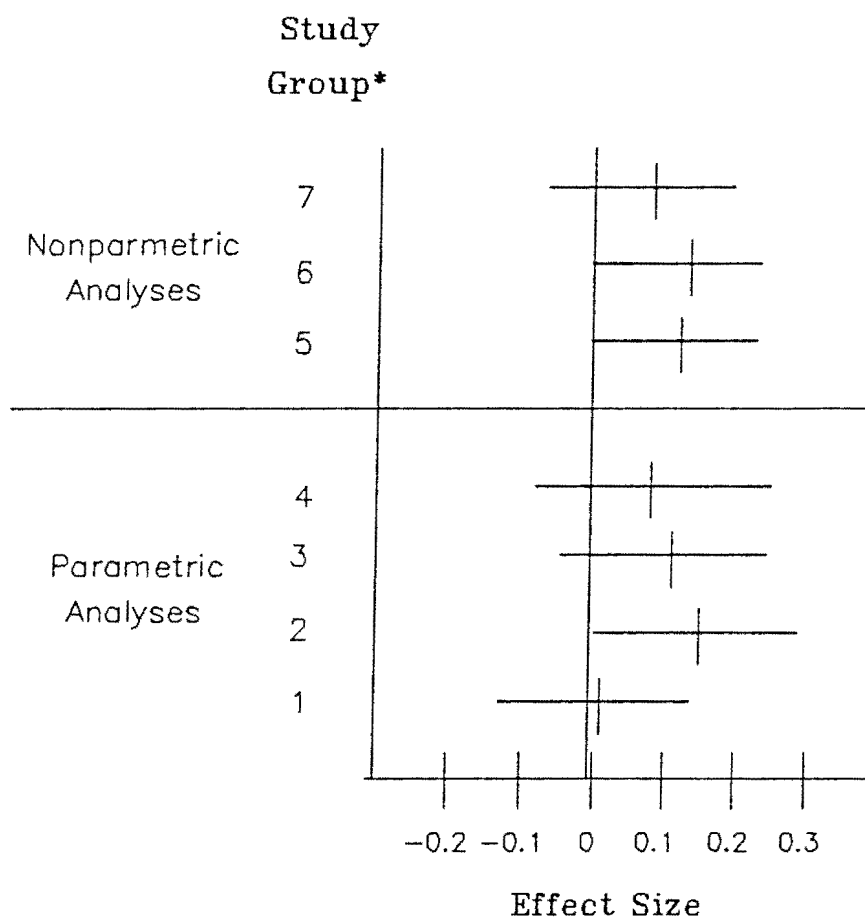
Table 10

Summary of Estimated Effect Sizes

| Study Group[a] | Mean Estimated Effect Size | 95% Confidence Interval |
|---|---|---|
| Parametric Analyses | | |
| 1 | 0.007 | -0.116 - 0.131 |
| 2 | 0.150 | 0.010 - 0.290 |
| 3 | 0.113 | -0.031 - 0.257 |
| 4 | 0.082 | -0.088 - 0.252 |
| Nonparametric Analyses | | |
| 5 | 0.117 | 0.000 - 0.226 |
| 6 | 0.134 | 0.000 - 0.232 |
| 7 | 0.080 | -0.060 - 0.200 |

[a]Study groupings:
1 = All studies in parametric analysis (Studies 1-11).
2 = Homogeneous studies in parametric analysis (Study grouping 1 minus Study 3 and Study 7).
3 = Study grouping 2 minus Study 4 (neonates).
4 = Study grouping 2 minus Study 8 (creatinine clearance).
5 = All studies in nonparametric analysis (Studies 6-18).
6 = Studies in nonparametric and homogeneous parametric analyses (Studies 6, 8-18).
7 = Studies only in nonparametric analysis (Studies 12-18).

Figure 2.  Effect Size Estimates and 95% Confidence Intervals
For Parametric and Nonparametric Analyses.



Study

Group*

Nonparmetric
Analyses

Parametric
Analyses

7

6

5

4

3

2

1

-0.2 -0.1  0  0.1  0.2  0.3

Effect Size

* As listed in Table 10.

SUMMARY

Aminoglycoside antibiotics are important agents in the
treatment of serious to life-threatening bacterial infections.
However, the use of these antibiotics is hampered by an association
with nephrotoxicity.  Research efforts have been undertaken to
determine whether any of the aminoglycoside antibiotics is less
nephrotoxic than the others.  Comparisons of two aminoglycoside
antibiotics in particular, gentamicin and tobramycin, have produced
equivocal results.  Some data have suggested that gentamicin is
associated with nephrotoxicity to a greater degree than tobramycin
while other data have suggested no difference.  Published reviews of
the empirical gentamicin and tobramycin comparisons  have been as
equivocal with regard to their comparative nephrotoxicity  as the
empirical research they covered.  However, none of the published
reviews applied systematic meta-analytical techniques.

In the investigation reported here, meta-analytical techniques
were used to assess the empirical research comparing the nephrotoxicity
of gentamicin and tobramycin in humans.  Specifically, effect sizes
were estimated using the parametric approach of standardized mean
differences.  In addition, a modified vote-counting procedure was used
in those situations where there was insufficient information for
parametric analysis.  When all studies in the parametric analysis were
included, there appeared to be no difference in the degree of
nephrotoxicity between gentamicin and tobramycin; however, when only

homogeneous studies were included, it appeared that gentamicin may indeed be associated with nephrotoxicity to a slightly greater degree. Interestingly, effect size estimates derived using the modified vote-counting method produced similar results and interpretations.

REFERENCES

Begg, C. B. (1985).A measure to aid in the interpretation of published
   clinical trials.  Statistics in Medicine, 4, 1-9.

Brown, A. E., Quesada, O., & Armstrong, D.  (1982).  Minimal
   nephrotoxicity with cephalosporin-aminoglycoside combinations in
   patients with neoplastic disease.  Antimicrobial Agents and
   Chemotherapy, 21, 592-594.

Burkle, W. S. (1986).  Is tobramycin less nephrotoxic than gentamicin?
   Clinical Pharmacy, 5, 514-516.

Chan, S. S., Sacks, H. S., & Chalmers, T. C. (1982).  The epidemiology
   of unpublished randomized control trials.  Clinical Research, 30,
   234A.

Chalmers, T. C., Berrier, J., Sacks, H. S., Levin, H., Reitman, D., &
   Nagalingam, R. (1987a).  Meta-analysis of clinical trials as a
   scientific discipline. II:  Replicate variability and comparison of
   studies that agree and disagree.  Statistics in Medicine, 6,
   733-744.

Chalmers, T. C., Levin, H., Sacks, H. S., Reitman, D., Berrier, J.,
   Nagalingam, R. (1987b).  Meta-analysis of clinical trials as a
   scientific discipline. I:  Control of bias and comparison with
   large co-operative trials.  Statistics in Medicine, 6, 315-325.

Chalmers, T. C., Smith, H. Jr., 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.

Collins, R., Gray, R., Godwin, J., & Peto, R. (1987).  Avoidance of
   large biases and large random errors in the assessment of moderate
   treatment effects:  the need for systematic overviews.  Statistics
   in Medicine, 6, 245-250.

Cone, L. A. (1982).  A survey of prospective, controlled clinical
   trials of gentamicin, tobramycin, amikacin, and netilmicin.
   Clinical Therapeutics, 5, 155-162.

Cooper, K., & Bennett, W. M. (1987).  Nephrotoxicity of common drugs
   used in clinical practice.  Archives of Internal Medicine, 147,
   1213-1218.

Darr, M. S. , & Elenbaas, J. K. (1982).  Update on drug therapy: IX.
   Gentamicin and tobramycin nephrotoxicity.  Journal of the American

<u>Medical Womens Association</u>, <u>36</u>, 310-316.

Demets, D. L. (1987). Methods for combining randomized clinical trials: strengths and limitations. <u>Statistics in Medicine</u>, <u>6</u>, 341-348.

DerSimonian, R., & Laird, N. (1986). Meta-Analysis in clinical trials. <u>Controlled Clinical Trials</u>, <u>7</u>, 177-188.

Donta, S. T., & Lembke, L. A. (1985). Comparatived effects of gentamicin and tobramycin on excretion of N-acetyl-beta-D-Glucosaminidase. <u>Antimicrobial Agents and Chemotherapy</u>, <u>28</u>, 500-503.

Einarson, T. R., McGhan, W. F., Bootman, J. L., & Sabers, D. L. (1985). Meta-analysis: quantitative integration of independents research results. <u>American Journal of Hospital Pharmacy</u>, <u>42</u>, 1957-1964.

Eysenck, H. J. (1978). An exercise in mega-silliness [Letter to the editor]. <u>American Psycologist</u>, <u>33</u>, 517.

Excerpta Medica. (1984). <u>Guide to the classification and indexing system</u> (3rd ed.). Amsterdam: Elsevier Science Publishers B.V.

Fee, W. E., Vierra, V., & Lathrop, G. R. (1978). Clinical evaluation of aminoglycoside toxicity: tobramycin versus gentamicin, a preliminary report. <u>Journal of Antimicrobial Chemotherapy</u>, <u>4</u>(Suppl. A), 31-36.

Feig, P. U., Mitchell, P. P., Abrutyn E., Brock, S. M., Carney, W. R., Graeber, C. W., Horak, E., Lyons, R. W., & Maher, J. F. (1982). Aminoglycoside nephrotoxicity: a double blind prospective randomized study of gentamicin and tobramycin. <u>Journal of Antimicrobial Chemotherapy</u>, <u>10</u>, 217-226.

Fiske, D. W. (1983). The meta-analytic revolution in outcome research. <u>Journal of Consulting and Clinical Psychology</u>, <u>51</u>, 65-70.

Fong, I. W., Fenton, R. S., & Bird, R. (1981). Comparative toxicity of gentamicin versus tobramycin: a randomized prospective study. <u>Journal of Antimicrobial Chemotherapy</u>, <u>7</u>, 81-88.

Furberg, C. D., & Morgan, T. M. (1987). Lessons from overviews of cardiovascular trials. <u>Statistics in Medicine</u>, <u>6</u>, 295-303.

Fye, W. B. (1987). The literature of American internal medicine. <u>Annals of Internal Medicine</u>, <u>106</u>, 451-460.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. <u>Educational Research</u>, <u>5</u>, 3-8.

Glass, G. V., & Kleigl, R. M. (1983). An apology for research integration in the study of psychotherapy. Journal of Consulting and Clinical Psychology, 51, 28-41.

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

Godfrey, K. (1985). Simple linear regression in medical research. New England Journal of Medicine, 313, 1629-1636.

Goodwin, C. S. (1980). Comparative nephrotoxicity of gentamicin and tobramycin. Current chemotherapy and infectious disease: Proceedings of the 11th International Congress of Chemotherapy and the 19th Interscience Conference on Antimicrobial Agents and Chemotherapy (pp. 612-613). Washington, D.C.: American Society for Microbiology.

Gruppo Italiano Per Lo Studio Della Streptochinasi Nell'Infarto Miocardico (GISSI). (1986). Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Lancet, 1, 397-401.

Hedges, L. V. (1987). Commentary. Statistics in Medicine, 6, 381-385.

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

Hendrickson, N. M., & Amerson, A. B. (1986). Evaluation of drug reviews. American Journal of Hospital Pharmacy, 43, 2468-2473.

Hennekens, C. H., Buring, J. E., & Hebert, P. R. (1987). Implications of overviews of randomized trials. Statistics in Medicine, 6, 397-402.

Horwitz, R. I. (1987). The experimental paradigm and observational studies of cause-effect relationships in clinical medicine. Journal of Chronic Disease, 40, 91-99.

Hubler, G. L. (Ed.). (1984, May). Northwestern Memorial Hospital Pharmacy and Therapeutics Bulletin. (Available from Northwestern Memorial Hospital Department of Pharmacy, Superior Street and Fairbanks Court, Chicago, Illinois 60611.

ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. (1988). Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. Lancet, 2, 349-360.

Itsarayoungyuen, S., Riff, L., Schauf, V., Hamilton, L., Otrembiak, J., & Vidyasagar, D. (1982). Tobramycin and gentamicin are equally safe

for neonates:   results of a double-blind randomized trial with quantitiative assessment of renal function.  Pediatric Pharmacology, 2, 143-155.

Kahlmeter, G., Hallberg, T., & Kamme, C. (1978).  Gentamicin and tobramycin in patients with various infections - nephrotoxicity. Journal of Antimicrobial Chemotherapy, 4(Suppl. A), 47-52.

Kahlmeter, G., & Dahlager, J. I. (1984).  Aminoglycoside toxicity - a review of clinical studies published between 1975 and 1982. Journal of Antimicrobial Chemotherapy, 13(Suppl. A), 9-22.

Keys, T. F., Kurtz, S. B., Jones, J. D., & Muller, S. M.   (1981). Renal toxicity during therapy with gentamicin or tobramycin.  Mayo Clinic Proceedings, 56, 556-559.

Kruse, K. W. (1983).  Online searching of the pharmaceutical literature.  American Journal of Hospital Pharmacy, 40, 240-253. Kumin, G. D. (1980).  Clinical nephrotoxicity of tobramycin and gentamicin:  a prospective study.  Journal of the American Medical Association, 244, 1808-1810.

L'Abbe K. A., Detsky, A. S., & O'Rourke, K. (1987). Meta-analysis in clinical research.  Annals of Internal Medicine, 107, 224-233.

Leviton, L. C., & Cook, T. D. (1981).  What differentiates meta-analysis from other forms of review.  Journal of Personality, 49, 231-236.

Lietman, P. S., & Smith C. R. (1983).  Aminoglycoside nephrotoxicity in humans.  Reviews of Infectious Diseases, 5(Suppl. 2), S284- S293.

Light, R. J. (1987).  Accumulating evidence from independent studies: what we can win and what we can lose.  Statistics in Medicine, 6, 221-231;

Littenberg, B. (1988).  Aminophylline treatment in severe, acute asthma.  Journal of the American Medical Association, 259, 1678-1684.

Louis, T. A., Fineberg, H. V., & Mosteller, F. (1985).  Findings for public health from meta-analysis.  Annual Review of Public Heath, 6, 1-20.

Madsen, P. O., Kjaer, T. B., & Mosegaard, A.   (1976).  Comparison of tobramycin and gentamicin in the treatment of complicated urinary tract infections.  Journal of Infectious Diseases, 134(Suppl.), S150-S152.

Matzke, G. R., Lucarotti, R. L., & Shapiro, H. S.   (1983).  Controlled comparison of gentamicin and tobramycin nephrotoxicity.  American Journal of Nephrology, 3, 11-17.

Meyer, R. D. (1986). Risk factors and comparisons of clinical nephro-
    toxicity of aminoglycosides. American Journal of Medicine,
    80(Suppl. 6B), 119-125.

Mintz, J. (1983). Integrating research evidence: a commentary on
    meta-analysis. Journal of Counseling and Clinical Psychology, 51,
    71-75.

Moore, R. D., Smith, C. R., Lipsky, J. J., Mellits, E. D., & Lietman,
    P. S. (1984). Risk factors for nephrotoxicity in patients treated
    with aminoglycosides. Annals of Internal Medicine, 100, 352-357.

Morgan, D. B., & Will, E. J. (1983). Selection, presentation, and
    interpretation of biochemical data in renal failure. Kidney
    International, 24, 438-445.

Mulrow, C. D. (1987). The medical review article: state of the
    science. Annals of Internal Medicine, 106, 485-488.

Mulrow, C. D., Thacker, S. B., & Pugh, J. A. (1988). A proposal for
    more informative abstracts of review articles. Annals of Internal
    Medicine, 108, 613-615.

National Library of Medicine. (1987). Medical Subject Headings,
    Annotated Alphabetic List, 1987 (NLM Report No. NLM-MED 86-03).
    Bethesda, Maryland: Author.

O'Brien, P. C., & Shampo, M. A. (1981). Comparing two proportions:
    the relative deviate test and chi-square equivalent. Mayo Clinic
    Proceedings, 56, 513-515.

Pancoast, S. J. (1988). Aminoglycoside antibiotics in clinical use.
    Medical Clinics of North America, 72, 581-612.

Perry, C. A. (1986). Online information retrieval in pharmacy and
    related fields. American Journal of Hospital Pharmacy, 43,
    1509-1524.

Peto, R. (1987). Why do we need systematic overviews of randomized
    trials? Statistics in Medicine, 6, 233-240.

Ravel, R. (1978). Clinical Laboratory Medicine. Chicago: Year Book
    Medical Publishers.

Rosenthal, R. The "file drawer problem" and tolerance for null
    results. Psychological Bulletin, 86, 638-641.

Pancorbo, S., Compty, C., & Heissler, J. (1982). Comparison of
    gentamicin and tobramycin nephrotoxicity in patients receiving
    individualized-pharmacokinetic dosing regimens. Biopharmaceutics
    and Drug Disposition, 3, 83-88.

Sacks, S. S., Berrier, J., Reitman, D., Ancona-Berk, V. A., & Chalmers, T. C. (1987). Meta-analysis of randomized controlled trials. New England Journal of Medicine, 316, 450-455.

Schentag, J. J. (1983). Specificity of renal tubular damage criteria for aminoglycoside nephrotoxicity in critically ill patients. Journal of Clinical Pharmacology, 23, 473-483.

Schentag, J. J., Plaut M. E., & Cerra, F. B. (1981). Comparative nephrotoxicity of gentamicin and tobramycin: pharmacokinetic and clinical studies in 201 patients. Antimicrobial Agents and Chemotherapy, 19, 859-866.

Smith, C. R., Lipsky, J. J., Laskin, O. L., Hellmann, D. B., Mellits, E. D., Longstreth, J., & Lietman, P. S. (1980). Double-blind comparison of the nephrotoxicity and auditory toxicity of gentamicin and tobramycin. New England Journal of Medicine, 302, 1106-1109.

Smith, C. R., & Lietman, P. S. (1982). Comparative clinical trials of aminoglycosides. In A. Whelton & H. C. Neu (Eds.), The Aminoglycosides (pp. 497-509). New York: Marcel Dekker.

Strube, M. J., & Hartmann, D. P. (1983). Meta-analysis: Techniques, applications, and functions. Journal of Consulting and Clinical Psychology, 51, 14-27.

Thacker, S. B. (1988). Meta-analysis: a quantitative approach to research integration. Journal of the American Medical Association, 259, 1685-1689.

Tousignaut, D. R. (Ed.). (1987). IPA Thesaurus and Frequency List (4th ed.). Bethesda: American Society of Hospital Pharmacy.

Wade, J. C., Petty, B. G., Conrad, G., Smith, C. R., Lipsky, J. J., Ellner, J., & Lietman, P. S. (1978). Cephalothin plus an amino-glycoside is more nephrotoxic than methicillin plus an aminoglycoside. Lancet, 2, 604-606.

Walker, A. M., Martin-Moreno, J. M., & Artalejo, F. R. (1988). Odd man out: a graphical approach to meta-analysis. American Journal of Public Health, 78, 961-966.

Walker, B. D., & Gentry, L. O. (1976). A randomized, comparative study of tobramycin and gentamicin in treatment of acute urinary tract infections. Journal of Infectious Diseases, 134(Suppl.), S146-S149.

Wilkinson, L. (1985). Systat [computer program]. Evanston, IL: Systat, Inc. (Version 2).

Wittes, R. E. (1987). Problems in the medical interpretation of

overviews.  <u>Statistics in Medicine</u>, <u>6</u>, 269-276.

Yusuf, S. (1987).  Obtaining medically meaningful answers from an
    overview of randomized clinical trials.  <u>Statistics in Medicine</u>, <u>6</u>,
    281-286.

Yusuf, S., Collins, R., Peto, R., Furberg, C., Stampfers, M. J.,
    Goldhaber, S. Z., & Hennekens, C. H. (1985).  Intravenous and intra-
    coronary fibrinolytic therapy in acute myocardial infarction:
    overview of results on mortality, reinfarction and side-effects from
    33 randomized controlled trials.  <u>European Heart Journal</u>, <u>6</u>,
    556-585.

## Approval Sheet

The thesis submitted by J. Russell Teagarden has been read and approved by the following committee:

Dr. Jack A. Kavanaugh, Director
Professor of Curriculum and Human Resource Development

Dr. Ronald R. Morgan
Associate Professor and Director of School of Psychology Program

Ms. Michelle Parker
Nurse Research Clinician
Northwestern Memorial Hospital, Chicago, Illinois

The final copies have been examinied by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval by the Committee with reference to content and form.

The thesis is therefore accepted in partial fullfillment of the requirements for the degree of Master of Arts.

_____11/20/88_____          _____*Jack A. Kavanagh*_____
Date                            Director's Signature