

Linking evidence-based medicine therapeutic summary measures to clinical decision analysis

Benjamin Djulbegovic*, Iztok Hozo[#], Gary H. Lyman^{\$}

Divisions of Bone Marrow Transplantation*
and Medical Oncology/Hematology^{\$}
H. Lee Moffitt Cancer Center and Research Institute
at the University of South Florida, Tampa, FL, USA

Department of Mathematics[#]
Indiana University Northwest
Gary, IN, USA

Correspondence:

Dr. Iztok Hozo
Associate Professor of Mathematics
Department of Mathematics
Indiana University Northwest
3400 Broadway
Gary, IN 46408
phone: (219) 980-6980
fax: (219) 981-4247
e-mail: ihozo@iunhaw1.iun.indiana.edu
www: <http://www.math.iun.indiana.edu>

Abstract

Objective: Evidence-based medicine (EBM) seeks to improve clinical practice by evaluating the quality of clinical evidence and ensuring that only the “best” evidence from clinical research is used in the management of individual patients. EBM has contributed to our understanding of the meaning of the benefit and harm of treatment as reported in the literature and it is often promoted as an aid to clinical decision making. However, EBM therapeutic summary measures reflect only a single dimension of clinical decision making. The purpose of this work is to show how EBM therapeutic summary measures can be effectively incorporated into medical decision making.

Design: The effective application of the therapeutic summary measures advocated by EBM requires their integration into the framework of clinical decision analysis. Clinical decision analysis involves not only the identification and specification of the probabilities of clinical events but also the assessment of their relative values or utilities. We present here several analytical models for the integration of EBM therapeutic summary measures within the framework of clinical decision analysis.

Main results: As expected, our analysis demonstrated that treatment should never be administered if its harm is greater than its efficacy generally expressed as relative risk reduction. Likewise, a diagnostic test should never be ordered if the therapeutic harm is greater than its efficacy. Intervention is always favored if the number needed to treat to avoid one adverse outcome (NNT) is smaller than the number needed to treat to harm one individual (NNH). When faced with a choice between two therapeutic options the action threshold above which an intervention is favored can be

expressed in terms of the harm inflicted (H) as $H \times \text{NNT}$ or NNT/NNH . If a patient's preferences are taken into account as relative value judgements (RV) of adverse events relative to that of therapeutic events, the action threshold is defined as $\text{NNT} \times (\text{RV}/\text{NNH})$.

Conclusions: In the setting of clinical decision making, EBM summary measures derived from population studies can be effectively used to define diagnostic and therapeutic action thresholds that may help in the management of individual patients.

Key Words: evidence-based medicine, decision analysis

Introduction

Evidence-based medicine (EBM) has emerged as a powerful problem-oriented approach to the practice of medicine that seeks to improve patient care by considering the quality of clinical evidence (1). In the category of therapeutics, the main focus of EBM is to evaluate treatment effect usually expressed as one of several therapeutic summary measures (2). EBM has been advanced as an important tool in clinical decision making which may aid physicians in selecting one treatment alternative over another (3, 4). Recommendations are often made concerning the preferred management strategy based on a comparison of the relative benefit and harm associated with competing treatment alternatives (5, 6). However, it is unclear how this understanding of treatment benefit and harm should actually relate to a specific clinical decision. For example, should we choose treatment that is more efficacious or one that is less harmful? What is the minimal therapeutic benefit at which a treatment is still worth administering? What is the maximal acceptable harm at which a treatment is still considered worthwhile? Meaningful answers to these questions as well as the application of EBM to everyday clinical practice can be achieved by linking therapeutic summary measures to the methods of formal decision analysis (7, 8).

Evidence-based therapeutic summary measures

Several summary measures have been introduced to express treatment effect in terms of either therapeutic benefit or harm (2). These summary measures relate to the morbidity and mortality of disease and the toxicity of treatment. In general, treatment can exert a beneficial effect by either reducing the risk of a poor outcome or by

increasing the chance of good outcome (2). Similarly, treatment can exert a harmful effect by either increasing the risk of a poor outcome or reducing the chance of a good outcome. For the purpose of brevity we will only concentrate on the most popular measures of therapeutic benefit and harm.

Popular indices of therapeutic benefit include: a) the treatment effect generally expressed as either the absolute or relative change in the rate of events and b) the number of patients who need to be treated to prevent one bad outcome or attain one good outcome (NNT) (5, 9). Treatment effect is commonly expressed as either the absolute risk difference (ARD) between event rates in the two groups, i.e., $ARD = Risk_1 - Risk_2$, or as the proportional relative risk reduction (RRR) in event rates, i.e., $RRR = (Risk_1 - Risk_2) / Risk_1 = 1 - Risk_2 / Risk_1$ (2, 5). Alternatively, NNT represents the reciprocal of the difference in event rates between the treatment alternatives such that $NNT = 1 / (Risk_1 - Risk_2) = 1 / ARD$ (5, 9).

The harmful effects of treatment can be presented in a similar way. The common way to express this is to assess the rates of adverse effects due to treatment or to calculate the NNH (the number of patients who must be treated for one to experience a harmful event). This can be expressed as absolute difference between two harms (AHD) as $NNH = 1 / (Harm_1 - Harm_2) = 1/AHD$ (2, 5).

One should note that these measures of benefit and harm are population-based and often derived from randomized controlled trials (10, 11). However, these population-derived therapeutic measures have been increasingly advocated in medical decision making in individual patients (5, 12). Failure to relate these population-based evidence-

based therapeutic measures to the care of individual patients has been one of the major criticisms of the EBM movement (10, 11). In this article, we show that it is possible to relate EBM therapeutic summary measures to practical action thresholds in clinical decision making for individual patients. By practical action threshold we refer to a clinical situation when two different management strategies have the same potential value, the decision often described as a "toss-up" (13) or the point of indifference with regard to a choice between available management actions (14). In this way, the threshold becomes a guide for action- when the threshold is crossed, the values of management strategies change helping the decision maker to select one strategy over another.

Decision Analysis

The true challenge of clinical medicine is that of effective decision making under conditions of considerable uncertainty. Common uncertainties encountered include the diagnosis, the benefit of treatment as well as the harm of treatment. Formal decision analysis represents an explicit, quantitative method of clinical decision making that involves the distinction between the probabilities of events and their relative values (15, 16). The value associated with each clinical outcome can be expressed in different units such as length of life, morbidity or mortality rates, absence of pain, cost, or the strength of individual or societal preference for an outcome often expressed in terms of utilities (14-17). Note that we are using here a very broad definition for utilities associated with clinical outcomes in the manner proposed by Pauker and Kassirer (17) who followed its usage according to principles of classic decision analysis (15, 18). In medical literature,

it has become customary to equate the term "utility" with a measure of strength of the patient's preference for outcome (19). In this paper, the use of utilities may or may not include patients' preferences or judgement values toward a given clinical outcome (8). Indeed, we will show how outcomes expressed as morbidity or mortality can be integrated within a patient's value judgements to arrive at the optimal clinical decision.

In choosing between several competing clinical scenarios, decision theory holds that the optimal choice rests with the selection of the strategy associated with the optimal (greatest or smallest) expected value, calculated by averaging values across all possible outcomes, weighted by the corresponding probabilities (15, 16). Therefore, the preferred management strategy is the one associated with the optimal expected value of utility and is not directly dictated by the value of individual strategy outcomes.

Health outcomes can be expressed using evidence-based therapeutic summary measures

As noted above, EBM therapeutic summary measures refer to the effects of treatment on morbidity or mortality (2). EBM measures of therapeutic benefit or harm, therefore, relate to the modification of outcome values or utilities (7) (see above and Appendix). Outcome values can also be expressed as the proportion of patients who are free from the consequences of disease or the harm of the treatment (14, 17, 20, 21). The treatment is designed to modify the negative impact of disease on health status but often is associated with treatment-related harm as a trade-off. The preferred management strategy is the one with the greatest expected value, which depends on

the outcome value or utility as well as on the probability of each outcome.

We utilize the example of Pauker and Kassirer (14, 17) and of Rigelman and Schroth (22) representing the measures of benefit and harm as affecting the outcomes or utilities. Alternatively (23) EBM summary therapeutic measures may be assumed to affect the probability of clinical event instead of its outcome. However, the model shown in the Appendix (7, 20, 24) yields the same results under both scenarios. Therefore, applications of our model, including any conclusions and recommendations, remain unchanged.

Integration of EBM within a decision analytic framework

As detailed above, in the setting of clinical uncertainty, treatment summary measures utilized in EBM cannot alone tell us if one treatment alternative is preferred over another. Guyat et al (8) and Dowie (25, 26) have argued that the effective application of EBM requires integration within the framework of a decision analysis model. We have recently demonstrated that it is possible to link EBM measures of therapeutic efficacy to the framework of formal decision analysis both when comparing two treatments as well as when comparing treatment with observation (7, 20, 24). In so doing, we shift the focus from that of EBM measures of treatment efficacy to the more clinically relevant issue of the optimal choice among possible treatment alternatives. In addition to considering the measures of treatment effect, decision analysis allows us to address the trade-off that almost always exists between treatment benefit and harm whether measured in terms of mortality, morbidity or cost. While accurate measures of treatment efficacy are necessary to clinical decision making, they are not sufficient, as

they do not address the trade-off between benefit and harm essential for optimal patient care. Here we will illustrate the integration of EBM measures into clinical decision analysis providing important relationships between the summary measures of EBM and the trade-off inherent in clinical decision making.

We will consider decision-making in both the prophylactic/adjuvant and treatment settings, as well as in clinical situations when the diagnosis is certain, and when it is not. Finally, we will also address these issues in the setting of a diagnostic test.

Prophylactic/adjuvant setting

When comparing an intervention with no intervention, a direct link between EBM and decision analysis is clearly expressed in the equations outlined below (7, 20, 24). If the net harm associated with treatment can be represented in the same units as the net benefit, the threshold probability of disease (p_t) at which the expected value of treatment is exactly the same as the expected value of no treatment can be shown to be (17):

$$p_t = \text{Net Harm} / (\text{Net Benefit} + \text{Net Harm})$$

The clinical setting of comparing an intervention with no intervention is shown in Figure 1. If the harm associated with treatment is H and the event rate, e.g., mortality, without and with treatment is M and M_{Rx} respectively, then the absolute risk difference (ARD) and relative risk reduction (RRR), or efficacy (E) are given by:

$$ARD = M - M_{Rx}, \text{ and}$$

$$RRR = E = (M - M_{Rx}) / M$$

$$NNT = 1 / ARD$$

The threshold probability of disease at which we should be indifferent between treatment vs. no-treatment (Fig 1, Appendix) is shown to be (20, 24):

$$p_t = H / (E * M) = H / [M - M_{Rx}] = H / [S_{rx} - S] = H / ARD = H * NNT$$

This expression shows how different outcome data can be related to action threshold. That is, regardless whether data are expressed in terms of morbidity (M and M_{Rx} , as defined above), survival (S and S_{rx} represent disease-specific survival in those without and with treatment, respectively), absolute difference in outcomes (ARD) or NNT, we obtain the same threshold value. Note that our model (Appendix) refers to the measures of benefits as those related to the effect of the disease (with or without treatment) and to the measures of harm as those related to the treatment only. This understanding of benefit and harm corresponds to our usual clinical practice. However, the reader should be cautioned that when using survival data he/she should employ disease-specific survival data rather than overall survival, because overall survival may include not only effect of disease while patient is being treated, but also other causes of death.

It follows from the above equation that, if the probability that the patient has the disease is less than p_t , then treatment is not indicated. If the probability is greater than p_t , treatment should be given. If we assume that a patient on the no-treatment arm is actually on placebo, as it is commonly the case in randomized trials, the above

relationship becomes:

$$p_t = (H_{\text{active treatment}} - H_{\text{placebo}}) * NNT = NNT/NNH$$

Therefore, treatment should only be administered if we anticipated more good than harm, i.e., if $NNT < NNH$. One should also note that this relationship also holds for the situation where the diagnosis is certain. We should administer the treatment only when it satisfies conditions such that $NNT < NNH$, or $NNT < 1/H$ derived from relationships between measures of benefit and harm (24). Note that harm should be expressed as H only if we assume that alternative management (such as placebo, or no-treatment) of a patient without disease does not affect patient's health status at all, which in our notification was chosen to be equal to 1. If that assumption is not true, than difference in harm or NNH between two treatments should be used (see Appendix).

Such relationships allow physicians to tailor their treatment decisions to individual patients rather than to average patients from clinical trials (10) (8). Evidence can be suitably applied to the management of individual patients if used within a decision analytic framework (8). Clinical decisions can be assessed in the context of an individual patient most readily through a sensitivity analysis of the variables considered(20, 24) (addressed later in this paper). In this way we can test the stability of the action threshold based on changing assumptions about the evidence.

Clinical example: Kearon et al(27) recently reported a study in which they randomized patients who already completed a 3 month course of warfarin to determine if longer anticoagulation would be beneficial in the prevention of deep venous thrombosis (DVT)

recurrence. From this study, we found that the *NNT* for the prophylaxis of DVT recurrence is 4, i.e., 4 patients need to be treated with warfarin for 1 year in order to prevent one episode of DVT. The study concluded that a considerable risk of DVT recurrence exists beyond the typical 3 months of treatment with warfarin and that longer duration of anticoagulation might be necessary. However, the optimal duration of treatment needs to be interpreted in light of not only the benefit but also the harm of warfarin treatment.

While many would argue that the *NNT* of 4 represents a very effective therapy, this measure alone does not provide an answer to the question whether this treatment is better than the alternative management strategy of observation without active treatment. The crucial clinical question is: How good is one treatment strategy in comparison with another one when both benefit and harm are taken into consideration? To begin to address the clinical question whether to give warfarin or not, we note in the study by Kearon et al(27) that the annual risk of major bleeding was 3.8% (compared to zero in placebo arm) representing the $NNH = 26$. If we assume that the avoidance of DVT and bleeding complications represent approximately the same value to the patient, warfarin should be administered if the probability of DVT recurrence is greater than 15% (4/26). In this study, the recurrence rate for DVT was 27.4% per year suggesting that warfarin treatment should be continued beyond the initial 3 months of treatment in a typical patient (10) meeting the eligibility criteria described in the study by Kearon et al (27).

Now, let us consider a patient at increased risk for bleeding with continued warfarin use because of heavy alcohol intake. Although such patients were excluded

from the clinical trial discussed above, it is possible to individualize treatment decision making by applying the above equations to the patient's specific circumstances. Using published data, we may assume that the risk for bleeding in the patient under consideration is increased at least by 2.7 fold (28, 29), i.e., $NNH = 10 [1/(0.038*2.7)]$. This translates into a new action threshold of .4 ($NNT/NNH = 4/10$). Thus, the risk for DVT should exceed 40% per year to justify continued administration of warfarin. The optimal duration of anticoagulation could very well be only a few months for this patient.

It is equally important to ask “What is the highest *NNT* at which treatment is still worth administering?” As noted above, the treatment should only be considered if $NNT < NNH$ (7, 20). We recommend that the *NNT* not be used without concomitant data on treatment harm. In the example above, warfarin should not be administered if the $NNT > 26$ (or in the case of the patient at increased therapeutic harm because of heavy alcohol intake, warfarin should not be used if the $NNT > 10$). At or above this *NNT*, the harm of treatment would always outweigh the benefit, assuming harms and benefits are valued equivalently.

This approach also presents an answer to a question in the recent article by Steiner (11): “For $NNT > 1$, what is the minimal therapeutic benefit at which treatment is worth administering?” In addressing the application of population-based therapeutic measures to the care of individual patients, Steiner lamented that since we cannot be sure who will benefit from treatment, “all you can say is that on the basis of best available evidence, everything possible is being done to prevent an adverse effect” (11). As shown here, treatment is worth considering if $NNT < 1/H$ or if $NNT < NNH$ (20, 24).

On the other hand, the following question can be asked: “How much harm is

acceptable knowing the efficacy of treatment?”

Clinical example: A recent meta-analysis of chemotherapy for early-stage breast cancer indicated that we would need to treat 56 women with lymph node-negative disease, aged 60-69 years, in order to prevent one breast cancer recurrence (30). Is this benefit worth the harm associated with chemotherapy? The above relationship reveals that treatment is justified only if the harm is $<1/NNT$, which is 1.8% in this case. An average mortality associated with conventional adjuvant chemotherapy in postmenopausal women is reported to be 0.43% but it may vary from 0 to greater than to 3.2% (31). While on average, benefit of adjuvant therapy outweighs its risk for this group of patients, it also appears that there are patients in whom the benefit of adjuvant chemotherapy for breast cancer is not worth the risk. It is the responsibility of the physician to individualize treatment risk for each patient and to provide recommendations based on both risk and benefit. Sensitivity analyses varying EBM summary data obtained from population studies (24) may help examine how various co-morbid conditions and prognostic factors will affect decisions in the individual patient (10).

Clinicians often ask: “What happens if I don’t intervene?” and “If I intervene, will I do more good than harm?” The equations described above, e.g., $p_t = H/(E*M)$ in the prophylactic setting and $E_t = H / M$ in the therapeutic setting where $p=1$ provide a method to relate clinically relevant thresholds to knowledge concerning the three key parameters of medical decision making: a) the natural history of the disease (M), b) the harm inflicted by treatment (H) and c) the benefit (efficacy) of treatment (E). Note that for $H > M * E$, p becomes greater than 1. This confirms a traditional clinical principle: “Never administer treatment if its harm is greater than its efficacy ” and is

mathematically equivalent to the venerable Hippocratic principle “First, do no harm”(24).

Treatment Comparisons

The clinical setting of comparing one intervention with another is a variation of the scenario discussed in the previous section (Figure 1). When one treatment is compared to another, the threshold probability reduces to the following equation (Appendix):

$$p_t = NNT*(H_1 - H_2) = NNT/NNH = (H_1 - H_2) / [M * (E_1 - E_2)]$$
$$= (H_1 - H_2) / (M_{Rx2} - M_{Rx1}) = (H_1 - H_2) / (S_{Rx1} - S_{Rx2})$$

In this formulation, if the probability of a disease is greater than p_t , then treatment 1 should be given while if the probability of a disease is less than the threshold probability, treatment 2 is favored. This equation stresses the importance of knowledge concerning the natural history of the disease along with the benefit and harm of competing treatment strategies. As in the previous model, we can perform sensitivity analysis to obtain further insight into how EBM summary measures affect our management choices. Note that in this formulation $NNT=1 / (Mrx2-Mrx1) = 1 / (Srx1-Srx2) = 1 / ARD$, and $NNH = 1 / (H1 - H2) = 1 / AHD$.

As discussed earlier, Srx refers to disease-specific survival.

Clinical example: Using currently available data on benefit and harm in the management of high-risk breast cancer, we have shown that in women with ≥ 9 axillary lymph positive nodes, high-dose chemotherapy with stem cell rescue may be justified if the probability of a breast cancer relapse exceeds 29-40% at five years (20). However,

the decision to use high-dose vs. conventional adjuvant chemotherapy was very sensitive to the efficacy data utilized, none of which were obtained from randomized controlled trials. Under assumptions of low treatment efficacy for high-dose chemotherapy, we found that conventional adjuvant chemotherapy was equally attractive alternative (20). This analysis underscores the importance of reliable, high-quality data in decision analysis, such are those collected in prospective randomized clinical trials. Many decisions will be data-sensitive, i.e., the better the data, the better the decision will be. In the case of breast cancer, efforts to obtain better data are currently under way. While this is, undoubtedly, an important and necessary legacy of EBM, it is insufficient for optimal medical decision making.

What is the minimal difference in efficacy between treatment 1 and treatment 2 at which treatment 1 is worth considering? If treatment 1 is more toxic, how much more efficacious should treatment 1 be than treatment 2 in order to offset its greater risk? These issues can be addressed by considering the following expressions derived from the threshold equations above (Appendix): Treatment 1 is favored if the following inequality holds(20):

$S_{Rx1} > S_{Rx2} + (H_1 - H_2)$
or
$E_1 > E_2 + (H_1 - H_2) / M$

As expressed mathematically in these equations, we should administer only the more effective treatment that would provide better disease-specific survival than the alternative therapy adjusted for the harm difference between two treatments options.

Clinical example: We considered these relationships in determining the minimal efficacy of high-dose chemotherapy in high-risk breast cancer (20). We found that for a mortality difference between high-dose and conventional adjuvant chemotherapy of 5% in breast cancer with more than 9 positive axillary lymph nodes, bone marrow transplantation can be justified only if its efficacy is at least 30% greater than conventional adjuvant chemotherapy (20).

Integration of patient's preferences into clinical decision making

So far, we have shown how "hard" data on benefit and harm can be related to treatment action thresholds. These calculations assume that patients value the outcomes associated with the disease and adverse effects of treatment equally. However, patients often express different value judgements or preferences toward the positive and negative consequences of alternative management strategies. The method we presented can readily accommodate patient preferences or value judgements into the decision making process. Guyat et al(8) have demonstrated that the patient's value the avoidance of disease events that treatment is designed to prevent and the avoidance of the adverse effects of treatment differently and this can be expressed as the relative value (RV):

$$\text{Relative Value (RV)} = \frac{(1 - \text{value of adverse event})}{(1 - \text{value of disease event})}$$
$$= \text{value of avoiding treatment toxicity} / \text{value of avoiding disease outcome}$$

When this definition of patient preferences is adopted within our model we obtain

the following relationships:

$$P_t = NNT * [RV * H] \text{ or } RV * NNT / NNH$$

$$E_t = RV * H / (p * M) \text{ or } E_t = RV * H / M \text{ if } p = 1$$

In the case of diagnostic certainty, i.e., $p = 1$, treatment should be administered only if $NNT < NNH / RV$, or if efficacy of treatment is greater than E_t , as shown above.

Clinical example: Ezekowitz and Levine (32) recently performed a comprehensive literature review attempting to summarize data on the benefit and risk of antithrombotic agents in the prevention of stroke in patients with atrial fibrillation. They found that in patients between 65 and 75 years age, with no other risk factors, the annual risk for stroke was 4.3%, 1.1% and 1.4% in those receiving placebo, warfarin and aspirin respectively (32). At the same time, they found that the frequency of major bleeding was 1.2%, 1.0% and 1.0% annually for the warfarin, aspirin, and placebo groups, respectively (32). Does the benefit of treatment with warfarin or aspirin justify the potential harm (life-threatening bleeding)?

Applying the above formulas in the setting of diagnostic certainty (we know that the patient has atrial fibrillation), we find that if the choice was between warfarin or aspirin vs. no-treatment, the benefit overwhelmingly favors treatment (NNT=31 for warfarin vs. placebo, NNT=34 for aspirin vs. placebo; NNH=500 for warfarin vs. aspirin, and no net harm for aspirin vs. placebo according to the data presented by Ezekowitz and Levine (32)). If our decision was between warfarin and aspirin, we obtain that NNT=333 and NNH= ∞ , in favor of warfarin.

However, this recommendation does not take into account patients' values toward increased life-threatening bleeding relative to the prevention of stroke. Guyat et al (8) reports that, on average, patients consider 1 stroke equivalent to 5 episodes of serious gastrointestinal bleeding. Using formula shown above, Guyat et al. computed the relative value (RV) of life-threatening bleeding vs. experiencing stroke as 0.744 (8). Incorporating this RV, we would choose warfarin for $NNT < 672$ ($500/0.744$). Since there was no difference in harms between warfarin and aspirin ($NNH=\infty$), an additional consideration of patient values may help reinforce our decision to administer warfarin over aspirin.

Diagnostic Testing

When a diagnostic test is considered one of the clinical options, the question becomes whether to a) treat immediately, b) perform the test and base treatment on the test result, or c) continue observation without treatment or testing. The decision tree in figure 2 illustrates this clinical situation. Pauker and Kassirer (14) solved this decision tree to define two action thresholds, the testing threshold (p_{tt}) and the treatment threshold (p_{rx}). If the probability of the disease is larger than p_{rx} , the treatment should be administered without testing, if the probability of disease is smaller than p_{tt} , we should continue observing the patient and if the probability of the disease is between the two thresholds, the test should be performed and the patient treated or observed based on the test result.

If benefit and harm are expressed as evidence-based therapeutic measures (Fig 2), Pauker and Kassirer's threshold formulas (14) can be written as(7, 24):

$$p_{tt} = \frac{(1 - S_p) \cdot H_{rx} + H_{te}}{(1 - S_p) \cdot H_{rx} + S \cdot (E \cdot M - H_{rx})} \text{ and}$$

$$p_{rx} = \frac{S_p \cdot H_{rx} - H_{te}}{S_p \cdot H_{rx} + (1 - S) \cdot (E \cdot M - H_{rx})},$$

where S is test sensitivity, S_p is test specificity and H_{te} represents harm associated with diagnostic test. Examination of these formulas show that when the harm of treatment, H_{rx} , is larger than the benefit of treatment, $E \cdot M$, the term $S \cdot (E \cdot M - H_{rx})$ is negative, indicating that a testing threshold, p_{tt} , is not defined. This provides an intuitive corollary that a diagnostic test should never be ordered if the harm of treatment is greater than or equal to its benefit. This important clinical axiom was not readily apparent from the original threshold model (14). In fact, p_{tt} will be undefined unless the net benefit of treating those with a positive test, i.e., $[S \cdot (E \cdot M - H_{rx})]$, is greater than or equal to the harm of testing H_{te} .

The reader should note that the equations presented here assume that the diagnostic test is not perfect, that is, it is associated with certain false negative and false positive rate, as well as with certain harms (15). Of interest is to examine the behavior of these equations under assumption of expected value of perfect information, which is defined "as the difference between the averaged-out outcome value with a test and the averaged-out outcome value without a test when the test reveals the true disease state with certainty and it is assumed to have no risk"(15). That is, if we assume that the test is perfect, its sensitivity and specificity are equal to 1, and harm associated with its administration $H_{te}=0$. Under these circumstances all patients with positive test will have the disease, and all patients with negative test will not have the disease, and applying

the test will harm none. In other words, we will treat our patients only if we are absolutely certain that he/she has disease, and we will never order the test if the probability of disease is zero. Indeed, by solving the decision tree shown in Fig 2 under these assumptions, the treatment threshold (p_{rx}) is equal to one and the testing threshold (p_{tt}) becomes equal to zero.

Clinical example: Substantial disagreement exists concerning which patient should be treated with cholesterol lowering drugs in the primary prevention of coronary heart disease (CHD)(33, 34). The debate hinges on the accuracy of various methods to correctly predict who is at risk for CHD(34). The sensitivity and specificity of the various methods reported vary from 52% to 98% and from 37% to 96%, respectively(34). The efficacy of statins in the primary prevention of CHD is 30.5% at 5 years, and harm, expressed as the percentage of patients who had to discontinue medications during this five-year period was about 0.3%(35). Utilizing the above equations, the benefit of treatment is seen to outweigh its harm if the estimated risk of CHD is above 3.1% at 5 years using Sheffield tables, 30% at 5 years using modified Sheffield tables and above 23% at 5 years using criteria of the Joint Euro Task force to predict the risk of CHD (34). These dramatic differences in results reflects the large differences in the specificity of the various risk assessment methods (34).

On the other hand, if we are interested in determining the risk of CHD at which a patient should be offered testing to assess his or her own CHD risk, the test threshold should be calculated. Using the above formula as well as the data cited above, testing should be offered to any patient judged to be at greater than 0.56%, 1.3%, or 4.5% risk for CHD at 5 years depending on the method used (34). Thus, this relatively simple

method can shed considerable light on such important health issues as testing and treating patients at risk for CHD(36).

The reader should note that the same units should be used for benefit and harm in applying the equations presented here. In the above example we used morbidity data (non-fatal CHD and liver function abnormalities). However, if mortality data are used, action thresholds drop virtually to zero because of the low harm of statins. Since results using two different type of units (morbidity data, in which one can argue that prevention of myocardial infarction is not the same as liver function abnormalities vs. mortality data) produce dramatically different recommendations, the right course of action for any individual patient would be to elicit his or her preference values toward negative and positive consequences of using statins in prevention of CHD. Once the relative value judgements [RV] are elicited, they can be used in our formulas to calculate action thresholds.

We should note here that the calculation of action thresholds help us determine the probability of the disease at which treatment benefit outweighs its risk or, in the case of diagnostic testing, at which the probability of the disease at which ordering the test would be the optimal course of action. Glasziou and Irwig (37) demonstrated that an individual patient's risk can be obtained from large inception cohort studies in which major risk factors were defined. In the CHD example, the individual risk of coronary heart disease can be obtained from Framingham's equation which calculates the risk of CHD based on gender and history of smoking , hypertension, left ventricular dysfunction, the presence or absence of diabetes and total and HDL-cholesterol levels (38). Using some or all of these factors, individual CHD risk can be derived. This can

then be contrasted with calculated threshold probabilities to guide us in subsequent actions (e.g. observe vs. treatment vs. testing).

Discussion

Evidence-based movement has been particularly successful to highlight the fact that "not all evidence is created equal" (39) and that some evidence is closer to the truth than others. Hence, in the hierarchy of evidence related to the treatment, the high-power randomized trials and/or meta-analysis based on individual patient data is considered more valid evidence than the one collected in non-randomized trials or in anecdotal fashion (5, 40, 41). However, as stressed here, understanding the quality of evidence alone is not sufficient for effective clinical decision making.

The integration of evidence about the beneficial and harmful effects of alternative management strategies within a decision analytic model may improve clinical decision-making. This fact has been recently acknowledged by the Evidence-based Medicine Working Group of the Cochrane Collaboration (8). This group also developed a method for determining the threshold NNT to facilitate treatment recommendations for specific patients groups (8). This method, which also includes patients' preferences, is a special case of the model applicable to clinical situations when the diagnosis is certain (17, 20, 24). One should note that use of relative patients' values developed and recommended by the Evidence-based Medicine Working Group of the Cochrane Collaboration (8), and integrated in our model, is not equivalent to preference elicitation using standard gamble scenarios or time trade-off method commonly advocated in decision-theory literature (19). The fact is, however, that these standard methods are time-consuming and cumbersome to use, and have not penetrated in every day medical practice despite

that they have been developed decades ago. To facilitate effective bedside decision-making, simpler techniques are needed (42, 43). Development and elicitation patients values toward adverse treatment vs. disease effect is one such development (8) that is easily adopted within simple decision models presented in this paper.

We should also note that Riegelman and Schroth (22) provided several useful derivations of "adjusted" NNT that allows for the inclusion of multiple harms and benefits. They also demonstrated how other outcome measures such as life expectancy, cost-effectiveness or the results of decision analysis can be expressed using the NNT concept (22). Their method closely resembles that of Guyat et al's (8) and is also only applicable to a clinical situation when the diagnosis is certain. Interestingly, Willan et al.(44) and Schulzer and Mancini (45), using different modeling assumptions, derived a relationship between measures of harm and benefit similar to our NNT/NNH derivation. Their model is also applicable only to situations when diagnosis is absolutely certain. None of the works cited, however, attempted to relate measures of benefit and harm to practical action thresholds or to address the issue of diagnostic testing (22, 37, 44, 45). While, on the surface, all of these methods are different, they all eventually converge in their findings. The model presented here and elsewhere [7] using formal decision analysis to integrate EBM therapeutic summary measures, in fact, represents a general model within which other models described in the literature can be readily accommodated. Despite the power and credibility of the approach described, all these methods, including ours, assume a constant reduction in relative risk and fixed adverse effects (37). The reader should check these assumptions before acting upon results provided by any of these methods.

Furthermore, we want to highlight another important aspect of the work presented here, and that is striking dependence of calculation of the results on definition of benefit and harms. Different formulations for derivation of action thresholds are obtained when benefits and harms are expressed in terms of relative or absolute measures, respectively. In our model, NNT relate to the effect of the disease on health outcome (for example, treated patients vs. those not-treated, or treated with one therapy vs. treated with another therapy) and NNH specifically refers to treatment adverse effects. These definitions are directly related to definitions of benefit and harm in the decision tree (see Appendix). This distinction between NNT and NNH is often not clear in the contemporary usage of these EBM measures. For example, it is customary to denote negative NNT as NNH (46). While this notification may be correct, it would require change in the definition of benefits and harms (see Appendix), and consequently derivation of action threshold may change.

Incidentally, we note that people make different decisions when measures of benefit and harm are presented to them in different formats (e.g. relative risk reduction vs. absolute risk difference vs. NNT)(47-49). It would be interesting to investigate results obtained by our normative model with respect to actual physicians' decisions and way how information on benefit and harm was presented to them.

Conclusions

A practical method for the integration of commonly used EBM summary measures of therapeutic effect within the context of decision analysis is illustrated in two common clinical situations (prophylaxis and treatment comparison) as well as in situations when the diagnosis is certain and when it is not. When EBM therapeutic and

diagnostic summary measures are linked to decision analysis, some important principles of clinical decision making are affirmed e.g., never order a diagnostic test or administer treatment if its harm is greater than its relative risk reduction.

The above relationships were derived for relatively simple (two choice) clinical situations and for the most common EBM treatment measures. More complex relationships will arise when faced with multiple diagnostic and therapeutic strategies, with multiple outcomes and with other EBM therapeutic summary measures. Although the method presented here can be extended to embrace multiple adverse effects that may be associated with a given treatment (see Appendix), more complex modeling would often be required under these circumstances. In recent years, we have increasingly witnessed the use of complex decision models to help with policy-making decisions rather than with bedside decisions (50). In doing so, decision-analytic models commonly analyze events over or beyond time for which no direct evidence exist and use assumptions or data which are of poor quality. For example, there is high-quality evidence from randomized trials that bone marrow transplant is superior to conventional therapy in the treatment of myeloma at five years (51). However, decision-analytic models often continue to simulate and compare two treatment modalities beyond the time point at which high-quality evidence exists (52). This results in paradox between EBM and decision-analysis: while on one hand EBM measures are inadequate for optimal decision-making and require integration within decision-analytic framework, paradoxically the high quality evidence can be lost within the decision-analytic model that integrates a number of other assumptions based on the low-quality or even non-existing evidence. The solution, then, may be to integrate EBM measures within a

simple model that does not go beyond time point for which (high-quality) empirical data exists. This is another reason why our simple model may be used at the bedside decision making. Nevertheless, our method is not intended to replace clinical judgement but to supplement it. For the class of problems presented here, we believe that these methods provide a practical and educational tool to help improve clinical decision-making.

Appendix: Threshold relationships

Comparison of an Intervention to No intervention

The analytical solution of the tree in Figure 1 for the treatment of a single disease involves multiplication of the outcomes of the tree by its corresponding probabilities and solving for the probability of a disease at which we should be indifferent to the two strategies (p_t). This represents a typical clinical situation with uncertain diagnosis, e.g. whether to administer anticoagulants to a patient suspected of pulmonary embolism (17, 24), or adjuvant chemotherapy in a patient who underwent surgery for breast cancer(20).

The threshold probability of disease or relapse (p_t) at which the expected value of treatment equals the expected value of no treatment is the solution to the equation:

$$p * [1 - Mrx - H] + (1-p) * [1 - H] = p * [1 - M] + (1-p) * [1],$$

or
$$p * [(1 - Mrx - H) - (1 - M)] + (1 - (1 - H)) = H$$

where $[(1 - Mrx - H) - (1 - M)] = M - Mrx - H$ is the net benefit from treatment in those with the disease (outcome in those treated – outcome in those not treated) and $[1 - (1 - H)]$ is the net harm from treatment in those without the disease (outcome in those treated – outcome in those not treated)(14, 17). The reader should note that net benefit of treatment is restricted to patients who have the disease, and net harm applies to those patients without the disease (see reference (17) for details).

As explained in the text, **H** refers to the harm associated with treatment, and **M** and **M_{Rx}** to morbidity/mortality, without and with treatment, respectively. As illustrated in the text, all of these parameters need to be expressed as probabilities on a scale 0 to 1. The

difference between M and M_{Rx} is equal to absolute risk difference in event rates (ARD) ($M - M_{Rx} = ARD$). The analytic derivation of net benefits shown here is equivalent to Glasziou's and Irwig's axiomatic definition of net benefits(37).

As discussed in the text, the solution of the tree depends on definition of benefits and harms. Our model decomposes utilities into effects of the disease (with or without treatment) and the effect of treatment. Therefore, harm [(e.g. $NNH=1/(H1-H2)$)] will relate only to the adverse effect of treatment, and benefit (e.g. NNT) to the effect of the disease which may or may not be treated [(e.g. $NNT=1/(M-Mrx)$ or $1/(Mrx2-Mrx1)$] (see, later and the text).

It is important to note that our model assumes that Mrx (morbidity/mortality on treatment) and Hrx (treatment-related morbidity/mortality) are independent events and that the probability of both effect occurring simultaneously (e.g. while on tamoxifen patient cannot die of breast cancer and endometrial cancer at the same time) is negligible and may be omitted (24). In most cases, the results under these assumptions do not significantly differ from the results when these assumptions are not taken into account. For details on differences in derivations of the model under different assumptions of a condition of independence, the reader is referred to reference (24).

Now, from the equations above, we derive the following:

$$p_t * [M - Mrx] = H$$

$$p_t = H / [M - Mrx] = H / (E * M) = H / [Srx - S] = H / ARD,$$

where, in those with disease, Srx is the disease-specific survival in those treated and S

is the survival in those not treated. Finally, since $NNT = 1 / ARD$, we have:

$$p_t = H * NNT$$

Comparison of an Intervention to placebo

If we assume that a patient is taking placebo, net harm is equal to

$$(1 - H_{active\ treatment}) - (1 - H_{placebo}) = H_{active\ treatment} - H_{placebo}, \text{ and since}$$

$$NNH = 1 / (H_{active\ treatment} - H_{placebo}) :$$

$$p_t = NNT / NNH$$

Since p_t must be less than or equal to one, it also follows

$$(H_{treatment} - H_{placebo}) * NNT \leq 1$$

or
$$(H_{treatment} - H_{placebo}) \leq 1 / NNT = M - M_{rx} = ARD$$

and
$$NNT \leq NNH$$

Each of these inequalities may have its specific applicability, depending upon the type of a clinical situation.

Comparison of One Treatment with Another

Following the same steps as illustrated above, the threshold probability of disease or relapse (p_t) at which the expected value of treatment 1 equals the expected value of treatment 2 is shown to be(20):

$$p_t = (H_1 - H_2) / (M * (E_1 - E_2)) = (H_1 - H_2) / (M_{Rx2} - M_{Rx1})$$

$$= (H_1 - H_2) / (S_{Rx1} - S_{Rx2}) = NNT * (H_1 - H_2) = NNT / NNH$$

where H_1 and H_2 represent the harm inflicted with treatments 1 and 2 respectively; E_1 and E_2 represent the effectiveness of treatments 1 and 2, generally expressed as the relative risk reduction; M represents morbidity/mortality without treatment, while M_{Rx1} and M_{Rx2} represent morbidity/mortality with treatments 1 and 2, respectively and p_t is the disease probability representing treatment action threshold. S_{Rx1} and S_{Rx2} refer to disease-specific survival with treatment 1 and 2, respectively. This formulation is particularly appealing if data are expressed in terms of survival, which is commonly the case in oncology practice. Note that in this formulation $NNT=1/(M_{Rx2}-M_{Rx1})=1/(S_{Rx1}-S_{Rx2})=1/ARD$, and $NNH=1/(H_1-H_2)=1/AHD$.

The following inequalities can also be derived from the threshold equations above (20): Treatment 1 is favored if either of the following inequalities holds:

$$E_1 > E_2 + (H_1 - H_2) / M$$

$$S_{Rx1} > S_{Rx2} + (H_1 - H_2)$$

Choice between withholding treatment, testing or treating without testing

The text also provides a solution for a choice between withholding treatment, treating without testing, or performing a test that will determine the further action(14).

The analytical solution of this tree provides two probabilities: the probability of a disease at which we should be indifferent between testing and withholding treatment (p_{tt}) and the probability of a disease at which we should be indifferent between testing and treatment (p_{rx}).

The solution of the decision tree in Fig 2 follows the same procedure illustrated above. Alternatively, the formulas shown in the text may be derived simply replacing the net benefit and net harm in the original Pauker and Kassirer model (14) with the evidence-based therapeutic summary measures shown above.

Integration of patient's preferences within threshold model

If we assume that a patient may expresses certain value judgements toward target events (morbidity/mortality without treatment) ($q_{\text{target}}=1 - \text{value of experiencing target event} = \text{value of avoiding target event}$)(8), and adverse events of the treatments ($q_{\text{adverse event(AE)}}=1 - \text{value of experiencing adverse event} = \text{value of avoiding adverse event}$), the threshold expression for the case of comparing treatment 1 vs treatment 2 can be defined as:

$$p * [1 - q_{\text{target}} * M * (1 - E1) - q_{\text{AE}} * H1] + (1 - p) * [1 - q_{\text{AE}} * H1] = p * [1 - q_{\text{target}} * M * (1 - E2) - q_{\text{AE}} * H2] + (1 - p) * [1 - q_{\text{AE}} * H2]$$

Solving this equation for p we obtain

$$p_t = \text{NNT} * [q_{\text{AE}} / q_{\text{target}} * (H1 - H2)] = \text{RV} * \text{NNT} / \text{NNH} = \text{NNT} * [\text{RV} / \text{NNH}]$$

where RV is the relative value of adverse events relative to target events $\text{RV} = q_{\text{AE}} / q_{\text{target}}$ (8). Note that this expression assumes only one adverse event which is identical in both treatments arms, but which may occur at different frequencies (e.g. life-threatening bleeding with aspirin vs. warfarin). However, if we assume that a patient may experience more than one adverse event (8) (e.g. treatment 1 and 2 can inflict two types of adverse events) the above threshold equation can be formulated as:

$$p * [1 - q_{target} * M * (1 - E1) - q_{AE1} * H1 - q_{AE2} * H2] + (1 - p) * [1 - q_{AE1} * H1 - q_{AE2} * H2] = p * [1 - q_{target} * M * (1 - E2) - q_{AE1} * H3 - q_{AE2} * H4] + (1 - p) * [1 - q_{AE1} * H3 - q_{AE2} * H4]$$

Again, using same procedure as above, we find the threshold at which the expected value of treatment 1 is equal to the expected value of value of treatment 2:

$$p_t = NNT * [(RV1 * H1 + RV2 * H2) - (RV1 * H3 + RV2 * H4)]$$

$$= NNT * [RV1 * (H1 - H3) + RV2 * (H2 - H4)] = NNT * [RV1 / NNH1 + RV2 / NNH2]$$

or, in general form, for n adverse effects, this equation can be expressed as:

$$p_t = NNT * [RV1 / NNH1 + RV2 / NNH2 + RV3 / NNH3 + \dots + RVn / NNHn]$$

The reader is also referred to references (20) and (24) for further technical details of derivations shown in this paper.

Legends:

Fig 1. Integration of evidence-based therapeutic summary measures within a decision analytic model. Choice between management strategy 1 (treatment 1) vs. management strategy 2 with possible clinical events (probabilities) and outcomes (relative values or utilities) associated with these events are shown. Management strategy 2 relates to a no-treatment strategy, placebo and treatment 2, respectively. For example, in the first model if treatment is selected, it can be administered to those patients with or without disease. If it is given to the patient with the disease, outcome will be determined by effect of treatment on the disease (Mrx) and its adverse effects (H). Note how evidence-based therapeutic summary measures affect outcomes associated with different management strategies and how these effects vary as assumptions of the model vary. See Appendix and the text for details on the analytical solution of the decision tree. Abbreviations: E-treatment efficacy; Mrx-morbidity/mortality with the treatment; M-morbidity/mortality without treatment; H or Hrx-harm of the treatment; Hpl-adverse effects on placebo; p-the probability of disease.

Fig 2. Integration of evidence-based therapeutic summary measures within a decision analytic model. In this clinical situation, the choice is between administering treatment, performing a diagnostic test and withholding therapy. Abbreviations: E-treatment efficacy; Mrx-morbidity/mortality with the treatment; M-morbidity/mortality without treatment; Hrx-harm of the treatment; Hte-harm associated with performing a diagnostic test; Sp-test specificity; S-test sensitivity; p-the probability of disease (see Appendix, Fig 1 and the text for details). Adapted from Pauker and Kasirer (14)

Acknowledgment

We are particularly grateful to Drs. Steven H. Woolf, Yael Cohen and Terence Hadley for their detailed critique of this manuscript and providing us with numerous suggestions that help significantly improved the value of this work.

References

1. **Sackett DL, Haynes RB.** On the need for evidence-based medicine. *Evidence-based medicine.* 1995;1:5-6.
2. **Anonymous.** Evidence-based medicine. Glossary. Terms used in therapeutics. *Evidence-based medicine.* 1998;3:33.
3. **Richardson WS, Wilson MC, Nishikawa J, Hayward RSA.** The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club.* 1995;Nov/Dec:A12-A13.
4. **Rosenberg W, Donald A.** Evidence-based medicine: an approach to clinical problem solving. *BMJ.* 1995;310:1122-1126.
5. **Sackett DL, Richardson WS, Rosenberg W, Haynes BR.** *Evidence-based medicine. How to practice and teach EBM.* New York: Churchill Livingstone Inc; 1997.
6. **McQuay HJ, Moore AR.** Using numerical results from systemic reviews in clinical practice. *Ann Intern Med.* 1997;126:712-720.
7. **Djulbegovic B, Hozo I.** Linking Evidence-based Medicine to Clinical Decision Analysis. *Med Decision Making.* 1998;18:464 (abstract).
8. **Guyatt GH, Sinclair J, Cook DJ, Glasziou P.** Users' guides to the medical literature. XVI. How to use a treatment recommendation. *JAMA.* 1999;281:1836-1843.
9. **Laupacis A, Sackett DL, Roberts RS.** An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med.* 1988;318:1728-

- 1733.
10. **Feinstein AR, Horwitz RI.** Problems in the "evidence" of "evidence-based medicine. *Am J Med.* 1997;103:529-535.
 11. **Steiner JF.** Talking about treatment: the language of populations and the language of individuals. *Ann Intern Med.* 1999;130:618-622.
 12. **Friedland DJ, Go AS, Davoren JB, et al.** *Evidence-based medicine. A framework for clinical practice.* Stamford, CT: Appleton & Lange; 1998.
 13. **Kassirer J, Pauker S.** The toss-up. *N Engl J Med.* 1981;305:1467-9.
 14. **Pauker SG, Kassirer J.** The threshold approach to clinical decision making. *N Engl J Med.* 1980;302:1109 - 1117.
 15. **Weinstein M, Fineberg H, Elstein A, et al.** *Clinical Decision Analysis.* Philadelphia: WB Saunders Co; 1980.
 16. **Zarin D, Pauker SG.** Decision analysis as a basis for medical decision making:the three of Hippocrates. *J Med Philosophy.* 1984;9:181-213.
 17. **Pauker S, Kassirer J.** Therapeutic decision making: a cost benefit analysis. *N Engl J Med.* 1975;293:229 -234.
 18. **Bell DE, Raiffa H, Tversky A.** *Decision making. Descriptive, normative, and prescriptive interactions.* Cambridge: Cambridge University Press; 1988.
 19. **Sox HC, Blatt MA, Higgins MC, Marton MC.** *Medical Decision Making* Boston: Butterworths; 1988.
 20. **Djulbegovic B, Hozo I, Fields K, Sullivan D.** High-dose chemotherapy in the adjuvant treatment of breast cancer:benefit risk analysis. *Cancer Control.*

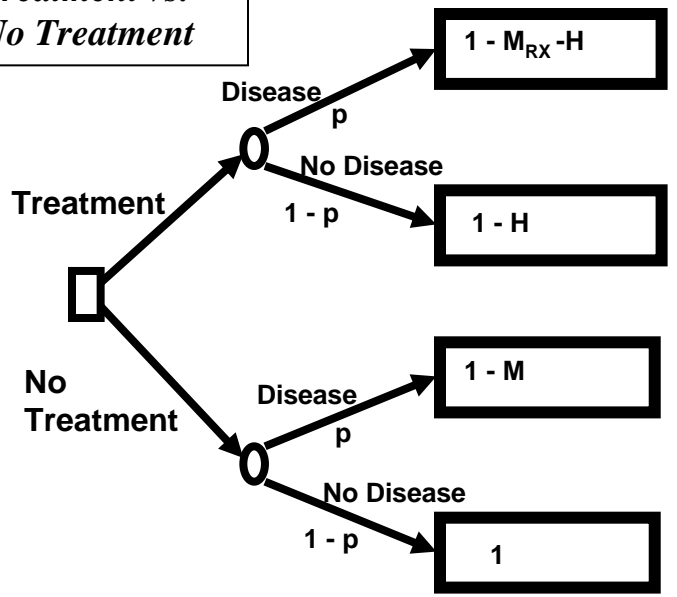
- 1998;5:394-405.
21. **Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS.** Primer on decision analysis:part 3-estimating probabilities and utilities. *Med Decis Making.* 1997;17:136-141.
 22. **Riegelamn R, Schroth WS.** Adjusting the number needed to treat: incorporating adjustment for the utility and timing of benefits and harms. *Med Decision Making.* 1993;13:247-252.
 23. **Detsky AS, Krahn MD, Redelmeier DA, Naimark D.** Primer on decision analysis:part 2-building a tree. *Med Decision Making.* 1997;17:126-135.
 24. **Hozo I, Djulbegovic B.** Using Internet to calculate clinical action thresholds. *Comp Biomed Res.* 1999;32:168-185.
 25. **Dowie J.** 'Evidence-based', 'cost-effective' and 'preference-driven' medicine: decision analysis based medical decision making is the pre-requisite. *J Health Serv Res Policy.* 1996;1:104-113.
 26. **Dowie J.** The 'number to treat' and the 'adjusted NNT' in health care decision-making. *J Health Serv Res Policy.* 1998;3:44-49.
 27. **Kearon C, Gent M, Hirsh J, et al.** A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. *N Engl J Med.* 1999;340:901-7.
 28. **Glasziou P, Guyatt GH, Dans AL, Dans LF, Strauss S, Sackett DL.** Applying the results of trials and systematic reviews to individual patients. *Evidence-based medicine.* 1998;3:165-166.

29. **McMahan Da, Smith DM, Carey MA, Zhou XH.** Risk of major hemorrhage for outpatients treated with warfarin. *J Gen Intern Med.* 1998;13:311-6.
30. **Early Breast Cancer Trialists Collaborative Group.** Polychemotherapy for early breast cancer: an overview of the randomized trials. *Lancet.* 1998;352:930-42.
31. **Colleoni M, Price KN, Castiglione-Gertsch M, Gelber RD, Coates AS, Goldhirsch A.** Mortality during adjuvant treatment of early breast cancer with cyclophosphamide, methotrexate, and fluorouracil. *Lancet.* 1999;354:130-131.
32. **Ezekowitz MD, Levine JA.** Preventing stroke in patients with atrial fibrillation. *JAMA.* 1999;281:1830-1835.
33. **Durington PN, Prais H, Bhatnagar D, et al.** Indications for cholesterol-lowering medications: comparison of risk-assessment method. *Lancet.* 1999;353:278-81.
34. **Ramsey LE, Wallis EJ, Haq IU, et al.** Coronary risk methods and cholesterol lowering. *Lancet.* 1999;353:1095-1098.
35. **Shepherd J, Cobbe SM, Ford I, Group. atWoSCPS.** Prevention of coronary heart disease with pravastatin in men with hypercholesterlaemia. *N Engl J Med.* 1995;333:1301-07.
36. **Grover S.** Gambling with cardiovascular risk: picking the winners and the losers. *Lancet.* 1999;353:254.
37. **Glasziou PP, Irwig LM.** An evidence based approach to individualising treatment. *BMJ.* 1995;311:1356-9.
38. **Anderson KM, Wilson PWF, Odell PM, Kannel WB.** An updated coronary risk

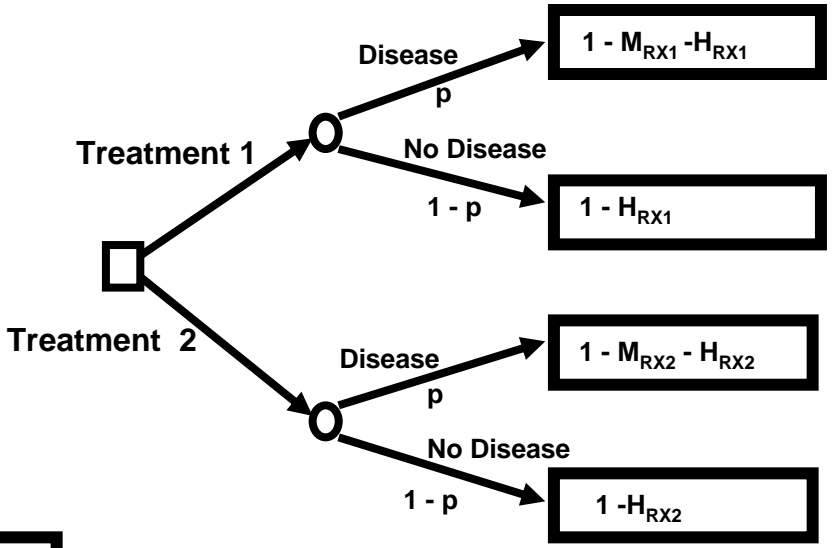
- profile. A statement for health professionals. *Circulation*. 1991;83:356-362.
39. **Djulbegovic B, Morris L, Lyman GH.** Evidentiary challenges to evidence-based medicine. *J Eval Clin Practice*. 2000:in press.
 40. **Sackett D.** A science for the art of consensus. *J Natl Cancer Inst*. 1997;89:1003-1005.
 41. **Sackett DL.** Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. 1989;95(2 supplement):2S-4S.
 42. **Charles C, Gafni A, Whelan T.** Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Soc Sci Med*. 1997;44:681-692.
 43. **Charles C, Whelan T, Gafni A.** What do we mean by partnership in making decision about treatment? *BMJ*. 1999;319:780-2.
 44. **Willan AR, O'Brien BJ, Cook DJ.** Benefit-risk ratios in the assessment of the clinical evidence of a new therapy. *Controlled Clin trials*. 1997;18:121-130.
 45. **Schulzer M, Mancini GBJ.** 'Unqualified success' and 'unmitigated failure': number-needed-to-treat-related concepts for assessing treatment efficacy in the presence of treatment-induced-adverse events. *Int J Epidemiol*. 1996;25:704-712.
 46. **Altman DG.** Confidence intervals for the number needed to treat. *BMJ*. 1998;317:1309-12.
 47. **Feinstein A.** Invidious comparisons and unmet clinical challenges. *Am J Med*. 1992;92:117-120.

48. **Forrow L, Taylor WC, Arnold RM.** Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med.* 1992;92:121-124.
49. **Schwartz L, Woloshin S, Black W, Welch H.** The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med.* 1997;127:966-72.
50. **Weisntein MC.** High-priced technology can be good value for money. *Ann Intern Med.* 1999;130:857-858.
51. **Djulbegovic B, Loughran TP, Hornung C, et al.** The quality of medical evidence in hematology-oncology. *Am J Med.* 1999;106:198-205.
52. **Sonnenberg FA, Beck RJ.** Markov models in medical decision making: a practical guide. *Med Decis Making.* 1993;13:322-338.

Treatment vs. No Treatment



Treatment 1 vs. Treatment 2



Treatment vs. Placebo

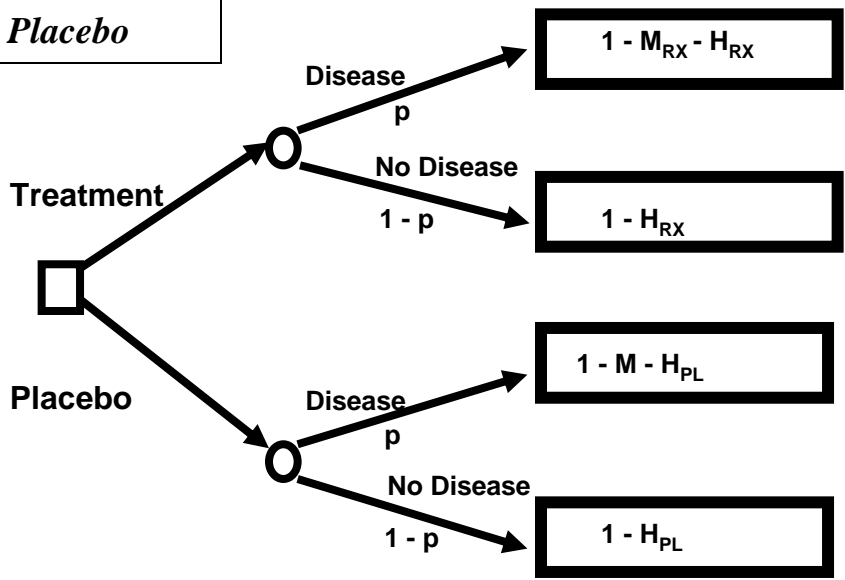


FIGURE 1

