Journal of Applied Research in Memory and Cognition 2 (2013) 149–157

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

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

Target article

Simple rules for detecting depression☆

Mirjam A. Jenny[a,*], Thorsten Pachur[a], S. Lloyd Williams[b], Eni Becker[c], Jürgen Margraf[b]

[a] Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany
[b] Ruhr University Bochum, Department of Clinical Psychology and Psychotherapy, Universitätsstrasse 150, 44780 Bochum, Germany
[c] Radboud University Nijmegen, Faculty of Social Sciences, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 6 December 2012
Received in revised form 5 June 2013
Accepted 16 June 2013
Available online 24 June 2013

Keywords:
Depressed mood
Fast and frugal trees
Medical decision making
Screening
Beck Depression Inventory

ABSTRACT

Depressive disorders are major public health issues worldwide. We tested the capacity of a simple lexicographic and noncompensatory *fast and frugal tree* (FFT) and a simple compensatory unit-weight model to detect depressed mood relative to a complex compensatory logistic regression and a naïve maximization model. The FFT and the two compensatory models were fitted to the Beck Depression Inventory (BDI) score of a representative sample of 1382 young women and cross validated on the women's BDI score approximately 18 months later. Although the FFT on average inspected only approximately one cue, it outperformed the naïve maximization model and performed comparably to the compensatory models. The heavier false alarms were weighted relative to misses, the better the FFT and the unit-weight model performed. We conclude that simple decision tools—which have received relatively little attention in mental health settings so far—might offer a competitive alternative to complex weighted assessment models in this domain.

1. Introduction

Clinical depression, which is characterized by sadness and loss of interest, affects approximately 1.9–12% of the world's population (Andrade et al., 2003; Kessler et al., 2003; World Health Organization, 2001). Depression can lead to a reduced quality of life and productivity (World Health Organization, 2001), harm the immune system (Herbert & Cohen, 1993), and increase stroke and suicide mortality (Everson, Roberts, Goldberg, & Kaplan, 1998; Inskip, Harris, & Barraclough, 1998).

Given the risks associated with untreated depressive disorders, it is important to have tools available to detect them early and reliably, based on the available indicators or cues (e.g., crying, feeling hopeless). Such detection tools should be simple to keep the extent and cost of the assessment procedure to a minimum. They also should be user friendly, allowing professionals without specific training—such as general practitioners, who are often the first point of contact for people with symptoms of depression, and non-experts (e.g., school or military officials)—to screen certain populations for depression.

In this article, we examine the capacity of simple decision models to detect depressed mood as assessed by the Beck Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), a commonly used screening tool. Specifically, we compare a fast and frugal tree (FFT)—which often limits information search—with two compensatory models, namely a unit-weight model and a logistic regression model, as well as with a simple baseline model. While compensatory models integrate all of the available information to make a categorization, noncompensatory models such as FFTs can decide on the basis of a single piece of information.

2. The role of FFTs in assessing health in medical settings: is more always better?

In medical decision making, errors in diagnosing a patient's health status can have severe and possibly lethal consequences. This may explain why doctors tend to gather more rather than less information when making decisions. But is more information always better? Green and Mehr (1997) suggest that this may not always be the case. They compared a simple decision tree for deciding whether to send a patient suffering from chest pain to the

☆ We thank Jan Woike and Michael Lee for discussing methodological aspects of our analyses, Noortje Vriends for her input on the clinical practice of assessing clinical depression, Eva-Lotta Brakemeier and Laura Martignon for comments on an earlier version of this paper, and Laura Wiles and Susannah Goss for editing the manuscript. Mirjam Jenny was supported by Swiss National Science Foundation Grant 100014_138174/1 granted to Jörg Rieskamp, and Thorsten Pachur was supported by Grant HE 2768/7-1 from the German Research Foundation (DFG) as part of the priority program on New Frameworks of Rationality (SPP 1516).

* Corresponding author.
E-mail addresses: jenny@mpib-berlin.mpg.de (M.A. Jenny), pachur@mpib-berlin.mpg.de (T. Pachur), s.lloyd.williams@rub.de (S. Lloyd Williams), e.becker@psych.ru.nl (E. Becker), juergen.margraf@ruhr-uni-bochum.de (J. Margraf).

2211-3681/$ – see front matter © 2013 Society for Applied Research in Memory and Cognition. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jarmac.2013.06.001

coronary care unit with a more complex statistical method and found that both methods were equally able to differentiate between patients with and without a heart attack. Relatedly, Fischer et al. (2002) found that a simple decision tree was able to compete with a complex regression-based method for assessing children's risk of pneumonia.

Does the potential of simple decision models extend to the mental health domain? Mental health is usually assessed using extensive procedures, such as lengthy structured interviews (Margraf, Schneider, Soeder, Neumer, & Becker, 1996). One frequently used tool to screen for depressed mood is the BDI (Steer, Cavalieri, Leonard, & Beck, 1999), which encompasses 21 questions. Because practitioners have difficulty remembering all criteria of major depression, there have been calls for shortened manuals (Bowers, Jorm, Henderson, & Harris, 1992; Krupinski & Tiller, 2001). The performance of such shortened procedures (e.g., Margraf, 1994) sometimes converges with that of more extensive procedures (Zimmerman et al., 2010). Whooley, Avins, Miranda, and Browner (1997) compared a simple two-question instrument with more complex approaches, and found that the simple instrument had similar (or even higher) discriminability in detecting depression.

Developing simple and robust decision methods has also been a key endeavor in decision science. In his seminal work, Dawes (1979; Dawes & Corrigan, 1974; see also Einhorn & Hogarth, 1975) showed that simple unit-weight models (which consider only the sign of a cue and weight all cues equally) often outperform regression models (which have differential weights) in prediction. More recently, Gigerenzer and colleagues (Gigerenzer & Goldstein, 1996; Gigerenzer, Hertwig, & Pachur, 2011; Gigerenzer, Todd, & the ABC Research Group, 1999; Katsikopoulos, 2010; Pachur, 2010; Pachur, Hertwig, & Rieskamp, in press) demonstrated that robustness in prediction can also be achieved by lexicographic and noncompensatory mechanisms with simple search, stopping, and decision rules. Moreover, simple lexicographic strategies are often used in professional decision making (Garcia-Retamero & Dhami, 2009; Pachur & Marinello, 2013). In lexicographic mechanisms, cues are inspected following a specific hierarchy, usually defined by cue validity or "diagnosticity." Search is stopped and a decision is made as soon as the inspected cue has a particular value. These mechanisms are *noncompensatory* because once information search is stopped, cues lower in the cue hierarchy cannot compensate for cue information further up in the hierarchy. Cue values are neither weighted nor added.

We examined the capacity of FFTs—simple categorization mechanisms that have received considerable attention in decision research (Gigerenzer & Selten, 2001; Luan, Schooler, & Gigerenzer, 2011; Martignon, Katsikopoulos, & Woike, 2008, 2012), but have not been tested in the mental health domain—to detect depressed mood. FFTs can be surprisingly accurate. In computer simulations based on 30 real-world data sets, Martignon et al. (2008) found their performance to be comparable with that of logistic regression and complex classification trees. FFTs are effective because they refrain from differentially weighting all pieces of information and instead focus on a few best pieces of information. This makes them less susceptible to overfitting—that is, they adjust less to unsystematic variability in the data and thus perform better when applied to new data than do methods that differentially weight many pieces of information (Gigerenzer & Brighton, 2009; Katsikopoulos, 2011; Luan et al., 2011; Martignon et al., 2008). Additionally, FFTs may be more readily accepted by clinicians as decision aids than more complex methods (Adams & Leveson, 2012; Elwyn, Edwards, Eccles, & Rovner, 2001; Katsikopoulos, Pachur, Machery, & Wallin, 2008). In sum, because of their accuracy, transparency, accessibility, and simplicity, FFTs could prove to be promising detection models in the domain of mental health (see Marewski & Gigerenzer, 2012, for a thorough, nontechnical discussion of these issues).

Below, we construct an FFT for detecting depressed mood (as assessed by the BDI) and compare its performance with that of a logistic regression model, which is frequently used as a benchmark for categorization (Dhami & Harries, 2001; Kee et al., 2003; Smith & Gilhooly, 2006), a unit-weight model (Dawes, 1979; Einhorn & Hogarth, 1975), and a naïve maximization model (which predicts all cases to be nondepressed and serves as a baseline model). We conducted this prescriptive test using cross validation—that is, we fitted the models to one data set and then tested their accuracy in predicting outcomes in another.

Our analysis contributes to the literature in several ways. First, although FFTs have been tested as descriptive models (i.e., how well they can describe people's decisions; Dhami & Ayton, 2001; Dhami & Harries, 2001; Kee et al., 2003; Smith & Gilhooly, 2006; Snook, Dhami, & Kavanagh, 2011), only a few investigations have subjected them to prescriptive testing (Fischer et al., 2002; Green & Mehr, 1997; Martignon et al., 2008). In this study, we investigated whether the advantages of using FFTs in medical decision making discussed above also holds for predicting depression. Second, to date, the studies by Martignon et al. (2008) and Luan et al. (2011) are the only ones to have tested FFTs by means of out-of-sample prediction (rather than fitting). Finally, we examine—to our knowledge for the first time—the performance of the various models under different weighting schemes of misses and false alarms using real-world data.

3. Overview of the present study

To this end, we drew on data from the Dresden Predictor Study (Trumpf et al., 2010), a prospective epidemiological study in which a representative sample of young women was surveyed twice at an 18-month interval. The FFT, the logistic regression model, and the unit-weight model were provided with data on a set of five BDI items obtained at the first time point ($t1$) and fitted to the women's depression status according to the full BDI at $t1$. The crucial test was how well the models, using the women's responses to the five BDI items at the second time point ($t2$), would be able to detect depressed mood according to the BDI at $t2$. The correlation of the women's depression status (i.e., the criterion values) between $t1$ and $t2$ was $\Phi = .34$; that of the cue values was $\Phi = .31$. Thus, although the $t1$ and $t2$ data sets stemmed from the same group of people, there was nevertheless considerable variability across the two time points, making the data a good test bed for cross validation (which is often used in judgment and decision making research; e.g., Glöckner & Pachur, 2012).

4. Method

The Dresden Predictor Study data set (Trumpf et al., 2010) comprises a representative sample of young women (average age = 18.8 years, range 18–25 years) from the general population. Respondents were recruited using an unweighted random sampling procedure (see Trumpf et al., 2010, for details). The BDI was completed by 1382 respondents at $t1$ and by 1392 respondents at $t2$ (approximately 18 months later). All but 10 of the respondents at $t2$ also completed the BDI at $t1$.

4.1. Criterion

The models inferred the "depressed mood" status of each respondent, as determined by the full BDI. A respondent was considered as having clinically depressed mood if her BDI score (based on all 21 items) exceeded 17 points (which, according to Beck, Steer, & Garbin, 1988, indicates mildly to severely depressed mood; the maximum number of points is 63). Based on this definition, which