

## Appendix 10. Outcome measures in studies on depressive disorders

The absence of consensus and precision, both in the definition and use of outcome measures, makes it difficult to generalise protocols, compare different studies and extrapolate research to clinical practice<sup>187</sup>.

The need to use uniform outcome measures in research on depressive pathology in order to compare and interpret the results of different research led the *McArthur Foundation Mental Health Research Network on Depression* to create a working group in 1988, which proposed conceptual definitions and operational criteria for identifying the changes caused by various treatments in the course of depressive disorders.

This effort has not been reflected in subsequent research, and there continues to be disparity in the operational criteria used in clinical trials (different scales, different cut-off points, different required durations to meet criteria, etc.). As a result, and in absence of a quantifiable biological marker, there still is no standard that allows comparing the outcomes of treatments for the major depressive disorder<sup>188</sup>.

To measure outcomes, continuous or discontinuous measures can be used, and when we find differences, they can be statistically significant or not, and clinically relevant or not.

### Statistical differences and clinical relevance

There is broad inconsistency in defining and using the terms that designate relevant changes in the course of the depressive illness, and this inconsistency generates difficulties for usefully interpreting, comparing and applying the outcomes of different studies. One review by Prien et al.<sup>187</sup> observed that the method used the most in studies for measuring outcomes was the statistical difference, and up to 32% of studies used statistical significance tests as the sole criteria for assessing the effects of a treatment. However, the statistical test by itself does not indicate if a change has clinical relevance, and it does not define a change point in the course of the depressive illness that can be applied in subsequent research or in clinical decisions. From the clinical point of view, statistical significance does not resolve all the questions marks that have to be answered, wherefore we may find associations that are statistically possible but conceptually sterile.

### Continuous measures versus discontinuous measures

There are some noteworthy scales among the instruments used to assess mood disorders. They are based on the severity and duration of the symptoms, which in turn help to define the state of the patient and his evolution.

The most-used scales are the Hamilton scale, the Beck scale and the Montgomery-Asberg scale. In some trials, the scales are used as continuous measures of outcome (average scores of the group at the end of the trial), although this has the disadvantage of providing results that are difficult to transfer to clinically solid figures, such as absolute differences and NNTs, or the number of patients needed to treat<sup>189</sup>.

In other studies, the measures are frequently dichotomised by establishing a minimum reduction of the symptomatology. Thus, a 50% reduction in the HAM-D 17 or MADRS score is defined as response in various studies<sup>190, 191</sup>. These measures can be arbitrary and have

uncertain clinical relevance. Thus, a 50% reduction in the symptomatology of mild depression is relevant, while in severe depression, the patient remains at a moderate level of disability despite the response, wherefore it would not be clinically important<sup>190</sup>. Patients who have a lower severity at the start are going to have a faster response than those who start with a higher severity score<sup>192</sup>.

Another approach sometimes used in trials on antidepressants is to calculate the proportion of patients with a score below a predefined cut-off value (for example, less than 7 in the Hamilton scale) and to consider these patients to be “recovered”<sup>188</sup>. This approach may be more useful, because it is based on a clinical definition of recovery.

In relation to the application of all these discontinuous measures, a consensus about the cut-off points to be used, their conceptual difference and the operational criteria that would allow elaborating the categories would mean great advances in the design, interpretation and comparison of studies on treatments and on the natural evolution of the illness, in addition to facilitating the development of research and clinical practice guidelines.

Some authors have made their proposals based on the following principles for delimiting them<sup>188</sup>:

~ The definitions are based on observable phenomena such as the severity of the symptoms or the functional state.

~ The change points are independent of whether or not treatment is given and of what treatment is administered.

~ The time to which the concepts refer is independent of the number of episodes that the patient may have had.

Below we will discuss the five necessary concepts for designing the change points in the course of the major depressive illness: recovery, remission, response, relapse and recurrence. These concepts are based fundamentally on operational criteria proposed according to ranges of scores of standardised assessment scales and to the time factor<sup>192, 197</sup>.

For a coherent definition of these terms, it is also necessary to define episode, for which it will in turn be necessary to take into account the number, severity and duration of the symptoms. Thus, an episode is a period longer than N days during which the patient has a sufficient number of symptoms so as to meet the illness diagnostic criteria (DSM-IV/CIE-10)<sup>188</sup>. An episode does not end until the patient recovers.

## Response and partial remission

The period in which considerable improvement is observed, but in which the presence of somewhat more than minimum symptoms persist, is called a partial remission<sup>188</sup>. A partial remission can be spontaneous or due to a treatment.

Response can be understood as the point of onset of the partial remission as a result of a treatment<sup>188, 193</sup>. It has been defined as a decrease of more than 50% in the base score of a standardised scale<sup>194, 195, 197</sup>. It makes sense to establish this category, because it may lead a professional to not increase the therapeutic dose (if the response is due to a treatment) or to not initiate it (if the partial remission appeared spontaneously).

## Full remission

It is a period of time in which the patient is asymptomatic<sup>192, 194</sup>, thereby taking into account that it does not mean the total absence of symptoms. HAM-D 17 below 7, BDI below 8 and duration of less than 6 months are used as operational diagnostic criteria<sup>184</sup>. It has been seen that, using the HAM-D 17, a score of 7 or less establishes the difference between patients with or without depression<sup>193</sup>.

## Recovery

Recovery is remission above a certain period of time<sup>188</sup> in which a patient is asymptomatic and has only one or two mild symptoms. The term is usually used to designate recovery of the episode and not of the illness. The concept “recovery” denotes a return to normalcy or to the usual state, and it is the primary objective of the treatment for acute depressive episodes<sup>160, 187, 194</sup>. The terms recovery and remission have occasionally been used interchangeably<sup>187</sup>.

## Relapse

It is a worsening of an apparently controlled episode, to the point where diagnostic level criteria are reached again, and it occurs during remission and before recovery.

## Recurrence

It is the appearance of a new episode in a recovered patient<sup>188</sup>.

### Operational criteria of outcomes in depression.

		HAM-D 17	BDI
Clinical ranges	Asymptomatic	7 points	8 points
	Completely symptomatic	15 points	15 points
Duration	Episode	2 two weeks completely symptomatic	4 two weeks completely symptomatic
	Full remission	2 and <6 months asymptomatic	3 weeks and < 4 months asymptomatic
	Recovery	6 months asymptomatic	4 months asymptomatic

Source: Frank *et al*<sup>188</sup> and Keller *et al*<sup>193</sup>.