Journal of Affective Disorders 156 (2014) 36–45

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Review

The status of sleep abnormalities as a diagnostic test for major depressive disorder

C.L. Arfken [a,*], A. Joseph [a], G.R. Sandhu [a], T. Roehrs [b], A.B. Douglass [c], N.N. Boutros [a]

[a] Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI 48207, USA
[b] Henry Ford Sleep Disorders & Research Center, Henry Ford Health System & Department of Psychiatry and Behavioral Neurosciences, Wayne State University, Detroit, MI 48207, USA
[c] University of Ottawa, Department of Psychiatry and Royal Ottawa Mental Health Center, Ottawa, ON, Canada

ARTICLE INFO

Article history:
Received 13 September 2013
Received in revised form
21 October 2013
Accepted 1 December 2013
Available online 16 December 2013

Keywords:
Diagnosis
Major depressive disorder
Polysomnography
REM sleep
Slow-wave sleep
Sleep efficiency

ABSTRACT

Background: Psychiatry lags other fields in development of diagnostic tests.
Methods: A literature review and meta-analysis was conducted to ascertain if polysomnographic abnormalities (REM density, REM latency, sleep efficiency, slow wave sleep, stage 1 and stage 2 sleep) warrant additional effort to develop them into a clinical diagnostic test for major depressive disorder (MDD). The 31 publications meeting inclusion criteria were then classified into one of three progressive steps using guidelines for evaluating the clinical usefulness of a diagnostic test.
Results: Most of the abnormalities found in MDD patients, when compared to healthy controls, occurred in the expected direction with moderate effect sizes but with substantial publication bias and heterogeneity. Eleven studies compared abnormalities in MDD to other psychiatric disorders (step 2a), and four studies provided data on the sensitivity or specificity of the findings in differentiating among the psychiatric disorders that frequently appear on the same differential diagnostic list as MDD (step 2b). No multicenter trial has been conducted prospectively to test the clinical utility of the diagnostic test (step 3).
Limitations: Only published articles in the English language were used.
Conclusions: Sleep studies for the detection of MDD appear replicable with a moderate effect size. However, additional step 1 studies are needed to define the sensitivity and specificity. The heterogeneity of sleep recording, scoring techniques, and MDD must also be addressed.

Contents

1. Background . . . 36
2. Methods . . . 38
3. Results . . . 40
4. Discussion . . . 41
5. Limitations . . . 43
Role of funding source . . . 44
Conflict of interest . . . 44
Acknowledgements . . . 44
References . . . 44

1. Background

Laboratory tests are an essential part of the practice of modern medicine. They can be used to confirm a diagnosis, provide supportive evidence for one diagnosis versus another, and rule out other diagnoses. The last 50 years of biological research into the

* Correspondence to: 3901 Chrysler Drive, Ste 1B, Detroit, MI 48201, USA. Tel.: +1 313 993 3490; fax: +1 313 577 8823.
E-mail address: carfken@med.wayne.edu (C.L. Arfken).

0165-0327/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jad.2013.12.007

pathophysiology of psychiatric disorders have yielded a number of highly replicable abnormalities. These abnormalities have the potential for being developed into clinically useful diagnostic tests. Although psychiatrists do use laboratory tests to rule out general medical conditions as causes for mental disorders, they are rarely used to differentiate between primary psychiatric disorders. As a field, psychiatry has lagged behind other medical specialties in developing laboratory tests according to well-defined epidemiological principles.

Laboratory tests in psychiatry tend either not to be developed into diagnostic tools (e.g., P300 evoked response in schizophrenia) or to be disseminated before their validity is fully documented (e.g., Quantified electroencephalography (Q-EEG) (Nuwer, 1989). The premature release of such tests leads to disappointment of the medical community and premature abandonment of the test. Moreover, when tests are used out of context, they may hinder the diagnostic and treatment process and increase the cost of management unnecessarily (Steffens and Krishnan, 2003).

On the other hand, an American Psychiatric Association task force report (American Psychiatric Association (APA), 1991) indicated that QEEG was particularly useful in detecting slow wave abnormalities and concluded that clinical replication and sharing of normative and patient databases were necessary for the advancement of this field. They further stated that standards for training and use of the technology in psychiatry were urgently needed. Unfortunately, the situation has not changed appreciably since that publication.

The development of ancillary diagnostic procedures is important to help the field move forward, as diagnosis in psychiatry remains the major limiting step in biological research and treatment studies (van Praag, 1997). We have proposed a four-step process for developing laboratory-based diagnostic tests in psychiatry (Boutros and Struve, 2002; Boutros and Arfken, 2007; Arfken et al., 2009). The four-step approach is based on published guidelines for deciding the clinical usefulness of diagnostic tests (Sackett et al., 1991) and the criteria specified by the Standard for Reporting Diagnostic tests (STARD) (Bossuyt et al., 2003; Bruns, 2003). Based on significant feedback from journal reviewers, we have now modified the approach into a 3-step progression with steps 1 and 2 each having 2 sub-steps.

For step 1a, a biological variable is observed to be deviant from healthy controls in a particular patient population. The demonstration of test–retest reliability of the finding using blinding procedures is an essential component of this early step. Replication of the finding by the same or collaborating groups is important but confirmation by independent groups is essential for this test to move into the next step of development. For step-1b, information is needed on sensitivity and specificity of the biological variable. Such information would be obtained from ROC (receiver operating characteristic) curves (Zweig and Campbell, 1993) which plot the sensitivity and false positive rates across different "cut-points" of the biological variable. In general, a good test has high sensitivity to the presence of the target disorder and high specificity (excludes normals). This is an essential step as establishing the sensitivity and specificity of a particular variable allows comparing it to other variables as well as composite variables.

Step 2a involves demonstrating the potential clinical usefulness of the specific finding. The objective at this step is demonstration of differences between the target patient population and appropriate comparison groups, that is, patients with diagnoses that commonly appear on the differential diagnostic list of the target disorder. This is an important point, as a biological abnormality may be common to two disorders that hardly ever appear on the same differential diagnostic list (e.g., schizophrenia and dementia in a young adult). Such findings would be of scientific interest but would not significantly influence the diagnostic potential of the biological variable. Alternatively, an abnormality that is equally common to similar disorders (e.g., psychotic bipolar-I disorder and schizophrenia) is less likely to be useful clinically. Abnormalities that occur with significantly different prevalence between disorders to be differentiated are more likely to contribute to the diagnostic process and should progress to step 2b. Similar to step 1b, sensitivity and specificity of the biological variable are to be established.

These data should allow the estimation of the added diagnostic value resulting from incorporating the test into the work-up of a particular patient. The choice of the "gold standard" or reference test is an essential component of this step. This is the standard against which the test being developed will be measured. The currently accepted gold standard in psychiatric diagnosis is the "Best Estimate Diagnosis" (Kosten and Rounsaville, 1992). Best Estimate Diagnosis is reached by agreement among a number of experts relying on multiple sources of information and with a standardized scale with demonstrated validity and reliability. At this step (step 2b), the clinical characteristics of the patient group identified by the test are usually further delineated. Due to the heterogeneous nature of psychiatric disorders, it would be naïve to expect any one biological test to be able to perfectly identify all patients that are classified into a currently defined diagnostic and statistical manual of mental disorders (DSM) based category (e.g., major depressive disorder). It is much more likely that a particular test will be able to identify one or more sub-groups within these categories. Defining the clinical characteristics of the sub-group that is identifiable by a particular test would be very important for the test to be considered for clinical use. Factors such as effects of illness duration, severity, and the effects of medications should also be defined. At this step, the test would be considered "promising" for development as a diagnostic test (Arfken et al., 2009).

The final step (step 3) defines the clinical application of the test and helps standardize the technique used in large and multicenter clinical trials. Multicenter trials should pave the road towards standardization of laboratory procedures used to conduct the test as well as providing data regarding cost effectiveness and impact on both short-term and long-term clinical outcomes. Studies in earlier steps depend on smaller samples of subjects who are usually locally formed. However, step 3 studies would begin to develop larger normative databases that can eventually be used for comparison with an individual's data. Development of such databases can be challenging and will require collaboration among research groups concerned with the specific test being developed.

We have previously documented that the four-step (and by extension the currently proposed 3-step) approach can be useful in determining the stage of development of a biological finding into a clinically useful laboratory test (Boutros et al., 2005). In that report, the reported increased theta activity in the resting EEGs of individuals with attention deficit/hyperactivity disorder (ADHD) was shown to be a promising finding for development into a clinical test. Large multicenter studies are needed before the actual clinical dissemination of the test. In a subsequent study, we assessed the progress of spectral EEG abnormalities in schizophrenia patients (Boutros et al., 2008). We similarly concluded that despite a highly significant deviation compared to normals, there was a lack of studies examining its use as a diagnostic test.

The purpose of the current report is to examine, in a similar manner, the status of development of polysomnographic deviations as a diagnostic tool for major depressive disorders (MDD). The knowledge that sleep is disturbed in mood-disordered patients is well-established Diaz-Guerrero et al. (1946). As early as the mid-1960s a number of groups provided evidence that the sleep of MDD patients could differ significantly from healthy