

## Choice of Evaluation Criteria in a Clinical Trial

Michel Bourin<sup>1\*</sup> and Ricardo P Garay<sup>2,3</sup>

<sup>1</sup>Neurobiology of Mood disorders, University of Nantes, Nantes, France.

<sup>2</sup>Pharmacology and Therapeutics, Craven, Villemoisson-sur-Orge, France.

<sup>3</sup>CNRS, National Centre of Scientific Research, Paris, France.

Received: September 23, 2016; Accepted: October 5, 2016; Published: October 27, 2016

\*Corresponding author: Michel Bourin, Neurobiology of Mood disorders, University of Nantes, 98 rue Joseph Blanchart, 44100 Nantes, France; E-mail: michel.bourin@univ-nantes.fr

### Abstract

The choice of proper evaluation criteria is a key aspect of any clinical trial. The efficacy, tolerance and safety of symptomatic treatments can be evaluated by assessing clinical symptoms and/or signs (vital functions). For life-threatening diseases, the relevant evaluation criterion is morbi-mortality. However, this criterion is rarely chosen because it requires very long-term studies and a large number of patients. Then, a biomarker or other intermediary criteria can be used as a substitute endpoint. The intermediary criterion should correlate well with morbi-mortality (clinical criterion). In these last years, the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) have been publishing scientific guidelines defining efficacy and safety evaluation criteria for the phase I and II “proof of concept” Randomized Clinical Trials (RCTs) and the phase III “pivotal trials” which are required for marketing authorization. The present paper describes the rationale underlying the choice of evaluation criteria, together with the advantages and limitations of their application in some selected therapeutic indications.

**Keywords:** Clinical trials; Evaluation criteria; Intermediary criteria; Judgment criteria;

### Introduction

Evaluation criteria include outcome measurements and assessment tools designed to evaluate the efficacy, safety and tolerability of a new treatment. Classically, the choice of evaluation criteria in a clinical trial was left to the discretion of the investigator. In these last years, the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) have been publishing scientific guidelines on how to interpret and apply the requirements for the demonstration of quality, safety and efficacy of health interventions in a large number of therapeutic indications [1, 2]. These regulatory guidelines define the efficacy outcomes and assessment tools for phase II “proof of concept” Randomized Clinical Trials (RCTs) and phase III “pivotal trials” which are required for marketing authorization (for definition of “proof of concept” studies see [3, 4], for “pivotal trials” see [5]).

A “therapeutic indication” defines the symptom or disease and the population for which the health intervention is intended.

Generally speaking, the primary outcome measure evaluates the efficacy of a given intervention on the main diagnostic criteria used to define the therapeutic indication (i.e. psychotic symptoms for schizophrenia; fast blood glucose and glycated hemoglobin for diabetes). Secondary outcome measures include individual components of the diagnostic criteria (specific symptoms, biomarkers and/or others), together with parameters evaluating quality of life, satisfaction, social and professional functioning, and burden for families and caregivers. “Safety studies” explore undesirable effects, particularly those related to mechanism of action.

Taken into account the above elements, we describe here the rationale underlying the choice of evaluation criteria. This means, which outcome measurements (clinical symptom severity, changes in vital signs, biochemical markers and imaging) may be chosen to evaluate treatment as a function of the type of therapeutic indication.

### Evaluation criteria for symptomatic treatments

Many therapeutic indications concern a given clinical symptomatology, including disorders in vital functions (clinical signs). Symptomatic treatment is also the rule for disorders where the physiopathological mechanisms are badly known, such as mental disorders. The primary outcome measurement of a symptomatic treatment is intended to evaluate the symptom severity and/or the functional disorders defining a therapeutic indication. Secondary outcome measurements may include single symptoms or signs, quality of life assessments and drug consumption (e.g., many evaluation methods are available to quantify drug consumption in chronic rheumatic diseases [6]).

Many studies, expert meetings and manuals have been dedicated to define a “core” of symptoms and signs of different diseases [7] (for mental disorders see [8]). Questionnaires concerning “core symptoms” have been edited, together with scales evaluating symptom severity, e.g., the Hamilton Anxiety Rating Scale (HAM-A) [9] and the Positive and Negative Syndrome Scale (PANSS) for schizophrenia [10]. Scale scores are evaluated with categorical or continuous scales. A typical categorical clinical scale is based on the following symptom severities: 0 = normal or absent, 1 = mild, 2 = moderate, 3 = severe and 4 = extremely

severe. A visual analog scale may also allow the quantification of subjective symptoms (pain, anxiety, pruritus, hunger, vigilance, etc.). The patients provide answers to the questionnaire items, which then allow evaluating scale scores. Diagrams, figures and drawings are useful in certain circumstances such as pain in children, ophthalmologic pathologies and Parkinson's disease.

### Evaluation criteria for life-threatening diseases

The most relevant judgment criterion for life-threatening diseases is morbi-mortality. However, this criterion is rarely chosen since its evaluation requires very long-term studies, including a large number of patients and dramatically increasing resource needs and associated costs.

**Intermediary criteria:** By contrast with the above clinical endpoints (assessing how a patient feels or functions), a biomarker or other intermediary criteria can be used as a substitute endpoint for life-threatening diseases [11-15]. For instance, blood pressure is not a clinical criterion, but epidemiological studies have shown that an increase in blood pressure is an important risk factor of cardiovascular disease [16]. Therefore, blood pressure is a pertinent intermediary criterion for cardiovascular disease, and it is the primary outcome measure to evaluate drugs for arterial hypertension. Similarly, fast blood glucose and glycated hemoglobin are reliable intermediary criteria for diabetes complications and primary outcome measures in many diabetes trials.

The use of an intermediary criterion likely reduces trial's time frame, resources and costs. The intermediary criterion should correlate well with the clinical criterion (morbidity and mortality). This implies that it plays a reasonably physiopathological role in the disease that epidemiologic studies have demonstrated a relation with the clinical criterion, and that therapeutic trials have compared treatments effects on both criteria.

The intermediary criterion should be reliably, reproducible and precisely measurable. Its prevalence should be higher than the corresponding clinical criterion, and its variation more rapid. If the statistical power of the trial is not sufficiently high, the inclusion of a higher number of patients and the increase of the time frame will increase costs, with a risk that medical advances might make it useless.

**Recent developments:** Our understanding of molecular pathways of heterogeneous diseases (e.g., some types of cancer, neurodegenerative disorders) has evolved, and some specific biomarkers are now available to identify responders to a given treatment [12]. Newer clinical trial designs have been developed to assess clinically meaningful endpoints in biomarker-enriched populations, and the number of modern, molecularly driven clinical trials is steadily increasing [13].

Functional imaging plays an important role in cardiovascular disease and imaging outcomes are integrated in cardiovascular clinical trials [14]. The same can be said for modern neuro imaging, including Magnetic Resonance Imaging (MRI) or amyloid imaging with Positron Emission Tomography (PET), which tend to correlate quantitatively with disease progression [11]. Imaging

is also of help to assess therapeutic efficacy for cancers [15].

**Future developments:** A number of therapeutic indications urgently need the development of reliable intermediary criteria for treatment evaluation. Heart failure and arterial hypertension are two examples among many.

Clinical signs of heart failure may vary depending on the causal factors and the use of diuretics. The discrimination point between normality and heart failure is difficult to determine, even by using objective signs. In fact, the zone of non-normality cannot exactly correspond to the beginning of heart failure. Moreover, a rhythm disorder can complicate the diagnosis.

Many intermediary criteria have been used in heart failure, and all raise problems. Clinical symptomatology and functional signs have no correlation with mortality and morbidity. Hemodynamic studies are precise, but they are invasive, frequently involve a single measurement and exhibit little correlation with functional signs. Ejection fraction can be measured by hemodynamic, angioscintigraphic or ultrasonographic approaches [17]. The measurements are reproducible, but the interpretation can be problematic in the event of akinetic zones. Measurement of the systolic ejection fraction concerns not only the heart, but also the peripheral sector, and no clear correlation with morbi-mortality (and quality of life). Ergometric data can be obtained with a bicycle, treadmill or by measurement of VO<sub>2</sub> max. However, this raises a problem as to the reliability and determination (20 to 25% variation), the effect of training and the fact that VO<sub>2</sub> max is rarely attained in cases of heart failure. Heart size is a simple measurement but difficult to reproduce and the correlation is non-linear. For rhythm disorders there is little correlation.

The above considerations clearly indicate the strong need of further exploration of novel analytical methods in heart failure, as well as development of predictive and prognostic biomarkers and more personalized treatments. The metabolism of the failing heart is significantly impaired from its baseline state may be a future target not only for biomarker discovery but also for pharmacologic interventions [18]. Finally, the choice of inclusion criteria should consider the facility of performing the study and the subsequent generalization of the results.

Concerning arterial hypertension, blood pressure can be chosen as main evaluation criterion if the purpose of the trial is to lower it, and as intermediary criterion if the purpose is to show the usefulness of lowering it (in which case the main criterion is the occurrence of a cardiovascular event or coronary mortality).

The limit between normality and arterial hypertension depends on the inter-individual variability in at least two factors [19]. First, methodological errors and bias can take place during blood pressure measurements, in spite of the usual recommendations for preventing them. Second, biological factors such as stress conditions or physical activity can increase blood pressure, and should be avoided by measuring blood pressure at rest and in the absence of stress conditions. These two elements lead to 20% blood pressure variations in normotensive subjects, which may explain why 30% of such patients react to a placebo.

Multiple ambulatory measurements provide a better blood pressure assessment, thus reducing intra-individual variability and the number subjects to include in the trial [20]. Moreover, ambulatory measurement of blood pressure is a better indicator of cardiovascular risk and is better correlated with left ventricular hypertrophy [20]. However, no precise standard of normality exists for ambulatory blood pressure. Although there is a clear advantage in clinical pharmacology for ambulatory blood pressure measurement, this is not a routine assessment in clinical trials. Ambulatory blood pressure may be a selection criterion for patient inclusion, but cannot be used as primary outcome measurement for monitoring the trial. Indeed, assessment of cardiovascular morbi-mortality should be the primary outcome measurement, whenever possible.

**Limitations of using intermediary criteria:** One should be cautious with intermediary criteria that have not been validated, particularly in psychiatric pathology in which it is difficult to make reasonable correlations as to the course of a disease on the basis of purely and strictly biological criteria. Certain subjectivity may enter in the interpretation of “objective” results from some complementary examinations (e.g. radiographic ones). Doubtful examinations should be repeated. Moreover, there is a risk of using an objective examination for indirect assessment of a very subjective symptom. For instance, bone scintigraphy, which appears to be a very objective examination, is a poor mean of evaluating a disease such as rheumatoid arthritis.

In the absence of treatment, the intermediary and the clinical criterion can be unrelated. Therefore, the correspondence between the two criteria should be verified. Unfortunately, this correspondence is often unknown, in which case epidemiologic or multivariate (multiple risk factors) studies may be useful. During treatment, discordances can be related to the effect of chance, errors in physiopathology between the intermediary and the clinical criterion, a slow effect of the treatment (cholestyramine) or treatment leading to an iatrogenic pathology. The possibility of an error due to chance should always be considered.

### Factors that influence decision-making

**Quality of the evaluation criteria:** The quality of the evaluation criteria and their assessment tools is essential to obtain precise and reliable information in a clinical trial [21, 22]. A new judgment criterion, either direct or indirect, should be validated in preliminary studies. The judgment criterion is considered to be sensitive if it allows the detection of very significant improvements or deteriorations of the patient's condition. It is essential to check the sensitivity of a new criterion before accepting a negative result. The criterion is regarded as specific if it does not detect supposed improvements (false-positive results) due to outside factors. Consistency relates to the reproducibility of measurements made by the same observer and to the concordance of measurements made by different observers. It is necessary to employ the same trained instructors. The criterion is considered to be stable if there is no variation over time in a given individual, or labile if there is variation. It is sometimes useful to calculate means (e.g. for blood pressure).

**Chronology of measurements:** Regarding longitudinal studies, the evaluation criteria should integrate variations in outcome measurements, either as absolute differences with baseline or as percentage of improvement. Repeated measurements before treatment allow for a better assessment of the basal state in case of labile outcome measurements (e.g., blood pressure). Repeated measurements during treatment allow assessment of changes in criteria and the influence of the time factor as well as time/treatment interactions. An overall analysis is therefore performed.

**Type of treatment:** The objective of curative treatment is shortening the course of a disease. Thus, the evaluation criterion should integrate disease duration (which supposes the problem of normality and cure) and morbi-mortality. For preventive treatment, the criterion must relate to the frequency of occurrence of the pathologic state in question (primary or secondary prevention), which raises the problem of the detection method to be used and the length of the disease course. The purpose of symptomatic treatment is to relieve a symptom without affecting the course of the disease. The judgment criterion depends on the intensity of the symptom or the duration of the effect achieved. Finally, the objective of palliative treatment is to delay the unavoidable progression of the disease toward death. The judgment criterion is the time of occurrence of this event (rate of survival).

The type of treatment influences the significance of hypercholesterolemia as biomarker of cardiovascular disease. Statins are the cornerstone of dyslipidemia management, and observational studies have shown that low levels of High-Density Lipoprotein Cholesterol (HDL-C) are associated with increased rates of cardiovascular diseases. However, high triglyceride levels may persist in some patients despite statin therapy, and epidemiologic and clinical studies have suggested that elevated triglyceride levels are a biomarker of cardiovascular disease [23]. Consistent with these findings, recent genetic evidence from mutational analyses, genome-wide association studies, and Mendelian randomization studies provide robust evidence that triglycerides and triglyceride-rich lipoproteins are involved in the causal pathway for atherosclerotic CV disease. On the other hand, it remains unclear whether strategies aimed at increasing HDL-C in addition to background statin therapy will further reduce risk. The AIM-HIGH (Atherothrombosis Intervention in Metabolic Syndrome With Low HDL/High Triglycerides: Impact on Global Health Outcomes) trial, which compared combined niacin/simvastatin with simvastatin alone, failed to demonstrate an incremental benefit of niacin among patients with atherosclerotic CVD and on-treatment low-density lipoprotein cholesterol values <70 mg/dl, but this study had some limitations[24].

### Quality of life studies

Several questionnaires have been developed to investigate quality of life in clinical trials (usually as secondary outcome criteria). Among these, we can mention the QVS (Quality of Life and Health), the PQVS (Profile of the Quality of Subjective Life), the OCAPI (Optimization of the Choice of Antihypertensive in

First Intention), and the APQV (Quality Adjusted Life Years). The APQV questionnaire is particularly efficient and can be used as an economic tool. A single figure indicates the length and quality of survival. For example, the length of life that an individual will sacrifice to obtain a better quality of life can be estimated (an additional year of life = +1, and a year of life with disease manifestations < 1) [25].

## Conclusions

The efficacy, tolerance and safety of symptomatic treatments can be evaluated by assessing the severity of clinical symptoms and/or disorders in vital functions (clinical signs). For life-threatening diseases, the relevant evaluation criterion is morbidity and mortality (clinical criteria). However, such clinical criteria are rarely chosen because they require very long-term studies and a large number of patients. Then, a biomarker or other intermediary criteria can be used as a substitute endpoint. The choice of intermediary (often biomarkers) rather than clinical criteria is warranted for practical, economic or ethical reasons. If absent, a serious preliminary study of the relation between the intermediary and the clinical criterion is required. It is also necessary to obtain validation even after drug registration on a largest population.

## References

- EMA. Clinical efficacy and safety guidelines. European Medicines Agency. 2016.
- FDA. Guidances (Drugs). Clinical/Medical. U.S Food and Drug Administration. 2016.
- Kendig CE. What is Proof of Concept Research and how does it Generate Epistemic and Ethical Categories for Future Scientific Practice?. *Sci Eng Ethics*. 2016;22(3):735-753. doi: 10.1007/s11948-015-9654-0.
- Schmidt B. Proof of Principle studies. *Epilepsy Res*. 2006;68(1):48-52. doi: 10.1016/j.epilepsyres.2005.09.019.
- Smithy JW, Downing NS, Ross JS. Publication of pivotal efficacy trials for novel therapeutic agents approved between 2005 and 2011: a cross-sectional study. *JAMA Intern Med*. 2014;174(9):1518-1520. doi: 10.1001/jamainternmed.2014.3438.
- Constant F, Guillemin F, Herbeth B, Collin JF, Boulange M. Measurement methods of drug consumption as a secondary judgment criterion for clinical trials in chronic rheumatic diseases. *Am J Epidemiol*. 1997;145(9):826-833.
- WHO. International Statistical Classification of Diseases and Related Health Problems(10 Edition). World Health Organization. 2016.
- APA. Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing. 2013.
- Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol*. 1959;32(1):50-55.
- Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261-276.
- Broich K, Weiergraber M, Hampel H. Biomarkers in clinical trials for neurodegenerative diseases: regulatory perspectives and requirements. *Prog Neurobiol*. 2011;95(4): 498-500. doi: 10.1016/j.pneurobio.2011.09.004.
- Takebe N, McShane L, Conley B. Biomarkers: exceptional responders-discovering predictive biomarkers. *Nat Rev Clin Oncol*. 2015;12(3):132-134. doi: 10.1038/nrclinonc.2015.19.
- Kim ES, Atlas J, Ison G, Ersek JL. Transforming Clinical Trial Eligibility Criteria to Reflect Practical Clinical Application. *Am Soc Clin Oncol Educ Book*. 2016;35: 83-90. doi: 10.14694/EDBK\_155880.
- Razzouk L, Farkouh ME. Imaging outcomes in cardiovascular clinical trials. *Nat Rev Cardiol*. 2009;6(8):524-531. doi: 10.1038/nrcardio.2009.104.
- Erickson BJ, Buckner JC. Imaging in clinical trials. *Cancer Inform*. 2007;4:13-18.
- Hansson L. The Hypertension Optimal Treatment study and the importance of lowering blood pressure. *J Hypertens Suppl*. 1999;17(1): S9-13.
- Masotti CS, Bonfranceschi P, Rusticali G, Rusticali F, Pierangeli A. Left ventricular dynamics after aortic valve replacement: a long-term, combined radionuclide angiographic and ultrasonographic study. *Tex Heart Inst J*. 1992;19(2): 97-106.
- Marcinkiewicz-Siemion M, Ciborowski M, Kretowski A, Musial WJ, Kaminski KA. Metabolomics - A wide-open door to personalized treatment in chronic heart failure? *Int J Cardiol*. 2016;219: 156-63. doi: 10.1016/j.ijcard.2016.06.022.
- Tu K, Campbell NR, Chen ZL, Cauch-Dudek KJ, McAlister FA. Accuracy of administrative databases in identifying patients with hypertension. *Open Med*. 2007;1(1): e18-26.
- Perloff D, Sokolow M, Cowan R. The prognostic value of ambulatory blood pressures. *JAMA*. 1983;249(20):2792-2798.
- Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001;323(7303): 42-46.
- Flecha OD, Douglas de Oliveira DW, Marques LS, Goncalves PF. A commentary on randomized clinical trials: How to produce them with a good level of evidence. *Perspect Clin Res*. 2016;7(2):75-80. doi: 10.4103/2229-3485.179432.
- Budoff M. Triglycerides and Triglyceride-Rich Lipoproteins in the Causal Pathway of Cardiovascular Disease. *Am J Cardiol*. 2016;118(1):138-145. doi:10.1016/j.amjcard.2016.04.004.
- Michos ED, Sibley CT, Baer JT, Blaha MJ, Blumenthal RS. Niacin and statin combination therapy for atherosclerosis regression and prevention of cardiovascular disease events: reconciling the AIM-HIGH (Atherothrombosis Intervention in Metabolic Syndrome With Low HDL/High Triglycerides: Impact on Global Health Outcomes) trial with previous surrogate endpoint trials. *J Am Coll Cardiol*. 2012;59(23): 2058-2064. doi: 10.1016/j.jacc.2012.01.045.
- Lacey L, Bobula J, Rudell K, Alvir J, Leibman C. Quality of Life and Utility Measurement in a Large Clinical Trial Sample of Patients with Mild to Moderate Alzheimer's Disease: Determinants and Level of Changes Observed. *Value Health*. 2015;18(5): 638-645. doi: 10.1016/j.jval.2015.03.1787.