Eur Arch Psychiatry Clin Neurosci (2001) 251 : 279 – 283

ORIGINAL PAPER

U. Ptok · C. Seeher · F. Jessen · A. Papassotiropoulos · R. Heun

Inter-rater reliability of family history information on psychiatric disorders in relatives

Received: 13 September 2001 / Accepted: 28 September 2001

Abstract The family history method in psychiatric family studies is an important and necessary way of obtaining information on family members who are not available for personal interview. Studies on the validity of this method have shown that family history information on psychiatric disorders in relatives is neither accurate nor sensitive but highly specific. However, its inter-rater reliability has rarely been assessed, even though this is a prerequisite for adequate validity.

In the present investigation we examined the inter-rater reliability of family history information obtained with a semi-structured and symptom-oriented interview. Forty informants were interviewed twice by two different raters within 3 and 20 days.

The inter-rater reliability was found to be good for dementia (kappa=0.82, 95 % CI=0.61–1.00), alcohol related disorders (kappa=0.93, 95 % CI=0.80–1.00), for depressive disorders (kappa=0.72, 95 % CI=0.42–1.00), anxiety disorders (kappa=0.75, 95 % CI=0.41–1.00) and any psychiatric disorder (kappa=0.79, 95 % CI=0.66–0.91).

We concluded that the family history interview is a useful family study instrument that can be applied reliably by different raters for frequent psychiatric disorders.

Key words inter-rater reliability · family history questionnaire · family study

Dr. U. Ptok (✉) · C. Seeher · F. Jessen ·
A. Papassotiropoulos · R. Heun
Department of Psychiatry
University of Bonn
Sigmund-Freud-Str. 25
53105 Bonn, Germany
Tel.: +49-228/287-6367
E-Mail: u.ptok@uni-bonn.de

Introduction

Family studies in psychiatric research are performed to evaluate the relevance of possible familial factors in psychiatric disorders. Even though it would be desirable to obtain personal interviews from all relatives of a patient, this is often impossible. Some family members may already have deceased by the time of the interview or may not be available for other reasons. In those cases, surrogate information has to be used to obtain psychiatric diagnoses for the relevant subjects. In order to optimize the acquisition of family history information, semi-structured interviews were developed (Andreasen et al. 1977, Breitner et Folstein 1984, Silverman et al. 1986) and tested for their validity (Andreasen et al. 1977, 1986, Silverman et al. 1986, Heun et al. 1996a, Heun et al. 1998b).

The inter-rater reliability is a relevant factor in family studies since in the usual setting there are a number of different raters that contribute to the data acquisition and an enormous bias could result from a poor reliability of the instrument. The joint-rater reliability was shown to be good by Andreasen et al. (1977) but does not allow sufficient conclusions about the inter-rater reliability in a family study setting. As investigated by Zimmerman et al. (1988), inter-rater agreement was shown to be excellent for most diagnoses. Even though a high reliability is a necessary prerequisite for a sufficient validity, the inter-rater reliability has rarely been assessed and not been tested for the dementia questionnaires or for any of the German versions of family history interviews.

We therefore decided to test the inter-rater reliability of the family history, dementia and dementia risk interviews for information on relatives in a family study setting in a German sample.

Methods

Forty patients of the Department of Psychiatry at the University of Bonn were recruited consecutively in 1999 and asked for permission to interview a family member about the patients' psychiatric family history. The family members were then interviewed twice by two different raters within a time frame of 3–20 days.

Forty patients and 40 informants participated in the investigation. Information was collected on 230 first degree relatives. Demographic information on patients, informants and first degree relatives is given in Table 1.

As the study instrument we used two semi-structured interviews that were adapted from the Family History Questionnaire (Andreasen et al. 1977), the Dementia Questionnaire and the Dementia Risk Questionnaire (Breitner et Folstein 1984, Silverman et al. 1986) and translated into German (by R. H.) with the generous permission of the authors. The family history interview is designed like a diagnostic interview intended to obtain lifetime psychiatric diagnoses for each first degree relative of an index person. It starts with a list of all relevant subjects and then systematically asks for the psychiatric history of each person in a symptom-oriented way according to diagnostic criteria as required in ICD10. The combined dementia/ dementia risk interview lists every first degree relative as well and asks for their history of memory disorders. The aim is to identify dementia and take into account associated clinical features in order to obtain a specific diagnosis whenever a memory disorder is or has been present. Examples of questions of the German versions of the interviews are attached in the appendix. The full interviews are available from the authors.

Interviewers were seven medical students in their last year of medical school who had undergone thorough training including a four week practical on a psychiatric ward as well as 10 supervised training interviews prior to the investigation. Interviews were distributed evenly among the interviewers to avoid fixed rater combinations. A full randomization of rater combinations or interview distribution was not possible due to a limited availability of raters. Information on the psychiatric history of the index subjects was not collected to keep the raters blind to the index subjects' diagnoses.

The interview information was used to create ICD-10 diagnoses for all 230 first degree relatives that were named by the interview partners. ICD-10 diagnoses were made by two independent experienced psychiatrists who were blind to the second or the first interview sequence, respectively, and to the index diagnosis or the relationship of the subject and the index person.

Data analysis

Complete identity of ICD-10 diagnoses in first and second interview was considered too stringent for comparisons. Therefore, diagnostic categories were formed as indicated in Table 2. For example, presenile and senile dementia of Alzheimer type, vascular dementia with acute onset, multi-infarct dementia as well as mild cognitive disorder were combined as "dementia" and single and recurrent depressive episodes as well as dysthymia were combined as "depressive disorders". Disorders that were diagnosed less than five times in both interviews were included in the category "any psychiatric disorder" but do not appear as individual categories.

Diagnostic agreement between first and second interviews was measured using kappa values including

Table 1 Description of the index sample, of the informant sample and of the sample of relatives on whom family history information was recorded

	Number	% female	Mean age	Relationship with index subject	% living
Index subjects	40	57.5	52.6 (range 21–88) SD 16.9	–	100
Informants	40	55.0	57.25 (range 21–94) SD 14.5	spouse 28 (70%) sibling 1 (2.5%) parent 8 (20%) child 3 (7.5%)	100
Relatives	230	43.9	51.1 (range 4–96) SD 21.8	spouse 7 (3%) sibling 102 (44.3%) parent 69 (30%) child 52 (22.6%)	67.4

Table 2 Frequencies of psychiatric diagnoses of relatives in first and second interviews

ICD-10 disorder	ICD-10 codes	Frequency in first interview	Frequency in re-interview	Kappa (95 %CI)
Dementia	F00–F06	9	9	0.88 (0.73–1.00)
Dementia of Alzheimer type	F00	5	7	0.83 (0.60–1.00)
Vascular dementia	F01	3	2	0.80 (0.41–1.00)
Alcohol-related disorder	F10	7	8	0.93 (0.80–1.00)
Depressive disorder	F32–F34	8	5	0.72 (0.42–1.00)
Depressive disorder, recurrent or single episode	F32–F33	3	4	0.85 (0.57–1.00)
Depressive disorder, single episode	F32	2	3	0.39 (0.00–0.94)
Depressive disorder, recurrent	F33	1	2	0.67 (0.05–1.00)
Anxiety disorder	F40	3	5	0.75 (0.41–1.00)
Any psychiatric disorder		30	28	0.79 (0.66–0.91)

95% confidence intervals. Kappa values indicate chance-corrected proportional agreement (Altman 1994). Criteria for the evaluation of kappa were chosen according to Landis and Koch (Landis and Koch 1977).

Results

Of the 230 relatives, 28 obtained at least one psychiatric diagnosis in the first and 25 obtained at least one psychiatric diagnosis in the second set of interviews.

Exact birth dates of relatives were given in 112 cases in the first and in 131 cases in the second interview.

Table 2 shows the frequency of psychiatric disorders in both interviews and the reliability of diagnostic categories as well as of individual psychiatric diagnoses, i. e., kappa values and 95% confidence intervals.

The inter-rater reliability appeared to be good for dementia (kappa=0.82, 95% CI=0.61–1.00), alcohol related disorders (kappa=0.93, 95% CI=0.80–1.00), depressive disorders (kappa=0.72, 95% CI=0.42–1.00), anxiety disorders (kappa=0.75, 95% CI=0.41–1.00) and any psychiatric disorder (kappa=0.79, 95% CI=0.669–0.91). There was also good inter-rater reliability for dementia of Alzheimer type and for vascular dementia (kappa=0.83, 95% CI=0.60–1.00 and kappa=0.80, 95% CI=0.41–1.00, respectively). However, the inter-rater reliability was poor for the diagnostic subgroups of depressive disorders (depressive disorder, single episode: kappa=0.39, 95% CI=0.00–0.94; depressive disorder, recurrent: kappa=0.67, 95% CI=0.05–1.00).

Discussion

In the present investigation, the inter-rater reliability of psychiatric diagnoses and dementia diagnoses derived from structured interviews appeared to be good for the diagnostic categories of depressive disorders, alcohol-related disorders, anxiety disorders and dementia. This is in line with previous publications (Andreasen et al. 1977 and Zimmerman et al. 1988).

Andreasen, however, tested the joint-rater reliability of the family-history research diagnostic criteria. In contrast to that, we tested the reliability of the actual instrument for different ratings of the same informant in separate interviews with different raters. The investigation of the joint-rater reliability cannot answer the question if two raters in separate interviews would still come to the same diagnosis, which was the aim of our investigation. This seems to be a relevant question since it could be argued that different raters might obtain different information from the same informant due to a different relationship between the two interview partners or factors like age, sex, race or place of the interview.

Other than in the study done by Zimmerman et al. (1988), in our investigation, diagnoses were not assigned by those conducting the interviews but by experienced psychiatrists who used the interview information later. This represents a practicable way in which large scale epidemiologic studies can be conducted, when information is collected from many informants and possibly at many different locations. Since the conduction of all interviews by experienced psychiatrists is no longer possible in such a setting, it is especially important to be able to rely on well-structured non-ambiguous instruments that can be used by less experienced raters for later central evaluation by experts.

Comparing diagnostic subgroups, we found that the more specific diagnoses of dementia of Alzheimer type and vascular dementia were diagnosed with a high inter-rater reliability. This is of interest especially for family studies of dementing disorders. Looking at depressive disorders, the inter-rater agreement of family history information seems to be sufficient for the combined category but not for the differentiation of single or recurrent depressive episodes or the severity of the depression.

Previous investigations have examined the family history method using Research Diagnostic Criteria (RDC) (Andreasen et al. 1977, Zimmerman et al. 1988). RDC are especially suited to diagnose subjects without access to information from a full diagnostic personal interview. In spite of the clear advantages of the simplifications of RDC, the use of a well-known and widely used classification system such as ICD10 allows comparison with other studies or even within studies (some subjects who were interviewed in person might be diagnosed according to ICD10) and makes the results more easily understandable to those not involved in epidemiological research or not familiar with RDC. ICD10 as a hierarchically structured diagnostic system seems to be a good alternative when broad diagnostic categories are chosen (Heun et al. 1997).

A limitation of our approach is the relatively short time between interview and re-interview. One could criticize that the re-interview would be confounded by the informant being able to recall the exact answers provided during the first interview. We chose this tight time frame to minimize the drop out between interviews. Even though some interviews were performed at the informants' home after the index patient had been discharged from hospital, it appeared to be easier to motivate informants to attend a re-interview when the index person was still hospitalized and the interview was performed during a hospital visit of the informant. In the investigation by Zimmerman et al. (1988), both interviews were performed within one week which means there was even less time between interviews on average.

There are other aspects that could possibly bias the information obtained in a test-retest setting. Informants might try to mobilize additional information between interview and re-interview in order to be able to answer the questions more accurately. It could also be hypothesized that informants would tend to report less psychiatric symptoms in the re-interview to avoid another long interview session the second time. Examples supporting both possibilities were found in our sample. There were

more exact birthdates given in the second set of interviews, but fewer psychiatric symptomatology was reported at the second interview. This seems to indicate that the quality and accuracy of the interview information could profit from restricting the number of interviews and thus interview time and from allowing the informant some time to prepare for the interview.

Previous evaluations of the family history method with the same diagnostic instruments have revealed a high specificity but low sensitivity of this method (Heun et al. 1996a) as well as the need to assess a number of different sources of possible bias (Heun et al. 2000). Interinformant reliability was shown to be low for depression but better for dementia (Heun et al. 1998a), the validity of family history information improved with increasing severity of a disorder (Heun et al. 1998b), the sensitivity of family history information was higher in families of affected subjects than in control families and was improved when more informants were interviewed (Heun et al. 1998a, 2000). In spite of these confounds and limitations that should be accounted for, the need to obtain family history information was clearly confirmed in two studies on selection effects in both general population and relatives of patients that showed the higher availability of subjects without psychiatric disorders for personal interview compared with those with a psychiatric disorder (Heun et al. 1996b, 1997). The limitations of the validity of the family history interviews is not caused by a lack of inter-rater reliability. The inter-rater reliability of the tested instruments is good and allows for use by different raters whenever surrogate information is the only way of obtaining information in a family study setting.

Conclusion

The inter-rater reliability of family-history information on psychiatric disorders and dementing disorders obtained in a test-retest setting with the help of standardized questionnaires is good for the combined diagnostic categories "dementia", "alcohol-related disorder", "depressive disorder" and "anxiety disorder" as well as for the diagnostic subgroups of dementia of Alzheimer type and vascular dementia. These instruments can therefore be administered reliably in a family study setting with different raters.

References

1. Altman DG (1994) Practical Statistics for Medical Research. Chapman and Hall. London
2. Andreasen NC, Endicott J, Spitzer RL, Winokur G (1977) The family history method using diagnostic criteria. Reliability and validity. Arch Gen Psychiatry 34:1229–1235
3. Andreasen NC, Rice J, Endicott J, Reich Th, Coryell W (1986) The family history approach to diagnosis. How useful is it? Arch Gen Psychiatry 43:421–429
4. Breitner JCS, Folstein MF (1984) Familial Alzheimer Dementia: a prevalent disorder with specific clinical features. Psychol Med 14:63–80
5. Heun R, Hardt J, Burkart M, Maier W (1996a) Validity of the family history method in relatives of gerontopsychiatric patients. Psychiatry Research 62:227–238
6. Heun R, Burkart M, Maier W (1996b) Selection biases during recruitment of patients and relatives for a family study in the elderly. J Psychiat Res 6:491–504
7. Heun R, Hardt J, Müller H, Maier W (1997) Selection bias during recruitment of elderly subjects from the general population for psychiatric interviews. Eur Arch Psychiatry Clin Neurosci 247:87–92
8. Heun R, Müller H (1998a) Interinformant reliability of family history information on psychiatric disorders in relatives. Eur Arch Psychiatry Clin Neurosci 248:104–109
9. Heun R, Müller H, Papassotiropoulos A (1998b) Differential validity of informant-based diagnoses of dementia and depression in index subjects and in their first-degree relatives. Soc Psychiatry Psychiatr Epidemiol 33:510–513
10. Heun R, Kockler M, Papassotiropoulos A, Ptok U, Fimmers R (2000) Subject and informant characteristics influence the reliability and validity of family history information: an analysis based on the generalized estimating equations approach. Int J Methods Psychiatr Res 9:60–67
11. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174
12. Silverman JM, Breitner JCS, Mohs RC, Davis KL (1986) Reliability of the family history method in genetic studies of Alzheimer's disease and related dementias. Am J Psychiatry 143:1279–1282
13. World Health Organisation (1992) ICD-10. The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines. WHO, Geneva, Switzerland
14. Zimmerman M, Coryell W, Pfohl B, Stagl D (1988) The reliability of the family history method for psychiatric diagnoses. Arch Gen Psychiatry 45:320–322

Appendix

Examples of questions included in the German version of the Family History Questionnaire

6a Gab es bei der/ dem Mutter, Vater, Schwester, Bruder, Tochter, Sohn oder Ehepartner von (Name des Indexprobanden) jemals eine Zeit von mindestens 2 Wochen Dauer, während der eine depressive, traurige, hoffnungslose, niedergeschlagene Stimmung vorlag?

NEIN
JA Name:
Wann (Zeiträume):
Welche Erkrankung:

6b Gab es bei der/dem Mutter, Vater, Schwester, Bruder, Tochter, Sohn oder Ehepartner von (Name des Indexprobanden) jemals eine Zeit von mindestens 2 Wochen Dauer, während der bei dieser Person Freudlosigkeit und Interesselosigkeit bestand?

NEIN
JA Name:
Wann (Zeiträume):
Welche Erkrankung:

Falls die Frage 6a oder 6b mit JA beantwortet wurde → Kapitel DEPRESSION auf Seite 12! Ansonsten gehen Sie zu Frage 7a auf Seite 21.

DEPRESSION

In dieser/der schlimmsten Phase

D1 kein Appetit

NEIN
JA Name:

D2 Gewichtsverlust (z. B. 1 kg/Woche), ohne es zu wollen

NEIN
JA Name:

D3 jemals deutlich mehr Appetit
D4 Gewichtszunahme (z. B. 1 kg/Woche)
D5 Einschlafschwierigkeiten
D6 Durchschlafschwierigkeiten
D7 Früherwachen
D8 zu viel geschlafen
D9 kaum Energie, erschöpft und abgespannt
D10 Stimmungstief am Morgen
D11 langsamer gesprochen oder sich bewegt
D12 ständige Unruhe, auf und ab gehen
D13 kein Interesse am Geschlechtsleben
D14 Interesse an fast allen (auch schönen) Dingen verloren
D15 Freudlosigkeit
D16 Äußerungen über Wertlosigkeit, Sünde und Schuld
D17 Minderwertigkeitsgefühle
D18 deutlich weniger Selbstvertrauen
D19 erhebliche Konzentrationsschwierigkeiten
D20 Klagen über sehr langsame Gedanken
D21 fast täglich Schwierigkeiten, selbst Alltagsentscheidungen zu treffen
D22 besonders viel über den Tod im Allgemeinen nachgedacht
D23 Wunsch zu sterben
D24 dachte daran, Selbstmord zu begehen, und äußerte, wie
D25 Selbstmordversuch unternommen

Wenn bei D1–D25 drei oder mehr Fragen mit JA beantwortet wurden → weiter mit Frage D26 auf Seite 18! Ansonsten gehen Sie zu Frage 7a auf Seite 21.

D26 Gab es Phasen, in denen die Symptome (vorlesen aus D1–25) gleichzeitig vorhanden waren?

Falls JA → weiter mit Frage D27 auf Seite 19! Ansonsten gehen Sie zu Frage 7a auf Seite 21.

D27 Wann trat dies zum ersten und wann zum letzten Mal auf?
D28 Wie lange dauerte die längste Phase?
D29 Wie viele Phasen gab es?
D30 Deswegen im Krankenhaus oder ärztlich behandelt?
D31 Deswegen Medikamente?
D32 Ist Ihnen bekannt, ob Mutter, Vater, Schwester, Bruder, Tochter, Sohn oder Ehepartner von (Name des Indexprobanden) wegen dieser Erkrankung nicht hat arbeiten können oder zur Schule gehen können?

Examples of questions of the Dementia Questionnaire

4a Gab es bei Mutter, Vater, Schwester, Bruder, Tochter, Sohn oder Ehepartner von (Name des Indexprobanden) Anzeichen für Gedächtnisverlust, Orientierungsstörung, Verwirrtheit, unerwartete Änderungen im Verhalten, Schwierigkeiten mit dem Lesen oder Schreiben, Probleme, Worte zu finden, oder damit, daß Dinge erzählt wurden, die keinen Sinn ergaben?

NEIN
JA Name:

Falls JA, weiter mit Frage 5a.
5a Welches war der höchste Bildungsabschluß, den (Name des betreffenden Verwandten) erreicht hat?
5b Welches war die höchste berufliche Stellung, die (Name des betreffenden Verwandten) erreicht hat?

Hat (Name des betreffenden Verwandten) jemals Probleme gehabt mit

6a dem Gedächtnis?
6b Verwirrtheit?
6c dem Verstehen von Situationen und Erklärungen?
6d der Erinnerung an Namen von Leuten?
6e dem Erkennen von bekannten Gesichtern?
6 f sich in unbekannter Umgebung zurechtzufinden?
6 g sich im Haus zurechtzufinden?
6h dem Behalten einer kurzen Liste (z. B. Einkaufsliste)?
7. Haben die Probleme mit dem Gedächtnis plötzlich oder langsam angefangen?
8. Verschlechterte sich das Gedächtnis eher kontinuierlich oder war die Verschlechterung abrupt?
9. Gab es irgendein Ereignis vor dem abrupten Beginn oder der plötzlichen Verschlechterung des Gedächtnisses?
10. Wurde ein Arzt wegen der Gedächtnisprobleme konsultiert?

Falls JA

11. Welche Diagnose liegt vor?