Quality of Life Research **12 (Suppl. 1)**: 3–16, 2003.

The significance of quality of life in health care

Robert M. Kaplan
Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA, USA (E-mail: rkaplan@ucsd.edu)

Accepted in revised form 9 February 2002

Abstract

This paper compares a traditional biomedical model with an outcomes model for evaluating health care. The traditional model emphasizes diagnosis and disease-specific outcomes. In contrast, the outcomes model emphasizes life expectancy and health-related quality of life. Although the models are similar, they lead to different conclusions with regard to some interventions. For some conditions, diagnosis and treatment may reduce the impact of a particular disease without extending life expectancy or improving quality of life. Older individuals with multiple co-morbidities may not benefit from treatments for a particular disease if competing health problems threaten life or reduce quality of life. In preventive medicine, diagnosis of disease is made more difficult because of ambiguity, uncertainty, lead-time bias, and length bias. In some circumstances, successful diagnosis and treatment may actually reduce life expectancy or overall life quality. Example applications of the outcomes model from clinical policy analysis, individual decision making and shared decision-making are offered. The outcomes model has received little attention in dental health care but may have parallels to applications in other areas of medicine.

Key words: Outcomes research, Quality of life measurement, Shared decision making, Utility

Abbreviations: DALY – disability adjusted life year; QALY – quality adjusted life years

Introduction

This article has three objectives. First, I will differentiate an Outcomes Model from a traditional Biomedical Model. Second, I will suggest that the traditional model leads to over-diagnosis and perhaps to excessive costs in health care. Third, I will propose that new methods of medical decision making, involving both patients and providers, can contribute to the solutions for these problems.

Biomedical and Outcomes Model

Medicine is the art of diagnosis and treatment. Preventive medicine is often regarded as the speciality for the early detection of treatment of disease. The traditional Biomedical Model, which is oriented toward disease, depends on measures of disease process. Diagnosis typically involves identification of pathology and uses measures of blood chemistry, tissue damage, and so forth. Successful intervention occurs when a disease is eradicated. We sometimes refer to this model as the 'find it–fix it' approach [1]. Diagnosis is used to find disease pathology and treatment is used to fix it. The traditional model reflects traditional thinking. Since the time of Sir Isaac Newton, linear thinking has been the predominate view of the world. Newton focused his attention on discrete components of the world and assumed that these operated with independence from one another.

The industrial revolution, which began in England in the 18th century, set the intellectual tone that dominated nearly all fields for the next two centuries [2]. Three concepts characterize linear

thinking: reductionism, analysis, and mechanism. Reductionism is the belief that everything that we experience is made up of component parts. Just as an automobile represents contributions from many different factories, we assume that human beings are also a conglomeration of component parts. Reductionistic science involves taking things apart. The parts become smaller and smaller until the scientist arrives at the ultimate basic elements which are no longer divisible. Reductionists believe that in order to understand something, it must be disassembled. Scientists often describe the parts as functioning independently of one another.

Analysis is the process by which things are divided into their components. These things may be tangible such as the human body or a machine; however, ideas can also be disassembled. Mechanism, the third basic component of linear thinking, is the belief that cause and effect can be described by one relationship. If x causes y, we may understand the mechanism of y by manipulating x. For example, if sun exposure causes red skin, we can recreate the red skin by placing a person in the sunlight. The sunlight is the mechanism that causes sunburn. Investigators rarely accept explanations at this global level. Instead, they search for finer and finer mechanisms that explain relationships at a more basic level. For example, they seek to understand the basic cellular events responsible for skin tone changes in response to sunlight.

In contrast to this linear thinking, there has been a trend toward 'systems' thinking. Complexity is fundamental in science. In the 17th century, Descartes proposed reductionism as a remedy to being overwhelmed by information. According to Descartes, complicated phenomenon could be understood by dividing them into their component parts [3]. It was assumed that this division would not distort the phenomenon under study. This approach has lead to many productive sciences. On the other hand, it is also apparent that there usually are dense interconnections between components in virtually all sciences [4]. In contrast to mechanistic understanding, systems thinking considers the whole rather than the individual parts.

In a system, the functioning of each part cannot be understood independently of the functioning of other parts. The value of individual parts is lost when the whole is disassembled. For example, an automobile broken down into component parts cannot be used to transport people. Similarly, a human body may not live or function adequately with some vital parts removed. The parts cannot function without the rest of the whole. A human eye cannot see if it is removed from the body, just as a steering wheel does not direct an automobile when it is removed from the machine [5]. Traditional scientific analysis tries to understand organisms by taking them apart and examining each part separately. This can be useful in determining the structure, but may not help us understand function.

The Outcomes Model

Systems thinkers recognize that the goal of health care is to help people live longer and feel better. This approach, known as the Outcomes Model, is similar to the traditional Biomedical Model in many ways. However, finding and fixing disease does not necessarily lead to the best patient outcomes. There may be occasions in which diagnosis does not contribute to improved life expectancy or quality of life. In fact, we will review occasions in which diagnosis and treatment may lead to losses in health status.

One of the important distinctions between the traditional Biomedical Model and the Outcomes Model is the value placed on patients' self-reports. The traditional Biomedical Model conceptualizes most problems as similar to acute disease. Acute diseases can typically be diagnosed and successfully treated, and are often identified through a biological test. With good testing, how patients report the experience may be of little value. Most of the information required to diagnose and treat the condition can be identified in the laboratory. The acute disease model dominates how we have developed health care, including the constructions of hospitals, the development of training programs, and the creation of medical sub-specialties [6]. The difficulty is that since about 1950, the major burden on our health care system has been chronic disease.

Chronic diseases typically have multiple causes; most people who have one chronic condition typically have other chronic diseases as well. The Medical Outcomes Study (MOS), for example, recruited patients who had one of six chronic

disease states. However, over 90% of the participants had other chronic conditions in addition to the one that placed in a category for the study [7]. In contrast to acute diseases that last a brief interval of time, chronic conditions are usually not cured. As a result, patients must adapt to their disease and psychological or social factors are of key importance. Patient interpretation of the condition and adaptation to the problem cannot be ignored.

The Outcomes Model also places greater emphasis on epidemiological data. In contrast to the traditional Biomedical Model that emphasizes identification of a basic disease mechanism, the Outcomes Model focuses on determinants of patient outcome. Sometimes, the exact biological model is unknown. For example, some people believe treatments are not of value unless the biological pathway underlying the disease is understood. The Outcomes Model recognizes that biologic pathways may never be fully understood [8]. Further, some behavioral risk factors affect health outcomes through a variety of different biological pathways. One example concerns the effects of tobacco use. For years, researchers attempted to identify the impact of tobacco use upon specific organs. Separate studies presented the effects of cigarette smoking upon lung cancer, heart disease, emphysema, oral cancers, and so on. By looking at the disease-specific impact of smoking and emphasizing the specific biological models, the total impact of tobacco use was underestimated. The outcomes approach links tobacco use to deaths from all causes, and to reductions in quality of life. Considered from this perspective, the impact of tobacco use is huge, accounting for an estimated 19% of all premature deaths [9]. Figure 1 shows that tobacco use is far and away the leading cause of preventable death in the United States.

One of the most important differences is in how the models define a unit of benefit. The traditional model usually links benefit to changes related to a diagnosis. For example, outcome might be assessed

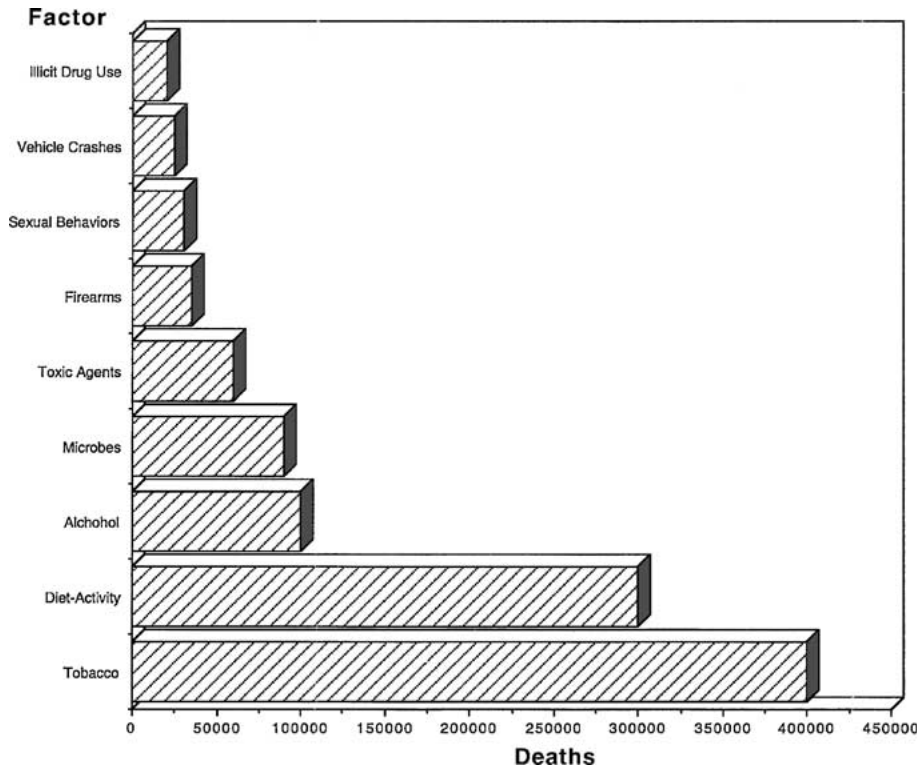

Figure 1. Actual Causes of Death – US 1990. Adapted from: McGinnis & Foege, JAMA, 270: 2207-12, 1993.

by changes in blood pressure, tumor size, or death from a specific disease. The traditional Biomedical Model often focuses on the small picture at the expense of the big picture. Much of contemporary preventive cardiology is based on observations from the Coronary Primary Prevention Trial, or CPPT [10]. In this experimental trial, men were randomly assigned to either take a placebo or to use a drug known as cholestyrmanine. Cholestyrmanine can significantly lower serum cholesterol. In this particular trial, cholestyrmanine produced an average total cholesterol reduction of 8.5%. In comparison to men using placebo, men in the treatment group experienced 24% fewer heart attack deaths and 19 fewer heart attacks.

One of the crucial features that make the Outcomes Model and the traditional Biomedical Model different is in how they measure the outcome. The CPPT showed a 24% reduction in cardiovascular mortality in the treated group. The absolute proportion of patients who died of cardiovascular disease was similar in the two groups. In the placebo group, there were 38 deaths among 1900 participants (2%). In the cholestyrmanine group, there were 30 deaths among 1906 participants (1.6%). In other words, taking medication for 6 years reduced the chances of dying from cardiovascular disease from 2 to 1.6%. The diagnosis-specific medical model focuses on cardiovascular deaths because the medicine was designed to reduce deaths from heart disease. Considering all causes of death, there was essentially no benefit of treatment. At the end of the study, 3.7% of those in the placebo group had died and 3.6% of those in the cholestyrmanine group had died. Since the publication of the CPPT, virtually every study has obtained the same result. Cholesterol lowering may reduce the chances of dying of heart disease, but does not reduce the chances of dying prematurely. The Outcomes Model does not take cause of death into consideration. From the outcomes perspective, the focus is on whether or not the patient is alive [11]. If a medication reduces the chances of dying of one disease while increasing the chances of dying of another, it is not regarded as effective [12]. Since virtually all treatments have the potential to produce harm as well as benefit, the Outcomes Model may be the most appropriate to evaluate benefits of treatment.

Medicine as a cognitive science

The traditional Biomedical Model treats disease as a binary variable. People are sick or they or not; however, most chronic diseases are gradual processes. The threshold for deciding whether or not someone has the disease can be ambiguous. This occurs not only in the definition of the disease, but also in the interpretation of clinical data [13]. Using their experience, clinicians examine and interpret clinical information. Like any judgment, these perceptions are not always reliable. For example, it is known that physicians are highly variable in their interpretation of clinical data. They disagree with one another when examining the same clinical information [14]. Further, they disagree with themselves when presented with the same information at two points in time. There are many examples to support this claim. For example, one study gave cardiologists high quality angiograms and asked them to say if the stenosis in the left anterior descending artery was greater than 50%. This judgment is important because it is usually the threshold for revascularization of the coronary arteries. The study showed that the clinicians disagreed with one another in about 60% of the cases [15]. In another study, cardiologists were given the same angiograms at two different points in time. At the second assessment, they disagreed with their own first judgment in between 8 and 37% of the cases [16].

Another study evaluated the reliability of pathologist-assessed ductile carcinoma *in situ* (DCIS). Six pathologist subjects were given written guidelines and examples of each of the problems they were looking for. Following this training, these experienced pathologists were given 24 high quality slides of breast tissue. There was considerable variability in the propensity to see DCIS. For example, one pathologist saw cancer in 12% of the slides while another saw DCIS in 33% of the same slides. Among 10 slides where at least one pathologist saw DCIS, no two pathologists had the same pattern of identification. One pathologist saw cancer in 8 of the 10 cases while another saw DCIS in only three. One case was diagnosed by only one pathologist, and only two cases were seen by all six [17]. These variations in diagnostic patterns imply that patients with the same problem, going to different doctors, may get different diagnoses. Table 1 summarizes recent

Table 1. Examples of studies on clinician agreement for studies in health status, radiology, and pathology

Reference	Problem	Comparison	Reliability	Comments
Health status				
Rothwell et al. [18]	Which domains of the SF-36 are important for patients with multiple sclerosis	Patients vs. clinicians	Poor agreement on which dimensions are important	Measures of disability were poorly correlated with patient rated quality of life
Shiels et al. [19]	Evaluation of intervention to improve reliability for severity of illness judgments	25 clinicians reviewed 14 patient records before and after training	Interclass correlations low before intervention and improved little with intervention	Reliability problems are not easily remedied through training
Unsworth et al. [20]	Discharge decisions for stroke patients	13 multidisciplinary teams (74 clinicians) rated 50 hypothetical cases	Poor correspondence of recommendations across teams	Recommendations may depend on characteristics of teams as much as on characteristics of patients
Pathology				
Kendall et al. [21]	Endometrial cancer diagnosis	Five pathologists rated 100 endometrial biopsies	κ's ranged from 0.67 to 0.89. For atypical hyperplasia $\kappa = 0.47$	Individual pathologists attend to different features.
Sylvester et al. [22]	Site variability in postmortem blood alcohol determination	Blood alcohol was measured from six sites in nine subjects (after death)	Taking samples from different sites produced different estimates of blood alcohol	Choice of site for sample can affect results.
Frierson et al. [23]	Histological grading of infiltrating ductal breast cancer	Six surgical pathologists rated 75 infiltrating ductal tumors	κ's ranged from 0.43 to 0.74 for histological grade	Normalizing mitotic counts resulted in only slight improvement
Radiology				
Drapé et al. [24]	Agreement on MR images for articular cartilage abnormalities in osteoarthritis	Multiple observers rated images from 43 patients	Interobserver reliability = 0.80	Ratings were significantly correlated with anatomic findings
Naitoh et al. [25]	Use of intraoral radiography to detect dental caries	Six observers rates 93 tooth surfaces	κ for interobserver agreement = 0.43	Agreement for diagnosis of caries modest using two different methods
Bright et al. [26]	Evaluation of rotator cuff disease	Six reviewers with different levels of training rated radiographs of 40 patients twice (separated by 4 months)	κ for inter observer reliability ranged from 0.01 to 0.75. For intraobserver, κ's ranged from 0.26 to 0.80	Reliability was higher for those with more training
Mussurakis et al. [27]	Detection and differentiation of breast cancer using contrast enhanced MRI	Three radiologists reviewed MRIs from 57 women	Agreement between radiologists was only moderate	Agreement better for detection than for differentiation
Dahlström and Lindvall [28]	Panoramic radiography to assess temporomandibular joint disease (TMD)	Two oral radiologists reviewed 50 TMD and 20 non-TMD cases on multiple occasions	Interobserver agreement varied between 0.22 and 0.65	Agreement improved some after calibration training
Jarvik et al. [29]	MR of for lumbar disk disease	Three radiologists reviewed 34 consecutive patients with back pain	κ's for extrusion present ranged from 0 to 0.78	Agreement for presence or absence of extrusion was only modest

studies from health status assessment, radiology, and pathology suggesting poor reliability of diagnostic judgment.

Small area variability

One of the consequences of this variation in diagnostic patterns is that health care expenditures can vary greatly across geographic areas. *The Dartmouth Atlas of Health Care* has documented remarkable variation of Medicare expenditures for Medicare recipients in various regions of the United States [30]. For example, health care expenditures for Medicare recipients in southern California, southern Texas and Florida are about twice as high per recipient as they are in other regions, such as New Mexico and parts of the Pacific Northwest [30]. The Medicare program spends almost twice as much per recipient in Boston, Massachusetts as it does in New Haven, Connecticut. Yet systematic investigations show that people in Boston enjoy at least the same level of health outcome as those in New Haven [30]. In fact, some evidence suggests that patients are more likely to be rehospitalized for the same conditions in Boston than in New Haven [31]. These findings have important implications. It is commonly believed that rates of medical care expenditures are driven by medical care need. The variation studies suggest that there is room for providers to make different decisions about what care is required. These decisions may be influenced by training, availability of hospital beds and methods of reimbursement.

Quality of life

The Biomedical Model and the Outcomes Model focus on different measures for the evaluation of health care. The traditional model is regarded as successful if disease is found and fixed. The Outcomes Model suggests that resources should be used to help people live longer and feel better. Finding and fixing disease may, or may not, contribute to this objective. In order to quantify the benefits of health care, it is necessary to build a comprehensive model of health benefit. Traditional measures of health outcome were very general. They included life expectancy, infant mortality, and disability days. The difficulty with these indicators is that they did not reflect most of the benefits of health care. For example, life expectancy and infant mortality are good measures because they allow for comparisons between programs with different specific objectives. The difficulty is that neither is sensitive to minor variations in health status. Treatment of most common illnesses may have relatively little effect on life expectancy. Infant mortality, although sensitive to socioeconomic variations, does not register the effect of health services delivered to people who are older than 1 year.

Survival analysis is an attractive generic measure of health status. Survival analysis gives a unit of credit for each year of survival. Suppose, for example, that a person has a life expectancy of 80 years and dies prematurely at age 50. In survival analysis, they are scored as 1.0 for each of the first 50 years and zero each year there after. The problem is that years with disability are scored the same as those years in perfect health. For example, a person with severe arthritis who is alive is scored exactly the same as someone in perfect health. To address this problem, we have proposed adjusted survival analysis. Using this method, we can summarize outcomes in terms of quality adjusted life years (QALYs). In quality adjusted survival analysis, years of wellness are scored on a continuum ranging from 0 for death to 1.0 for full function.

QALYs are measures of life expectancy with adjustments for quality of life [32–34]. QALYs integrate mortality and morbidity to express health status in terms of equivalents of well-years of life. If a woman dies of breast cancer at age 50 and one would have expected her to live to age 75, the disease was associated with 25 lost life years. If 100 women died at age 50 (and also had a life expectancies of 75 years) 2500 (100×25 years) life years would be lost.

Death is not the only outcome of concern in cancer. Many adults suffer from the disease leaving them somewhat disabled over long periods of time. Although still alive, the quality of their lives has diminished. QALYs take into consideration the quality of life consequences of these illnesses. For example, a disease that reduces quality of life by one half will take away 0.5

QALYs over the course of one year. If it affects two people, it will take away 1 year (equal 2×0.5) over a 1-year period. A pharmaceutical treatment that improves quality of life by 0.2 for each of five individuals will result in the equivalent of one QALY if the benefit is maintained over a 1-year period. The basic assumption is that 2 years scored as 0.5 add up to the equivalent of 1 year of complete wellness. Similarly, 4 years scored as 0.25 are equivalent to one completely well year of life. A treatment that boosts a patient's health from 0.5 to 0.75 produces the equivalent of 0.25 QALYs. If applied to four individuals, and the duration of the treatment effect is 1 year, the effect of the treatment would be equivalent to one completely well year of life. This system has the advantage of considering both benefits and side effects of programs in terms of the common QALY units. Although QALYs are typically assessed for patients, they can also be measured for others, including caregivers who are placed at risk because they experience stressful life events. The Institute of Medicine recommended that population health metrics be used to evaluate public programs and to assist the decision-making process [35].

In addition to health benefits, programs also have costs. Resources are limited, and good policy requires allocation to maximize life expectancy and health related quality of life. Thus, in addition to measuring health outcomes, costs must also be considered. Methodologies for estimating costs have now become standardized [32]. From an administrative perspective, cost estimates include all costs of treatment and costs associated with caring for any side effects of treatment. Typically, economic discounting is applied to adjust for using current assets to achieve a future benefit. From a social perspective, costs are broader and may include costs of family members staying off work to provide care. Comparing programs for a given population with a given medical condition, cost-effectiveness is measured as the change in costs of care for the program compared to the existing therapy or program, relative to the change in health measured in a standardized unit such as the QALY. The difference in costs over the difference in effectiveness is the *incremental cost-effectiveness*, and is usually expressed as the cost/QALY. Since the objective of all programs is to produce QALYs, the cost/QALY ratio can be used to show the relative efficiency of different programs [33].

In the following sections I will offer three examples of how the Outcomes Model may lead to different interpretation than the traditional Biomedical Model. The first concerns international public policy. The second example is relevant to clinical policy, and the third concerns individual patient decision making.

International policy

The World Health Organization and the World Bank have recently adopted a methodology similar to QALYs. Using data on disease and disability, they have proposed the disability adjusted life year (DALY) as a metric for prioritizing world health needs. DALYs are scored in the opposite direction as QALYs, but are conceptually quite similar. Figure 2 shows the relationship between years of life lost and QALYs lost for a variety of different health problems. Some of the most important data points are in the upper left corner of the graph. Conditions such as osteoarthritis and depression are not major causes of life years lost; however, they are among the leading causes of DALYs lost [36].

When using DALYs as a measure of outcome, it is apparent that some of the major health threats have received too little attention. For example, we are tremendously concerned about infectious problems such as the Ebola virus. However, the impact of these infectious problems is relatively minor in relation to mental illness, which is the number one threat to worldwide health. Other problems, such as those associated with tobacco use, are also major contributors to the number of DALYs lost. Over the next 30 years, it is expected that DALYs lost to infectious diseases will decline while those lost to major chronic diseases will increase. In effect, the disease burden patterns of Westernized developed countries will begin emerging in the developing world [36].

Clinical policy

Another distinction between the Outcomes Model and the traditional Biomedical Model is in the

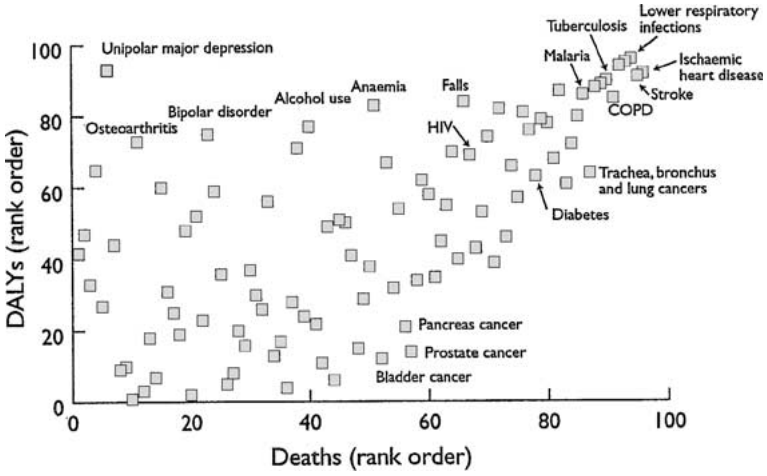

Figure 2. Comparison of global burden of illness by deaths alone (*x*-axis) vs. DALYs lost (*y*-axis). From Murray and Lopez (1996; p. 25).

definition of disease. The traditional view is that disease is binary; the disease is regarded as either present or absent. If present, treatment is typically indicated. The Outcomes Model regards disease as a process. Although biological abnormalities might be detected, they are not considered problematic unless they threaten the life expectancy or may reduce health-related quality of life. Biological abnormalities that will not affect either life expectancy, or life quality, are called pseudo-disease [37]. Pseudo-disease is very common. As a result, efforts to screen populations for health problems will result in a lot of 'disease' and may produce significant expenditures on treatment. However, it is not clear that population health will improve. Organizations such as the American Heart Association, the American Lung Association, and the American Cancer Society (ACS) [38] argue that mass screenings for disease is necessary because observed disease represents only the tip of the iceberg. Clearly, greater screening will produce more cases. On the other hand, what will be detected includes both true disease and pseudo-disease.

Acknowledging that disease is a continuum requires that a specific threshold for diagnosis be defined. Changing the threshold might greatly increase the number of people designated for treatment. For example, the threshold for the treatment of high blood cholesterol used to be a total cholesterol of 240 mg/dl. About 20% of American adults fall in this category. When the definition was recently changed to 200 mg/dl, the proportion of American adults who needed treatment increase to over 50% [13]. The Outcomes Model attempts to estimate the likelihood that each patient will benefit from treatment. An individual with a total cholesterol of 201 mg/dl might be regarded as requiring treatment according to the traditional model; however, the person may be exposed to the risks of treatment with very little individual potential to benefit. The Outcomes Model attempts to develop decision models that maximize QALYs for each patient.

In order to understand some of the differences in the way information is perceived by the two models, it is necessary to consider two biases: lead time bias and length bias.

Lead time bias

Cancer screening may result in early detection of disease. Survival is typically calculated from the date that disease is documented until death. Since screening is associated with earlier disease detection, the interval between detection and death is longer for screened cases than for unscreened cases. Epidemiologists refer to this as lead time bias. Figure 3 illustrates this bias.

Imagine that two men each develop prostate cancer in 1985 and die in 1997. Hypothetically, the

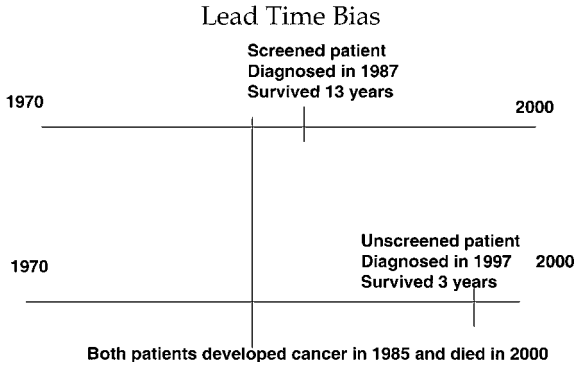

Figure 3. Example of lead time bias. The two lines show two different patients, each patient develops prostate cancer in 1985 and dies 12 years later in 1997. However, it appears that the patient shown on the top line survives longer because the disease was detected earlier.

progression of the cancer is identical in these two men. The man illustrated on the top line of Figure 3 was screened in 1987 and the cancer was detected. After this diagnosis, he lived 10 additional years before his death in 1997. The man shown on the lower line did not receive screening, and developed symptoms of urinary retention in 1994. After this, he lived three additional years. Survival for the man on the top appears to be much longer than that for the man on the bottom, even though the interval between developing cancer and dying is exactly the same. Referring back to Figure 3 which showed changes in survival among those diagnosed with prostate cancer according to the ACS, these data prompted the conclusion, 'Over the past 30 years, the survival rate for all stages combined has increased from 50 to 87%' [38]. The ACS attributes these changes to advances in cancer diagnosis and treatment.

Observational (nonrandomized) studies are often unable to separate lead time bias from treatment effect. It has been suggested that increased survival associated with screening can be attributed to lead time, and not to early detection and treatment [39, 40]. The only way to eliminate lead time bias is to perform clinical trials in which men are randomly assigned to either treatment or control groups and followed for many years. To date, there has been no randomized clinical trial evaluating the benefits of screening for prostate cancer. As a result, the ACS statement on increased survival cannot be confirmed nor refuted.

Length bias

Tumors progress at different rates. Some cancers are very-slow growing while other tumors progress very rapidly. Some cases may regress, remain stable, or progress so slowly that they never produce a clinical problem during an ordinary lifetime. These cases might be described as pseudo-disease because they are not clinically important [41]. The probability that disease is detected through screening is inversely proportional to the rate of progression. For example, with rapidly progressing disease, early detection may not produce a clinical benefit because cases are detected too late. On the other hand, diseases with very long preclinical phases are more likely to be detected by screening. However, diseases that are progressing extremely slowly may never cause clinical problems. Ironically, advances in screening technology have a greater likelihood of detecting cases for which a clinical manifestation will never materialize [42].

It is possible that some of the apparent benefits of screening and treatment for cancer are actually attributable to lead time and length bias. If this were true, then the greater incidence of detected disease would not be reflected in reduced mortality rates. This appears to be the case for prostate cancer. Current data suggest that, despite increases in screening, mortality rates of prostate cancer have remained relatively constant over the last two decades [43]. The same holds for ovarian cancer, breast cancer, colon cancer, and most other malignancies (except lung cancer).

Technology will improve disease detection rates. Newer approaches adjust raw prostate specific antigen (PSA) level by gland density [44] or use ratios of free to complexed PSA [45]. These approaches are still under evaluation, but it is likely that they will identify more cases at an earlier stage. Although they may identify some men who will benefit from early treatment, they will also find a larger number of men who would have died never knowing they had prostate cancer.

In summary, we typically assume that the more sensitive the test, the more it will contribute to population health status. However, tests can also do harm because false-positive tests can lead to other investigation that might be physically or psychologically harmful [46]. As a result, patients

cannot be given a clear and definitive answer to questions about whether they should be tested.

Individual decision making

The Outcomes Model shifts the focus of health care from finding and fixing disease to maximizing QALYs. Although the outcomes model and the traditional Biomedical Model are similar in many ways, they lead to very different approaches to care. According to the traditional Biomedical Model, medicine is about diagnosis and treatment (finding and fixing). According to the Outcomes Model, medicine is about making decisions that will maximize the quality adjusted life expectancy.

Perhaps the best example of contrast between the two models concerns the diagnosis and treatment of prostate cancer. Prostate cancer is an important health problem; it is the second leading cause of cancer death among men (behind lung cancer). The epidemiology is interesting because there may be a large reservoir of undetected cases [42]. The National Center for Health Statistics reports that there were 132,000 new cases in 1992 [47]. The ACS reported that there were 334,500 new cases in 1997 [38]. National data suggest that there were 34,000 deaths from prostate cancer in 1996 while the ACS projected 41,000 expected deaths in 1997. There are significant differences of opinion about whether the public should invest in screening programs for prostate cancer. The American Urological Association and the ACS have promoted large-scale screening of all men older than age 50 [48]. These organizations suggest a yearly screening using digital-rectal exams or PSA. The State of California enacted legislation in 1998 requiring physicians to advise men about the benefits of prostate cancer screening. Other organizations, including the American College of Physicians, argue that such screening programs may be of limited benefit [49a, b] and that they may be costly, accounting for about 5% of all health care costs [50].

One of the challenges is to determine whether there really is an epidemic of prostate cancer. Figure 4 shows changes in prostate cancer incidence and mortality between 1976 and 1994. The

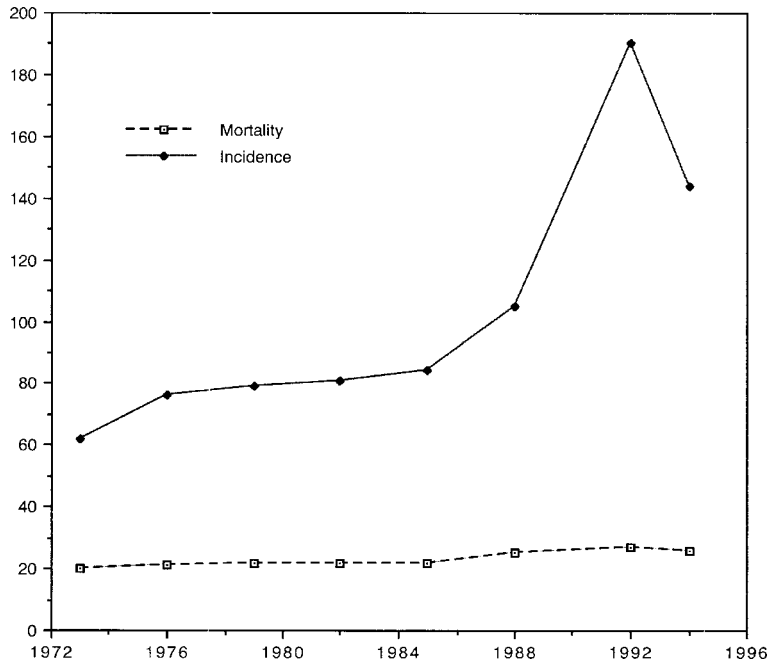

Figure 4. Prostate cancer incidents and mortality per 100,000 men in the US population from 1973 to 1994. Data from SEER, Cancer Statistics Review, 1973–1994 (NIH publication no. 97-2789).

number of reported prostate cancer cases doubled over this interval. Following concern about the value of screening, there has been a recent downturn in incidence. However, mortality from prostate cancer has remained relatively constant. One explanation for this apparent discrepancy is that there is a reservoir of undetected prostate cancer. Many of the undetected cases are unlikely to lead to ill health or death [37].

Ambiguity

The Outcomes Model recognizes the significant ambiguity surrounding many treatment choices. Although we may be able to find and attempt to fix prostate cancer, the real challenge is in deciding if diagnosis and treatment is valuable.

There have been several simulations of the benefits of screening and treatment. There are at least three methods available to screen for prostate cancer: digital rectal exams, trans-rectal ultra sound, and PSA. About 3% of all men will die of prostate cancer; however, autopsy studies show that for men in their mid-seventies, about 40% have prostate cancer [41]. Better diagnostic procedures will identify more men who have the condition. For those who do have disease there are three options: radical prostatectomy (surgical removal of the prostate gland), external beam radiation, and watchful waiting.

For men who choose radical prostatectomy, it is unclear whether there is a survival benefit [49a, b]. They may gain some relief knowing that they have chosen the most aggressive option, although, there are consequences. Among men receiving radical prostatectomy, about 40% will become incontinent, and 30% of these will have incontinence that requires the use of pads or clamps. Sixty percent of the men who will undergo prostatectomy will become impotent, and only about 11% will have had sexual intercourse in the 30 days prior to the interview [51].

The traditional model encourages treatment for those with a diagnosis (find it–fix it). The Outcomes Model recognizes another option: watchful waiting. Watchful waiting involves monitoring the condition without treatment; treatment can be initiated if the disease changes. Understanding the value of watchful waiting requires an understanding of the natural history of disease. Computer simulations of cohorts of 68-year-old men suggest that the risk of distant metastasis is about 5 per 100 patient years. The median time to metastasis is about 14 years. During the 14-year interval, 58% of the men will die of other causes prior to the development of metastatic problems from their prostate cancer. For those who do develop metastases, hormonal therapy can provide control of symptoms, and can delay disease progression long enough that many of the men die of other causes prior to serious complications from their prostate cancer [52].

Using QALYs as an outcome measure, simulations suggest there are few benefits of screening. For example, Krahn et al. [53] estimated the population benefit for programs to screen 70-year-old men for prostate cancer. They found that the benefits, on average, were improvements in the life expectancy between a few hours and 2 days. However, when they adjusted the life expectancy for quality of life, they discovered that screening programs reduced quality adjusted life days. The reason for this negative impact is that screening identifies many men who would have died of other causes. These men, once identified with prostate cancer, are then likely to engage in a series of treatments that would significantly reduce their quality of life. For these men, the treatment causes harm without producing substantial benefits.

Shared decision making

The Outcomes Model recognizes that there is ambiguity surrounding many health care decisions. Although there are a few cases in which the choice of treatment is obvious, many choices in health care involve complex decisions in which there are risks and benefits for several different alternatives. Earlier several studies on small area variations were summarized. These studies suggest that the likelihood of getting certain medical procedures such as coronary artery bypass surgery, surgery for prostate cancer, or end of life treatment, vary dramatically across geographic areas. Physicians are significantly more likely to prefer aggressive care in some communities than in others [30]. The treatments that patients receive are largely a function of physician choice.

Patients are rarely told about the ambiguity of the data, and are rarely involved in the decision process.

One of the best examples involves mastectomy vs. lumpectomy for women with well-defined breast cancer. The reason this particular example is valuable is that the probability of surviving lumpectomy vs. mastectomy (plus radiation) is approximately equal [54]. In other words, a woman has the same chance of surviving her breast cancer with each of these alternatives. The real issue for the women is their desire to retain breast tissue. For some women the cosmetic advantage of maintaining the breast is very important; however, some women may feel more comfortable knowing that a larger amount of potentially cancer-prone tissue is removed.

Studies of the use of mastectomy vs. lumpectomy have shown that there is remarkable variability across US communities. Provo, Utah, for example, has the highest mastectomy rate in United States. Roughly half of the women with breast cancer in this community get mastectomies. On the other hand, Paterson, New Jersey has the lowest rate. Women in Provo are about 25 times more likely to have mastectomy than women in Patterson. The difference is largely physician choice because the evidence shows that the likelihood of surviving breast cancer is about the same for these two communities.

The shared decision making paradigm incorporates patient preferences into the decision process. The crucial factor is one of quality of life. Individual preferences for life with and without a breast can be taken into consideration. The most important challenge in using this shared decision making approach is in finding ways to communicate risk information. The first step in developing shared decision making tools requires a systematic review of the literature. Quality of studies must be taken into consideration and meta-analysis must identify the most probable results of different procedures. Effort must be devoted to communicating risk information so that it can be comprehended. This is best accomplished by reporting raw frequencies, or relative risks, rather than odds ratios. Interactive videos showing patients who have experienced different options may be of value, although very few studies have systematically evaluated these programs [55]. Perhaps the most important feature of the shared decision making paradigm is that it recognizes the unique information that each partner contributes. Health care providers may be best able to explain the treatment options and the probabilities for various outcomes. On the other hand, the effects of treatments may be very personal and reaction to these effects may differ from person to person. Only the patient is able to value these outcomes.

Summary

The traditional Biomedical Model and the Outcomes Model differ in a variety of ways. One of the most important distinctions is in the focus of attention. The traditional model emphasizes disease pathology and treatment. According to this model, the function of health care is to detect problems by identifying pathology. Once identified, treatment is initiated. The Outcomes Model focuses on the impact of detection and treatment. Often, identification of pathology and treatment result in improved patient outcomes; however, there may be cases in which identifying disease does not result in better patient outcomes. For example, there are many circumstances in which disease, if left undetected, has no impact on life expectancy or quality of life. As a result of this ambiguity, providers and patients must face difficult decisions about what treatment should or should not be initiated. The change in focus of attention may affect decisions at the public policy level, the clinical policy level, and in individual decision making.

Acknowledgements

Supported in part by grants R01 HS 09170 from the AHCPR, P60 AR 40770 from the NIH, TPRH-98-119-01 from the American Cancer Society. Some sections of this paper were adapted from: R.M. Kaplan. Shared medical decision-making: A new paradigm for behavioral medicine. Ann Behav Med 1999; 21(1): 3–11. Based on keynote address at the University of Pittsburgh Conference on Chronic Disorders, 'Quality of Life: In the Eye of the Beholder' October 25, 1998.

References

1. Kaplan RM. Shared medical decision making: A new paradigm for behavioral medicine. Ann Behav Med 1999; 21(1): 3–11.
2. Ackoff RL. Systems thinking and thinking systems. Syst Dyn Rev 1994; 10(2–3): 175–188.
3. Carter RB. Descartes' Medical Philosophy: The Organic Solution to the Mind-Body Problem. Baltimore MD: Johns Hopkins University Press, 1988.
4. Checkland P. Systems theory and management thinking. Am Behav Sci 1994; 38(1): 75–91.
5. Ackoff RL, Gharajedaghi J. Reflections on systems and their models. Syst Res 1996; 13: 13–23.
6. Holman HR, Lorig KR. Patient education: Essential to good health care for patients with chronic arthritis (editorial, comment). Arthritis Rheum 1997; 40(8): 1371–1373.
7. Ware JE Jr, Bayliss MS, Rogers WH, Kosinski M, Tarlov AR. Differences in 4-year health outcomes for elderly and poor, chronically ill patients treated in HMO and fee-for-service systems. Results from the Medical Outcomes Study. JAMA 1996; 276(13): 1039–1047.
8. Kaplan RM, The Ziggy theorem: Toward an outcomes-focused health psychology. Health Psychol 1994; 13(6): 451–460.
9. McGinnis JM, Foege WH. Actual causes of death in the United States. JAMA 1993; 270(18): 2207–2212.
10. Coronary Primary Prevention Trial Group. The Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. JAMA 1984; 251(3): 351–364.
11. Golomb BA. Cholesterol and violence: Is there a connection? Ann Intern Med 1998; 128(6): 478–487.
12. Kaplan RM. Behavior as the central outcome in health care. Am Psychol 1990; 45: 1211–1220.
13. Fisher ES, Welch HG. Avoiding the unintended consequences of growth in medical care: How might more be worse? JAMA 1999; 281(5): 446–453.
14. Eddy DM. Clinical decision making: From theory to practice. Principles for making difficult decisions in difficult times. JAMA 1994; 271(22): 1792–1798.
15. Zir LM, Miller SW, Dinsmore RE, Gilbert JP, Harthorne JW. Interobserver variability in coronary angiography. Circulation 1976; 53(4): 627–632.
16. Detre KM, Wright E, Murphy ML, Takaro T. Observer agreement in evaluating coronary angiograms. Circulation 1975; 52(6): 979–986.
17. Schnitt SJ, Connolly JL, Tavassoli FA, et al. Interobserver reproducibility in the diagnosis of ductal proliferative breast lesions using standardized criteria. Am J Surg Pathol 1992; 16(12): 1133–1143.
18. Rothwell PM, McDowell Z, Wong CK, Dorman PJ. Doctors and patients don't agree: Cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. Br Med J (Clin Res Ed) 1997; 314(7094): 1580–1583.
19. Shiels C, Hutchinson A, Eccles M, Smoljanovic L. Accuracy and reliability of assessment of severity of illness before and after an educational intervention. J Eval Clin Pract 1996; 2(4): 265–271.
20. Unsworth CA, Thomas SA, Greenwood KM. Rehabilitation teams decisions on discharge housing for stroke patients. Arch Phys Med Rehabil 1995; 76(4): 331–340.
21. Kendall BS, Ronnett BM, Isacson C, et al. Reproducibility of the diagnosis of endometrial hyperplasia, atypical hyperplasia, and well-differentiated carcinoma. Am J Surg Path 1998; 22(8): 1012–1019.
22. Sylvester PA, Wong NA, Warren BF, Ranson DL. Unacceptably high site variability in postmortem blood alcohol analysis. J Clin Path 1998; 51(3): 250–252.
23. Frierson HF Jr, Wolber RA, Berean KW, Franquemont DW, Gaffey MJ, Boyd JC. Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. Am J Clin Pathol 1995; 103(2): 195–198.
24. Drapé JL, Pessis E, Auleley GR, Chevrot A, Dougados M, Ayral X. Quantitative MR imaging evaluation of chondropathy in osteoarthritic knees. Radiology 1998; 208(1): 49–55.
25. Naitoh M, Yuasa H, Toyama M, Shiojima M, Ushida M, Iida H. Observer agreement in the detection of proximal caries with direct digital intraoral radiography. Oral Surg Oral Med Oral Path Oral Radiol Endod 1998; 85(1): 107–112.
26. Bright AS, Torpey B, Magid D, Codd T, McFarland EG. Reliability of radiographic evaluation for acromial morphology. Skeletal Radiol 1997; 26(12): 718–721.
27. Mussurakis S, Buckley DL, Coady AM, Turnbull LW, Horsman A. Observer variability in the interpretation of contrast enhanced MRI of the breast. Br J Radiol 1996; 69(827): 1009–1016.
28. Dahlström L, Lindvall AM. Assessment of temporomandibular joint disease by panoramic radiography: Reliability and validity in relation to tomography. Dentomaxillofaci Radiol 1996; 25(4): 197–201.
29. Jarvik JG, Haynor DR, Koepsell TD, Bronstein A, Ashley D, Deyo RA. Interreader reliability for a new classification of lumbar disk disease. Acad Radiol 1996; 3(7): 537–544.
30. Wennberg JE. The Dartmouth Atlas of Health Care in the United States. Hanover, NH: Trustees of Dartmouth College, 1998.
31. Fisher ES, Wennberg JE, Stukel TA, Sharp SM. Hospital Readmission Rates for Cohorts of Medicare Beneficiaries in Boston and New Haven. New England J Med 1994; 331(15); 989–995.
32. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
33. Kaplan RM, Anderson JP. The general health policy model: An integrated approach. In: Spilker B (ed), Quality of Life and Pharmacoeconomics in Clinical Trials. New York: Raven, 1996; 309–322.
34. Kaplan RM, Ganiats TG, Sieber WJ, Anderson JP. The Quality of Well-Being Scale: Critical similarities and differences with SF-36. Int J Qual Health Care 1998; 10(6): 509–520.

35. Field MJ, Gold MR. Summarizing Population Health, Washington DC, Institute of Medicine, 1998.
36. Lopez AD, Murray CC. The global burden of disease, 1990–2020 (news). Nat Med 1998; 4(11): 1241–1243.
37. Black WC, Welch HG. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. New Engl J Med 1993; 328(17): 1237–1243.
38. American Cancer Society. Cancer Facts and Figures-1999, Atlanta: American Cancer Society, 1999.
39. Eddy DM. Screening for breast cancer. Ann Int Med 1989; 111(5): 389–399.
40. Fletcher SW. Whither scientific deliberation in health policy recommendations? Alice in the Wonderland of breast-cancer screening. New Engl J Med 1997; 336(16): 1180–1183.
41. Black WC, Welch HG. Screening for disease. Am J Roentgenol 1997; 168(1): 3–11.
42. Welch HG, Black WC. Using autopsy series to estimate the disease 'reservoir' for ductal carcinoma in situ of the breast: How much more breast cancer can we find? Ann Intern Med 1997; 127(11): 1023–1028.
43. Landis SH, Murray T, Bolden S, Wingo PA. Cancer statistics, 1999. CA Cancer J Clin 1999; 49(1): 8–31.
44. Benson MC, Whang IS, Pantuck A, Ring K, Kaplan SA, Olsson CA. Prostate specific antigen density: A means of distinguishing benign prostatic hypertrophy and prostate cancer. J Urol 1992; 147(3 Pt 2): 815–816.
45. Smith DS, Catalona WJ. The nature of prostate cancer detected through prostate specific antigen based screening. J Urol 1994; 152(5 Pt 2): 1732–1736.
46. Kaplan RM. Breast cancer screening: When to begin? In: Blechman AE, Brownell KD (eds), Behavioral Medicine and Women: A Comprehensive Handbook, New York: Guilford Publications, 1998; 213–220.
47. Wingo PA, Ries LA, Rosenberg HM, Miller DS, Edwards BK. Cancer incidence and mortality, 1973–1995: A report card for the U.S. Cancer 1998; 82(6): 1197–1207.
48. Middleton RG. Prostate cancer: Are we screening and treating too much? (Editorial, comment). Ann Intern Med 1997; 126(6): 465–467.
49. (a) Coley CM, Barry MJ, Fleming C, Fahs MC, Mulley AG. Early detection of prostate cancer. Part II: Estimating the risks, benefits, and costs. American College of Physicians. Ann Intern Med 1997; 126(6): 468–479; (b) Coley CM, Barry MJ, Fleming C, Mulley AG. Early detection of prostate cancer. Part I: Prior probability and effectiveness of tests. The American College of Physicians. Ann Intern Med 1997; 126(5): 394–406.
50. Optenberg SA, Thompson IM. Economics of screening for carcinoma of the prostate. Urol Clin North Am 1997; 17(4): 719–737.
51. Fowler FJ Jr, Barry MJ, Lu-Yao G, Wasson JH, Bin L. Outcomes of external-beam radiation therapy for prostate cancer: A study of Medicare beneficiaries in three surveillance, epidemiology, and end results areas. J Clin Oncol 1996; 14(8): 2258–2265.
52. Fleming C, Wasson JH, Albertsen PC, Barry MJ, Wennberg JE. A decision analysis of alternative treatment strategies for clinically localized prostate cancer. Prostate Patient Outcomes Research Team. JAMA 1993; 269(20): 2650–2658.
53. Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for prostate cancer. A decision analytic view. JAMA 1994; 272(10): 773–780.
54. Fisher B. Cancer surgery: A commentary. Cancer Treat Rep 1984; 68(1): 31–41.
55. Frosch D, Kaplan RM. Shared medical decision making. Am J Prev Med 1999; 17(4): 285–294.

Address for correspondence: Robert M. Kaplan, Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA 92093, USA
Phone: +1-858-534-6058; Fax: +1-858-534-7517
E-mail: rkaplan@ucsd.edu