

Debate

Outcomes research in the development and evaluation of practice guidelines

Louise Pilote ^{*1} and Ira B Tager ²

Address: ¹Division of Clinical Epidemiology, Research Institute of the McGill University Health Center, Montreal, Canada and ²Division of Public Health Biology & Epidemiology, School of Public Health, University of California at Berkeley, Berkeley, USA

E-mail: Louise Pilote* - louise.pilote@mcgill.ca; Ira B Tager - ibt@uclink.berkeley.edu

*Corresponding author

Published: 25 March 2002

Received: 1 October 2001

BMC Health Services Research 2002, **2**:7

Accepted: 25 March 2002

This article is available from: <http://www.biomedcentral.com/1472-6963/2/7>

© 2002 Pilote and Tager; licensee BioMed Central Ltd. Verbatim copying and redistribution of this article are permitted in any medium for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Practice guidelines have been developed in response to the observation that variations exist in clinical medicine that are not related to variations in the clinical presentation and severity of the disease. Despite their widespread use, however, practice guideline evaluation lacks a rigorous scientific methodology to support its development and application.

Discussion: Firstly, we review the major epidemiological foundations of practice guideline development. Secondly, we propose a chronic disease epidemiological model in which practice patterns are viewed as the exposure and outcomes of interest such as quality or cost are viewed as the disease. Sources of selection, information, confounding and temporal trend bias are identified and discussed.

Summary: The proposed methodological framework for outcomes research to evaluate practice guidelines reflects the selection, information and confounding biases inherent in its observational nature which must be accounted for in both the design and the analysis phases of any outcomes research study.

Background

The development of practice guidelines

In clinical medicine, variations exist that do not appear to be related to variations in the clinical presentation and severity of disease [1–3]. In response, practice guidelines have been developed in an attempt to reduce the wide practice variations and, through this process, to increase the appropriateness and quality of medical care and to reduce health care costs [4–8].

Despite the publication and dissemination of practice guidelines [9], there has been relatively little evaluation of the application and impact of clinical practice guidelines

[9–14]. Some of the difficulty in the evaluation of these guidelines relates to the methods that were used to develop them [9]. Guidelines have often have been developed before adequate data have been available to assess the relationship between clinical practice patterns and desired clinical outcomes. Nevertheless, there have been some reviews of practice guideline evaluation [15,16].

While epidemiological designs are commonly used to evaluate the effectiveness of health care interventions, never has this been discussed in the context of outcomes research. We propose the use of a methodological frame-

work for outcomes research to evaluate practice guidelines.

Methodological issues with the measurement of practice variations

In the debate about reasons to promote the development of practice guidelines, few have questioned whether the variations are real, or alternatively, whether they are simply a function of methodological flaws in the measurement of medical practices themselves, the result of variations in practice patterns across groups of patients with a similar diagnosis, or both. Furthermore, few studies have addressed whether practice variations, in fact, lead to outcome variations. Finally, little attention has been paid to the identification and measurement of initial conditions, that is, the potentially confounding factors and effect modifiers of the practice patterns outcomes relationship.

Measurement of practice pattern variation

The measurement of medical practice patterns is susceptible to error. Measurement error may affect the validity of medical practice measurement in three major ways. First, it may lead to selection bias, in that subjects are selected to belong to a certain group based on an erroneous diagnosis. Secondly, it may lead to misclassification of exposure (information bias), in that patients treated with a specific practice pattern are classified in the wrong diagnostic group. Thirdly, it may lead to misclassification of outcomes, in that patients with a given outcome are classified in the wrong diagnostic group.

Potential problems with the measurement of practice variations relate to the mechanisms that underlie the choice of groups that are compared in studies of practice variations. These mechanisms must be defined clearly to minimize selection bias. In many studies of practice variations, populations are arbitrarily divided according to hospitals, regions, counties, or countries. Little information is available about the factors that lead these groups to go to a particular hospital, live in a particular region, go to a particular doctor, etc. The population base

from which each comparison group is derived should, in principle, be quite similar for all groups. Basically, if the groups are drawn from a similar population, unmeasurable and potentially confounding variables are more likely to be equally distributed between groups.

In addition, the measurement of practice variations cannot be valid without information on relevant "initial conditions". Initial conditions are all confounding factors and effect modifiers, other than the treatment/practice patterns, that may cause or influence the clinical outcomes of interest. These factors may explain practice variations among groups that do not share similar initial conditions. To evaluate practice patterns-outcomes associations, potential confounders must be identified and controlled for in the analysis.

Aside from clinical presentation and severity of illness; the initial conditions to be identified and characterized as completely as possible include physician, patient, and practice environment factors (Table 1). Measurement of such factors is essential to minimize the chance of a systematic error following confounding biases and effect modification (Figure 1).

Identification and measurement of outcomes of interest

Limitations to the development and evaluation of practice guidelines also include the absence of a clear concept of the targeted outcomes and the paucity of outcomes data to support these guidelines [17]. There appears to be only a weak relationship between the purpose of guidelines and many of the outcomes usually measured in clinical research, that is, the source of evidence for guideline development (evidence-based). The initial goals of establishing practice guidelines – to reduce costs and enhance the quality and appropriateness of treatment – are, in fact, rarely the basis for guideline development, since little data is available for these outcomes. To some degree, the development of guidelines has been driven by the availability of data on clinical outcomes, such as morbidity and mortality, rather than those outcomes related to the primary goals of the guidelines.

Table 1: Initial conditions to be taken into account when making inferences about practice patterns-outcomes associations

Patient factors	Physician factors	Environmental factors
Demographics Severity of illness	Physicians' beliefs and interpretation of medical evidence	Medical insurance status (fee-for-service vs. managed care)
Heterogeneity of disease expression	Medical uncertainty	Availability of needed procedure
Patient preference	Specialist vs. generalist Physicians' skills	Proximity of medical centers with appropriate expertise
Patient attitude towards disease	Inappropriate use Nature of group	

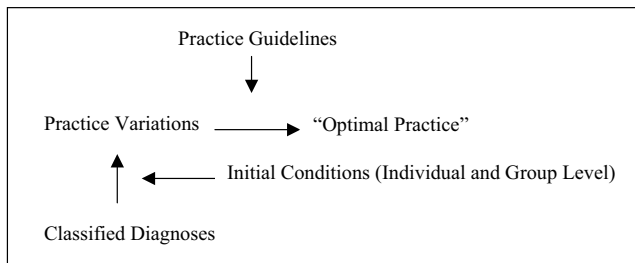


Figure 1

The evaluation of practice guidelines

Throughout the development of practice guidelines, the major deficiency has been the lack of an evaluative method [18–27]. Thus, we suggest a methodological framework for outcomes research to be applied to evaluate practice guidelines. Outcomes research evaluates practice patterns as they occur in actual clinical settings. This type of research can describe practice patterns, evaluate their divergence from practice guidelines and determine the effect of practice variations on outcomes. Outcomes research is necessarily observational in nature and, although observational studies have been used to evaluate health care interventions, the proposed methodological framework has yet to be applied to outcomes research.

Why should outcomes research be used to evaluate and validate practice guidelines? The primary goal of practice guidelines is the consistent adherence by physicians to practice patterns that achieve the "best" outcomes at the lowest cost. Outcomes research evaluates practice patterns as they occur in actual clinical settings, and is thus the logical method to evaluate practice guidelines. In fact, outcomes research and practice guidelines are connected through concepts that relate to efficacy and effectiveness research (Figure 2). Efficacy studies, which normally complement practice guideline development, are those performed in highly selected groups of patients to investigate if a particular intervention works under controlled conditions set by the study investigators. In contrast, outcomes research evaluates practice as it occurs in actual clinical settings [28]. Research in these settings is called effectiveness research because the investigators have limited control over the conditions that qualify the practice settings. The difference between efficacy and effectiveness research can be summarized as follows: does it work at all (efficacy) or does it work in the real world (effectiveness)? Thus, there exists a dynamic process in which evidence from both effectiveness and efficacy studies feeds into the development and evaluation of practice guidelines, as depicted in Figure 2.

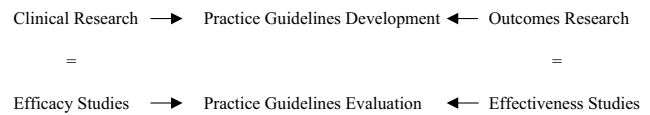


Figure 2
Relationship between outcomes research and practice guidelines

Most practice guidelines are derived from efficacy studies rather than effectiveness studies. Therefore, it is not surprising that practice guidelines are not fully applicable in actual clinical practice. We suggest that effectiveness studies be used not only as a method to evaluate practice guidelines but also as a basis for their development. These could include both observational studies and effectiveness trials. Outcomes research better reflects practice in the real world and may make guidelines more likely to be applied. However, to date, little attention has been paid to the epidemiological underpinnings of the methods used to conduct outcomes research.

Discussion

We will first propose a methodological framework for outcomes research. Then, we will show how it can be used to evaluate practice guidelines. Finally, we will address the limitations of the proposed methodological framework.

Generic epidemiological issues in outcomes research

In the proposed methodological framework, the generic issues related to outcomes research will be discussed in sequential order. In outcomes research, the first step is to identify the study population and the groups (hospitals, providers, regions, etc.) that will be compared. The next step is the measurement of practice patterns and outcomes. After groups are compared on the basis of the treatment they receive and outcomes of interest, associations are sought between practice patterns and the various measures of outcome. This step of the methodological framework raises issues of confounding bias because not all factors that can confound these associations are measured and controlled or even known. The presence or absence of confounding bias can be affected by the other sources of bias namely selection and information biases. Lastly, we discuss the issue of temporal trends. In the evaluation of practice guidelines, the measurement of practice patterns may not be contemporaneous with the publication of practice guidelines. This may explain and even lead to the frequently observed discrepancy between the actual practice and what the guidelines state that it should be. Finally, two particularities of outcomes research 1) the presence of ecological exposures in individual level studies and 2) the common use of large administrative databases are discussed.

Table 2: Epidemiological model for outcomes research to evaluate practice guidelines

Outcome =	Exposure +	Confounders +	Error
<ul style="list-style-type: none"> • Clinical outcomes (for example, disease development, complications) • Quality • Cost • Appropriateness 	<ul style="list-style-type: none"> • Practice patterns 	<ul style="list-style-type: none"> • Extraneous factors that can influence practice patterns and outcomes, that is, initial conditions 	<ul style="list-style-type: none"> • Misclassification of diagnosis, practice patterns and outcomes • Error in measurement of practice patterns and outcomes • Unmeasured factors

Specification of the model

Definition of the elements of the proposed epidemiological model for outcomes research

In the proposed model for outcomes research designed to evaluate practice guidelines, the outcome of interest can be a disease (Table 2). For example, if the practice patterns that are being studied pertain to coronary revascularization, complications such as mortality and reinfarction after acute myocardial infarction may constitute the outcome of interest. Finally, the consequences of different practice patterns on medical resources (cost, quality and appropriateness) may be another possible outcome of interest.

In the studies of outcome research, practice patterns, (which constitute the exposure in the proposed model), range from the use of medication, diagnostic tests and therapeutic procedures to the length of hospital stay, transfer to other facilities and/or scheduled physicians visits. The primary goal of outcomes research is the evaluation of the effects of the selected practice patterns on the outcomes of interest. Consequently, any inference made about this association must be evaluated as a function of the potential selection, information (measurement error) and confounding biases. A limitation of outcomes research as it is most often performed is the lack of attention given to the measurement of each of the elements of the epidemiological model shown in Table 3. The basis of the proposed methodological framework will be the identification of generic sources of potential bias that relate to each element of the proposed model.

Selection bias

Since outcomes research is observational in nature, the choice of the study population and of the compared groups is highly susceptible to selection bias. As applied to outcomes research, selection bias is defined as a distortion in the estimate of the practice patterns outcomes association due to the way that subjects are selected for inclusion in the study population and in the different groups to be

compared [29]. A major consequence of selection bias is the potential confounding of inferences made about practice patterns-outcomes associations. This occurs when some characteristics of the subjects related to practice patterns or clinical outcomes influence the selection or exclusion of individual subjects, groups of subjects or practice environments.

The selection process should be such that patients included in the study population come from the same target population [30]. Furthermore, patients or study members should have a similar probability of being selected and included in the actual population. Inclusion and exclusion criteria must be clearly defined in order to characterize the actual population as precisely as possible. Judging the internal validity of a study is more feasible when there is a detailed account of how the individuals were selected to become members of the actual population. Finally, the study population, also needs to be carefully characterized so that the inferences derived from the analysis of the study population can be evaluated for both internal validity (based on the data analyzed in the study) and external validity (the extent to which results obtained from the data analyzed in a particular study can be generalized to populations outside of the study). Any systematic differences between those actually studied and the source (target) population could result in biased estimates of the impact of a practice pattern on a clinical outcome.

In many studies of outcomes research, groups exposed to different practice patterns are compared. The identification of such groups of patients is sought to assess the impact of different practice patterns on various outcomes in actual clinical settings and, as previously mentioned, can be used to assess practice guidelines. Because of such study design, it becomes unclear as to what the target population precisely is. Is it the group (the set of patients in a given environment) or is it the individuals receiving the various practice patterns within each group? For example, in a study of regional variations in the treatment of acute

myocardial infarction in the U.S., the treatment of patients (practice patterns) was compared across different regions of the U.S. In this study, one wishes to generalize the findings about practice patterns-outcomes associations to all individuals with acute myocardial infarction (individual level). One also wishes to generalize the effect of the exposure, which is in this case practice patterns, to those prevalent in a given region (ecological level).

The presence of these two levels, the individual and the ecological levels, introduces an added level of complexity in terms of the assessment of the effect of the exposure on outcome. When comparing practice patterns across regions using individual data, there is a certain degree of correlation brought about by the clustering of practice patterns that needs to be taken into account. Such a correlation is very difficult to quantify. In contrast, when assessing the effect of the exposure at the individual level, there are ecological factors (initial conditions particular to a given region) that need to be taken into account. The data originating from studies with mixed design, which are often the design of outcomes research studies, need to be analyzed with special attention to the degree of correlation between the individual covariates and to the presence of ecological exposure variables.

Another potential source of selection bias is the choice of the groups to be compared, which depends on the criteria used to divide the groups. Individuals included in the groups to be compared should have the same probability of being included in these groups. Not infrequently in outcomes research, geographic criteria (such as country, regions, hospitals) are used because such criteria allow the identification of clinically comparable groups that receive very different treatments, whose resulting outcomes can then be assessed. However, such a process must be scrutinized for the possibility of selection bias other than the treatments that are being evaluated. Such selection bias would make groups not comparable as to clinical and other factors that could affect outcomes.

The presence of a biased selection process could lead to confounding bias when practice patterns-outcomes associations are assessed. Such a situation may occur when the study groups are not comparable with regard to some characteristics of the subjects related to practice patterns or clinical outcomes that influenced the selection or exclusion of individual subjects, groups of subjects or practice environments. For example, in the same study of regional variations in the treatment of acute myocardial infarction, census regions of the U.S. were arbitrarily chosen as a basis for comparison. In this example, patients with similar risk of developing the outcome of interest, which is defined here as a complication after acute myocardial infarction, may not have had the same probability of being

included in the different groups to be compared. Confounders may then bias the practice patterns/outcomes association if the selection of different risk groups is related to practice patterns.

Selection bias can also affect the assessment of outcomes. Potential sources of this bias include loss to follow-up or missing data. Follow-up data is difficult to obtain in outcomes research studies, which often rely on administrative databases for data acquisition. Linkage, either of different databases or of the same database over time, is often performed [31]. A failure to link the databases for a number of individuals presents a problem equivalent to having data missing for these individuals.

Information bias

The second step in outcomes research studies is the measurement of practice patterns and of the outcomes of interest. Here, issues of information bias must be considered. Information bias can be defined as a distortion of the potential practice patterns outcomes association due to misclassification of subjects with regard to practice patterns, outcome measures or both, or due to measurement error [29].

There are two major ways in which practice patterns can be misclassified. They relate to the sensitivity and specificity of the tests that are used for the diagnosis for which practice patterns are being evaluated and for the classification of the outcomes of interest. The measurement of the different practice patterns and their related outcomes largely depend on the identification of a group of patients who have a given diagnosis and require a given treatment. The characteristics that make a diagnosis more amenable to outcomes research are the following: 1) a precise diagnostic definition, 2) a diagnostic test with high sensitivity and specificity, 3) reproducibility among different individuals and locations, 4) easily coded, 5) related to a procedure, and 6) common and costly, so that it is likely to be collected in large, administrative databases frequently used in outcomes research. Because of such requirements, only a limited number of clinical conditions are amenable to outcomes research. Acute myocardial infarction is an example of a diagnosis that can be made with a high level of certainty because it has a precise diagnostic definition and well-defined diagnostic criteria, which, when taken together, have high sensitivity and specificity for the correct classification of patients. Therefore, it is easy to identify a study population that, in fact, has this disease and to describe their treatment. Thus, in order to minimize the misclassification of relevant practice patterns, the methods used to classify the disease and the outcomes that relate to the practice patterns under investigation must have high sensitivity and specificity [29,31,32].

Given the principles underlying the measurement of practice patterns and outcomes, how are the measurements generally made in outcomes research studies? The measurement of the exposure (practice patterns) in outcomes research is valid only if it corresponds to the "true" practice as performed in the clinical setting. Again, practice can only be "true" if the diagnosis is correct. The identification of both patients with the disease of interest and their treatment requires a source of information that has the features of a diagnostic test.

In outcomes research, administrative databases are often used as an information source to identify a study population and to obtain data on exposure. The database coding of diagnoses and procedures can be used as a "diagnostic test" to identify the clinical condition for which practice patterns will be described and to classify the practice patterns themselves and the outcomes of interest. Such a "diagnostic test" will have higher sensitivity and specificity values for some diagnoses than for others.

For example, administrative database coding will have higher sensitivity and specificity for procedure-related diagnoses (such as hip fracture) because the diagnostic code is related to a major operation and is likely to be recorded for administrative purposes. In contrast, a diagnostic criterion for osteoarthritis can be quite vague and administrative coding is likely to have very low sensitivity and specificity for this diagnosis.

The use of databases as a diagnostic test must be validated in all outcomes research studies, especially those using administrative databases. Methods to validate these databases include chart reviews, *a priori* coding systems or both. These validation methods ensure that coding is as accurate and reproducible as possible, thus allowing the database to be used as a diagnostic test to identify the study population and the practice patterns and the outcomes in outcomes research. However, these validation methods are rarely used.

Finally, appropriate measures of outcomes that will serve to evaluate practice guidelines must be identified. This presents a problem because most practice guidelines aim to reduce practice variations, which will, in turn, lead to improved appropriateness and quality of care. However, how appropriateness and quality of care are measured is controversial and will not be discussed here [33–97]. Nevertheless, defining the outcomes that will be used to evaluate practice guidelines is a crucial step in this process.

Quality of life and functional status measures constitute another group of outcome measures that should be included for the evaluation of practice guidelines. These dimensions of outcomes have received more attention from

health providers, while consumers have become more concerned about outcomes of care. However, these outcomes also are difficult to measure, because they rely heavily on patient interviews and questionnaires. They are likely to vary with patient expectations, culture, and climate and are thus potentially to be measured with error and be misclassified. A few reliable, valid instruments have been developed to assess health-related quality of life [91,92], but such instruments are not easily used to collect this information from large databases. There is a need to develop instruments to measure these types of outcomes, whether they are conversion factors for existing databases (such as using length of stay as a proxy for cost) or new measures that could easily be integrated in administrative databases. Such measures could include estimates of functional class or severity of illness.

At present, many outcomes research studies measure mortality and disease-specific morbidity. The validity of the measurement of these outcomes is limited by the type of database that is used. For example, using death registries to obtain causes for death is a notoriously invalid source for this type of information. There are many examples of poor correlation between cause of death as established by death registries versus disease registries. Death certificates in New York City during 1992 were assessed to determine the accuracy and frequency of reporting tuberculosis as a cause of death. Of 310 persons who died with active tuberculosis in 1992 (based on a disease-specific registry), only 34% had tuberculosis listed on their death certificate. Thus, in this example, as in many others like it, using death certificates led to an inaccurate measure of disease burden [98].

Confounding bias

In outcomes research terms, confounding bias is present when the effect of the practice variations on the outcomes of interest is distorted because of the effects of extraneous variables (variables that are causally associated with the practice variations and the outcomes of interest) [29]. This issue is crucial in outcomes research because, while outcomes research shares the purpose of a clinical trial (to evaluate different treatments), it primarily uses observational methods – investigators conducting outcomes research have limited control over potentially confounding factors (the initial conditions of individual groups of patients). Because outcomes research builds on existing practice variations and analyses the natural ongoing experiment, there is ample opportunity for confounding bias to invalidate any inference made about practice patterns outcome associations [99]. For example, variations in practice patterns could reflect variation not only in the use of a given procedure but also in the severity of disease. Assignment of patients to certain procedures on the basis of the severity of illness makes sense clinically, but in out-

comes research, it is a common and important source of confounding if the procedure is either efficacious or particularly harmful in high-risk patients. Many indices have been developed to measure the severity of illness when using existing databases to correct for such confounding, but one can never be sure that this type of confounding has been entirely controlled [100,101]. This presents an intrinsic limitation of outcomes research.

Avoidance of confounding bias is limited by the source of data used to describe practice patterns, particularly when observational data, such as the large Medicare administrative databases, are used to compare outcomes among patients who receive different treatments. The potential for confounding bias arises because many factors other than the treatment under evaluation may affect patient outcomes. These factors include comorbid diseases, severity of illness, and patient, physician and environmental factors. Such factors are likely to influence treatment decisions but are difficult to capture fully in recorded data. Researchers cannot adjust for imbalances in prognostic factors that are unmeasured or poorly categorized and administrative data, in particular, may lack the precise and accurate coverage of clinical details needed to permit full and fair adjustments. Further data collection might solve this issue, but it is not always possible to collect additional information. Standard statistical modeling can attempt to adjust for the known differences between the groups, but this might not be sufficient for unmeasured differences.

Several alternative methods have been suggested. One method is subgroup analysis [102] to adjust for unmeasured differences between groups of individuals who differ on known risk factors. Another method consists of the use of instrumental variables [103,104]. Instrumental variables are observable factors that influence treatments but do not directly affect patient outcomes. This approach uses the so-called instrumental variables to mimic a randomization of patients to different likelihoods of receiving alternative treatments. McClellan *et al.* [103] applied this methodology to assess whether more aggressive use of invasive cardiac procedures improved outcomes in the elderly. In this study, the instrumental variable was the distance of the patient's residence from the nearest hospital with on-site angiography. The authors noted lower mortality among elderly individuals who received more aggressive treatment than among those treated more conservatively.

Temporal trend bias

We propose a bias called a "temporal trend bias" that is particular to the use of outcomes research to evaluate practice guidelines. This bias results from the inability to control for secular trends. It reflects the fact that by the time practice guidelines are published and disseminated,

new treatments and technology are being incorporated into clinical practice. Thus, it is difficult to identify a pure application of a practice guideline whose application is not undermined by recent advances in medicine and technology. For example, we evaluated the effect of a specific set of guidelines on return to work after acute myocardial infarction. The use of these guidelines had been successful in a university setting; this study assessed their use in a community setting. During the 5 years that elapsed between these two studies, practices changed. The use of guidelines was less successful in the community not only because they did not influence practice but also because usual care had grown closer to the proposed guidelines [105].

Ecological exposure in individual level studies

A frequently encountered particularity of outcomes research study design is the presence of both ecological exposure and individual level covariates in the same analysis. Because the unit of analysis is a group, but inferences are made about the impact of a given practice pattern on individual outcomes, many outcomes research analyses have elements of both individual and ecological analyses [106]. In our study of regional variations in the treatment of acute myocardial infarction, measures describing practice patterns at the regional level, ecological exposure, (proportion of patients receiving angiography, angioplasty, and coronary artery bypass surgery) were linked to the outcome measures of mortality adjusting for individual level variables that measured severity of disease. Then, inferences were made about the use of these procedures at the patient level. Although the unit of analysis is the region, which would demand an ecological analysis, there are individual level covariates, which are likely to be correlated within each region, that need to be taken into account.

When group measures are used that contain individual-level variability with some degree of correlatedness (within region) and aggregate-level variability (between regions), specific analytic tools must be used. It has been suggested that hierarchical logistic regression modeling be used to examine the interplay between sources of variation in the use of health-care services, that is, between ecological-level and individual-level sources. This type of modeling is designed to separate true variability across areas from observed variability. An application of this method is the work by Gatsonis *et al.* [107] who found that practice variations across regions of the U.S. in the use of angiography after acute myocardial infarction were largely explained by differences in patient characteristics and geographic region. However, states that had more on-site availability of angiography still tended to have higher angiography rates after accounting for between-region and within-region variability. After analysis for sources of var-

iability, more reliable inferences about the associations between practice patterns and outcomes can be made.

Sources of data

The application of the proposed methodological framework for outcomes research largely depends on the sources of data that are used to evaluate the effect of the practice variations on outcomes [56]. Most commonly, the study design is a retrospective cohort analysis and the dataset that is used has been obtained either for administrative purposes (discharge databases) or for a randomized clinical trial that addressed a different question [108]. Less often, a prospective cohort study is designed to evaluate a particular set of practice guidelines [109]. Although a prospective design provides more control in data collection than a retrospective analysis, both designs are subject to selection, information and confounding biases.

The ideal database to use for the evaluation of practice guidelines is one that allows the precise measurement of the practice patterns (exposure) and outcomes (disease) as well as the measurement of potential confounders (severity of illness, precision of diagnosis, socioeconomic characteristics). Unfortunately, such a database probably does not exist. The strength of administrative databases, such as that of Medicare is that they allow the observation of large numbers of patients for which practice patterns can be evaluated as they occur in actual clinical practice. Furthermore, administrative databases allow the observation of practice patterns outcomes associations in large numbers of unselected patients.

However, the limitations of such databases include the missing information about potential confounding factors, such as severity of illness, and the limited ability to measure exposure and outcome accurately. Many databases that are not designed for clinical research either mismeasure patient outcomes or fail to capture outcomes that are important to both physicians and patients (such as quality of life and functional status). The control of these biases was the basis of the methodological framework for outcomes research proposed in this chapter.

The application of outcomes research methods to practice guideline evaluation

The application of outcomes research methods to practice guideline evaluation can accomplish several goals. One important goal is the evaluation of practice guidelines, that is, to determine to what extent the guidelines accomplished their primary goals after their dissemination. We have suggested the model of chronic disease epidemiology as the methodological framework for outcomes research to evaluate practice guidelines.

The steps to evaluate practice guidelines using outcomes research when the basic design is a retrospective cohort study are summarized in Figure 3. Some limitations to the application of this model exist. The reasons for the inability of the proposed methodological framework to deal completely with the intrinsic biases in outcomes research are listed in Figure 4. They relate mostly to the databases usually used in studies of outcomes research.

Summary

The proposed methodological framework for outcomes research to evaluate practice guidelines reflects the selection, information and confounding biases inherent in its observational nature which must be accounted for in both the design and the analysis phases of any outcomes research study. Indeed, a major limitation of outcomes research is the inability to account for unobserved heterogeneity that directly correlates with practice patterns and/or health outcomes. This may lend bias to any inferences made about practice variations and outcomes. "Researchers cannot correct for the subtle reason doctors choose one treatment over another for a particular patient. That bias, in turn, can undermine the entire premise of outcomes research" [110]. These are intrinsic properties of outcomes research that can be dealt with only in part, by applying the principles of chronic disease epidemiology. Thus, this proposed methodology can serve as a framework for the conduct of outcomes research in the evaluation of practice guidelines but its application will be limited.

1. Can a large database be identified that contains information on practice patterns for the treatment of a condition for which practice guidelines have been developed?
2. Is the database suitable for guideline evaluation in terms of the following criteria?
 - a. Can a precise diagnosis be made using the available data?
 - b. Can criteria be established to allow for the creation of comparison groups with different practice patterns?
 - c. Are there data to ensure the comparability of the groups?
 - d. Can practice patterns be measured?
 - e. Can practice patterns be identified according to those prescribed by practice guidelines?
 - f. Are there any data on patient, physician, and environmental factors that could explain deviations from practice prescribed by practice guidelines and that could help validate any inference made about practice patterns-outcomes associations?
 - g. Are outcomes of interest related to the purpose of clinical guidelines to enhance the quality, appropriateness, and effectiveness of health care, available and measured with precision?
 - h. Are the incidence rates or prevalence of the outcomes of interest large enough to allow meaningful practice patterns-outcomes associations?

Figure 3
Steps to evaluate practice guidelines using outcomes research

- ## Lack of control over data quality.
- ## Lack of control over what is being collected; lack of data on practice patterns; lack of data on initial conditions; lack of data on all outcomes of interest for guideline evaluation.
- ## Difficulty with the measure of correlated data.
- ## Limited availability of statistical methods to deal with ecological exposures in individual level studies.
- ## Limited number of diagnoses amenable to be studied through this method.
- ## Outcomes research works when natural experiments can be observed. Only for a few conditions, the argument is as follows: treatments are so varied, and the doctor's choice so unpredictable, that database records approximate those obtained from arbitrary assignment in clinical trials.

Figure 4
Reasons for the inability of the proposed methodological framework to deal with biases in outcomes research

Competing interests

none declared

References

- Wennberg JE, Freeman JL, Shelton RM, Bubolz TA: **Hospital use and mortality among Medicare beneficiaries in Boston and New Haven.** *N Engl J Med* 1989, **321**:1168-1173
- Wennberg J, Gittlesohn A: **Small area variations in health care delivery.** *Science* 1973, **182**:1102-1108
- Leape LL, Park RE, Solomon DH, et al: **Does inappropriate use explain small-area variations in the use of health care services?** *JAMA* 1990, **263**:669-672
- Eddy DM: **Three battles to watch in the 1990s.** *JAMA* 1993, **270**:520-526
- Eddy DM: **Practice policies: What are they?** *JAMA* 1990, **263**:877-880
- Eddy DM: **Practice policies: Where do they come from?** *JAMA* 1990, **263**:1267-1272
- Eddy DM: **Practice policies: Guidelines for methods.** *JAMA* 1990, **263**:1839-1841
- Eddy DM: **Guidelines for policy statements: The explicit approach.** *JAMA* 1990, **263**:2239-2243
- Audet A-M, Greenfield S, Field M: **Medical practice guidelines: current activities and future directions.** *Ann Intern Med* 1990, **113**:709-714
- Marwick C: **New health care research agency reflects interest in evaluating quality.** *JAMA* 1990, **263**:929-930
- King SH: **Clinical guideline development.** *Department of Health and Human Services. Rockville, Agency for Health Care Policy and Research* 1990, 1-7
- Fuchs VR: **Managed care and merger mania.** *JAMA* 1997, **277**:920-921
- McGuire LB: **A long run for short jump: Understanding clinical guidelines.** *Ann Intern Med* 1990, **113**:705-708
- Fletcher RH, Fletcher SW, Ed: **Clinical practice guidelines.** *Ann Intern Med* 1990, **113**:645-646
- Grimshaw JM, Russel IT: **Achieving health gain through clinical guidelines II: Ensuring guidelines change medical practice.** *Quality in Health Care* 1994, **3**:45-52
- Mowatt G, Lib D, Grimshaw JM, et al: **Getting evidence into practice: The work of the Cochrane Effective Practice and Organisation of Care group (EPOC).** *The Journal of Continuing Education in the Health Professions* 2001, **21**:55-60
- Ellwood PM: **Shattuck lecture - Outcomes management. A technology of patient experience.** *N Engl J Med* 1988, **318**:1549-1556
- Gifford F F: **Outcomes research and practice guidelines: Upstream issues for downstream users.** *Hastings Centre report* 1996, 38-44
- Tower EC, Ballin S: **Outcomes research and practice guideline development in health care reform.** *Circulation* 1994, **90**:2607-2608
- Roberts RG: **Marriage of practice guidelines and outcomes research.** *Am Family Phys* 1995, **51**:1385-1386
- Ellrodt AG, Conner L, Rieinger M, Weingarten S: **Measuring and improving physician compliance with clinical practice guidelines.** *Ann Intern Med* 1995, **122**:277-282
- Eddy DM: **Comparing benefits and harms: The balance sheet.** *JAMA* 1990, **263**:2497-2505
- Logan RL, Scott PJ: **Uncertainty in clinical practice: Implications for quality and costs of health care.** *Lancet* 1996, **347**:595-598
- Saltman D: **Guidelines: In search of certainty.** *Med J Aust* 1997, **166**:62
- Povar G: **Profiling and performance measures. What are the ethical issues?** *Med* 1995, **33**:S60-JS68
- Hadorn DC, Baker D: **Development of the AHCPH-Sponsored heart failure guideline: Methodologic and procedural issues.** *J Quality Improvement* 1994, **20**:539-554
- Kravitz RL, Laouri M, Kahan JP, Guzy P, Sherman T, Hilborne L, Brook RH: **Validity of criteria used for detecting underuse of coronary revascularization.** *JAMA* 1995, **274**:632-638
- Tanenbaum SJ: **Sounding board: What physicians know.** *N Engl J Med* 1993, **329**:1268-1270
- Kleinbaum DG, Kupper LL, Morgenstern H: **Epidemiologic research. Principles and quantitative methods.** *New York, Van Nostrand Reinhold* 1982, **Chapter 12**:
- Kleinbaum DG, Kupper LL, Morgenstern H: **Epidemiologic research. Principles and quantitative methods.** *New York, Van Nostrand Reinhold* 1982, **Chapter 10**:
- Roos LL, Wald R, Wadja A, Bond R, Hartford K: **Record linkage strategies, out patient procedures and administrative data.** *Medical Care* 1996, **34**:570-582
- Greenland S: **Interpretation and choice of effect measures in epidemiologic analyses.** *Am J Epidemiol* 1987, **125**:761-768
- Chassin MR: **Quality of care.** *N Engl J Med* 1996, **335**:1060-1063
- Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, for the Evidence-Based Medicine Working Group: **Users' guides to the medical literature. A method of grading health care recommendations.** *JAMA* 1995, **274**:1800-1804
- Trobe JD, Fendrick AM: **The effectiveness initiative.** *Arch Ophthalmol* 1995, **113**:715-717
- Feussner JR: **Evidence-based medicine: New priority for an old paradigm.** *J Bone Miner Res* 1996, **11**:877-882
- Brownman GP, Levine MN, Mohide A, Hayward RSA, Pritchard KI, Gafni A, A Laupacis: **The practice guidelines development cycle: A conceptual tool for practice guidelines development and implementation.** *J Clin Oncol* 1995, **13**:502-512
- Resenkov L, Chediak J, Hirsh J, Lewis HD: **Antithrombotic agents in coronary artery disease.** *Chest* 1989, **95**(Suppl):52S-72S
- Pilote L, Racine N, Hlatky MA: **Differences in the treatment of myocardial infarction in the United States and Canada: A comparison of two university hospitals.** *Arch Intern Med* 1994, **154**:1090-1096
- Rouleau JL, Moya LA, Pfeffer MA, et al: **A comparison of management patterns after acute myocardial infarction in Canada and the United States.** *N Engl J Med* 1993, **328**:779-784
- Mark DB, Naylor CD, Hlatky MA, et al: **of medical resources and quality of life after myocardial infarction in Canada and the United States.** *N Engl J Med* 1994, **331**:1130-1135
- Van der Werf F, Topol EJ, Lee KL, et al: **Variations in patient management and outcomes for acute myocardial infarction I the United States and other countries: Results from the GUSTO trial.** *JAMA* 1995, **273**:1586-1591
- The GUSTO Investigators: **An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction.** *N Engl J Med* 1993, **329**:673-682
- The GUSTO Angiographic Investigators: **The effects of tissue plasminogen activator, streptokinase, or both on coronary-artery patency, ventricular function and survival after acute myocardial infarction.** *N Engl J Med* 1993, **329**:1615-1622
- American Hospital Association guide to health care field.** *Chicago: American Hospital Association* 1993
- Neter J, Wasserman W, Kutner MH: **Applied linear statistical models.** *Homewood, Ill: Richard D. Irwin* 1990, 466-470

47. Lee KL, Woodlief LH, Topol EJ, et al: **Predictors of 30-day mortality in the era of reperfusion for acute myocardial infarction: Results from an international trial of 41,021 patients.** *Circulation* 1995, **91**:1659-1668
48. Little RJA: **Regression with missing X's: A review.** *J Am Stat Assoc* 1992, **87**:1227-1237
49. Teo KK, Yusuf S, Furberg CD: **Effects of prophylactic antiarrhythmic drug therapy in acute myocardial infarction: An overview of results from randomized controlled trials.** *JAMA* 1993, **270**:1589-1595
50. Lamas GA, Pfeffer MA, Hamm P, Wertheimer J, Rouleau J-L, Braunwald: **Do the results of randomized clinical trials of cardiovascular drugs influence medical practice?** *N Engl J Med* 1992, **327**:241-247
51. Yusuf S, Sleight P, Held P, McMahon S: **Routine medical management of acute myocardial infarction: Lessons from overviews of recent randomized controlled trials.** *Circulation* 1990, **82(Suppl)**:II-117-II-134
52. Rogers WR, Bowlby LJ, Chandra NC, et al: **Treatment of myocardial infarction in the United States (1990 to 1993): Observations from the National Registry of Myocardial Infarction.** *Circulation* 1994, **90**:2103-2114
53. Topol EJ, Ellis SG, Cosgrove DM, et al: **Analysis of coronary angioplasty practice in the United States with an insurance-claims database.** *Circulation* 1993, **87**:1489-1497
54. The Cardiology Working Group: **Cardiology and the quality of medical practice: A response.** *JAMA* 1991, **265**:496-498
55. Winters WL: **Cardiology and the quality of medical practice: A response.** *JAMA* 1991, **265**:496-8
56. Jollis JG, Ancukiwicz M, DeLong ER, Pryor DB, Muhlbauer LH, Mark DB: **Discordance of databases designed for claims payment versus clinical information systems: Implications for outcomes research.** *Ann Intern Med* 1993, **119**:844-850
57. Every NR, Larson EB, Litwin PE, et al: **The association between on-site cardiac catheterization facilities and the use of coronary angiography after acute myocardial infarction.** *N Engl J Med* 1993, **329**:546-551
58. Blustein J: **High technology cardiac procedures: The impact of service availability on service use in New York State.** *JAMA* 1993, **270**:344-349
59. Rogers WJ, Baim DS, Gore JM, et al: **Comparison of immediate invasive, delayed invasive and conservative strategies after tissue-type plasminogen activator: Results of the Thrombolysis in Myocardial Infarction (TIMI) Phase II-A trial.** *Circulation* 1990, **81**:1457-1476
60. Barbash GI, Roth A, Hod H, et al: **Randomized controlled trial of late in-hospital angiography and angioplasty versus conservative management after treatment with recombinant tissue-type plasminogen activator in acute myocardial infarction.** *Aqm J Cardiol* 1990, **66**:538-545
61. Wennberg JE: **The paradox of appropriate care.** *JAMA* 1987, **258**:2568-2569
62. Yusuf S, Zucker D, Peduzzi E, et al: **Effect of coronary artery bypass graft surgery on survival: Overview of 10-year results from randomized trials by the Coronary Artery Bypass Graft Surgery Trialists Collaboration.** *Lancet* 1994, **344**:563-570
63. Topol EJ, Holmes DR, Rogers WJ: **Coronary angiography after thrombolytic therapy for acute myocardial infarction.** *Ann Intern Med* 1991, **114**:877-885
64. Rogers WJ, Baim DS, Gore JM, et al: **Comparison of immediate invasive, delayed invasive and conservative strategies after tissue-type plasminogen activator: Results of the Thrombolysis in Myocardial Infarction (TIMI) Phase II-A trial.** *Circulation* 1990, **81**:1457-1476
65. The TIMI Research Group: **Immediate vs. delayed catheterization and angioplasty following thrombolytic therapy for acute myocardial infarction: No additional benefits from immediate percutaneous coronary angioplasty.** *JAMA* 1988, **260**:2849-2858
66. Simoons M, Arnold AE, Betriu A, et al: **Thrombolysis with tissue plasminogen activator in acute myocardial infarction: No additional benefits from immediate percutaneous coronary angioplasty.** *Lancet* 1988, **1**:197-203
67. Topol EJ, Califf RM, George BS, et al: **A randomized trial of immediate versus delayed elective angioplasty after intravenous tissue plasminogen activator in acute myocardial infarction.** *N Engl J Med* 1987, **317**:581-588
68. Califf RM, Topol EJ, Stack RS, et al: **Evaluation of combination thrombolytic therapy and timing of cardiac catheterization in acute myocardial infarction: Results of Thrombolysis and Angioplasty in Myocardial Infarction - phase 5 randomized trial.** *Circulation* 1991, **83**:1543-1556
69. Lee KL, Califf RM, Simes J, Van de Werf F, Topol EJ: **Holding GUSTO up to the light: Global utilization of streptokinase and tissue plasminogen activator for occluded coronary artery.** *Ann Intern Med* 1994, **120**:876-881
70. Ridker PM, Donnel CJ, Marder VJ, Hannekens CH: **A response to "holding GUSTO up to the light".** *Ann Intern Med* 1994, **120**:882-885
71. Breiman L, Friedman JH, Olshen RA, Stone CJ: **Classification and regression trees.** Belmont, Calif, Wadsworth International Group 1984
72. Chambers JM, Hastie TJ: **Statistical models in S.** New York, Chapman & Hall 1993, 414
73. Spertus JA, Weiss NS, Every NR, Weaver WD: **The influence of clinical risk factors on the use of angiography and revascularization after acute myocardial infarction.** *Arch Intern Med* 1995, **155**:2309-2316
74. Henning H, Gilpin EA, Covell JW, Swan EA, O'Rourke RA, Ross J: **Prognosis after acute myocardial infarction: A multivariate analysis of mortality and survival.** *Circulation* 1979, **59**:1124-1136
75. The International Study Group: **In-hospital mortality and clinical course of 20, 891 patients with suspected acute myocardial infarction randomized between alteplase and streptokinase with or without heparin.** *Lancet* 1990, **336**:71
76. Emod M, Mock MB, Davis KB, et al: **Long-term, survival of medically treated patients in the Coronary Artery Surgery Study (CASS) Registry.** *Circulation* 1994, **90**:2645-2657
77. Pryor DB, Shaw LR, McCants CB, et al: **Value of the history and physical in identifying patients at increased risks for coronary artery disease.** *Ann Intern Med* 1993, **118**:81-90
78. Maichels KB, Yusuf S: **Does PTCA in acute myocardial infarction affect mortality and reinfarction rates? A quantitative overview (meta-analysis) of the randomized clinical trials.** *Circulation* 1995, **91**:476-485
79. Califf RM, Harrell FE, Lee KL, et al: **The evolution at medical and surgical therapy for coronary artery disease: A 15-year perspective.** *JAMA* 1989, **261**:2077-2086
80. Alderman EL, Bourassa MG, Cohen LS, et al: **Ten-year follow-up of survival and myocardial infarction in the randomized Coronary Artery Surgery Study.** *Circulation* 1990, **82**:1629-1646
81. Ferguson JJ: **Meeting highlights, American Heart Association 68th Scientific Sessions, Anaheim, California, November 13 to 15 1995.** *Circulation* 1996, **93**:843-846
82. Detre KM, Peduzzi R, Hammermeister KE, Murphy ML, Hultgren HN, Takaro T: **Five-year effect of medical and surgical therapy on resting left ventricular function in stable angina: Veterans Administration Cooperative Study.** *Am J* 1984, **53**:444-450
83. Holmes DR, Bates EX, Kleimen NS, et al: **Contemporary reperfusion therapy for cardiogenic shock: the GUSTO-I trial experience.** *J Am Coll Cardiol* 1995, **26**:668-674
84. Meyer J, Merx W, Dorr R, Lamoertz H, Bethge C, Effert S: **Successful treatment of acute myocardial infarction shock by combined percutaneous transluminal coronary revascularization (PTCR) and percutaneous transluminal coronary angioplasty (PTCA).** *Am Heart J* 1982, **103**:132-134
85. Lee L, Bates ER, Pitt B, Walton JA, Laufer N, O'Neil WW: **Percutaneous transluminal coronary angioplasty improves survival in acute myocardial infarction complicated by cardiogenic shock.** *Circulation* 1988, **78**:1345-1351
86. Granger CB, Califf RM, Armstrong PV, et al: **Non-invasive testing is done only in low risk patients with unstable angina and non-Q wave myocardial infarction (MI): Results from GUSTO-IIa.** *J Am Coll Cardiol* 1996, **27(Suppl)**:181A abstract
87. BARI Investigators: **Protocol for the Bypass Angioplasty Revascularization Investigation.** *Circulation* 1991, **84(Suppl V)**:V-1-27
88. Hlatky MA, Charles ED, Nobrega F, et al: **Initial functional and economic status of patients with multivessel coronary artery disease randomized in the Bypass Angioplasty Revascularization Investigation (BARI).** *Am J Cardiol* 1995, **75**:34C-41C

89. Hlatky MA, Boineau RE, Higginbotham MB, et al: **A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index).** *Am J Cardiol* 1989, **64**:651-654
90. Nelson CL, Herndon JE, Mark DB, Pryor DB, Califf RM, Hlatky MA: **Relation of clinical and angiographic factors to functional capacity as measured by the Duke Activity Status Index.** *Am J Cardiol* 1991, **68**:973-975
91. Stewart AL, Greenfield S, Hays RD, et al: **Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study.** *JAMA* 1989, **262**:907-913
92. Stewart AL, Hays RD, Ware JE: **The MOS short-form general health survey. Reliability and validity in a patient population.** *Med Care* 1988, **26**:724-735
93. Lipset SM: **Continental divide. The Values and Institutions of the United States and Canada.** *New York, Routledge* 1991, 1-18
94. Anderson RT, Aaronson NK, Wilkin D: **Critical review of the international assessments of health-related quality of life.** *Qual Life Res* 1993, **2**:369-395
95. Guillemin F, Bombardier C, Beaton D: **Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines.** *J Clin Epidemiol* 1993, **46**:1417-1432
96. Feinstein AR, Sosin DM, Wells CK: **The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer.** *N Engl J Med* 1985, **312**:1604-1608
97. Kirshner B, Guyatt G: **A methodological framework for assessing health indices.** *J Chron Dis* 1985, **38**:27-36
98. Washko RM, Frieden TR: **Tuberculosis surveillance using death certificate data, New York City, 1992.** *Public Health Reports* 1996, **111**:251-255
99. Wen SH, Hernandez R, Naylor CD: **Pitfalls in nonrandomized outcomes studies.** *JAMA* 1995, **274**:1687-1691
100. Deyo RA, Cherkin DC, Ciol MA: **Adapting a clinical comorbidity index for use with ICD-9-CM administrative database.** *J Clin Epidemiol* 1992, **45**:613-619
101. Daley J, Schwartz M: **Developing risk-adjustment methods.** *In: Risk Adjustment for Measuring Health Care Outcomes* 1994, 199-238
102. Breslow N, Day N: **The analysis of case-controlled studies.** *Lyon, IARC* 1980, 338
103. McClellan M, McNeil BJ, Newhouse JP: **Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables.** *JAMA* 1994, **272**:859-866
104. McClellan M, Newhouse JP: **The marginal cost-effectiveness of medical technology: A panel instrumental-variables approach.** *J Econometrics* 1997, **77**:39-64
105. Pilote L, Thomas RJ, Dennis C, Goins P, Houston N-Miller, Kraemer HC, C Leong, Berger WE, Lew H, Heller RS, Rompf J, Debusk RF: **Return to work after uncomplicated myocardial infarction: A trial of practice guidelines in the community.** *Ann Intern Med* 1992, **117**:383-389
106. Schwartz S: **The fallacy of the ecological fallacy: the potential misuse of a concept and consequences.** *Am J Pub Health* 1994, **84**:819-824
107. Gastonis CA, Epstein AM, Newhouse JP, S-Normand L, McNeil BJ: **Variations in the utilisation of coronary angiography for elderly patients with an acute myocardial infarction.** *Med Care* 1995, **33**:625-642
108. Green J, Wintfield N: **How accurate are hospital discharge data for evaluating effectiveness of care?** *Med Care* 1993, **31**:719-731
109. Gleason PP, Kapoor WN, Stone JR, Lave JR, Obrosky DS, Schulz R: **Medical outcomes and antimicrobial costs with the use of the American Thoracic Society guidelines for outpatients with community-acquired pneumonia.** *JAMA* 1997, **278**:32-39
110. Greenfield S: **The state of outcomes research: Are we on target?** *N Engl J Med* 1989, **320**:1142-1143

Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1472-6963/2/7/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright



Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

editorial@biomedcentral.com