Korean Version of the Diagnostic Interview for Genetic Studies: Validity and Reliability

Eun-Jeong Joo, Yeon-Ho Joo, Jin-Pyo Hong, Samuel Hwang, So-Jin Maeng, Jin-Hee Han, Byung-Hwan Yang, Yu-Sang Lee, and Yong Sik Kim

The Diagnostic Interview for Genetic Studies (DIGS), developed in 1994 by the National Institute of Mental Health (NIMH), was translated into Korean and tested for reliability and diagnostic validity. Concurrent validity was tested using the Structured Clinical Interview for DSM-IV (SCID) and clinical diagnoses in 53 patients, most of whom had either schizophrenia or bipolar disorder. Inter-rater reliability was tested in 24 patients. Test-retest reliability was also tested in 17 patients. Overall and specific diagnostic validity for the Korean version of DIGS (DIGS-K) was excellent for most diagnoses. Inter-rater and test-retest reliability for overall and specific diagnoses also ranged from fair to excellent. For schizoaffective disorder, the test-retest reliability of DIGS-K was in a fair range, although the level was lower than that of other diagnoses. However, its diagnostic validity and inter-rater reliability was below fair range. In conclusion, DIGS-K appears to be a reliable interview for major psychiatric disorders.

NUMEROUS STUDIES on genetic factors of major psychiatric disorders have been conducted globally, but they have yet to identify definite genes for schizophrenia or mood disorders.[1-3] One of the problems may lie in the difficulty of defining psychiatric phenotypes. It is largely unclear as to which and how clinical features should be assessed to make a genetically relevant diagnosis for psychiatric disorders. To address such a fundamental issue, in 1994 the National Institute of Mental Health (NIMH) Genetics Initiative developed the Diagnostic Interview for Genetic Studies (DIGS),[4] a semistructured clinical interview designed for collecting a comprehensive database of psychiatric symptoms, signs, and psychiatric history for genetic studies. DIGS primarily focuses on the assessment of major mood disorders, psychotic disorders, and substance abuse.

DIGS has unique features.[4] It has a poly-diagnostic capacity, which makes it possible to extract multiple diagnoses according to different diagnostic systems. It is designed to facilitate the collection and storage of a comprehensive database of information. It allows for a detailed assessment of the course of the illness, chronology of psychotic and mood syndromes, and comorbidity. All of these data are collected both qualitatively and quantitatively.

DIGS has become one of the most widely used interview tools for psychiatric genetic studies. Using the standard method in the assessment of phenotype for psychiatric disorders may be one way to resolve the issue of genetic complexity in psychiatric disorders. Most psychiatric disorders are so complex that genetic studies generally require a large sample size to prove any statistical significance.[5] By using the same assessment tool, such as DIGS, researchers could analyze multiple genetic studies collectively and make more reliable comparisons of results. Needless to say, international collaborative studies require the use of a unified standardized method.

DIGS has already been translated into several languages, e.g., Hindi, Japanese, French, Chinese, etc. Several studies, including the original report by the NIMH, already have reported good reliability of DIGS and its translated versions.[4,6-8] Currently, in Korea, there is vigorous interest in a genetic approach among psychiatric researchers and the number of publications pertaining to genetic studies with Korean patients is rapidly increasing. However, since most of the studies use clinical diagnosis as a phenotype, proving the scientifically reliability of the phenotype has been a

From the Eulji University, School of Medicine, Department of Psychiatry, Eulji Hospital, Seoul; University of Ulsan College of Medicine, Department of Psychiatry, Asan Medical Center, Seoul; Seoul National University, School of Medicine, Department of Psychiatry and Behavioral Science, Seoul National University Hospital, Seoul; Catholic University of Korea, College of Medicine, Department of Psychiatry, St. Vincent Hospital, Suwon; Hanyang University, College of Medicine, Department of Neuropsychiatry, The Mental Health Research Institute, Seoul; and the Yong-In Psychiatric Research Institute, Yong-In Mental Hospital, Yongin, Korea.

Supported by a grant from the Korean Academy of Schizophrenia and sponsored by Lilly Korea.

Address reprint requests to Yong Sik Kim, M.D., 28 Yongon-Dong Chongno-Gu, Seoul 110-744, Korea.

0010-440X/04/4503-0007$30.00/0
doi:10.1016/j.comppsych.2004.02.007

difficult task. In turn, a reliable interview tool for genetic studies such as DIGS has recently become a definite necessity in Korea. DIGS has thus been translated into Korean and tested for its reliability and validity in this study.

METHOD

Translation Process

The English version of DIGS (version 2.0) was translated into Korean. The translation process entailed consultations from linguists of Korean and English. A special committee for the translation procedure consisted of three psychiatrists, a bilingual psychologist, and a linguist of Korean language. The committee underwent a rigorous revision process of the preliminary translation of DIGS. Psychiatrists then applied the revised version of the translated DIGS to several patients and the questions that were difficult to understand were further modified. While some questions with correct literal meaning were not readily understood because of the differences in culture and social background, most modifications were kept to a minimum in order to retain the original meaning of the question. Back-translation was done and critically compared with the original version of DIGS. There was no crucial difference between the original version and the back-translated version of DIGS and only minor changes were added to the translated DIGS based on the comparison. The final version of the Korean version of DIGS (DIGS-K) has been formatted to follow the identical layout of the original version of DIGS 2.0. Several phrases or sentences in italics have been added which denote DIGS-K specific modification: for example, "*Buddhism*" has been added as one of the choices of religion.

Standardization Process

Patients. Patients were recruited from both psychiatric wards and outpatient clinics for the study. Patients with schizophrenia, bipolar disorder (type I), schizoaffective disorder, major depression, and alcohol-related disorder were included. Most patients were recruited from two university hospitals, one psychiatric hospital, and one general hospital. All of the patients submitted written consent forms to participate in this study.

Interviewers. Interviews with DIGS were conducted in a semistructured manner, allowing further questions and explorations as needed. Interviewers, who consisted of psychiatric nurses with more than 3 years of clinical experience in the psychiatric ward, assessed psychopathology and history of illness. Interviewers also drew timelines for the lifetime history of psychiatric illness and evaluated reliability of the whole interview. Before the study, the interviewers received intensive training for DIGS interview using the DIGS manual and exercise interviews with real patients.

Items of standardization. Concurrent validity, inter-rater reliability, and test-retest reliability were examined to confirm the validity and reliability of DIGS-K. Concurrent validity was evaluated in two ways. First, DSM-IV diagnosis extracted from DIGS-K interview was compared with DSM-IV diagnosis extracted from Structured Clinical Interview for DSM-IV (SCID) interview. SCID was selected for the comparison because the Korean version of SCID has already proven to have good inter-rater reliability.[9] Second, DSM-IV diagnosis extracted from DIGS-K interview was compared with clinical diagnosis using DSM-IV criteria. Clinical diagnoses were extracted from medical records and the evaluation of the psychiatrist in charge of each patient. Inter-rater reliability was evaluated by comparing DSM-IV diagnoses based on the information of DIGS-K interview recorded by two different interviewers. During the interview, one interviewer conducted the DIGS interview while the other merely observed the interview. Test-retest reliability was evaluated by comparing DSM-IV diagnoses based on the information of two independent interviews of the same patient. Interviews for test-retest reliability were always conducted by the same interviewer. A second interview was done after different time spans according to the patients' convenience. The timespan between the initial and retest interviews therefore varied from 1 week to 50 weeks. DSM-IV diagnoses for all cases were made by a psychiatrist blind to any clinical information on the patients, whose decision was only based on the information written on the DIGS-K book. All of the above consistencies were calculated as an overall kappa value as well as an individual kappa value of specific psychiatric diagnosis.

Analysis

For the analysis, 5×5 tables were constructed based on five diagnostic categories, i.e., schizophrenia, schizoaffective disorder, bipolar disorder (type I), major depression, and alcohol dependence. Consistency of DSM-IV diagnoses from different interviews was calculated as Cohen's kappa value. Overall kappa coefficient was calculated from these 5×5 tables for concurrent validity, inter-rater reliability, and test-retest reliability. Individual kappa coefficient was calculated from 2×2 tables, which were constructed based on the presence or absence of the specific psychiatric disorder. Consistency was considered as "excellent" if kappa was greater than 0.75, "fair to good" if 0.40 to 0.75, and "poor" if less than 0.40.[10]

RESULTS

Fifty-three patients were included in the study. Patients with schizophrenia, bipolar disorder (type I), major depression, schizoaffective disorder, and alcohol dependence were recruited randomly. Most patients had either schizophrenia ($n = 22$) or bipolar disorder, type I ($n = 16$). Diagnostic validity of DIGS-K was evaluated in terms of consistency of DSM-IV diagnosis with SCID and clinical interview. Consistency was evaluated with Cohen's kappa value for overall diagnoses and individual diagnosis. Overall concurrent validity with SCID was excellent (kappa = 0.81). The kappa values for schizophrenia, bipolar disorder (type I), major depression, and alcohol dependence were also in the excellent range (0.87 to 0.90). However, for schizoaffective disorder, concurrent validity was not as good (kappa = 0.31) and kappa value between DIGS-K diagnosis and clinical diagnosis was also poor (kappa = 0.25). The same trend was

Table 1. Validity and Reliability of DIGS-K: Overall and Specific Kappa Values

	Total	ALD	BPD	MDD	SZ	SA
Concurrent validity						
DIGS-K *v* SCID	0.81	0.88	0.87	0.90	0.89	0.31
DIGS-K *v* clinical diagnosis	0.84	1.00	0.78	1.00	0.85	0.25
SCID *v* clinical diagnosis	0.92	0.88	0.91	0.90	0.96	0.66
No. of subjects	53	5	16	5	22	5
Inter-rater reliability	0.79	1.00	1.00	0.83	0.60	0.50
No. of subjects	24	5	5	3	7	4
Test-retest reliability	0.82	—	0.74	1.00	1.00	0.43
No. of subjects	17	0	6	1	8	2

Abbreviations: ALD, alcohol dependence; BPD, bipolar disorder (type I); MDD, major depression; SZ, schizophrenia; SA, schizoaffective disorder.

observed between SCID diagnosis and clinical diagnosis (kappa = 0.66). In addition, overall and individual kappa values between DIGS-K diagnosis and clinical diagnosis, with the exception of schizoaffective disorder, were in the excellent range of 0.78 to 1.00. Overall and individual kappa values between SCID diagnosis and clinical diagnosis were even better, ranging from 0.88 to 1.00, with the exception of schizoaffective disorder (Table 1).

A total of 24 patients were included to examine inter-rater reliability. The results are provided in Table 1. The overall kappa coefficient was 0.79. The kappa value for individual diagnosis varied: it was excellent for alcohol dependence, bipolar disorder (type I), and major depression (1.00, 1.00, and 0.83, respectively), but just fair for schizophrenia and schizoaffective disorder (0.60 and 0.50, respectively).

A total of 17 patients were included in the determination of test-retest reliability. The overall kappa value was 0.82 (Table 1). The kappa value for individual diagnosis varied from 0.43 to 1.00. Notably, schizoaffective disorder showed the lowest test-retest reliability (kappa = 0.43) compared to other reliability values and no data were available on alcohol dependence.

DISCUSSION

The English version of DIGS is a comprehensive psychiatric interview tool primarily for the genetic studies of major psychiatric disorders. There has been a need to develop the Korean version of DIGS in order to carry out more reliable and standardized genetic studies in Korea as well as international collaborative efforts. DIGS-K will definitely make national, as well as worldwide, comparisons of psychiatric genetic studies easier to conduct. As DIGS mainly focuses on major psychiatric disorders, such as schizophrenia and bipolar disorder (type I), our standardization process has been conducted mostly with those patients. A typical interview using DIGS-K took about 2 hours to complete. Review of other medical information and proper recording of the DIGS-K book took another 1 to 2 hours. As a whole, 3 to 4 hours were generally required to complete DIGS-K with one patient. Appropriate breaks were allowed during the interview to help the patient concentrate on the interview. Consistent with the instruction in the DIGS manual, the section for screening schizotypy was skipped for all interviews in this study since most of our patients were already identified as having a psychosis. Future studies that mostly deal with major psychosis may consider skipping the schizotypy section to save time. Since DIGS interviews take a long time to complete, interviews need to be conducted flexibly to accommodate the patient's ability to focus and maintain attention.

The validity and reliability of DIGS-K were tested using DSM-IV diagnosis extracted from the DIGS-K interview. DIGS has been developed not only to make DSM-IV diagnoses, but also to collect comprehensive information about psychopathology and lifetime psychiatric history. Therefore, merely looking at the agreement of DSM-IV diagnosis of the DIGS-K interview may not be sufficient to prove the reliability and validity of DIGS-K as a whole. However, evaluating agreement for all of the information on DIGS-K is simply not feasible due to the sheer number of items. For this reason, DSM-IV diagnosis has been

used alternatively for the reliability study in other language versions of DIGS.[6,7]

For schizophrenia and bipolar disorder, kappa coefficients were excellent enough to give diagnostic confidence for DSM-IV diagnosis of DIGS-K. The inter-rater reliability was excellent for overall and bipolar disorder (type I), major depression, and alcohol dependence. Inter-rater reliability for schizophrenia and schizoaffective disorder was lower than other diagnoses, but it was still in the fair range. The lower inter-rater reliability for schizophrenia seems to be partially due to the lower kappa for schizoaffective disorder, as most disagreement came from the confusion with schizoaffective disorder. Overall test-retest reliability was excellent and individual test-retest reliability for specific diagnosis varied in the range of 0.43 to 1.00. While the kappa values varied somewhat, they were still in the excellent to fair range. Nonetheless, the sample size may not have been sufficiently large to confirm the reliability with confidence. In particular, no alcohol dependence patient and only one patient with major depressive disorder were included for test-retest reliability in this study.

The kappa coefficients of schizoaffective disorder in terms of diagnostic validity, inter-rater reliability, and test-retest reliability were usually much lower than for other diagnoses. The values were not sufficiently high to prove validity and/or reliability. Similar results have been reported previously by other researchers.[4,7] The original reliability study on DIGS has reported much lower test-retest reliability for schizoaffective disorder (kappa = 0.31 to 0.45) compared to other diagnoses such as major depression, bipolar disorder (type I), and schizophrenia (kappa = 0.75 to 0.96).[4] The reliability study on DIGS-French version also reported relatively lower kappa value for inter-rater reliability of schizoaffective disorder (kappa = 0.60 to 0.87) than other diagnoses (kappa = 0.85 to 1.00). Test-retest reliability of schizoaffective disorder (kappa = 0.38 to 0.48) was lower than other diagnoses (kappa = 0.62 to 0.65).[7] The reason for lower kappa values of schzoaffective disorder could be as follows: first, the number of patients was relatively small, and second, it is difficult to accurately assess the duration of mood symptoms and/or psychotic symptoms and their overlap as in the case of schizoaffective disorder. Hence, duration of symptoms, which is crucial DSM-IV criteria in differentiating schizoaffective disorder from other diagnoses, may not have been assessed reliably.

In this study, all DSM-IV diagnoses were made by one psychiatrist, even though DIGS-K interview and recording of DIGS-K books were completed by different interviewers. In turn, DSM-IV diagnoses were made consistently based on the same clinical knowledge and DSM-IV experience. Previous studies on DIGS reliability were based on DSM diagnosis made by DIGS interviewers themselves. Considering the actual genetic research setting in Korea, even just one genetic study is likely to involve numerous researchers, including psychiatric nurses or psychologists, administering DIGS-K interviews. It is well known that psychiatric diagnoses may vary greatly depending on the psychiatric training background and amount of clinical experience of the interviewers, even if they apply the DSM-IV criteria equally. Different researchers can make different DSM-IV diagnoses based on the same information obtained from DIGS interview. Therefore, formulating ways to reach a consistent DSM-IV diagnosis based on the DIGS-K interview is thought to be a more relevant issue in terms of reliability and validity of DIGS-K.

In summary, DIGS-K was proven to be a reliable and valid interview tool for major psychiatric disorders in Korea. DIGS-K will be an excellent tool not only for genetic studies but also for other researches on major psychiatric disorders. Researchers will be able to collect detailed and comprehensive lifetime histories of psychiatric illness and psychopathology for current and past episodes. The relatively low diagnostic validity and reliability for schizoaffective disorder may be overcome if researchers strive to define the exact duration of symptoms at the time of DIGS-K interview. Actually, DIGS has been designed to reveal and resolve such diagnostic issues by asking the same question in many different ways throughout the whole interview. However, difficulties with extracting and identifying pertinent information, especially about the possible overlap between mood and psychotic symptom, still seem to exist. Therefore, diagnosis consensus meeting and various sources of clinical information may be useful for making more accurate diagnoses and resolving the discrepancy of

diagnosis. DIGS-K will be made available at the website of Korean Academy of Schizophrenia (*http://www.schizophrenia.or.kr/*) for free to any researchers having academic purpose only.

ACKNOWLEDGMENT

We would like to thank Jee-Duk Lee and Hye-Sik Yoon for their excellent help with psychiatric interviews. We express our deepest gratitude to all patients who agreed to participate in this study.

REFERENCES

1. Owen MJ, Holmans P, McGuffin P. Association studies in psychiatric genetics. Mol Psychiatry 1997;2:270-273.

2. Gershon ES, Badner JA, Goldin LR, Sanders AR, Cravchik A, Detera-Wadleigh SD. Closing in on genes for manic-depressive illness and schizophrenia. Neuropsychopharmacology 1998;18:233-242.

3. NIMH, Clinical Neurogenetic Branch. Genetics and mental disorders. Report of the MIMH's Genetic Workgroup. Biol Psychiatry 1999;45:559-602.

4. Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, et al. Diagnostic interview for genetic studies: rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry 1994; 51:849-859.

5. Risch N, Merikangas K. The future of genetic studies of complex human disease. Science 1996;273:1516-1517.

6. Deshpande SN, Mathur MNL, Das SK, Bhatia T, Sharma S, Nimgaonkar VL. A Hindi version of the diagnostic interview for genetic studies. Schizophr Bull 1998;24:489-493.

7. Preisig M, Fenton BT, Matthey ML, Berney A, Ferrero F. Diagnostic interview for genetic studies (DIGS): inter-rater and test-retest reliability of the French version. Eur Arch Psychiatry Clin Neurosci 1999;249:174-179.

8. Berney A, Preisig M, Matthey M-L, Ferrero F, Fenton BT. Diagnostic Interview for Genetic Studies (DIGS): Inter-rater and test-retest reliability of alcohol and drug diagnoses. Drug Alcohol Depend 2002;65:149-158.

9. Han OS, Hong JP. SCID-I; Structured Clinical Interview for Axis I Disorder of DSM-IV. Seoul, Korea: Hana Medical, 2000.

10. Fleiss JL. Statistical Methods for Rates and Proportions. Ed. 2. New York, NY: Wiley, 1981.