ABSTRACT

Background: The Structured Clinical Interview for DSM-IV (SCID) is a highly reliable diagnostic instrument used worldwide. However, there is little data as to its reliability and validity outside of the U.S.

Objective: To create a Russian version of the World Mental Health (WMH) SCID and to test its validity among Russian Jewish émigrés in the U.S.

Method: The author, a bilingual Board Certified psychiatrist who has been trained in the application of the original English SCID and WMHSCID, supervised the translation and adaptation of the WMH SCID into Russian. A convenience sample, consisting of 35 subjects, was interviewed by two clinicians trained by the author, yielding 54 diagnoses. All interviews were audio taped and blindly reviewed by the author, who served as the gold standard.

Results: 32 subjects met criteria for one or more of 11 DSM-IV depressive and anxiety disorder diagnoses. There was very good inter-rater agreement; median kappa was 0.75; seven disorders had kappas ranging from 0.65 (Depressive Disorder Not Otherwise Specified) to 1.0 (dysthymia and agoraphobia). Sensitivity was 88.9% and specificity was 77.1%, compared to the “gold standard” diagnosis.

Conclusion: Initial data suggest that the Russian version of the WMH SCID is a valid instrument.

INTRODUCTION

Over the past decades, structural clinical interviews for mental health disorders, e.g., the Diagnostic Interview Schedule (DIS) (1) and Schedule for Affective Disorders and Schizophrenia (SADS) (2) have been developed to improve the accuracy of clinical diagnoses. The Structured Clinical Interview for DSM (SCID) had been developed in the late 1980s (3). It has been explicitly described (3) and validated (4) in English in the early 1990s. As opposed to the DIS, which is a fully structured instrument used in large-scale epidemiological research by trained lay interviewers, SADS and SCID are used by trained clinicians. The SCID proved to be a very reliable psychometric instrument for establishing accurate psychiatric diagnoses. A Multisite Test-Retest Reliability study (4) showed that for most major diagnostic categories, kappas for current and lifetime diagnoses in the patient samples were above 0.60, with an overall weighted kappa of 0.61 for current and 0.68 for lifetime diagnoses. Kranzler et al. (5) assessed the concurrent, discriminant and predictive validity of substance use disorders and common co-morbid SCID diagnoses. The study showed good-to-excellent validity for substance use disorders, moderate validity for major depression, and poor validity for anxiety disorder diagnoses. It is of note that the creators of the SCID warn that “because SCID relies on clinician’s judgment, its reliability is highly dependent on the training and skills of the interviewer” (3, 4). Some studies, (e.g., 6), reported good reliability for anxiety disorders as well — probably due to the high level of skill of the interviewers.
SCID is a state-of-the-art tool, efficiently reviewing criteria for multiple psychiatric disorders within a span of about 90 minutes. It has been widely used in clinical and epidemiological research around the globe (3-5). The SCID incorporates overview, screening questions, and specific questions within diagnostic modules, guidelines about additional probes, operational criteria for DSM-IV, and a system of rating for the severity of symptoms. The critical feature of the SCID is its decision tree design that guides clinicians through hypothesis testing while the interview is underway. The interview skip structure maximizes coverage and time efficiency. The interviewer arrives at a final diagnosis based on the results of the testing and on clinical judgment. The application of the SCID requires familiarity with the “Diagnostic and Statistical Manual of Mental Disorders” and training in its application.

Since the original English version of SCID was created in 1989 there have been multiple modifications of the instrument to address different tasks or settings. Those modifications include in particular Patient vs. non-Patient versions, versions for Axis I Disorders vs. Axis II Disorders. Also, SCID has been modified as a more advanced classification of Mental Disorders — DSM-IV — replaced the previous classification (the DSM-III R for which the original version of SCID had been created). However, every version of SCID, including WMH SCID (7) translated by us into Russian retains its major features described above, which makes it an excellent diagnostic tool.

The translation and validation of the SCID into foreign languages remains a very important task for several reasons. It is important in its provision of data on the prevalence of mental illness abroad that would be comparable to the epidemiological data in the U.S. and other countries. These versions are also important in their use in high risk populations of non-English speaking immigrants within the U.S.; in particular the Russian version is an important assessment of mental disorders in the population of Russian Jewish émigrés. In a previous study, we showed that older Russian-Jewish émigrés living in the U.S. were extremely vulnerable to depressive disorders (8), thus demonstrating the applicability of DSM-IV concepts to a Russian-speaking population.

It had been noted that the adequacy of a diagnostic instrument in a given culture does not guarantee its reliability or validity in another population (9). Valid use of diagnostic instruments across cultures requires a careful adaptation process that goes beyond mere language translation (10). Use of the diagnostic instrument in a culture different from the culture for which the instrument was created presents multiple challenges. We used an adaptation model which evaluates the cultural equivalency of the instrument in five dimensions (9-11). These dimensions are: (1) semantic (the meaning of each item is to be similar in the language of each culture); (2) content (each item is to assess a content which is relevant to each culture); (3) technical (a similar effect is to be achieved by the measuring techniques in the different cultures); (4) criterion (the interpretation of the results of the measure is to be similar when evaluated in accordance with the established norms of each culture); and (5) conceptual (the same theoretical construct is to be evaluated in the different cultures).

Literature is available on the cultural adaptation of fully structured non-clinician administered diagnostic instruments such as DIS (10), as well as semi-structured clinician administered instruments such as Kiddie SADS (12). Wilson and Young pointed out difficulties in applying SCID interviews to non-Western populations: specifically in application to Chinese patients due to the large cultural differences (13). Hence, it is even more striking that despite its worldwide use, studies on cultural adaptation and the reliability of non-English versions of SCID are practically non-existent. We have been able to identify the study addressing cultural adaptation and reliability of the Portuguese version of SCID (14), and the study addressing adaptation of only one module, namely Anxiety Disorder module of the SCID, into Hungarian (15). Inter-rater reliability for the SCID in Norway has been reported, but the relevant study did not provide data on the translation and cultural adaptation of the SCID (6). To our knowledge there is one non-validated Russian translation of the classic version of SCID (Michael First, personal communication). This version had been used in the study of PTSD among Russian Afghan War veterans (16).

Previous research has shown that the level of psychopathology is increased among Russian Jewish émigrés (e.g., 8, 17). There is a need, therefore, for reliable diagnostic instruments for practitioners to be able to accurately diagnose these disorders.

As a collaborator in the World Mental Health (WMH) initiative (or WMH Survey) (18) the author was responsible for producing a professionally translated Russian version of the “WMH SCID” (7) (Russian WMH SCID). The author translated the patient version of the SCID for Axis I Disorders, which included the following diagnoses: Major Depressive Disorder (MDD), Dysthymic Disorder, Depressive Disorder Not Otherwise Specified, Depressive Disorder due to a General Medical
Condition, Substance Induced Depressive Disorder; Panic Disorder, Agoraphobia without History of Panic Disorder (AWOPD), Social Phobia, Specific Phobia, Post-Traumatic Stress Disorder (PTSD), Generalized Anxiety Disorder (GAD), Anxiety Disorder due to a General Medical Condition, Substance Induced Anxiety Disorder, and alcohol abuse/dependence or substance abuse/dependence. Initial validation of the Russian WMH SCID assessing 12 month prevalence of above mentioned diagnoses among Russian Jewish émigrés (“Russian SCID project”) in the United States, in New York City, have been undertaken.

Main indexes used in the evaluation of an assessment tool are inter-rater reliability (or “diagnostic agreement”) and indexes that involve comparison of the diagnoses derived by raters to a “gold standard” diagnosis. The “gold standard” diagnosis is typically an “expert” diagnosis or a “consensus” diagnosis. The aforementioned indexes establish “diagnostic accuracy”: or the percent of diagnoses consistent with “gold standard” diagnosis, “sensitivity”: or the agreement on diagnoses present as determined by “gold standard” diagnosis, and “specificity,” which is agreement on diagnoses absent as determined by “gold standard” diagnosis.

It has been important to create an efficient training program. As the author did not have prior familiarity with the SCID, it had been decided to train the author during the first phase. For the training of the raters we used procedures similar to those utilized by other researchers. In particular, in the study by Ventura et al. (19) aimed to create and assess high quality SCID training program, a User Guide for SCID was utilized by trainees. Training videotapes, including videotapes with the “gold standard” consensus ratings, and tapes observing an experienced rater interviewing the patients were viewed. Symptoms and items elicited by SCID on a training video were judged as present or absent with consensus rating being a gold standard. Diagnostic accuracy after completion of training was 82%.

In the study on Portuguese adaptation of the SCID (13), the training of raters included reading the interview translated into Portuguese, discussion of the material being read, role playing, and applying interviews to the inpatients. The last step of training was repeated until 70% agreement on the main diagnosis has been reached. Authors report that agreement between results obtained with SCID was statistically significant for 10 of the 12 specific diagnostic categories studied.

Both above mentioned studies implemented practice interviews at the end of training. In the current paper we present a detailed report on the initial validation of the Russian WMH SCID for Axis I Mental Disorders.

METHODS
SAMPLE
The subjects were recruited through the Russian Language Health Services (RLHS) and the Department of Psychiatry at Beth Israel Medical Center (BIMC), New York. RLHS is a triage program in the primary care setting. The program is run by a licensed full time social worker, and assists Russian émigrés who are not fluent in English and are seeking mental health and/or other health services by providing them with an initial assessment. Many of the clients using RLHS services are older non-English speaking Russian émigrés and, consistent with our previous study (7), are very vulnerable to the stress of immigration and have high prevalence of undiagnosed psychiatric disorders, in particular depressive disorders.

Since the focus of the project was to assess the validity of the Russian SCID administered by trained interviewers, efforts were made to identify and recruit the clients who were suspected by RLHS staff to have psychiatric disorders even if they did not receive psychiatric services per se (e.g., seeing internist only).

A total of 35 subjects were recruited and interviewed yielding 54 diagnoses due to high comorbidity in the sample. Seventeen were current clients of RLHS. Three were active clients in the Outpatient Division of the Department of Psychiatry at BIMC, formally known to RLHS. Three subjects were recruited from the pool of callers who used the “September 11 hotline” opened by RLHS for Russian émigrés soon after the September 11 terrorist attack. One subject was an inpatient in the Psychiatry Unit at BIMC. Eleven subjects learned about the study by word of mouth from other respondents living in the same housing project and volunteered to participate. All subjects were not known to the interviewers or to the author and a great majority of them (31 out of 35) had never been assessed psychiatrically. Except for one inpatient subject, the interviewers and the expert were blind to the source of subject referral.

The subjects in our study had been resident in the US for varied lengths of time: from one year to more than 20 years. They came from different regions of the former Soviet Union: mostly from the Slavic Republics (Russia, Ukraine and Belarus) and the Middle Asia region of the former Soviet Union (e.g., Uzbekistan). For all the subjects, Russian was a primary language. We did not notice
differences in the flow of the interview depending on the part of the former Soviet Union the subjects came from or based on their length of stay in the U.S.

The project was approved by the Institutional Review Boards of the BIMC and SUNY at Stony Brook. All participating subjects signed informed consent, which included separate consents for interview and for audio taping. All interviews were audio taped. One interview could not be completed because the subject became physically exhausted. Since by this time most of the assessment was completed, the subject was included into the analysis.

**INSTRUMENT**

The mental health interview on which interviewers had been trained and which had been administered to the subjects was the Russian version of the Structured Clinical Interview for DSM-IV (SCID) for the Axis I disorders, WMH version (18).

Like other versions of SCID, the WMH SCID includes an introductory overview followed by a screening module. It then focuses on depressive disorders, anxiety disorders and alcohol and substance disorders; an overall Clinical Global Impression score is given. Thus, if a respondent answers positively to the screening item, the corresponding diagnostic module is then assessed. The severity scales incorporated into WMH SCID were also translated and back translated, in the same way as the rest of WMH SCID, including Montgomery-Asberg Depression Rating Scale (MADRS) for MDD and Dysthymia (20), the Panic Disorder Severity Scale (PDSS) (21), Liebowitz Social Anxiety Scale for Social Phobia (22), Marks Fear Questionnaire for Agoraphobia and Specific Phobia (23) and Structured Interview for the Hamilton Anxiety Rating Scale (SIGH-A) (24, 25) for GAD. The validation of the diagnostic instrument, such as the Russian version of WMH SCID, applied to the validation of the instrument as a whole and did not include separate validation of the above mentioned severity scales.

**Translation and cultural adaptation of the WMH SCID.**

The second step was the translation of the instrument. The English WMH version and quantitative scales were translated into Russian jointly by a professional translator working together with the author, back-translated to English by another translator and checked and corrected by the author.

Below we provide some examples of overcoming challenges in different dimensions to achieve cultural equivalence. Examples of semantic and technical challenges are applicable to both immigrant and non-immigrant (residing in Former Soviet Union) population. Examples of content and criterion challenges are relevant only to non-immigrant population.

I. **Semantic:**

1. While the word that would be a literal translation of “functioning” exists in the Russian language, it is never used in the particular context such as its use in DSM. (It is used to say, e.g., “normal functioning of cardiovascular system,” “normal functioning of machinery,” etc.). Therefore, adequate semantic translation required using two different words. One word needed back translates as “function,” e.g., in Panic Disorder Severity Scale “impairment/interference in carrying out function at work, or social function as a result of Panic Disorder.” The second word back translates as “activities of living,” as it is used numerous times for DSM criteria of impairment in functioning for multiple disorders: “symptoms cause… impairment of social, occupational, or other important areas of functioning.”

2. The word “discouraged” is impossible to translate...
precisely into Russian without using two-sentence phrasing (in general English is a more laconic language than Russian): “[did you ever feel] that all your efforts are futile and there is no sense to continue trying?” Since semantically precise translation would disrupt the flow of the interview, as a compromise we translated “discouraged about how your life is going” as “feeling down about how your life is going,” thus conveying meaning by preserving part of the sentence “about how your life is going.”

Culturally relevant idioms were used when appropriate (e.g., translation of the question about panic attacks), “Have these attacks ever come on completely out of the blue…” was translated using similar Russian idiom: “As if a thunderstorm from the clear sky.”

II. Content: in the English SCID, the screening question for alcohol disorders sounds, “Has there been any time in your life when you had five or more drinks (beer, wine, or liquor) (on one occasion/at the one time)?” Adequate Russian translation would back translate the word “drink” as “portion.” However, due to cultural patterns of drinking in Russia, “portion” could mean as much as 100-200 grams of vodka. Therefore, the cultural equivalent utilized reflected the absolute amount of alcohol, rather than “drinks” (“150 gram of hard liquor or equivalent amount of other alcoholic beverages”).

III. Technical: Russian respondents suffering from depression tend to express emotional distress in somatic terms (7). Thus, in training, we have emphasized that raters should be aware of this cultural tendency and make a special effort to elicit depressive symptoms other than, e.g., sleep disturbance or fatigue, through adequate probing.

IV. Criterion: the societal tolerance to heavy alcohol consumption is significantly higher in Russian culture; also many people don’t have cars. Thus, the value of certain items is not the same as it is in the American culture: the item related to “problems with a law” is less applicable because arrest for disorderly conduct would be much less likely. The item related to driving while intoxicated is much less relevant as is the possibility for trouble with a law for “Driving While Intoxicated” (DWI), so common in the U.S. However, we preserved all the items from the English version of SCID, and because only one item is sufficient for meeting symptom criteria we put heavy reliance on the item related to intra familial conflict — which is very common in the families of people with alcohol problems (26).

V. Conceptual: we didn’t face challenges related to conceptual equivalence.

PROCEDURE

Raters. Two psychiatry trainees participated in the project. Both trainees were bilingual and worked as physicians in the former Soviet Union prior to coming to the U.S. Both trainees had knowledge of the American psychiatric diagnostic system (DSM-IV).

Training Program. During the preparation phase prior to in-person training, the trainees reviewed the training materials. These materials included teaching videotapes and a User Guide in English.

1. Days 1-3 of training included an introduction of the SCID-IV structure: the role of the overview section, flexibility vs. adherence to format, scoring system (skip-outs, symptom ratings), and discussion of particular modules. This training has been conducted in Russian and utilized the translated Russian version of the WMH SCID. It also included role-playing in Russian in which one of the trainees would act as a “subject” based on a prewritten scenario with a history and responses to critical questions. Trainees’ performances during the role play was supervised by the trainer and guided as necessary. All arising questions were immediately discussed.

2. On day 4 training videotapes (in English, produced in Harvard University) of the SCID-IV, the WMH version, were reviewed. Trainees had to rate the symptoms and make diagnoses. One of the videotapes presented a full length interview and the other videotapes presented segments demonstrating use of severity scales for depression, panic disorder and generalized anxiety disorder. Ratings of trainees were compared to those of Harvard University raters and discrepancies were discussed.

3. On day 5 of training, each trainee conducted practice interviews in English with the actual patients in the in-patient psychiatry unit. Trainees were blind to the patients’ diagnoses. Two interviews were conducted by trainee A and one interview was conducted by trainee B. Interviews were supervised by the method of a joint interview (13), i.e., the author observed interviews directly and was allowed to ask additional questions only after the interview had been completed and the ratings were done. Then, interviews and ratings were carefully analyzed and extensive feedback was provided. The feedback included comments on the technique and accuracy of ratings by the trainees. In all cases, trainees were able to make the same diagnosis...
as the author — operationalized as a “gold standard” diagnosis.

**Interviews of the study.** Trainee A conducted 23 assessments and trainee B conducted 12 assessments. Raters filled a hard copy of the SCID and all interviews were audio taped. All audiotapes were blindly reviewed by the author. Trainees’ diagnoses had been compared to the “gold standard” diagnosis by the author.

**DATA ANALYSIS**

Inter-rater reliability of the instrument was assessed by analyzing agreement between diagnoses assigned by a rater and the author. Inter-rater agreement was calculated as Cohen’s kappa (27, 28). Kappa coefficients represent the amount of agreement between pairs of ratings (e.g., agreement on diagnosis five out of 10 times) adjusted for chance agreement. Chance is based on the likelihood that a specific diagnosis will occur in a sample. For example, if in the sample of 100 patients, 50 are given diagnoses of depression, a rater is going to be correct at least 50% of time if he or she arbitrarily gives everybody a diagnosis of depression. Criteria proposed by Landis and Koch (29) were used to interpret the Kappa coefficients: excellent reliability (Kappa > 0.75) good reliability (Kappa = 0.59-0.75), fair reliability (Kappa = 0.40 - 0.58), and poor reliability (Kappa < 0.40).

**RESULTS**

Twenty-two females and 13 males participated in this study. The age of the subjects ranged from 43 to 77 years with the majority of subjects (69%) being in their 50s or 60s. The results are shown in Table 1. Agreement was good to excellent for all diagnoses, with the exception of social phobia and specific phobia. These were the least prevalent disorders in this sample (one case of each), which made it difficult to estimate inter-rater agreement for these diagnoses precisely. Four subjects known to the mental health system (three referred from psychiatric outpatient clinic and one referred from inpatient unit) had a chart diagnosis consistent with the “gold standard” diagnosis (MDD, dysthymia, Panic Disorder, Generalized Anxiety Disorder). This provides additional support for the validity of the Russian WMH SCID.

Comparative data to other studies are shown in Table 2. The WHO SCID allows deriving diagnoses from ratings or from the clinical impression indicated by the interviewer at the end of the SCID worksheet (in the great majority of the cases it would be the same diagnosis as the one produced by ratings). For the purposes of this report, we used the diagnoses based on clinical impressions made by the trainees and compared them to the diagnoses made by the author after listening to the audiotapes. Different studies use different ways to assess inter-rater agreement; some studies use more than one approach. We chose to assess reliability on specific diagnoses (6, 13), (i.e., if participant had three diagnoses, we would make three assessments for this participant), as opposed to agreement on specific items (16), or combined diagnosis (6), or main diagnosis (13) because it was most consistent with the objectives of our study.
Only three subjects did not have any positive responses to any of the screening questions and did not have symptoms of any psychiatric disorders over the past 12 months. Thirty-two subjects had positive responses to one or more screening questions and had at least sub-syndromal symptoms of any psychiatric disorder. Twenty-four subjects met the DSM criteria for one or more psychiatric diagnoses according to the author’s opinion. There were a total of 54 diagnoses (varying from one to six diagnoses per subject). Probably because of the nature of the sample, no subject required assessment for the past 12 months for alcohol or a substance use disorder.

There was very good agreement on the diagnoses that did not differ significantly for the two trainees. Overall diagnostic accuracy was 77.4%. There was an excellent sensitivity of 88.9%. For the 54 cases, when the author made a diagnosis of a particular disorder, the trainees made the same diagnosis in 48 cases. In the six remaining cases, trainees did not arrive at the gold standard diagnosis. In five of these cases, trainees did not diagnose a disorder. In one case, a trainee made a very similar but different diagnosis of depressive disorder due to general medical condition instead of the diagnosis of Major Depressive Disorder (subject developed depression after being diagnosed with colon cancer. In that case, however, the depression was not due to the physiological effect of her cancer).

There was a higher accuracy on the diagnosis of depressive disorders — 84.4% and a lower but acceptable accuracy on the diagnosis of anxiety disorders — 73.8%.

**DISCUSSION**

There are different ways of checking the reliability of a psychometric instrument in general and particularly in a foreign language. Major methods include use of audiotapes (6), videotapes (16), joint interviews (13), and test-retest design (4). Test-retest design has limitations in that it is not a naturalistic diagnostic procedure. In particular, subjects would recognize the structure of the interview and on repetition could have chosen to deny symptoms formally reported in order to cut the interview short (4).

The method of the joint interview (6) entails the trained rater conducting the interview and making the diagnosis, while the expert, who is present during the interview, may ask additional questions after the interview is completed and may arrive at his own expert diagnosis.

Although audio taping as opposed to videotaping precludes the expert from the opportunity to observe the patient, in particular to evaluate visible emotional distress due to psychiatric disorder, our method of checking reliability of the SCID using non-selective listening of audio-tapes by the expert has been utilized in other studies (6) and is a valid method of checking reliability of the psychometric instrument, in particular the foreign language version. It is of note that the unique feature of the SCID WMH version is that it incorporates within the modules quantitative clinician-rated scales for many of the disorders which then are used as a “back-up” — while the SCID-directed interview is underway, thereby decreasing the chance of obtaining insufficient information.

All audio-taped interviews, in the opinion of the author, provided sufficient information for arriving at a diagnosis. The discrepancies emerged as a result of differences in interpreting and rating available data, in particular over-diagnosing by raters of “near-threshold” cases. In our opinion, from a clinical point of view, the advantage of an audio-taped one-to-one interview as opposed to a joint interview is that one-to-one interviews increase the comfort and ease of both the interviewer and respondent alike.

Direct comparison with other studies on English, and Portuguese study on non-English SCID, is complicated due to the fact that different versions of SCID had been used in different populations and different diagnoses were assessed (in particular Psychotic Disorders, which were not included into WMH SCID); so such comparison should be treated with caution. Diagnostic accuracy in our project was 77.4% and in a study by Ventura et al. (16) of the English SCID, striving to create a state-of-the-art training program, accuracy was 82% — which was classified by the authors as “very good” (16). Table 2 provides direct comparison of inter-rater agreement on specific diagnoses with other studies. Apparently, these data show very favorable results compared to other studies. Study on the reliability of the Portuguese version of SCID is not included as it assessed a different range of diagnoses; it did provide kappa coefficient for MDD 0.88 which is only slightly better than the 0.76 in our study. As diagnosis of MDD is highly common in our sample we also calculated the sensitivity and specificity for MDD which was 86% and 90 % respectively in our study compared to 84% and 91% accordingly in the study by Basco et al. on the English SCID (30). We suggest that the success of the project of the adaptation of SCID to the Russian language was facilitated by the fact that the principal investigator is a bilingual psychiatrist with many years of clinical experience, knowledgeable about both DSM-IV and Russian culture.

It is of note that, consistent with our previous work (7), only a small proportion of participants with psychiatric
disorder(s) (4 out of 32) had been receiving psychiatric services, thus illustrating high unmet needs in this population.

The study has certain limitations and the results should be considered as preliminary. The major limitation is the small size of our sample. However, we believe that this limitation is compensated in part by high “caseness”, that is to say the high total number of diagnoses in this very sick population. Another potential sample limitation is that it was a particular group of people, with very limited age range as well as with unusually high comorbidity (54 diagnoses). Also, due to the nature of the sample, no subjects in our study were diagnosed with alcohol or substance use disorders.

Another potential limitation of the study is that initial validation had been done in Russian speaking subjects living in the U.S. as opposed to their country of origin. Potential language issues could also be applied to the main three clinicians in the study as their working language is English. The research shows (31) that in order for tests to be effectively translated and adapted in different cultural contexts they need to have demonstrated validity and reliability in their home countries.

A further limitation of the study was that pre-training materials such as the User Guide and teaching videotapes were in English and practice interviews were conducted with English speaking patients.

CONCLUSION
The initial results suggest that the Russian version of the WMH SCID is a valid instrument. The Russian SCID can improve diagnostic accuracy in Russian Jewish émigrés. Future studies should use larger samples of Russian speaking émigrés from the former Soviet Union, as well as Russian speaking subjects living in their country of origin. Also, future research should address different populations, in particular those with alcohol and substance use disorders.

Future research should use more rigorous methodology: utilizing higher standard translation and back translation procedures (independent translation by two translators, then creating one edited version, and back translation by two independent translators), as well as using consensus diagnosis as opposed to expert diagnosis as a “gold standard.” Interviews should be evaluated by several equally trained SCID experts in Russia. Future studies should also utilize scales such as Rater Applied Performance Scale (32), thus providing more objective data.

The Russian version of the SCID WMH version is available from the author upon request.

Acknowledgements
The author is very thankful to Dr. Charles Webb and Dr. Igor Galynker for their assistance with organizing the project and their very valuable suggestions; to Dr. Evelyn Bromet for invaluable comments and assistance in the preparation of the manuscript; to Roman Kotov for his help with statistics; and the author expresses deep thanks to the raters: Drs. Natalia Osmanova and Alisa Shakhverdi.

References
2. Endicott J, Spitzer RL. A diagnostic interview: The schedule for affective disorders and schizophrenia. Arch Gen Psychiatry 1978; 35:837-844.
27. Cohen J. Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. Psychological Bull 1968; 70: 213-220.