



# 2.1 Clinical Decision Support

## CHAPTER OUTLINE

- 2.1.1 The Nature and Cognitive Aspects of Human Decision Making
  - 2.1.1.1 *General*
  - 2.1.1.2 *Medical*
- 2.1.2 Decision Science
  - 2.1.2.1 *Decision Analysis*
  - 2.1.2.2 *Probability Theory*
  - 2.1.2.3 *Utility and Preference Assessment*
  - 2.1.2.4 *Cost-Effectiveness Analysis (CEA)*
  - 2.1.2.5 *Test Characteristics*
- 2.1.3 Application of Clinical Decision Support
  - 2.1.3.1 *Types of Decision Support*
  - 2.1.3.2 *Users of Decision Support (Including Clinicians and Patients)*
  - 2.1.3.3 *Implementing, Evaluating and Maintaining Decision Support Tools*
- 2.1.4 Transformation of Knowledge into Clinical Decision Support Tools
  - 2.1.4.1 *Knowledge generation*
  - 2.1.4.2 *Knowledge acquisition*
  - 2.1.4.3 *Knowledge modeling*
  - 2.1.4.4 *Knowledge representation*
  - 2.1.4.5 *Knowledge management and maintenance*
- 2.1.5 Legal, Ethical, and Regulatory Issues
- 2.1.6 Quality and Safety Issues
- 2.1.7 Supporting Decisions for Populations of Patients

## 2.1.1 THE NATURE AND COGNITIVE ASPECTS OF HUMAN DECISION MAKING

Samuel Alfano

Most of decision theory is concerned with identifying the best decision to make (in practice, there are situations in which “best” is not necessarily the optimum, and may also include values within a specific or approximate range), assuming an ideal decision maker who is fully informed, able to compute with perfect accuracy, and fully rational. The practical application of this prescriptive approach (how people ought to make decisions) is called **decision analysis**, and aimed at finding tools, methodologies and software to help people make better decisions. The most systematic and comprehensive software tools developed in this way are called decision support systems (CDS).

Decision analysis often includes the concept of expected value. That is, when faced with a number of options, each of which could give rise to more than one possible outcome, the rational procedure is to identify the likelihood and value (positive or negative) of all possible outcomes associated with each option. You can then multiply the likelihood and value to calculate an expected value. The action to be chosen should be the one that gives rise to the highest total expected value. See Sect. 2.1.2.1 for examples of expected value.

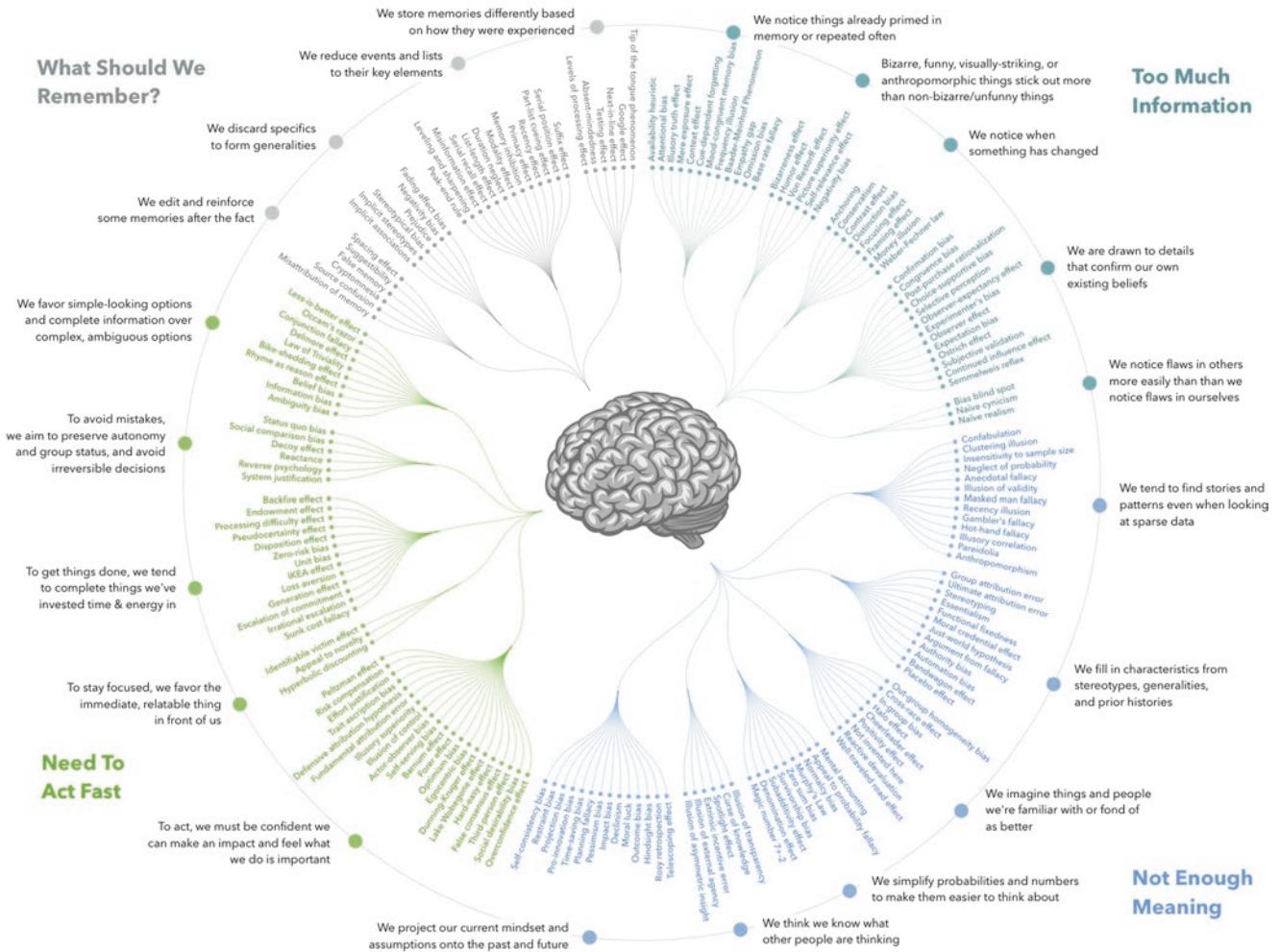
### 2.1.1.1 General

**Cognitive biases** are tendencies to think in certain ways that can lead to systematic deviations from a standard of rationality or good judgment. Making decisions based on opinion is subject to predictable patterns of bias. These are effects of information-processing rules (i.e., mental shortcuts), called **heuristics**, that the brain uses to produce decisions or judgments (Figure 3-1).

Some common Heuristics that lead to bias are listed below.

The original version of this chapter was revised. A correction to this chapter can be found at [https://doi.org/10.1007/978-3-030-16230-6\\_17](https://doi.org/10.1007/978-3-030-16230-6_17)

COGNITIVE BIAS CODEX



DESIGNHACKS.CO · CATEGORIZATION BY BUSTER BENSON · ALGORITHMIC DESIGN BY JOHN MANOOGIAN III (JM3) · DATA BY WIKIPEDIA attribution · share-alike

FIGURE 3-1

Cognitive biases, categorized. Illustration by John Manoogian III, CC-SA

**Availability Bias**

This is the tendency to overestimate the probability of unusual events because of recent or memorable experiences. These memories can be influenced by how recent the memories are or how unusual or emotionally charged they may be. “The last patient I saw with atypical chest pain died of a heart attack 3 days later. I better rule out cardiac disease in this patient” is an example of this type of bias.

**Representativeness Bias**

This is the tendency to overestimate unusual diseases or conditions due to matching pieces of the typical picture of that disease. This bias helps to quickly limit the possibilities and reduce the decision making time. “The patient has bilateral knee pain, we should check an ANA to make sure he does not have lupus” is an example of representativeness.

**Anchoring Bias**

The tendency to rely too heavily, or “anchor”, on one trait or piece of information when making decisions (usually the first piece of information acquired on that subject). This usually leads to a failure to adjust the probability of a disease based on new information. “Even

though he had abnormalities on his ECG, he has a previous diagnosis of peptic ulcer disease, so I don't think he has cardiac disease" is an example of this type of bias.

### Value Based Bias

The tendency to over or under estimate the probability of an outcome based on the perceived value associated with that outcome. We tend to add more weight to information which supports a more valuable outcome. "Even though this patient's headache is certainly a migraine, we should do a CT scan to rule out a brain tumor" is an example of this type of bias.

### Confirmation Bias

The tendency to search for, interpret, focus on and remember information in a way that confirms one's preconceptions. We tend to believe and interpret data in a way that supports our beliefs and initial conclusions. "I think this patient has esophageal reflux disease, he responded well to treatment, so I don't think his ECG abnormalities are significant enough to warrant further investigation" is an example of this type of bias.

### Automation Bias

The tendency to depend excessively on automated systems which can lead to erroneous information that can override or interfere with correct decisions. This can be seen frequently when electronic medical record systems are installed and has been referred to as "task blindness". "I thought it was unusual dose of medication to give to my patient, but the order was confirmed in the system and no medication alerts were triggered." is an example of this bias.

These biases affect belief formation, business and economic decisions, medical decision making, and human behavior in general. They arise as a replicable result to a specific condition.

## 2.1.1.2 Medical

**Clinical reasoning** is the cognitive process that clinicians use to discard or confirm a hypothesis. Several models of this process have been proposed.

**Additive model:** In this model, clinicians assign a positive weight to findings that tend to confirm a diagnosis and negative weights to findings that disconfirm the diagnosis. The clinician's decision to discard or accept a hypothesis is based on the sum of the diagnostic weights. Clinical prediction rules use this model. A clinical example would be the Caprini DVT risk<sup>1</sup> scoring where numerical values are added to the score depending on the patient's history. In many cases though, Physicians probably assign weights to probable conclusions based on the clinical findings. This may happen subjectively and subconsciously.

**Bayesian model:** Bayes' theorem is the method to calculate the probability of a condition based on the prevalence or pretest probability of the condition and other related events, such as diagnostic test results. Bayes' theorem informs the clinician how to adjust his degree of confidence in a hypothesis when new information becomes available. At some point, he may conclude that the probability is low enough to discard the hypothesis.

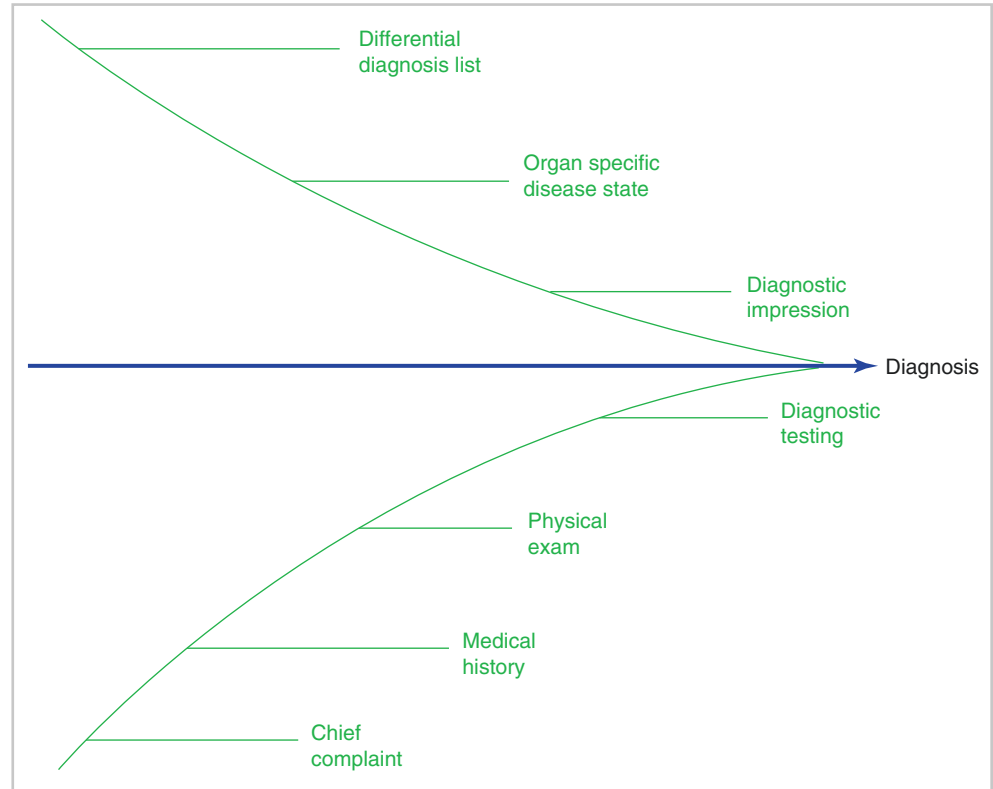
**Algorithmic model:** Algorithms are commonly used to represent the logic of diagnosis. Clinicians follow an internal flow sheet with branching logic as they test a hypothesis. A series of *no* branches eventually leads to discarding a hypothesis and a series of *yes* answers may lead to accepting a diagnosis.

**Blois' Funnel** illustrates the notion that when a doctor begins the diagnostic process for a given patient's condition, he initially brings the widest range of skills and experiences to bear (the area under the upper-left curve on the graph). The patient begins that encounter by describing and presenting a wide, but often vague, range of complaints and symptoms (the area above the lower-left curve on the graph). As the examination continues, a complex set of interactions occur between the two people as they communicate. Body language, targeted questions, readings, measurements, etc., are exchanged as the two arrive at a specific diagnosis (represented by the

<sup>1</sup> Bahl V, Hu HM, Henke PK, Wakefield, TW, Campbell, DA, Caprini, JA. A Validation Study of a Retrospective Venous Thromboembolism Risk Scoring Method. *Ann Surg* 2010. 251(2):344–350.

FIGURE 3-2

Blois' Funnel illustrates the process of finding a diagnosis. Initially, the physician maintains a broad differential diagnosis and the patient elaborates a chief complaint. By applying information from the medical history, physical exam and diagnostic testing, the physician is able to narrow his differential into a diagnostic impression



convergence of the two curves on the right side of the graph). Thus, the practitioner's knowledge, skills, and tools must connect with the patient's complaints, testimony, and anatomy to interoperate technically, semantically, and with orchestrated processes, possibly involving multiple teams, just to produce the working diagnosis (Figure 3-2).

## 2.1.2 DECISION SCIENCE

Samuel Alfano

### 2.1.2.1 Decision Analysis

Patients and Providers often make decisions based on the expected value or utility of the outcome of their decisions. **Expected value** is the summation of the independent probabilities of events. **Expected utility** includes expected value but also takes into account mitigating factors like risk aversion, personal preferences, or circumstances. Though the expected value of a high risk game may be zero or negative to play (given the low rate of return), someone might still choose to play if he gets pleasure out of taking risk, is desperate for money, or for other reasons. In order to calculate the expected value or utility, it is necessary to have a working knowledge of the principles of probabilities.

### 2.1.2.2 Probability Theory

1. **Probability Notation.** The probability that event  $A$  will happen is written as  $P(A)$ . The probability that event  $A$  will NOT happen can be written as  $P(!A)$  or  $P(\sim A)$  or  $P(A')$  or simply  $P(\text{not } A)$ .
2. **Conditional Probabilities** The probability of event  $A$  happening given that  $B$  has happened is notated with a vertical pipe. If  $A$  is the probability of having HIV and  $B$  is the probability of using IV drugs,  $P(A|B)$  is the probability of having HIV given IV drug use.

3. **Addition Rule:** The probability of “A” plus the probability of “not A” must add up to one.  $P(A) + P(A') = 1$ .

(a) Corollary: If another event, B is related to event A, the probability of B is equal to the probability of A and B plus the probability of not-A and B.

$$P(B) = P(A \text{ and } B) + P(A' \text{ and } B)$$

4. **Multiplication Rule:** The probability of A and B occurring is equal to the probability of A times the probability of B, given A.

$$P(A \text{ and } B) = P(A) \cdot P(B|A)$$

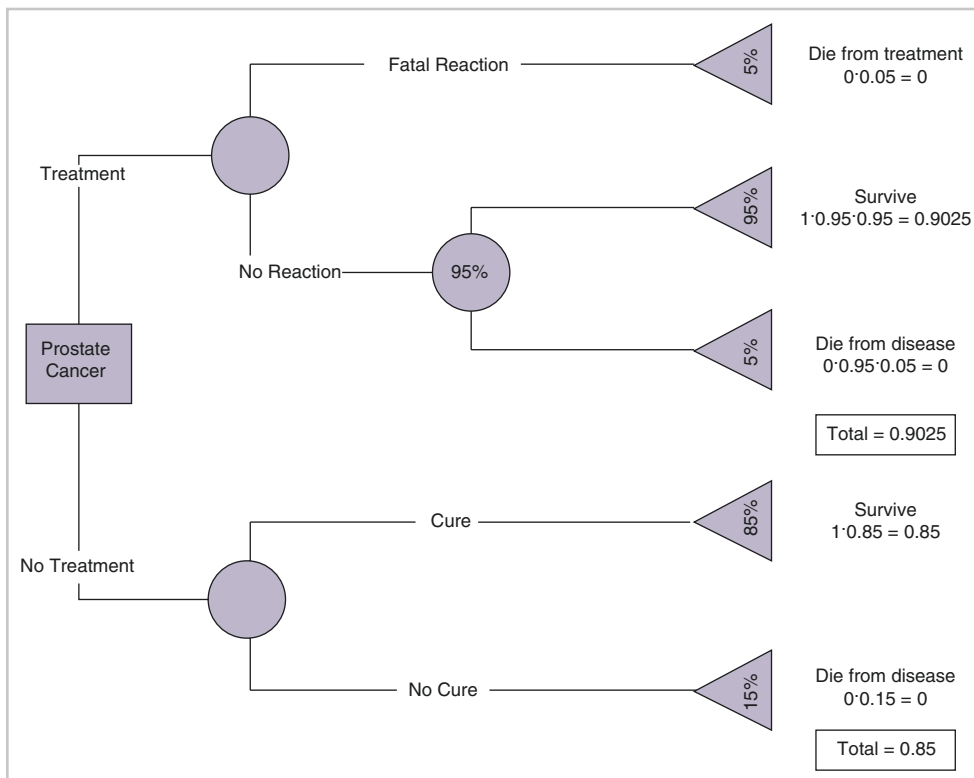
5. **Outcomes Rule:** All of the possible outcome probabilities in a specific instance with multiple outcomes add up to one.

$$P(A) + P(B) + P(C) \dots = 1$$

Sequential events can be described as a **decision tree** or graph which can be used to model a decision using the sum of conditional probabilities. Rules governing decision trees:

1. Decision nodes are designated by a square and represent branching points in the decision tree.
2. Chance nodes are designated by a circle and represent the probability of a specific outcome occurring.
3. Each branch is assigned a probability.
4. The probability of all branches of a node must add up to 1.
5. Outcome nodes are designated by a triangle.
6. Outcomes are assigned a value (cost, utility, Quality Adjusted Life Years, relative value, etc.)
7. If life or death are the outcomes of interest then life = 1 and death = 0

Clinical example: Your patient has prostate cancer. The disease is fatal in 15% of untreated patients. The remaining 85% have spontaneous improvement. When patients are given treatment Y, survival improves from 85 to 95%. Unfortunately, 5% of treated patients have a fatal reaction. The question is whether treatment produces better results than non-treatment (Figure 3-3).



**FIGURE 3-3**

Decision tree



Example Calculations (see chart below)

Expected Value of treatment branch = 0.9025

Fatal Reaction =  $0 \cdot 0.05 = 0$

No reaction =  $((1 \times 0.95) + (0 \times 0.05)) \cdot 0.95 = 0.9025$

Sum of both =  $0 + 0.9025 = 0.9025$

Expected Value of no treatment branch = 0.85

Cure =  $1 \cdot 0.85 = 0.85$

No cure =  $0 \cdot 0.15 = 0$

Sum of probabilities of both =  $0.85 + 0 = 0.85$

This model very slightly favors giving treatment Y despite a 5% chance of fatal reaction. What if the probability of treatment complications or fatality due to treatment Y increases? At what threshold does the therapy become too risky and confer no additional benefit? Also defined as **sensitivity analysis**, what-if analysis is a technique used to determine how projected performance is affected by changes in the underlying assumptions.

### 2.1.2.3 Utility and Preference Assessment

**Utility** can be used in decision analysis to adjust the value of a choice based on a patient's perceived utility of a potential outcome. Methods include Standard Gamble, Time Trade Off, Visual Analog, and Quality Adjusted Life Years. The expected value of gambling is the sum of the expected probabilities. One can compare this to the expected utility of not gambling.

Expected Value vs Expected Utility:

A \$10 bet gives you a 1 in 80 chance of winning \$1000. If you gamble and win, you get \$1000, however, if you gamble and lose, you get nothing. If you do not gamble, you keep your \$10.

Expected value of doing nothing (not gambling) =  $1 \cdot \$10 = \$10$

Expected value of gambling \$1000  $\cdot (1/80) + 0 \times (79/80) = \$12.50$

This example favors gambling. In the long run, gambling will produce a more favorable outcome than not gambling. However, expected utility is a function of value and also risk aversion, personal preferences and circumstances. If you desperately need \$10, you might choose not to gamble. If you are a risk taker or have lots of money, you might choose the gamble.

Patient preferences require more input than straight probability. Two outcomes may not have same effectiveness or impact. (e.g. surgical debulking vs. chemotherapy). In addition, the numerically favorable therapy is not always the one the patient prefers. Some decisions require subjective input by the patient. Consider:

Therapy A: 50% chance of 10 year survival: expected value = 5 years

Therapy B: 90% chance of a 2 year survival: expected value = 1.8 years

Utility can be used in decision analysis to define the "value" of an outcome node. One can adjust the value of the outcome based on the perceived utility of that outcome for that patient. There are several common approaches:

#### Standard Gamble

In the Standard Gamble, the patient is assumed to have a chronic health condition and is offered a potentially disease-altering treatment, however the treatment is not without risk. The patient is given two alternatives: (1) continue life with the current medical condition; (2) choose the intervention, even though there is a defined risk of death. If the patient strongly

chooses either option, the risk of death is adjusted up or down and the question is posed again. At some point, the patient reaches a **point of indifference**, where he really can't choose between the two options. That resulting value is the utility.

As an example, suppose a patient has chronic renal failure on maintenance hemodialysis and we are trying to determine the utility of renal transplant. The patient is asked, "would you rather have renal failure or have a transplant? The risk of dying from a transplant is 90%". The patient recoils and says that he would much prefer dialysis to a 90% chance of death. What if the risk of death were only 0.05%? At this point, the patient gladly accepts the transplant. Next, we go back to the patient and say, "what if the risk of death were 25%". Again, the patient refuses transplant. We keep modifying the risk of death up and down until we arrive at 2%. At that point, the patient can't really decide which option is better. He has reached his point of indifference, and the utility of transplant is determined to be 2%.

### Time Trade Off

In a Time Trade Off (TTO), sick patients estimate how many years of their life they would be willing to give up to live a certain number of years in full health. For example, you have a chronic disease and are told that you have 10 years left to live. In connection with this you are also told that you can choose to live these 10 years in your current (less than healthy) state or that you can choose to give up some life years for a shorter period in full health. How many years in full health do you think is of equal value to 10 years in your current health state?

The TTO utility (the indifference point) is the length of remaining life in perfect health divided by the length of remaining life with the evaluated health state. In the above example, one might choose to give up 5 years of life in their current state in order to live a more healthy life. The TTO utility would be 5 years/10 years or 0.5.

### Visual Analog

When using this approach, patients are asked to rate different health states on a marked or unmarked scale where 0 = death and 100 = perfect health.

### Quality Adjusted Life Year

The quality-adjusted life-year (QALY) is a measure of the value of health outcomes. Since health is a function of length and quality of life, the QALY was developed as an attempt to combine the value of these attributes into a single number. The QALY calculation is simple: the change in utility value induced by the treatment is multiplied by the duration of the treatment effect to provide the number of QALYs gained. QALYs can then be incorporated with medical costs to arrive at a ratio of cost/QALY. This parameter can be used to compare the cost-effectiveness of any treatment. Some believe that there are states of health that are worse than death so QALY can have a negative value.

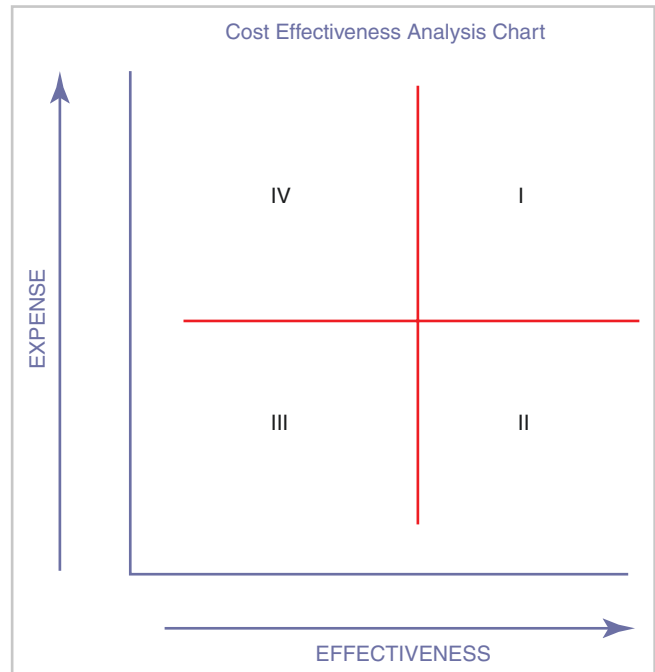
## 2.1.2.4 Cost-Effectiveness Analysis (CEA)

**Cost-effectiveness analysis (CEA)** is a form of economic analysis that compares the relative costs and outcomes (effects) of two or more courses of action. Cost-effectiveness analysis is distinct from cost-benefit analysis, which assigns a monetary value to the measure of effect. Cost-effectiveness analysis is often used in the field of health services, where it may be inappropriate to monetize health effect. Typically the CEA is expressed in terms of a ratio where the denominator is a gain in health from a measure (years of life, premature births averted, sight-years gained) and the numerator is the cost associated with the health gain. The most commonly used outcome measure is quality-adjusted life years (QALY). **Cost-utility analysis** is similar to cost-effectiveness analysis. Cost-effectiveness analyses are often visualized on a cost-effectiveness plane consisting of four quadrants (Figure 3-4). Knowing cost and utility/value of an outcome allows you to calculate the Incremental Cost/Effectiveness Ratio (ICER). Comparing the calculated ICER to the "willingness to pay" can help to determine if a therapy is cost effective. Assuming two options, A and B, ICER is defined as:

$$\text{ICER} = (\text{Cost of A} - \text{Cost of B}) / (\text{Effect of A} - \text{Effect of B})$$

**FIGURE 3-4**

Cost effectiveness analysis



### 2.1.2.5 Test Characteristics

Besides understanding the probabilities of treatments and disease states, it is equally important to understand the diagnostic testing that is used to make decisions and its limitations.

**Sensitivity** is defined as the probability that a test result will be positive when the disease is present (true positive rate). The true positive rate is equal to the number of true positives divided by the total number of sick patients in the population. Recall that the number of sick patients is equal to the true positives as well as the false negatives. A negative result in a test with high sensitivity is useful for ruling out disease. A high sensitivity test is reliable when its result is negative, since it rarely misdiagnoses those who have the disease. A test with 100% sensitivity will recognize all patients with the disease. A negative test result would definitively *rule out* presence of the disease in a patient.

Since sensitivity does not account for false positives, a positive result in a test with high sensitivity is not useful for ruling in disease.

$$\text{Sensitivity} = \text{True Positive} / (\text{True Positive} + \text{False Negative})$$

**Specificity** is defined as the probability that a test result will be negative when the disease is not present (true negative rate). The true negative rate is equal to the number of true negatives divided by the number of people in the population who do not have the disease. Recall that the number of healthy patients without disease is equal to the true negatives plus the false positives. A positive result in a test with high specificity is useful for ruling in disease. The test rarely gives positive results in healthy patients. A test with 100% specificity will read negative, and accurately exclude disease in all healthy patients. A positive result signifies a high probability of the presence of disease. A negative result in a test with high specificity is not useful for ruling out disease.

$$\text{Specificity} = \text{True Negative} / (\text{False positive} + \text{True Negative})$$



	DISEASE	NO DISEASE
TEST +	8	10
TEST -	2	980

In this example, the prevalence of the disease is 1% (i.e. because 10 of 1000 people have it)

$$\text{Sensitivity} = 8/(8 + 2) = 80\%$$

$$\text{Specificity} = 980/(980 + 10) = 99\%$$

Screening tests tend to be highly sensitive and relatively inexpensive. False positives are common, and the person ordering the test must use this knowledge when interpreting the results. Confirmatory tests tend to be more expensive and are designed to minimize false positives (i.e. have high specificity). The two by two table (above) is also called a **confusion matrix**.

**Likelihood ratios** are used for assessing the value of performing a diagnostic test. They use the sensitivity and specificity of the test to determine whether a test result usefully changes the probability that a disease exists.

The **Negative Likelihood Ratio**, sometimes written as  $LR(-)$ , is defined as the ratio between the probability of a negative test result given the presence of the disease and the probability of a negative test result given the absence of the disease, i.e.  $(\text{False negative rate} / \text{True negative rate}) = (1 - \text{Sensitivity}) / \text{Specificity}$ .

The **Positive Likelihood Ratio**, written as  $LR(+)$ , is defined as the ratio between the probability of a positive test in the presence of disease and the probability of a positive test in the absence of disease, i.e.  $(\text{True positive rate} / \text{False positive rate}) = \text{Sensitivity} / (1 - \text{Specificity})$ .

A likelihood ratio of greater than 1 indicates the test result is associated with the disease. A likelihood ratio less than 1 indicates that the result is associated with absence of the disease.

#### Positive and Negative Predictive Values

While sensitivity and specificity are characteristics of the test being used to diagnose a disease, **positive predictive value (PPV)** and **negative predictive value (NPV)** are predictive characteristics of the patient given certain test results.

The **PPV** is the probability that the disease is present when the test is positive and the **NPV** is the probability that the disease is not present when the test is negative.

$$\text{PPV} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{NPV} = \text{TN} / (\text{TN} + \text{FN})$$

Sensitivity and specificity assume the disease state is known, and are used to determine if the test performs as intended. PPV and NPV are more useful in clinical practice, where the test result is known and the disease state is being sought.

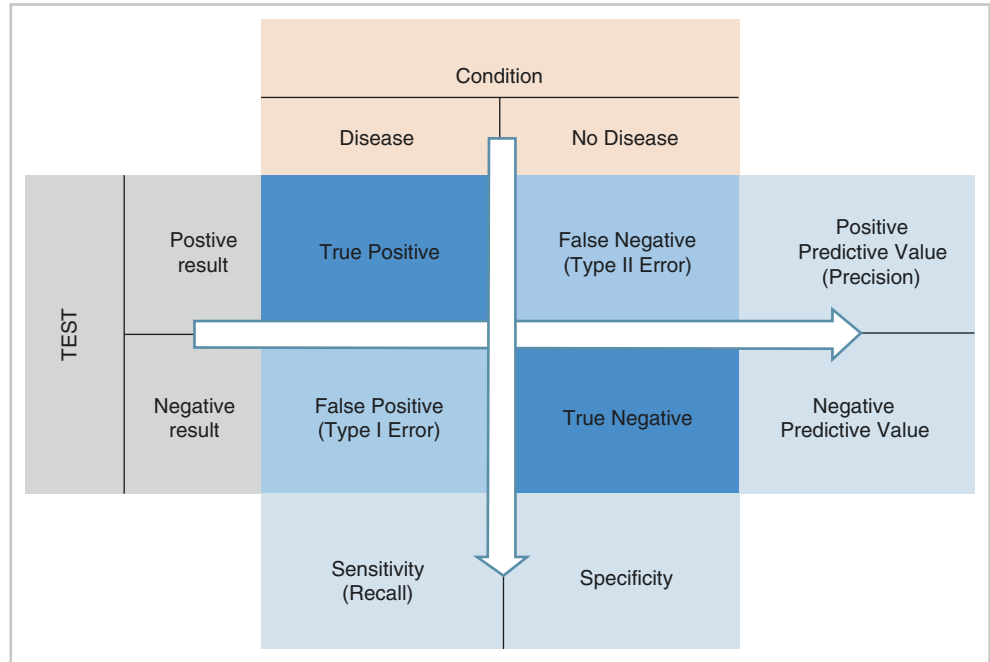
Another way to state the relationship between these values:

Sensitivity: Given the patient has disease, what is the probability of a positive test?

Specificity: Given the patient does not have the disease, what is the probability of a negative test?

PPV: Given a positive test, what is the probability the patient has disease?

**FIGURE 3-5**  
Relationship between test results and presence of condition



NPV: Given a negative test, what is the probability the patient does not have disease?

In statistics, a null hypothesis ( $H_0$ ) is a statement that one seeks to disprove with evidence to the contrary. Most commonly it is a statement that the phenomenon being studied produces no effect or makes no difference. Usually, an experimenter frames a null hypothesis with the intent of rejecting it: that is, intending to run an experiment which produces data that shows that the phenomenon under study does make a difference. In most cases there is a specific alternative hypothesis ( $H_a$ ) that is opposed to the null hypothesis, in other cases the alternative hypothesis is not explicitly stated, or is simply that the null hypothesis is false.

A type I error (or error of the first kind) is the incorrect rejection of a true null hypothesis (i.e. a false positive). This error usually leads one to conclude that a supposed effect or relationship exists when in fact it doesn't.

A type II error (or error of the second kind) is the failure to reject a false null hypothesis (i.e. a false negative). This error leads one to conclude that a relationship does not exist when it truly does (Figure 3-5).<sup>2</sup>

Increasing prevalence decreases negative predictive value of the test. Even though a test predicts the disease as absent, the more prevalent the disease, the less likely the test is right. This means that a negative test isn't very good with a common disease—and a positive test isn't very good for a rare disease.

For every possible cut-off point or criterion value you select to discriminate between the two populations (disease or no disease) there will be some cases where the disease is incorrectly classified. Suppose we have a diagnostic test which yields a result between 1 and 5. The healthy population is centered around 2.5 with a standard deviation of 0.5. Similarly, the sick population has a mean of 4 with a similar standard deviation. A plot of the two populations is shown in Figure 3-6. When designing our test, where should the cut-off value be? If we set it at 3, some of the healthy population (i.e. those with values from 3 to 3.5) will be falsely characterized as sick (i.e. false positives). Similarly, if we set the cut-off at 3.5, some of the sick patients will be categorized as healthy (i.e. false negatives).

<sup>2</sup> A mnemonic to remember this is "Ladies First". A type I error, is a girlfriend's mistake. She assumes that a relationship exists when it truly doesn't. The type II error is the boyfriend's mistake. He assumes that there is no relationship when there really is. -Ed.

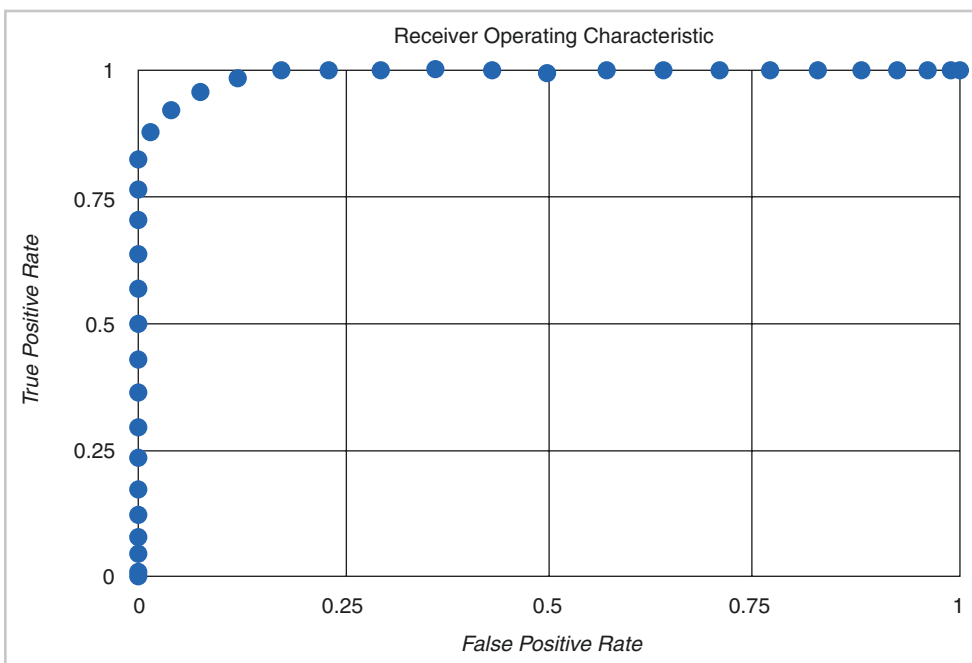
When you select a lower criterion value, then the true positive fraction and sensitivity will increase. On the other hand the false positive fraction will also increase, and therefore the true negative fraction and specificity will decrease. This can be depicted as a **receiver operating characteristic (ROC) plot** (Figure 3-7).

A ROC space is defined by FPR and TPR as  $x$  and  $y$  axes respectively, which depicts relative trade-offs between true positive (benefits) and false positive (costs). Since TPR is equivalent to sensitivity and FPR is equal to  $1 - \text{specificity}$ , the ROC graph is sometimes called the sensitivity vs  $(1 - \text{specificity})$  plot. Each prediction result or instance of a confusion matrix represents one point in the ROC space.

The best possible prediction method would yield a point in the upper left corner or coordinate  $(0,1)$  of the ROC space, representing 100% sensitivity (no false negatives) and 100%



**FIGURE 3-6**  
Example distribution of healthy and sick patients in a population. Note the overlap in test results between 3-3.5



**FIGURE 3-7**  
Construction of the receiver operating characteristic (ROC) curve based on true positive rate (TPR) and false positive rate (FPR). Each point on this curve represents a particular value for TPR and FPR

specificity (no false positives). The (0,1) point is also called a **perfect classification**. A completely random guess would give a point along a diagonal line (the so-called line of no-discrimination) from the left bottom to the top right corners (regardless of the positive and negative base rates). The diagonal divides the ROC space. Points above the diagonal represent good classification results (better than random), points below the line represent poor results (worse than random).

### Bayes' Theorem

Bayes' theorem helps overcome many well-known cognitive errors in diagnosis, such as ignoring the base rate, probability adjustment errors (conservatism, anchoring and adjustment) and order effects. It is calculated as the probability of disease given a positive test. This is equal to the probability of a positive test given disease (sensitivity) multiplied by the probability of disease (prior probability) divided by the probability of a positive test.

Bayes's theorem is stated mathematically as the following form:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Suppose a drug test is 99% sensitive and 99% specific. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability he or she is a user? To solve this, we set A=probability that he is a drug user and B=probability that the test is positive. Therefore, P(A|B) is the probability that he is a user given that his test is positive.

To calculate the probability that the test will be positive (e.g. P(B)), we have to add up all the true positives (i.e. users who test positive) and the false positives (non-users who test positive), resulting in the following equations

$$\begin{aligned} P(\text{User} | \text{Test}+) &= \frac{P(\text{Test}+ | \text{User}) \cdot P(\text{User})}{\text{true positives} + \text{false positives}} \\ &= \frac{P(\text{Test}+ | \text{User}) \cdot P(\text{User})}{P(\text{Test}+ | \text{User}) \cdot P(\text{User}) + P(\text{Test}+ | \text{Nonuser}) \cdot P(\text{Nonuser})} \\ &= \frac{0.99 \cdot 0.005}{0.99 \cdot 0.005 + 0.01 \cdot 0.995} \\ &\approx 33.2\% \end{aligned}$$

Despite the apparent accuracy (high sensitivity and specificity) of the test, if an individual tests positive, it is more likely that they do *not* use the drug than that they do.

This surprising result arises because the number of non-users is very large compared to the number of users; thus the number of false positives (0.995%) outweighs the number of true positives (0.495%). To use concrete numbers, if 1000 individuals are tested, there are expected to be 995 non-users and 5 users. From the 995 non-users,  $0.01 \times 995 \cong 10$  false positives are expected. From the 5 users,  $0.99 \times 5 \cong 5$  true positives are expected. Out of the 15 positive results, only 5, about 33%, are genuine. The importance of specificity can be illustrated by showing that even if sensitivity is 100% and specificity is at 99% the probability of the person being a drug user is  $\approx 33\%$  but if the specificity is changed to 99.5% and the sensitivity is dropped down to 99% the probability of the person being a drug user rises to 50%.

### Odds and Likelihood Ratios

$$\text{Odds of } X = \frac{P(X)}{1 - P(X)}$$

For example, a coin toss has a 50% chance of heads, which means that the odds of heads are 1:1. Events with 1:3 odds have probability of 0.25. Odds, likelihood ratio, sensitivity and specificity are related as follows:

Post-test odds = (pretest odds) x (positive likelihood ratio).

Positive likelihood ratio (LR+) = sensitivity / (1-specificity) = TPR/FPR.

Negative Likelihood ratio (LR-) = (1-sensitivity) / specificity = FNR/TNR

$$LR = \frac{\text{Probability of result in diseased persons}}{\text{Probability of result in nondiseased persons}}$$

$$LR (+) = \frac{\text{Probability that test is positive in diseased persons}}{\text{Probability that test is positive in nondiseased persons}} \\ = \frac{\text{Sensitivity}}{1 - \text{Specificity}}$$

$$LR (-) = \frac{\text{Probability that test is negative in diseased persons}}{\text{Probability that test is negative in nondiseased persons}} \\ = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

Example:

	Disease	No Disease
Test +	19	7
Test -	2	439

Sensitivity =  $19/(19 + 2) = .904$

Specificity =  $439/(7 + 439) = 0.984$

LR (+) = Sensitivity / (1 - specificity) = 57.6

Pre-test odds of having disease are  $(19 + 2)/(7 + 439) = 0.047$  and if the test is positive, post test odds are  $0.047 \times 57.6 = 2.71$ . Since odds =  $P(x) / (1-P(x))$  you can determine that  $P(x) = 73\%$ . That is, a positive test means the patient has a 73% chance of having the disease and a 27% chance of not having the disease sensitivity / (1-specificity)

### Fagan Nomogram

The Fagan nomogram (Figure 3-8) is a tool for estimating how much the result on a diagnostic test changes the probability that a patient has a disease (1-sensitivity) / specificity. To use this tool  $LR(+) = \text{sensitivity} / (1 - \text{specificity}) = 57.6$ .

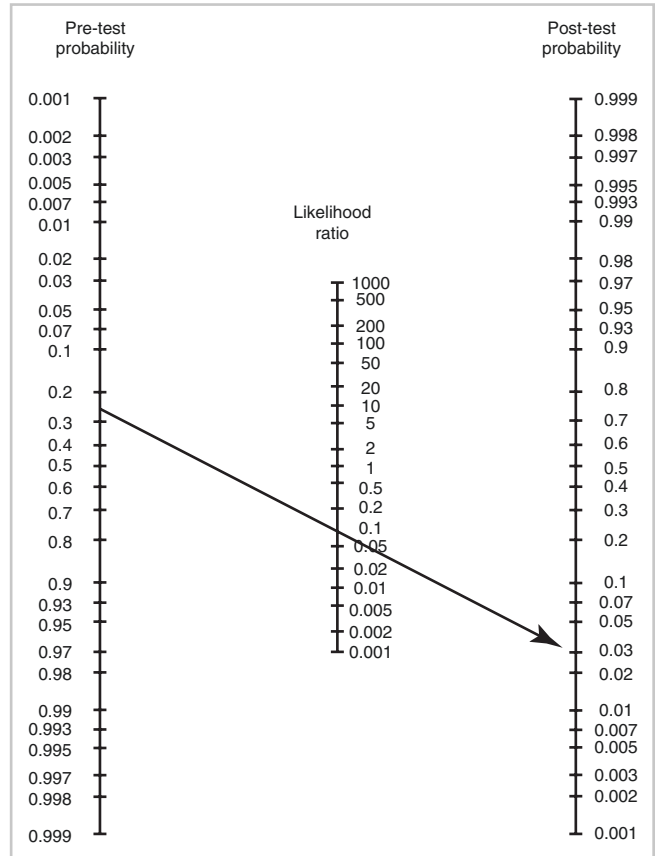
Draw a line connecting the pre-test probability and the likelihood ratio. Extend this line until it intersects with the post-test probability. The point of intersection is the new estimate of the probability that your patient has this disease (Figure 3-8).

### Precision and Recall

In pattern recognition and information retrieval, precision (also called positive predictive value) is the fraction of retrieved instances that are relevant, while recall (also known as sensitivity) is the fraction of relevant instances that are retrieved. Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity. In simple terms, high precision means that an algorithm returned substantially more relevant results than irrelevant, while high recall means that an algorithm returned most of the relevant results.

**FIGURE 3-8**

The Fagan nomogram for calculating the post-test probability from the pre-test probability and the likelihood ratio. After Fagan TJ. Nomogram for Bayes theorem. *N Engl J Med* 1975; 293 (5):257–61



## 2.1.3 APPLICATION OF CLINICAL DECISION SUPPORT

Lisa Masson

### 2.1.3.1 Types of Decision Support

**Clinical Decision Support (CDS)** is any electronic tool that provides structured guidance. By this broad definition, CDS is any means to reduce the cognitive burden of patient care.

Often, team members seeking to solve an organizational problem suggest building an alert. Technology is not always the answer to the problem; when it is, the technology that is the answer is not always an alert.

CDS includes general references, specific guidelines, and a broad assortment of interventions.<sup>3</sup> Order sets and data visualization techniques also fall under the umbrella of CDS. Broad groups of CDS include: (1) information management; (2) focused attention (alerts); and (3) patient-specific recommendations.<sup>4</sup>

Benefits of CDS include preventing errors, optimizing decision making and improving care processes. An example of preventing errors is automatic dosage calculations in the ordering of medications. A cardiovascular risk calculator embedded in an Electronic Health Record (EHR) can be used to optimize decision making. An example of improving care

<sup>3</sup> Berner ES. Clinical decision support systems: State of the Art. AHRQ Publication No. 09-0069-EF. Rockville, Maryland: Agency for Healthcare Research and Quality. June 2009.

<sup>4</sup> Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The Benefits of Health Information Technology: A Review of the Recent Literature Shows Predominantly Positive Results. *Health Affairs* 30, No. 3 (2011): 464–471.



FIGURE 3-9

Example of an embedded calculator, invoked when the physician types a certain phrase asking for the information

processes is automatically providing flow sheets for appropriate home support during the discharge process (Figure 3-9).

A **knowledge-based clinical decision support** system includes several elements. First, the knowledge base is comprised of clinical information such as drug interactions and guidelines. Second, the system will consider each individual patient's information including problem lists, medications, allergies, age, weight, and lab results. Finally, the CDS has a communication mechanism to convey relevant information to the physician. This information could include: lists of possible diagnoses, warnings of potential drug interactions, clinical guidelines suggesting workup and treatment options, and preventative care reminders.

CDS interventions have shown benefits in every area of healthcare. In preventative care, there are provisions for immunizations and appropriate screening examinations. In patients with established diseases, disease management guidelines embedded in the EHR may support medical care. CDS may aid in the process of establishing a diagnosis by making suggestions which match signs and symptoms. Treatment guidelines and drug dosage recommendations, as well as drug-drug interactions, once painstakingly memorized, are now so commonplace as to be expected benefits of an EHR. CDS is leveraged for provider efficiency by offering care plans to standardize care and minimize length of stay. These pathways help prevent common hospital complications such as deep venous thrombosis and decubitus ulcers by automating the process of writing admission orders via pre-populated order sets. Duplicate testing alerts and drug formulary guidelines help reduce costs.

FIGURE 3-10

An example of Arden Syntax for encoding of a Medical Logic Module (MLM) which gives a warning if hematocrit has dropped by 5 or is more than 30

```

maintenance:
  title: Alert on low hematocrit;;
library:
  purpose: Warn provider of new or worsening anemia.;;
knowledge:
  type: data-driven;;
data:
  blood_count_storage := event {'complete blood count'};
  hematocrit := read last {'hematocrit'};
  previous_hct := read last { {'hematocrit'} where it occurred before
    the time of hematocrit};;
evoke: blood_count_storage;;
logic:
  if hematocrit is not number then conclude false; endif;
  if hematocrit <= previous_hct-5 or hematocrit < 30 then conclude true;
  endif;;
action:
  write "The patient's hematocrit ("|| hematocrit ||") is low or
    falling rapidly.";;
end:

```

Clinical practice guidelines (CPGs, see Sect. 2.2.3) aim to improve quality of care, reduce unjustified practice variations and reduce healthcare costs.<sup>5</sup> Effective guidelines can be seamlessly incorporated into the clinical workflow. CPGs have been enacted as **computer-interpretable guidelines (CIGs)**. To be effective, the CIGs must be recognized as relevant to the physician in their daily practice. A goal of a CIG is to match the patient data, and provide evidence-based guideline support in an appropriate fashion during a care provider's work.

CIGs face barriers in translating from English to a computer language due to different types of ambiguity. **Syntactic ambiguity** exists if the sentence is constructed such that it could have more than one meaning: I'm glad I'm a man, and so is Lola (Ray Davies) or I shot an elephant in my pajamas (Groucho Marx). **Semantic ambiguity** refers to one word having different meanings; a mole can be a lesion on one's skin or a small burrowing mammal. **Pragmatic ambiguity** arises from inconsistent or conflicting recommendations. Additional vagueness arises from **under-specification** (using the word *children* rather than specifying ages 0–18 years), **strength qualifiers** (should be effective) and **passive voice** (should be performed). To address these issues, a set of building blocks such as tasks, rules, nodes or frames is used to create a model. Formal language must be used to clarify the actual guidelines within the constraints of the model and this language must be machine-interpretable.

Many active CIG approaches are currently in place with unique features. No standard CIG language or model exists. Most approaches model guidelines via a **Task-Network Model (TNM)**, which contains a flowchart of specific tasks.

**Arden Syntax**, a language for encoding medical knowledge, is most useful for simple guidelines. Unlike most of the other approaches, Arden Syntax does not use TNM. In contrast, the guidelines in Arden Syntax are a collection of Medical Logic Modules (MLMs). Each MLM represents a single decision. These independent modular rules render Arden Syntax less than ideal for complex multi-step guidelines. Arden Syntax is most appropriate for representing simple alerts in reminder systems (Figure 3-10).

5 Mor Peleg. Computer Interpretable Clinical Guidelines: A methodological review. Journal of Biomedical Informatics. Volume 46, Issue 4. August 2013: 744–763.

### 2.1.3.2 Users of Decision Support (Including Clinicians and Patients)

To realize the benefits of CDS, the interventions must tackle the **five rights of clinical decision support**. They are the right **(1) information (2) person (3) format (4) channel and (5) time**. The **right information** should be patient specific, and accurate. The **right person** refers to displaying the information to the appropriate team member who can act, whether it be the physician, the nurse, the pharmacist, or the hospitality suite. The **right format** may be less obvious. Would this data be easier to digest as a dashboard or as a pop up box? Can a guideline be protocolized (such as ordering annual flu shots for appropriate populations) and simply enacted by staff, or should it be an order set for a clinician to choose? Multiple channels exist beyond the graphic user interface of the EHR. For example, the **right channel** for decision support may be provided via the patient portal or e-mail. The **right timing** of the information is critical for acceptance. The information should be actionable. Don't tell a physician about a drug-drug interaction after the medication has been administered; warn her before the order is placed. Don't suggest a clinical pathway during the history taking portion of the exam, as this might be forgotten in the complexities of patient care, but rather in the orders section during the active cognitive process pertinent to this knowledge (Figure 3-11).

If an alert is chosen as the method of CDS, the alert can be **interruptive or non-interruptive**. The modeless alert appears in the background or on the toolbar. It stays on the screen available for use but allows other activities, and is retrievable when the clinician is ready. An example is an unread lab result indicator on the toolbar. A **modal** is a pop-up alert, such as a reminder, which requires acknowledgment before continuing (See figs. 3-12, 3-13 and 3-14).

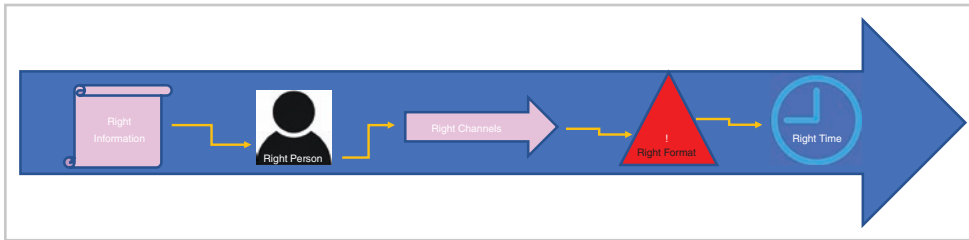


FIGURE 3-11

The five rights of clinical decision support: Right information, person, format, channel and time

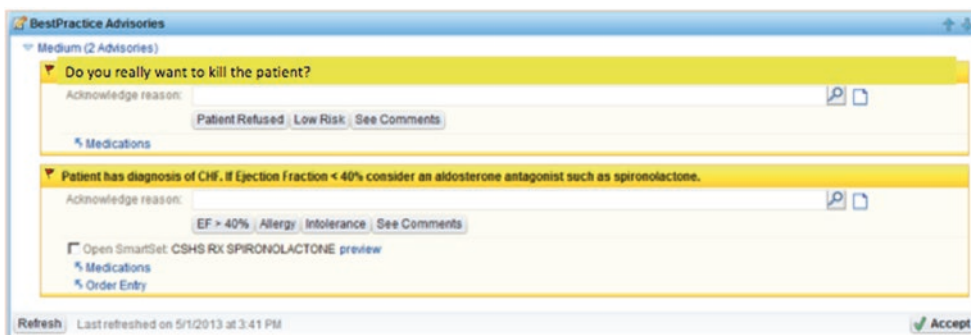


FIGURE 3-12

An example of a pop-up alert

#### MIPS Diabetes Eye Exam

This quality measure alert identifies patients 18 through 75 years of age with a diagnosis of diabetes (type 1 and type 2) who did not have a retinal or dilated eye exam by an eye care professional this year or did not have a negative retinal or dilated eye exam (negative for retinopathy) in the year prior. Please consider if further action is required for your patient.

[Click here to show more](#)

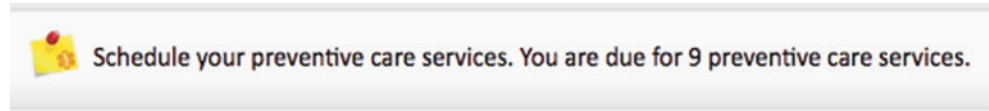
This recommendation is not a substitute for clinical judgment. Any trademarks identified are owned by their respective owners and used solely for reference purposes and do not imply an endorsement by such owners of the Stanson alerts. This intervention was independently developed and funded by Stanson Health and is licensed for use in an electronic health record by Cedars-Sinai Health System. Initial creation date: 12/29/2016; last updated: February 2017.

FIGURE 3-13

An example of a passive or modeless alert. The clinician is not asked to interact with the alert and can simply continue her workflow

FIGURE 3-14

Passive alert in a personal health record



The decision support may simply supply information, issue reminders, correct errors or suggest changes in care plans. The urgency and impact of the support are considerations in assessing the five rights. The best accepted support matches the clinician's intentions while relieving cognitive burden or memory. This is exemplified in a reminder to order a mammogram while the patient is being seen in an unrelated visit for flu-like symptoms. Neither the rushed clinician nor the ill patient would think to obtain the mammography order. Well-orchestrated CDS would have the order suggested or pended so it automatically occurs for the patient on completing the visit. Ease of use is critical for acceptance.

### 2.1.3.3 Implementing, Evaluating and Maintaining Decision Support Tools

Review of recent literature indicates predominantly positive results from technology in health care delivery.<sup>6</sup> In a systematic review of literature from July 2007 to February 2010, 92% of the articles had positive conclusions. Smaller practices, for whom the overhead was higher, are slowly beginning to enjoy benefits of health information technology as well.

Examples of positive findings in clinical decision support abound. CDS for erythropoietin dosing in a dialysis clinic reduced the staff time spent on anemia management by half while clinical outcomes remained unchanged.<sup>7</sup> In an inpatient setting, an alert to reduce unnecessary RBC transfusions efficiently showed a reduction in transfusion costs without an increase in patient length of stay or mortality.<sup>8</sup> Evidence has shown that CDS can alter our decisions and actions thus reducing medication errors, increasing ordering of screening measures, and adhering to evidence based practice.

Despite all the evidence that CDS is beneficial, studies on CDS remain limited. Publication bias limits reporting of CDS that were harmful or offered no benefit. Most research still concentrates on the process of care and decision making. There are few controlled studies focused on patient outcomes, and more are needed.

Moreover, in any and every institution, vocal physicians tout their dissatisfaction with their EHR. Commonly avowed hindrances include poor usability, excessive uncompensated time spent interfacing with a keyboard, alert fatigue, and patient dissatisfaction caused by the physician spending time looking at a screen.

There are several factors affecting a clinician's acceptance of CDS suggestions, and ways to overcome limitations. To be effective, CDS must be carefully planned. When crafting a framework for CDS, it is helpful to consider the five rights (see above). Successful CDS includes the matching the user intention. For example, reminders to do things physicians intended to do are usually more accepted than an alert to reconsider. In terms of user control, disruptiveness and risk, too many alerts truly can cause a "wall of yellow" or "alert fatigue". Robert Wachter describes popular software in place at University of California, San Francisco, with advanced CDS. Alert fatigue allowed a common antibiotic to be delivered to a patient at a 39-fold overdose.<sup>9</sup> Alerts had been shown to a resident physician, nursing and pharmacy staff. All missed the alert buried among other warnings. Just like the wailing of multiple car alarms in an urban setting, the deafening noise ceases to have meaning.

<sup>6</sup> Buntin MB, Burke MF, Hoaglin MC, Blumenthal. The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health Aff (Millwood)*. 2011;30(3):464–71.

<sup>7</sup> Miskulin DC, et al. Computerized Decision Support for EPO Dosing in Hemodialysis Patients. *Am J Kidney Dis*. 2009 Dec;54(6):1081–8.

<sup>8</sup> Goodnough LT, et al. Improved blood utilization using real-time clinical decision support. *Transfusion*. 2014 May;54(5):1358–65.

<sup>9</sup> Wachter R. *The Digital Doctor: Hope, Hype and Harm at the Dawn of Medicine's Computer Age*. New York, NY: McGraw-Hill; 2015.

One way to overcome fatigue is the use of tiered alerts, which have been shown to improve acceptances of alerts. Tiered alerts vary in the degree of alert disruptiveness. User options are modified on the basis of the seriousness of the situation prompting the alert. For example, an alert with low-risk is shown passively while a life-threatening drug-drug interaction is displayed automatically and cannot be overcome. An alert in the middle may show up automatically but have an override option.<sup>10</sup>

### Implementation

Design and implementation play a large role in the success of CDS. The culture and leadership of the arena must also be considered. As CDS is used more, the potential for harm from CDS rises proportionally. Analysis has shown problems stem from system implementation issues as opposed to incorrect recommendations or intrinsic flaws. The Joint Commission emphasizes proper implementation, which involves users to resolve workflow and process problems before the CDS is live. Users need to be trained as well as have their performance monitored.

It cannot be overemphasized that a key issue with the use of CDS is workflow integration. Workflow assessment should occur early in the process. Clinicians should be consulted throughout the entire process of CDS design and implementation. Process improvement should drive workflow changes; implementation teams need to be aware of multiple workflow patterns. Departments, clinics and in many cases, individual physicians exhibit distinct requirements. A successful CDS is modified, as possible, to meet a physician's needs.

CDS interface issues are present in both data entry and output. A CDS integrated into an EHR is far more likely to be used than a stand-alone system requiring duplicate data entry. A clinician immersed in the processes required in the EHR is unlikely to seek advice by logging into a separate system, and typing in redundant data, but may be quite likely to accept the advice when conveniently presented to him during his usual documentation and order entry. Exceptions to this rule exist, as many users prefer other resources, or believe information outside the EHR is more accurate or comprehensive, such as when physicians use their smartphones to access UpToDate or Epocrates rather than use the EHR's embedded database. Some CDS is found by going to the internet for well known calculators that may not be integrated in the EHR. Examples include cardiac risk, fracture risk, and suicide risk calculations.

CDS requires local customization. Standards and transferability are ubiquitous issues. No national standards exist for which evidence-based guidelines should be built into CDS. As mentioned above, the goal is to embed guidelines, and the methods and scope remains to be defined and standardized. Similarly, there are no national standards for clinical data definitions. For example, there are different definitions of "tachycardia". How fast is too fast? Or what is anemia? How do you order a chest radiograph? Is it a CXR or a chest x-ray? Even when guidelines are accepted, local implementation varies by system. CDS cannot be used "out of the box." Much of the cost of CDS implementation is clinician time to select and design content, recognizing local definitions, language, needs and workflows.

### Maintenance

CDS remains an ongoing quality project. The practice of medicine changes, with new pharmaceuticals, new guidelines, and new technology. CDS must remain fluid and highly accurate. Should CDS become misleading, users will mistrust the entire system.

A fundamental source of error is incorrect data in the patient record. This is unfortunately common. As many people access one chart, the data upon which CDS is based must be verifiable and accurate. Any one person who is sloppy can lead to erroneous CDS in the future. (See Figure 3-15)

The Leapfrog group was created in November 2000 in response to the 1999 Institute of Medicine report "To Err is Human." This well publicized report underscored the cost and frequency of adverse drug events. The Leapfrog Group chose to address Computer Physician Order Entry (CPOE) as their first task due to the potential to lower patient harm from

<sup>10</sup> Slabodkin G. Cedars-Sinai reduces unnecessary care using EHR alerts. Health Data Management. May 22 2017.



**FIGURE 3-15**

Inappropriate alerts can induce provider dissatisfaction



medications. This decision was based on the fact that adverse drug events (ADEs) are one of the leading causes of iatrogenic patient injury. They noted examples such as prescribing a beta blocker to an asthma patient, or a medication metabolized by the kidney to a patient with compromised renal function.

Compounding the problem is the capricious nature of knowledge. On the first day of medical school, we are told that half of what we learn will be proven to be false. Unfortunately, we don't know which half. CDS specific problems occur as guidelines change, and require maintenance and monitoring. Maintaining CDS is as critical as the initial implementation. AHRQ has funded a CDS Consortium providing an online database of guidelines which may be adapted for local use.<sup>11</sup>

The end result is that CDS requires physician cooperation. Time pressures, arguably, are the greatest influence in the use or disuse of the CDS. Traditionally, physicians value their autonomy; however, this culture is changing, as the norm is to not only accept information from outside sources, but to insist on verification of our memories. CDS is not foolproof. The clinician must have a good sense of what should be right, and recognize that the CDS is only a tool. Clinical acumen and experience still override CDS, yet clinicians feel angst in the face of potential liability when overriding CDS.

When evaluating a CDS, look to the clinics and hospital, not the usability lab. It should be evaluated for its use in practice, not a controlled environment. Osheroff,<sup>12</sup> in his book, created the mnemonic METRIC or "Measure Everything That Really Impacts Customers." Customers are defined as all stakeholders including clinicians, patients and the care delivery organization. Outside of academic medical systems, there are few evaluations of CDS. Only a few randomized trials have been performed due to expense and level of difficulty. These are invariably sponsored by industry.

CMS incentivized electronic prescribing as part of Meaningful Use. E-prescribing has been shown to decrease errors and adverse drug events even without medication decision support. This has been attributed to elimination of hand writing illegibility and reduction of incomplete prescriptions by mandating a structured entry form. There is little evidence that patient safety has improved with this type of decision support. The best evidence for benefit is shown for drug-disease interaction checking and drug dosing.

<sup>11</sup> See <https://healthit.ahrq.gov/ahrq-funded-projects/clinical-decision-support-cds>

<sup>12</sup> Osheroff JA., Teich JA, et al. Improving Outcomes with Clinical Decision Support: An Implementer's Guide. 2nd Edition. Chicago, IL: HIMSS, 2012: p. 15.



Technological developments in the United States continue to facilitate the use of CDS. There has been an increased purchase and use of EHR systems. Funding and policy initiatives to improve systems, standardization and interoperability have also been created. Currently, the Commission for Certification of Healthcare Information Technology (CCHIT) is developing standards for CDS. In addition, healthcare workers now have the use of Internet resources and the promise of broad dissemination of CDS interventions.

In summary, over the last several years it has been recognized that a well-designed and implemented CDS can improve health care quality, increase efficiency and reduce healthcare costs. CDS should not be looked as a substitute for a clinician. Instead, it is an intervention which requires close examination of its goals, delivery and user audience. Clinicians need to understand its benefits and limits in order for a CDS to be optimized. This includes understanding the difficulties in designing and implementing CDS. CDS implementation requires integration into workflow. This requires compliance of the user. CDS alerts and recommendations will fail if this is not addressed. Researchers need to examine the informatics, structural, cognitive and workflow issues that have led to suboptimal CDS design or implementation which have led to limited use and effectiveness. Vendors must focus on the knowledge learned from research and development efforts that have focused on clinician efficiency. It is important to carefully look at the evaluations of commercial CDS systems in community settings to improve optimization of design, implementation and impact.

## 2.1.4 TRANSFORMATION OF KNOWLEDGE INTO CLINICAL DECISION SUPPORT TOOLS

Lisa Masson

### 2.1.4.1 Knowledge generation

The challenge presented in translating **knowledge** into computer systems is an obstacle to the application of clinical decision support in healthcare.

What is knowledge?

Knowledge may be explained as a subjective assimilation of information, merged with experience, situational awareness, and intellect. Wisdom assumes there is added awareness, morality and insight.

In medicine, the main source of knowledge is original research. Generating knowledge requires a cyclic process of publication, review, acceptance, and more publications.<sup>13</sup> Ultimately, knowledge is acquired from the intersection of evidence and preferences.

### 2.1.4.2 Knowledge acquisition

From the beginning of time with pen and paper, through the information age, the amount of literature published has increased at a terrifying rate. In 1913, Penzoldt of Erlangen, a wise German teacher, offered advice relating to the enormous growth of medical literature.<sup>14</sup> No mortal can keep up. The information age has moved us from sequestered information systems to a seemingly unrestrained exchange of information. Between 1665 and 2009, more than 50 million scientific papers were published, and approximately 2.5 million are added each year (Figure 3-16).<sup>15</sup>

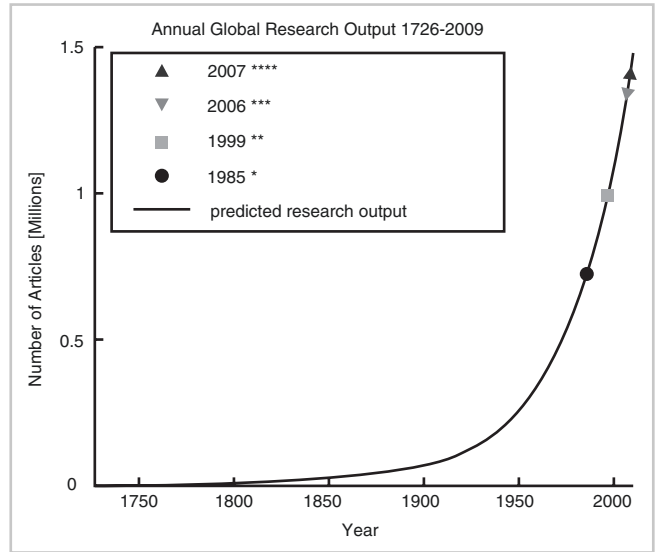
<sup>13</sup> William Hersh. *Information Retrieval: A Health and Biomedical Perspective*. Springer. 2009.

<sup>14</sup> December 25, 2013 *The Growth of Medical Literature* JAMA. 2013;310(24):2680. doi:<https://doi.org/10.1001/jama.2013.5484>.

<sup>15</sup> Jinha, Arif E. **Article 50 million: An estimate of the number of scholarly articles in existence** *Learned Publishing* 23(3):258–263 July 2010.

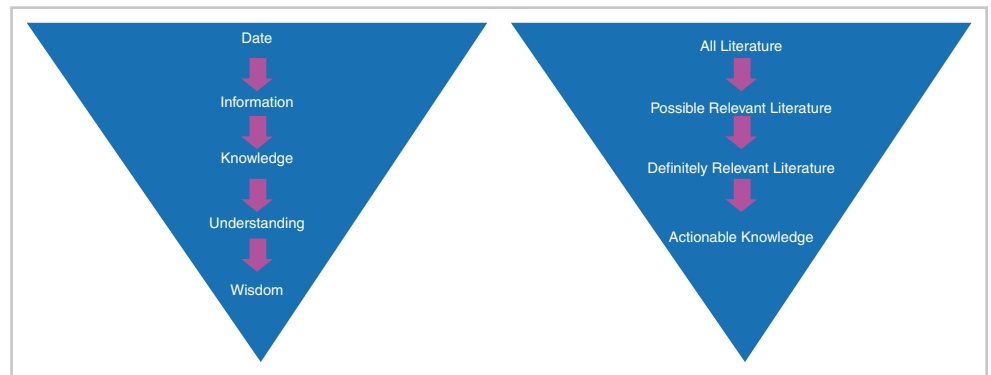
**FIGURE 3-16**

Estimated annual global research article output at 3% annual growth (from Jinha)



**FIGURE 3-17**

The pyramid of knowledge and the funnel of knowledge acquisition



The estimates suggest that our 1.8 zettabyte (which is 1.8 trillion gigabytes) of data will continue to grow with the number of servers managing our world’s data increasing an order of magnitude in the next decade.<sup>16</sup>

How is this lavish abundance of information turned into knowledge?

The Data-Information-Knowledge-Wisdom pyramid (funnel) emphasizes that as data becomes more refined, it also becomes more actionable. (Initially made prominent by Russell Ackoff in 1989, but the pyramid is recorded as far back as 1934 was referenced in a song by Frank Zappa “Packard Goose”).<sup>17</sup> The emphasis in this pyramid (or funnel) is on the knowledge being actionable. Remember there is far more to knowledge than data (Figure 3-17).<sup>18</sup>

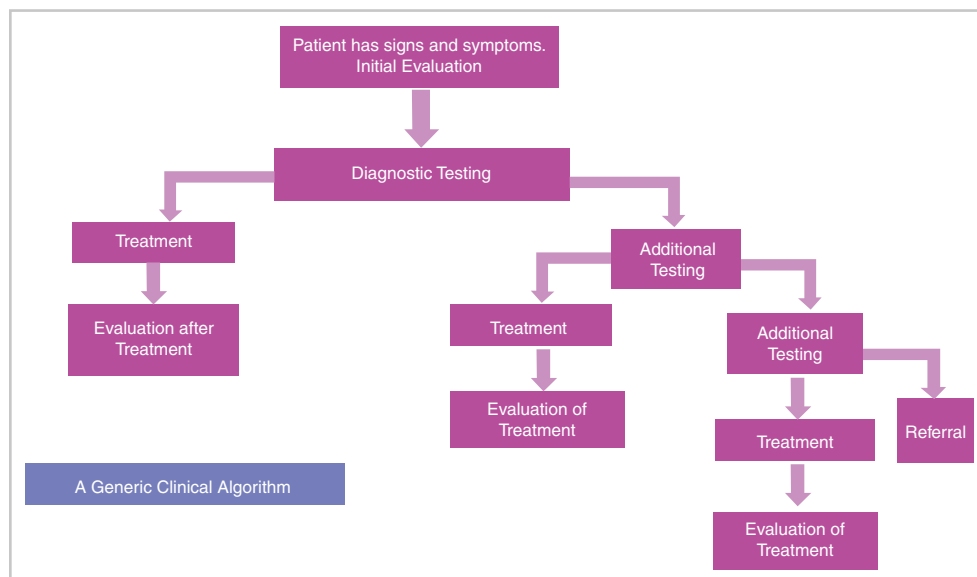
<sup>16</sup> Mearian, Lucas. World’s data will grow by 50x in the next decade, IDC study predicts. Computerworld from IDG. June 28, 2011.

<sup>17</sup> From Frank Zappa: Packard Goose Lyrics:

“Information is  
Not knowledge  
Knowledge is  
Not wisdom  
Wisdom is not truth  
Truth is not beauty  
Beauty is not love  
Love is not music  
Music is the best...  
Wisdom is the domain  
Of the wis  
(which is extinct).”

Read more: Frank Zappa—Packard Goose Lyrics | MetroLyrics.

<sup>18</sup> Weinberger, David. The Problem with the Data-Information-Knowledge-Wisdom Hierarchy. Harvard Business Review. February 02, 2010.



**FIGURE 3-18**  
Example clinical algorithm

In most cases, no single factor is sufficient to make a decision. Scientific evidence is complicated and often contradictory. Neither the patient nor the physician is without emotion (yet). Cultural, personal and temporal biases are always at play, consciously or unconsciously. Education and experience color the decision process. Other constraints include regulatory and financial concerns. Knowledge, and decision making, is muddled at best.

### 2.1.4.3 Knowledge modeling

**Knowledge Modeling** is a process of creating computer interpretable model of knowledge or standard specifications about a kind of process. This knowledge model can only be computer interpretable when it is expressed in some knowledge representation language or data structure that the software can interpret and that can be stored in a database or a file system.

The four general approaches to creating a knowledge model are: **(1) Clinical algorithms** **(2) Bayesian statistics** **(3) Production Rules** and **(4) Scoring and heuristics**. The current approaches avail themselves of the advances in electronic health record technology.

A clinical algorithm is defined as a systematic process through an ordered sequence of steps, with each step dependent on the outcome of the previous step. A clinical algorithm follows a path through a flow chart. A flow chart is composed of different types of nodes. Data is gathered at the information nodes; decision is made at decision nodes. Benefits of clinical algorithms include the ease of encoding the knowledge. In a flowchart, the knowledge is clearly expressed. Conspicuous limitations in the algorithmic approach is the inability to pursue new treatments or etiologies and the lack of accounting for prior results. Clinical algorithms are the precursors to clinical guidelines (Figure 3-18).

A modelling tool should realistically represent uncertainty and should be adaptive. Bayes' theorem calculates the likelihood of an event based on prior probability and new information, thus permitting inference from known quantities to make predictions and to learn from the new data. Bayes' theorem relies on two assumptions: **(1) conditional independence** (i.e. there is no relationship between different findings for a given disease); and **(2) mutual exclusivity** of conditions.

**Bayes' theorem**, as you learned in Sect. 2.1.2, tells us that the probability of a disease given one or more findings can be calculated from the prior probability of the disease and the probability of findings occurring in the disease. Unfortunately, Bayesian analysis is limited because findings in a disease are not conditionally independent and the diseases themselves are not mutually exclusive. If there are many diagnostic findings important to a diagnosis, the computational complexity increases rapidly.

**Production Rules** encode the knowledge as “if-then” rules. The system brings together the evidence from different rules to arrive at a conclusion. A rule-based system can use backward chaining or forward chaining. Backward chaining starts with a goal (or presumed diagnosis), and asks questions to determine if there is data to support that answer. In forward chaining the computer follows a defined path, similar to an algorithmic approach, to reach a conclusion.

MYCIN, developed as Edward Shortliffe’s PhD dissertation in the 1970s, was an innovative system in its use of production rules. MYCIN used backward-chaining deduction system to help physicians diagnose infections. The system assembled the observations entered by the clinician to arrive at a recommendation. As more rules were added, the recommendations improved. Shortliffe noted that a major lesson from the work on expert systems is that large knowledge bases must be built by successive additions of inductive and deductive steps.<sup>19</sup>

A production rules based system allows for maintenance with additions, deletions and modifications to rules. However, rule bases are large and difficult to maintain. MYCIN had 400 rules to cover two types of bacterial infection. The approach is best suited to a constrained domain. Also note the system was developed in an era prior to modern computers and graphical user interfaces. A session with MYCIN could take over 30 min to enter information. Personal computers had not yet been developed to allow for clinical interfaces.

In the quest to design a more comprehensive decision support system, a more scalable approach was necessary. Using a method of **scoring and heuristics**, knowledge is represented as profiles of findings found in diseases. These are then measured for importance and frequency for each disease.

INTERNIST-1 is a good example of using scoring and heuristics. This system was originally intended to mimic the expertise of an expert diagnostician. The novel approach in this knowledge model was in the way it responds to the user with follow up questions to narrow the field of possible diagnoses. In addition, the system scored findings from the history, exam and laboratory. The scoring included the likelihood of a disease given the finding (with zero being non-specific and five being pathognomonic) and the frequency of the finding given the disease (with one being rare and five being always). Properties included taboos, such as a male not being pregnant and not having ovarian cancer. This expert system performed equally as well as the experts in the *New England Journal of Medicine Cases*.<sup>20</sup>

The principles of INTERNIST-1 are used in DxPlain, still in existence and available for use today.<sup>21</sup>

Limitations to INTERNIST-1 included the long learning curve and time consuming data entry. The knowledge base was incomplete. Importantly, diagnostic dilemmas did not fit the information needs of most clinicians. INTERNIST-I was unable to construct a differential diagnosis spanning multiple problem areas and unable to explain its “thinking”.<sup>22</sup>

#### 2.1.4.4 Knowledge representation

As systems evolved through the 1980s and 1990s, it became apparent that the information provided did not meet the wants of the medical practices. Diagnostic support systems are less effective than therapeutic decision support systems.<sup>23</sup> Generally speaking, artificial intelligence and expert systems promised in the 1980s have yet to live up to the promise. As technology advances at a rapid rate, striving to help clinicians, diagnostic errors continue in medical practice. Thus the evolution through the 1990s to Clinical Decision Support (CDS) with recognized value in the electronic health records. **Clinical algorithms, Bayesian**

19 Buchanan B, Shortliffe E. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison Wesley, Reading, MA 1984.

20 Miller RA, Pople HE, Myers JD. Internist-1, an experimental computer-based diagnostic consultant for general internal medicine. *N Engl J Med*. 1982;307(8):468–76.

21 Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain. An evolving diagnostic decision- support system. *JAMA*. 1987;258(1):67–74.

22 Miller, RA, Pople HE, Myers, JD. *N Engl J Med* 1982; 307:468–476.

23 Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*. 2005;293(10).

statistics, Production Rules and Scoring and heuristics remain in use. These concepts are being implemented in large scales electronic health records.<sup>24</sup>

### 2.1.4.5 Knowledge management and maintenance

Maintaining knowledge bases for the foundation for CDS is not simple. Which knowledge base will be used? The abundance of published literature can not be consumed by a single human or committee. Information retrieval requires an understanding of databases, precision and recall. Translating information into decision support is a rapidly advancing science. Once in place, clinical decision support requires maintenance. Maintenance requires ongoing measurements and metrics examining the tools in place. These metrics may examine change in behavior, times the knowledge was acknowledged, or a change in outcomes. Maintenance requires that the decision support adapts to new clinical guidelines. Decisions around CDS require local governance, communication, training and support. Health systems must adapt appropriately to manage the complexity of CDS. Increasingly, this tsunami of information is managed by third party vendors. Examples of vendors of CDS and knowledge base solutions include Zynx, Lexicomp and Stanson Health.

Clinical decision support remains distinct from clinical decision automation. However, advances in natural language processing will aid knowledge acquisition by translating medical literature into a machine-usable format.

## 2.1.5 LEGAL, ETHICAL, AND REGULATORY ISSUES

Clinical decision support (CDS) systems are in widespread use in many institutions, bringing with them a host of legal, ethical and regulatory issues.

Software developers clearly have responsibility to reduce the risk of using a CDS to a level that is as low as reasonably practicable (ALARP). Despite an absence of case law, it seems reasonable that a supplier of medical systems would have some responsibility to patients who may suffer an adverse outcome as well as to providers who use the system in good faith.

Many systems contain End-User Licensing Agreements (EULAs) which claim that the software is provided “AS IS” without any claim or warranty of quality, fitness for purpose, completeness, accuracy or freedom from errors. These systems require users to acknowledge the EULA before using it. It is unclear to what extent these EULAs would protect a developer in the case where a CDS harmed a patient.

### Information Overload

In large, well-established institutions (such as the Veterans Administration), Electronic Health Records (EHRs) have been in use for many years, which means that the amount of data available for any given patient is voluminous. With the advent of regional health information exchanges (HIEs), even a non-institutional clinician now has access to vast amounts of patient data. From a practical standpoint, it would be nearly impossible for a clinician to review all this data in a timeframe reasonable to provide care. However, since all this data is readily available, a clinician who overlooks a critical detail could be held liable. A wily attorney could demonstrate the relative ease of searching the HIE portal to find the relevant information.

This risk becomes even more acute when a physician is a part of a care team. Suppose an internist orders a test which shows an abnormally low hemoglobin. The patient is referred to a specialist who orders more tests. Since the internist is still involved with the patient’s care, he daily reviews all new laboratory results, even those tests which he did not order. If the results of the specialist’s tests are not addressed, who is at fault? In the past, the internist would simply claim that he was unaware of the abnormal results of the test that he did not

<sup>24</sup> Osheroff J, Teich J, Levick D, Saldana L, Velasco F, Sittig D, Rogers K, Jenders R. (2012). *Improving Outcomes with Clinical Decision Support: An Implementer’s Guide* (2nd Ed.). Chicago, IL: Healthcare Information and Management Systems Society (HIMSS).

order. However, since modern EHRs provide excellent auditing procedures, it would be trivial to demonstrate that the internist had viewed the results at a certain time and date. As the number of providers increases, the liability could conceivably extend to each of them.

#### **Copy and Paste**

Most EHR progress notes are composed largely of boilerplate. This may arise as the result of a common note template or because of the use of “macros” which are short phrases which automatically expand into a larger string. In some cases, notes contain text that is wholesale copied from other provider notes, which may be outdated, inaccurate or irrelevant. However, since the EHR provides the only written record of the patient encounter and is digitally signed, it would be very difficult for a clinician to disavow responsibility. Moreover, this creates billing and reimbursement problems. If notes are not significantly different from 1 day to the next, a payor may argue that the provider did not provide any service and could deny payment for that day.

#### **Order Sets**

Just as notes can be copied and pasted without thinking, orders can be placed according to miswritten or incomplete protocols. For example, suppose an EHR vendor creates an order set for patients who are short of breath. Included in the orders is a D-Dimer test, which has an unfortunately high false-positive rate. If this test is ordered indiscriminately over a large number of patients, providers will feel obligated to order follow-up tests (such as CT scans of the chest) in order to explain the abnormal D-Dimer results, causing unwarranted exposure to radiation and needlessly increasing the cost of care.

#### **Alert Fatigue**

Clinical Decision Support systems have grown in scope and complexity over the years. Institutions and EHR vendors are motivated to keep the number of alerts as high as possible, relying on clinicians to differentiate the important alarms from background noise. It is unclear to what extent a provider will increase his liability when a bad outcome can be linked to an alert that was ignored. Similarly, there are many instances when a CDS failed to fire because of a programming error. The liability for missing a clinical opportunity rests on the provider, even though it is his custom to rely on the CDS to protect him.

#### **Privacy Breaches**

The Health Insurance Portability and Accountability Act (HIPAA, see Sect. 3.1.4.1) codifies many privacy rules. Clinicians (or non-clinicians) who inappropriately or inadvertently access patient records can be fired or face legal retribution.

## **2.1.6 QUALITY AND SAFETY ISSUES**

There is now ample evidence that clinical decision support (CDS) systems can make meaningful contributions to patient care. Every informaticist hopes that that contribution is positive, or at least mostly positive. For this reason, great effort must be spent ensuring the quality and safety of CDSs. (See Sect. 2.1.5, for legal and ethical issues surrounding CDS)

Fox<sup>25</sup> describes four primary approaches to safety and quality in CDS

1. Use rigorous software engineering to ensure reliability
2. Systematic quality control for the medical component of the CDS
3. Hazard management during system operation
4. Comprehensive auditing to allow quality reviews

#### **Quality Engineering**

Software is created within a “development lifecycle” (see Sect. 3.5.6) under certain quality standards, such as ISO 9000. Unfortunately, no software is perfect and errors will always

<sup>25</sup> Fox J, Thomson R. Clinical decision support systems: A discussion of quality, safety and legal liability issues. Proc AMIA Symp. 2002. 265–9.



exist. There will always be a trade-off between costs of development, ease of use and maximizing safety.

Fox suggests the use of Hazards and Operability Analysis (HAZOP) to analyze risk and to stratify into categories.

1. **Risk level 1.** There are significant and avoidable hazards that could be caused by the CDS. For example, the CDS may recommend a dangerous medication or intervention.
2. **Risk level 2.** There is no direct hazard associated with the CDS, but it may lead to a situation where a beneficial intervention is overlooked. For example, the CDS normally advises the clinician to check lead levels in healthy children, but fails to do so.
3. **Risk level 3.** There is no direct or indirect hazard with the CDS, but it fails to anticipate future conditions. For example, it fails to recognize that prolonged use of antibiotics may lead to *C. difficile* colitis.
4. **Risk level 4.** There are no identified risks.

It is important to remember that there are two components of quality engineering for a CDS. On one hand is the technology aspect, such as user interface, data access, storage, presentation methods, etc. On the other hand is the medical knowledge repository upon which recommendations are given. The problem with medical knowledge is that it is relatively dynamic and the software is only able to encapsulate the consensus of experts at a given time. As time goes on and new discoveries are made, recommended therapies may change. The developer of CDS has to ensure that its advice remains as complete and up-to-date as possible.

In addition, it seems prudent to have a medical professional (or committee) review the CDS recommendations periodically. This review should include evaluation of the CDS rules in plain text as well as applying the CDS to known test cases to check for a desired outcome. An auditing trail should be available for forensic analysis if the system does not behave as expected.

### Safety Concerns

A CDS that is well-designed and properly implemented with high-grade medical information can still give bad advice, especially in atypical situations or patients with unusual combinations of diseases. For example, consider a patient with acutely decompensated heart failure (HF) as well as diabetic ketoacidosis (DKA). The CDS rule for DKA may recommend an intravenous fluid bolus, while the CDS rule for HF would list IV fluid as a contraindication. Unlike a human, the CDS does not know which rule it should favor, or to what extent. Its output could be unpredictable. Therefore, it is vital to test the system with as many cases as possible so as to minimize the unexpected failures.

Errors may also arise from the method in which a CDS is used.<sup>26</sup> Commercial CDS will not be tailored to the host institution, and may recommend therapies or drugs that are not locally available. Moreover, the in-house health information technology (HIT) staff may not have sufficient training or permissions to modify the CDS.

Errors may also arise from the method in which the CDS is used. For example, with a paper system, the physician may examine the patient at the bedside and immediately write medication orders. If the patient objects to the medication because of allergy or some other reason, he can quickly change to another option. However, a complicated electronic prescribing system removes the physician from the bedside and prevents direct interaction with the patient. Without that two-way communication, the problem may not be detected until much later, if at all.

Order entry (OE) systems require input of structured data, and may trigger an error when information is incomplete. Unfortunately, the very structuring of data may prevent nuances of orders to be placed. For example, most OE systems require a flow rate for IV fluids (e.g. 1000 mL/h). In the setting of trauma, the first liter is often given “wide open” (i.e. as fast as

26 Coiera E, Westbrook JI, Wyatt JC. The safety and quality of decision support systems. *IMIA Yearb Med Inform* 2006:20–25.

gravity allows). Many OE systems will only accept a numerical flow rate, and not “wide open” as a rate of infusion, which essentially prohibits placing the order as desired.<sup>27</sup>

In institutions where the CDS has been present for longer times, users become completely reliant on the automation it provides. Many users will trust the CDS to do accurate drug-drug interaction checking and will be less vigilant because they assume that the computer is able to check orders more thoroughly than they could themselves. Similarly, when the CDS recommends a certain care path, users will accept the advice, even if it runs contrary to their own training because they assume that either the computer is correct, or there is some institutional motivation to proceed a certain way. The opposite problem is alert fatigue, where the CDS gives so many false-positive alerts, that the user begins to ignore all alerts.

#### ■ Safety Recommendations

1. Usage of the CDS should be limited to persons who are properly trained to understand its capacities and its limitations
2. The CDS should provide robust auditing functionality so that if it does not function as expected, debugging will be much easier
3. End users should be in direct contact with the CDS maintainers to resolve problems before they affect other patients.
4. The medical knowledge of the CDS must be periodically reviewed for accuracy, efficacy and currency. Numerous test cases should be applied to see how the system responds to rare or unusual combinations of conditions.
5. Usability testing must be done on each of the warnings to ensure that the advice is reasonable for the setting in which it is used. If it appears that certain warnings are being ignored, it may be necessary to review the appropriateness of that warning to that setting.
6. When users become too familiar with CDS, they give up autonomous thought and trust the CDS too much. Alternatively, if the CDS gives inappropriate warnings too frequently, they will ignore it.

## 2.1.7 SUPPORTING DECISIONS FOR POPULATIONS OF PATIENTS

Public and widespread reporting of infectious disease enables early detection and control of epidemics. For example, when a cluster of cases is detected and reported to a public health agency, it can educate and warn local providers about diagnostic and therapeutic options.

The majority of Clinical Decision Support (CDS) systems are aimed at improving the health of individual patients. While Electronic Health Records (EHRs) are commonly used to aggregate data, they are only now gaining purchase for helping to inform decisions on population health. In stage 2 of the EHR incentive program (i.e. Meaningful Use), one of the menu options included transmission of structured data to a public health agency. Standards and protocols for this type of data transfer are still evolving. The hope is that automated public health data collection systems can be used to detect outbreaks more rapidly than traditional systems which rely on human analysis of disease trends.

There are seven elements of public health informatics.<sup>28</sup>

<sup>27</sup> In response to problems like this, users will enter the largest number they can before triggering the high rate alarm. At my institution, this is 4000 mL/h. But this still does not address the issue. What should the nurse do gravity only results in a flow rate of 3000 mL/h? Must she use a pump? The OE system would imply she should.

<sup>28</sup> Savel TG, Foldy S. The Role of Public Health Informatics in Enhancing Public Health Surveillance. *MMWR* 2012. 61:20–24.

1. Planning and system design—identifying readily available information sources that best inform a particular surveillance goal.
2. Data collection—combining data from various sources and differing formats, including both structured and unstructured data
3. Data management—identifying and correcting data anomalies and inconsistencies;
4. Analysis—Using statistics and data visualization techniques to understand the data; creating alarms and thresholds for initiating public warnings.
5. Interpretation—determining the actual utility of the surveillance data by reviewing it in the context of other available data
6. Dissemination—optimizing the broadcast of this information to the intended audiences
7. Application to public health programs—linking health care providers and public health workers to this new information and assessing its value in clinical decision making.

In the past, public health agencies were primarily concerned with confirmed diagnoses, such as the number of cases of culture-proven methicillin resistant *Staphylococcus aureus*. However detection can be enhanced by simply searching for clusters of symptoms. According to the World Health Organization, **syndromic surveillance** is “the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice”<sup>29</sup>

For example, a Health Information Exchange (HIE) collects Emergency Department triage data from several hospitals and aggregates the information into a single repository. Automated review of the free-text patient complaints reveals a higher-than-expected rate of gastrointestinal illness. This triggers the release of a public health advisory. Meanwhile, epidemiologists determine that the cluster of cases are likely related to a particular eatery. Inspection of the restaurant reveals rotavirus and it is temporarily closed in order to protect the public health.

Surveillance data does not always originate in medical encounters. In 2008, Google famously claimed that it could predict influenza trends by analyzing search data.<sup>30</sup> Unfortunately, their prediction model was not so successful during subsequent years.<sup>31</sup> This highlights the utility of using surrogate markers (such as search engine flu queries) to predict actual disease.

---

29 WHO | Public health surveillance [Internet]. Who.int. 2017 [cited 4 March 2017]. Available from: [http://www.who.int/topics/public\\_health\\_surveillance/en/](http://www.who.int/topics/public_health_surveillance/en/)

30 Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. *Nature* 2009. 457:1012–1014.

31 Butler D. When Google got flu wrong US outbreak foxes a leading web-based method for tracking seasonal flu. *Nature* 2013. 494:155–156.