We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Clinical Epidemiology: Principles Revisited in an Approach to Study Heart Failure

Ana Azevedo

University of Porto Medical School & Institute of Public Health of the University of Porto
Portugal

1. Introduction

Clinical research studies in general aim to help answer questions that patients most frequently ask their physicians (or these ask themselves): What is wrong with me (diagnosis)? What can you do for me (treatment)? Will I get better (prognosis)?

Diagnosis, prognosis and treatment are obviously related, although not so simply as might be thought at first sight, with prognosis and treatment being dependent on a previously established diagnosis. In fact, the information produced in any of these three activities of clinical practice influences the others and they are much interdependent. For example, prognostic information provides the final confirmation of the diagnosis in some cases, and response to treatment can be used as evidence in favor of a suspected diagnosis. Also, a diagnosis is often immediately connoted with a certain prognosis and the need for certain treatments, and these may be more important to the patient than the diagnosis itself. A good example of this is cancer, with its expected ominous outcomes and fearful treatments.

Another type of concern raised in clinical practice, but of more general interest, is usually addressed in more wide scope epidemiological studies (What caused my illness (etiology)?) and will not be approached in this chapter.

The purpose of this chapter is to review the essentials of clinical epidemiology as a bridging discipline that provides information useful to care for individual patients. The approach to the theme is based on an overview of the modern probabilistic approach to diagnosis, prognosis and assessment of disease management in heart failure.

Heart failure is a complex syndrome with a large and increasing burden that poses interesting and at times unresolved challenges in all the issues that are to be technically discussed. We aim at providing the concepts and guiding the development of competencies necessary for using the medical literature and making clinical decisions.

2. Diagnosis

As Edmond A. Murphy put in his claim for the need for a theory of Medicine, "There is probably no more important field in Medicine than diagnosis and none more difficult to teach. It seems astonishing that it is not attracting hundreds of theorists. We have a crisis of medical care on our hands, and the need to optimize the efficiency of diagnosis is obvious.

What work has been done on diagnosis and by whom? Some statisticians have developed algebraic models; but since they have never seen the diagnostician at work, the models are hopelessly unrealistic. Some few clinicians have accepted the idea that diagnosis is a straightforward application of Bayes' theorem. (…) These imported approaches will not do because they do not start from, and attempt to refine, how the process works in clinical practice. The diagnostic process is a sequential strategy in which the facts are nonindependently and nonidentically distributed, usually collected not singly but in groups, and with an end point constrained by urgency, compassion, cost and redundancy. No useful solution to that challenge is likely to be successful unless the first goal is to specify what the clinician is trying to do" (Murphy, 1997).

The clinical diagnosis of heart failure is unreliable and current recommendations for diagnosis warrant the objective demonstration of cardiac structural or functional abnormalities, usually by echocardiogram (Dickstein et al., 2008). However, its syndromic nature implies that symptoms and signs are the fundamental basis of diagnosis and echocardiographic measurements are also susceptible to measurement error and are strongly observer-dependent. To complicate things further, objective evidence of diastolic dysfunction of some form is currently recommended for the diagnosis of heart failure with preserved ejection fraction (Paulus et al., 2007), which used to be an exclusion diagnosis. This contrasts with the past reliance mainly on left ventricular systolic dysfunction as the underlying cardiac functional abnormality to explain a clinical picture of heart failure, with the exception of valvular heart disease. Diastolic function is technically more difficult to characterize by echocardiogram. Symptoms and signs of heart failure and objective evidence of cardiac dysfunction must both be present for a diagnosis of heart failure to be established, and in case of doubt response to treatment can be considered (Dickstein et al., 2008).

There is no consensual gold standard for the diagnosis of heart failure and the current best reference is an expert's opinion based on clinical, laboratory and functional data.

Plasma concentrations of natriuretic peptides are useful biomarkers in the diagnosis of heart failure. B-type natriuretic peptide (BNP) and N-terminal pro-BNP (NT-proBNP) rise in response to an increase in myocardial wall stress. Evidence exists supporting their use for diagnosing, staging, making hospitalization/discharge decisions, and identifying patients at risk for clinical events. The evidence for their use in monitoring and adjusting drug therapy is less clearly established. There is no definitive cut-off value recognized for either of the two natriuretic peptides for the diagnosis of HF. A normal concentration in an untreated patient has a high negative predictive value and makes HF an unlikely cause of symptoms (Bettencourt, 2005; Dickstein et al., 2008).

2.1 Concordance

Diagnosis starts with clinical history collection and registration.

To assess the relative completeness and validity of data sources for evaluating the quality of care, 1270 patients with at least one of a set of chronic diseases were sampled from 39 American medical organizations and surveyed. Self-reported information and ambulatory care record data were compared to assess concordance (Tisnado et al., 2006).

In this study, the prevalence of previous diagnosis of heart failure was 13% according to the medical record, 9% according to patient's self-report and 18% according to one or both

sources. The proportion of cases in which medical records and patient's self-report agreed regarding previous history of heart failure was 86%, which might seem high. However, if the data in medical records and patient's self-report were truly independent, that is if having a diagnosis of heart failure registered in the medical record was in no way related with the probability of the patient reporting such diagnosis (the extreme, and for this matter absurd, situation of independence, the null hypothesis), then by chance alone agreement could be observed in some cases. This effect of chance is usually the most difficult to understand, but the point is in the concept of statistical independence.

Consider the raw data presented in the following table:

		Medical record		
		Heart failure	No heart failure	Total
Patient's	Heart failure	50 [a]	64 [b]	114
self-report	No heart failure	115 [c]	1041 [d]	1156
	Total	165 [g]	1105 [g]	1270

Table 1. Cross-classification of diagnosis of heart failure according to data source. In concordance tables, paired data are presented so the number of observations is the number of subjects but for each subject two variables are being presented at the same time; for instance, in the first cell 50 cases are counted with heart failure registered in the medical record and also reported by the patient. (Note: the data presented in this table were derived for the purpose of presentation but were not published as such in the original paper (Tisnado et al., 2006) and it is possible they are not real.)

Under independence, one would expect to observe a distribution of heart failure *versus* no heart failure that was the same by strata of what was observed in the other data source. Specifically, if 165/1270, that is 13%, of patients are considered to have heart failure according to the medical record, this distribution would be 13% with heart failure among the 114 who self-reported heart failure (n=14.82) and 13% with heart failure among the 1156 who did not (n=150.28). Thus, the expected number of cases in cell "a" would be 14.82 and the expected number of cases in cell "d" would be 1005.81, and the agreement would be 100x(14.82+1005.81)/1270=80%.

Given the low prevalence of heart failure, even in this sample of patients with chronic diseases, the probability of no heart failure is so high that the agreement expected by chance becomes very high. In other words, regardless of the other data source, in any data source the likelihood of no heart failure is so high that the probability of both data sources reporting no heart failure is also high.

The kappa coefficient aims at quantifying concordance beyond that expected by chance alone, to avoid overestimation of true concordance when looking at absolute agreement. The underlying idea is that concordance varies from 0 to 100% (no cases concordant to all cases concordant). The absolute agreement is the proportion of concordant cases directly observed (in this case 86%). If it was true that the information registered in both data sources was independent of each other, by chance alone one would expect to have observed concordant classification in 80% of cases. So, 86% is only 6% higher than expected by chance under independence. Departing from the expected agreement due to chance, the maximum increase in agreement up to perfect concordance is 20% and kappa expresses the 6% increase as a proportion of this maximum possible: 6/20=0.3 (Fig. 1).
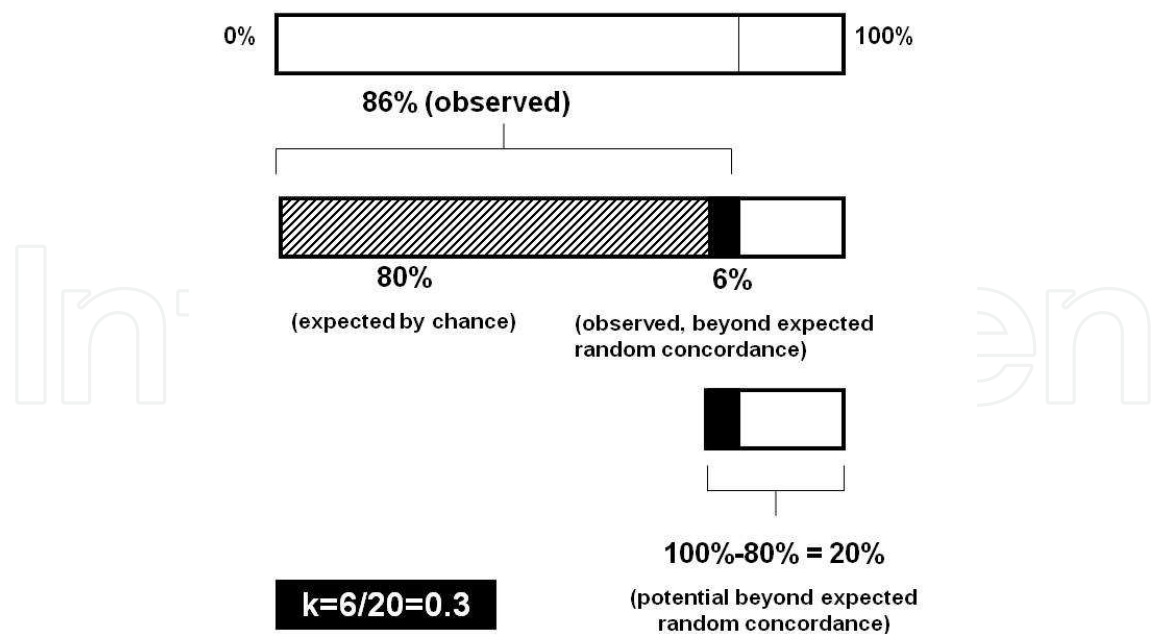
Fig. 1. Interpretation of kappa coefficient.

Assuming the meaning and interpretation of the kappa coefficient is now clear, its definition is: k=(absolute agreement – expected agreement) / (100 – expected agreement). In our example, k=(86-80)/(100-80)=0.3.

It is clear that kappa is zero when the absolute agreement is equal to that expected by chance and its maximal value is 1, when the absolute agreement is 100%, that is, perfect concordance. In theory it can take up negative values if the observed absolute agreement is lower than that expected by chance, but if this happens in situations where you expect some agreement, check for errors in coding of variables in your database before you accept the result.

There are suggestions of ranges of values for kappa to be interpreted as excellent, good, fair or poor agreement but no universal solution exists. The same value might be considered excellent if we are assessing the concordance between scales to measure a subjective and imprecise phenomenon, while it may be unacceptably low when assessing for example the concordance between two laboratory methods to measure the same protein.

Many software packages are available to calculate kappa coefficients, along with confidence intervals estimation, and for more complex scenarios than a two-rater or two-method classification of a dichotomous variable.

Obviously, the kappa coefficient is adequate to assess the concordance in categorical variables that are expected to have the same value. It can be used to quantify the inter- or intra-observer reproducibility (between observers that rate the same subjects or repeated measurements by the same observer), an instrument's precision, or, as in the case presented, the agreement between different methods of assessing the same construct.

The agreement between two observers/methods in assessing a continuous variable cannot be approached with the calculation of a kappa coefficient, unless the results are categorized, resulting in loss of information. The correlation coefficient between two continuous variables, such as for example two measurements of the same phenomenon by different methods or different observers, measures the extent to which they are linearly related, and concordant

observations are obviously expected to be strongly correlated. However, very high correlations do not necessarily reflect agreement. An extreme example illustrates this clearly: the correlation between x and $2x$ is perfect, yet for no subject will they be equal except when x=0.

Over 20 years ago, Bland and Altman proposed a graphical method to assess agreement between methods of clinical measurement (Bland & Altman, 1986). This method is based on the graphical display of the difference between the two raters/methods (Y axis) against the average of both values for each subject (X axis).

Let us see how this applies with an example. A frequently raised question in clinical practice that has important implications on therapeutic decisions in the management of heart failure is that estimates of ejection fraction vary between alternative methods of assessing it, namely standard two-dimensional echocardiogram, nuclear imaging perfusion studies and magnetic resonance.

Mistry et al. determined left ventricular ejection fraction and end-diastolic volumes in 150 patients treated for acute ST-elevation myocardial infarction using four imaging studies – standard echocardiography (standard echo), contrast echocardiography (contrast echo), single-photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI) (Mistry et al., 2010). Fig. 2 depicts a sample of Bland-Altman plots for left ventricular ejection fraction and end-diastolic volume from this study.
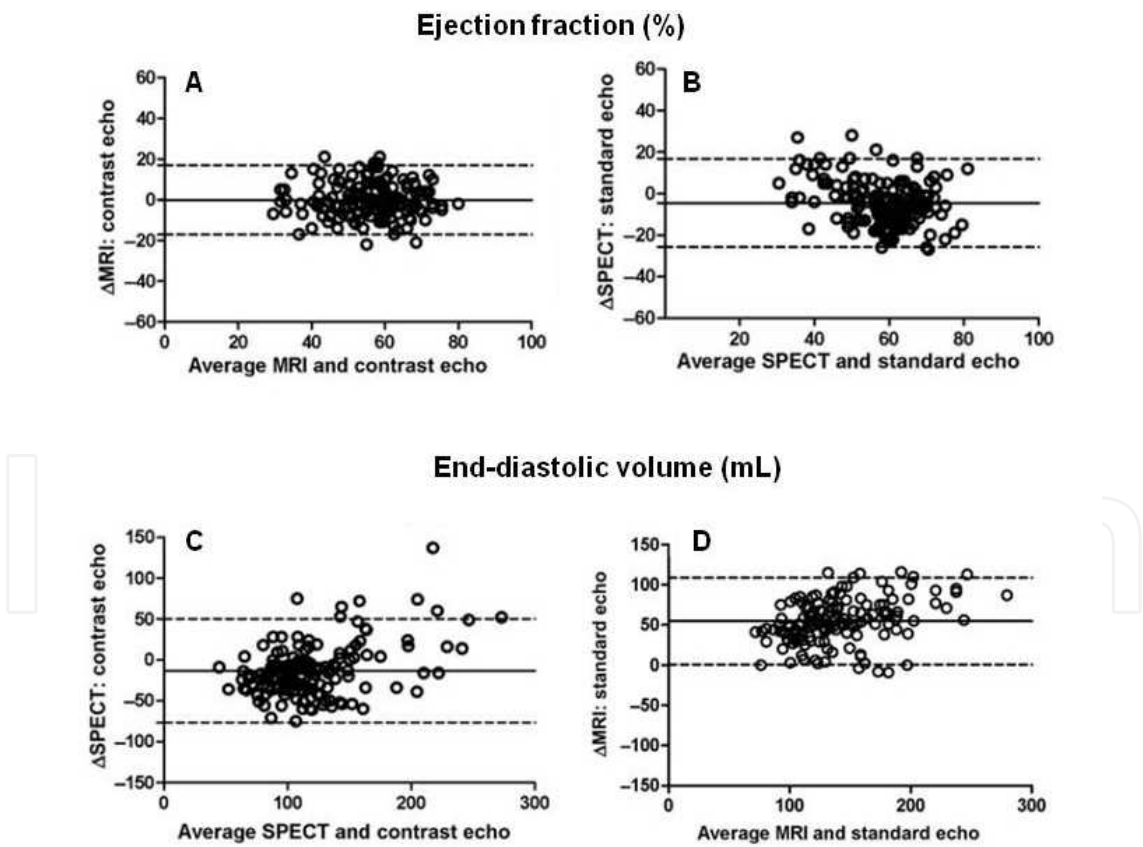

Fig. 2. Agreement in estimation of left ventricular ejection fraction and end-diastolic volume between standard echocardiography (standard echo), contrast echocardiography (contrast echo), single-photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI). Adapted from (Mistry et al., 2010) with kind permission of Elsevier.

With the graph and a few simple calculations, one can extract several informations. Firstly, is the mean difference close to zero? Under perfect concordance, not only would the mean of individual differences be zero but all differences (for all subjects) should be zero. Secondly, is the distribution of differences symmetrical in relation to the horizontal line at the mean of differences? Such symmetry argues in favor of random variation explaining the dispersion. The correlation coefficient between measurements' difference and their mean tests for the presence of systematic error (bias), with higher values suggesting more severe bias. Thirdly, what are the boundaries for variation in differences? These are usually presented and quantified using 95% limits of agreement which are no more than the limits of the interval between 1.96 times the standard deviation of the difference above and below the mean difference. Of course that this works well if the width of variation of differences does not differ much by the average value from the two methods.

In some cases (Fig. 2, Panel C), the width of the variation in differences is larger for higher values of the variable and then it is best to use the logarithm of the initial variables which corresponds to assessing the relative instead of absolute differences.

Bland–Altman analysis of ejection fraction measured by all four imaging modalities showed generally low mean differences but wide limits of agreement. Left ventricular end-diastolic volume was systematically larger when assessed by MRI and, when assessed by SPECT in comparison with contrast echo, it was higher by SPECT only for severely dilated cavities (Fig. 2). While it is known that MRI is the most valid method, it is relevant to assess concordance of alternative methods with the standard since MRI is not always feasible.

The reproducibility of clinical findings in heart failure is important not only at the time of diagnosis but also to interpret changes or absence of change over time, namely because physical examination findings, specifically the jugular venous pressure coupled with biomarker trends, are useful in timing discharge planning or making therapeutic decisions. For example, in order to assess how large the variation between measurements of NT–pro-BNP can be in patients with clinically stable heart failure, we measured its plasma concentration at rest repeatedly at 3-week intervals, in 118 patients. The results supported the clinical use of NT–pro-BNP in the monitoring of patients with HF with high NT–pro-BNP levels (>1,300 pg/mL). In these patients, variations between around 30% more and 30% less than the baseline can be expected without clinical improvement or deterioration; therefore, only changes larger than these should be valued. In patients with lower mean values, the variability was even larger, but in those patients the answer to the appropriateness of monitoring the biomarker over time is not as relevant (Araújo et al., 2006).

Lack of concordance can result from lack of precision or systematic error. Its correct interpretation depends on judgment and familiarity with the question being studied. For example, in the study cited above where the relative completeness and validity of ambulatory medical record and patient's self-report as data sources for evaluating the quality of care setting was assessed (Tisnado et al., 2006), when data on echocardiogram as a delivered service were analyzed, that is, whether the patient had undergone an echocardiograhic examination, the concordance between medical record and patient's self-report was very low (absolute agreement 55%, kappa=0.1). This could have happened for several reasons, some representing systematic error in medical records, such as the absence of registration of ordered tests, the fact that an ordered test might not have been done, etc, systematic error in patient's self-report, such as the patient not knowing that the test

performed was an echocardiogram, etc, or by lack of precision of both data sources. The point is that concordance is not a good measure of validity, even if one of the methods/observers can be considered the standard against which the other(s) is(are) being assessed. Measurement of validity particularly applied to diagnosis will be addressed in the next section of this chapter.

2.2 Validity

Validity of a piece of information for diagnosis, be it a fact collected by clinical history, a finding (or lack thereof) in physical examination or the result of an ancillary test, is assessed by confronting that information with the true state regarding presence or absence of a certain disease one is aiming to confirm or rule out.

The first challenge is that very often there is no good standard to define the true state. Sometimes, only the test of time or response to therapy can bring a definite conclusion regarding a hypothetical diagnosis, but these are affected by other factors, such as competitive risks, effectiveness of interventions, determinants of the treatment decisions and assessment of the patient's condition after some follow up time. Also, even if there is a standard with which to compare a test whose validity one aims to assess, if the standard is not perfect, and it seldom if ever is, apparent lack of sensitivity of the new test may result from mere lack of specificity of the standard (false positive cases in assessment with the standard are actually well classified as negative by the new test, but since the standard is our reference, we see the test result as false negative) and apparent lack of specificity of the test may result from lack of sensitivity of the standard.

In comparing with the standard, the accuracy of the test results is usually quantified with two measures of the proportion of correctly classified cases: sensitivity is the proportion of true cases that are considered positive by the test and specificity is the proportion of true non-cases that are considered negative by the test.

Specificity, like sensitivity, is often considered an intrinsic property of a test and therefore independent of the population under study. However, as specificity is determined by unaffected individuals who have positive results, it is in fact dependent on the characteristics (even subclinical) of this comparison population (Rutjes et al., 2005). It is critical to evaluate the study design from which the specificity of a test has been determined and to consider whether the test can be used more appropriately to distinguish one disease from another or to distinguish the presence or absence of disease. Also, sensitivity is vulnerable to variation depending on the spectrum of severity of the cases studied (Lunet et al., 2009; Ransohoff & Feinstein, 1978). For example, due to these spectrum effects and characteristics of the comparison population, published values for sensitivity and specificity of a long list of history, physical examination and ancillary tests for the diagnosis of heart failure as the cause of dyspnea in the emergency department (Wang et al., 2005) do not necessarily apply to the same clinical findings in primary care or in an epidemiologic study in the general population.

All clinicians understand that predictive value is critical for moving beyond the simplicity of sensitivity and specificity for interpretation of test results. In simple terms, the starting point of a clinical encounter immediately influences the probability of the patient being affected by a certain disease one may be trying to diagnose. Whether the patient came by his initiative due to a complaint or referred by another colleague or was actually called for a

screening procedure, the clinician, more or less consciously depending on personal characteristics and the circumstances, immediately elaborates a list of possible diagnoses, ordered by the probability of being the right diagnosis for that case. He then works from there to gather additional information that will help reorder this list, hopefully bringing some hypotheses to become such remote possibilities that they are excluded and a few, preferably one, to such high probability that it is considered the final diagnosis.

In everyday clinical practice with individual patients, quantitative probability theory is usually not explicitly used. However, diagnostic reasoning, as described in the previous paragraph, involves a probabilistic approach and takes into account the validity of tests when incorporating their results in the process of diagnosis. To interpret any diagnostic test, one must have information not only about the test's characteristics but also about the patient (or a population with similar characteristics). Incomplete epidemiological information that facilitates estimation of pretest probability certainly contributes to the challenge (Bianchi & Alexander, 2006). Few tests are inherently accurate enough to "rule in" or "rule out" disease effectively in all cases. We should look at results as altering disease probability. This requires estimation of a pretest probability that will be adjusted up or down by the test results (Bianchi & Alexander, 2006). This is bayesian logic, which uses an adjustment factor called the likelihood ratio to convert a pretest probability into a posttest probability (Grimes & Schulz, 2005).
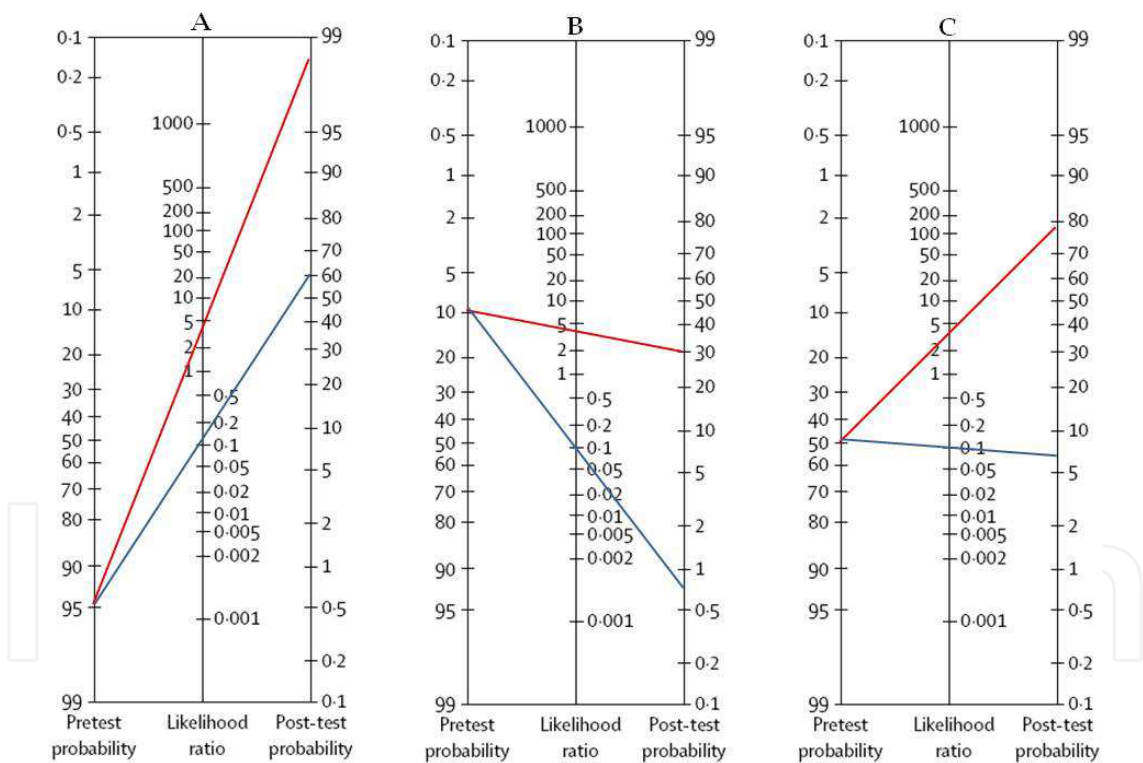

Fig. 3. Change from pre- to posttest probability, after a BNP test result is obtained, considering a positive (≥100pg/mL, red line) or negative (<100pg/mL, blue line) test result, with likelihood ratios of approximately 4 and 0.1, respectively (Wang, 2005), in three hypothetical scenarios in the emergency department: A – acute lung edema (pretest probability assumed to be 95%); B – unspecific malaise in an old patient without previous diagnosis of heart disease (pretest probability assumed to be 10%); C – aggravated dyspnea in a heavy smoker, with chronic obstructive lung disease and past history of myocardial infarction (pretest probability assumed to be 50%).

The key feature of the likelihood ratio is that, unlike traditional indices of validity, it incorporates all four cells of a 2-by-2 table (Grimes & Schulz, 2005). Likelihood ratios help clinicians to navigate zones of clinical uncertainty. Building on an accurate pretest probability of disease, likelihood ratios from ancillary tests can refine clinical judgment.

Clinicians should be wary of ordering tests when the pretest probability of disease is high or low. Tests are unlikely to alter disease probability and will only confuse the situation: unexpected results will usually be false-positives or false-negatives. Consider for example the scenarios represented in Fig. 3. In the case of an acute lung edema (Panel A), in which the clinical diagnosis is generally very accurate, a high BNP level will increase our certainty of the diagnosis from 95% to almost 99%, which is irrelevant. A low BNP value, on the other hand, does not exclude the diagnosis of heart failure and would still leave us thinking that the diagnosis is more likely (66%) to be heart failure than not. The correct clinical attitude would be to treat the acute lung edema immediately, avoiding the delay and cost of the test. If, on the contrary, the pretest probability is low (Panel B), then a low BNP value would only tell us what we already know and a high BNP would more likely be a false positive than represent heart failure. It is in the case of uncertainty (Panel C depicts equal probability of the diagnosis being heart failure or not, the maximum uncertainty) that the test is more able to change our thought: a high BNP value will yield a predictive value of 80%, which in the emergency department and for the hypothesis of heart failure is enough to decide treating as such, while a low BNP will practically exclude heart failure or at least lower its probability so much as to guide the diagnostic work-up in alternative directions.

Likelihood ratios enable clinicians to interpret and use the full range of diagnostic test results, not only dichotomous. For each test result, the likelihood ratio is the ratio between the probability of that result among cases to the probability of that same result among non-cases. Thus, test results with likelihood ratio close to 1, say between 0.5 and 2, are not informative because they are practically as likely to occur in cases as in non-cases and do not change the probability in the diagnostic reasoning. High likelihood ratios above 1 increase the probability of disease and low likelihood ratios below 1 decrease it. Going one step further from dichotomous test results, likelihood ratios can help deal with grey zones. For example, natriuretic peptides have a very low likelihood ratio for low plasma values, say <30 pg/mL, in untreated patients and reasonably high likelihood ratio for high values, say >250 pg/mL (Bettencourt, 2005; Wang, 2005). Between these two cut points, it is not such an informative test.

Ruling disease in or out (or considering subsequent decisions on management) depends on a comparison of posttest probability with thresholds for further action based on factors such as severity of disease, risks of further testing, or side effects of treatment. The posttest probability is the predictive value. Test results cannot be said to have predictive value; only a test result in a given patient (or population) has predictive value (Bianchi & Alexander, 2006).

Recognizing that most tests are imperfect and therefore do nothing more than adjust probability, which may or may not "rule in" or "rule out" the disease depending on the situation, protects against the misconception that a result can be interpreted without considering pretest probability.

The medical history and physical examination remain fundamentally important. Indeed, a precise assessment of the chance of disease can be far more important than the likelihood ratios of sophisticated, usually expensive and sometimes dangerous tests. Although clinical diagnosis might not necessarily be more accurate than ancillary testing, its accuracy has a striking effect on the interpretation of any test results that follow. An accurate pretest probability and subsequent testing can greatly improve clinical diagnosis.

Tests can build on each other in sequence as long as they are independent. Test independence means that the result from one test cannot bias the outcome of the next, such that the posttest probability after one test becomes the pretest probability of the subsequent test.

The purpose of diagnostic work-up is to assess whether the probability of disease is above or below the treatment threshold. Tests should only be used when they will affect management. If a clinician's pretest probability of disease securely rules in or out a diagnosis, further testing is unwarranted. More testing should be considered only in the murky middle zone of clinical uncertainty. The location of the decision thresholds along the continuum of diagnostic certainty needs to be determined before testing is done and should be tailored to the specific patient. Using a nomogram or a simple calculation, a clinician can estimate how high or low a likelihood ratio would have to be to shift the pretest probability down to exclude the diagnosis or up to begin treatment. If no test result would achieve this shift in probability, the test should not be done (Grimes & Schulz, 2005).

2.3 Early diagnosis (and screening?)

The increasingly deeper understanding of the pathophysiology of heart failure led to the definition of stages of heart failure (Hunt et al., 2001), considering asymptomatic cardiac dysfunction as an intermediate step to the development of overt heart failure. Long before this paradigm was established, the importance of asymptomatic left ventricular systolic dysfunction, one of the most important cardiac abnormalities underlying heart failure, was recognized. This recognition is related not only to its frequency, with a prevalence at least as high as that of symptomatic heart failure, but also to the fact that inhibition of the renin-angiotensin-aldosterone system could delay or prevent progression to symptomatic heart failure.

However, screening asymptomatic patients for heart failure remains controversial. In the Cardiovascular Health Study, only 9% of elderly patients who ultimately developed systolic heart failure had a reduced left ventricular ejection fraction on study enrollment, on average 5.5 years before (Gottdiener et al., 2000).

Biomarkers, such as the already mentioned natriuretic peptides, could potentially play this role; however, the cost-effectiveness and target populations for these strategies remain unsettled (Betti et al., 2009). For example, in the Olmsted County cohort, with a low prevalence of left ventricular systolic dysfunction (1.1%), 24% of the population would require an echocardiogram based on raised BNP concentrations and the vast majority of these echocardiograms (96%) would reveal an ejection fraction over 40%. The performance of BNP and NT-proBNP for the detection of left ventricular systolic dysfunction in the community is fair, mainly because of the low specificity, compromising the potential usefulness of the test as a screening procedure. Therefore, BNP testing for screening for left

ventricular systolic dysfunction in the general population is not recommended (Bettencourt, 2005).

Effective primary and secondary prevention to decrease the burden of heart failure can be expected to be attained through adherence to existing guidelines and reduction of the financial and psychosocial barriers that impair adherence to prescribed medical therapy and lifestyle changes recommendations. In clinical practice, it is the practitioner's responsibility to search with clinical examination alone for latent structural heart disease and manifest heart failure, in a case-finding more than screening approach (Raffle & Muir Gray, 2007). Such screening can be accomplished by asking a simple series of questions related to the occurrence of such symptoms as easy fatigability, functional limitations, and development of lower extremity swelling (Ramani et al., 2010). This approach would contribute to better refine the pretest probability of ancillary tests which would no longer be done for screening but rather for diagnosis.

In general, the goodness of attempting an early diagnosis does not depend only on the existence of a valid test for identification of the altered state one is interested in identifying, but also on the ability to define an appropriate target population, that is, with a pretest probability upon which the test results can turn out to have acceptable predictive value. Moreover, early diagnosis should be considered only if effective treatment can be offered and the natural course of the disease changed by this intervention. Such effects on outcomes are best assessed using experimental approaches for complex interventions, that is, the intervention being tested should be the fact that early diagnosis is attempted and all consequences thereof. Obviously, such studies warrant a considerable investment of resources, adequate sample size and a favorable prior odds of successful long-term negative and positive predictive values. An interesting challenge is to create conditions in which these effects can be understood using observational or quasi-experimental research, namely using real-world data of actual practices and their relation with outcomes.

2.4 Utility of a diagnostic test

The rationale as to when a test should be applied requires a judgment that among patients to whom the test is administered, the costs of the illness, both monetary and physical, along with the cost of the test and the errors that arise when it does not classify patients accurately, will be exceeded by the costs of the illness, had the test not been done (Weiss, 2006).

The utility of a diagnostic test is multidimensional and its comprehensive assessment should take into account reproducibility, validity, feasibility, acceptability by subjects, costs and effects on decisions and clinical outcomes. This means that the answer may vary from place to place, institution to institution, physician to physician, and patient to patient.

Earlier sections of this chapter have reviewed issues of reproducibility and validity. We will now briefly address the effect of diagnostic tests on decisions and clinical outcomes, illustrating with an example.

The etiology of systolic heart failure affects prognosis and treatment. In newly diagnosed cardiomyopathy, the exclusion of underlying coronary artery disease and myocardium that might benefit from revascularization is critical. Patients with coronary artery disease

and concomitant heart failure have a worse prognosis than those with nonischemic cardiomyopathy, but myocardial function may substantially improve after revascularization, highlighting the importance of making the appropriate diagnosis early and accurately.

Current European guidelines for the diagnostic work-up of acute and chronic heart failure recommend that coronary angiography should be considered in heart failure patients with a history of exertional angina or suspected ischemic left ventricular systolic dysfunction, following cardiac arrest, and in those with a strong risk factor profile for coronary heart disease, and may be urgently required in selected patients with severe heart failure (shock or acute pulmonary oedema) and in patients not responding adequately to treatment (Dickstein et al., 2008).

Angiographic evidence of atherosclerosis does not necessarily mean that revascularization will be beneficial. Left ventricular dysfunction in patients with coronary artery disease can improve substantially and even normalize after coronary artery bypass grafting (CABG) surgery, presumably due to recovery of function by hibernating myocardium, which is defined as myocardial tissue with abnormal function but maintained cellular function. The assessment of myocardial viability with the use of single-photon-emission computed tomography (SPECT) or low-dose dobutamine echocardiography is commonly performed to predict improvement in left ventricular function after CABG, and numerous studies have suggested that the identification of viable myocardium also predicts improved survival after CABG (Allman et al., 2002). However, studies that suggested an association between myocardial viability and outcome were retrospective in nature, and it is uncertain in most of these studies whether the decision to perform CABG may have been driven by the results of the tests, whether adjustment for key baseline variables was adequate, and whether patients who did not undergo CABG received aggressive medical therapy for heart failure. Therefore, the efficacy of this approach is uncertain.

The Surgical Treatment for Ischemic Heart Failure (STICH) trial was designed to compare the efficacy of medical therapy alone with that of medical therapy plus CABG in patients with angiographic documentation of coronary artery disease amenable to surgical revascularization and with left ventricular systolic dysfunction (ejection fraction ≤35%) (Velazquez et al., 2011). In the initial design of the trial, viability testing with SPECT was required for the enrollment of patients. However, due to unfeasibility of this requirement, the protocol was subsequently revised to make viability testing optional and to allow the use of either SPECT or dobutamine echocardiography for viability testing. Investigators at all study centers were strongly encouraged to perform viability testing in every patient, but the decision to perform the test was left up to the recruiting investigators. This resulted in only around half of patients in the trial undergoing viability testing. The differences in baseline characteristics between patients who underwent viability testing and those who did not undergo such testing suggest that at least some patients may have been selected for testing on the basis of clinical factors.

The risk of death was strongly and significantly lower in patients with than without viability (hazard ratio, 0.64; 95% confidence interval, 0.48–0.86), but after adjustment for other prognostic baseline variables the between-group difference was no longer significant (Bonow et al., 2011).

In patients without viability, the hazard ratio of all-cause death for CABG *versus* medical therapy alone was 0.70 (95% confidence interval: 0.41-1.18) and in patients with viability 0.86 (95% confidence interval: 0.64-1.16), p for interaction=0.53, which was interpreted as evidence that the effect of CABG does not depend on the existence of viability (Bonow et al., 2011). Since in the STICH trial there was overall no significant difference between medical therapy alone and medical therapy plus CABG with respect to this primary end point (Velazquez et al., 2011), this may not be the best endpoint to assess an interaction with viability. Patients assigned to CABG, as compared with those assigned to medical therapy alone, had lower rates of death from cardiovascular causes and of death from any cause or hospitalization for cardiovascular causes (Velazquez et al., 2011); the authors did not find a significant interaction between myocardial-viability status and medical *versus* surgical treatment with respect to the rates of death from any cause or from cardiovascular causes or the rate of death or hospitalization for cardiovascular causes either (Bonow et al., 2011).

Despite the strengths of the study described in comparison with the previous literature, the shortcomings acknowledged by the authors in the STICH trial leave some doubt regarding the effect of viability studies on clinical outcomes. The direct response to this question would warrant the randomization of patients to undergo viability testing or not. A pragmatic approach to estimate effectiveness would leave to the physicians' discretion the decision of management after knowing the result (or not).

Cost-effectiveness analyses are out of scope of this chapter.

3. (Evidence-based) Management

The fact that the decision to order a viability test, and then acting upon the observed result, is dependent on clinical characteristics and on the clinicians' impression of the benefit a particular patient may derive from that approach introduces confounding when comparing outcomes of patients managed this way with those of patients managed otherwise. This is called confounding by indication (the indication for the intervention being tested, or the clinical "hunch" that such indication exists).

Confounding by indication is one of the main reasons why a randomized controlled trial is the study design that creates best conditions for valid causal inferences in attributing an effect in outcomes to an intervention being tested, pharmacological or of other nature. The high position of randomized clinical trials in the hierarchy of study designs for intended effects of therapy derives from this requisite.

Jan Vandenbroucke argues that we may have been deluding ourselves about their unique superiority because they start with much higher prior odds than most observational research (Vandenbroucke, 2008). In fact, for obvious reasons, clinical trials are conducted only under conditions of high probability of success (benefit of the intervention), specifically 1:1, that is, "equipoise" (Djulbegovic et al., 2000) in the sense that it is at least as likely to be beneficial as not, which contributes to the much lower risk of not standing the test of time than observational research which is conducted under much lower prior odds (Vandenbroucke, 2008).

This is not to suggest that this hierarchy is senseless or useless. This author suggests that we need two different hierarchies, the hierarchy of discovery and explanation and that of

evaluation, to assess the evidence of studies designed and conducted for different purposes. Etiologic researchers should pursue low probability hypotheses because these may lead to new insights, particularly if they are able to take advantage of detecting what was wrong when they go down the wrong way (Vandenbroucke, 2008).

In conclusion, the widely accepted hierarchy of levels of evidence in clinical research must be interpreted specifically for the context of evaluation of interventions, be them pharmacological or as "macro" as the organization of services.

It is not the purpose of this text to review extensively the principles and characteristics of clinical trials and several sources of literature can be found regarding this issue. I suggest introductory epidemiology books for a more rapid overview of the fundamental concepts and principles in study design, Haynes et al's Clinical Epidemiology for a thorough review from the clinical point of view with sensible calls of attention useful for interpretation (Haynes et al., 2006), and Friedman and DeMets' references for more applied approach based on case studies covering also execution issues (DeMets et al., 2006; Friedman et al., 2010).

In aiming to provide the concepts and develop competencies necessary for using the medical literature and making clinical decisions, as announced at the beginning of this chapter, we will briefly review an issue that is often misunderstood by clinicians, leading to misinterpretation of the results of trials, and which is intimately related with issues of precision and validity discussed in previous sections.

When, as is the case for most outcomes in heart failure, there are several interventions with well documented benefit and clearly recommended, alternative options for the same effect are tested under non-inferiority and equivalence hypotheses, raising the need to define minimally important differences, which must be established based on knowledge of the natural history of disease guided by clinical judgment. One major condition for credibility of trials is complete preplanning of every aspect of the trial and advance registration and documentation of everything that was preplanned (Laine et al., 2007), including the *a priori* definition of this threshold of effect (Moher et al., 2010).

If a study is designed to test whether a new treatment is significantly different from the control intervention at the 5% significance level, with the null hypothesis being no difference between groups (relative risk = 1 or difference = 0), the absence of a significant difference, that is, failure to reject the null hypothesis, is not synonymous with the interventions being clinically equivalent. Absence of proof of difference is not proof of absence of difference (Haynes et al., 2006). Minimally important differences from the clinical point of view must be defined beforehand and then the results must be interpreted in light of these thresholds and considering the whole width of the confidence interval of the effect estimate (Haynes et al., 2006).

Let us consider an example. In chronic heart failure, inhibition of the renin–angiotensin–aldosterone system by angiotensin-converting enzyme inhibitors improves survival and decreases morbidity, improving exercise capacity, quality of life, and left ventricular function and size (Konstam et al., 1993; Konstam et al., 1992). Although a cornerstone in the treatment of heart failure, angiotensin-converting enzyme inhibitors are underused, partly due to side effects. If proven at least similarly efficacious to angiotensin-converting enzyme

inhibitors, angiotensin-receptor blockers could at least be considered an alternative due to their superior tolerability. The primary objective of the HEAVAN study was to test the hypothesis that valsartan, in comparison with enalapril, is at least as effective on exercise capacity, measured as distance walked during a 6-minute walk test, in heart failure patients stabilised on an angiotensin-converting enzyme inhibitor (Willenheimer et al., 2002).
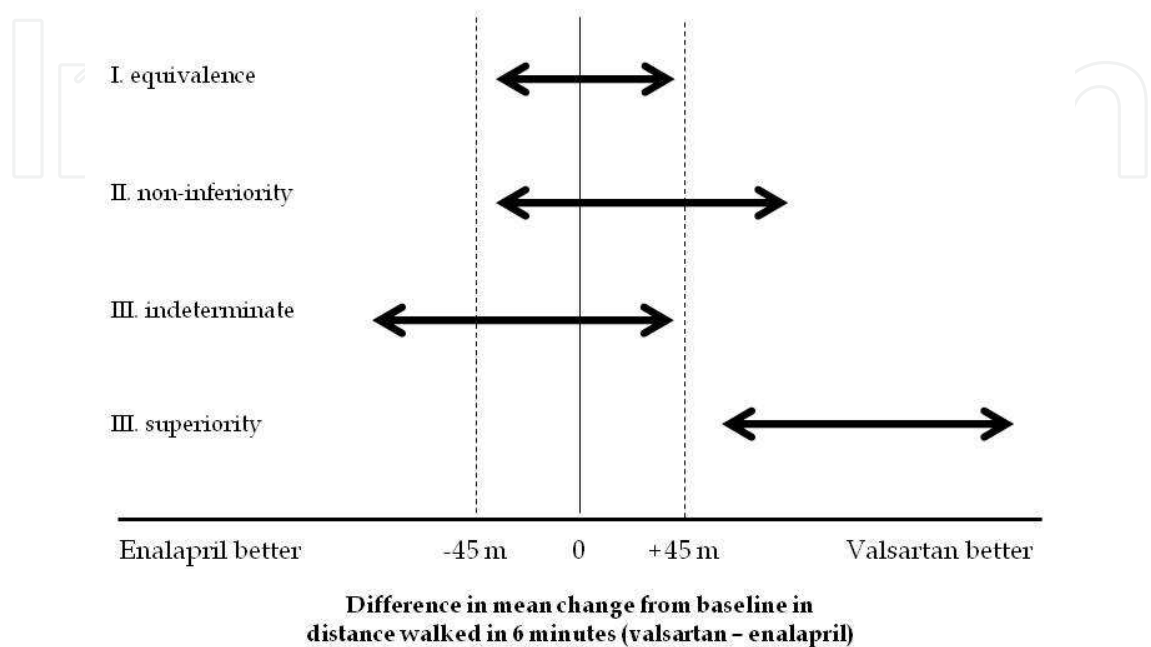

Fig. 4. Schematic representation and interpretation of 4 possible results of the HEAVAN study. The horizontal two-headed arrows represent the 95% confidence interval of the difference between the valsartan and enalapril treatment groups in the change, from baseline to after treatment, in distance walked in 6 minutes.

Non-inferiority was defined as a treatment effect of valsartan, with respect to mean change from baseline in the distance walked during the 6-minute walk test, better than 45 m less than that of enalapril.

A distance of 45 m was chosen based on an expected average baseline 6-min-walk test distance of 450 m, since a difference of 10% in this distance is not considered clinically relevant (Willenheimer et al., 2002). The 10% threshold must at least cover the imprecision of the outcome measurement, if not more, as long as there is fundament from previous data. For example, if one was to study the effect of some intervention on the change in NT-proBNP in chronic heart failure patients, only variations beyond 30% of baseline, upwards or downwards, should be considered clinically important, given the (apparently random) fluctuation of this biomarker under clinical stability, as described above (Araújo et al., 2006).

In the HEAVAN study, the null hypothesis under test is not that there is no difference between treatments. Rather, this is an example of a situation where a one-sided test best suits the underlying reasoning. Unilateral questions warrant one-sided answers (Haynes et al., 2006). The four hypothetic scenarios represented in Fig. 4 show a situation of equivalence (I), with the limits of the confidence interval of the difference between treatments not surpassing 45 m in either direction; non-inferiority (II), with the lower limit of the

confidence interval excluding the possibility of enalapril being more than 45 m more beneficial than valsartan, that is, the data are compatible with valsartan being better than or equivalent to enalapril, but not inferior; scenario III is the most vulnerable to misinterpretation, since the fact that zero difference is not excluded from the range of the confidence interval does not mean that the two interventions are not different and in fact in the situation represented in the figure, the superiority of enalapril, beyond 45 m, cannot be excluded based on the data; superiority (IV) implies that the whole confidence interval is beyond the minimal clinically important difference of 45 m.

Reaching target doses of angiotensin-converting enzyme inhibitors was once the main (process) objective when treating systolic heart failure patients. Angiotensin receptor blockers were initially candidates to replace angiotensin-converting enzyme inhibitors in case of low tolerability, as described above, but there is now evidence that adding an angiotensin receptor blocker to angiotensin-converting enzyme inhibitor leads to a further clinically important reduction in relevant cardiovascular events in such patients (McMurray et al., 2003), possibly even when a beta-blocker and an aldosterone antagonist are concurrently prescribed (Weir et al., 2008). The current state-of-the-art management of heart failure involves use of multiple drugs, which are not free from side-effects particularly when used together and in high doses, with a lower benefit being generally obtained if the highest tolerated dose up to the target is not used. Moreover, patients in need of such care are old and with multiple comorbidities, but it is important to emphasize that the recommended schemes can be tolerated with benefit by a large proportion of patients under the care of experienced professionals.

The complexity of the heart failure syndrome, together with the increasingly recognized need for demonstration of cost-effectiveness of interventions, motivates research to assess the effect of specialized multidisciplinary heart failure management programmes. The heterogeneity of the content and organization of these programmes in large part justifies conflicting results (McDonagh et al., 2011). On the other hand, observational approaches to study the effect of such interventions have been threatened by serious selection bias and confounding by indication (Azevedo et al., 2002). However, experimental evidence has been accumulating supporting the benefit of a range of models of programmes and it is generally accepted that specialized teams of different kinds are more successful in achieving higher rates of patients under recommended prognosis modifying drug and device therapy, with drug doses closer to the recommended targets and overall benefit for patients namely in reducing hospitalizations. Consequently, heart failure management programmes are recommended for patients with HF recently hospitalized and for other high-risk patients (Dickstein et al., 2008).

4. Prognosis

Physicians need to counsel patients about prognosis to enable informed decisions about medications, devices, transplantation, and end-of-life care.

Heart failure has an ominous prognosis, particularly after an acute heart failure episode requiring hospitalization. Half of patients admitted with acute heart failure are readmitted within 6 months (Bettencourt et al., 2004; Jong et al., 2002). Analysis of 100 000 cases of acute decompensated heart failure in the United States revealed that in-hospital mortality after

hospital admission ranges from 5% to 8%, with 1-year mortality averaging 40% to 60% (Adams Jr et al., 2005). It has been general practice to discharge patients according to improvement in symptoms, but, given the reproducibility and validity of clinical examination discussed above, it becomes clear that this decision threshold suffers from severe inter-observer variability. Some studies have tried to identify patients at higher risk of death and/or readmission who might benefit from more intensive therapy. Also for this prognosis issue, natriuretic peptides were good candidates for objective assessment of risk since they decrease in close correlation with falling wedge pressures, and correlate with functional capacity.

We followed a sample of 156 patients consecutively admitted to the hospital due to acute decompensated heart failure and discharged alive, excluding acute coronary syndromes, for the primary end-point of death or hospital readmission for 6 months. The plasma concentration of NT-proBNP at admission was not associated with the endpoint, in contrast with that at discharge, suggesting that it is the change in response to therapy during hospitalization, that matters most. Thus, to refine the risk stratification with the dynamic perspective of change in neuro-humoral activation, we studied the effect of the relative change in NT-proBNP from baseline to discharge. Categories of this new variable were defined according to a change of at least 30%, in agreement with the clinical meaning that can be attributed to the time variation discussed above. The results are depicted in Fig. 5 and clearly show an increase in risk that can be predicted from the change in NT-proBNP during hospitalization and which was independent of clinical signs of volume overload at discharge, thus confirming that new information is being obtained from this biomarker (Bettencourt et al., 2004).
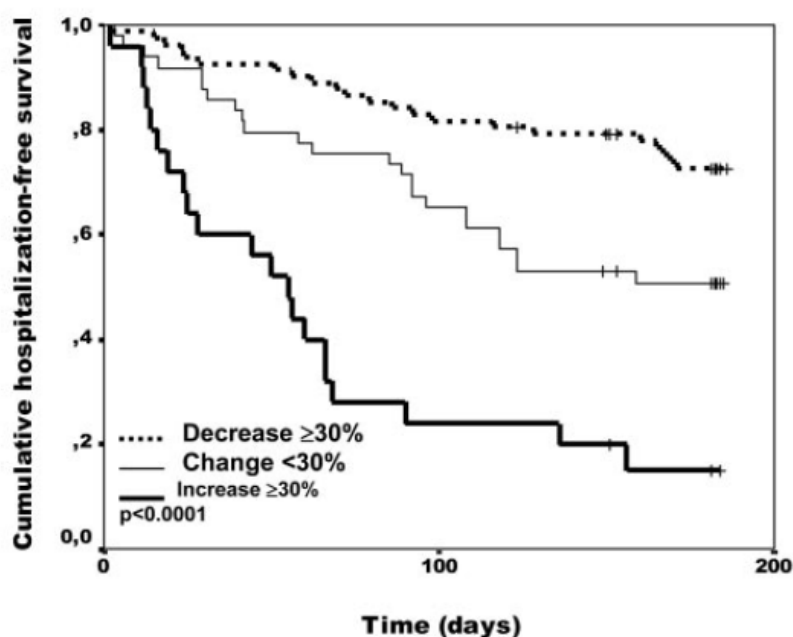

Fig. 5. Cumulative hospitalization-free survival according to patterns of response of NT-proBNP (decreased by ≥30% of baseline value, changed by <30%, increased by ≥30%). Reprinted from (Bettencourt et al., 2004) with kind permission of Wolters Kluwer Health.

Many clinical, laboratory and functional variables have been identified as associated with prognosis in chronic and acute heart failure patients. Clinical prediction rules include simultaneous consideration of several factors in predicting the prognosis of individual patients.

Evaluation of prediction models should consider two attributes. Discrimination is related to higher values of the predicted value of the outcome being obtained among those who actually develop the outcome. This is exactly what the area under a ROC curve represents, specifically the probability that a random person with the outcome has a higher value of the measurement than a random person without the outcome (Altman & Bland, 1994). Calibration goes a step forward and measures to which extent the model predicts well what will happen. This is usually done with a goodness-of-fit test such as the Hosmer-Lemeshow statistic (Lemeshow & Hosmer, 1982) or by simply comparing predicted and observed numbers of events, usually by deciles of predicted risk. Deviations in absolute risk prediction are important when applying a model for clinical decision-making and suggest that recalibration might be necessary, which consists in correcting the baseline risk function with data from the population in which the model is to be used while importing the coefficients (if discrimination is good).

Most existing models to predict the risk of death or urgent transplantation in heart failure have features that may limit their applicability. These models relied on either peak oxygen consumption or invasive measures of cardiac performance, and most have not been validated in another sample than the one used for its development. An exception is the Seattle Heart Failure Model (SHFM) which allows prediction of survival of heart failure patients with the use of easily obtained clinical characteristics. The model provides an accurate estimate of mean, 1-, 2-, and 3-year survival and is unique in allowing estimation of effects of adding medications or devices to a patient's regimen (Levy et al., 2006), potentially contributing to the better prediction of mortality than clinical characteristics alone, because medications and devices are critically altered by physicians to improve the chances of survival of their patients. The model was developed using previously collected data in a cohort of 1125 patients with predominantly left ventricular systolic heart failure [the Prospective Randomized Amlodipine Survival Evaluation (PRAISE)] and 9942 patients from other 5 cohorts were used to prospectively validate the model. PRAISE was a randomized trial of amlodipine *versus* placebo among patients in the United States and Canada with ejection fraction below 30% and New York Heart Association functional class IIIB to IV heart failure (Packer, 1996, as cited in (Levy et al., 2006)).

The SHFM and 3 other predictive models, namely Acute Decompensated Heart Failure National Registry (ADHERE), the American Heart Association Get With The Guidelines-Heart Failure score (GWTG-HF), and Association of Health Aging and Body Composition Heart Failure score (ABC), were all calculated in each of 2472 consecutive patients hospitalized with acute heart failure. The authors compared predicted and observed mortality and also compared the predicted mortality with the observed composite end point, including death, heart transplantation, or implantation of left ventricular assist device (Nakayama et al., 2011). For all the outcomes assessed, the SHFM had highest discrimination as indicated by a higher area under the receiver operating characteristic curve (Fig. 6).
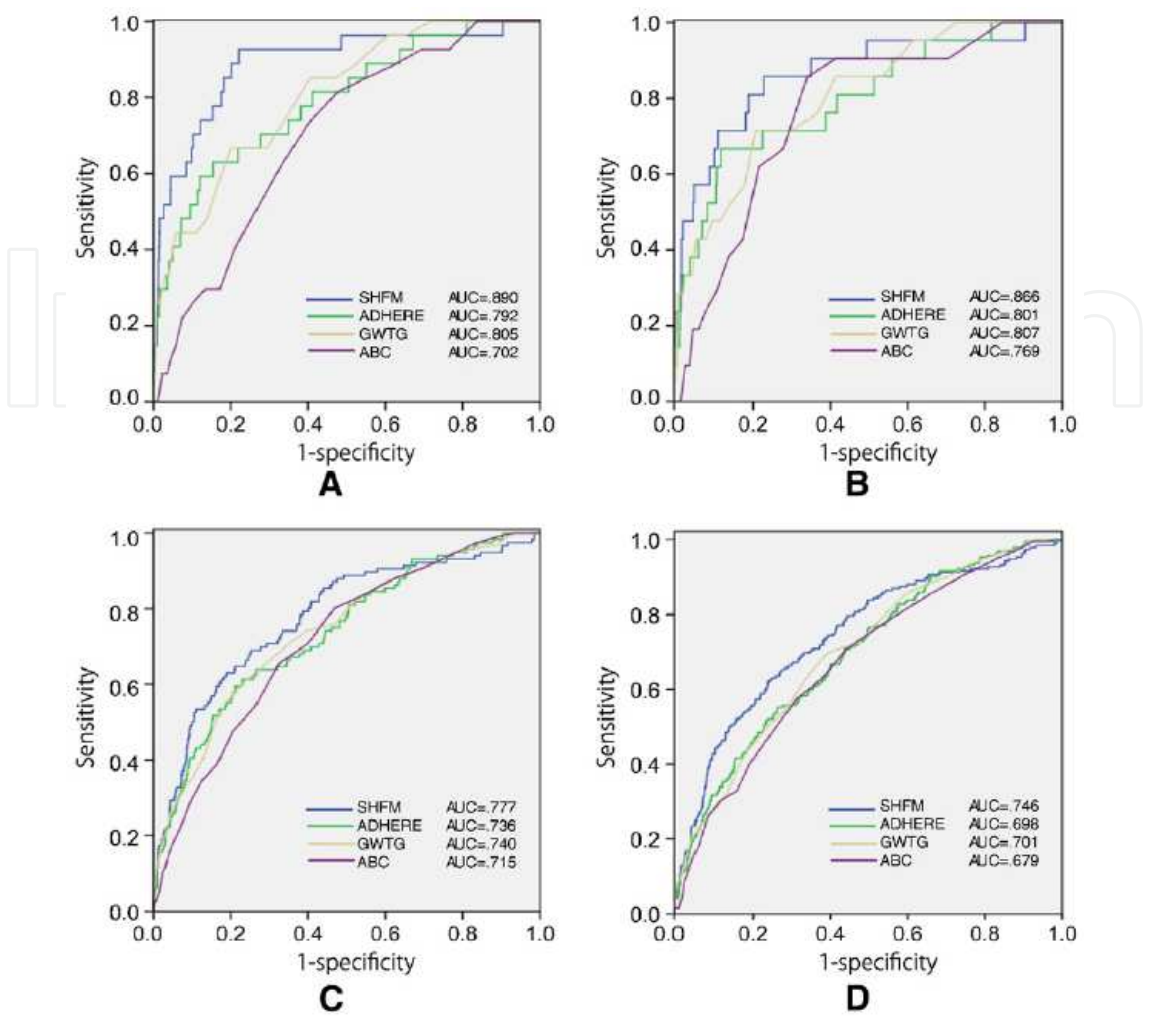
Fig. 6. Area under the receiver operating characteristic curves (AUC) for combined end point of death, heart transplantation, or left ventricular assist device implantation for Seattle Heart Failure Model (SHFM), Acute Decompensated Heart Failure National Registry (ADHERE), the American Heart Association Get With The Guidelines-Heart Failure score (GWTG-HF), and Association of Health Aging and Body Composition Heart Failure score (ABC), for *(A)* in-hospital death, and combined end points at *(B)* 30 days and *(C)* 1- and *(D)* 2 years of follow-up. Reprinted from (Nakayama et al., 2011) with kind permission of Elsevier.

Calibration analysis, presented in Fig. 7, shows absence of important or significant differences between predicted and observed events, supporting the validity of this prognostic model for populations quite different from those of the original study, and indicating that the SHFM is also an adequate risk prediction model in patients with milder heart failure.

Much of the accumulated evidence on the prognosis of heart failure resulted from the prospective assessment of outcomes in patients included in the negative control arm of randomized clinical trials in which a patient population might have been limited because of strict enrollment criteria, resulting in the exclusion of patients with severe conditions and comorbidities. With the development and generalization of access to informatic resources, large administrative databases and electronic medical records have increasingly been used to fit risk prediction models to assess prognosis. Major issues raised by these settings and

data sources are the generalizability of results from randomized trials and completeness and accuracy of data from large administrative databases or electronic medical records.
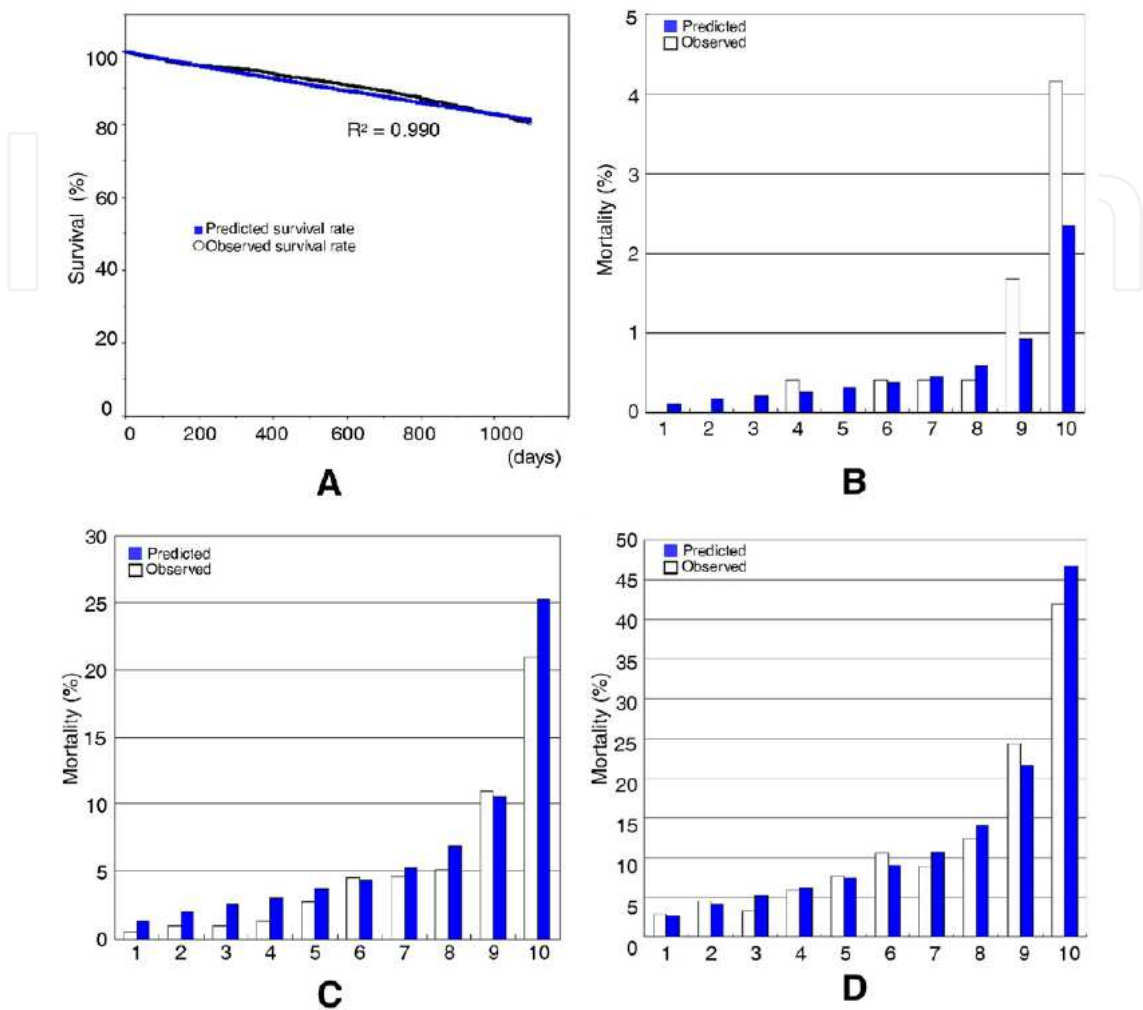

Fig. 7. (A) Comparison of predicted and observed survival for the Seattle Heart Failure Model (SHFM). Predicted (blue) *versus* observed (white) survival rate at each day plotted during follow-up period of ≤3 years. Calibration plots for composite outcome at (B) 30 days and (C) 1 and (D) 2 years for SHFM. Hosmer-Lemeshow chi-square was 7.21 (p = 0.51), 11.15 (p = 0.19), and 5.04 (p = 0.74) at 30 days and 1 and 2 years, respectively. Reprinted from (Nakayama et al., 2011) with kind permission of Elsevier.

Technological evolution led to the possibility of telemonitoring patients and continuously collecting physiological data through implanted devices. These large amounts of data, representing repeated measures over time, raise new challenges for the inclusion of all those time-varying data as independent variables in risk prediction models.

5. Conclusion

In contrast to other major cardiovascular diseases in developed countries, heart failure is a growing problem as the population ages and survivors of myocardial infarction live longer. The successful management of this population depends on risk factor reduction via lifestyle

modification and application of currently established guidelines. During the past 3 decades, a combination of pharmacological, device-based and surgical treatment modalities has tremendously enhanced the survival and quality of life of patients with heart failure. Technological developments improve our capacities but also raise new challenges for their proper use.

The concepts and tools reviewed in this chapter help clinicians and researchers deal with uncertainty. Continued application of these principles in clinical practice and research is vital for optimal medical care for heart failure patients. These are, however, general principles that apply to clinical practice in other areas and heart failure was merely an example setting.

6. Acknowledgment

7. References

Adams Jr, K. F.,Fonarow, G. C.,Emerman, C. L.,LeJemtel, T. H.,Costanzo, M. R.,Abraham, W. T.,Berkowitz, R. L.,Galvao, M. & Horton, D. P. (2005). Characteristics and outcomes of patients hospitalized for heart failure in the united states: Rationale, design, and preliminary observations from the first 100,000 cases in the acute decompensated heart failure national registry (ADHERE). *Am Heart J*, Vol.149, No.2, pp.209-216, 0002-8703.

Allman, K. C.,Shaw, L. J.,Hachamovitch, R. & Udelson, J. E. (2002). Myocardial viability testing and impact of revascularization on prognosis in patients with coronary artery disease and left ventricular dysfunction: A meta-analysis. *J Am Coll Cardiol*, Vol.39, No.7, pp.1151-1158, 0735-1097.

Altman, D. G. & Bland, J. M. (1994). Statistics notes: Diagnostic tests 3: Receiver operating characteristic plots. *BMJ*, Vol.309, No.6948, pp.188.

Araújo, J. P.,Azevedo, A.,Lourenço, P.,Rocha-Gonçalves, F.,Ferreira, A. & Bettencourt, P. (2006). Intraindividual variation of amino-terminal pro-b-type natriuretic peptide levels in patients with stable heart failure. *Am J Cardiol*, Vol.98, No.9, pp.1248-1250, 0002-9149.

Azevedo, A.,Pimenta, J.,Dias, P.,Bettencourt, P.,Ferreira, A. & Cerqueira-Gomes, M. (2002). Effect of a heart failure clinic on survival and hospital readmission in patients discharged from acute hospital care. *Eur J Heart Fail*, Vol.4, pp.353–359.

Bettencourt, P.,Azevedo, A.,Pimenta, J.,Friões, F.,Ferreira, S. & Ferreira, A. (2004). N-terminal–pro-brain natriuretic peptide predicts outcome after hospital discharge in heart failure patients. *Circulation*, Vol.110, No.15, pp.2168-2174.

Bettencourt, P. M. (2005). Clinical usefulness of B-type natriuretic peptide measurement: Present and future perspectives. *Heart*, Vol.91, No.11, pp.1489-1494.

Betti, I.,Castelli, G.,Barchielli, A.,Beligni, C.,Boscherini, V.,De Luca, L.,Messeri, G.,Gheorghiade, M.,Maisel, A. & Zuppiroli, A. (2009). The role of N-terminal pro-brain natriuretic peptide and echocardiography for screening asymptomatic left ventricular dysfunction in a population at high risk for heart failure. The PROBE-HF study. *J Card Fail*, Vol.15, No.5, pp.377-384, 1071-9164.

Bianchi, M. & Alexander, B. (2006). Evidence-based diagnosis: Does the language reflect the theory? *BMJ*, Vol.333, pp.442-445.

Bland, J. & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, Vol.1, No.8476, pp.307-310.

Bonow, R. O.,Maurer, G.,Lee, K. L.,Holly, T. A.,Binkley, P. F.,Desvigne-Nickens, P.,Drozdz, J.,Farsky, P. S.,Feldman, A. M.,Doenst, T.,Michler, R. E.,Berman, D. S.,Nicolau, J. C.,Pellikka, P. A.,Wrobel, K.,Alotti, N.,Asch, F. M.,Favaloro, L. E.,She, L.,Velazquez, E. J.,Jones, R. H. & Panza, J. A. (2011). Myocardial viability and survival in ischemic left ventricular dysfunction. *N Engl J Med*, Vol.364, No.17, pp.1617-1625.

DeMets, D.,Furberg, C. & Friedman, L. (2006). *Data monitoring in clinical trials: A case studies approach* Springer, ISBN 978-0387-20330-0, New York.

Dickstein, K.,Cohen-Solal, A.,Filippatos, G.,McMurray, J. J. V.,Ponikowski, P.,Poole-Wilson, P. A.,Strömberg, A.,van Veldhuisen, D. J.,Atar, D.,Hoes, A. W.,Keren, A.,Mebazaa, A.,Nieminen, M.,Priori, S. G. & Swedberg, K. (2008). ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2008. *Eur Heart J*, Vol.29, No.19, pp.2388-2442.

Djulbegovic, B.,Lacevic, M.,Cantor, A.,Fields, K.,Bennett, C. & al., e. (2000). The uncertainty principle and industry-sponsored research. *Lancet*, Vol.356, pp.635-638.

Friedman, L.,Furberg, C. & DeMets, D. (2010). *Fundamentals of clinical trials* (Fourth edition), Springer, ISBN 978-1441915856, New York.

Gottdiener, J. S.,Arnold, A. M.,Aurigemma, G. P.,Polak, J. F.,Tracy, R. P.,Kitzman, D. W.,Gardin, J. M.,Rutledge, J. E. & Boineau, R. C. (2000). Predictors of congestive heart failure in the elderly: The Cardiovascular Health Study. *J Am Coll Cardiol*, Vol.35, No.6, pp.1628-1637, 0735-1097.

Grimes, D. A. & Schulz, K. F. (2005). Refining clinical diagnosis with likelihood ratios. *Lancet*, Vol.365, No.9469, pp.1500-1505, ISSN 0140-6736.

Haynes, R.,Sackett, D.,Guyatt, G. & Tugwell, P. (2006). *Clinical epidemiology: How to do clinical practice research* (Third edition), Lippincott Williams & Wilkins, ISBN 0-7817-4524-1, Philadelphia.

Hunt, S. A.,Baker, D. W.,Chin, M. H.,Cinquegrani, M. P.,Feldmanmd, A. M.,Francis, G. S.,Ganiats, T. G.,Goldstein, S.,Gregoratos, G.,Jessup, M. L.,Noble, R. J.,Packer, M.,Silver, M. A.,Stevenson, L. W.,Gibbons, R. J.,Antman, E. M.,Alpert, J. S.,Faxon, D. P.,Fuster, V.,Jacobs, A. K.,Hiratzka, L. F.,Russell, R. O. & Smith, S. C. (2001). ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult: Executive summary a report of the American College of Cardiology/American Heart Association task force on practice guidelines (committee to revise the 1995 guidelines for the evaluation and management of heart failure). *Circulation*, Vol.104, No.24, pp.2996-3007.

Jong, P.,Vowinckel, E.,Liu, P. P.,Gong, Y. & Tu, J. V. (2002). Prognosis and determinants of survival in patients newly hospitalized for heart failure: A population-based study. *Arch Intern Med*, Vol.162, No.15, pp.1689-1694.

Konstam, M.,Kronenberg, M.,Rousseau, M.,Udelson, J.,Melin, J.,Stewart, D.,Dolan, N.,Edens, T.,Ahn, S. & Kinan, D. (1993). Effects of the angiotensin converting enzyme inhibitor enalapril on the long-term progression of left ventricular dilatation in patients with asymptomatic systolic dysfunction. SOLVD (Studies of Left Ventricular Dysfunction) investigators. *Circulation*, Vol.88, No.5, pp.2277-2283.

Konstam, M.,Rousseau, M.,Kronenberg, M.,Udelson, J.,Melin, J.,Stewart, D.,Dolan, N.,Edens, T.,Ahn, S. & Kinan, D. (1992). Effects of the angiotensin converting enzyme inhibitor enalapril on the long-term progression of left ventricular dysfunction in patients with heart failure. SOLVD investigators. *Circulation*, Vol.86, No.2, pp.431-438.

Laine, C.,Horton, R.,DeAngelis, C.,Drazen, J., Frizelle, F. A., Godlee, F., Haug, C., Hébert, P. C., Kotzin, S, Marusic, A, Sahni, P, Schroeder, T. V., Sox, H. C., Van der Weyden, M. B. & Verheugt, F. W. (2007). Clinical trial registration—looking back and moving ahead. *N Engl J Med*, Vol.356, pp.2734-2736.

Lemeshow, S. & Hosmer, D. J. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol*, Vol.115, pp.92-106.

Levy, W. C.,Mozaffarian, D.,Linker, D. T.,Sutradhar, S. C.,Anker, S. D.,Cropp, A. B.,Anand, I.,Maggioni, A.,Burton, P.,Sullivan, M. D.,Pitt, B.,Poole-Wilson, P. A.,Mann, D. L. & Packer, M. (2006). The Seattle heart failure model. *Circulation*, Vol.113, No.11, pp.1424-1433.

Lunet, N.,Peleteiro, B.,Carrilho, C.,Figueiredo, C. & Azevedo, A. (2009). Sensitivity is not an intrinsic property of a diagnostic test: Empirical evidence from histological diagnosis of Helicobacter pylori infection. *BMC Gastroenterology*, Vol.9, No.1, pp.98, 1471-230X.

McDonagh, T. A.,Blue, L.,Clark, A. L.,Dahlström, U.,Ekman, I.,Lainscak, M.,McDonald, K.,Ryder, M.,Strömberg, A. & Jaarsma, T. (2011). European Society of Cardiology Heart Failure Association standards for delivering heart failure care. *Eur J Heart Fail*, Vol.13, No.3, pp.235-241.

McMurray, J. J. V.,Östergren, J.,Swedberg, K.,Granger, C. B.,Held, P.,Michelson, E. L.,Olofsson, B.,Yusuf, S. & Pfeffer, M. A. (2003). Effects of candesartan in patients with chronic heart failure and reduced left-ventricular systolic function taking angiotensin-converting-enzyme inhibitors: The CHARM-ADDED trial. *Lancet*, Vol.362, No.9386, pp.767-771, 0140-6736.

Mistry, N.,Halvorsen, S.,Hoffmann, P.,Müller, C.,Bøhmer, E.,Kjeldsen, S. E. & Bjørnerheim, R. (2010). Assessment of left ventricular function with magnetic resonance imaging vs. Echocardiography, contrast echocardiography, and single-photon emission computed tomography in patients with recent ST-elevation myocardial infarction. *Eur J Echocardiogr*, Vol.11, No.9, pp.793-800.

Moher, D.,Hopewell, S.,Schulz, K. F.,Montori, V.,Gøtzsche, P. C.,Devereaux, P. J.,Elbourne, D.,Egger, M. & Altman, D. G. (2010). Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*, Vol.63, No.8, pp.e1-e37, 0895-4356.

Murphy, E. (1997). *The logic of medicine.* (Second edition), The Johns Hopkins University Press, ISBN 0-8018-5538-1, Baltimore.

Nakayama, M.,Osaki, S. & Shimokawa, H. (2011). Validation of mortality risk stratification models for cardiovascular disease. *Am J Cardiol*, Vol.108, pp.391-396.

Paulus, W. J.,Tschöpe, C.,Sanderson, J. E.,Rusconi, C.,Flachskampf, F. A.,Rademakers, F. E.,Marino, P.,Smiseth, O. A.,De Keulenaer, G.,Leite-Moreira, A. F.,Borbély, A.,Édes, I.,Handoko, M. L.,Heymans, S.,Pezzali, N.,Pieske, B.,Dickstein, K.,Fraser, A. G. & Brutsaert, D. L. (2007). How to diagnose diastolic heart failure: A consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the heart failure and echocardiography associations of the European Society of Cardiology. *Eur Heart J*, Vol.28, No.20, pp.2539-2550.

Raffle, A. & Muir Gray, J. (2007). What screening is, and is not, In: *Screening: Evidence and practice*. A. Raffle and J. Muir Gray. pp. 33-57, Oxford University Press, ISBN 978-0-19-921449-5, New York.

Ramani, G. V.,Uber, P. A. & Mehra, M. R. (2010). Chronic heart failure: Contemporary diagnosis and management. *Mayo Clin Proc*, Vol.85, No.2, pp.180-195, ISSN 1942-5546.

Ransohoff, D. & Feinstein, A. (1978). Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*, Vol.299, pp.926 - 930.

Rutjes, A. W. S.,Reitsma, J. B.,Vandenbroucke, J. P.,Glas, A. S. & Bossuyt, P. M. M. (2005). Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem*, Vol.51, No.8, pp.1335-1341.

Tisnado, D. M.,Adams, J. L.,Liu, H.,Damberg, C. L.,Chen, W.-P.,Hu, F. A.,Carlisle, D. M.,Mangione, C. M. & Kahn, K. L. (2006). What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Med Care*, Vol.44, No.2, pp.132-140, 0025-7079.

Vandenbroucke, J. P. (2008). Observational research, randomised trials, and two views of medical science. *Plos Med*, Vol.5, No.3, pp.e67.

Velazquez, E. J.,Lee, K. L.,Deja, M. A.,Jain, A.,Sopko, G.,Marchenko, A.,Ali, I. S.,Pohost, G.,Gradinac, S.,Abraham, W. T.,Yii, M.,Prabhakaran, D.,Szwed, H.,Ferrazzi, P.,Petrie, M. C.,O'Connor, C. M.,Panchavinnin, P.,She, L.,Bonow, R. O.,Rankin, G. R.,Jones, R. H. & Rouleau, J.-L. (2011). Coronary-artery bypass surgery in patients with left ventricular dysfunction. *N Engl J Med*, Vol.364, No.17, pp.1607-1616.

Wang, C. S.,FitzGerald, J. M.,Schulzer, M.,Mak, E. & Ayas, N. T. (2005). Does this dyspneic patient in the emergency department have congestive heart failure? *JAMA*, Vol.294, No.15, pp.1944-1956.

Weir, R. A. P.,McMurray, J. J. V.,Puu, M.,Solomon, S. D.,Olofsson, B.,Granger, C. B.,Yusuf, S.,Michelson, E. L.,Swedberg, K.,Pfeffer, M. A. & Investigators, C. (2008). Efficacy and tolerability of adding an angiotensin receptor blocker in patients with heart failure already receiving an angiotensin-converting inhibitor plus aldosterone antagonist, with or without a beta blocker. Findings from the candesartan in heart failure: Assessment of reduction in mortality and morbidity (CHARM)-ADDED trial. *Eur J Heart Fail*, Vol.10, No.2, pp.157-163.

Weiss, N. (2006). *Clinical epidemiology. The study of the outcome of illness.* (Third edition), Oxford University Press, ISBN 978-0-19-530523-4, New York.

Willenheimer, R.,Helmers, C.,Pantev, E.,Rydberg, E.,Löfdahl, P. & Gordon, A. (2002). Safety and efficacy of valsartan versus enalapril in heart failure patients. *Int J Cardiol*, Vol.85, No.2-3, pp.261-270, 0167-5273.

Epidemiology - Current Perspectives on Research and Practice

Edited by Prof. Nuno Lunet

ISBN 978-953-51-0382-0

Hard cover, 208 pages

Publisher InTech

Published online 13, March, 2012

Published in print edition March, 2012

This special issue resulted from the invitation made to selected authors to contribute with an overview of a specific subject of their choice, and is based on a collection of papers chosen to exemplify some of the interests, uses and views of the epidemiology across different areas of research and practice. Rather than the comprehensiveness and coherence of a conventional textbook, readers will find a set of independent chapters, each of them of a great interest in their own specialized areas within epidemiology. Taken together, they illustrate the contrast between the attempt to extend the limits of applicability of epidemiological research, and the "regular" scientific activity in this field or an applied epidemiology. Epidemiologists with different levels of expertise and interests will be able to find informative and inspiring readings among the chapters of this book.

INTECH

open science | open minds