Chapter 8

General Discussion

INTRODUCTION

In this Chapter, the study is critically reviewed as a whole. Consecutively, a global clinical impression is given and the design and implementation of the study, the psychometric methodology applied and the utility of the new structured approaches in clinical practice and scientific research are discussed. The major conclusions are outlined and recommendations are made for further research.

GLOBAL CLINICAL IMPRESSIONS

In the course of this study, a number of subjective observations made during the interviews led to some global clinical impressions. It seemed that many of the participating BZD users underestimated their liability of being dependent on BZDs, despite their often long-term BZD use, their tendency to secure a ready supply of BZDs and the general knowledge that tranquillizers can be addictive. While viewing their complaints in terms of disease symptoms, they usually tended to legitimate their BZD use by referring to the expertise of the doctor who had prescribed them. During the interviews, patients were confronted with questions implying that BZD users may experience withdrawal symptoms between successive intakes of these drugs. Some patients volunteered that they were not aware of this possibility. Without this insight, such withdrawal symptoms can be expected to reinforce further BZD use instead of leading to the acknowledgment of dependence. However, in order to obtain BZDs on prescription, a patient must have a connection with a physician. Consultations could be turned into a working alliance aimed at promoting the awareness of BZD dependence and offering guidance and support in a process of reduction and discontinuation of BZD use. The fact that informed consent was given by the majority of patients in this study, suggests that the medical

context of BZD use can effectively be used to reach patients with the objective of attaining such a working alliance.

DESIGN AND IMPLEMENTATION OF THE STUDY

Selection of settings and subjects

The representativeness of the study population and the generalizability of the findings depend on the selection of the study samples. Nine general practices, seven psychiatric outpatient departments, six community-based outpatient addiction centres (CBACs) and two self-help groups participated in this study. As the general practices were first approached by an informative letter about the study and a subsequent telephone call, the number of refusals was considerable before a visit could be arranged to deliberate potential participation in the study. However, when a visit took place, it led to participation in almost all cases. The course of recruitment might have caused some selection bias in the GP sample, because the attitude towards cooperation is likely to reflect the attitude of the GP towards the clinical management of BZD use. It is unknown to what extent this might influence the course of BZD dependence in these GP populations. There were no refusals from the other populations. The psychiatric outpatient departments were selected from different backgrounds, associated with general hospitals as well as psychiatric hospitals. All the known Dutch self-help groups concerned with the addictive use of medications were included. The CBACs were all divisions of one addiction institute, which may be considered a typical example of a Dutch addiction institute. Except for the reasonable doubt pointed out with respect to the GP population, the populations in this study can be assumed to be representative.

The patient selection process depended on the efforts of many different (practice)

assistants (especially at the general practices), who invited patients to cooperate. Although great care was taken to inform these assistants as fully as possible about the procedure, the selection process progressed more efficiently in some settings than in others. However, in all settings, the majority of patients initially identified as BZD users were reached with a request to participate in the study. Considering the time investment of approximately three hours to complete the two interviews, fairly high response rates (67 to 70%) and low drop-out rates (0 to 7%) were attained. These acceptable response rates and the fact that participating patients were reassured that their data would be anonymized to prohibit any personal consequences of the investigation, support the assumption that the results of this study are fairly generalizable.

Interview-related reliability and validity considerations

This study was carried out by 18 different interviewers, who had a university background in either medicine (n=8) or clinical psychology (n=10). During the investigations, there were differences in the efficacy of the interviewers. Despite the fact that the interviewers completed a series of training sessions for the administration of the structured interviews, i.e. the Benzodiazepine Dependence-Structured Diagnostic Interview (Bendep-SDI), Schedules for Clinical Assessments in Neuropsychiatry (SCAN) and Addiction Severity Index-Revised (ASI-R), it is possible that certain differences between interviewers caused some interviewer-related bias.

The application of the instruments went well in general. The administration of the Bendep-SRQ, Bendep-SDI and SCL-90 did not require any special skills and did not cause problems. The administration of the SCAN was more complicated in some cases. In the case of poly-dependent and sometimes intoxicated CBAC patients, it was sometimes an elaborate task to complete both the sections on psychoactive substance use. It is questionable whether cases with polydependence give reliable answers to the SCAN questions for all the separate

substances which are being used. For example, to what degree is a polydependent drug user able to distinguish BZD withdrawal symptoms from alcohol, heroine and cocaine withdrawal symptoms, when the use of these substances displays an irregular pattern and the interactions are unpredictable?

The SCAN and ASI-R were originally developed for the assessment of alcohol and illicit drug users. Some parts the ASI-R therefore had little relevance for BZD users without alcohol and drug problems. For example, all the GP patients had a criminality problem severity score of zero. In the ASI-R section 'drug problems', a sole category 'medications' referred to the BZDs together with all other medications. In the SCAN, the BZDs are assigned to the category 'sedatives', which was purposefully reserved in this study for BZDs only. The SCAN and ASI-R lack specificity with respect to BZD use due to their general approach to all psychoactive substances. The structured questions in the SCAN and ASI-R do not differentiate between dependence and disease symptoms. This demonstrates a clinical problem that has been reported frequently.[1-4] Symptoms experienced by a patient who tries to control, delay, reduce or discontinue BZD use, can be attributed to dependence, but in some cases, they might be relapsing disease symptoms. This duality is also reflected in the relationship between the patient and physician in the case of BZD use, whereas for other substances, such a relationship can concentrate on the dependence problem only.

In the assessment of the concurrent and discriminant validity of the Bendep-SRQ scales, the above-mentioned lack of specificity of the SCAN and ASI-R with respect to BZD use has to be taken into account. Nonetheless, in Chapters 4, 5 and 6 they were used as comparative measures in the factor analyses carried out for this purpose, because no other more specific instruments were available. To improve the specificity of the SCAN results with respect to BZD use, the Rasch homogeneous DSM-III-R and ICD-10 BZD dependence scales, constructed and assessed in Chapter 3, were used in these factor analyses. The specificity of

the ASI-R problem severity scores, which reflect the clinical judgement of the interviewer with respect to a whole problem section on a scale from 0 to 9,[5,6] could not be improved. So-called composite ASI scores, which have been propagated for scientific use in former studies on the ASI,[7-10] were not used in this study because they are based on variable sets of items, which are selected by inflating Cronbach's alpha as much as possible. The construct therefore has to be changed on each occasion to achieve high internal consistency. Furthermore, there are differences between the constituting variables with regard to the contents and the number of the response categories, which makes the assigned numerical weights quite arbitrary. To derive an ASI-R problem severity score the interviewer weighed the responses to all the questions in a section and chose the most relevant point on the problem severity scale. The subjective nature of this procedure calls for cautious interpretation of the factor analyses in which the ASI-R problem severity scores were included. The Maximum Likelihood factor analyses, attempted initially in the Chapters 4 and 6, were inconclusive, because no satisfactory goodness of fit could be established for any factor solution. The factor-analytical approach was therefore restricted in these cases to Principal Axis factor analyses without a goodness of fit test. However, in Chapter 5, a satisfactory goodness of fit was found in the CBAC population. A possible explanation for this fit is greater reliability and validity of the ASI-R problem severity scores in this sample of patients with alcohol and drug dependence problems, for which the ASI was originally designed.

PSYCHOMETRIC METHODOLOGY

Rasch Homogeneity

Psychiatric rating-scales are usually assessed in terms of reliability and validity in

accordance with the 'classical test theory'.[11] The validity of a scale is judged by comparison with other measures or a chosen external standard, such as an expert clinical judgement. However, the conceptual homogeneity of items in a supposed scale is not addressed by the classical test theory. The 'latent trait theory' or 'item response theory' has provided a theoretical framework to assess the coherence between the underlying construct, i.e. the latent trait, and the response behaviour on a set of items.[12]

The main reason for applying the Rasch model in this study was to be able to use the sum scores of the Bendep-SRQ scales. It is customary to divide questionnaire items into subsets (so-called scales) and to use the sum of the item scores as a measure for the underlying construct. However, the item sum score is only a sufficient statistic for the underlying construct or latent trait, if it reflects all information that is contained in the item scores. If more or different information is contained in the item scores, then more and other measures should be used. On the basis of this requirement and some other plausible assumptions (1. uni-dimensionality of the underlying trait; 2. continuous strictly monotone increasing item characteristic curves (ICC's); 3. local stochastic independence) the Rasch model was derived by the Danish statistician G. Rasch in 1960.[13] Formal proof of the Rasch model was later given by Fischer.[14] The use of the item sum score can only be justified by testing whether the Rasch model holds true for the item sets for which sum scores are used and precludes the application of any other scaling model. For example, in the case of the normal ogive model, which was proposed by Lawley[15,16] and discussed in more detail by Lord and Novick,[17] the sum score is not a sufficient statistic for the latent trait. Furthermore, the use of a sum score is certainly not justified if the Rasch assumption of continuous strictly monotone increasing ICCs does not hold true. In the case of questionnaires, the assumption of continuous single peaked ICCs might hold true instead. The PARELLA model, a unidimensional latent trait model for dichotomous items, is based on this alternative

assumption,[18-20] but does not justify the use of sum scores.

The development of test statistics for the Rasch scaling model have made it possible to test whether the assumptions of the Rasch model hold true. The Rasch Scaling Computer Program (RSP) has been developed to carry out these tests on sets of dichotomous items.[21,22] Analogously, a PARELLA computer program became available in 1994. Although polytomous Rasch models have been outlined theoretically by Andersen (1977) and Masters (1982),[12,23] a Rasch Scaling Program which can be applied to a set of polytomous items has not yet been developed. The use of the RSP therefore required dichotomization of the Bendep-SRQ items, which caused a considerable reduction in scale discriminability.

Reliability

The benefits of applying the dichotomous Rasch model with regard to the scalability of the Bendep-SRQ scales were accompanied by a loss of reliability. The limited number of items on the Rasch scales further suppressed the subject discriminability, as was demonstrated by some moderate KR-20 values. The item discriminability coefficients (IDC) and test-retest correlations were more convincing, as they were not affected by the limited number of items on the scales. The former parameter was newly developed in this study as a reliability measure for the item sum score, in addition to the well-known concept of internal consistency (i.e. subject discriminability) which reflects the reliability of the subject sumscore. Reviewing the whole situation, the reliability outcomes were considered to be sufficiently good.

Construct Validity

The items on the Rasch scales were arranged in a specific order based on increasing Rasch scale values, to reflect increasing severity of the underlying dimension, i.e. the latent trait. This offered a new approach to assess the construct validity of the Bendep-SRQ scales. Interpretation of the specific item order and the contents of the items made it possible to formulate theoretical rationales that reflect a more thorough understanding of the underlying dimensions. Obviously, this interpretation was subjective and the theoretical rationales given might be challenged by alternative ones. However, the structure provided in this process of interpretation by the Rasch scaling model was lacking before and could sufficiently guide the debate among clinical experts to reach consensus with respect to the formulation of such theoretical rationales. In Chapter 3, this approach to construct validity already proved to be useful in the comparison of the DSM-III-R and ICD-10 BZD dependence criteria, as it provided an empirically supported rationale for the systematic differences between the prevalences of the DSM-III-R and ICD-10 BZD dependence diagnoses found in Chapter 2.

Standardization

The Rasch homogeneity of the Bendep-SRQ scales provided the opportunity to apply and investigate a new standardization method, referred to as 'latent trait standardization', which transforms raw scores into latent trait scores. This required a Rasch model with the additional assumption of a normally distributed latent trait, which was tested by computation of the test statistic R0, while in the classical standardization method, the normality of the transformed distribution is merely assumed. In contrast with classical standardization, Rasch latent trait standardization has a sound theoretical basis. It tests all essential assumptions and yields estimated latent trait scores which properly reflect the underlying dimension of the scale. Chapter 7 showed that it was possible to analyse why the assumption of a normally distributed latent trait did not hold true for some of the Bendep-SRQ scales. Therefore, this standardization method offers important clues about how to improve Rasch scales.

UTILITY OF THE NEW STRUCTURED APPROACHES

The utility of the above-mentioned DSM-III-R, ICD-10 and Bendep-SRQ scales, with respect to the assessment of BZD dependence in clinical practice and scientific research, can be viewed from different angles.

Good psychometric properties can be considered as a basic requirement for the utility of a scale. The scalability, reliability and validity of the scales in this study were thoroughly assessed in order to fulfil this requirement. In clinical practice and scientific research different aspects of utility are emphasized. In clinical practice utility depends on the amount of time and training which is required to administer the scale regularly. Administration of the

substance dependence sections of the SCAN, which comprise the criteria of the DSM-III-R and ICD-10 BZD dependence scales**,** requires substantial training and time investment. The utility of these scales is therefore limited in clinical practice, but acceptable in scientific research when more time is available. On the other hand, administration of the Bendep-SRQ does not require any special training and takes a limited amount of time. The utility of the Bendep-SRQ was further improved by the computation of norms (Chapter 7) and the development of an on-line version on internet (*http://baserv.uci.kun.nl/~fzitman/Bendep-SRQ.html*), which automatically provides the sum scores after administration.

Regular use of rating scales is still not customary in general and psychiatric practice. Even the use of an instrument with good psychometric and practical properties requires an attitude which accepts aids and appliances. In the working alliance between a BZD user and a physician, the latter is not only challenged by needing to acknowledge BZD dependence as an alternative source of the symptoms, but he must also be prepared to treat BZD dependence and to use appropriate psychometric instruments for this purpose. The latter can be considered as an additional dimension to the working alliance, because it provides patient and physician with the means to assess the treatment process objectively and to create a collective frame of reference on this level.

MAJOR CONCLUSIONS

In this study, structured approaches to BZD dependence, based on the general criteria of the substance dependence syndrome and the specific criteria for BZD dependence, were evaluated.

On the basis of the DSM-III-R and ICD-10 criteria to diagnose BZD dependence, high

prevalence rates were encountered in different outpatient samples, suggesting that BZD dependence is a major health problem.[24] It should be noted that these findings rely on the validity of this diagnostic approach, which endorses the assumption that the elements of the substance dependence syndrome are homogeneous.[25,26]

In subsequent investigations, the assumption of homogeneity was challenged by means of Rasch modelling. Rasch homogeneity was required to demonstrate the scalability of the DSM-III-R and ICD-10 BZD dependence criteria and Bendep-SRQ items.[27-30] This starting-point marks a fundamental change in test methodology that favours the use of latent trait models, which was already advocated by Duncan-Jones et al. in 1986.[31] The application of this new methodology resulted in the delineation of Rasch homogeneous DSM-III-R, ICD-10 and Bendep-SRQ BZD dependence scales. It became apparent that some criteria, e.g. the withdrawal-related criteria, had to be removed from the DSM-III-R and ICD-10 constructs in order to uphold Rasch homogeneity. The present sets of BZD dependence criteria in the DSM-III-R and ICD-10 are therefore not in accordance with the above-mentioned homogeneity assumption, as specific revisions of these sets had to be made to regain this objective. The question remains as to whether these findings are valid with respect to BZD dependence only, or to substance dependence in general. If the latter is true, this will have consequences on the definition of substance dependence in future DSM and ICD editions. If not, BZD dependence can be separated as a different kind of dependence, due to its medical context.[27]

Rasch modelling was also applied in the developmental process of the Bendep-SRQ. Starting with a larger set of items, which were more specific to BZD dependence, four Rasch homogeneous Bendep-SRQ scales were delineated. Rasch homogeneity was a condition for subsequent reliability and validity assessments of these scales. To conclude the development of the Bendep-SRQ, Rasch modelling with the additional assumption of a normally

distributed latent trait appeared to be a useful and more advanced method of standardization.

In this study as a whole, Rasch modelling proved to be a valuable new methodology in the development of useful structured approaches to BZD dependence: it yielded stable and interpretable results. The evaluation of empirical BZD dependence data by means of this latent trait scaling model resulted in comprehensible psychometric constructs, which can be applied to achieve more uniformity and to avoid ambiguity in the assessment of BZD dependence.[32-34] The most tangible result of these new approaches was the multidimensional profile of BZD dependence provided by the sum scores of four Rasch homogeneous Bendep-SRQ scales. This can contribute to clinical management and applied scientific research with respect to BZD dependence.

RECOMMENDATIONS FOR FURTHER RESEARCH

The Bendep-SRQ can be improved further. The reliability and validity would gain from adding new items to the Bendep-SRQ, formulated in line with the theoretical rationales of the Bendep-SRQ scales (see Chapter 4). Formulating appropriate new items to bridge 'gaps' in the Rasch scales could improve 'equal item spacing' and prevent rejection of the Rasch model with the additional assumption of a normally distributed latent trait.

A model in which the sum of the original item scores, i.e. the item scores before dichotomization, is a sufficient statistic for the subject parameter, is known as the Partial Credit Model.[23] In 1977, Andersen proposed generalization of this model, in which the scale values of item score categories are estimated from the data.[35] When computer programs become available for these models, the data on the Bendep-SRQ items can be analysed using the original item response categories.As a result, scale discriminability and therefore the

reliability of the Bendep-SRQ scales would increase considerably.

Henceforth, the development of Rasch scales could be made more efficient by a 'circular' process of subsequent Rasch analyses on expanding item sets. Guided by interpretations of the specific item orders yielded by the Rasch analyses on item sets, new items could be formulated and added to these sets. The enlarged item sets could be administered again to the same patient sample, in order to subject their responses to Rasch analyses again and so on. By means of such a first stage, final Rasch scales with sufficient numbers of items and good psychometric properties can be constructed. In the following stage, these Rasch scales could be administered to a second patient sample to check the scalability, reliability and construct validity and to compare them to other instruments to assess concurrent and discriminant validity. This two-stage design would be less time-consuming and require fewer subjects for data acquisition. The results of this study, which lacked such a design, already showed that the

Rasch methodology has the potential to become a new standard in the field of test development.

The present study concentrated on outpatient BZD users from different outpatient settings. It would be worthwhile to repeat the investigations on psychiatric inpatients, general hospital inpatients and inpatients at drug- and alcohol centres. Theoretically, the Rasch model is claimed to be population-independent, so it can therefore be expected to hold true for such inpatient samples as well.[12] By comparing the Bendep-SRQ results between out- and inpatient populations, cut-off points on the Bendep-SRQ scales could be chosen as indicative for considering admission.

Longitudinal monitoring of BZD users with the Bendep-SRQ could yield valuable data to assess the predictive validity of the Bendep-SRQ scales. By following the course of continuing BZD use without any other intervention than repeated administration of the Bendep-SRQ, some of the scale scores might demonstrate predictive value with respect to spontaneous dose escalation or dose reduction. Combined with interventions such as patient education, motivational group therapy and/or a dose-reduction program, some of the scale scores might appear to have predictive value with respect to success or failure of these interventions. Based on the results and experience from such longitudinal studies with the Bendep-SRQ, a treatment protocol could be designed to guide the process of clinical decision-making, using the Bendep-SRQ profile of severity scores obtained through regular monitoring. The efficacy of such a protocol could be investigated in comparison with a control group that receives the customary treatment.

REFERENCES

1. Tyrer P, Rutherford D, Huggett T. Benzodiazepine withdrawal symptoms and propranolol. Lancet 1981;1:520-522.

2. Smith DE, Wesson DR. Benzodiazepine dependency syndromes. Journal of Psychoactive Drugs 1983;15:85-95.

3. Ashton H. Benzodiazepine withdrawal: outcome in 50 patients. Br J Addict 1987;82(6):665-671.

4. Ashton H. Protracted withdrawal syndromes from benzodiazepines. J Subst Abuse Treat 1991;8(1-2):19-28.

5. Hendriks VM. Addiction and psychopathology:a multidimensional approach to clinical practice. Rotterdam: V.M. Hendriks, 1990.

6. McLellan AT, Luborsky L, Cacciola J, Griffith J, Evans F, Barr HL, O'Brien CP. New data from the Addiction Severity Index. Reliability and validity in three centers. J nerv ment dis 1985;173(7):412-423.

7. Hodgins DC, el Guebaly N. More data on the Addiction Severity Index. Reliability and validity with the mentally ill substance abuser. J nerv ment dis 1992;180:197-201.

8. Hendriks VM, Kaplan CD, Limbeek Jv, Geerlings P. The addiction severity index: reliability and validity in a Dutch addict population. J Subst Abuse Treat 1989;6:133-141.

9. Amoureus M, P SR., van den Hurk AA., Schippers GM., Breteler MHM. The Addiction Severity Index in penitentiaries. International Journal of Offender Therapy and Comparative Criminology 1994;Vol 38(4):309-318.

10. Jong CAJd, Willems JCEW, Schippers GM, Hendriks VM. The addiction severity index: reliability and validity in a Dutch alcoholic population. Int J Addict 1995;30:5:605-616.

11. Nunnally JC. Wright RR and Gardner M, editors. Psychometric theory. New York: McGraw-Hill, 1967.

12. Verhelst ND. Itemresponstheorie. In: Eggen TJHM, Sanders PF (eds): Psychometrie in de praktijk. Arnhem: Cito Instituut voor Toetsontwikkeling, 1993:83-176.

13. Rasch G. Probabilistic models for some intelligence and attainment tests. Kopenhagen: Padagogischen Institut, 1960.

14. Fischer GH. Derivations of the Rasch Model. In: Fischer GH, Molenaar IW (eds): Rasch models: Foundations, recent developments and applications. New York: Springer Verlag, 1995:15-38.

15. Lawley DN. On problems connected with item selection and test construction. Proceedings of the Royal Society of Edinburgh 1943;61:273-287.

16. Lawley DN. The factorial analysis of multiple item tests. Proceedings of the Royal Society of Edinburgh 1944;62-A:74-82.

17. Lord FM, Novick MR. Statistical theories of mental test scores. Reading, Massachusetts: Addison-Wesley, 1968.

18. Hoijting H. A latent trait model for dichotomous choice data. Psychometrika 1990;55:641-656.

19. Hoijting H. The measurement of latent traits by proximity items. Applied psychological measurement 1991;15:153-170.

20. Hoijting H, Molenaar IW. Testing for DIF in a model with single peaked item characteristic curves: the

PARELLA model. Psychometrika 1992;57:383-398.

21. Glas CAW. RSP: Rasch Scaling Program. Groningen, The Netherlands: IEC ProGAMMA: 1993.

22. Glas CAW. RSP: Rasch Scaling Program User's Manual. Groningen, The Netherlands: IEC ProGAMMA: 1993.

23. Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982;47:149-174.

24. Kan CC, Breteler MHM, Zitman FG. High prevalence of benzodiazepine dependence in outpatient users, based on the DSM-III-R and ICD-10 criteria. Acta Psychiatr Scand 1997;96:85-93.

25. Edwards G. The alcohol dependence syndrome: a concept as stimulus to enquiry. Br J Addict 1986;81:171-183.

26. Edwards G, Gross MM. Alcohol dependence: provisional description of a clinical syndrome. Br Med J 1976;1:1058-1061.

27. Kan CC, Breteler MHM, Van der Ven AHGS, Zitman FG. An evaluation of the DSM-III-R and ICD-10 benzodiazepine dependence criteria using Rasch modelling. Addiction 1998;93(3):349-359.

28. Kan CC, Breteler MHM, Timmermans EAY, Van der Ven AHGS, Zitman FG. Scalability, reliability and validity of the Benzodiazepine Dependence Self-Report Questionnaire in outpatient benzodiazepine users. Compr Psychiatry 1999;40(4):283-291.

29. Kan CC, Breteler MHM, Van der Ven AHGS, Zitman FG. Assessment of benzodiazepine dependence in alcohol and drug dependent outpatients. (Submitted).

30. Kan CC, Breteler MHM, Van der Ven AHGS, Zitman FG. Cross-validation of the Benzodiazepine Dependence Self-Report Questionnaire in outpatient benzodiazepine users. (Submitted).

31. Duncan Jones P, Grayson DA, Moran PA. The utility of latent trait models in psychiatric epidemiology. Psychol Med 1986;16:391-405.

32. Miller NS, Mahler JC. Addiction to and dependence on benzodiazepines. Diagnostic confusion in clinical practice and research studies. J Subst Abuse Treat 1991;8:61-67.

33. Tyrer P. Benzodiazepine dependence: a shadowy diagnosis. Biochem Soc Symp 1993;59:107-119.

34. Linsen SM, Breteler MHM, Zitman FG. Defining benzodiazepine dependence: the confusion persists. European Psychiatry 1995;10:306-311.

35. Andersen EB. Sufficient statistics and latent trait models. Psychometrika 1977;42:69-81.