

Chapter

1

Introduction

Understanding Diagnosis and Evidence-Based Diagnosis

Diagnosis

When we think about diagnosis, most of us think about a sick person going to the health-care provider with a collection of signs and symptoms of illness. The provider, perhaps with the help of some tests, names the disease and tells the patient if and how it can be treated. The cognitive process of diagnosis involves integrating information from history, observation, exam, and testing using a combination of knowledge, experience, pattern recognition, and intuition to refine the possibilities. The key element of diagnosis is assigning a name to the patient's illness, not necessarily deciding about treatment. Just as we name a recognizably distinct animal, vegetable, or mineral, we name a recognizably distinct disease, so we can talk about it and study it.

Associated with a disease name might be a pathophysiologic mechanism, histopathologic findings, a causative microorganism (if the disease is infectious), and one or more treatments. But more than two millennia before any of these were available, asthma, diabetes mellitus, gout, tuberculosis, leprosy, malaria, and many other diseases were recognized as discrete named entities.

Although we now understand and treat diabetes and malaria better than the ancient Greeks, we still diagnose infantile colic, autism, and fibromyalgia without really knowing what they are. We have anything but a complete pathophysiologic understanding of schizophrenia, amyotrophic lateral sclerosis, and rheumatoid arthritis, all diseases for which treatment (at present) can only be supportive and symptomatic, not curative. Diagnosing a disease with no specific treatment may still help the patient by providing an explanation for what is happening and predicting the prognosis. It can benefit others by establishing the level of infectiousness, helping to prevent the spread of disease, tracking the burden of disease and the success of disease control efforts, discovering etiologies to prevent future cases, and advancing medical science.

Assigning each illness a diagnosis is one way that we attempt to impose order on the chaotic world of signs and symptoms. We group diagnoses into categories based on various shared characteristics, including etiology, clinical picture, prognosis, mechanism of transmission, and response to treatment. The trouble is that homogeneity with respect to one of these characteristics does not imply homogeneity with respect to the others, so different purposes of diagnosis can lead to different disease classification schemes.

For example, entities with different etiologies or different pathologies may have the same treatment. If the goal is to make decisions about treatment, the etiology or pathology may be irrelevant. Consider a child who presents with puffy eyes, excess fluid in the ankles, and a large amount of protein in the urine – a classic presentation of the nephrotic syndrome. In medical school, we dutifully learned how to classify nephrotic syndrome in

1: Introduction: Understanding Diagnosis and Evidence-Based Diagnosis

children by the appearance of the kidney biopsy: there were minimal change disease, focal segmental glomerulosclerosis, membranoproliferative glomerulonephritis, and so on. “Nephrotic syndrome,” our professors emphasized, was a syndrome, not a diagnosis; a kidney biopsy to determine the type of nephrotic syndrome was felt to be necessary.

However, minimal change disease and focal segmental glomerulosclerosis make up the overwhelming majority of nephrotic syndrome cases in children, and both are treated with corticosteroids. So, although a kidney biopsy would provide prognostic information, current recommendations suggest skipping the biopsy initially, starting steroids, and then doing the biopsy later (if at all), only if the symptoms fail to respond or frequent relapses occur. Thus, if the purpose of making the diagnosis is to guide treatment, the pathologic classification that we learned in medical school is usually irrelevant. Instead, nephrotic syndrome is classified as steroid-responsive or nonresponsive and relapsing or non-relapsing. If, as is usually the case, it is steroid-responsive and non-relapsing, we will never know whether it was minimal change disease or focal segmental glomerulosclerosis, because it is not worth doing a kidney biopsy to find out.

There are many similar examples where, at least at some point in an illness, an exact diagnosis is unnecessary to guide treatment. We have sometimes been amused by the number of Latin names that exist for certain similar skin conditions, all of which are treated with topical steroids, which makes distinguishing between them rarely necessary from a treatment standpoint. And, although it is sometimes interesting for an emergency physician to determine which knee ligament is torn, “acute ligamentous knee injury” is a perfectly adequate emergency department diagnosis because the treatment is immobilization, ice, analgesia, and orthopedic follow-up, regardless of the specific ligament injured.

Disease classification systems sometimes have to expand as treatment improves. Before the days of chemotherapy, a pale child with a large number of blasts (very immature white blood cells) on the peripheral blood smear could be diagnosed simply with leukemia. That was enough to determine the treatment (supportive) and the prognosis (grim) without any additional tests. Now, there are many different types of leukemia based, in part, on cell surface markers, each with a specific prognosis and treatment schedule. The classification based on cell surface markers has no inherent value; it is valuable only because careful studies have shown that these markers predict prognosis and response to treatment.

For evidence-based diagnosis, the main subject of this book, we move away from discussions about how to classify and name illnesses toward the process of estimating disease probabilities and quantifying treatment effects to aid with specific clinical decisions.

Evidence-Based Diagnosis

The term “Evidence-based Medicine” (EBM) was coined by Gordon Guyatt around 1992, [1] building on work by David Sackett and colleagues at McMaster University, David Eddy [2], and others [3]. Guyatt et al. characterized EBM as a new scientific paradigm of the sort described in Thomas Kuhn’s 1962 book *The Structure of Scientific Revolutions* [1, 4]. Although not everyone agrees that EBM, “which involves using the medical literature more effectively in guiding medical practice,” is profound enough to constitute a “paradigm shift,” we believe the move from *eminence-based* medicine [5] has been a significant advance.

Oversimplifying greatly, EBM involves learning how to use the best available evidence in two related areas:

1: Introduction: Understanding Diagnosis and Evidence-Based Diagnosis

- **Estimating disease probabilities:** How to evaluate new information, especially a test result, and then use it to refine the probability that a patient has (or will develop) a given disease.
- **Quantifying treatment effects:** How to determine whether a treatment is beneficial in patients with (or at risk for) a given disease, and if so, whether the benefits outweigh the costs and risks.

These two areas are closely related. Although a definitive diagnosis can be useful for prognosis, epidemiologic tracking, and scientific study, in many cases, we may make treatment decisions based on the *probability* of disease. It may not be worth the costs and risks of testing to diagnose a disease that has no effective treatment. Even if an effective treatment exists, there are probabilities of the disease so low that it's not worth testing or so high that it's worth treating without testing. How low or high these probabilities need to be to forgo testing depends on not only the cost and accuracy of the test but also the costs, risks, and effectiveness of the treatment. As suggested by the title, this book focuses more intensively on the probability estimation (diagnosis) area of EBM, but it also covers quantification of the benefits and harms of treatments as well as evaluation of screening programs in which testing and treatment are impossible to separate.

Estimating Disease Probabilities

While diagnosis is the process of naming a disease, testing can be thought of as the process of obtaining additional information to refine disease probabilities. While most of our examples will involve laboratory or imaging tests that cost money or have risks, for which the stakes are higher, the underlying process of obtaining information to refine disease probability is the same for elements of the history and physical examination as it is for blood tests, scans, and biopsies.

How does new information alter disease probabilities? The key is that the *distribution* of test results, exam findings, or answers to history questions *must vary depending on the underlying diagnosis*. To the extent that a test or question gives results that are more likely with condition A than condition B, our estimate of the probability of condition A must rise in comparison to that of condition B. The mathematics behind this updating of probabilities, derived by the eighteenth-century English minister Thomas Bayes, is a key component of evidence-based diagnosis, and one of the most fun parts of this book.

Quantifying Treatment Effects

The main reason for doing tests is to guide treatment decisions. The value of a test depends on its accuracy, costs, and risks; but it also depends on the benefits and harms of the treatment under consideration. One way to estimate a treatment's effect is to *randomize* patients with the same condition to receive or not to receive the treatment and compare the outcomes. If the treatment's purpose is to prevent a bad outcome, we can subtract the proportion with the outcome in the treated group from the proportion with the outcome in the control group. This *absolute risk reduction (ARR)* and its inverse, the *number needed to treat (NNT)*, can be useful measures of the treatment's effect. We will cover these randomized trials at length in Chapter 8. If randomization is unethical or impractical, we can still compare treated to untreated patients, but we must *address the possibility that there are other differences* between the two groups—an interesting topic we will discuss in Chapter 9.

1: Introduction: Understanding Diagnosis and Evidence-Based Diagnosis

Dichotomous Disease State (D+/D−): A Convenient Oversimplification

Most discussions of diagnostic testing, including this one, simplify the problem of diagnosis by assuming a dichotomy between those with a particular disease and those without the disease. The patients with disease, that is, with a positive diagnosis, are denoted “D+,” and the patients without the disease are denoted “D−.” This is an oversimplification for two reasons. First, there is usually a **spectrum of disease**. Some patients we label D+ have mild or early disease, and other patients have severe or advanced disease; so instead of D+, we could have D+, D++, and D+++. Second, there also is usually a **spectrum of nondisease** (D−) that includes **other diseases** as well as varying states of health. Thus, for symptomatic patients, instead of D+ and D−, we should have D1, D2, and D3, each potentially at varying levels of severity, and for asymptomatic patients, we will have D− as well.

For example, a patient with prostate cancer might have early, localized cancer or widely metastatic cancer. A test for prostate cancer, the prostate-specific antigen, is much more likely to be positive in the case of metastatic cancer. Further, consider a patient with acute headache due to subarachnoid hemorrhage (bleeding around the brain). The hemorrhage may be extensive and easily identified by computed tomography scanning, or it might be a small “sentinel bleed,” unlikely to be identified by computed tomography and identifiable only by lumbar puncture (spinal tap).

Even in patients who do not have the disease in question, a multiplicity of potential conditions of interest may exist. Consider a young woman with lower abdominal pain and a positive urine pregnancy test. The primary concern is an ectopic (outside the uterus) pregnancy. One test commonly used in these patients, the β -human chorionic gonadotropin (β -HCG), is lower in women with ectopic pregnancies than in women with normal pregnancies. However, the β -HCG, is often also low in patients with abnormal intrauterine pregnancies [6].

Thus, dichotomizing disease states can get us into trouble because the composition of the D+ group (which includes patients with differing severity of disease) as well as the D− group (which includes patients with differing distributions of other conditions) can vary from one study and one clinical situation to another. This, of course, will affect results of measurements that we make on these groups (like the distribution of prostate-specific antigen results in men with prostate cancer or of β -HCG results in women who do not have ectopic pregnancies). So, although we will generally assume that we are testing for the presence or absence of a single disease and can therefore use the D+/D− shorthand, we will occasionally point out the limitations of this assumption.

Generic Decision Problem: Examples

We will start out by considering an oversimplified, **generic medical decision problem** in which the patient either has the disease (D+) or does not have the disease (D−). If he has the disease, there is a quantifiable benefit to treatment. If he does not have the disease, there is an equally quantifiable cost associated with treating unnecessarily. A single test is under consideration. The test, although not perfect, provides information on whether the patient is D+ or D−. The test has two or more possible results with different distributions in D+ individuals than in D− individuals. The test itself has an associated cost.

Here are several examples of the sorts of clinical scenarios that material covered in this book will help you understand better. In each scenario, the decision to be made includes

1: Introduction: Understanding Diagnosis and Evidence-Based Diagnosis

whether to treat without testing, to do the test and treat based on the results, or to neither test nor treat. We will refer to these scenarios throughout the book.

Clinical Scenario #1: Sore Throat

A 24-year-old graduate student presents with a sore throat and fever that has lasted for 1 day. She has a temperature of 39°C, pus on her tonsils, and tender lymph nodes in her anterior neck.

Disease in question: Strep throat

Test being considered: Rapid antigen detection test for group A streptococcus

Treatment decision: Whether to prescribe penicillin

Clinical Scenario #2: At-Risk Newborn

A 6-hour-old term baby born to a mother who had a fever of 38.7°C is noted to be breathing a little fast (respiratory rate 66). You are concerned about a bacterial infection in the blood, which would require treatment as soon as possible with intravenous antibiotics. You can wait an hour for the results of a white blood cell count and differential, but you need to make a decision before getting the results of the more definitive blood culture, which must incubate for many hours before a result is available.

Disease in question: Bacteria in the blood (bacteremia)

Test being considered: White blood cell count

Treatment decision: Whether to transfer to the neonatal intensive care unit for intravenous antibiotics

Clinical Scenario #3: Screening Mammography

A 45-year-old economics professor from a local university wants to know whether she should get screening mammography. She has not detected any lumps on breast self-examination. A positive screening mammogram would be followed by further testing, possibly including biopsy of the breast.

Disease in question: Breast cancer

Test being considered: Mammogram

Treatment decision: Whether to pursue further evaluation for breast cancer

Clinical Scenario #4: Sonographic Screening for Fetal Chromosomal Abnormalities

In late first-trimester pregnancies, fetal chromosomal abnormalities can be identified definitively using chorionic villus sampling (CVS). CVS entails a small risk of accidentally terminating the pregnancy. Chromosomally abnormal fetuses tend to have larger nuchal translucencies (a measurement of fluid at the back of the fetal neck), absence of the nasal bone, or other structural abnormalities on 13-week ultrasound, which is a noninvasive test. A government perinatal screening program faces the question of who should receive the screening ultrasound examination and what combination of nuchal translucency, nasal bone examination, and other findings should prompt CVS.¹

Disease in question: Fetal chromosomal abnormalities

Test being considered: Prenatal ultrasound

Treatment decision: Whether to do the definitive diagnostic test, chorionic villus sampling (CVS)

¹ A government program would also consider the results of blood tests (serum markers).

1: Introduction: Understanding Diagnosis and Evidence-Based Diagnosis

Preview of Coming Attractions

In Chapters 2, 3, and 4 of this book, we will focus on testing to diagnose **prevalent** (existing) disease in symptomatic patients. In Chapter 5, we will cover test reproducibility, then in Chapter 6, we will move to risk prediction: estimating the probability of incident outcomes (like heart attack, stroke, or death) that are not yet present at the time of the test. In Chapter 7, we will cover combining results from multiple tests. Throughout, we will focus on using tests to guide treatment decisions, which means that the disease (or outcome) under consideration can be treated (or prevented) and, under at least some conditions, the benefits of treatment outweigh the harms. Chapters 8 and 9 are about quantifying these benefits and harms. Chapter 10 covers studies of screening programs, which combine testing of patients not already known to be sick with early intervention in an attempt to improve outcomes. Chapter 11 covers the parallels between statistical testing and diagnostic testing, and Chapter 12 covers challenges for evidence-based diagnosis and returns to the complex cognitive task of diagnosis, especially the errors to which it is prone.

Summary of Key Points

1. The real meaning of the word “diagnosis” is naming the disease that is causing a patient’s illness.
2. This book is primarily about the evidence-based evaluation and use of medical tests to guide treatment decisions.
3. Tests provide information about the likelihood of different diseases when the distribution of test results differs between those who do and do not have each disease.
4. Using a test to guide treatment requires knowing the benefits and harms of treatment, so we will also discuss how to estimate these quantities.

References

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*. 1992;268(17):2420–5.
2. Eddy DM. The origins of evidence-based medicine – a personal perspective. *Virtual Mentor*. 2011;13(1):55–60.
3. Smith R and Rennie D. Evidence based medicine – an oral history. *BMJ*. 2014;348:g371.
4. Kuhn TS. *The structure of scientific revolutions*. Chicago: University of Chicago Press;1962. xv, 172pp.
5. Isaacs D and Fitzgerald D. Seven alternatives to evidence based medicine. *BMJ*. 1999;319(7225):1618.
6. Kohn MA, Kerr K, Malkevich D, et al. Beta-human chorionic gonadotropin levels and the likelihood of ectopic

pregnancy in emergency department patients with abdominal pain or vaginal bleeding. *Acad Emerg Med*. 2003;10(2):119–26.

Problems

1.1 Rotavirus testing

In children with apparent viral gastroenteritis (vomiting and diarrhea), clinicians sometimes order or perform a rapid detection test of the stool for rotavirus. No specific antiviral therapy for rotavirus is available, but rotavirus is the most common cause of hospital-acquired diarrhea in children and is an important cause of acute gastroenteritis in children attending childcare. A rotavirus vaccine is recommended by the CDC’s Advisory Committee on Immunization Practices. Under what circumstances would it be worth

1: Introduction: Understanding Diagnosis and Evidence-Based Diagnosis

doing a rotavirus test in a child with apparent viral gastroenteritis?

1.2 Probiotics for Colic

Randomized trials suggest that breastfed newborns with colic may benefit from the probiotic *Lactobacillus reuteri* [1]. Colic in these studies (and in textbooks) is generally defined as crying at least 3 hours per day at least three times a week in an otherwise well infant [2]. You are seeing a distressed mother of a breastfed 5-week-old who cries inconsolably for about 1–2 hours daily. Your physical examination is normal. Does this child have colic? Would you offer a trial of *Lactobacillus reuteri*?

1.3 Malignant Pleural Effusion in an old man

An 89-year-old man presents with weight loss for 2 months and worsening shortness of breath for 2 weeks. An x-ray shows a left pleural effusion (fluid around the lung). Tests of that fluid removed with a needle (thoracentesis) show undifferentiated carcinoma. History, physical examination, routine laboratory tests, and noninvasive imaging do not disclose the primary cancer. Could “metastatic undifferentiated carcinoma” be a sufficient diagnosis or are additional studies needed? Does your answer change if he has late-stage Alzheimer’s disease?

1.4 Axillary Node Dissection for Breast Cancer Staging

In women with early-stage breast cancer, an axillary lymph node dissection (ALND) to determine whether the axillary (arm pit) nodes are involved is commonly done for staging. ALND involves a couple of days in

the hospital, and is often followed by some degree of pain, swelling, and trouble moving the arm on the dissected side. If the nodes are positive, treatment is more aggressive. However, an alternative to this type of staging is to use a genetic test panel like OncoTypeDX® to quantify the prognosis. A woman whose two oncologists and tumor board all said an ALND was essential for staging (and therefore necessary) consulted one of us after obtaining an OncoTypeDX recurrence score of 7, indicating a low-risk tumor. An excerpt of the report from her test is pasted below:

Five-year recurrence or mortality risk (95% CI) for OncoTypeDX score = 7, by treatment and nodal involvement. (Numbers come from post hoc stratification of subjects in randomized trials comparing tamoxifen alone to tamoxifen plus chemo.)

Treatment	Number of nodes involved (based on ALND)		
	No nodes+	1–3 Nodes+	≥4 Nodes+
Tamoxifen	6% (3%–8%)	8% (4%–15%)	19% (11%–33%)
Tamoxifen + Chemotherapy		11% (7%–17%)	25% (16%–37%)

Assuming the OncoTypeDX report accurately summarizes available evidence, do you agree with her treating clinicians that the ALND is essential? What would be some reasons to do it or not do it?

References

1. Sung V, D’Amico F, Cabana MD, et al. *Lactobacillus reuteri* to treat infant colic: a meta-analysis. *Pediatrics*. 2018;141(1).

2. Benninga MA, Faure C, Hyman PE, et al. Childhood functional gastrointestinal disorders: neonate/toddler. *Gastroenterology*. 2016. doi: 10.1053/j.gastro.2016.02.016.