Chapter 1

Introduction to Evidence-based Infectious Diseases

Dominik Mertz, Nick Daneman, and Fiona Smaill

The purposes of this first chapter are to provide brief overviews of the scope of the third edition of this book as well as evidence-based infectious diseases (EBID) practice, and to introduce the approach we implemented to reflect the level of evidence supporting recommendations made in this book.

1.1 What is Evidence-based Medicine?

Evidence-based medicine was born in the 1980s of the last century [1,2]. David Sackett, the founding chair of the Department of Clinical Epidemiology and Biostatistics at McMaster University, defined *evidence-based medicine* as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of patients" [3]. One of the key aspects of evidence-based medicine is a focus on randomized clinical trials (RCTs) for assessing treatment, which now is a standard requirement for the licensing of new therapies.

1.2 Evidence-based Infectious Diseases (EBID)

The field of infectious diseases, or more accurately the importance of illness due to infections, played a major role in the development of epidemiological research in the 19th and early 20th centuries. Classical observational epidemiology was derived from studies of epidemics—infectious diseases such as cholera, smallpox, and tuberculosis. Classical epidemiology was nevertheless action-oriented. For example, John Snow's observations regarding cholera led to his removal of the Broad Street pump handle in an attempt to reduce the incidence of cholera. Pasteur, on developing an animal vaccine for anthrax, vaccinated a number of animals with members of the media in attendance [4]. When unvaccinated animals subsequently died, while vaccinated animals did not, the results were immediately reported throughout European newspapers.

In the era of clinical epidemiology, it is notable that the first true RCT is widely attributed to Sir Austin Bradford Hill's 1947 study of streptomycin for tuberculosis [5]. In subsequent years, and long before the "large simple trial" was rediscovered by the cardiology community, large-scale trials were carried out for polio prevention as well as tuberculosis prevention and treatment.

Infectious diseases were at the frontiers of both classical and clinical epidemiology, but is current infectious diseases practice evidence-based? We believe the answer is "somewhat." We have excellent evidence for the efficacy and side effects of many modern vaccines and antiviral drugs for treatment of HIV and Hepatitis C. Furthermore, non-inferiority

Evidence-Based Infectious Diseases, Third Edition. Edited by Dominik Mertz, Fiona Smaill, and Nick Daneman.

trials are mandatory for new antibiotics to receive approval from the FDA and other regulatory authorities for specific indications. This being said, the current use of many anti-infectives are not supported by high-level RCT data, and head-to-head comparisons of different anti-infectives and/or durations of treatment are largely missing. Thus, the acceptance of before-and-after data to prove the efficacy of antibiotics for syndromes such as bacterial meningitis is ethically appropriate and recommended in guidelines despite the fact that no RCT data exists. Therefore, it is not surprising that recommendations in Infectious Diseases Society of America (IDSA) guidelines are primarily based on low-quality evidence derived from non-randomized studies or expert opinion [6].

Furthermore, in treating many common infectious syndromes—from sinusitis and cellulitis to pneumonia—we have many very basic diagnostic and therapeutic questions that have not been optimally answered. How do we reliably diagnose pneumonia? Which antibiotic is most effective and cost-effective? Can we improve on the impaired quality of life that often follows such infections as pneumonia? Furthermore, there may not be a single "best" antibiotic for pneumonia, in contrast to treatment algorithms for myocardial infarction that apply uniformly to the majority of patients. Much of the "evidence" that guides therapy in infectious diseases, particularly for bacterial diseases, may not be clinical, but exists in the form of a sound biologic rationale, the activity of the antimicrobial against the offending pathogen, and the penetration at the site of infection (pharmacodynamics and pharmacokinetics). Still, despite having a sound biologic basis for choice of therapy, there are many situations where better RCTs need to be conducted and where clinically important outcomes, such as symptom improvement and health-related quality, are measured.

How, then, can we define *EBID*? Paraphrasing David Sackett, *EBID* may be defined as "the explicit, judicious and conscientious use of current best evidence from infectious diseases research in making decisions about the prevention and treatment of infection of individuals and populations." It is an attempt to bridge the gap between research evidence and the clinical practice of infectious diseases. Such an "evidence-based approach" may include critically appraising evidence for the efficacy and safety of a treatment option. However, it may also involve finding the best evidence to support (or refute) use of a diagnostic test to detect a potential pathogen. Additionally, EBID refers to the use of the best evidence to estimate prognosis of an infection or risk factors for the development of infection. EBID therefore represents the application of research findings to help answer a specific clinical question. In so doing, it is a form of knowledge transfer, from the researcher to the clinician. It is important to remember that use of research evidence is only one component of good clinical decision-making. Experience, clinical skills, and a patient-centered approach are all essential components. EBID serves to inform the decision-making process. For the field of infectious diseases, a sound knowledge of antimicrobials and microbiologic principles are also needed.

1.3 Posing a Clinical Question and Finding an Answer

The first step in practicing EBID is posing a clinically driven and clinically relevant question. To answer a question about diagnosis, therapy, prognosis, or causation, we can begin by framing the question [2]. The question usually includes a brief description of the patients, the intervention or exposure, the comparison, and the outcome (PICO). For example, if asking about the efficacy of antimicrobial-impregnated catheters in intensive care units [7], the question can be framed as follows: "In critically ill patients, does the use of antibiotic-impregnated catheters, compared with regular vascular catheters, reduce central line associated infections?" After framing the question, the

second step is to search the literature. The most time-efficient approach is to search for evidence-based synopses and systematic reviews in a first step. Systematic reviews can be considered as concise summaries of the best available evidence that address sharply defined clinical questions. If there are no synopses or systematic reviews that can answer the clinical question, the next step is to search the primary literature itself, which, of course, is much more time-consuming. After finding the evidence the next step is to critically appraise it.

1.4 Evidence-based Diagnosis

Let us consider the use of a rapid antigen detection test for group A streptococcal infection in throat swabs. The first question to ask is whether there was a blinded comparison against an accepted reference standard. By *blinded*, we mean that the measurements with the new test were done without knowledge of the results of the reference standard.

Next, we would assess the results. Traditionally, we are interested in the sensitivity (proportion of reference-standard positives correctly identified as positive by the new test) and specificity (the proportion of reference-standard negatives correctly identified as negative by the new test).

Ideally, we would also like to have a measure of the precision of this estimate, such as a 95% confidence interval on the sensitivity and specificity, although such measures are unfortunately rarely reported in the infectious diseases literature.

Note, however, that while the sensitivity and specificity may help a laboratory to choose the best test to offer for routine testing, they do not necessarily help the clinician manage the patient. Thus, faced with a positive test with known 95% sensitivity and specificity, we cannot infer that our patient with a positive test for group A streptococcal infection has a 95% likelihood of being infected. For this, we need a positive predictive value, which is calculated as the percentage of true positives among all those who test positive. If the positive predictive value is 90%, then a positive test would suggest a 90% likelihood that the person is truly infected. Similarly, the negative predictive value is the percentage of true negatives among all those who test negative. Both positive and negative predictive value change with the underlying prevalence of the disease, hence such numbers cannot be generalized to other settings.

A more sophisticated way to summarize diagnostic accuracy, which combines the advantages of positive and negative predictive values while solving the problem of varying prevalence, is to quantify the results using likelihood ratios. Like sensitivity and specificity, likelihood ratios are a constant characteristic of a diagnostic test and independent of prevalence. However, to estimate the probability of a disease using likelihood ratios, we additionally need to estimate the probability of the target condition (based on prevalence or clinical signs). Diagnostic tests then help us to shift our suspicion (pretest probability) about a condition depending on the result. Likelihood ratios tell us how much we should increase the probability of a condition for a positive test (positive likelihood ratio) or reduce the probability for a negative test (negative likelihood ratio). More formally, likelihood ratio positive (LR+) and negative (LR–) are defined as:

$$\mathrm{LR+} = \frac{\begin{array}{l}\text{odds of a positive test}\\ \text{in an individual with}\\ \text{the condition}\end{array}}{\begin{array}{l}\text{odds of a positive test}\\ \text{in an individual}\\ \text{without the condition}\end{array}}. \qquad (1.1)$$

$$\mathrm{LR-} = \frac{\begin{array}{l}\text{odds of a negative test}\\ \text{in an individual with}\\ \text{the condition}\end{array}}{\begin{array}{l}\text{odds of a negative test}\\ \text{in an individual without}\\ \text{the condition}\end{array}}. \qquad (1.2)$$

A positive likelihood ratio is also defined as sensitivity/(1 – specificity), and the negative likelihood ratio as (1 – sensitivity)/specificity.

Having found that the results of the diagnostic test appear favorable for both diagnosing or ruling out disease, we ask whether the results of a study can be generalized to our patients. We might also call this "external validity" or "generalizability" of the study. Here, we are asking the question: "Am I likely to get the same results as in this study in my own patients?" This includes such factors as the severity and spectrum of patients studied, technical issues in how the test is performed outside the research setting, but also the epidemiology of pathogens in your area that affects pre-test probabilities—a unique additional challenge we face in infectious diseases.

Important caveats, however, are that (a) there may be no appropriate reference standard, and (b) the spectrum of illness may dramatically change the test characteristics, as may other co-interventions such as antibiotics. For example, let us assume that we are interested in estimating the diagnostic accuracy of a new commercially available polymerase chain reaction (PCR) test for the rapid detection of *Neisseria meningitidis* (*N. meningitidis*) in spinal fluid. The reference standard of culture may not be completely sensitive. Therefore, use of an expanded reference ("gold") standard might be used. For example, the reference standard may be growth of *N. meningitidis* from the spinal fluid, demonstration of an elevated white blood cell count in the spinal fluid along with gram-negative bacilli with typical morphology on Gram stain, or elevated white blood cell count along with isolation of *N. meningitidis* in the blood. It is also important to know in what type of patients the test was evaluated, such as the inclusion and exclusion criteria, as well as the spectrum of illness. Given that growth of microorganisms is usually progressive, test characteristics in infectious diseases can change depending when the tests are conducted. For example, PCR conducted in patients who are early in their course of meningitis may not be sensitive as compared to patients who presented with late-stage disease.

1.5 Evidence-based Treatment

The term *evidence-based medicine* has become largely synonymous with the dictum that only double-blinded RCTs give reliable estimates of the true efficacy of a treatment. For the purposes of guidelines, "levels of evidence" have been proposed, with a hierarchy from large to small RCTs, prospective cohort studies, case-control studies, and case series. In newer iterations of these "levels of evidence," a meta-analysis of RCTs (without statistical heterogeneity, indicating that the trials appear to be estimating the same treatment effect), are touted as the highest level of evidence for a therapy.

In general, clinical questions about therapy or prevention are best addressed through RCTs. In observational studies, the choice of treatment may have been influenced by extraneous factors that influence prognosis (so-called "confounding factors"). One of the most important confounding factors when comparing treatment options in an observational study is confounding by indication, that is, the treatment decision is made based on how the patient presents. For example, patients who appear more severely sick may receive predominantly treatment A as the treating physicians believe that treatment A is better than treatment option B. Given the inferior prognosis of patients receiving predominantly treatment A, this treatment option may appear inferior to treatment B, which was mostly given to less severely ill patients. Statistical methods exist to "adjust" for identified potentially confounding variables, and we can use propensity scores to adjust for confounding by indication. However, not all such factors are known or accurately measured.

An RCT, if large enough, deals with such extraneous prognostic variables by equally apportioning them to the two or more study arms by randomization. Thus, both known and unknown confounders are distributed roughly evenly between the study arms. For example, a RCT would be the appropriate design to assess whether dexamethasone administered prior to antibiotics reduces mortality in adults who have bacterial meningitis [8]. We would evaluate the following characteristics of such a study: who was studied, was there true random assignment, were interventions and assessments blinded, what was the outcome, and can we generalize to our own patients?

When evaluating clinical trials, it is important to ensure that assignment of treatment was truly randomized. Studies should describe exactly how the patients were randomized, and how the allocation was concealed. It is especially important here to distinguish allocation concealment from blinding. Allocation of an intervention can always be concealed even though blinding of investigators, participants, or outcome assessors may be impossible. Consider an RCT of antibiotics versus surgery for appendicitis. Blinding participants and investigators after patients have been randomized would be difficult as sham operations are ethically problematic. However, allocation concealment occurs before randomization. It is an attempt to prevent selection bias by making certain that the investigator has no idea to what arm (antibiotics versus surgery) the next patient enrolled will be randomized. In many trials, this is done through a centralized randomized process whereby the study investigator is given the assignment after the patient has been enrolled. In some trials, the assignment is kept in envelopes. The problem with this is that, if the site investigator (or another clinician) has a preference for one particular intervention over another, the possibility for tampering exists.

The degree of blinding in a study should also be considered. It is important to recognize that blinding can occur at multiple levels such as the investigators, other health care providers, the patients, the outcome assessors, the data monitoring committee, the data analysts, and even the manuscript writers [9]. Describing a clinical trial as "double-blinded" is vague if, in fact, blinding can occur at so many different levels. It is better to describe who was blinded than using generic terms.

Similarity of groups at baseline should also be considered to assess whether differences in prognostic factors at baseline may have had an impact on the result. A careful consideration of the intervention is also important. We can ask what actually constitutes the intervention—was there a co-intervention that really may have been the "active ingredient"?

Follow-up is another important issue. It is important to assess whether all participants who were actually randomized are accounted for in the results. The expectation nowadays is that the analysis is based on the intention-to-treat population, which is the most conservative approach. That is, all patients randomized are accounted for and are analyzed with respect to the group to which they were originally allocated. For example, an individual in our hypothetical appendicitis trial who was initially randomized to antibiotics but later received surgery would be considered in the analysis to have received antibiotics.

In a next step, we examine the results of the RCT. Consider a randomized controlled trial of two antibiotics A and B for community-acquired pneumonia. If the mortality rate with antibiotic A is 2% and that with B is 4%, the absolute risk reduction is the difference between the two rates ($4 - 2 = 2\%$), the relative risk of A versus B is 0.5, and the relative risk reduction is 50% ($2/4 \times 100 = 50\%$). In studies with time-to-event data, the hazard ratio is measured rather than the relative risk, and can be thought of as an averaged relative risk over the duration of the study. These risk estimates are all commonly reported with a 95% confidence interval (CI) as a measure of precision. A 95% CI that does

not cross 1.0 (for a relative risk or hazard ratio) or 0 (for the absolute risk reduction) has the same interpretation as a p value of <0.05, and we would declare these results as "statistically significant." Unlike the p value, the 95% CI gives us more information regarding the size of the treatment effect. Importantly, the lack of a statistically significant difference between two treatment options does not imply equal efficacy: The 95% CI presents a range of plausible treatment effects. As this plausible effect can be either superior or inferior to the comparison group, the study must be considered indeterminate rather than assuming non-inferiority. It is also important to be aware that statistical significance and clinical importance are not synonymous. A small study may miss an important clinical effect, whereas a very large study may reveal a small but statistically significant difference of no clinical importance. In well-designed studies, researchers prespecify the size of a postulated "minimum clinically important difference" and power the study accordingly rather than solely relying on statistical significance.

A more practical way of determining the size of a treatment effect is to translate the absolute risk reduction into its reciprocal, the number needed to treat (NNT). In this example, the number needed to treat is the number of patients who need to be treated to prevent one death. It is the inverse of the absolute risk reduction (1/0.02), which is 50. Therefore, if 50 patients are treated with antibiotic B instead of A, one death would be prevented. A 95% CI can be calculated on the NNT, although this should only be done if the 95% CI of the absolute risk reduction is not crossing 0.

It is important to determine if all patient-important outcomes were considered in the RCT. For example, a RCT of a novel immunomodulating agent for patients with severe West Nile virus disease would need not only to consider neurologic signs and symptoms, but also to assess functional status and health-related quality of life. When deciding whether the results of a RCT can be applied to your patients, the similarity in the setting and patient population needs to be considered. Finally, you must consider whether the potential benefits of the therapy outweigh the potential risks.

Rather than relying on individual RCTs, it is generally preferable to try to identify systematic reviews on the topic. Systematic reviews, however, also need to be critically evaluated. First, you must ensure that the stated question of the review addresses the clinical question that you are asking. Similar to critical appraisal of RCTs, you should assess the validity of the systematic review itself, in particular, the comprehensiveness of the search strategy, how rigorously the search of titles, abstracts, and full texts were conducted–optimally by at least two investigators independently–and whether the statistical analysis were appropriate. Furthermore, it is an expectation that the authors of the systematic review have critically appraised the studies included in their review, preferably by using the Cochrane risk of bias tool [10]. The most helpful systematic reviews would also GRADE [11] the certainty of evidence made, and as such, provide the reader with an objective assessment of strength of recommendation that can be made based on the identified evidence, an approach adapted in this book for recommendations made by the authors (see section 1.7).

In examining a treatment in the field of infectious diseases, a few other caveats are in order. For many infections, there may be a very strong historic and biologic rationale to treat; in such cases, an RCT using placebo will be unethical. Furthermore, many infections may be too rare to study in RCTs, and some infected populations (such as injection drug users) may be difficult to enroll into treatment studies [12]. Observational methods, such as case-control or cohorts to examine therapies or durations associated with cure or relapse, may be the most appropriate methods in these circumstances. Second, while individual patient RCTs are held up as an ideal, it may be more sensible to study many infections through so-called "cluster randomization" in which the unit of randomization may be a

hospital, school, neighborhood, or family. In particular, in large simple trials that involve a change in policies, for example, in infection prevention and control, this study design may be more appropriate than an individual-patient RCT because infections are transmissible between patients and there may be herd effects of prevention methods. Third, infectious diseases are more dynamic than other illnesses due to new emerging pathogens, and so we frequently encounter new illnesses with no direct body of evidence to guide management (e.g., Middle East respiratory syndrome coronavirus [MERS-CoV]). Furthermore, the effectiveness of antimicrobials is not static, but rather diminishes over time due to the development of antimicrobial resistance, and so the absolute risk reductions and numbers needed to treat in a trial may not be accurate in the future, which is in contrast to, for example, the effect of acetylsalicylic acid in the setting of stroke. Finally, decisions on prevention of infection during outbreaks by newly emerging pathogens must be made when little to nothing is known about this new urgent public health threat. As a consequence, the need for immediate implementation of prevention and treatment approaches often overrides the possibility of conducting studies in order to obtain high-level evidence.

1.6 Evidence-based Assessment of Prognosis

Many studies about risk factors and outcomes for infectious diseases are published, but the quality is variable. The best designs for assessing these are cohort studies in which a representative sample of patients is followed, either prior to developing the infection (to determine risk) or after being infected (to determine outcome). Patients should be assembled at a similar point in their illness ("inception cohort"), and follow-up should be sufficiently long and complete. Important potential confounding prognostic factors should be measured and adjusted for in the analysis. As with clinical trials, the outcome measures are a relative risk, absolute risk, or hazard ratio associated with a particular infection or prognostic factor. For example, to assess the outcome of patients with MERS-CoV infection, we would optimally want an inception cohort of individuals with a laboratory confirmed diagnosis as early in the course of the disease as possible. These individuals would then be followed prospectively. In general, as diagnostic tests improve, our ability to detect early disease will improve.

A challenge unique to infectious diseases is that many infections are transmissible. Thus, a case of disease is, by definition, also a risk factor of disease for others, which complicates research in this field further.

1.7 Our Approach to Reflect the Level of Evidence in this Book

As outlined earlier, RCTs in infectious diseases research are still rare compared to other specialties such as cardiovascular and oncology. Optimally, we would conduct a systematic review for all prognostic factors of interest, diagnostic approaches as well as therapeutic interventions. By doing so, we would assess the risk of bias in included studies, and would be able to GRADE the certainty of evidence and strength of recommendation [11]. Obviously, such an approach would not be feasible; however, we aimed to consider systematic reviews and certainty of evidence assessments made in these reviews wherever possible, and—where no systematic review data was available—authors as content experts were asked to assess the level of evidence themselves based on their best knowledge of the evidence available. Finally, in order to reflect the level of evidence, we used specific wording to reflect the level of evidence of therapeutic or diagnostic recommendations made in this book.

If the evidence for a certain recommendation is backed up by several well-conducted RCTs, we are using statements such as "we recommend," "it is recommended," or "one

should." If there is some evidence available, however, mostly observational studies or small RCTs with wide confidence intervals and/or at high risk of bias, we are making a weak suggestion using terms such as "we suggest," "one might," or "may be used." If the recommendation is based on expert opinion or guideline recommendations only, without a reasonable amount of supporting evidence, we are using terms such as "experts in the field suggest," "expert opinion is," or "some guidelines recommend." By using these terms, we are hoping to make it clearer how confident we are that a specific recommendation should be followed.

1.8 Other Major Changes in the Third Edition of this Book

In addition to consistent wording to reflect the level of evidence supporting our recommendations, the third edition has several new chapters to reflect what had been emerging in terms of important infectious diseases topics over the last few years. These new chapters discuss health-care associated pneumonia, *Clostridium difficile* infection, and antimicrobial stewardship. In order to fit these new chapters into the book, we needed to cut the other chapters significantly, which resulted in more concise and condensed text. The chapters on long-term care in special populations, diarrhea, and infections in thermally injured patients have been omitted from the current version.

Acknowledgements

We would like to acknowledge the two editors of the first and second edition, Mark Loeb and Marek Smieja, who were not involved in this third edition. Their contributions are still reflected in Chapter 1, where we left some of the previous content unchanged.

References

1 How to read clinical journals: I. Why to read them and how to start reading them critically. Can Med Assoc J. 1981;124(5):555–8.
2 Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. JAMA. 1993;270(17):2093–5.
3 Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ. 1996;312(7023):71–2.
4 Dubos R. Pasteur and Modern Science. Washington: ASM Press; 1998.
5 Daniels M, Hill AB. Chemotherapy of pulmonary tuberculosis in young adults: an analysis of the combined results of three Medical Research Council trials. Br Med J. 1952;1(4769):1162–8.
6 Khan AR, Khan S, Zimmerman V, Baddour LM, Tleyjeh IM. Quality and strength of evidence of the Infectious Diseases Society of America clinical practice guidelines. Clin Infect Dis. 2010;51(10):1147–56.
7 Detsky AS, Abrams HB, Forbath N, Scott JG, Hilliard JR. Cardiac assessment for patients undergoing noncardiac surgery: A multifactorial clinical risk index. Arch Intern Med. 1986;146(11):2131–4.
8 de Gans J, van de Beek D. Dexamethasone in adults with bacterial meningitis. N Engl J Med. 2002;347(20):1549–56.
9 Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. JAMA. 2001;285(15):2000–3.

10 Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
11 Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328(7454):1490.
12 Darenberg J, Ihendyane N, Sjolin J, Aufwerber E, Haidl S, Follin P, et al. Intravenous immunoglobulin G therapy in streptococcal toxic shock syndrome: a European randomized, double-blind, placebo-controlled trial. Clin Infect Dis. 2003;37(3):333–40.