Review

Screening for bipolar spectrum disorders: A comprehensive meta-analysis of accuracy studies

André F. Carvalho [a,*], Yemisi Takwoingi [b], Paulo Marcelo G. Sales [a], Joanna K. Soczynska [c], Cristiano A. Köhler [d], Thiago H. Freitas [a], João Quevedo [e,f], Thomas N. Hyphantis [g], Roger S. McIntyre [c,h], Eduard Vieta [l]

[a] Translational Psychiatry Research Group, Faculty of Medicine, Federal University of Ceara, Fortaleza, CE, Brazil
[b] Department of Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK
[c] Mood Disorders Psychopharmacology Unit, University of Toronto, Toronto, ON, Canada
[d] Memory Research Laboratory, Brain Institute (ICe), Federal University of Rio Grande do Norte (UFRN), Natal, RN, Brazil
[e] Laboratory of Neurosciences, Graduate Program in Health Sciences, Health Sciences Unit, University of Southern Santa Catarina, Criciúma, SC, Brazil
[f] Center for Experimental Models in Psychiatry, Department of Psychiatry and Behavioral Sciences, The University of Texas Medical School at Houston, Houston, TX, USA
[g] Department of Psychiatry, University of Ioaninna, Ioaninna, Greece
[h] Departments of Psychiatry and Pharmacology, University of Toronto, Toronto, ON, Canada
[l] Bipolar Disorders Unit, Clinical Institute of Neurosciences, Hospital Clinic, IDIBAPS, University of Barcelona, CIBERSAM, Barcelona, Catalonia, Spain

ARTICLE INFO

ABSTRACT

Background: Bipolar spectrum disorders are frequently under-recognized and/or misdiagnosed in various settings. Several influential publications recommend the routine screening of bipolar disorder. A systematic review and meta-analysis of accuracy studies for the bipolar spectrum diagnostic scale (BSDS), the hypomania checklist (HCL-32) and the mood disorder questionnaire (MDQ) were performed.

Methods: The Pubmed, EMBASE, Cochrane, PsycINFO and SCOPUS databases were searched. Studies were included if the accuracy properties of the screening measures were determined against a DSM or ICD-10 structured diagnostic interview. The QUADAS-2 tool was used to rate bias.

Results: Fifty three original studies met inclusion criteria (N=21,542). At recommended cutoffs, summary sensitivities were 81%, 66% and 69%, while specificities were 67%, 79% and 86% for the HCL-32, MDQ, and BSDS in psychiatric services, respectively. The HCL-32 was more accurate than the MDQ for the detection of type II bipolar disorder in mental health care centers (P=0.018). At a cutoff of 7, the MDQ had a summary sensitivity of 43% and a summary specificity of 95% for detection of bipolar disorder in primary care or general population settings.

Limitations: Most studies were performed in mental health care settings. Several included studies had a high risk of bias.

Conclusions: Although accuracy properties of the three screening instruments did not consistently differ in mental health care services, the HCL-32 was more accurate than the MDQ for the detection of type II BD. More studies in other settings (for example, in primary care) are necessary.

© 2014 Elsevier B.V. All rights reserved.

Contents

* Corresponding author. Tel./fax: +558532617227.
  E-mail address: andrefc7@terra.com.br (A.F. Carvalho).

1. Introduction

The diagnosis of bipolar disorders in most circumstances is not straightforward and requires a careful assessment of its longitudinal course. Almost three-third of individuals with bipolar disorders report having received a misdiagnosis at least once, while a proper diagnosis takes on average 10 years from the initiation of affective symptoms (Drancourt et al., 2013; Lish et al., 1994). Evidences also indicate that bipolar disorder is prevalent and frequently under-recognized in primary care (Cerimele et al., 2014; Culpepper, 2014). Furthermore, depressive symptoms and episodes more frequently predominate in the longitudinal course of bipolar disorders (Judd et al., 2003); this results in a significant proportion of individuals with BD being misdiagnosed as having unipolar depression (Hirschfeld and Vornik, 2004). These patients misdiagnosed as having major depressive disorder are more likely to receive antidepressant monotherapy (Matza et al., 2005) which may result in manic switches, cycle acceleration, and possibly heightened suicidality (Bond et al., 2008; Ghaemi et al., 2004; Undurraga et al., 2012).

The use of self-report screening instruments for bipolar disorder that are both time- and cost-effective may aid in the timely recognition of this illness. In the last several years four self-report questionnaires have been developed to screen for bipolar spectrum disorders, namely the mood disorders questionnaire (MDQ) (Hirschfeld et al., 2000), the bipolar spectrum diagnostic scale (BSDS) (Ghaemi et al., 2005), the hypomanic checklist (HCL-32) (Angst et al., 2005) and the mood swings questionnaire/survey (MSQ/MSS) (Parker et al., 2008; Parker et al., 2006). These screening tools are readily available for clinical use. Briefly, the MDQ screens for a lifetime history of (hypo) mania with 13 yes/no questions reflecting DSM-IV criteria. These questions are followed by a single yes/no question asking whether the symptoms clustered in the same period. The final question evaluates the level of impairment resulting from the symptoms. The MDQ developers recommended a cut-off score of seven endorsed symptoms that co-occurred and caused at least moderate impairment. (Hirschfeld et al., 2000) The BSDS consists of two parts. The first part is a paragraph containing 19 statements describing several manifestations of bipolar disorder. Each affirmatively checked sentence is counted as 1 point. The second part of the BSDS is a single multiple-choice question asking respondents how well the paragraph describes their behavior (very well or almost perfect – 6 points; fairly well – 4 points; to some degree but not in most respects – 2 points; not really at all – 0 points). In the initial

study, a cut-off point of 13 yielded the best balance of sensitivity/specificity (Ghaemi et al., 2005). In the HCL-32, after a brief introduction, the respondent is instructed to think of a period when he/she was in a "high" state and answer 32 yes/no questions about their mood and behavior during that period. Each 'yes' response is scored 1, whereas each 'no' answer is scored as 0. In the initial study, the authors suggested a cut-off score of 14 (Angst et al., 2005).

Notwithstanding several influential publications recommend the routine screening in clinical practice (Anderson et al., 2012; Chessick and Dimidjian, 2010; Frye, 2011; Loganathan et al., 2010), concerns have been raised regarding the validity and applicability of these screening tools (Phelps and Ghaemi, 2006; Zimmerman, 2012; Zimmerman et al., 2010). Phelps and Ghaemi (2006) used previously published data on sensitivity and specificity of the MDQ and BSDS to estimate positive and negative predictive values at varying prevalence levels using Bayesian statistical concepts. At lower prevalence or low prior clinical probability (for example, in primary care), high negative predictive values were verified indicating that both instruments effectively rule out bipolar disorders. However, in these contexts the positive predictive value significantly dropped resulting in a higher number of 'false positives'.

The BSDS, HCL-32 and MDQ have been the most extensively investigated screening tools for bipolar spectrum disorders in accuracy studies and epidemiological surveys. Therefore, the overarching aims of this report were to perform a systematic review and meta-analysis to evaluate and compare the diagnostic accuracy of these three screening tools in different clinical settings. Our secondary objective was to investigate the effect of pre-defined potential sources of heterogeneity on estimates of test performance.

2. Method

2.1. Search strategy and selection of studies

Studies were identified through three methods. First, we conducted comprehensive computerized literature searches in five bibliographical databases – MEDLINE, EMBASE, Cochrane CENTRAL, PsycInfo and SCOPUS – from inception to January 9th, 2014. Search strings are provided in the Supplementary material S1 that accompanies the online edition of this article. Second, this search strategy was augmented through tracking citations of included articles in Google Scholar. Finally, references of relevant reviews were examined