

# IDENTIFYING THE CONCEPTS CONTAINED IN OUTCOME MEASURES OF CLINICAL TRIALS ON DEPRESSIVE DISORDERS USING THE INTERNATIONAL CLASSIFICATION OF FUNCTIONING, DISABILITY AND HEALTH AS A REFERENCE

Thomas T. Brockow,<sup>1</sup> Kathrin Wohlfahrt,<sup>1</sup> Andreas Hillert,<sup>2</sup> Szilvia Geyh,<sup>3</sup> Martin Weigl,<sup>3,4</sup>  
Thomas Franke,<sup>1</sup> Karl Ludwig Resch<sup>1</sup> and Alarcos Cieza<sup>3</sup>

From the <sup>1</sup>Spa Medicine Research Institute, Bad Elster, <sup>2</sup>Centre of Behavioural Medicine, Prien am Chiemsee, <sup>3</sup>ICF Research Branch, WHO FIC Collaborating Center (DIMDI), IMBK, Ludwig-Maximilians-University, Munich, Germany and <sup>4</sup>Department of Physical Medicine and Rehabilitation, University of Munich, Munich, Germany

**Objectives:** First, to systematically identify the concepts contained in outcome measures of trials on depressive disorders using the International Classification of Functioning, Disability and Health (ICF) as a reference. Secondly, to explore differences in the use of ICF categories across different intervention types. Thirdly, to examine which and how often health status measures have been applied in trials on depressive disorders.

**Methods:** Randomized controlled trials between 1991 and 2000 were located in MEDLINE and selected according to predefined criteria. The outcome measures were extracted and the concepts contained in the outcome measures were linked to the ICF.

**Results:** A random sample of 203 (50%) of 406 eligible studies were included. The 5 most used ICF categories (range 88–94%) were *sleep functions (b134)*, *emotional functions (b152)*, *energy and drive functions (b130)*, *thought functions (b160)* and *higher-level cognitive functions (b164)*, all belonging to the *body functions* component. The use of ICF categories did not vary across different intervention types. A total of 126 different health status measures were extracted. The Hamilton Rating Scale for Depression was the most used health status measure applied in 80% of the studies.

**Conclusion:** Concepts about execution of tasks/actions, participation in life situations, and the influence of the environment were under-represented in the outcome assessment of trials on depressive disorders. These observations indicate that most trials were limited in their ability to assess more global individual outcomes.

**Key words:** depressive disorders, outcome assessment, systematic review, ICF.

J Rehabil Med 2004; suppl. 44: 49–55

*Correspondence address:* Alarcos Cieza, Department of Physical Medicine and Rehabilitation, University of Munich, Marchioninstr. 15, DE-81377 Munich, Germany. Tel: +49 89 2180 78216. Fax: +49 89 2180 78230. E-mail: Alarcos.Cieza@med.uni-muenchen.de

## INTRODUCTION

Depressive disorders fulfilling the DSM-III-R (1), DSM-IV (2), or ICD-10 diagnostic criteria (3) for major depressive disorder (MDD) and dysthymic disorder (DD) are psychological conditions with a relatively high prevalence. The point prevalence for MDD in the industrialized world ranged from 2 to 4% (4–6) in the community, 5–10% in primary care, and 10–14% in medical inpatients care (4). Prevalence increases 2–3 times in each setting when only some of the MDD criteria are fulfilled (4, 6).

Patients with depressive disorders have substantial limitations in physical (7), interpersonal (8), social (8), and occupational functioning (9). The extent of functional limitations is often equal to or even greater than those of chronic general medical conditions (8–10). The concept of functioning have been included for diagnosing depressive disorders both in WHO's ICD-10 (3) and in APA's DSM-IV (2). In 2020, depressive disorders will be the second leading cause for disability-adjusted life years (11).

Despite convincing evidence of the relationship between depressive psychopathology and functional limitations (9, 10), concepts of individual function were used less frequently than concepts of body function in the outcome assessment of depressive disorders. Depression scales are among the oldest and best established health status measures, many of them are more than 20 years old (12–14), a few are about 40 years or even older (15–17), and are mainly symptom-based (18). Only a few functional concepts such as “indecisiveness” (14, 16, 17), “social withdrawal” (13, 14, 16), or “work difficulties” (13, 15, 16, 19) are considered in some scales.

Based on the new International Classification of Functioning, Disability and Health (ICF) (20) which was endorsed by the World Health Assembly in May 2001 as a reference or common language of functioning and health, it is now possible to identify, quantify and compare the concepts contained in different outcome measures (21).

The objective of this systematic review therefore was to identify and quantify the concepts contained in the outcome measures of randomized controlled trials (RCTs) for interventions in depressive disorders using the ICF as a reference tool. The specific aims were: (i) to determine the frequency of ICF

categories linked to the concepts contained within the outcome measures; (ii) explore differences in the use of ICF categories across different intervention types; and (iii) examine which and how often health status measures have been applied in trials on depressive disorders.

## METHODS

### Study design

A systematic review was performed in 3 steps: step 1, selection of studies, step 2, outcome measures extraction and, step 3, linkage of the concepts contained within the outcomes measures to the corresponding categories of the ICF. All steps were carried out by 2 independent reviewers.

In step 1, selection of studies, RCTs between the years 1991 and 2000 were located in MEDLINE<sup>®</sup>, Silver Platter, 2000 Edition, using Dickersin's et al. (22) highly precise search strategy (sets 1–8). Thereafter, the Dickersin search was combined with a condition-specific search strategy using the “and” operator.

To locate trials on depressive disorders, the explode-functions for “Depressive-Disorders”, “Depression”, “Dysthymic Disorder”, and “Seasonal-Affective Disorder” with all subheadings and the free text words “depress\*” and “dysthymic” restricted to the abstract and title section were combined using the “or” operator.

All searches were limited to English articles. The abstracts were checked applying general and condition-specific eligibility criteria. For the selected trials the original study reports were ordered and reviewed applying again the same eligibility criteria. The finally included studies entered step II of the review.

MEDLINE<sup>®</sup> was used for study location, because in a direct comparison of MEDLINE<sup>®</sup> and PsychINFO<sup>®</sup> for the years 1998–2000, 89 potentially eligible studies selected by abstract checking in MEDLINE<sup>®</sup> could not be retrieved by PsychINFO<sup>®</sup> opposed to 59 by MEDLINE<sup>®</sup>.

A study met general eligibility, if the study design was a RCT, the experimental intervention had a therapeutic aim, the study was on humans, the report was in English and, if none of the following exclusion criteria were fulfilled: randomized *n* of 1 study, reviews, secondary analyses, psychometric studies, primary prevention studies (healthy population at risk), and mode of action studies. In the case of multiple publications, the paper with the highest impact factor was included.

To identify the appropriate study population condition-specific eligibility criteria were applied. The diagnosis of MDD, DD, or double depression (DD with superimposed major depressive episodes) according to the DSM-III-R (1) or DSM-IV diagnostic criteria (2) or the diagnosis of depressive episode, recurrent depressive disorders, or dysthymia according to the ICD-10 (3) or related terms or classifications with a comparable meaning (e.g. endogenous depression for MDD, chronic minor depression for DD, etc.) has to be reported to describe the study population. Studies on populations with minor depression/minor depressive disorder according to the diagnostic criteria of DSM-III-R/DSM-IV, bipolar disorders, cyclothymic disorders, mood disorder due to a general medical condition (e.g. post-stroke depression (23)) or a substance (e.g. withdrawal from cocaine) and risk patients (e.g. history of depressive illness in first-degree relatives) were excluded.

In step 2, outcome measures extraction, all types of outcome measures and some study characteristics on study population and intervention were extracted.

Outcome measures included clinician-rated and self-rated questionnaires, single item measures on different domains such as nausea, dry mouth, insomnia, dizziness, etc., biochemical test (e.g. liver or kidney function, drug blood level, etc.), and test batteries on memory or intelligence. If the items of a questionnaire were not specified in the publication, we obtained the questionnaire by reference checking, searches in databases or books on health status measures (18, 24), e-mail consultation with the developers of the questionnaire in demand, or internet searches, and then the items were extracted.

Study characteristics included parameters such as subtype of condition (MDD, DD, double depression, or combination of these categories),

age, the country of performance of the trial, and the type of experimental intervention (“drug therapy”, “physical or rehabilitative therapy”, or “different therapeutic modalities”). “Different therapeutic modalities” was chosen in studies with more than 1 experimental group and different intervention types such as comparisons of psychotherapy vs pharmacotherapy with an antidepressant (25, 26).

In the third step, linkage of the concepts contained within the outcome measures to the corresponding categories of the ICF, the concepts contained within the outcome measures were extracted and linked to the most specific ICF category by 2 independent health professionals according to a recently developed set of 10 linking rules (21). If the outcome measure was a laboratory test (e.g. drug blood level of an antidepressant) the goal (compliance) rather than the concept was extracted. Questionnaires not available in English language and those evaluating the effect of an intervention on other units than patients were excluded from linkage, for example the Camberwell Family Interview assessing emotional function in family members of psychiatric patients (27). Concepts of outcome measures that could not be linked to the ICF were documented and classified in 2 ways: (i) If a concept of an outcome measure was not sufficiently specified to make a decision which ICF category the concept should be linked to, the “not definable” option was chosen (linking rule 9). To give an example, unspecified concepts such as “productivity”, “functioning”, “enjoying life”, “health”, or “somatic complaints” were considered not to be definable for linking. (ii) If a concept of an outcome measure was not represented by the ICF, the option “not covered” was chosen (linking rule 10). To give an example, concepts such as “would kill myself if I had a chance” extracted from the Beck Depression Inventory (16) or “wanted to do the opposite of what your relatives wanted ...” extracted from the Social Adjustment Scale-Self Report (28) were considered not to be covered by the ICF.

Consensus between the 2 health professionals was used to decide which ICF category should be linked to each item/concept of the questionnaires. To resolve disagreements between the 2 health professionals concerning the selected categories, a third person trained in the linking rules was consulted. In a discussion led by the third person, the 2 health professionals that linked the item stated their pros and cons for the linking of the concept under consideration to a specific ICF category. Based on these statements, the third person made an informed decision.

To control the plausibility of the linkage procedure the concepts of the outcome measures assigned to the same single ICF category were analysed (e.g. the concepts “feeling inadequate, inferior, worthless, no good” (19), “feeling disappointed in myself” (16), “feeling quite guilty” (16), “feeling miserable/comfortable” (29), “continuous feeling of inner tension or intermittent panic” (12), which were linked to the ICF category *emotional functions* (b152)).

### Analyses

Descriptive statistics were used to examine the frequency of ICF categories linked to the concepts contained in the outcome measures. Large-scale cross tables generated from a SQL-database (SQL-Server 2000) were thereby analysed. If the same ICF category was assigned repeatedly in a study, the category was counted only once.

ICF categories are presented on the second level. If a concept of an outcome measure was linked to a third or fourth level ICF category, the overlying second level category was considered. The ICF is organized in a hierarchical scheme, so that the lower-level category shares the attributes of the higher-level category (20). Only ICF categories with a frequency equal or greater than 10% are shown (pre-set frequency).

## RESULTS

In step 1, 2434 studies were located by the search strategy, 834 studies were preliminarily selected by abstract checking, 406 studies fulfil the eligibility criteria by screening the respective original papers. A computer-generated random sample of 203 studies (50%) was drawn and included into the review.

In step 2, 126 different questionnaires (different versions and subscales of a questionnaire were considered as 1 questionnaire)

were extracted. Among them we identified 19 condition-specific, 102 domain-specific, and 5 generic questionnaires. Domain-specific questionnaire considered dimensions such as anxiety (30), beliefs and attitudes (31), coping (32), functional status (33), life satisfaction (34), mania (35), personality (36), self-esteem (37), side-effects of psychotropics (38), or social adjustment (28). At least 1 health status questionnaire per trial was used in 197 or 97% of the studies. The few remaining trials without health status assessment considered physiological or laboratory outcomes such as physiological sleep parameters, neuroendocrine parameters (39) or heart rate parameters (40). Type and frequency of the 20 most used questionnaires are shown in Table I. The Hamilton Rating Scale for Depression (15) was applied in 80% of the studies representing the most used health status measure. Domains of single-item measures with a frequency of at least 20% were nausea, dry mouth, headache, dizziness, insomnia, constipation, heart rate, sweating, blood pressure, diarrhoea, weight, somnolence, agitation, compliance (drug blood level), and anxiety, some of them were nearly exclusively assessed in the context of adverse effects evaluation.

In 177 or 87% of studies the study population consisted of patients with MDD, in 6 or 3% of studies the patients had DD and in 1 study the patients had double depression. The remaining study populations consisted of patients with different condition-subtypes such as MDD and DD or MDD and double depression.

Most trials were conducted in the USA (44%), followed by multi-national trials (9%) and trials conducted in France and UK, each with a frequency of 5%. Drug therapy was the most frequently used intervention type with a prevalence of 77%

(157 studies), followed by physical or rehabilitative therapy with a prevalence of 16% (32 studies). The remaining 14 trials (7%) dealt with 2 experimental interventions and different intervention types (“different therapeutic modalities”). “Physical or rehabilitative therapy” included psychotherapy (11 studies), light therapy (7 studies), electroconvulsive therapy (7 studies), managed care (6 studies), and acupuncture (1 study). “Managed care” was assigned to “physical or rehabilitative therapy” because this category include complex organized treatment programs (41–44) as usual applied in rehabilitation medicine.

In step 3, a total of 24 114 concepts were extracted from the outcome measures. Of the 24 114 concepts extracted, a third person was ordered 497 times to resolve differences between the 2 health professionals in their assignment of the ICF categories. 21 561 or 89% of concepts could be linked to the ICF, 1660 or 7% of concepts were considered not to be sufficiently specified for an assignment to the ICF (“not definable option”), and 893 or 4% of concepts were considered to be not covered by the ICF. Of these, 717 concepts describing personal characteristics, 114 referred to “locus of control”, 219 to “coping”, 20 to “personality”, 56 to “life satisfaction”, and 308 to “self-esteem”. 86% ( $n = 18\,599$ ) of the assignable concepts were linked to the component “body functions”, only 12 concepts to the component *body structures*, 12% ( $n = 2483$ ) to the component *activities and participation*, and 2% ( $n = 460$ ) to the component *environmental factors*.

The concepts contained in the outcome measures were linked to 454 different ICF categories. Of these there were 250 categories which were linked to the component *body functions*, 7 to the component “body structures”, 173 to component *activities and participation* and 24 to the component *environmental factors*.

Five questionnaires were excluded from linking. In 3 cases the questionnaire was not available in English, in 1 case the questionnaire evaluates family members, and in 1 case only the name of the questionnaire was mentioned without a reference given or an item-description in the paper.

Table II shows the relative frequency in percentage of ICF categories linked to the concepts contained in the outcome measures as well as differences in the use of ICF categories across different intervention types. The 5 most used ICF categories were *sleep functions* (b134), *emotional functions* (b152), *energy and drive functions* (b130), *thought functions* (b160) and *higher-level cognitive functions* (b164). All these categories were highly prevalent reaching frequencies varying between 88% and 94%, all belong to the *body functions* component, and all represented concepts contained in the 2 most frequently used depression scales of this review, the Hamilton Rating Scale for Depression (15) and the Montgomery-Asberg Depression Rating Scale (12) (Table I). The clinical contents underlying these categories were the following: (i) Concepts such as “wake up much earlier than I need” (13), “periods of wakefulness” (45), “experience of reduced duration or depth of sleep” (12), or “waking up more tired in the morning” (16) were linked to *sleep functions*. (ii) *Emotional*

Table I. Type and frequency of the 20 most used health status measures

Outcome measure	Type	n (%)
Hamilton Rating Scale for Depression*	cs	162 (80)
Clinical Global Impressions	ds	79 (39)
Montgomery-Asberg Depression Rating Scale*	cs	59 (29)
Beck Depression Inventory	cs	30 (15)
Hamilton Anxiety Scale	ds	19 (9)
Symptom Check List*	ds	19 (9)
UKU-Side Effects Rating Scale*	ds	10 (5)
Mini-Mental State Examination*	ds	9 (4)
Quality of Life Enjoyment and Satisfaction Questionnaire	ds	9 (4)
Zung Self-Rating Depression Scale	cs	9 (4)
Covi Anxiety scale	ds	8 (4)
Social Adjustment Scale*	ds	8 (4)
Short Form Health Survey*	g	8 (4)
Bech-Rafaelsen Melancholia Scale	cs	7 (3)
Adjective Mood Scale	ds	5 (3)
Inventory for Depressive Symptomatology	cs	5 (3)
Longitudinal Interval Follow-Up Evaluation	ds	5 (3)
Profile of Mood States	ds	5 (3)
Raskin Depression Scale	cs	5 (3)
Wechsler Adult Intelligence Scale*	ds	5 (3)

\* Different versions or subscales. cs = condition-specific, ds = domain-specific, g = generic.

Table II. Relative frequency in percentage of ICF categories linked to the concepts contained in the outcome measures

ICF code	ICF category title	AT (n = 203)	DT (n = 157)	PRT (n = 32)	DMT (n = 14)
b110	Consciousness functions	86	90	59	92
b126	Temperament and personality functions	82	82	75	92
b130	Energy and drive functions	92	94	75	100
b134	Sleep functions	94	96	78	100
b140	Attention functions	48	54	28	21
b144	Memory functions	26	29	15	14
b147	Psychomotor functions	84	87	62	92
b152	Emotional functions	93	94	81	92
b156	Perceptual functions	50	54	25	57
b160	Thought functions	91	94	75	92
b164	Higher-level cognitive functions	88	94	53	92
b167	Mental functions of language	38	43	12	35
b180	Experience of self and time functions	24	17	37	64
b210	Seeing functions	22	26	6	14
b215	Functions of structures adjoining the eye	22	25	6	14
b240	Sensations associated with hearing and vestibular function	41	49	9	28
b270	Sensory functions related to temperature and other stimuli	20	22	9	14
b279	Additional sensory functions, other specified and unspecified	43	49	15	42
b280	Sensation of pain	84	88	56	92
b330	Fluency and rhythm of speech functions	50	52	34	57
b340	Alternative vocalization functions	44	50	12	42
b410	Heart functions	56	64	15	50
b420	Blood pressure functions	28	34	3	7
b430	Haematological system functions	14	17	3	
b440	Respiration functions	78	84	40	100
b455	Exercise tolerance functions	17	18	6	21
b460	Sensations associated with cardiovascular and respiratory functions	62	65	37	71
b510	Ingestion functions	70	77	31	71
b515	Digestive functions	37	42	9	35
b525	Defecation functions	84	89	53	100
b530	Weight maintenance functions	82	85	59	92
b535	Sensations associated with the digestive system	86	90	59	92
b550	Thermoregulatory functions	11	12	9	
b620	Urination functions	50	57	15	50
b640	Sexual functions	82	86	53	92
b650	Menstruation functions	14	17	3	7
b699	Genitourinary and reproductive functions, unspecified	42	43	31	57
b735	Muscle tone functions	15	18	3	7
b765	Involuntary movement functions	34	38	12	21
b780	Sensations related to muscles and movement functions	52	54	37	57
b830	Other functions of the skin	37	44	12	14
d115	Listening	38	43	12	35
d177	Making decisions	82	84	62	92
d230	Carrying out daily routine	38	43	18	21
d335	Producing non-verbal messages	23	17	31	64
d550	Eating	13	15	6	
d570	Looking after one's health	62	60	53	92
d599	Self-care, unspecified	16	16	9	28
d640	Doing housework	11	7	18	35
d720	Complex interpersonal interactions	18	12	25	64
d760	Family relationships	10	8	9	35
d770	Intimate relationships	10	10	9	14
d799	Interpersonal interactions and relationships, unspecified	19	12	34	57
d859	Work and employment, other specified and unspecified	81	83	65	92
d920	Recreation and leisure	53	56	28	64
e399	Support and relationships, unspecified	44	50	15	35

AT = all trials; DT = drug therapy; PRT = physical or rehabilitative therapy; DTM = different therapeutic modalities.

functions included concepts such as “feeling no interest in things” (46), “depressed mood” (12), “inability to feel” (12), “feeling miserable” (29), “feeling guilty” (16), “feeling hopeless” (35), or “feeling worthless” (19). (iii) *Energy and drive functions* included concepts such as “loss of energy” (15),

“difficulty in getting started” (12), “listless” (47), “feeling driven to overeat” (48), or “loss of appetite” (15). (iv) *Thought functions* included concepts such as “flight of thoughts” (35), “slowness of thoughts” (15), “coherence of thoughts” (35), or “difficulties in collecting or sustaining thoughts” (12). (v)

Higher-level cognitive functions included concepts such as “impairment in judgement” (49), “loss of insight” (15), “preparation for suicide” (12), or “solving problems” (33).

The most often used ICF category within the component *activities and participation* was *making decisions* (d177), followed by *work and employment*, *other specified and unspecified* (d859) with a frequency of 82% and 81%, respectively. The clinical contents underlying “making decisions” included concepts in the meaning of *difficulty in making decisions/indecisiveness*. Concepts such as “being effective in one’s job” (19), “being ashamed of how you do your work” (28), “unable to work” (15), “can work about as well as before” (16), “being able to do your work” (28) were linked to the category *work and employment, other specified and unspecified*. Other categories with a frequency of at least 50% were *looking after one’s health* (d570) and “recreation and leisure” (d920). *Looking after one’s health* included concepts such as “drug compliance”, “adherence to treatment”, “following medical advice”. *Recreation and leisure* included concepts such as “participating in sports”, “loss of interest in hobbies”, “difficulty in relaxing”, “socializing”, mainly in the sense of visiting friends, relatives or neighbours.

Only 1 category of the *environmental factors* component reached a frequency of at least 10%. This was the case for *support and relationships, unspecified* (e399) including concepts such as “requiring urging, guidance, and reassurance” (15), “needing help with bathing and dressing” (13) or “needing supervision to prevent hurting self or others” (49).

The ICF categories *temperament and personality functions, energy and drive functions, sleep functions, emotional functions, and thought functions* were used in more than 75% of all drug-therapy ( $n = 157$ ), physical- or rehabilitative-therapy studies ( $n = 32$ ), and studies with different therapeutic modalities ( $n = 14$ ).

All chapters (one-level classification of the ICF) were represented by at least 1 ICF category except for *structures of the nervous system* (s1), *the eye, the ear and related structure* (s2) and *structures involved in voice and speech* (s3).

## DISCUSSION

Using the ICF as a reference it was possible to identify and quantify the concepts within the outcome measures used in RCTs for interventions in depressive disorders. Most concepts within the outcome measures could be linked to the ICF and those who could not be linked were mostly not specified in enough detail for an assignment.

Only a very small portion of concepts was considered not covered by the ICF. In these cases the content of the concepts did not lie in the defined universe of the ICF. For example, health status measures on dimensions such as “locus of control” (31), coping (32), “personality” (36), life satisfaction (34), or self-esteem (37) include personal concepts that are not covered by the current ICF and could therefore not be linked.

The data show, that foremostly concepts about psychological

but also on physiological functions were more often represented in outcome measures of trials on depressive disorders compared to concepts dealing with executions of tasks or actions or the participation of an individual in life situations. ICF categories such as *self-care* (d599), *housework* (d640), *family* (d760) or intimate relationships (d770) reached only frequencies just above the pre-set 10% level. Additionally, *environmental factors* with an impact on functioning in depressive disorders were hardly addressed in outcome measures of trials on depressive disorders. ICF categories such as *time-related changes* (e245), *light* (e240), *individual attitudes of family member* (e410) and friends (e420) or *individual attitudes of health professionals* (e355) did not even reach the pre-set frequency level.

Nearly all ICF categories showed higher frequencies in drug trials than in trials on physical or rehabilitative therapy (non-pharmacological trials). In drug trials different health status measures were used more often simultaneously (data not shown). A reason might be, that drug trials are often sponsored by pharmaceutical companies enabling a greater methodological effort.

Surprisingly, different patterns in the use of ICF categories in drug trials compared with non-pharmacological trials could not be found. The reason might be, that independent of the intervention type nearly all therapeutic options aimed primarily to influence depressive symptomatology.

Our results reflect the “state of art” of concepts used in outcome measures of trials on depressive disorders over the last decade. Secular trends in the use of outcome measures or ICF categories could not be confirmed (data not shown). From a conceptual point of view, the results might be “biased” by a reduced use of functional concepts and *environmental factors* with an impact on individual life of patients with depressive disorders. More the biological than the social or environmental perspective of the disease was represented in outcome measures of studies on depressive disorders. As a classification system encompassing all aspects of human health, the ICF has the capacity to expand the outcomes commonly measured by psychiatric studies on depressive disorders. It is beyond the scope of this paper to discuss in detail whether the outcome measures used in the studies and therefore the concepts linked to the ICF are appropriate for specific study questions or whether or not they adequately reflect the burden of disease in patients with depressive disorders. However, our findings indicate a need to define “what should be measured” in clinical trials to allow for a more comprehensive and comparable comparison of patient populations across studies and interventions.

Due to limited personal and time resources we had to work with some methodological limitations such as the solely use of MEDLINE<sup>®</sup>, RCTs, studies published in English and outcome measures available in English. Furthermore, we had to draw a random sample of the eligible studies selected. The limitation of the review to RCTs might have introduced an overrepresentation of biological aspects in the outcome assessment of depressive

disorders. Future investigations should review longitudinal observational studies of those with depressive disorders to determine whether they include more outcomes from a social perspective.

In conclusion, the ICF provides a useful reference to identify and quantify the concepts within the outcome measures used in RCTs for interventions on depressive disorders. On the other hand, studies on the outcomes of depressive disorder trials were often limited by the conceptual limitations of their measures.

## REFERENCES

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (3rd edn, revised) (DSM-III-R). Washington: American Psychiatric Association; 1987.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders DSM-IV (4th edn). Washington: American Psychiatric Association; 1994.
- World Health Organization. International Statistical Classification of Diseases and Related Health Problem, Tenth Revision. Geneva: World Health Organization; 1992.
- Katon W, Schulberg H. Epidemiology of depression in primary care. *Gen Hosp Psychiatry* 1992; 14: 237–247.
- Girling DM, Barkley C, Paykel ES, Gehlhaar E, Brayne C, Gill C, et al. The prevalence of depression in a cohort of the very elderly. *J Affect Disord* 1995; 34: 319–329.
- Roberts RE, Lewinsohn PM, Seeley JR. Symptoms of DSM-III-R major depression in adolescence: evidence from an epidemiological survey. *J Am Acad Child Adolesc Psychiatry* 1995; 34: 1608–1617.
- Wells KB, Golding JM, Burnam MA. Psychiatric disorder and limitations in physical functioning in a sample of the Los Angeles general population. *Am J Psychiatry* 1988; 145: 712–717.
- Fredman L, Weissman MM, Leaf PJ, Bruce ML. Social functioning in community residents with depression and other psychiatric disorders: results of the New Haven Epidemiologic Catchment Area Study. *J Affect Disord* 1988; 15: 103–112.
- Ormel J, VonKorff M, Ustun TB, Pini S, Korten A, Oldehinkel T. Common mental disorders and disability across cultures. Results from the WHO Collaborative Study on Psychological Problems in General Health Care. *JAMA* 1994; 272: 1741–1748.
- Spitzer RL, Kroenke K, Linzer M, Hahn SR, Williams JB, deGruy FV, et al. Health-related quality of life in primary care patients with mental disorders. Results from the PRIME-MD 1000 Study. *JAMA* 1995; 274: 1511–1517.
- Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: Global Burden of Disease Study. *Lancet* 1997; 349: 1498–1504.
- Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979; 134: 382–389.
- Carroll BJ, Feinberg M, Smouse PE, Rawson SG, Greden JF. The Carroll rating scale for depression. I. Development, reliability and validation. *Br J Psychiatry* 1981; 138: 194–200.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey MB, et al. Development and validation of a geriatric depression screening scale: a preliminary report. *J Psychiatr Res* 1983; 17: 37–49.
- Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960; 23: 56–62.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatr* 1961; 4: 561–571.
- Zung WWK. A self-rating depression scale. *Arch Gen Psychiatry* 1965; 12: 63–67.
- McDowell I, Newell C. Measuring health. A guide to rating scales and questionnaires. 2nd edn. New York: Oxford University Press; 1996.
- Mason BJ, Kocsis JH, Leon AC, Thompson S, Frances AJ, Morgan RO, et al. Measurement of severity and treatment response in dysthymia. *Psychiatr Annals* 1993; 23: 625–631.
- WHO. ICF-International Classification of Functioning, Disability and Health Geneva: World Health Organisation; 2001.
- Cieza A, Brockow T, Ewert T, Amman E, Kollerits B, Chatterji S, et al. Linking health-status measurements to the international classification of functioning, disability and health. *J Rehabil Med* 2002; 34: 205–210.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286–1291.
- Andersen G, Vestergaard K, Lauritzen L. Effective treatment of poststroke depression with the selective serotonin reuptake inhibitor citalopram. *Stroke* 1994; 25 (6): 1099–1104.
- Salek S. Compendium of quality of life instruments. Chichester: Wiley; 1999.
- Mynors Wallis LM, Gath DH, Lloyd Thomas AR, Tomlinson D. Randomised controlled trial comparing problem solving treatment with amitriptyline and placebo for major depression in primary care. *BMJ* 1995; 310: 441–445.
- Brown C, Schulberg HC, Madonia MJ, Shear MK, Houck PR. Treatment outcomes for primary care patients with major depression and lifetime anxiety disorders. *Am J Psychiatry* 1996; 153: 1293–1300.
- Vaughn C, Leff JP. The measurement of expressed emotion in families of psychiatric patients. *Br J Social Clin Psychol* 1976; 15: 157–165.
- Weissman MM, Bothwell S. Assessment of social adjustment by patient self-report. *Arch Gen Psychiatry* 1976; 33: 1111–1115.
- von Zerssen D. Clinical self-rating scales of the Munich Psychiatric Information System. In: Satorius N, Ban TA, eds. Assessment of depression. Berlin: Springer-Verlag; 1986, p. 270–303.
- Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol* 1959; 50–55.
- Weissman AN. The Dysfunctional Attitudes Scale: a validation study. *Dissertation Abstracts Int* 1979; 60: 1389B–1390B.
- Beckham EE, Adams RL. Coping behavior in depression: report on a new scale. *Behav Res Ther* 1984; 22: 71–75.
- Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil* 1993; 74: 531–536.
- Neugarten BL, Havighurst RJ, Tobin SS. The measurement of life satisfaction. *J Gerontol* 1961; 16: 141–142.
- Bech P. Rating scales for affective disorders: their validity and consistency. *Acta Psychiatr Scand Suppl* 1981; 295: 1–101.
- Cloninger CR. The Tridimensional Personality Questionnaire, Version IV. St. Louis: Washington University School of Medicine, Department of Psychiatry; 1987.
- Rosenberg M. Society and the adolescent self-image. Princeton, NJ: Princeton University; 1965.
- Lingaerde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K. The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr Scand* 1987; 76 (suppl 334): 1–100.
- Poland RE, McCracken JT, Lutchmansingh P, Tondo L. Effects of low-dose dexamethasone on sleep EEG patterns, plasma cortisol, and the TSH response to TRH in major depression. *Pharmacopsychiatry* 1993; 26: 79–83.
- Rechlin T, Weis M, Claus D. Heart rate variability in depressed patients and differential effects of paroxetine and amitriptyline on cardiovascular autonomic functions. *Pharmacopsychiatry* 1994; 27: 124–128.
- Katon W, Robinson P, Von Korff M, Lin E, Bush T, Ludman E, et al. A multifaceted intervention to improve treatment of depression in primary care. *Arch Gen Psychiatry* 1996; 53: 924–932.
- Katon W, Von Korff M, Lin E, Simon G, Walker E, Unutzer J, et al. Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Arch Gen Psychiatry* 1999; 56: 1109–1115.
- Llewellyn Jones RH, Baikie KA, Smithers H, Cohen J, Snowdon J, Tennant CC. Multifaceted shared care intervention for late life depression in residential care: randomised controlled trial. *BMJ* 1999; 319: 676–682.

44. Katzelnick DJ, Simon GE, Pearson SD, Manning WG, Helstad CP, Henk HJ, et al. Randomized trial of a depression management program in high utilizers of medical care. *Arch Fam Med* 2000; 9: 345–351.
45. Parott AC, Hindmarch I. The Leeds Sleep Evaluation Questionnaire in psychopharmacological investigations-review. *Psychopharmacology* 1980; 71: 173–179.
46. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. *Psychol Med* 1983; 13: 595–605.
47. McNair DM, Lorr M, Droppleman L. The manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service; 1971.
48. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory for Depressive Symptomatology (IDS): psychometric properties. *Psychol Med* 1996; 26: 477–486.
49. Endicott J, Spitzer RL, Fleiss JL, Cohen J. The global assessment scale: a procedure for measuring overall severity of psychiatric disturbances. *Arch Gen Psychiatry* 1976; 22: 766–771.