A Framework and a Coding Manual for the Investigation of the Interpretation and Utilization of Evidence

Vimla L. Patel, PhD, DSc
patel@hebb.psych.mcgill.ca

José F. Arocha, PhD
francisc@med.mcgill.ca

Cognitive Studies in Medicine: Centre for Medical Education,
McGill University, Montreal, Canada

These studies were supported in part by funds from the National Library of Medicine through the Shared Internet Server for Delivering Guidelines (Columbia University, Edward Shortliffe, PI) subcontract to Vimla L. Patel, McGill University, American College of Physicians-American Society of Internal Medicine, and the Social Science and Humanities Council of Canada.

Correspondence should be sent to:
Vimla L. Patel
Centre for Medical Education
1110 Pine Ave. West
Montreal, QC
H3A 1A3

Canada
Phone (514)398-4989
Fax (514) 398-7246

1. Introduction: Evidence-based Medicine and the Nature of Evidence

Although the term "evidence" is commonly used in many contexts, nowhere is evidence more important than in science and some professional fields (e.g., law), where claims to validity of theories or hypotheses are based on their degree of empirical support. We require that the use of medical procedures or drugs be supported on scientific evidence, rather than on expert opinion, on the assumption that scientific evidence is more credible than other types of evidence (e.g., personal, anecdotal). Although it is sometimes expressed that scientific knowledge claims to validity are not more secure than other types of knowledge, the overwhelming success of science in the last few hundred years speaks loudly of the importance of scientific evidence in our society.

Sciences that have not reached the level of maturity of the harder sciences, still debate the role that evidence plays in their development and their application to real-world problems. In the health domains, evidence-based medicine (EBM) has been advertised as a possible integrative force for medical practice (Davidoff 1999). Some reasons for this are that EBM would provide the best quality care, as it is based on sound empirical support, and that it would serve to standardize the practice of medicine, by allowing physicians to follow clinical practice guidelines that are built on scientific evidence. There are, however, some reservations as to the applicability of EBM in clinical practice. Two arguments are often aired. One argument is that there is not enough scientific evidence to support clinical practice. Only a few well known problems can be explicated on sound scientific evidence (Reilly, Hart, and Evans 1998). The second argument is that given that scientifically sound clinical evidence is based on populations, its application to particular patients is not straightforward, and may not even suggest the most appropriate course of action for the physician and the most benefit for the patient.

2. The Role of Evidence in Health Sciences

Medicine has been conceived of as a science and as an art. The scientific aspects of medicine consist of the wealth of knowledge coming from the biomedical researcher's laboratory and from the epidemiologist desk. The "artistic" aspects derive from the physician's experience, in the form of intuitions, rules of thumbs, heuristics, and "remindings" of past experiences with particular patients. These two aspects, in some way, are reconciled in medical practice, to the extent that medicine makes successful use of both scientific and experiential knowledge (Enkin and Jadad 1998). However, as scientific knowledge of medicine accumulates, there has been an increasing emphasis on the science of medicine, in detriment to its art. The need to establish the foundations for EBM is asserted on the basis of the superiority of scientific knowledge over experiential knowledge and on the necessity to standardize clinical practice. Although no physician is against the idea of EBM per se, there has been a long debate over the alleged supremacy of EBM over traditional medicine.

EBM proponents argue in favor of promoting the use of scientific research, such as results from randomized controlled trials, as the foundation for clinical practice. Critics of EBM vary in their claims, ranging from questioning the validity of current scientific clinical studies (Miettinen 1998) to the relevance of population-based evidence for particular patients to emphasizing the importance of experience-based knowledge of the individual physician (McDonald 1996) to pointing out the impervious influence of environmental factors on health and disease(Aveyard 1997).

Despite the divergence of opinions, it is now widely accepted that clinical practice would benefit from a greater consideration of evidence than it has been given up to now. Even those who question the validity of today's clinical evidence propose, not its dismissal, but the improvement of clinical studies (Miettinen 1998). Furthermore, proponents of EBM have revised their notion of what constitutes evidence by acknowledging the multiple character of evidence, which in their view now includes expert opinion and practical knowledge. Consider the following definition of EBM:

> Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. By individual clinical expertise we mean the proficiency and judgement that individual clinicians acquire through clinical experience and clinical practice. Increased expertise is reflected in many ways, but especially in more effective and efficient diagnosis and in the more thoughtful identification and compassionate use of individual patients' predicaments, rights, and preferences in making clinical decisions about their care. By best available external clinical evidence we mean clinically relevant research, often from the basic sciences of medicine, but especially from patient centered clinical research into the accuracy and precision of diagnostic tests (including the clinical examination), the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative, and preventive regimens. External clinical evidence both invalidates previously accepted diagnostic tests and treatments an d replaces them with new ones that are more

> powerful, more accurate, more efficacious, and safer." (Sackett et al. 1996, p. 71)

Thus, the definition includes the use of experience-based knowledge about medicine, as well as properly conducted research studies. Furthermore, the consideration of other factors have been investigated by the use of methodologies and theories that rest on qualitative paradigm. This type of research (Green and Britten 1998) is becoming increasingly accepted as a valid way of research in various fields of health care. What needs to be done is to examine the extent to which each form of evidence (experience-based or external) contributes to interpreting and evaluating evidence. This involves an elucidation of the concepts of evidence and an investigation on the interpretation and use of evidence in clinical practice.

3. Approaches to the Study of Evidence

The study of evidence has been approached from various viewpoints. For the most part, these approaches have remained in isolation from one another. One of the most commonly used perspective has been to look at evidence in terms of the place it takes in an argument (Kuhn 1991). An advantage of this perspective is that it provides a domain-general way to investigating the use of evidence that can be accommodated to various disciplines, such as philosophy, psychology, anthropology, etc. Research on evidence as argument covers studies in philosophy (Toulmin, Rieke, and Janik 1984), psychology (Kuhn 1991), medical pragmatic reasoning (Dickinson 1998; Horton 1998). The approach thus allows a truly interdisciplinary approach to the study of evidence as argumentation. A limitation is that it focuses only on those aspects of evidence interpretation and use that rely on a verbal, symbolic, level, where other approaches are more useful.

A set of perspectives derives from the study of the psychological process involved in acting in the world. There are several sources for these perspectives. One such perspective originates in cognitive science studies (e.g., scientific reasoning) , where the emphasis has been on the memory and knowledge structures and the strategies that are used by people when generating or evaluating evidence. The cognitive science perspective encompasses several paradigms. First, the symbolic paradigm, where the emphasis has been on the study of high-level cognition (e.g., problem-solving, decision-making). Second, a "situated" perspective where the emphasis is on how the environment is used by people to supports their decisions and problem solving. Third, a control perspective emphasizing action in the world (i.e., how people are able to have control over aspects of the world that are important to them) and which takes account of both the symbolic and the situated within a single framework while providing new insights into the nature of cognition and action. A fourth perspective, in the health care domain, has been concerned with determining the role that scientific evidence plays in medical decision making (e.g., in screening, diagnosis, and treatment). A useful conceptual tool for understanding the nature and utilization of evidence has been provided by Evans and Gadd (1989; Patel, Evans, and Kaufman 1989). This framework is used in this manual to characterize the types of medical knowledge and how evidence can be interpreted within it.

The next section presents a short overview of the argumentative perspective on evidence interpretation and use. We will not dwell on this perspective, as other publications (Dickinson 1998) present a thorough description of such an approach. Next, we describe the basic models for action- and cognition-based investigations of the interpretation of evidence. These models are then extended to and expanded by the study of the nature and structure of medical knowledge Finally, we present the coding manual, based on our framework. The coding manual attempts to unify the theoretical frameworks into a multi-dimensional coding tool.

3.1. Evidence as Argumentation

Researchers in various fields (e.g., Dickinson 1998; Horton 1998; Kuhn 1991; Stein and Miller 1993) have conceived of the study of evidence in terms of the role that evidential information plays in argumentation. The focus of these researchers has been diverse. While some researchers (e.g., (Dickinson 1998; Horton 1998) have taken a normative approach by developing a framework for the categorization of evidence based on its argumentative strength, cognitive scientists (Kuhn 1992; Stein and Miller 1993) have carried out empirical studies of how people reason while using evidence to support their reasoning. Within this perspective, evidence is viewed as a component in a process of reasoning, where several rhetorical components of an argument (e.g., data, claim, counter-argument) are identified. This perspective attempts to understand the extent to which people’s argumentation in favor of a claim (e.g., a hypothesis), matches sound argumentation as dictated by a normative perspective, based on the theory of natural reasoning (Toulmin 1958). The theory of natural reasoning was developed to account for informal reasoning and argumentation. Although sound practical reasoning should not collide with formal reasoning (as exemplified by standard logic), the former is not constrained by the rigid limitations of logical derivations (as we saw in the case of Popper’s model, confirmation or disconfirmation of a hypothesis requires more than following a prescribed logical procedure).

Within cognitive perspective, researchers have emphasized the nature and development of the reasoning process (e.g., strategies, developmental trends in human argumentation), rather than the types of evidence that are used. (Kuhn 1991) investigated the skills that children and adults use when arguing about various topics (e.g., causes of crime, educational failure) and provided evidence in support of their arguments. She used an open interview format in which subjects were questioned on scenarios dealing with the causes of social problems (e.g., crime, educational failure). She categorized evidence into either evidence and pseudo-evidence. Evidence consisted of arguments that presented supporting information in either causal or empirical forms. identified a number of types of evidence used by people, such as covariation (the most common form of evidence), analogy, Pseudoevidence consisted of scenarios describing how something might happen but without showing that it actually happens. Pseudoevidence can be seen simply as a form of re-description of the problem. She found that less than half of the subjects used real evidence to support their arguments, preferring instead to use various forms of pseudo-evidence.

Looking at evidence as part of argumentation provides a form of both describing the argumentative process involved in EBM and the possibility of prescribing some form of sound argumentation. However, in order to claim prescriptive status, an approach has to be securely preferred in terms of the quality of the processes or the outcomes it is prescribing. Given that no guarantee can be made that by following the argumentative approach one produces better decisions, in this paper, we emphasize the cognitive aspects over the argumentation aspects[1].

However, given that evidence is always presented as part of an argument, it would be insightful to investigate how the form of an argument constrains the acceptability of evidence. One may hypothesize that the structure of an argument, and the role that evidentiary statements play in such structure, provide useful information for designing evidence-based guidelines. This information can be used in the design of textual information that optimizes the quality of an argument, by using supporting evidence that is likely to be accepted by guideline users.

3.2. Evidence in the Study of Cognition and Action

A piece of data constitutes evidence to the extent that a person recognizes it and represents it as such. In this regard, evidence is primarily a perceptual or cognitive category. This means that evidence, as a cognitive representation, is subject to the same constraints that any cognitive representation is.

Evidence is subject to constraints that derive from the way people function in everyday life. Since people are intentional systems (e.g., they act to produce some desired result), evidence is constrained by such intentions or goals. In other words, in examining data, the aspects of the data that become the focus of attention may change, depending on the intentions or goals of the person (e.g., to support a hypothesis, look for data that is consistent with the hypothesis). Thus, it is important to understand the person's intentions or goals when developing a framework for understanding evidence. An intentional perspective on the

---

[1] Normative approaches to other aspects of human behavior, which are also said to be normative, include decision-theoretic approach to decision making. However, the same reservations apply to this approach (Allais and Hagen 1979; Bunge 1985).

investigation of evidence calls for a framework capable of explicating how intentions interact with perceptions of the evidence and the actions that ensue from this interaction. It also calls for an understanding of the environmental constraints on both perception and action. A model will be presented next that we think may serve as the basis of an account that takes care of all of these aspects.

3.3. A Basic Perceptual/Cognitive Model of Evidence Interpretation

Understanding any human activity, such as evaluating evidence, involves explicating the interplay between knowledge, goals, perceptions, and actions and how these are constrained by the physical and social and cultural environment. Goals are desired states of some perceived environmental variables (e.g., some perceived state of the world outside the person). They determine how the evidence is interpreted and used by focusing the one's attention toward some parts of the environment (e.g., searching for a particular relationship between two variables) instead of others (e.g., making sure that the variables have been appropriately identified). Goals are also important in the evaluation of evidence, as they serve as standards for comparison with perceptions of the situation.

Perceptions are representations inside a person of some current state of the world. Although many of our everyday perceptions are unimportant to us, since they affect us in little or no way, there are also many perceptions that we care about and that we try to bring to a desired state (e.g., a goal). These are called controlled perceptions. To identify the perceptions that someone is controlling, it is important to understand, first that perceptions correspond to properties of things, rather than the things themselves (e.g., an object) and that these vary (i.e., can be in various states). For example, although we normally talk about controlling a car (i.e., an object) in a winding road, we actually control some variable aspect of the car, such as its speed or its position on the road. The same occurs with evidence: people control some variable aspect of the evidence, not the evidence itself. This is important because it accounts for the great variability that exists in the interpretation of same piece of evidence. As evidence is composed of possibly many variables, some of which may be controlled, different people may focus on different variable aspects of the evidence.

When attempting to explain people's behavior in everyday life, most often we are concerned with understanding not what actions they perform, but what intentions (i.e., goals) exist behind those actions. Thus, we ordinarily name a person's actions in terms of the apparent intentions of the person, not in terms of his or her movements. For instance, when we see someone sitting at a desk, looking downward with a pen in his or her hand and ask, 'What are you doing?' the answer we would look for would be more likely something like 'Writing a paper,' or 'Doing my taxes' rather than 'making inscriptions on a piece of paper by holding a pen in my hand and moving it from left to right.' Similarly, when investigating human behavior, the level of description that we find appropriate is dependent on the intentions of the actor. One reason for this is that there are many possible variations in actions that realize the same intention. Writing a paper can be done by typing on a keyboard or by talking aloud to a tape recorder and then transcribing the tape. Such dissimilar patterns of actions can be meaningfully described as 'writing a paper'. A second reason to pay particular attention to the intentions of the actor is that actions can produce consequences that are unintended by the actor. Furthermore, in many cases, observers cannot distinguish between intended and unintended consequences. When explaining behavior, we would like to distinguish between what the person is trying to do (e.g., reaching for a cup at the other end of the table) from accidental consequences of behaving (e.g., spilling the coffee, falling over the table).

An interesting aspect of this distinction arises from the fact that unintended consequences are not simply random, but are systematically related to what the person is doing, although they may be accounted for by the environmental factors. An example is driving. The movements of the steering wheel are systematically related to attempting to keep the car on course, but they are accounted for by variations in road conditions, effect of wind, and other environmental effects.

Any task to be carried out involves a goal or, more likely, a set of goals, where the goal (e.g., write a paper), rather than the actions performed (e.g., scribbling on a piece of paper), define the task. Goals are structured in hierarchies, where the reaching of a goal serves as a means to achieve another goal (goals that serve to achieve another goal are sometimes called "sub-goals"). Goals can be set as points in a continuum (e.g., write 3 papers per year), ranges in a continuum (e.g., generate a document of between 3000-5000 words) or programs specified as IF-THEN condition/action pairs (e.g., IF there is a paper on the topic, THEN-review it; IF there is no guideline, THEN write a new one). Goals can also be specified in 'fuzzier' terms, such as 'greater', lower', 'more' or 'better.' It is important to note that goals are not fixed, and furthermore, they are likely to change, either by refining them, replacing them or abandoning them.

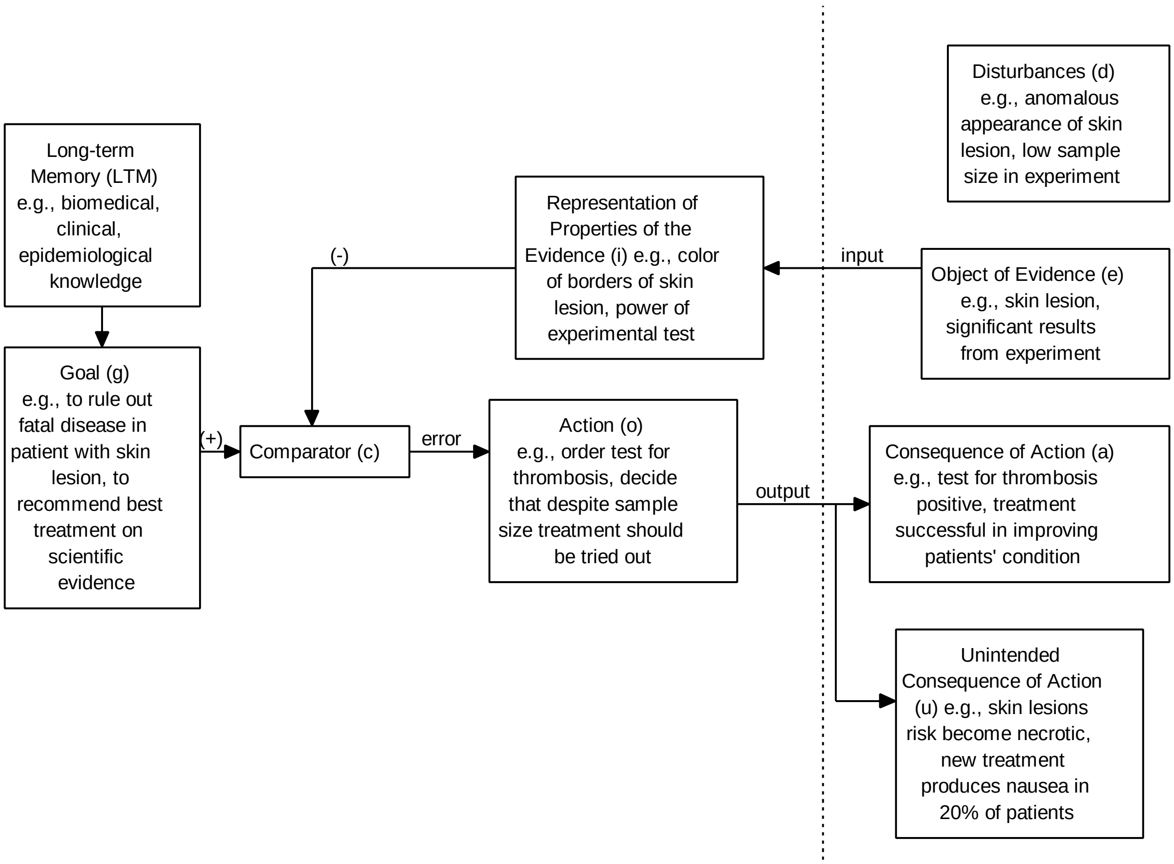

Figure 1. A goal-based model of interpretation of evidence

A goal-based theoretical framework serves as a foundation for characterizing how people interpret the world around them, including the interpretation and use of evidence. The relationships among these functions form a causal loop, where each function concurrently modifies the next function in the loopin particular ways. A current perception of the environment is matched against a goal. If there is a disparity between the goal and the perception, an action takes place that brings the perception in alignment with the goal. These actions produce consequences, some of which are intended (i.e., getting closer to or realizing the goal) and some of which are unintended (i.e., produce unwanted affects that need to be further removed by other actions). For the process of goal attainment to be successful, the actions have to vary in any way needed to change the environment in the intended direction, and therefore change our perception of the world. Thus, variability of actions is a necessary condition for success (i.e., consistently producing the desired consequences). During this process, environmental factors frequently involve disturbances that need to be overcome to reach the goal. These disturbances refer to anything that makes the attainment of the goal difficult, typically resulting from the changing and unpredictable nature of the environment (e.g., some live cultures to be used today in an experiment die; the preparation of the experimental materials suffers a delay; experimental subjects drop out of the study). In most circumstances, these disturbances are successfully removed by "appropriate" actions, so that goals are achieved (e.g., the experiment provides the desired results). These actions cannot be planned as they have to change since environmental conditions often change rapidly and unpredictably. However, as the time frame between the setting of the goal and the action increases, the possibility of prediction, and therefore planning, also increases (e.g., by having a mental model underlying some outcome).

The following example may illustrate how the model can be used to understand how people interpret evidence. Two expert dermatologists, who observed an elderly patient with an ulcerated lesion in her legs have a discussion. The experts agree on the significance of the lesions (they both see the lesions), but markedly differ in the focus of attention they give to variable aspects of the ulcers. Whereas one physician focuses on the size and the bilateral location of the ulcers (they're in both legs), the other physician focuses on their colour and degree of inflammation. The example illustrates how the same piece of patient evidence was used by the two physicians to support their preferred interpretation, by focusing on different aspects of the same evidence. This suggests that evidence has a variable character and not, as it is often assumed, a fixed easily observed nature. What was critical about the main piece of patient evidence was not just that the ulcers were there, but that their various characteristics (e.g., their colour; their size, their location, their degree of inflammation) suggested different hypotheses to each physician depending on what variable aspect was the focus of attention. Why were these two dermatologists focusing on different aspects of the ulcers? A hypothesis is that at a higher level they may be concerned about different things. One physician may be concerned about the possible death of the patient by focusing his attention to the more severe aspects of the evidence, the more symptomatic of a "systemic" problem, and trying to rule out this possibility (e.g., the bilateral location may suggests some thrombosis involving the larger arteries which may lead to death). This may have led him to focus on the bilateral location while not paying too much attention to the degree of inflammation of the ulcers (a more localized problem, not likely leading to death). The other physician, instead, may have been more concerned with getting the diagnosis right to treat the patient successfully, thus focusing on the more salient properties of the ulcers (e.g., size and color). We can then look at evidence interpretation as a function of the goals of the two physicians (e.g., saving the patient's life vs. getting the correct diagnosis).

In this process, long-term memory (e.g., prior knowledge of the medical domain, years of clinical experience, familiarity with the topic of research) also plays an important role in the investigation of expertise effects on evidence interpretation, since people with different quality or degrees of expertise will interpret information differently (Allen, Arocha, and Patel 1998; Arocha and Patel 1995). Representation involves the encoding of the evidential information present in external objects (e.g., a clinical case finding) or symbols (e.g., an experimental result from a clinical trial) and the construction of a mental image of these. In this process, a distinction is made between the "literal" representation of the information gathered from the external objective or symbolic stimuli and the inferential representation of the stimuli, which is dependent on our goals, prior knowledge, and experience. The mental representation that is constructed results from the "interaction" between what is actually there in the world (e.g., a skin lesion), and our knowledge in memory, which is composed of traces of our past experience (e.g., prior knowledge of the domain or prior exposure to similar clinical conditions serves to focus our attention on particular aspects of the lesion, such as location, while de-emphasizing others, like size, or color). As someone develops into an expert in a domain, his or her representations of those stimuli increasingly become more dependent on specialized knowledge and memory and less on "literal" representations. Furthermore, experts also become more capable of making finer discriminations, which results in greater "flexibility" in interpretation. However, they sometimes become "inflexible" thinkers. This may occur, for instance, when a problem looks extremely familiar and automatic, semi-conscious routines are applied to its interpretation and solution. Research (Patel, Arocha, and Kaufman 1994) has reported cases of experts who can retrieve well known problem solutions from memory without realizing that the problem contained some anomalous features that remained undetected, sometimes reporting recalling problem features that were not in the problem.

4. Strategies and Operations in Evidence Generation and Interpretation

The study of the use of strategies in generating and interpreting evidence has been carried out for several decades. These studies often focus on the ways that people (both lay people and scientists) tend to show that their hypotheses are correct. One early motivation for this research was the normative thesis, held by the philosopher Karl Popper (1959; Shahar 1997), that confirmation strategies—that is attempting to provide data that supports one's hypothesis—is logically invalid and that the only sound strategy is to search for information that disconfirms or falsifies one's hypothesis. Popper argued that the logic of confirmation, e.g., as it is practiced by scientists, amounts to the logically fallacious form known of affirming the consequent. That is, $p \supset q,\ q\ \therefore\ p$[1]. He instead argued for disconfirmation, whose logically sound form is $p \supset q,\ \neg\ p$, or modus tollens[2]. Although Popper's proposal may be logically sound, interpreting evidence—as in scientific research— is not only a logical task ; it involves much more than logic. Criticisms of Popper's argument have been forwarded by many philosophers, but perhaps the stronger one consists of the qualification that, in science we never have the pure logical form of the Popperian argument, but that instead the testing is done with the help of auxiliary hypotheses. The correct formulation of the argument should add then the auxiliary hypotheses. In this case, if there is a test of disconfirmation (i.e., the result is 5 q), this does not necessarily involve the rejection of p, as it may the auxiliary hypotheses that are at fault. Nonetheless, observing Popper's normative call will lead to emphasizing disconfirming strategies. The use of such strategies would force people to think of circumstances that would lead to a rejection of the hypothesis. By testing the hypothesis under this conditions, it is possible to conclude that the hypothesis is valid, as long as no disconfirming evidence is forthcoming. If the hypothesis is disconfirmed, then we can improve our prediction by testing whether the hypothesis itself or the auxiliary conditions are incorrect.

Popper's arguments became a motivating factor for carrying out investigations of a psychological nature (Chapman and Chapman 1969) which, not surprisingly, have shown that, like scientists, lay people rely on data that confirm their hypotheses. In general, these studies have shown that people have a tendency to use confirmatory strategies (sometimes called *confirmation bias*), by focusing their attention to the part of the evidence that supports their prior beliefs, even in the presence of evidence contrary to their hypothesis. Confirmation bias in both the general population and specific populations, such as university students and scientists, has been shown in many research studies on the matter. In the hypothesis testing process, one frequently find data that contradicts one's hypothesis. In these circumstances, people respond in various ways. Many of these are attempts to keep the hypothesis alive. Next, we will present a description of the various ways that people act in response to anomalous data. We illustrate these with examples taken from the history of science.

4.1. Actions in Response to Anomalous Data

The use of strategies such as confirmation or disconfirmation is more important when one is dealing with data that is inconsistent with one's hypothesis. In this case, people respond in a variety of ways. These responses can be viewed as actions to maintain a perception aligned with the intention (e.g., keeping one's hypothesis alive). Studies of a historical and psychological nature (Chinn and Brewer 1993) have shown that when dealing with anomalous data:

Ignoring of data: There are instances, both in the history of science and in experimental studies of reasoning that show that people sometimes ignore data that contradict their hypothesis. Ignoring data involves not paying attention to the data or considering unimportant. Among the more common are claims that would contradict fundamental tenets of a discipline (e.g., claims about a perpetual motion machine or about levitation and other ESP phenomena). There are, however, instances when ignoring data has resulted in negative effects for the discipline. An example of ignoring data in the history of science is provided by (Chinn and Brewer 1993). They recount the story of the extinction of large animals during the Cretaceous Tertiary period, proposed by Luis Alvarez and colleagues (Alvarez et al. 1980). They based their hypothesis on the larger than expected amounts of *iridium*, in the geological sediments of that period, which they could only account for as a result of the impact of a meteor or a comet on Earth. Months after the publication of their paper, still their fundamental finding was being ignored.

Rejection of data: Rejection of data occurs when the data is explained away as a result of factors that are peripheral to the real value of the data (e.g., the result of random variation; the incompetence of the person claiming to have produced the data; the malfunctioning of an instrument). Some reasons for rejecting data are methodological errors, the result of chance variations, ad-hominem arguments, fraud, and showing the implausibility of consequences if the data were accepted as evidence.

Methodological error: An example is given by the unexpected discovery, in 1965, of cosmic microwave background radiation (Penzias and Wilson 1966), which gave the first important empirical support to the Big Bang theory of the creation of the universe. This example is interesting because it illustrates a case of pure rejection of data with no expectations or hypothesis to account for the results. The discovery, made by Arno Penzias and Robert Wilson at Bell Laboratories, happened when they attempted to measure the intensity of radio waves from several parts of our galaxy. To do this, they first selected a small area of the radiation spectrum, which was assumed by scientists to be free of radiomagnetic activity. What they got, however, was a consistent amount of radiation coming from all directions of space. Penzias and Wilson disbelieving their results, rejected them as a side effect of some form of malfunctioning of their instrument (a gigantic radiowave antenna placed on top of a hill). First, they attributed the results to waste material from a couple of pigeons who had made their home of the antenna; they disassembled the antenna, clean it up, and put it back together. The same results occurred as before. They then attributed them to a defect in the metal structure of the antenna, which they fixed soon after by sealing the apparatus, with similar results after testing. Still not believing their results, they did extensive testing with the antenna, resulting in the same level of radiation anywhere to which they pointed the antenna. Only after all possible sources of error were disconfirmed, Penzias ad Wilson, were willing to accept the reality of their data, although had no explanation for the data[3].

Chance variations: Another illustration of data rejection is provided by the findings presented in 1915 by Alfred Wegener (1924) in support of the theory of the continental drift (now known as *plate tectonics*). The theory states that all continents were once a single landmass, which was later separated by the breaking up and moving away of the various pieces of land that now form the continents. Rejection of Wegener's theory and data delayed the development of the entire field of geology for several decades. The theory was supported by a number of pieces of data (e.g., the geological fit between the shorelines of South America and Africa; the finding of similar species of plants and animals across the Oceans; the similarities of geological formations between separated areas of the globe, such as the Appalachian mountains and the Scottish highlands), which were rejected on the basis of being the result of chance (Chinn and Brewer 1993).

Ad-hominem arguments: Other reasons for ignoring Wegener's data were that he did not have a university faculty appointment at the time of the publication of his book, was not a geologist by training, and furthermore was almost completely unknown in the geology community. Being unknown in the scientific community, was sufficient for mainstream scientists to reject his contribution. It was not until the late 1960s that Wegener's work was appreciated as revolutionary.

Unreasonableness of conclusions: Critics of Wegener also pointed out that if the evidence were accepted, the continents would not look the way they do now, since the force of tidal waves produced by the displacement of the oceanic crust would have to be enormous and have devastating consequences for the Earth (e.g., the Earth would stop its rotating movement). Wegener did not have a mechanism explaining how the drifting of the continents could have taken place, which left a window open for his critics to reject the theory out of hand. A mechanism was not proposed until the 1960s when new exploration techniques allowed the gathering of evidence from the bottom of the oceans.

Fraud: Although real fraud is observed in science more often than we hoped for, the rejection of data on the reasons of being fraudulently produced is sometimes shown when honestly gathered and valid data rejects strongly held beliefs. Fraud can be claimed when the data is judged as being "too good to be true," when they are dismissed as a hoax or a joke, or when they contradict beliefs so strongly believed, that fraud is the only explanation available.

Reinterpretation of data: It is frequently the case that scientists acknowledge the evidence and its relevance, but try to explain it in terms of their preferred theory, or some other explanation that does not conflict with their beliefs. For instance, some of the evidence that Wegener used to support his continental drift hypothesis were known to the scientific community. However, these data were given a different interpretation. The data on similarities among animal and plant species in physically separate regions (e.g., Africa and Brazil; Appalachian mountains and Scottish highlands) were explained in terms of a landbridge that was hypothesized to have existed between the continents. Thus, although some of the data presented by Wegener were known by the scientific community and accepted by most scientists, they were reinterpreted in terms of a different hypothesis.

Holding of data in abeyance: This involves acknowledging the data but holding any decision as to their validity until new relevant information is produced. Examples of this include the data concerning the orbit of Mercury (Chinn and Brewer 1993). Although the data regarding its orbit was inconsistent with Newtonian theory, the latter was not changed nor rejected to fit this observation; rather, the data was put on hold in the expectation that eventually the theory would account for the apparently anomalous evidence.

***Excluding data*:** This involves the exclusion of data from the scope of one's hypothesis, on the basis of its non relevance to the hypothesis.

By describing the types of responses to anomalous data, it is possible to develop some scheme for coding the various responses that people (e.g., scientists, physicians, medical students) produce to deal with evidence that is new or challenges their understanding of a particular hypothesis (Arocha and Patel 1995), which may suggests to possible errors in the evaluation or assessment of evidence-based medicine, not only by its users, but also by its creators. If such cases are not uncommon even in the more developed sciences, the occurrence in biomedical disciplines cannot be rejected or ignored. One reason is that they seem to point to a human tendency to act in particular ways to what is not in keeping with our own knowledge and experience. Studies in psychology and education have also shown the same or similar types of responses are displayed by people, ranging from elementary students to domain experts. Knowledge and beliefs thus determine the way people interpret and use evidence. In medicine, the structure of medical knowledge has an important role to play in this process.

4.2. Evidence and the Structure of Medical Knowledge

The extent to which evidence from epidemiological studies is directly relevant to interventions with individual patients is currently a hot topic of debate. However, as we have seen subjective factors in evidence interpretation make the evaluation of the evidence a far from straightforward matter. Studies conducted by our research team (Arocha and Patel 1995) which have focused on how people reason from prior information have shown that sometimes people fail to consider evidence in their decision making processes. They show a number of biases in their reasoning, including ignoring important information as evidence, explaining away the evidence to maintain their preferred hypothesis, or making selective use of evidence to support their hypothesis. This research is extremely relevant to the investigation of the EBM (e.g., in the development and use of clinical guidelines or the design and use of health decision making support systems). In this context, appropriate actions in the real world are based on accurate assessment of the situation, which includes selection of relevant cues, evaluation of expectancies, anticipation, and rapid assessment of the degree of urgency. Adequate knowledge of the domain is a necessary condition for developing the accurate assessment of situations, but having adequate knowledge is not sufficient. Possessing knowledge of typical problems and their solutions are also necessary, because they enable people to detect when expectations are violated. However, natural practice is usually very "messy" and typical problems are seldom observed. In these circumstances, practical experience with an assortment of different problems may be a necessary, although not sufficient, condition for successful performance, as they expose the expert to a variety of scenarios, which in turn serve to promote flexibility in the expert's problem solving and decision making. In more unfamiliar or complex cases, experts may differ among themselves about the interpretation of the case evidence, which has implications for diagnosis and treatment.

Given the complexity of the domain of medicine, there is a need for a framework for differentiating between different types of knowledge in medical contexts. Evans and Gadd (1989) proposed an epistemological framework that differentiates several levels at which clinical knowledge is organized in a medical problem-solving context. In prior research, we have made use of some of these levels to characterize clinical reasoning. To understand the nature and use of evidence, however, we will make use of two levels for which we found little use before. The first level identified by Evans and Gadd is the level of **empirium**. This level described the raw data of our perceptual experience, which can support numerous interpretations. We hypothesize that the most basic level of evidence resides at the level of empirium. This level includes fundamental biomedical facts, such as anatomic structures, as well as specific patient information, such as the skin coloration seen in a patient or the presence of absence of hair (Evans and Gadd 1989). The next level is that of **observations**, which are units of information that are recognized as potentially relevant in the problem-solving context, but that do not constitute clinically useful facts. Higher up the knowledge structure are **findings,** which are comprised of sets of observations that have clinical significance in the medical knowledge base. It is recognized that there are about 7,000 different findings. Establishing a finding reflects a decision made by a physician that an array of data contains a significant number of patient cues that need to be taken into account. Further up in the hierarchy, we find **facets**, clusters of findings that are suggestive of a particular type of problem, reflecting an underlying pathology (e.g., aortic insufficiency, hypometabolic state).

The following level is formed by **diagnosis**, the level of classification that subsumes and explains all levels beneath it. The final level is **global complexes**, which are contextual information or circumstances that affect the patient's behavior and possible treatment (e.g., age, sensitivity to allergenic substance, degree of autonomy).

Given that evidence can be construed in terms of the role of a statement in the argument structure, any level of the knowledge hierarchy can be used as evidence for some hypothesis. A characterization of evidence then can benefit from the level at which it is presented. Studies (Patel, Evans, & Kaufman 1989; Patel, Kaufman, & Magder 1996) have shown that to justify the diagnosis to a clinical problem, students, internists and expert physicians use different levels of the hierarchy. Students tend to rely on observations or findings (depending on their level of training); residents rely on findings and facets; whereas expert physicians rely on facets, and sometimes findings. This suggests that and provide evidence at different they reason at various levels of abstraction.

4.3. Towards a Definition of Evidence

We can summarize these approaches to the conceptualization of evidence presented above in terns of a series of requirements for calling a piece of data, evidence. Evidence is a piece of information that a) is a perceptual/conceptual representation inside a person; b) of a variable property of the environment; c) that is publicly available or acknowledged; and d) that is used in a chain of reasoning to support a conclusion (e.g., a hypothesis). The first requirement states that to be considered evidence, information has to be represented in someone's mind. From a psychological perspective, evidence is not outside the individual, but is present as a mental representation of something. The second requirement states that the evidence is always of a variable aspect of the perceived world; that is, it consists of properties or characteristics of some object or set of objects (e.g., position, size, color, relationship to another object, a particular sequence of states). The third requirement says that evidence has to be public; that is, it has to be directly perceived (e.g., results of a test) or agreed upon by other people (e.g., a patient's symptom). The last requirement is that this representation, in the form of a proposition, is used to support some conclusion (e.g., a hypothesis about a clinical case, a diagnosis) in a chain of reasoning.

5. Methodological Tools for the Study of Evidence Interpretation and Use

5.1. Experimental Procedure n Evidence Utilization

The model for the evaluation of evidence presented above will serve to described a procedure for the experimental investigation of intentional behavior, such as the generation or the evaluation of evidence. The procedure is based on the test for the controlled variable, TCV (Arocha and Patel 1996; Marken 1997; Powers 1973; Runkel 1990).

The experimental investigation of intentional behavior is based on the hypothesis that to understand what people are doing it is necessary to investigate the perceptions that people are controlling. This can only be done with a methodology that puts special emphasis on the identification of a person's intentions, perceptions, and actions and the environmental constraints that influence his or her behavior. This necessitates a different experimental approach than is commonly used in the behavioral sciences. Although a mapping between the traditional approach and the control perspective cannot be done (as the latter involves factors not considered in the former), in the control perspective, inputs are not independent variables, they are the perceptions that are under control (aligned with the person's intentions), as long as the subject is controlling the experimenter's intended perception. In other words, as long as the subject willfully and correctly understands the experimental task. This simply acknowledges that, notwithstanding the experimenter's intentions in designing the task and instructions given to the subjects, the results of the experiment rest on the subject's appropriate interpretation of the task instructions and on his or her willingness to do the task. It is important to understand what the intentions of the subject are and what perceptions are brought into alignment with his or her intentions. Traditional methodologies have no way of demonstrating this.

In its basic form, the different types of the traditional research design reduce to a basic common denominator: variations in output (i.e., the subject's responses) are traced back to variations in inputs (i.e., experimenter manipulated factors). The study is successful to the extent that there is statistically significant covariation between inputs and outputs. Methods of analysis used in this approach (e.g., ANOVA, GLM) are also based on the same input-output model, where variations in output are accounted for in terms of variations in input and where the aim is to find covariations (e.g., statistically significant correlations) between the experimenter-manipulated variables and the subject's responses. By emphasizing outputs (and their concomitant variation with inputs), traditional experimental approach has to handle a great deal of noise in the data, hence the use of statistics to discern real effects from random fluctuations. Also, as observed in the model, the path connecting inputs to outputs pass through the environment, not from the person. The implication of this is that while traditional methodology reveals something about the environmental constraints on behavior, they reveal little about the person's intentions or the perceptions they are trying to control (although it may suggest some possible controlled perceptions). And, as we saw, understanding people's intentions are basic to understanding what people are doing.

There is a reason for developing a new experimental methodology. According to the basic control model presented in Figure 1, a person's outputs have to vary in any way that is necessary to counteract the net effect of the often unpredictable disturbances on the controlled perceptions, if the goal is to be realized. The outputs themselves are therefore of less importance to the researcher than the perceptions the subjects are controlling. Furthermore, in contrast to the traditional approach, where the emphasis is on the concomitant variations between inputs and outputs, a control perspective looks for *lack of variation*. The reason for this derives from the theory, which states that the realization of intentions involves the alignment of the input (the perception) with the intention, by generating outputs that counteract the changes—produced by the disturbances—to the input.

Although the basic experimental control design is a bit more complex than the basic traditional design, its basic structure can be described in a few steps. The basic methodology consists of the following steps:

(a) The researcher hypothesizes about a perception that the subject is controlling. An example will serve as illustration: it is hypothesized that the subject, e.g., e physician, is concerned about the possibility of a thrombosis causing the purple coloration of the patient's ulcers. In this case, the subject is controlling for a *relationship* between the color of the ulcers and thrombosis. Controlling for that relationship may be driven by another controlled perception: saving the patient's leg (i.e., avoiding surgery). If the subject is indeed controlling for that relationship. then he or she would act to maintain it, e.g., by performing tests that confirms the presence of thrombosis.

(b) The researcher modifies some aspect of the evidence in such a way that it becomes inconsistent with the hypothesized controlled perception. The rationale for this is that any disturbance to the controlled perception will be resisted by the subject through actions that oppose the disturbance. Disturbances can be applied by, for instance, linking two unrelated pieces of data (e.g., by covariating them), removing part of the evidence or presenting some anomalous data that is inconsistent with thrombosis.

(c) The researcher measures the effects of the disturbance (e.g., data that s inconsistent with the hypothesis of thrombosis) by determining the subject's corrective action. The subject may not do anything (e.g., ignore data) or may do something instead. If the subject, for instance, is presented with a disturbance, e.g., a negative test for thrombosis (e.g., oppose the disturbance in some way, such as rejecting the data or questioning the credibility of the author of the data).

(d) If the subject does not act on the disturbance, he or she is not controlling that perception. In this case, the hypothesized controlled perception (as defined by the experimenter) varies or sometimes even covaries with the output. It may be possible, however, that the perception actually being controlled is very similar to the one hypothesized by the experimenter. Slight modification to the hypothesized controlled perception may result in the subject strongly opposing the disturbance.

(e) If the subject acts to remove the disturbance, look for what he or she might be doing to opposed the disturbance (e.g., rejecting data, reinterpreting data). The hypothesized perception is likely to be the one that is controlled by the subject.

In summary, a control perspective on the study of intentional action provides an experimental paradigm for the study of the interpretation and utilization of evidence. This paradigm is unique in that it attempts to find the specific variables that are being kept under control, rather than finding general responses averaged across subjects. We are currently studying the evidence interpretation as a function of the format presentation, based on the hypothesis that the way of that evidence is presented (e.g., in graphical or narrative forms), are aspects of the environment that are also under control, rather than only medically relevant factors (Wyatt 1999). It is also unique in the search for universal invariances (i.e., those that apply to all the member of the population of concern), where the emphasis is on determining what does not change after introducing disturbances to the hypothesized controlled variable.

The control model provides a series of categories that can be used to interpret observational data (e.g., as in naturalistic research). These categories are useful to identify possible controlled variables that can be tested later. In conjunction with more traditional, sample-based studies, it can provide important insights on the nature of evidence and its utilization. The next section will present a coding scheme for the analysis of evidence that is based on the model, including cognitive strategies and knowledge constraints.

5.2. Coding Scheme for the Analysis of Observational Data

As was argued in the previous sections, evidence can be looked at along various dimensions. First, we have the cognitive dimension, where evidence is identified in relationship with the intentions or goals of the person. In this context, evidence is a mental representation of some variable aspect of the world. A second dimension is the argumentative, where evidence is examined in terms of role that it plays in support of an argument to reach a conclusion. Third, we have the knowledge dimension, where evidence in health care is seen in terms of the type of knowledge represented in the evidence.

Table 1. Cognitive Dimensions of Variation for the Analysis of Evidence Interpretation and Use

Dimensions		
Goal-driven Dimension	**Strategy Dimension**	**Knowledge Dimension**
Goal	Confirming	Biomedical
Perception	Disconfirming	Epidemiological
Disturbances	Ignore data	Clinical
Actions	Reject	Traditional
(Memory)	Exclude	Observational
	Hold	
	Reinterpret	

5.2.1. Goal-driven Dimension

The perceptual cognitive dimension serves to categorize evidence in terms of their role that they play in the generation, assessment, and use of evidence from the point of view of the person interpreting or evaluating the evidence (i.e., the actor), rather than from a normative (e.g., argumentative dimension) or an epistemological (e.g., knowledge dimension) viewpoints. From this perspective evidence is a representation of some part of the environment that is either consistent or inconsistent with some theoretical or empirical statement.

There are two aspects to the cognitive dimension t hat we will consider here. First, evidence can be seen as a perceptual representation, often directly observable (e.(e.g., a pattern of X-rays on a slide). Second, as a cognitive representation generated from some linguistically mediated perception. The former tend to be perceptions of some low level character, such as the perception of some particular configuration (as in the X-rays example) whereas the latter tend to be perceptions at a higher level of processing, such as a program, e.g., logical or a temporal sequence of events. Still other perceptions are related in a subjective way with even higher levels of perception beyond the purely cognitive, such as the perceptions of principles (e.g., perception of principles of humane clinical practice) which may also influence evidence interpretation and use.

We have found that the control framework is useful in characterizing those aspects of evidence interpretation that are linked to action, whereas a models based on notions of long terms memory and comprehension processes seems more suitable to deal with domain-specific decision making.

5.2.2. The Psychological Nature of Evidence Interpretation

The interpretation of evidence is determined by the person's goals, his or her perceptions of the current situation, the situation itself as it relates to the person's perceptions and goals and the actions that the person can produce to generate the changes that he or she wants. All these in turn are influenced by the knowledge that the person possess. It is, therefore, important to examine all these aspects in the investigation of evidence interpretation and use. Some of these aspects can be directly observed, but others have to be inferred from observable data and the observer's knowledge of the cultural milieu.

Goals: Goal statements can be often gathered from the investigation of naturally occurring or experimentally generated protocols, such as think aloud protocols, explanations, and questionnaires, to the extent that people possess a clear representation of their goals and can or are willing to verbalize them. The more straightforward method consists of asking the subjects what their intentions are regarding a particular action. This can be accomplished by generating semi-structured interviews that allows the identification of the people's goal structures. The process begins by questioning , for instance, the key personnel in an organization on issues such as the major goals to be achieved or why such goals are important. The next step consists of determining what lower-level goals are needed to be achieved so that the major goals can be achieved, repeating this process until most of the goals, both higher and lower level ones, are covered. When an organization is involved, it is important to look at the printed material the organization has produced describing the objectives and main tasks of the organization. This provides a supra-individual goal structure which is then mapped against the individuals' structure of intentions.

Current perceptions/conceptions: Statements regarding a person's perceptions of an issue should be gathered also. These statements reflect mostly the high level perceptions rather than the lower level perceptions that are controlled to bring about the high level perceptions. Statements reflecting the person's assessment of a situation and that they explicitly relate to a goal or that, by inference, can be related to a goal fall within this category. For instance, in a process of guideline development, 'timeliness' can be a perception that is controlled by an organization, a low value of which can be brought into alignment with an intention (e.g., bringing the time of that takes to create a guideline down from 3 years to 1 year).

Disturbances to the attainment of goals: Recent research in cognitive science has stressed the inherent situated character of human behaviour. That is, people behave in specific environments, which impose constraints on how people think and behave. Although people may use their surroundings to their own advantage, the environment also gets in the way of reaching people's goals. In these challenging circumstances, environmental disturbances (e.g., lack of resources, opposition from other people, delays in the provision of a service by someone else, lack of reliable information) can make goal achievement difficult. These disturbances need to be overcome in order to achieve the desired goals. When collecting data on evidence interpretation, disturbing aspects of the environment have to be identified. These refer to those aspects that may influence the perceptions the people control and that therefore that have an effect on the state of those perceptions. Disturbances can be either observed or inferred from the verbal or observational data collected. It is important to note, however, that the actual nature of the disturbances need not be determined, only through their effect on impeding goal attainment. An excellent source of information about disturbances are people who has a lot of experience in a particular setting (e.g., laboratory researchers, expert clinicians). In organizations, it is important to question people at various levels of the organizational structure (e.g., senior and junior personnel) as they provide with potential disturbances at different degrees of granularity (e.g., lack of adequate budget vs. malfunctioning of a measurement apparatus).

Actions taken to counteract or overcome those obstacles: These actions are triggered by the perceived difference between the goal (what the person wants to achieve) and the perception of the current state of the world. It is possible from an analysis of verbal or behavioral data, to list the major obstacles that are expected to be encountered in the future, based on past problems. For each of the disturbances gathered, the analyzer should also list the actions that are taken to counteract it or overcome it. As pointed out before, an interesting characteristic of actions is that sometimes an observer cannot distinguish between actions that were intended from those that occur as a side effect of the intended action. Since actions are hierarchically organized, at some level of organization there are consequences that are unintended by the participant. In trying to overcome obstacles, some of the actions taken can generate other problems that need to be counteracted by still other actions. These can lead to complex action sequences (e.g., Patel, Kaufman, & Magder 1996). Knowing what actions may produce significant unintended consequences permits the planning of contingency strategies to avoid such consequences. It is possible to list the actions that people took to counteract past disturbances and prevent some of the disturbances to reoccur.

5.3. Strategies in Evaluation of Evidence

Confirmation and Disconfirmation: Confirmation is the generation or search for data that would support a one's hypothesis. If previous research is an indication of the use of this strategy, we could assume that this strategy is the most widely used by lay people, scientists, and professionals. Disconfirmation is the generation or search for information that would reject one's hypothesis. Although it has not been found to be very widely used, disconfirmation is sometimes used by expert scientists and physicians. An example in the medical field is presented by Patel, Groen, and Arocha (1990). In their study, physicians who were clinicians or researchers were given clinical problems to diagnose. Most expert physicians accounted for the clinical cases using a forward-directed pattern of reasoning (typical of expertise) using a confirmation strategy (Data $\rightarrow$ Hypothesis), such as "galactorrhoea, elevated TSH, and increased prolactin suggest myxoedema". However, one physician diagnosed the clinical case in a different way: This physician started by ruling out possible diagnoses (Data $\xrightarrow{Neg}$ Hypothesis) such as "elevated liver edge is not consistent with a primary hepatic process" and only diagnosed the problem after various hypotheses were ruled out. Such a strategy is particularly useful when solving difficult problems or making difficult decisions, as it allows the person to develop a more systematic approach to evaluating evidence. Thus, it is more likely to occur in situations of high uncertainty.

5.3.1. Operations on Anomalous Evidence

Ignoring data: This category is coded when there is no mention of the anomalous data in a protocol. The absence of comment regarding the evidence can be taken as an indication that the anomalous data is not considered in the subject's argument. When there are doubts as to the status of a piece of data, the researcher can obtain posthoc data to support the hypothesis that the evidence has been ignored.

Rejecting data: Any statement that explicitly rejects the anomalous data. Frequently, the rejection statement is supported by one or more reasons.

Methodological error: It is common in scientific evaluation to reject data based on methodological errors or inappropriateness of the methodology. Scientific criticism of research studies are very often based on some methodological or procedural problems.

Chance variations: Statements where a piece of data is dismissed and accounted for by random variations.

Ad-hominem arguments: Rejecting anomalous data based on the person or persons who reported the data.

Fraud: Disbelief in the data because of possible fraud committed while gathering, analyzing or reporting the evidence.

Unreasonableness of conclusions: Rejecting the interpretation of the evidence based a reasoning that, if followed, would produce unreasonable or absurd conclusions.

Ignoring or rejecting data are used in many instances of medical decision making involving expert physicians. Although the use of such strategies may serve the purpose of filtering non-relevant information, it may sometimes contribute to missing critical, albeit anomalous information. Of particular interest is the case ignoring or rejecting data because of the entrenchment of one's beliefs, as it may occur in experts who are so secure about their interpretation that they fail to notice anomalies. Cases have been reported of such occurrences in previous work (Patel, Arocha, and Kaufman 1994) where experts have been "blind" to the presence of information that does not fit a previously held schema for a problem.

Reinterpretation of data: Acknowledging the validity of the data, but reinterpreting it in terms of another hypothesis, not the one originally associated with the evidence.

Holding of data in abeyance: Decision not to believe or disbelieve the evidence and holding one's opinion after new data or a better explanation is found.

Excluding data: Acknowledging the data but keeping it outside of a theory's purview based on its perceived irrelevance to the theory.

5.3.2. Knowledge Dimension

In terms of the types of knowledge used as evidence we have found it necessary to include the various types of knowledge available to health care workers when solving problems. Evidence will be divided into the categories described below, based on the properties of the information given. It will be segregated according to the issues related to the nature of information used. Evidence is separated into the categories relating to its nature and the evaluation, the latter category being a meta-level analysis of the evidence (for the coding of evaluative statements).

In the presentation of evidence, there are seven categories:

Epidemiological: This type of evidence includes all evidence stated in terms of probabilities or the likelihood in a population, such as estimated survival rates or the average efficacy of a form of treatment that are provided independently of whether they are expressed in quantitative (e.g., there is 5% drop in mortality rate) or qualitative (e.g., the mortality rate drops a small but significant amount) terms. This is basically population-based evidence, such as that found in randomized clinical trial studies, independently of the scale of the study in terms of its sample size nor the quality of the research design. Two example statements of this type of information are presented below:

"...I just think the increase in mortality is so minimal as compared to the risks.....

"Just hold on a second, so you are going to have... 100 patients that you're looking at, so let's say, 25 have …uhhm, thrombotic infarct, lets say you take the best case scenario, you reduce the risk by 20%, so it's 20% of 25... so you get 4 patients roughly. Four to five patients who are going to benefit... with a small amount of benefit..."

Biomedical: This category includes all evidence relating a finding to its underlying physiology or pathophysiology. It also includes other sorts of data that link a finding to a specific outcome, for instance, the amount of time spent in a specific state if this temporal factor is expected to cause a patient to be in a specific state. In general, biomedical evidence is the sort of information that could be found in a textbook regarding the body in its different states or in literature on basic biomedical research (e.g., pathology). This also includes pharmacological information that relates to a drug's effect on physiology.

An example of biomedical evidence in the history of medicine can be seen in the discovery of helicobacter pylori by Marshall and Warren (1984). If can be demonstrated that bacterial causes ulcers, then it stands to reason that giving antibiotics to patients with ulcers will improve their condition. The mechanism of action does not need to be fully known for this type of evidence to be used. Another example is given in an excerpt from a protocol from a physician, who is reasoning about the functioning of a patient's heart:

"Yeah, it (repolarization) is everywhere, theyare wide all over, so it is a generalized wide complex, it is not a... normal QRS with an ST elevation in localized areas..."

Clinical: This is evidence that refers to basic categories of clinical knowledge, typically signs and symptoms and is most often gleaned from clinical encounters, or what might be called findings (Evans and Gadd 1989). In general, clinical evidence includes information gathered from experience in the medical profession, generalized from other cases, to form ideas about likely diagnoses or effective treatments. Clinical evidence is also found in textbooks and practice manuals. What differentiates it from biomedical textbook knowledge is that it does not involve underlying causal relationships (e.g., descriptions of pathophysiological processes). Most typically, clinical evidence involves either causal or conditional relations without the specification of mechanism (e.g., IF symptom, THEN do/give X). An example is given below:

"They don't present like that with an arrest, that's not what sudden death usually is."

Observational: This includes information gathered from interactions with patients. This category includes such general statement, such as a description of how a patient seems at a particular time, or the general impression of a patient without linking this information to any other sort of biomedical or clinical statements. In general it includes information that is not related to either clinical or biomedical information, but which nonetheless may be important for decision making in health care contexts. An example is give below:

"...neurally he is awake, alert, oriented times three, except for the day, he doesn't know it is Thursday but he knows it is November, and it is 1994."

Authority: This involves those statements that are used as evidence from authority. Although information citing authority can be hardly called evidence, we include it here given that on occasion, recourse to authority can serve to support one's argument protempore. Currently we have no evidence of its use in medical reasoning, but as an example we could think of a consultation between a non-expert physician and an expert, where the former accepts the recommendation of the latter, based on his or her reputation as a highly qualified physician. This sort of information will likely be found among the less experienced physicians or lay people, but this does not invalidate its existence and its presence cannot be ignored.

Tradition based: This refers to those statements referring to what one could call "traditional" beliefs whose factual validity or effectiveness has not been determined empirically (and may not be considered important) and would not be questioned. Examples of this sort of information would correspond to statements regarding the use of home remedies, ideological reasons for treatment or refusing treatment (e.g., religious beliefs).

Experiential/Association based: This type of evidence refers to data that results from the daily experience and that lack of a any relationship to the underlying pathophysiological knowledge or to shared clinical knowledge. These associations can be made by any person and are not limited to lay descriptions. It simply refers to the use of one's personal experience or other cases (be it friends, family, even a patient for whom the underlying physiology remained a mystery) where something was observed and that observation is now being used as evidence for claims made in a particular case. Sometimes, personal experience, through self-experimentation, is used to provide evidence for a hypothesis. There have been several known cases in the history of medicine (Altman 1986; Brown 1995). A recent one is the case of Barry Marshall who together with Warren discovered the helicobacter pylori, first tested on himself the bacterial cause of ulcers, by swallowing a mixture of a billion bacteria and water."

Other examples relate to general social or cultural experiences, such as the following excerpt from a physician discussing a patient with visible signs of suspicion and agitation "one of the things we noted yesterday was on his left arm he has got a tattoo. This is a guy who has gone through the camps and uh, waking up in this situation, there is usually an extensive paranoia which comes from a long history of self preservation."

5.4. Assessment and Evaluation of Evidence

Given that evidence is not transparent, as is shown by the frequent disagreements among experts (e.g., scientists, physicians) as to the role that a piece of evidence plays in an argument (e.g., a scientific explanation), it is often observed that people also disagree as to what constitutes *good* evidence. We can identify at least three categories that are often the focus of debate regarding the quality of evidence: validity (i.e., whether the evidence is valid in a given context), meaning (i.e., what the evidence "says"), and relevance (i.e., whether the evidence is applicable to a particular situation).

Validity: People question whether evidence is deemed to be correct or not. This involves both the idea of the correct citation of evidence as well as its validity in the assessment of a particular diagnosis or procedure. The validity of evidence is questioned, but this happens with some frequency where physicians disagree when assessing complex clinical cases (e.g., in clinical rounds).

Meaning: This refers to the way in which evidence should be interpreted in a specific case. Within this category, there are variations. For example, one may question how the evidence presented in support of a hypothesis should be viewed in relationship to the particular spectrum of symptoms of a given patient. Also, there maybe questions regarding meaning of statistical/epidemiological evidence, when applied to a particular case. The following exchange between three physicians illustrates this type of evaluation:

> *P1*: ...you have to look at 100 people and say what is the overall probability and what is the best case scenario in the group that actually gets treated.
>
> P2: That's Michael's concern and mine about using these global statistics for one patient.
>
> *P1*: You have to, you don't know what this guy has got, he is a global patient coming in.
>
> *P2:* If he comes in with ST segment elevation he has had an infarct, and you have given him streptokinase within 6 hours you decrease his mortality and improve left ventricular function. That is a very important number and you are lost in your numbers.
>
> *P3*: Yeah, but we're talking about this patient, a decrease in mortality is over a whole bunch of patients, in this patient, he is hemodynamically stable.

Relevance: This refers to how applicable evidence is to a certain circumstance. The evidence is not put into question, but only its relevance to the present case. This analysis of evidence brings into question the relative importance of pieces of evidence presented by or attributed to a patient. The following example illustrates this category:

> "so, what… I'm saying is… the cardiogram in this case doesn't help you, so the probability then of this is low, in fact, given the cases."

These two sorts of uses of evidence, presentation and analysis of, represent different levels of knowledge. The presentation of knowledge lies at a lower level than the analysis of it. The analysis involves the weighing of evidence in many different ways.

5.5. Examples of Coding for Evidence

The following examples look at two aspects of evidence coding, as described in the previous section: Coding on the nature of evidence and on the evaluation of evidence. The examples derive from the analysis of a clinical guideline to which actual protocols can be compared.

5.6. Clinical Practice Guidelines and the Nature of Evidentiary Knowledge

The use of clinical guidelines by physicians is supposed to provide a more scientifically sound way of assessing evidence in clinical practice. This is done by emphasizing results from epidemiological studies on a particular topic. However, as research has consistently found physicians do not make use of clinical guidelines at the point of care. One reason for the lack of utilization of guidelines may be the information provided in guidelines may not be Table 2 presents the number of occurrences of data that were used as evidence in terms of the knowledge dimension categories for data coding. The figures presented in the data show how can the coding be used to tabulate propositions from a clinical guideline. The expectation is that epidemiological evidence be used more frequently than other types of evidence, given the strong emphasis on EBM in current guidelines.

Table 2. Frequency of evidentiary statements related to the types of knowledge relied on by clinical practice guideline developers.

Type of Evidence	**Frequency**
Epidemiological	107
Biomedical	4
Clinical	4

As the table show, the vast majority of propositions refer to results of clinical trial and epidemiological studies. This is expected given the Biomedical evidence was used in only four occasions, An example of which is given next.

> "The newer, third-generation sensitive TSH assays may be more specific"
> Part 2, Page 8

Like biomedical evidence, clinical evidence was used only four times. In the example presented below, the guideline analyzed makes reference to physicians' experience in the clinics to adapt treatment according to their beliefs.

> "The efficacy of treating patients with subclinical hypothyroidism to prevent progression to overt hypothyroidism is not known. (Lack of Evidence) However, physicians *who believe* that significant morbidity accompanies progression to overt hypothyroidism *may wish to prescribe 1-thyroxine to all asymptomatic patients who have a TSH level of 10 mU/L or more*. Part 1, Page 4

The excerpts presented above show that the clinical guideline analyzed rests, almost exclusively, on epidemiological evidence. Little biomedical evidence was used, and when it was, it related to the testing used, not to the mechanism supporting the recommendation of the guideline. As clinical evidence is used largely when other types of evidence, considered stronger, are not available, it seems that the type of evidence most valued by physicians in developing guidelines is epidemiological. Little emphasis is placed on the biomedical processes underlying the pathophysiology of the disease covered by the practice guideline.

5.7. Evaluation of Evidence in Clinical Practice Guidelines

The review of the literature supporting the guideline generally screens out what is considered bad studies, in terms of poor experimental designs or other methodological factors. In this regard, a research evaluation process has already been done before the guideline is written. However, practice guidelines, often include evaluative comments to studies (in a very summarized form), which although not perfect, according to guideline developers, provide some useful information to the practitioner. The number of evaluative statements regarding treatments or procedures, presented in the analyzed guideline, are given in Table 3.

Table 3. Evaluative statements of evidence presented in the clinical guideline

Lack of Epidemiological Evidence	41
Relevance	31
Validity	29
Meaning	2

Lack of epidemiological evidence was raised as the most important problem in the evidence evaluated for the guideline, supporting the greater emphasis on epidemiological evidence given in the guideline creation process. An example of such evaluative statement is given below:

> "No randomized trials of early treatment for subclinical hyperthyroidism to prevent these complications have been done ..." Part 2, Page 8

Validity is another aspect highly asserted as evaluation of evidence presented. This category is used to classify statements that question the evidence in terms of its scientific quality. The next excerpt gives an illustration:

> "Other researchers have questioned these assumptions, citing flaws in the literature on which they are based." (Part 2, page 10).

Relevance questions the applicability of techniques, procedures, or treatments to particular guideline. An example is presented next:

> "Most studies focus on older women with markedly elevated TSH levels – the highest risk group – but most persons in whom subclinical hypothyroidism would be diagnosed in a screening program are younger, have mildly elevated TSH level, and have a lower risk." Part 2, Page 11

Meaning was the least raised concerned about the meaning of a evidentiary proposition with respect to a particular situation. An example is given below:

> "In a 60-year-old woman with a pretreatment total cholesterol level of 6.5 mmol/L and no other risk factor, a reduction of 0.6 mmol/L(8%) would decrease the risk for developing ischemic heart disease from 10% to 9%." Part 2, Page 19

In summary, the largest proportion of comments about the nature of the evidence referred to its absence, i.e. a need for further research in order to be able to make good evidence-based recommendations. There are also statements related to specific populations or situations to which the guidelines might apply (Relevance – 31). Validity was questioned 29 times. Meaning, i.e. relevance for a particular case, was only used twice and both times as illustration. Compared with the number of pieces of evidence (115), there are 101 questions (not including *Meaning*). This does not imply that each piece of evidence is questioned, but it might be taken as an indication of the need for additional evidence in order to be able to propose well-substantiated guidelines.

5.8. Examples of Interpretation and Use of Evidence in Clinical Contexts

5.8.1. Use of Evidence in Intensive Care Unit

This section presents some examples of how evidence is used n an intensive care unit. In contrast with the use of evidence in guidelines, the utilization of evidence in the ICU relies heavily on the more pragmatic aspects dealing with the patient state. The nature of the tasks, its constraints, determine in a large manner the types of evidence that are used. Since the goal of the ICU is typically to stabilize the patient, most of the evidence is used is used to this end. Table 4 presents the number of evidentiary statements found in verbal protocols gathered from physicians and residents in their daily rounds.

The evidence was divided into on-line and explanatory. The former refers to evidence generated during clinical problem solving, whereas the latter refers to evidence produced after the case was solved (i.e., during an explanation of the case).

The type of evidence most frequently used was clinical evidence, followed closely by biomedical evidence, whereas epidemiological evidence was not discussed at all.

Table 4. Evidence used in the Intensive Care Unit—A quantitative description of the number of times different types of evidence are used.

Type of Evidence	On-line Evidence	Explanatory Evidence
Clinical	33	4
Biomedical	27	4
Epidemiological	0	0
Authority-based	4	0
Observational	5	6

In the next section we present some examples that illustrate the coding of the knowledge dimension, regarding the types of evidence used in the ICU.

5.8.2. Use of evidence in Emergency Settings

Next, we present examples of evidence used to make decisions in an emergency setting. As before, evidence was divided depending on when was the evidence discussed. On-line evidence was generated during clinical problem solving whereas explanatory evidence was generated after the case was analyzed, as post-facto explanation of what is going on.

Again, the results show that clinical and biomedical evidence were used most and epidemiological evidence was not use much. A difference between ICU and ER is in the amount of observational evidence, which was used often. The difference may be accounted for by the time-factor in ER. As reliable data may be difficult to obtain timely, the ER physicians may resort to more data types to make decisions, rather than evaluate the data in a systematic fashion.

Table 5. Number of times that types of evidence were used during coded sections of transcribed audio tapes of emergency room physicians' decision making.

Type of Evidence	On-line Evidence	Explanatory Evidence
Clinical	42	2
Biomedical	38	1
Observational	18	1
Epidemiological	1	0
Authority-based	0	0

Following are several examples of some of the types of evidence that are used in the emergency room (ER). The examples serve to illustrate the nature of evidence used during clinical decision making in emergency settings.

Biomedical

This example is about a 91 year old female patient brought to the emergency room complaining of abdominal pain and anxiety:

> "…to make a long story short, she's got multiple air fluid levels, but no gross dilation of the series. She has previous abdominal surgery,

> cholecystectomy (?) and I forget what the other one is. The labs are fine. Her urine shows trace red cells.... So, she's NPO, she's dehydrated, no entry is down for the time being, there's no grossly dilated loose (?) bowel. There's air in the rectum."

The excerpt shows how the physician presents the case o by reference to underlying physiological issues involved.

Clinical

This example presents a discussion of the case of a 57 year old patient, who complains of palpitations during the last few hours with increase in frequency. The patient describes the palpitations as pauses of the heart.

> "…he was also complaining about this atypical left-sided chest pinching, not necessarily related to the pauses. Ah, he has no, the past cardiac history he has, he had pericarditis a number of years ago, there's a significant family history of coronary artery disease. He's a nonsmoker, no other medical problems. The only thing I see, his ECG is totally normal, his labs are normal…

Epidemiological

The following abstract presents a discussion of an ER physician who discusses the possibility of a patient suffering from pulmonary embolism:

> "…so no risk factors for? So just taking, understand what you don't want to miss, I mean, since you have a positive alternate diagnosis, your pretest probabilities are probably around those as well. Way down, you still have to consider it. So, does he have any risk factors for pulmonary embolism?"

> "…what's significant is that's it's a false negative?, so the positive predictive and negative predictive values are probably only about 60%... and people question whether you really need to do them. It's sort of something in the textbooks, on the other hand, vital signs themselves are also very poor in terms of deciding."

Authority-based

This example presents a resident physician who treats a patient based on information given to him by the specialist without supporting the decision with other argument.

> "…Over the last week…the cardiologists have told me to give him two units of blood. He's got brown stool that's OB positive, I'm going to transfuse him two units of blood, his I&R hasn't been seen yet, but it's always 2-3 and a half. I guess there's been a decision that we don't stop the cumadin…"

The second example presents a resident physician discussing a patient who complaints about low back pain.

> "… she's really complaining about is low back pain which seems like it's been an ongoing problem, but she says it's really bothering her. She's had already severaldoctors check it out, and everyone says it's mechanical type back pain which looks like what it is, but she's really complaining about it, and physical exam, ... vitals are, her vitals are fine, I didn't get a chance to write them down, sorry, she's not febrile, she... Head and neck exam, her throat is definitely erthemitus."

In this case, the resident is using the information provided by other physicians of her condition, which the resident acknowledges as probably correct.

Observational

The category of observational is exemplified by the following dialogue between a physician and a nurse. In the case, the nurse uses her perception of a patient who comes to the ER limping as information to the physician to do an X-ray of the patient's ankle.

> Nurse: ... a guy in triage is limping ... it's (his ankle) all swollen...
> Physician: He can't walk on it? He can't walk four steps?
> Nurse: Like that (demonstrating a limp)
> Physician: Ya, so x-ray him.

In this case, the information used forward reasoning decisionmaking originates in currently or previously observed situation. This observation is the used, typically with other information, to make decisions regarding patient care.

6. The Nature of Evidence in Clinical Guidelines: Analysis of two CPGs

In this study, we analyzed the types of evidence presented in two background papers supporting two clinical guidelines. The first guideline analyzed was the analysis of the paper by

Williams, J.W., Mulrow, C.D., Chiquette, E., Aguilar, C., Cornell, J. (2000). The second paper analyzed for evidence was Helfand (1998) on the screening for thyroid disease.

6.1. Method

6.1.1. Material

We analyzed the document "Newer pharmacotherapies for depression in adults and adolescents," a background paper developed by for American College of Physician/American Society of Internal Medicine (ACP/ASIM), as support for the clinical practice guideline of the same name. Depressive disorders are common, debilitating and recurring illness that cause enormous personal suffering to individuals and their families. Individuals suffering from such disease and who receive early treatment can benefit. Newer antidepressants are clearly effective in treating depressive disorders in a variety of settings, at least as effectively as older antidepressants. Given theirsimilar effectiveness, the selection of antidepressant agents should based on criteria such as anticipated tolerance and type of adverse side effects. Clinicians and patients have a considerable number of options for treatment because the difference between older and newer antidepressant drugs lies on types of side effects. Choosing one drug over another, therefore, depends largely on the patient characteristics and preferences.

6.2. Method of characterization of evidence

Evidence plays an important role in different medical decisions such as, in screening, diagnosis and treatment of the patient. Evidence comes from different sources; it is crucial for scientific explanation aimed at solving a clinical problem. Evidence is a representation of something that illuminates decision-making to support a conclusion, such as a diagnosis, a treatment, or a management plan. We have identified different categories of evidence in the medical field. A few examples of them are given here.

Epidemiological: It includes all types of evidence stated in terms of probabilities or the expected likelihood of occurrence in a population, such as estimated survival rates or the average efficacy of a form of treatment.

- The life time risk for major depressive disorder ranges from 10-25% for women and from 5-12% for men.
- The point prevalence rate of depression is 5-9% for women and 2-3% for men.
- Personal costs of depression include impairment in multiple areas of functioning.
- The estimated costs for treating depression and the costs due to lost productivity exceed 44 billion dollars in 1993.
- The world health organization estimates that major depression is the 4th most important

cause worldwide of loss in disability adjusted life years.

Clinical: It refers to basic categories of clinical knowledge, typically signs and symptoms and is most often gleaned from clinical encounters, or what might be called findings. Below, are some examples:

- Major depression is characterized by at least two weeks of depressed mood or loss of interest or pleasure in nearly all activities.
- The symptoms of depression should cause clinically significant distress or impairment in social, occupational, or other areas of function.
- The symptoms of dysthymia are appetite disturbance, insomnia/hypersomnia, decreased energy/fatigue, low self-esteem, decreased concentration or difficulty making decisions, or feelings of hopelessness.

Biomedical: It is related to a finding to its underlying physiology or pathophysiology.

Observational: It includes information gathered from interaction with patient. This category includes such general statement, as a description of how a patient seems at a particular time, or the general impression of a patient without linking this information to any other sort of biomedical or clinical statements.

Authority: This involves those statements that are used as evidence from authority.

Experiential /Association based: It refers to data that result from the daily experience and is not directly related to the underlying pathophysiological knowledge or shared clinical knowledge.

- Some patients are reluctant to take traditional antidepressant medications.
- Some patients are reluctant to engage in Psychotherapy.

6.3. Analysis

The documents were analyzed for the types of evidence used to support clinical recommendations. For instance statements that presented epidemiological evidence were identified and tabulated.

6.4. Results

Some preliminary results are presented below. These results show the frequency of evidentiary statements, presented in the background paper, which refer to the types of knowledge described earlier.

Table 6: Frequency of Evidence in the Depression Background Paper

Type of evidence	Frequency	Percentage
1. Epidemiological	205	93
2. Clinical	12	6
3. Biomedical	0	0
4. Observational	0	0
5. Experiential	2	1
6. Authority based	0	0
Total:	219	100

This table represents the number of occurrences of data that are used as evidence in terms of the knowledge dimension categories for data coding. The figures presented in the data show epidemiological evidence is used overwhelmingly more frequent than other types of evidence.

6.5. Analysis of Screening for thyroid disease, *Annals of Internal Medicine*, developed by for American College of Physician/American Society of Internal Medicine (ACP/ASIM).

6.5.1. Method of characterization of evidence

The method of characterizing the evidence was the same as that used above. The document was analyzed for the types of evidence used to support clinical recommendations. The results showed that all the supporting evidence was either epidemiological or biomedical. No other categories were identified. The following table represents the frequency that examples of each category were found. The data show that epidemiological evidence was used more frequently than other types of evidence.

Table 7 Frequency of Evidence in the Thyroid Background Paper.

Type of evidence	Frequency	Percentage
1. Epidemiological	150	90.90
2. Clinical	15	9.09
3. Biomedical	0	
4. Observational	0	
5. Experiential	0	
6.Authority based	0	

A few examples of different types of evidence found in the paper are given below:

Epidemiological

- The prevalence of overt thyroid dysfunction depends on the age and sex of the population.
- The prevalence of overt thyroid dysfunction was lower in younger women and men.
- Risk for osteoporosis in subclinical hyperthyroidism is scant.
- Two cross-sectional studies evaluated bone density in patients with subclinical hyperthyroidism associated with multinodular goiter.
- Women with low TSH level had an increased risk for hip or vertebral fracture over 4 years.

Clinical:

- Hyperthyroidism can cause weight loss, tremor, heat intolerance, and muscle weakness.
- Hypothyroidism can cause muscle cramps, dry skin, intolerance to cold, constipation, poor energy levels, fatigue, and mental slowness.
- Overt hyperthyroidism is associated with accelerated bone loss.
- Adverse effects of replacement doses of l-thyroxine include nervousness, palpitations, atrial fibrillation, and exacerbation of angina pectoris.
- Overt hypothyroidism is seldom diagnosed in persons with mildly elevated TSH level.

The potential complications of subclinical hypothyroidism are symptoms, hyperlipidemia, and progression to overt hypothyroidism.

Biomedical: None found. It is related to a finding to its underlying physiology or pathophysiology.

Observational: None found. It includes information gathered from interactions with patient. This category includes such general statement, such as a description of how a patient seems at a particular time, or the general impression of a patient without linking this information to any other sort of biomedical or clinical statements.

Authority based: None found. This involves those statements that are used as evidence from authority.

Experiential/Association: None found. It refers to data that result from the daily experience and that lack of any relationship to the underlying pathophysiological knowledge or to shared clinical knowledge.

7. Conclusions and Recommendations

This document presents a framework for the analysis ofevidence interpretation and utilization. The framework is based on a cognitive model of decision making in medicine, previously used to investigate decision processes in a variety of tasks. In t he model, we described the decision maker (e.g., goals, actions, perceptions), the strategies deployed in making decisions (confirmation, disconfirmation, data and hypotheses operations), and the structure of the knowledge-base on which decisions are made (observations, findings, facets, diagnoses, global complexes). From the framework, we derive two types of methodologies: (a) an experimental method for the investigation of decision making and evidence utilization "in vitro"; and (b) a coding scheme for the analysis of verbal data in the "in vivo" investigation of evidence interpretation and evaluation.

The experimental methodology presents an iterative design for conducting single-subject studies in laboratory conditions. The general questions that this methodology are designed to ask are What goals are people trying to achieve and by what means? The method involves the manipulation of an environmental variable (e.g., difficulty of a problem; relation between two procedures; chronological sequence of events) to examine how the person acts to counteract the manipulation. This is based on the hypothesis that if the person is to reach or maintain a goal, he or she will act to prevent environmental factors from disturbing the path to the goal. This methodology is useful, for instance, to understand what goals a user of clinical guidelines is attempting to maintain. For instance, a physician may have the goal of avoiding complications in a patient whereas another may be more concerned with treating the disease in the most efficient way. Achieving these goals may involve "doing different things" and therefore following different procedures. By assuming constant goals and averaging across people, standard methodologies cannot distinguish between these two situations. The application of a control methodology allows a more detail investigation of factors affecting clinical guidelines use. To this end, it helps identify (a) the goals of the guideline user; (b) the aspects of the perceivable world that the user relying on (i.e., keeping constant) to maintain his or her goal in the face of disrupting influences; (c) the actions, heuristics, and strategies that are deployed in pursuing the goal; and (d) the factors that may distract from achieving or maintaining the goal (i.e., disruptive influences). Although simple in principle, this framework can be applied to a variety of tasks and levels of abstraction by adding necessary factors (e.g., memory, expectations, complex models) as these are needed to understand a particular phenomenon.

The coding scheme was designed to categorize the cognitive aspects involved in evidence interpretation and use. It encompasses factors along three dimensions: Knowledge, strategies, and the logic of practical discourse. These dimensions cover most of the cognitive psychological variables likely to be involved in evidence interpretation.

Illustrative analyses of the framework were provided. The analyses show the various uses of evidence in terms of the nature of the knowledge involved in evidence interpretation and use. We analyzed two cases: (a) supporting information for the use of guideline procedures; (b) use of evidence in naturalistic situations; such as the ICU and emergency settings. As expected, the type of evidence used most in the guidelines was epidemiological. Very little clinically or biomedically-derived information was used. In the ICU and emergency settings, the type of evidence used most was clinical with a strong use of biomedically-based information. Further analyses, looking at the types of knowledge used in such settings are currently being carried out.

What lessons can we take from the framework? The framework provides a number of issues that need to be taken into account in the study of evidence interpretation and use. We will point out some of them:

7.1 Knowledge and the Nature of Evidence

The framework emphasizes the need to identify the nature of the evidence used. In the medical domain this takes the form of the nature of the knowledge used in medical tasks. A first classification is between biomedical knowledge and clinical knowledge. Extensive research has consistently found that medical practitioners use both types of knowledge differently. Clinical knowledge is used mostly in routine clinical practice while biomedical knowledge is used as explanations, especially in the presence of uncertainty.

Another distinction that needs to be present is the layered structure of medical knowledge. For this purpose, the epistemological model proposed in Evans and Gadd (1989) provides a way to analyze the level of concepts that are most likely to be generated by medical practitioners at different levels of expertise. Research has shown that experts (specialists with may years of experience) process clinical case information at higher levels of abstraction than more novice subjects (e.g., general practitioners). This has implications for the design of evidence-based support systems, such as clinical guidelines embedded within patient record systems. The information presented as "reminders" to physicians should be different for more experts than for less expert practitioners.

A third aspect regarding knowledge corresponds to the nature of evidence itself. As it has been shown in the examples presented above, data of different nature can be used as evidence to support decisions. Some data may be thought of as having questionable value, if analyzed from a purely rational viewpoint (such as making decisions based on the opinion of an authority figure in the field), but they are nonetheless used be cause they provide support for decisions when there is not other evidence available or there is ample disagreement as to what constitute the best course of action.

7.2 Goals and Evidence Interpretation

This brings us to the role that intentions have in the making of decisions, specially those involved in such complex situations as medical practice. Underneath the decisions, there is also the purpose for the decision. This is important because purposes (i.e., what the person wants to achieve) can determine what would be identified as evidence. This is done by directing attention to those aspects of the evidence that are most relevant to goal achievement. Consideration of goals and intentions are also important in the design and evaluation of clinical practice guidelines (Advani, Lo, and Shahar 1998; Shahar, Miksch, and Johnson 1996)

Evidence is not unitary. That is, a piece of data possesses variable properties, some of which can be the focus of attention, while other properties are ignored or relegated to secondary status. This is not simply saying that evidence can be interpreted in different ways, but that the evidence can be decomposed into variable properties along several dimensions. This would explain differences in interpretation made by experts in a given domain. This has implications for the development of guidelines, as the purpose of the guideline should be an important aspect in making clinical guidelines more adaptable to the user's needs.

7.3 Strategies for Evaluating Evidence

We have identified two major types of strategies to evaluate evidence: First, general strategies of confirmation and disconfirmation; and second, specific responses to inconsistent or anomalous data. Although the preferred strategy for evaluating data is confirmation, attempting to "falsify" a hypothesis may be more fruitful.

Responses to anomalous data seem to vary greatly among people. We saw some well-known cases in the history of science, although such responses also appear when evaluating clinical data. Furthermore, research suggests distinct variations along expertise, where the types of responses may change from novice to expert. Studies conducted with novices and experts have shown that early novices (e.g., beginning medical students) respond to anomalous data by simply ignoring them. More experienced students tend to reinterpret the anomalous data in terms of a known hypothesis. The developmental pattern of responses to data uncovered in research studies can be used in the development of evidence-based systems for both practice and education

The approaches to evidence interpretation and use described in this document lead to a number of principles, presented in the executive summary, that can be used by designers and users of clinical practice guidelines. They are reiterated here:

7.4 Principles for the Design, Use, and Evaluation of Evidence-based Tools

Principle 1: The structure of an argument and the role that evidentiary statements play in such structure should be taken into account when designing and using evidence-based information. The argumentative approach can be used to improve the acceptance of evidence by designing argument structures that people find more convincing. This can be used in the design of textual information that optimizes the quality and convincingness of an argument. However, research is needed to test the empirically test argumentative approach with end-users of medical information.

Principle 2: Since representations are controlled relative to a goal, decision tools such as clinical guidelines should stress the possible goals that are likely to be sought and the results that are need to be perceived in order to reach those goals. Goals should be an important consideration in the design of clinical guidelines, as evidence that is useful may vary with the purpose of the guideline (e.g., screening, diagnosis). By clearly establishing the purpose of the guideline, evidence can be chosen that best supports the purpose.

Principle 3: Although typically we interpret a procedure in terms of steps to be performed (i.e., what to do), it s important to realize that actions are likely to change from one situation to another. Instead of stressing actions, clinical guidelines should stress results of actions. For instance, clinical guidelines could be design such that they specify not only what procedures or tests to perform but what consequences such tests can have (e.g., what information is to be gained, what undesirable effects they may have).

Principle 4: As circumstances change and there is no single set of actions that bring about a goal, flexibility in the ways to achieve the goals should be designed into practice guidelines and other evidence-based tools. Every circumstance is different from one another. For instance two patients with the same diagnosis may respond differently to the same treatment. What worked with one patient may not work with the other or what worked with a patient once may not work a second time. This circumstantial variability suggests then that there is not "a right way" to proceed and therefore, evidence-based tools should allow flexibility in the application of recommendations for practice. Flexibility is also inherent in the timeliness of guideline information. In this respect, information in guidelines should be easily accessible at the right time, when is needed. In this way, just-in-time information makes more efficient the use of the guidelines by taken the burden of the user's memory to access later such information (it relieves cognitive load).

Principle 5: At the strategic level there are two ways in which evidence can be generated and evaluated: by searching and providing (a) supportive or (b) contrary evidence for a hypothesis. Given that supportive evidence is much more likely to be used in natural situations and that contrary evidence is logically stronger in evaluating a hypothesis, evidence-based clinical guidelines may be improved by providing reminders to information (e.g., patient signs and symptoms) that disconfirm the hypothesis.

Principle 6: There are different ways in which people respond to anomalous evidence. Those responses depend on the quality of the evidence (i.e., how good the evidence is) as well as on the degree of entrenchment of the prior beliefs and knowledge. This suggests that the evidence used in clinical guidelines should be assessed in the context of the many constraints (e.g., the state of current knowledge, the degree of risk of using the new evidence, patient-specific pattern of signs and symptoms) that influence the evaluation of the evidence, not on the quality of the evidence alone.

Principle 7: In practice, physicians base their decisions on their prior knowledge, which they bring to bear in decision making. Prior knowledge in the form of experiential knowledge (heuristics, analogies, reminders) is more pervasive than knowledge obtained indirectly, such as biomedical knowledge. That is, given a discrepancy between experiential knowledge and biomedical knowledge, it is more likely that physicians rely more on the former than on the latter. It is important then to design clinical guidelines that make use of both types of knowledge to be successfully used in decision making (e.g., implemented in decision support systems as reminders).

Principle 8: People process information at different levels of granularity. These different levels are related to the expertise of the person. At higher levels of expertise, physicians can use information the lack in detail but are more general (e.g., facets). Less than experts, in contrast, process information that is more detailed and specific. Information in clinical guidelines should therefore match the level of knowledge granularity with the level of expertise of the user. Evidence at higher levels of knowledge processed by physicians may be inadequate for less experienced physicians.

Principle 9: The previous principle brings us to the importance of expertise in interpretation, understanding and decision making. In general the level of expertise of the user should be an important consideration in the design and implementation of clinical decision support (e.g., practice guidelines).

Principle 10: Information also should match not only the expertise but also the purpose of the guideline. A guideline can have a heuristic role (i.e., when it plays a role in the decision) or a support role (i.e., when it is used to justify one's decisions or in teaching and learning). When used to make decisions, guidelines can contain only the information needed to reach the decision. However, information at a more detail level may be needed when guidelines are used as a learning/teaching tool.

In summary, this report describes a framework for the cognitive investigation of evidence interpretation and utilization. Since this is a first attempt at developing such a framework, we realize that the framework has limitation and will likely require modifications based on ongoing and future empirical work. Nonetheless, we believe that the framework can provide a starting point for both carrying out investigations of evidence uptake and developing design principles for guideline development and implementation. Our purpose is to improve the framework through an iterative process of theoretical refinement and empirical investigation, where there is a mutual influence between empirical research and theoretical development.

References

Advani, A., Lo, K., Shahar, Y. 1998. Intention-based critiquing of guideline-oriented medical care. *Proceedings of the AMIA Annual Symposium*:483-7

Allais, M., Hagen, O. 1979. *Expected utility hypothesis and the Allais paradox*. Dordretch, Holland: D. Reidel

Allen, V., Arocha, J., Patel, V. 1998. Evaluating evidence against diagnostic hypotheses in clinical decision making by students, residents, and physicians. *International Journal of Medical Informatics* 51

Altman, L. 1986. *Who Goes First? The Story of Self- Experimentation in Medicine*. New York: Random House

Alvarez, L., Alvarez, W., Asaro, F., Michel, L. V. 1980. Extraterrestrial cause of the Cretaceous-Tertiary extinction: Experimental result and theoretical interpretation. *Science* 208:1095-1108

Arocha, J. A., Patel, V. L. 1995. Novice diagnostic reasoning in medicine: Accounting for clinical evidence. *Journal of the Learning Sciences* 4:355-384

Arocha, J. F., Patel, V. L. 1996. Methodologies for cognitive assessment in medical informatics: Knowledge organization and evaluation of hypothesis and evidence. . In *COACH 21 Scientific Program Proceedings*, ed. J. Moehr, F. Lau, pp. 1-7. Edmonton, Alberta: Canada Health Informatics Association

Aveyard, P. 1997. Evidence-based medicine and public health. *Journal of Evaluation in Clinical Practice* 3:139-44

Brown, K. S. 1995. Testing the most curious subject—Oneself. *The Scientist* 9:1, 10

Bunge, M. 1985. *Philosophy of Science: Philosophy of Social Sciences and Technology*. Treatise on basic philosophy, VII. 8 vols.

Chapman, L. J., Chapman, J. P. 1969. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology* 72:271-280

Chinn, C. A., Brewer, W. F. 1993. The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research* 63:1-49

Davidoff, F. 1999. In the teeth of the evidence: the curious case of evidence-based medicine. *Mount Sinai Journal of Medicine* 66:75-83

Dickinson, H. D. 1998. Evidence-based decision-making: An argumentative approach. *International Journal of Medical Informatics* 51:71-81

Enkin, M. W., Jadad, A. R. 1998. Using anecdotal information in evidence-based health care: heresy or necessity? *Annals of Oncology* 9:963-6

Evans, D. A., Gadd, C. S. 1989. Managing coherence and context in medical problem-solving discourse. . In *Cognitive science in medicine: Biomedical modeling*, ed. D. A. Evans, V. L. Patel, pp. 211-255. Cambridge, MA: MIT Press

Evans, D. A., Patel, V. L., eds. 1989. *Cognitive science in medicine: Biomedical modeling*. Hillsdale, NJ: Lawrence Erlbaum

Green, J., Britten, N. 1998. Qualitative research and evidence based medicine. *Bmj* 316:1230-2

Helfand M. Screening for thyroid disease. Ann Intern Med. 1998; 129:144-158.

Horton, R. 1998. The grammar of interpretive medicine. *Canadian Medical Association Journal* 159:245-249

Kuhn, D. 1991. *The Skills of Argument* Cambridge: Cambridge University Press
Kuhn, D. 1992. Thinking as Argument. *Harvard Educational Review* 62:155-78
Marken, R. S. 1997. The dancer and the dance: Methods in the study of living control systems. *Psychological Methods* 2
McDonald, C. J. 1996. Medical heuristics: The silent adjudicators of clinical practice. *Annals of Internal Medicine* 124:56-62
Miettinen, O. S. 1998. Evidence in medicine: invited commentary. *Canadian Medical Association Journal* 158:215-221
Patel, V. L., Arocha, J. F., Kaufman, D. R. 1994. Diagnostic reasoning and expertise. *The Psychology of Learning and Motivation: Advances in Research and Theory* 31:137-252
Patel, V. L., Evans, D. A., Kaufman, D. R. 1989. Cognitive framework for doctor-patient interaction. . In *Cognitive science in medicine: Biomedical modeling*, ed. D. A. Evans, V. L. Patel, pp. 253-308. Cambridge, MA: MIT Press
Patel, V. L., Groen, G. J., Arocha, J. F. 1990. Medical expertise as a function of task difficulty. *Memory & Cognition* 18:394-406
Patel, V. L., Kaufman, D. R., Magder, S. A. 1996. The acquisition of medical expertise in complex dynamic environments. . In *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports and Games*, ed. K. A. Ericsson, pp. 127-165. Hillsdale, NJ: Lawrence Erlbaum
Penzias, A. A., Wilson, R. W. 1966. Determination of the Microwave Spectrum of Galactic Radiation. *Astrophysical Journal* 146:666-669
Popper, K. 1959. *The logic of scientific discovery* New York: Harper
Powers, W. T. 1973. *Behavior: The control of perception* Chicago: Aldine
Reilly, B. M., Hart, A., Evans, A. T. 1998. Part II. Evidence-based medicine: a passing fancy or the future of primary care? *Disease A Month* 44:370-99
Runkel, P. J. 1990. *Casting nets and testing specimens. Two grand methods of psychology* Westport, CN: Praeger
Sackett, D. L., Rosenburg, W. M. C., Gray, J. A. M., Haynes, R. B., Richardson, W. S. 1996. Evidence based medicine: What it is and what it isn't. *British Medical Journal* 312:71-72
Shahar, E. 1997. A Popperian perspective of the term 'evidence-based medicine'. *Journal of Evaluation in Clinical Practice* 3:109-16
Shahar, Y., Miksch, S., Johnson, P. 1996. An intention-based language for representing clinical guidelines. *Proceedings of the AMIA Annual Fall Symposium*:592-6
Stein, N. L., Miller, C. A. 1993. The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. . In *Advances in Instructional Psychology*, ed. R. Glaser, vol. 4, pp. 285-334. Hillsdale, NJ: Lawrence Erlbaum
Toulmin, S. 1958. *The Uses of Argument* Cambridge: Cambridge University Press
Toulmin, S., Rieke, R., Janik, A. 1984. *An Introduction to Reasoning* New York: Macmillan
Wegener, A. 1924. *The origin of continents and oceans* Trans. J.G.A. Skerl. London: Methuen
Williams, J.W., Mulrow, C.D., Chiquette, E., Aguilar, C., Cornell, J., Newer Pharmacotherapies for Depression in Adults and Adolescents: A Background Paper (Unpublished manuscript).
Wyatt, J. 1999. Same information, different decisions: format counts. *British Medical Journal*

318:1507-1502

Notes

[1]. If *p* (a hypothesis) implies *q* (an empirical derivation from the hypothesis, e.g., a result), and it happens that the result *q* is true, we cannot conclude that the hypothesis, *p*, is true, as it may be other reasons for *q*. The typical example is "If it rains, the grass in backyard will be wet. The grass is wet, therefore, it rained," which is a logically invalid argument since the grass maybe wet because of other causes.

[2]. If *p* implies *q* and *q* is false, then we can be assured that *p* is also false. To use the same example, If it rains, the grass in the backyard will be wet. The grass is, *not* wet therefore, it did not rained,"

[3]. Concurrently with Penzias and Wilson's discovery, a theory was proposed that predicted their results, based on the hypothesis that at the beginning of the universe a big explosion had occurred. Penzias and Wilson won the 1978 Nobel Prize for their discovery.