ICPE 2018
International Conference on Psychology and Education

DEVELOPMENT OF A METHOD FOR ASSESSING PERSONALIZED RISKS OF DEPRESSION

M.M. Danina (a), N.V. Kiselnikova (b)*, I.V. Smirnov (c)
*Corresponding author

(a) PhD, Laboratory of Psychotherapy and Counseling Psychology FGBNU «PI RAE», Mokhovaya str., 9-4, Moscow, Russia, mdanina@yandex.ru
(b) PhD, Laboratory of Psychotherapy and Counseling Psychology FGBNU «PI RAE», Mokhovaya str., 9-4, Moscow, Russia, nv_psy@mail.ru, +79036284022
(c) PhD, Head of Department Federal Research Centre "Computer Science and Control" of RAS, 60-letiya Oktyabrya avenue, RUDN University, Moscow, Russia, ivs@isa.ru, +74991359020

Abstract

Identification of depressive disorders risk factors still remains a relevant task, and the studies do not show sufficiently consistent data. The strongest risk factors are already known, but there is no general model of depression risk prediction and there are no calculators and predictive algorithms based on it. The results of attempts to develop complex predictive models and algorithms are considered in the paper. Advantages and limitations of available algorithms and calculators are discussed. It is shown that: 1) the most powerful risk factors are the same in different studies; 2) there is a great variety of both depression criteria and methods for measuring variables in predictive models and algorithms. That makes it impossible to compare the proposed models; 3) the way of information gathering may indirectly affect the findings: the answers given by subjects in the national study may differ from those that they might give in a specially organized study of depression. The absence of objective markers and using self-reports makes it difficult to make unambiguous conclusions from the presented studies; 4) the universality of predictive models for different cultures and especially Russian-speaking sample is questioned. Authors present a study aimed at developing a predictive algorithm for assessing depression risk among Russian-speaking adults based on an online health assessment system, that is collecting information about users history, current status and behavior. This online screening service will include identification of four groups of depression predictors and protective factors, as well as situational triggers that can contribute to depression development.

© 2018 Published by Future Academy www.FutureAcademy.org.UK

Keywords: Depression, personalized risks, factors of risk, method of assessing.

1. Introduction

Depression is the leading wide-spread disease observed in 9% of men and 17% of women in Europe. It's about 33.4 million people. The economic cost of depressive disorders associated with expenses for treatment and labor costs, was estimated at 136.3 million euros according to 2007 statistics. Since mental well-being is a key resource for learning, working productivity and active longevity, the necessary and priority measure for any state is investments in preventive care for the population.

According to numerous studies, as well as the WHO 2016 report (World Health Organization, 2016), the current healthcare system is not fully successful in dealing with depression, and evidence-based approaches to its diagnosis and treatment are in dire need of being completed by a strategy for public health support and preventive care. In accordance with the recommendations of WHO, diagnostic and preventive programs must meet the requirements of efficiency and economy, provide opportunities to obtain measurable results and minimally depend on the involvement of specialists in direct work.

These criteria are most fully met by online means of depression risk factors and early signs identification, as well as identification of high-risk groups of people.

2. Problem Statement

In recent decades, the field of knowledge associated with identification of depression risk factors is actively developing. Researchers have identified them and evaluated the effectiveness of prevention programs, especially Internet-based (Buntrock et al., 2016; Muñoz et al., 2010).

There are two dominant trends in assessing risk factors for major depression development in the population. The first trend is associated with identification of objective markers, which can be directly observed, among Internet users. For example, variables that indirectly indicate an actual or developing depressive symptoms in the subjects. Such markers are linguistic features of user texts, characteristics of their voice, photos in social networks and others (Cummins et al., 2015; Yates et al.,2017). The second trend relates to development of complex predictive models and algorithms. They predict the development of depression in healthy subjects over a certain period. Predictive models and algorithms are not aimed at identifying new risk factors, they are based on already known variables and just show their interaction.

The Chicago Adolescent Depression Risk Assessment project (CADRA) is dedicated to predicting depression in adolescents (Van Voorhees et al., 2008). The researchers used data from the National Longitudinal Study of Adolescent Health (Udry, 1998), that is, the data was not collected specifically for the development of a risk assessment model. Depression is detected with the standard scale of the Center for Epidemiological Studies-Depression (Lewinsohn et al., 1997).

An episode of depression during 1 year after the survey was recorded if data met the following criteria: 1) at least one primary symptom of depression (e.g., anhedonia) most of the time and 2) 4 additional depression symptoms according to DSM. Subjects who met the formal criteria for depression at the first wave of assessment were excluded from the predictive model.

Based on current articles on the features of the pathogenesis of the depressive disorder, a list of 119 variables was formed. The variables that were collected in both series of the longitudinal study were selected. The initial list of factors that were actually analyzed included 52 variables, e.g., demographic data, the level of parents' education; anthropometric characteristics, sufficiency of sleep; issues at various aspects

of relations with peers and family; beliefs of the adolescent about himself and about the world; behavioral activation; delinquency; emotions regulation; anxiety; depressed mood.

The mathematical model was constructed on the basis of boosted regression and split-sample validation, training sample, and a test sample. The predictive accuracy of the algorithm was determined using the concordance index and is equal to 0.75.

As a result of the analysis, the items that contributed the most to the occurrence of the depression episode within 1 year after the first examination (20 points, in descending order of contribution) were selected.

The described study has a number of limitations for using its results in our project. Such are the heterogeneity of both the initial evaluation methods and the resulting factors associated with the risk of depression in adolescents. Some of these factors, for example, are markers of current depressive state. Others were represented very extensively (for example, age and BMI). Also, many other variables that, according to studies, also affect development of depression (for example, lifestyle) were not included into the predictive model. In addition, this model was developed for adolescents and cannot be extrapolated to the depression risk prediction in adults.

In another project, a detailed multi-domain model of risk factors for depression and anxiety in preschool children was constructed (Hopkins et al., 2013). It was developed on the basis of the bioecological model of Bronfenbrenner (1989), the transactional model of child depression of Cicchetti and Toth (1998). Contextual variables (for example, stress), parents' variables (for example, parents' depression) and parenting style (for example, hostility), children variables (for example, temperament, attachment style) were categorized as specific factors. Factors that have been studied in the scientific field with respect to the risk of developing depression have also been included. There were observed such factors as socioeconomic status, stressful life events, family conflicts. The sample included 796 children of both sexes, with an average age of 4.44 years. Questionnaires, observations and interviews were used for a comprehensive assessment of the factors.

Techniques that have been used by researchers to diagnose variables, since collection of direct data for preschool-age subjects is difficult, are of particular interest. Context factors were investigated through the collection of socio-demographic information about the family of subjects. Socio-economic status was determined using a Four-Factor Index of Social Status. Demographic information included age, sex, race, parents' education and employment.

Stressful life events were measured using three questionnaires: Perceived Stress Scale; The Family Changes & Strains Scale; Parenting Stress Index - Short Form.

The study of conflicts in the family occurred using three questionnaires: The Family Environment Scale; Family Distress Index; Family Problem Solving / Communication Scales. Parental depression was measured using the Beck Depression Inventory and the Center for Epidemiologic Studies Depression Scale.

Variables that refer to children factors have been investigated using a variety of techniques, including questionnaires and observation. Parental support and hostility were measured using the Parent Behavior Inventory. The NICHD Three Boxes Task was also used. It is a 15-minute recording of parent-child interactions that are used to assess the parents' ability to create a support for the child's development,

including parameters such as supportive parent presence, respect for autonomy, quality of support, cognitive stimulation, confidence and hostility (NICHD Early Child Care Research Network, 2003).

Negative affect and Effective control in a child were measured using the Children's Behavior Questionnaire. The ability of the child to inhibit the dominant responses to the stimulus was diagnosed using a subtest of the Neuropsychological Assessment method. Methods of diagnostics of sensory regulation (Short Sensory Profile), attachment (Attachment Q-Sort), intelligence (Peabody Picture Vocabulary Test) were also used. Symptoms of anxiety and depression were identified using the Interview for Children-Parent Scale (Young child version) and Child Symptom Inventory, which are conducted with the child's parents.

The final predictive model was the same for both depression and anxiety, and consisted of 7 latent factors with multiple indicators: 1) parental depression, 2) effective control, 3) parental hostility, 4) parental support / involvement, 5) scaffolding, 6) sensory regulation, 7) symptoms of depression and anxiety. Three factors with a common indicator - stress, conflict and negative effect, as well as three independent factors with a single indicator - attachment, socioeconomic status and overwhelming control are singled out separately.

Thus, the model shows that 1) contextual factors, such as stress or family conflict, directly affect the symptoms of depression and anxiety in a child, and, along with the socio-economic status, indirectly affect the parents' depression and parental characteristics; 2) symptoms of parents' depression also directly affect the symptoms of depression in children and indirectly - on the characteristics of the children' variables; 3) parenting also has a direct influence, as well as indirect, mediated through attachment and temperament; 4) attachment and temperament directly affect the symptoms of depression in children, and also indirectly affect them through attachment style.

Despite the thoroughness and completeness of this model, we see its significant limitations. At first, parents are the source of information about the behavior and manifestations of the child, including the symptoms of depression. That is, it is not an objective way of measuring these variables. Also, the collected data relates mainly to the psychological and environmental factors of the child's development, while it is known that a significant role in depression development is played by features of health and lifestyle that are not included in the study. The last and most important issue limiting the use of this model in depression prediction for adults is its sampling frame.

In scientific articles, several attempts have been made to develop predictive algorithms for predicting the risk of depression in adult subjects. In one of the first projects - PredictD - a predictive algorithm was developed (Bellón et al., 2011). It was based on data from individuals seeking non-specialized medical care. The following data was collected: age, gender, occupation, educational level, family status, employment, ethnicity, own housing, living alone or with others, birth in or outside the country, satisfaction with living conditions, prolonged illness.

The occurrence of depression episodes during lifetime were determined by two positive statements in the Comprehensive International Diagnostic Interview (depression section). All subjects were retested at 6 and 12 months after the initial examination.

Data were also collected on the following factors: stress on paid and unpaid work, including a sense of control, difficulties with support and stress about lack of respect for the work performed; financial

difficulties; self-assessment of physical and mental health; alcohol use during the last 6 months (using the AUDIT test and the issue of alcohol problems or alcohol dependence treatment); use of recreational drugs; the quality of sexual and emotional relationships with partners or spouses; the presence of serious problems with physical or psychological health, including dependence on substances, disability in close people; difficulties in establishing contacts with people and maintaining close relations; children's experience of physical, emotional or sexual violence; faith in God or other spiritual practices; history of serious psychological problems or suicide in close relatives; anxiety or panic disorder in the past 6 months (Patient Health Questionnaire); satisfaction with neighbors and a sense of security at home and away from home; negative life experience (questionnaire List of Threatening Life Experiences Questionnaire); experience of discrimination over the past 6 months regarding sex, age, ethnicity, appearance, disability, sexual orientation; the adequacy of social support from family and friends.

Unlike similar later studies, the final predictive algorithm included significantly fewer factors: age, sex, educational level, difficulties in paid and unpaid work, physical health, mental health, close relatives with emotional problems, discrimination and episodes of depression in the past. The predictive accuracy of the algorithm was verified by a concordance index of 0.79.

Based on the research of D. Wang and colleagues (2013), an Internet calculator was developed to assess the risk of depression in online users. The predictive model itself, as in CADRA project, was constructed based on the Canadian National Longitudinal Study National Population Health Survey for 1994-1995 (Tambay & Catlin, 1995).

The presence of depression was diagnosed using the Composite International Diagnostic Interview Short Form for Major Depression, which is based on the DSM-III-R. Demographic and socioeconomic variables were selected, as well as self-assessment of general health and stress, movement restrictions, chronic illnesses, difficulty in moving, cognitive functioning, level of pain, frequency of physical activity, chronic stress, negative life events and traumatic childhood experience, work stress - based on the Job Content Questionnaire, self-esteem, mastery, antidepressants and pills against insomnia use during the last month, smoking, problem drinking, psychological distress during the past month, decreased mood or loss of interest during the past year, communication with professionals in the field of emotional or mental health during the past year, past episodes of depression, a family history of depression.

The data was analyzed using logistic regression modelling for men and women separately. It was found that age, family history of depression and episodes of depression in the past are strong predictors for both sexes. Sex-specific predictive algorithms also contained common factors, such as stress, traumatic child experiences, financial difficulties and expectations from other people. Low self-esteem, current mood problems and general health (including low levels of subjective health assessment and physical activity limitations) turned out to be more predictive for women, while for men such factors were physical illness (diabetes, fatigue, insomnia) were more relevant. A depressed state, loss of interest and communication with mental health professionals are predictors of depression development for women, antidepressants and pills against insomnia use - for men. Education and work stress, according to the study, are not significant predictors for depression and that contradicts the other data. Items with low significance were included in the final predictive algorithms. Authors state that the aim of the project is to predict the occurrence of

depression, rather than hypothesis testing, so every factor that can predict it should be taken into account. The predictive accuracy of the algorithm using the concordance index was 0.75.

A similar study was conducted based on the US National Epidemiological Survey on Alcohol and Related Conditions 2001-2003 (Wang et al., 2014). Depression was diagnosed using Alcohol Use Disorder and Associated Disabilities Interview Schedule, which reflects the diagnostic criteria for depression according to the DSM-IV.

Socioeconomic data was collected, e.g.: sex, age, family status, level of individual income, level of education, working status, race, living conditions. Among the specific variables there were such as family history of depression; self-assessment of the state of health and the presence of confirmed diagnoses and conditions, including atherosclerosis, hypertension, liver disease, chest pain, angina, tachycardia, heart attack or other diseases of the cardiovascular system, gastric ulcer, gastritis, arthritis; stressful life events during the past year; quality of life related to health in the last month; problems in childhood - measured using the Children's Trauma Questionnaire and the Conflict Tactics Scale; experience of discrimination - using the Experiences with Discrimination scales.

The predictive algorithm was developed using logistic regression. The risk group was 73% of the subjects who actually developed depression during the period before the second survey.

Universal risk factors included self-assessment of health, presence of depression in parents. Unique risk factors included gender, age, annual income, suicidal thoughts / attempts in the past, depression in siblings, depressive or anxious symptoms, impaired roles due to emotional problems, traumatic experiences, childhood maltreatment, experience of racial discrimination.

Projects aimed at the prevention of depression, in which developers create their own risk assessment tools for users, are also presented. An example of such a project for healthy young people with a genetic risk of depression and bipolar disorder is MindExpress (Wilde, 2014). Although the main goal of this project is preventive interventions, the developers have used an extensive database of questionnaires aimed at diagnosing risk factors for depression.

The 100 points of the final survey included: points aimed at measuring socio-demographic variables, Family History Screen; The Patient Health Questionnaire-9 for measuring current symptoms of depression with two additional questions about past episodes of depression and the perceived future risk of a depressive episode; a scale of neuroticism from a personal questionnaire (Neuroticism-Extroversion-Openness); a short version of the COPE for assessing styles of coping; points on the characteristics of social relations from the Household scale, Income and Labor Dynamics in Australia; Adolescent Alcohol and Drug Involvement Scale; List of Threatening Life Experiences measure for assessing traumatic experiences in the life of the subject; Measure of Parenting Styles to measure the experience of violent, controlling or depreciating styles of parenting in childhood or adolescence of subjects.

The results are divided into 8 groups - risk factors and targets for interventions within the online system: 1) Genetics, family and environment (includes family history of depression, family situation and genetic factors); 2) Styles of thinking; 3) Styles of coping; 4) Social relations; 5) Use of alcohol; 6) Use of cannabis; 7) Traumatic life events; 8) Family dynamics (parenting styles). No statistics is presented by authors, so it is difficult to assess the adequacy of the calculator.

The literature review allows drawing several conclusions: 1) the most powerful risk factors are the same in different studies; 2) there is a great variety of both depression criteria and methods for measuring variables in predictive models and algorithms. That makes it impossible to compare the proposed models; 3) the way of information gathering may indirectly affect the findings: the answers given by subjects in the national study may differ from those that they might give in a specially organized study of depression. The absence of objective markers and using self-reports makes it difficult to make unambiguous conclusions from the presented studies; 4) the universality of predictive models for different cultures and especially Russian-speaking sample is questioned.

3. Research Questions

What principles should be used as a basis for online diagnostics of personalized risks of depression?

4. Purpose of the Study

The purpose of this study is to develop a complex multifactorial predictive algorithm for personalized assessment of depression risk among online health assessment system users.

5. Research Method

We conducted theoretical analysis means investigation of current online health assessment systems and algorithms. The development of a predictive algorithm for assessing depression risk among adults will be based on a system, that is collecting information about users' history, current status and behaviour. The sources of data are: 1) mobile devices that monitor users' status (fitness trackers, smartphones and other wearables), 2) questionnaires filled in by users online, 3) information from social networks related to the activity and preferences of the user, 4) medical data obtained in the course of examinations and diagnostics in medical institutions.

In contrast to the studies described above, we plan to add protective factors into the algorithm, that is, variables that reduce the risk of depression. Also objective diagnostic data will be used, which, presumably, will make the algorithm more accurate. Another important feature of the future project is dynamic variables monitoring.

6. Findings

Analysis allowed us to identify the following groups of factors which will be used in the algorithm: 1. Biological: age, gender, marital status, family history of depression, menopause, sleep disorders (poor sleep quality, insomnia), chronic diseases and conditions (chronic pain, arthritis, heart disease, diabetes mellitus Huntington's disease, Parkinson's disease, Alzheimer's disease, dementia, thyroid disease, stroke, cancer, multiple sclerosis, Itenko-Cushing's disease, hypothyroidism, Addison's disease, Lyme disease, migraine, low blood pressure, acne, obesity, insufficient body weight, high level of cytokines); 2. Personal: Neuroticism, extraversion, stability, a sense of loneliness (including social isolation), financial difficulties, health, stress levels over the past month, self-satisfaction, self-efficacy, the intensity of using social networks, religion, the degree of religiosity, the presence and history of mental disorders, dysfunctional thinking; 3. Social and environmental factors: 1) the level of material well-being, negative experiences

experienced in childhood (emotional, physical or sexual violence, other types of traumatic events), a stressful experience experienced in adulthood more than one year ago (loss of a loved one / animal, parting with a loved one, an experienced catastrophe or a life threatening situation, physical / emotional / sexual violence), level of education, level of illumination in the environment (exposure to daylight, artificial lighting at night); 4. Behaviour: alcohol and / or drug abuse, smoking, irregular working hours, sleep disorders, motor problems, motor activity (degree), nutrition (balanced / unbalanced, regularity), long-term use of drugs (hypnotics, painkillers, sedative, hypotensive, steroid, interferon, beta-blockers, isotretinoin, oral contraceptives, anticonvulsant, antipsychotic, hormonal, against migraine, cardiac.

We assume that, in addition to risk factors, there may be triggers, situations that can create favorable conditions for the development of depression depending on the presence of risk factors. We identified the following groups of triggers: the birth of a child (pregnancy, the period after birth), actual stressful events (loss of work, loss of a loved one, parting, moving, housing, deteriorating living conditions, retirement, unplanned pregnancy, disaster, physical assault) , a chronic disease in close persons.

7. Conclusion

Identification of depressive disorders risk factors still remains a relevant task, and the studies do not show sufficiently consistent data. The strongest risk factors are already known, but there is no general model of depression risk prediction and there are no calculators and predictive algorithms based on it. The researchers are faced with the task of developing a system that relies both on objective markers (medical and behavioral information) and self-report data. Developing an online screening service for Russian-speaking users will include identification of four groups of depression predictors and protective factors, as well as situational triggers that can contribute to depression development.

Acknowledgments

References

Bellón, J. Á, Luna, J. D., King, M., ... & Torres-González, F. (2011). Predicting the onset of major depression in primary care: International validation of a risk prediction algorithm from Spain. *Psychological Medicine*, *41*(10), 2075-2088. doi:10.1017/s0033291711000468

Bronfenbrenner , U. (1989). Ecological systems theory. In R. Vasta (ed.), *Annals of child development* (Vol. 6, pp, 187–249). Greenwich, CT: JAI Press:.

Buntrock, C., Ebert, D. D., Lehr, D., … & Cuijpers, P. (2016). Effect of a Web-Based Guided Self-help Intervention for Prevention of Major Depression in Adults With Subthreshold Depression. *Jama*, *315*(17), 1854. doi:10.1001/jama.2016.4326.

Cicchetti, D., Toth, S. L. (1998). Perspectives on research and practice in developmental psychopathology. In *Handbook of child psychology* (Vol. 4, pp. 479-583). New York: Wiley.

Cummins, N., Scherer, S., Krajewski, J., … & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, *71*, 10-49. doi:10.1016/j.specom.2015.03.004

Hopkins, J., Lavigne, J. V., Gouze, K. R., Lebailly, S. A., & Bryant, F. B. (2013). Multi-domain Models of Risk Factors for Depression and Anxiety Symptoms in Preschoolers: Evidence for Common and

Specific Factors. *Journal of Abnormal Child Psychology*, *41*(5), 705-722. doi:10.1007/s10802-013-9723-2

Lewinsohn, P. M., Seeley, J. R., Roberts, R. E., & Allen, N. B. (1997). Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and Aging*, *12*(2), 277-287. doi:10.1037//0882-7974.12.2.277

Muñoz, R. F., Cuijpers, P., Smit, F., Barrera, A. Z., & Leykin, Y. (2010). Prevention of Major Depression. *Annual Review of Clinical Psychology*, *6*(1), 181-212. doi:10.1146/annurev-clinpsy-033109-132040

Udry, J.R. (1998) *The National Longitudinal Study of Adolescent Health (Add Health). Waves I and II, 1994–1996. Data Sets 48–50, 98, A1–A3.* Los Altos, Calif: American Family Data Archive, Sociometrics Corporation.

NICHD Early Child Care Research Network. (2003). Early child care and mother–child interaction from 36 months through first grade. *Infant Behavior and Development*, *26*(3), 345-370. doi:10.1016/s0163-6383(03)00035-3

Tambay J-L, Catlin G. (1995) Sample design of the National Population Health Survey. *Health Reports*, *7*(1), 29-38.

Van Voorhees, B. W., Paunesku, D., Gollan, J., Kuwabara, S., Reinecke, M., & Basu, A. (2008). Predicting Future Risk of Depressive Episode in Adolescents: The Chicago Adolescent Depression Risk Assessment (CADRA). *Annals of Family Medicine, 6*(6), 503–511. http://doi.org/10.1370/afm.887

Wang, J. L., Manuel, D., Williams, J., Schmitz, N., Gilmour, H., Patten, S. B., MacQueen, G., & Birney, A. (2013). Development and Validation of Prediction Algorithms for Major Depressive Episode in the General Population. *Journal of Affective Disorders*, *151*, 39–45. doi:10.1016/j.jad.2013.05.045

Wang, J., Sareen, J., Patten, S., Bolton, J., Schmitz, N., & Birney, A. (2014). A prediction algorithm for first onset of major depression in the general population: Development and validation. *Journal of Epidemiology and Community Health*, *68*(5), 418-424. doi:10.1136/jech-2013-202845

Wilde, A. (2014) Pilot study of MindExpressTM: An online risk factor-based tailored depression preventive program for young adults with a familial risk of major depressive disorder. In *15th international mental health conference mental health: innovation, integration, early intervention*.

World Health Organization. (2016). Preventing depression in the WHO European Region. In *Trimbos Institute. Netherlands Institute of Mental Health and Addiction., WHO Regional Offices for Europe*.

Yates, A., Cohan, A., & Goharian, N. (2017). Depression and Self-Harm Risk Assessment in Online Forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. doi:10.18653/v1/d17-1322