State of the Evidence: Current Status and Prospects of Meta-Analysis in Infectious Diseases

John P. A. Ioannidis\(^1\)\(^2\) and Joseph Lau\(^2\)

Meta-analysis is increasingly applied in infectious diseases to summarize clinical data and to evaluate the strength, diversity, and deficiencies of evidence for medical questions of interest. We present an overview of the current status of meta-analysis in the area of infectious diseases and the lessons learnt from its applications. Recently published meta-analyses show that several important areas of research on infectious diseases lack sufficient randomized evidence. Often evidence is scattered across a large number of small trials, making meta-analysis a promising way to integrate diverse results. Quality of trials in the field is often poor. There are several examples where evidence was accumulated primarily for marketing rather than for scientific purposes. Finally, meta-analyses are also raising the problem of what constitutes clinically significant treatment benefits, as well as interesting issues about the reproducibility of clinical evidence and its evolving nature. The increasing applications of meta-analytic methods in the study of infectious diseases may enhance data sharing and international collaborations.

Recently there has been a lot of interest, discussion, and debate about “evidence-based medicine” [1–3]. Evidence-based medicine may be defined as the systematic, quantitative, preferentially experimental approach to obtaining and using medical information. It places more emphasis on data from randomized, experimental trials than on data from other types of studies, such as case reports, case series, case-control, and cohort studies. It also emphasizes systematic integration of information on each clinical problem. This means that in evidence-based medicine, meta-analysis [4], the scientific method of critically appraising and quantitatively summarizing data from diverse studies, assumes a prominent role, next and in parallel to clinical trials.

Randomized clinical trials already have a half-century history in research on infectious diseases. Some of the first randomized clinical trials targeted infectious disease questions such as the management of hepatitis or chemotherapy for tuberculosis. Some of the first meta-analyses were also performed in the infectious disease field. The first precursor of meta-analysis (published in 1908 by Pearson in the British Medical Journal) was a quantitative summary of the relationship between typhoid vaccine inoculation status and the incidence and fatality of the disease [5].

In one of the first meta-analyses of randomized controlled trials in the modern era, Baum et al. showed in 1981 that strong evidence had accumulated to suggest that antibiotic prophylaxis was essential in reducing postoperative morbidity from colon surgery [6]. Furthermore, as shown in figure 1, enough substantial evidence had accumulated by the early 1970s to allow a definitive decision on the efficacy of prophylaxis. For 15 more years patients continued to be randomized to no-prophylaxis arms in randomized trials, although meta-analysis would have suggested that this was inefficient and unethical.

Many meta-analyses have already been performed on infection-related topics, and the specialist is likely to be increasingly exposed to this new methodology in the future. In this article we discuss some key issues that stem from an examination of current meta-analyses in infectious diseases and what they have to teach us from their strengths and deficiencies. We have selected to focus on a few issues of particular importance for the scientific and clinical progress of this field.

Quantity of the Evidence

Although it is hard to define what constitutes “enough” randomized evidence, several recently published meta-analyses in infectious diseases (1997–1998) suggest that randomized evidence is sometimes sparse or even nonexistent. Several “attempted” meta-analyses have failed to identify any randomized evidence for important questions and have resorted to compiling observational data. For example, there has been no randomized trial performed to determine whether antibiotics have a role in the prophylactic management of CSF fistulae [7]; or whether metronidazole is teratogenic [8]; or whether 3 days of antibiotic therapy for acute otitis media are as effective as longer courses of treatment [9]. There are numerous other examples, and potential topics extend beyond the

\(^{1}\)Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece; and \(^{2}\)Division of Clinical Care Research, Department of Medicine, New England Medical Center Hospitals, Tufts University School of Medicine, Boston, Massachusetts
Figure 1. Standard and cumulative meta-analyses of antibiotic prophylaxis compared with no treatment (controls) in colon surgery [6]. Wound infection is analyzed by outcome, and data have been updated since original meta-analysis. Risk ratio <1 favors use of prophylactic antibiotics, whereas risk ratio >1 suggests that no treatment is better. Left panel. Studies displayed chronologically by year of publication; for each, point estimate and 95% confidence interval are shown. Pooled results from all studies are shown at bottom (n represents study size). Right panel. Cumulative meta-analysis of same studies reveals that antibiotic prophylaxis efficacy could have been identified back in 1971, after 5 studies involving about 300 patients (n in this panel represents cumulative number of patients from included studies). Meta-analysis calculations are based on random-effects model.

questions targeted by “attempted” meta-analyses. In fact, for many important questions there will never be any meta-analyses, simply because there are no data to analyze.

The reasons for the lack of randomized evidence may be manifold. One is that entrenched perceptions are propagated by dogma with little or no data, such as recommendations for the optimal duration of therapy for otitis media. Such dogma hinders the performance of a randomized trial. Second, feasibility may be an issue. Single investigators and institutions might believe they lack the number of patients necessary for a randomized study design or lack the opportunities or resources for multicenter collaborations; this probably explains why studies of antibiotics for posttraumatic CSF fistulae have not been done.

Third, there may be ethical dilemmas. For interventions perceived to be potentially harmful, even if not supported yet by any strong data (such as the use of metronidazole or many other medications in pregnancy), there may be a perhaps falsely perceived ethical barrier to randomization. Finally, commitment is always important to consider. Clinical trials demand great time and effort and may not be among the best-rewarded academic investments for an investigator.

These considerations pertain to most medical fields. A peculiarity for infectious diseases—compared to other fields such as cardiology and oncology, where clinical trials are more firmly established—is that several infectious syndromes are rare in single practices and institutions. The need to approach uncommon problems through multicenter collaborations in randomized trials cannot be overemphasized. Moreover, there are ways that randomized evidence can be collected and appraised even for very rare syndromes [10]. Having some randomized data is better than having none at all.

Nevertheless, not only for rare syndromes, but even for some very common infectious conditions, the amount of randomized evidence that has been generated to date is relatively sparse. Evidence for diseases affecting patient populations in developing countries tends to be particularly scarce. For example, although about 1 million people, mostly children, die every year of pneumonia complicating measles, a meta-analysis [11] showed that there have been only 6 small trials, all performed a long time ago and with very poor design and analysis standards, to address the issue of whether antibiotics have any prophylactic role in measles.

Methodologists in other fields have commented that patient participation in clinical trials is fairly low; it has been quoted [12], for example, that only ~3% of patients with cancer participate in some clinical trial during their lives. Although the respective figures for infectious diseases are not available, probably the participation rate in clinical trials for infectious disease problems is much lower. Ex-
Quality of the Evidence

The quality of the evidence has sometimes suffered in studies of infectious diseases, although this is far from being an absolute statement. For example, we have typically scored HIV-related trials included in meta-analyses of antiretroviral therapy very high [16, 17]. This is probably a reflection of the fact that HIV research has grown mostly under the collaborative, multicenter trial model and has involved close cooperation between seasoned statisticians and methodologists and HIV clinicians in models such as those sponsored by the NIH under the AIDS Clinical Trials Group and Community Programs for Clinical Research on AIDS Mechanisms.

The quality of a trial is typically judged on the basis of its written report [18], but this may not necessarily reflect the true quality of its design and its conduct [19]. Nevertheless, a trial that is reported with glaring deficiencies probably had at least as many or even more problems and limitations with regard to its design and conduct. However, with the widespread implementation of standardized structured reporting of clinical trials, such as suggested by the CONSORT statement [20], ambiguities in the presentation of data, at least, will be lessened, and we will be able to deal with true defects in quality.

A variety of quality scores and scales have been proposed to evaluate clinical trials [18]. Important components of quality scales and quality scores may include rigorous randomization, double-blinding, allocation concealment, and respect for the intention-to-treat principle. Most trials follow proper randomization schedules, but this may not be the case for several small trials performed by industry-sponsored teams. Double-blinding is not always feasible, but examples abound in which it was not used even though it would have been both easy and important to implement.

More than one-half of the clinical trials included in meta-analyses of otitis media [9] and sinusitis [13] have been unmasked, whereas the subjective interpretation of the clinical response in these common infections makes blinding of paramount importance. Allocation concealment is often not guaranteed in trials, and when it is lacking, blinding is superfluous. An empirical evaluation in a different medical field has shown that lack of allocation concealment may inflate the treatment effect by as much as 40%, and lack of blinding may inflate the observed treatment effect by 20% [21]. With such degrees of bias, a poor treatment may easily approach the range of a major therapeutic breakthrough.

Finally, lack of attention to the intention-to-treat principle has also been common in infectious disease-related trials. Although the on-treatment approach makes sense for the evaluation of toxicity and of biological parameters [22], the use of on-treatment variants for the assessment of clinical efficacy opens the door to bias. The ingenuity of generating on-treatment variants can often be amazing. For example, a trial comparing 2 different antibiotics for acute sinusitis [23] used a modified intention-to-treat approach, an on-treatment approach, an intention-to-treat approach, and, as it becomes apparent while one reads the report, yet another unnamed approach. It is impossible to say from the published report [23] which one is used in each presented analysis.

Timing and Variety of the Evidence

A meta-analysis can be updated each time a new piece of evidence appears on a specific question. In cumulative meta-analysis [24, 25], data are entered successively in the order of their chronological appearance. Keeping a meta-analysis up to date can be a real challenge, especially when a field is moving very quickly. We can mention as an example our personal experience in performing a meta-analysis of antiretroviral changes among patients with prior exposure to antiretroviral treatment [17].

Our database of antiretroviral trials was set up in 1991, with a plan to keep updating meta-analyses on important questions of antiretroviral therapy. A meta-analysis of antiretroviral changes was “completed” by our team in early 1996 and formatted for submission for publication. At that time, the world of antiretroviral therapy was changing rapidly with the advent of protease inhibitors and triple-combination therapy. We felt that the meta-analysis needed updating even before the com-

exceptions exist, such as for human immunodeficiency virus (HIV) infection in the United States (>50,000 of the ∼1 million HIV-infected patients in this country have been randomized to date in National Institutes of Health [NIH]–sponsored randomized trials alone).

However, for several other common diseases, the trial participation rate is a millionfold less than that for cancer research. For example, only 878 patients have been randomized to date to address the efficacy of antibiotics versus placebo in acute community-acquired sinusitis [13], a condition that is most often of viral etiology. In 5 of the 6 pertinent clinical trials, patients were selected for recruitment on the basis of positive findings of culture, radiography, or computed tomography; in only 1 trial [14], with 192 patients, were participants recruited on the basis of the clinical picture alone, the typical means of diagnosis of sinusitis in the community.

One may estimate that in order to have 80% power to detect an improvement in the complete cure rate of 50% to 60% at 1 week, with a type I error (α) of 0.05, a trial would need to include ∼852 patients; for 90% power, ∼1114 patients—or even more to allow for losses to follow-up. In the United States alone, acute community-acquired sinusitis episodes amount to ∼30 million each year [15]. Moreover, even if larger trials were indeed performed, raw numbers of enrolled patients would not always tell the whole story, since the quality of study design, the pertinence of the question being addressed, and the selection of representative patient populations are also very important. Clearly, missed opportunities for obtaining more solid evidence abound in the field of infectious diseases, but many of them should be easy targets in the near future for talented clinical investigators who have been nurtured on the ideal of randomized evidence.
ments of the peer reviewers came back. We kept updating the meta-analysis during and after the peer-review process to keep it up to date.

At the same time, expert guidelines that were published on fast track by the Journal of the American Medical Association [26] soon became obsolete and had to be replaced; admittedly, they were considered obsolete even at the time of their publication. Major randomized trials such as the Delta [27] and ACTG 175 [28] showing the effectiveness of the combination of 2 nucleoside reverse transcriptase inhibitors were obsolete for clinical practice by the time they appeared in the Lancet [27] and the New England Journal of Medicine [28] in late 1996. For the nonexpert who retrieves articles from MEDLINE to manage cases, a naive interpretation of data that are seemingly very recent may even be dangerous in clinical practice.

A big challenge for meta-analysis now and in the future will be to keep calculations and comparisons up to date. Yet, still, how much can we trust evidence that was obtained from “archaeological” clinical trials performed several decades ago? What can we say of trials of antibiotic prophylaxis for measles done in the 1960s [11]? Are antibiotic trials performed in the 1970s, before the widespread acquisition of β-lactamases among common upper respiratory bacteria, relevant for the management of acute sinusitis? The half-life of the truth of randomized evidence is unknown [29], but perhaps in infectious diseases it is often of limited duration.

With changes in background medical management, the risk of severe outcomes is changing rapidly for many diseases. Most antimicrobial therapies depend to some extent on the susceptibility of the pathogens, and shifts in susceptibility patterns over time can be very substantial. Therefore, perhaps meta-analysis should try to adapt to such a picture of changing evidence and show the change in the magnitude of treatment effects over time, rather than simply trying to amass all evidence, giving equal weight to all pieces of data, old and new.

The variability of the treatment effect in different trials over time can be modeled with meta-regression analyses. These are regression analyses in which each study is the unit of information. Timing is only one of several independent variables that may be considered. For example, meta-regression analyses have been used to describe the decrease in the clinical effect of zidovudine monotherapy as a function of duration of follow-up in HIV-infected patients [16] and the dose-dependence of discontinuations of trimethoprim-sulfamethoxazole prophylaxis for Pneumocystis carinii pneumonia [30]. In these examples, duration of follow-up and dose are the independent variables of the respective meta-regressions.

A particularly interesting variable to consider is the risk of the major outcomes of the studied disease [31, 32]: is the treatment effect different in patient groups with different risks of getting the disease? An evaluation of >100 meta-analyses showed that the relative benefit varies significantly with the level of risk in about 14% of the cases [33]. By bringing together evidence from populations with diverse levels of risk, meta-analysis may also bring clinical evidence closer to clinical decision-making, where physicians take into consideration the patient’s risk before taking a course of action.

Ideally, meta-analyses also use individual-patient data from each of the combined trials. Data on predictors of risk may be used to generate risk scores at the individual-patient level [34]. Meta-analysis of individual-patient data promises improvements over meta-analysis of summary data in some situations, and there are already examples of studies of infectious diseases that apply this method [35, 36].

Rationale for the Evidence: Marketing Versus Medicine

Several meta-analyses on important clinical topics suggest that randomized trials in infectious disease therapeutics are often performed to assess, if not clearly promote, the role of specific antimicrobial agents, rather than to address the management of a clinical disease entity. The latter may often require the use of multiple agents in combination or sequential strategies, as in HIV therapeutics, or the use of a wide variety of different agents, each with a typical spectrum of activity, often overlapping one another, but potentially of variable cost and toxicity.

Figure 2 shows 80 randomized comparisons of antibiotics among themselves and with placebo from an evidence report of therapeutic choices for immunocompetent patients with uncomplicated acute sinusitis [15]. This seems to be a lot of evidence, but it is so fragmented that it is almost impossible to make sense of. An extensive variety of agents have been tested, including some (such as third-generation cephalosporins) that would probably be an extreme choice for this common, mild, often self-resolving clinical syndrome. Since most of these small to moderate-sized clinical trials are sponsored by the industry, with a clear financial incentive, it is not surprising that evidence is generated to support drug marketing rather than help physicians in medical decision-making. A meta-analysis suggests that there is currently no evidence that newer, broader-spectrum antibiotics are any better than amoxicillin in terms of clinical response [14].

Another example of fragmented evidence comes from multiple small, industry-sponsored trials of medical treatment for eradication of Helicobacter pylori. A recent meta-analysis [37] compiled eradication rates across 119 studies with a total of 6416 patients, an average of only 54 patients per study. A sample size of 54 is about 1% of the sample size required for a trial to be powered adequately enough to show equivalence among such highly effective regimens. By piecing together the big picture, meta-analysis can offer, to some extent, what these single studies have not. Of course, meta-analysis still will not help if the questions asked are not sensible.

A second corollary of the “evidence for marketing” approach is that the design or reporting of trials during and after drug
development may be subtly manipulated so as to favor the sponsored pharmaceutical agent. A meta-analysis [38] of trials comparing carbapenems against other antibiotics for the treatment of severe intra-abdominal infections showed clearly that carbapenems were more effective than the comparator agents in early published trials, but this trend completely disappeared in subsequent trials. This might possibly have been a chance finding or a true biological phenomenon (e.g., related to the development of resistance); but this evolution may reflect the fact that early trials used less potent comparator agents, whereas subsequent trials used more appropriate comparators.

When the relative efficacy of a treatment changes over time, one should also consider the possibility of publication lag and publication bias affecting the strength and the direction of the evidence. Publication lag [39] refers to delays in completion, submission after completion, and publication after submission for the reports of trials with negative, non-statistically significant results, as compared with trials that show that a new treatment is significantly superior to its control comparator. A detailed study among all 109 efficacy trials conducted by National Institute of Allergy and Infectious Diseases-sponsored investigators in the domain of HIV infection showed that negative trial findings were published 3.8 fold more slowly than positive ones [39].

In some cases, negative trial findings may never be published, resulting in a distorted picture of the evidence and overt publication bias [40]. Publication bias means that negative trial findings not only are delayed from appearing but never appear in print. Retrospective identification of publication bias is difficult, although some diagnostic approaches [41] have been proposed. Ideally, prospective registration of all trials in a given field should be the best way to eliminate the problem [42].

---

**Figure 2.** Randomized clinical trials of antibiotic treatment of acute sinusitis [15]. Number in each cell corresponds to number of randomized comparisons of respective antibiotics (figure modified from Agency for Health Care Policy and Research).
Replication of Meta-Analyses on the Same Topic

A new phenomenon in the field of evidence-based medicine is the appearance of several meta-analyses that address the same or similar questions. This should not be surprising, since meta-analysis has become a competitive field of medical research in which several teams may be working on similar projects. We recently witnessed this with regard to the question of whether single daily doses of aminoglycosides are as effective and safe as multiple daily doses of these drugs. This question has fueled a scientific and clinical debate for >20 years. A number of clinical trials have tried to address the relative merits of the 2 dosing regimens, but none was definitively powered to provide a final clinical answer.

Between 1996 and 1997, at least 8 meta-analyses appeared on this same topic [43–50]. A careful examination of these independent research efforts shows that their protocols were not all identical; for example, they differed in their criteria for inclusion and exclusion of studies. Sometimes they also differed in their choice of how to measure effects and in their methods of statistical analysis. Their final results and conclusions were very similar: all concluded that single daily doses seemed to be at least as good or possibly better than multiple doses. Nevertheless, the levels of statistical significance of the individual trials for similar endpoints were on both sides of the traditional $P = .05$ value.

Clinical trials are usually designed and interpreted in such a way that $P$ values above and below .05 distinguish whether the original null hypothesis is rejected or not. Meta-analyses, on the other hand, should be interpreted with emphasis more on the magnitude of the effect given by the point estimate, and the uncertainty around it, as suggested by the confidence intervals [51]. Since there is no equivalent to sample size and power calculations, meta-analysis is continuously open to updating with new pieces of evidence, as new trials are designed and performed on the same topic.

Conclusion: Prospective Directions

Published meta-analyses of studies of infectious diseases clearly reveal specific strengths and weaknesses of evidence in the field. The typical situation is that more and better evidence is needed. In many cases where understanding of the biology or of therapeutics is advancing at a rapid pace, new evidence needs to be obtained to supplement old evidence. Meta-analysis can show us what we already know, what we do not know, how much more and what kind of evidence we need, and whether pieces of evidence diverge and for what reason [52].

Perhaps more important, meta-analysis is evolving along with and promoting the development of a new global conscience in clinical research. There are very few clinical questions that are currently studied only by a single team. The same or similar questions are typically studied by several groups of investigators around the world at about the same time and/or successively. The answers obtained by these independent efforts are likely to be juxtaposed at some time. With the knowledge that such informal or formal meta-analyses are going to happen anyhow, it is probably better for investigators to try to anticipate them ahead of time [53] and to perform them in a scientifically robust, systematic way, rather than leave them to subjective a posteriori interpretation and the capriciousness of industry sponsors.

The International Cochrane Collaboration was launched 6 years ago as an effort to generate and disseminate regularly updated systematic reviews and meta-analyses in all disciplines of health care in medicine [54]. The effort has been heralded as equivalent in scope to the Human Genome Project [55]. An HIV/AIDS collaborative review group was launched in 1998 within the Collaboration, with the aim of performing meta-analyses on HIV therapeutics, prognostics, diagnostics, behavioral interventions, prevention, and health care delivery. Other groups within the Collaboration are already targeting other infectious disease disciplines such as tuberculosis, vaccines, and sexually transmitted diseases.

At the same time, we would like to see more examples of prospective meta-analyses in which a cluster of clinical trials is designed and launched with the explicit goal that they will be analyzed together by meta-analysis upon completion. This practice is still in its infancy in the area of infectious diseases, but in most cases of therapeutic development trials, prospective meta-analysis is a large missed opportunity. The current paradigm is that a company will sponsor several trials around the globe, typically more than a dozen and sometimes, such as in the recent development of sparfloxacin, more than a hundred. Apparently, the aim is to run many trials to ensure enough successes to satisfy the various regulatory requirements of different countries. Moreover, the first “big successes” are also selectively propagated at influential meetings and further disseminated by the attendant press coverage. It would have been much more scientifically robust to explicitly design the developmental plan for new therapeutic agents as prospective meta-analyses where all interested investigator groups around the globe would be able to contribute their data, with the explicit goal of weighing all pieces of evidence objectively and systematically [56].

Advancement of transparent, global collaboration and of data sharing are two of the immediate benefits that we can reap by promoting evidence-based standards and an all-inclusive philosophy of meta-analysis in the future.

References

39. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281–6.
50. Bailey TC, Little JR, Littenberg B, Reichley RM, Dunagan WC. A meta-