Epidemiology Deserves Better Questionnaires
by
The IEA European Questionnaire Group*

* The IEA European Questionnaire Group had the following members:

Wolfgang Ahrens, Ph.D. Bremen Institute for Prevention Research and Social Medicine, Bremen, Germany.
Jakob Bjørner, MD, Ph.D. The Health Institute, New England Medical Center, Boston, USA.
Mogens Grønvold, MD, Ph.D. Department of Health Services Research, University of Copenhagen, Denmark.
Karl Heinz Jöckel, Ph.D. Institute for Medical Informatics, Biometry and Epidemiology, University Clinics, Essen, Germany.
Lis Keiding, MD. Danish Institute of Clinical Epidemiology, Copenhagen, Denmark.
Jens M. Lauritsen, MD, Ph.D. Initiative for Accident Prevention, The County Council of Funen, Denmark.
Susan Levy-Desroches, MB, BS, MPH, Ph.D. Luxembourg.
Lenore Manderson, Ph.D., FASSA. Australian Centre for International and Tropical Health and Nutrition, The University of Queensland, Australia.
Frede Olesen, MD, Ph.D. Director of the Research Unit for Primary Care, University of Aarhus, Denmark.
Jørn Olsen, MD, Ph.D. The Danish Epidemiology Science Centre, University of Aarhus, Denmark.
Eero Pukkala, Ph.D. Finnish Cancer Registry, Helsinki, Finland.
Maria Manuela Morales Suárez-Varela, MD, Ph.D. Department of Public Health, University of Valencia, Clinical-Epidemiology Unit. Hospital "Dr. Peset". Valencia. Spain.
Manfred Wildner, MD, MPH. IBE, Ludwig-Maximilians-University, Munich, Germany.

All correspondance to Jørn Olsen, IEA Council Member. Danish Epidemiology Science Centre, Vennelyst Boulevard 6, 8000 Aarhus C, DK.
E-mail: jo@soci.au.dk, fax: +45 86131580

Contents

Abstract 2
Introduction 2
Why does questionnaire development attract so little attention? 3
Steps towards a valid questionnaire 5
Latent structure analysis 7
Towards the standardization of questionnaires 8
Four arguments in favour of better validation and standardization of questionnaires 8
Discussion and conclusions 9
References: 10

ABSTRACT

Background

Research findings are usually no better than thedata used. Since epidemiologists quite often have to rely upon self-reported data we should therefore make these data as good as possible. It is difficult to develop questionnaires of good standard. Too many epidemiological studies are based upon questionnaires which were developed without using existing knowledge; without proper pilot testing and without using proper expert advice. Several different questionnaires address the same topics and lack of available standardized questionnaires jeopardize comparability between studies and the ability to combine results from several studies.

Methods

This is an internet-based consensus paper from epidemiologists with expertise and interests in questionnaire development.

Conclusions

We believe much more emphasis and money should be given to the development of a selected number of standard questionnaires of high quality and we suggest that criteria should be developed for use in quality control related to questionnaire development.

We recommend a committee to be established to guide and inspire further work in this field. We also recommend that an archive of "ready for use" questionnaires in different languages should be made available, e.g. on the Internet.

INTRODUCTION

Use of questionnaires is an essential epidemiological tool. Epidemiological findings are often based partly or completely on responses to questionnaires, which are used extensively for collecting information on exposures, outcomes, modifiers and confounders. Obviously, the use of an invalid instrument is simply a waste of time and money. The adequate preparation of questionnaires is thus essential for the quality of data. Yet the attention given to

questionnaire development and validation is often inadequate, compared with the amount of time and resources devoted to study design, population selection and sampling.

Indeed, when it comes to questionnaire development, the concerns of willingness to respond, discriminatory power, comparability, responsiveness/reliability and validity of data seem sometimes to be forgotten. This problem appear to be inherent to many areas of medical science, whenever questionnaires are concerned; the valuable body of knowledge and experience of other disciplines, especially the cognitive and social sciences, is often ignored (1).

Most of the developmental work of relevance to epidemiology has been carried out in the area of outcome measures, usually of non-observable outcomes such as depression, headache and intelligence. Relatively little work has been undertaken to develop questionnaires related to exposures, with certain exceptions.

Unfortunately, even when well constructed questionnaires exist, they are not always used by other researchers and opportunities for improvement are thus missed. Frequently, questionnaires are designed with little if any consultation of past research experience with similar measurement tools. Even when the effort is made, copies of past questionnaires may be difficult or even impossible to locate and obtain. Apparently, the standards for storing questionnaires as part of research documentation vary from country to country. Much valuable experience and information is never transmitted, and the resulting duplication of effort is most unfortunate. Even when researchers do not start from scratch when designing questionnaires, they often believe that they have good ideas for improving an existing instrument by changing the wording or introducing different exposure categories. The result, however, is a loss of comparability with other studies, often without documented benefits.

A worse situation is the continuing use of instruments that are known to be invalid. For example, in occupational epidemiology, exposure check-lists are often close to useless. Yet some researchers consistently use these instruments.

The aim of this paper is to explore the problem of questionnaire development in epidemiology, and to propose ways in which the situation might be improved. We do not offer critiques of particular studies, although such critical reviews are of value to questionnaire design and validation. We argue that more attention needs to be given to the development of high quality questionnaires, and that less time and money should be spent on data collection based on unvalidated questions. In particular, we ask for better collective use of the experience of questionnaire constructors worldwide. We hope to open up the debate that is a necessary step towards increasing the rigour of epidemiological research and the value of its findings.

WHY DOES QUESTIONNAIRE DEVELOPMENT ATTRACT SO LITTLE ATTENTION?

The field of validated questionnaire development within epidemiology suffers from low status. The task is seen by many as time consuming and boring (with the possible exception of developing latent structure questionnaires for which the statistical and conceptual principles are challenging and exciting). This may indeed be true for the epidemiologist who fails to recognize the importance of developing accurate exposure measurements, which after all are essential if valid scientific conclusions are to be reached. It is our undocumented

observation that funding policies do not seem to encourage investments in areas in which the output is not highly visible.

It is not customary in epidemiology to seek copyright protection for questionnaires. Some may see this as an advantage; yet the lack of recognition for the creators of a high quality questionnaire can only further reduce the incentive to undertake developmental research. Until the value of the work is better appreciated, no real progress will be made.

The importance of questionnaire design is frequently ignored in the published epidemiological literature as well. Published results often fail to reproduce the exact wording of key questions used to define exposures or outcomes, nor do they always provide adequate information on how the data collection instruments were developed, or if procedures such as pretesting, validity checks, or pilot studies were used to ensure accuracy. Even when questionnaires are reported as validated, the meaning of this is usually unclear. In the absence of information on the development, pretesting and piloting of instruments, this could mean anything from "my partner did not complain when completing the questionnaire", to years of hard work spent in developing several versions of the questionnaire.

In many European countries, epidemiology is not a well established discipline, and there has been little investment in the development of an infrastructure. Many research units are small and have insufficient financial and human resources for long term projects. If public health is not a priority in a given country few resources are directed towards the development of epidemiological infrastructures. Furthermore, there is a clear preference to fund research in technical sciences (such as molecular biology) rather than in epidemiology (2). In addition, public health practice agencies often have neither the budget nor expertise for research. The international epidemiological societies are disparate, and lack resources for undertaking comprehensive research programmes of their own.

For most purposes the perfect questionnaire has yet to be developed and may well be an elusive goal. Nevertheless, the quality of a questionnaire can and should be examined on the basis of a set of criteria. We need to establish a set of generally accepted criteria.

Table 1. **Some reasons for the lack of good quality questionnaires**
1. Lack of economic incentives (no copyright tradition) 2. Lack of scientific incentives (low status) 3. Most units too small to take up the work 4. Lack of criteria for good quality questionnaires (3) 5. Difficult to locate and obtain developed questionnaires 6. Ignorance of the importance of accurate measurements

STEPS TOWARDS A VALID QUESTIONNAIRE

One way of ensuring the quality of data collected by questionnaire is to use only questionnaires which have been validated. A validated questionnaire is one which has undergone a validation procedure to show that it accurately measures what it aims to do, regardless of who responds, when they respond, and to whom they respond. Questions on smoking are valid if they provide accurate data on actual smoking behaviour at any time of the day or year, to any interviewer regardless of various personal ascriptions (e.g. age, sex, ethnicity) or when self-administered. A validation procedure examines these issues and should also include the ability to detect known associations between exposures and health.

Bias is to a degree inherent in the use of questionnaires, external to the questions themselves. Questionnaire validation should establish the range of and reasons for inaccuracies and potential sources of bias. The validation of questionnaires reduces bias by detecting ambiguities and misinterpretations which can then be minimized.

A questionnaire should not only provide valid responses but also be well accepted by the responders. Willingness to participate and to provide the wanted information is of course important.

In ideal situations, an independent "gold standard" is available when developing the questionnaire. This "gold standard" might be too expensive or too difficult to use in large scale data collection, but can be applied in a validation exercise. If smoking behaviour can be recorded by independent observers or reliable information exists from biomarkers, as is the case for short-term exposure, it is possible to test a questionnaire by comparing results from the two sets of information. The problem is, however, that the result may not be generalizable if respondents know that they are being observed. People may change their smoking behaviour as a result of surveillance, which would make results "true" (valid and reliable in the setting) but irrelevant (because the validity is time and place specific). The use of biological markers to test the reliability of responses (for example, urine tests to validate reports of chemotherapeutic compliance) may be useful for specific purposes.

In many cases, no "gold standards" are available, as is the case for alcohol consumption, stress exposure, exposure to sunlight, and for many other physical and chemical exposures. Alternative methods of measurement may exist, yet not necessarily closer to the "true values" than self-reports obtained by questionnaires. In these situations, we cannot estimate the extent of measurement error introduced by questionnaire data. Even so, a "non-gold measure" may still serve a useful purpose in estimating the extent to which external factors modify the response to the questionnaire (4,5). This relates to the second part of the validation procedure.

Rasch (6) used the term "specific objectivity" to describe a measuring instrument to which responses depend only upon the question asked, and on whether or not the conditions or events that are the subject of enquiry are present. Although this starting point is naive it sets the direction for the work. The setting and circumstances in which the questionnaire is administered play an important role. A good questionnaire seeks to minimise situational effects. Part of the process of questionnaire development is to obtain information on factors that influence response, especially those that may bias the key associations under study. If the intention is to measure the actual exposure to an environmental factor, the epidemiologist

should be aware of the bias that may arise if the questions give the impression that responders should only confirm exposures associated with adverse effects.

The sex, ethnicity and age of the interviewer routinely influence the validity of data, especially in research concerned with fertility, contraceptive practice, sexual health, HIV/AIDS, and women's health in general. This issue must be addressed at the point of data collection (e.g. with respect to choice of interviewer and/or decisions regarding self-reports, telephone surveys, and so on)(5-18).

Translation followed by independent back-translation and comparing the source with the translated text is an important method to test if questions are phrased unambiguously. No specific rules exist for how close independent translations and back-translations should be in order to be acceptable, and it would perhaps be impossible to set fixed rules. The aim is to ensure that ambiguity is minimal. Translation and back-translation is also a method of examining cultural differences in the understanding of underlying concepts (19,20). Such cultural differences exist both within and between countries, regions, ethnic groups and social classes. It is also important to recognise that differences of specificity and meaning exist between lay and biomedical uses of terms like "stress", "intoxication","anxiety", "tumour", and so on. Variation in response is compounded where subjective assessments are required, for example, in describing pain.

Repeated use of the same questionnaire within the same study population may help to test the reliability of responses and to identify ambiguity in questions, although altered recall of events or recall of previous response may be a disturbing factor. Repeating the same questionnaire while changing the circumstances from, say, a male to female interviewer, telephone interviewing to personal interviewing, interviewing to self-administration of questionnaires, and so on, may also be a useful way to identify ambiguity.

Besides validity of the questionnaire, reliability deserves special attention. Even if a questionnaire is not completely valid (which it hardly ever will be) the reliability of the instrument has a value of its own. If an instrument is reliable this offers the opportunity to compare results from one study with work that has been done by others. This is of special value if measurements are taken at different points in time or in different regions.

High reliability is particular important for the independent variables in statistical analyses: exposures, modifiers, and confounders. Low reliability in such determinants can bias the data analysis in unpredictable ways, causing both underestimation and overestimation of effects (21). Strangely, most efforts of providing measures of high reliability have been done for outcome measures.

A validated questionnaire has usually been investigated by comparing responses to different instruments and in different situations. However, a validated questionnaire should not be assumed to be valid under all circumstances; data quality control remains an important part of any project.

Table 2 **Important elements of the validation procedure**
• Compare with a "gold standard" • Compare with other sources of data • Examine reliability • Use translation and back translation to reduce ambiguity and aim at a high degree of "specific objectivity" • Examine feasibility: acceptability, time needed to respond, cost etc. • Examine variation in response due to data inquiry methods (self-administered, personal interview, telephone interview etc.)

LATENT STRUCTURE ANALYSIS

Questionnaires may be used to collect information on data that could be recorded directly (e.g. food intake) but are not because the actual measuring is too expensive, time consuming or likely to introduce bias. Questionnaires could also seek to gather information that is not available by other means. In such cases questionnaires should have construct validity i.e. questions and the questionnaire should have a clear relation to the underlying concept or theory.

A popular approach to measuring such concepts is to ask several questions that each is an indirect indicator of the concept and then combine the answers to these questions into a multiitem scale. When analysing such scales, we could also learn from the expertise developed within latent structure analysis, especially concerning statistical principles related to construct validity and item bias. If the concept exists and has a one dimensional structure, a single number of "positive" responses provides sufficient information on a latent disease such as depression. The questions (items) aimed at measuring a depression must act independently of each other given the scale, and not be influenced by external factors. Numerous possibilities of checking these conditions exist (22) and the best scales have been exposed to several types of analyses.

Criterion validity is used to describe a correlation between two scales measuring the same structure, often a reduced version versus a more complete set of questions, or between a scale and a "gold standard" e.g. a standardized cognitive test. As part of the evaluation, the predictive value of the obtained information, for example in the form of predicting outcomes, should be examined. If the scale deals with depression, for example, the outcome could be hospitalization, suicide attempts, absence from work, and so on.

TOWARDS THE STANDARDIZATION OF QUESTIONNAIRES

The development of standardized ways of asking questions on specific topics, that is, of standard questionnaires, might usefully be advanced by involving scholars from numerous disciplines in the establishment of a clearing house, a committee or working group acceptable to the broader scientific community. A standard questionnaire is a questionnaire recognized and approved by this body. The committee might commence by evaluating questionnaires for which substantial validation is already available, and continue by identifying exposures and outcomes for which valid measures need to be found. A classification of these measures in order of priority based on the importance and frequency of use of the data could be established.

The committee could consider the issues of cost, feasibility, and ethics and should be able to set up workshops or task forces to develop specific questionnaires of interest. The committee should furthermore set up an annotated questionnaire library, preferably on the Internet.

Funding will be needed to conduct reviews of existing material and to establish the necessary studies to carry this work forward, but it would be money well spent because of the wide benefit to the scholarly community and in terms of public health gains.

Table 3. **Elements in improving questionnaires**
• A standing committee or a department devoted to develop the field • Better education in questionnaire methodology • Agreement on criteria for a "standard questionnaire" • Editors should encourage publication of methodological research devoted to improve data collection in epidemiology • Establish a public archive of standardized questionnaires (a questionnaire bank) and make it available on the Internet

FOUR ARGUMENTS IN FAVOUR OF BETTER VALIDATION AND STANDARDIZATION OF QUESTIONNAIRES

1. The quality of data will be improved. Most questionnaires developed on an *ad hoc* basis contain errors; some of these errors can be detected in the validation process and subsequently corrected.

2. Comparability which is a prerequisite for meta-analyses and essential when trying to corroborate existing results will be enhanced. Our knowledge of the aetiology of and risk factors for particular health problems remains restricted by the lack of comparability of information from within different populations and over time. Descriptive epidemiology of this type has been very important for analytical research in cancer epidemiology and many other areas.

3. Duplication of effort will be reduced. Much of the time spent on developing questionnaires will be saved, improving the cost-effectiveness of epidemiological research. The time saved could be spent on data quality control.

4. The credibility of epidemiology as well as the quality and usefulness of information thus derived will be improved.

DISCUSSION AND CONCLUSIONS

Epidemiologists need to take the development of research instruments and the validity of questionnaire data more seriously. We should demand higher standards from ourselves and our colleagues. The development of questionnaires will, however, only receive the attention it deserves if editors, peer reviewers and funding agencies demand use of validated questionnaires.

Due to limited resources, it is unlikely that the work of assessing the potential for and of developing standard questionnaires could be organised and funded by the epidemiological community alone. It is a task that should be supported by the WHO or the European Union (EU). The EU Section on Public Health Research could take responsibility for strengthening these centralised facilities that would benefit epidemiology in all European countries (as well as in other countries). It is the responsibility of epidemiologists and their societies to advocate for this to occur.

Most of the work undertaken until now has concentrated on the validation of outcome questionnaires. Substantive work has been done in the areas of general health, quality of life and psychiatric disorders (23-43), evaluation of health care (44-54), pain (54-61) and other topics (62-79). In exposure assessment, work has been devoted to developing questionnaires in areas such as dietary intake (80-86), alcohol use (87-93), smoking (94) and occupational exposures (95-100). The assessment of exposure to indoor air pollutants has been the subject for some time (101) and is also described by a recent WHO-ECEH working group (102).

Although substantial effort has been devoted to questionnaire validation, much of the work has been done but too often in isolation. Epidemiologists too often repeat many of the old mistakes and learn little from past experience. We see a need to develop standardized questionnaires particularly on smoking (including passive smoking), alcohol use, use of tap water, consumption of medicine, physical exercise, occupational and domestic exposures, among others, and for outcomes such as subfecundity (23,24), sleeping disorders, and functional deficits.

Further work needs to be undertaken to assess current questionnaires and to develop recognised standards. We believe that much is to be gained if researchers, funding bodies and public health agencies gave higher priority to the field of questionnaire development. WHO has started part of the work process (103).

References:

1. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: The impact of memory and inference on surveys. Science 1987;236:157-61.
2. Gordon S. 1991. The History and Philosophy of Social Science. London: Routledge.
3. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. Clinical Therapeutics 1997;18:979-92.
4. Basso O, Olsen J, Bisanti L, Karmaus W, and the European Study Group on Infertility and Subfecundity. The performance of several indicators in detecting recall bias. Epidemiology 1997;8:269-79.
5. Ipsen J, Olsen J. Estimating sensitivity and specificity in order to correct for misclassification. Scand J Soc Med 1984;12:111-4.
6. Rasch G. On specific objectivity. Danish Yearbook of Philosophy. Vol 14. Munksgaard, Copenhagen 1977.
7. Bleek W. 1987. Lying informants: a fieldwork experience from Ghana. *Population and Development Review* **13**: 314-322.
8. Stone L and Campbell J G. 1984. The use and misuse of surveys in international development: an experiment from Nepal. *Human Organization* **43**: 2737.
9. McDowell I, Newell C. Measuring health. A guide to rating scales and questionnaires. Oxforxd University Press. Oxford 1996.
10. Olsen J. Measurement techniques. Questionnaire, telephone interview, clinical interview - validation and reliability. In: Headache classification and epidemiology. Olesen J (eds). Raven Press, New York 1994.
11. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Monographs on epidemiology and biostatistics, Vol. 21. Oxford, New York, Tokio. Oxford University Press, 1992 (351 pages).
12. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622-9.
13. Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. Qual Life Res 1993;2:451-9.
14. Streiner DL, Norman GR. Health measurement scales - A practical guide to their development and use, Oxford: Oxford University Press, 1995.
15. Joffe M. Validity of exposure data derived from a structured questionnaire. Am J Epidemiol 1992;135(5):564-70.
16. Coggon D. Questionnaire based exposure assessment methods. Science Total Environ 1995;168(2):175-8.
17. Baumann LJ, Adair EG. The use of ethnographic interviewing to inform questionnaire construction. Health Educ Quarterly 1992;19(1):9-23.
18. Heliövaara M, Aromaa A, Klaukka T, Knekt P, Joukamaa M, Impivaara O. Reliability and validity of interview data on chronic diseases. J Clin Epidemiol 1993;46(2):181-91.
19. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaption of health related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993;46:1417-32.
20. Ware JE, Jr., Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health Status questionnaires. Int J Technol Assess Health Care 1995;11:525-51.
21. Aickin M, Ritenbaugh C. Analysis of multivariate reliability structures and the induced bias in linear model estimation. Stat Med 1996;15:1647-61.

22. Kreiner S. Validation of index scales for analysis of surveydata: the symptom index. In: Dean K. (ed). Population health research. Linking theory and methods. Sage Publications London 1993.
23. Marchbanks PA, Peterson HB, Rubin RL, Wingo PA. Research on infertility: Definition makes a difference. Am J Epidemiol 1989;130:259-67.
24. Vlassoff C and Vlassoff M. 1978. Misreporting in rural fertility data: An analysis of husband-wife disagreement. *Journal of Biosocial Science* **10**: 437-444.
25. Febbo S, Burvill PW. Validation of an Italian translation of the 30-item general health questionnaire for use in Australia. Aust N Z J Psychiatry 1995;29:266-9.
26. Ahrens W, Jöckel K-H, Brochard P, Bolm-Audorff U, Grossgarten KH, Iwatsubo Y, Orlowski E, Pohlabeln H, Berrino F. Retrospective assessment of asbestos exposure. I. Case-control analysis of job-specific questionnaires and job exposure matrices. Int J Epidemiol 1993;22(Suppl. 2):83-95.
27. Piccinelli M, Bisoffi G, Bon MG, Cunico L, Tansella M. Validity and test-retest reliability of the Italian version of the 12-item general health questionnaire in general practice: A comparison between three scoring methods. Compr Psychiatry 1993;34(3):198-205.
28. Main DS, Pace WD. Measuring health: guidelines for reliability assessment. Fam Med 1991;23:227-30.
29. Politi PL, Piccinelli M, Wilkinson G. Reliability, validity and factor structure of the 12-item general health questionnaire among young males in Italy. Acta Psychiatr Scand 1994;90:432-7.
30. Walsh J, Joseph S, Lewis CA. Internal reliability and convergent validity of the depression-happiness scale with the general health questionnaire in an employed adult sample. Psychol Rep 1995;76:137-8.
31. Poole JL, Williams CA, Bloch DA, Hollak B, Spitz P. Concurrent validity of the health assessment questionnaire disability index in scleroderma. Arthritis Care Res 1995;8(3):189-93.
32. Lyons RA, Perry HM, Littlepage BNC. Evidence for the validity of the short-form 36 questionnaire (SF-36) in an elderly population. Age Ageing 1994;23(3):182-4.
33. Hollen PJ, Gralla RJ. Comparison of instruments for measuring quality of life in patients with lung cancer. Semin Oncol 1996;23(2):31-40.
34. Hyland ME, Crocker GR. Validation of an asthma quality of life diary in a clinical trial. Thorax 1995;50:724-30.
35. Groenvold M, Bjorner JB, Klee MC, Kreiner S. Test for item bias in a quality of life questionnaire. J Clin Epidemiol 1995;48(6):805-16.
36. Marquis P, Fayol C, Joire JE, Leplège A. Psychometric properties of a specific quality of life questionnaire in angina pectoris patients. Qual Life Res 1995;4:540-6.
37. Cockburn J, Luise TD, Hurley S, Clover K. Development and validation of the PCQ: A questionnaire to measure the psychological consequences of screening mammography. Soc Sci Med 1992;34(10):1129-34.
38. Egger MJ, Ward JR, Karg MB, Williams HJ, Reading JC and the Cooperative Systematic Studies of Rheumatic Diseases. Reliability and Validity of the CSSRD functional assessment survey in rheumatoid arthritis. Arthritis Care Res 1995;8(1):21-7.
39. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Qual Life Res 1993;2:169-80.
40. Carta MG, Carpiniello B, Cicone V, Sannais C, Paulis M, Rudas N. Standardization of a psychiatric screening test for use by general practitioners in Sardinia. Acta Psychiatr Scand 1993;87:342-4.

41. Kitamura T, Shima S, Sugawara M, Toda MA. Temporal variation of validity of self-rating questionnaires: repeated use of the general health questionnaire and Zung's self-rating depression scale among women during antenatal and postnatal periods. Acta Psychiatr Scand 1994;90:446-50.
42. Grossarth-Maticek R, Eysenck HJ, Boyle GJ. Method of test administration as a factor in test validity: the use of a personality questionnaire in the prediction of cancer and coronary heart disease. Behav Res Ther 1995;33(6):705-10.
43. Banks MH, Clegg CW, Jackson PR, Kemp NJ, Stafford EM, Wall TD. The use of the general health questionnaire as an indicator of mental health in occupational studies. J Occup Psychol 1980;53:187-94.
44. Groenvold M. Validation of a quality of life questionnaire for breast cancer patients. Koebenhavns Universitet, Koebenhavn 1996.
45. Groenvold M, Klee MC, Sprangers MAG, Aaronson NK. Validation of the EORTC QLQ-C30 quality of life questionnaire through combined qualitative and quantitative assessment of patient-observer agreement. J Clin Epidemiol 1997;50:441-50.
46. Gloth III FM, Walston J, Meyer J, Pearson BS, Pearson J. Reliability and validity of the frail elderly functional assessment questionnaire. Am J Phys Med Rehabil 1995;74(1):45-53.
47. Baker R. Development of a questionnaire to assess patients' satisfaction with consultations in general practice. Br J Gen Pract 1990;40:487-90.
48. Eisen SV, Dill DL, Grob MC. Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. Hosp Community Psychiatry 1994;45(3):242-7.
49. Sprangers MAG, Cull A, Bjordal K, Groenvold M, Aaronson NK. The European organization for research and treatment of cancer approach to quality of life assessment: guidelines for developing questionnaire modules. Qual Life Res 1993;2:287-95.
50. Carey RG, Seibert JH. A patient survey system to measure quality improvement: questionnaire reliability and validity. Medical Care 1993;31(9):834-45.
51. Rubin HR, Ware JE, Hays RD. The PJHQ questionnaire. Exploratory factor analysis and empirical scale construction. Medical Care 1990;28(9):S22-S29.
52. Wilding RJC, Osman YI. A technique for the construction of complete dentures in two visits. II. A pilot study to evaluate the opinions of practitioners and patients. Quintessence Int 1990;21(7):559-62.
53. Punch KF, Horner B. Stages in the development of a needs assessment questionnaire. J Nursing Staff Development 1991;July/August:176-80.
54. Boult C, Pacala JT, Boult LB. Targeting elders for geriatric evaluation and management: reliability, validity, and practicality of a questionnaire. Aging Clin Exp Res 1995;7:159-64.
55. Wessel J. The reliability and validity of pain threshold measurements in osteoarthritis of the knee. Scand J Rheumatol 1995;24:238-42.
56. Strand LI, Wisnes AR. The development of a Norwegian pain questionnaire. Pain 1991;46:61-6.
57. Burckhardt CS, Clark SR, Bennett RM. The fibromyalgia impact questionnaire: development and validation. J Rheumatol 1991;18(5):728-33.
58. Rasmussen BK, Jensen R, Olesen J. Questionnaire versus clinical interview in the diagnosis of headache. Headache 1991;31:290-5.
59. Boureau F, Luu M, Doubrère JF. Comparative study of the validity of four French McGill pain questionnaire (MPQ) versions. Pain 1992;50:59-65.

60. Holmström E, Moritz U. Low back pain - correspondence between questionnaire, interview and clinical examination. Scand J Rehab Med 1991;23:119-25.
61. Donaldson GW. The factorial structure and stability of the McGill pain questionnaire in patients experiencing oral mucositis following bone marrow transplantation. Pain 1995;62:101-9.
62. Lanata CF, Quintanilla N, Verastegui HA. Validity of a respiratory questionnaire to identify pneumonia in children in Lima, Peru. Int J Epidemiol 1994;23(4):827-34.
63. Wijkstra PJ, TenVergert EM, Van Altena R, Otten V, Postma DS, Kraan J, Koëter GH. Reliability and validity of the chronic respiratory questionnaire (CRQ). Thorax 1994;49:465-7.
64. McKinlay KP, Venables KM. Respiratory symptoms questionnaire for asthma epidemiology: validity and reproducibility. Thorax 1993;48:1289.
65. Torén K, Brisman J, Järvholm B. Asthma and asthma-like symptoms in adults assessed by questionnaires. Chest 1993;104:600-08.
66. Stucki G, Stucki S, Brühlmann P, Maus S, Michel BA. Comparison of the validity and reliability of self-reported articular indices. Br J Rheumatol 1995;34:760-6.
67. Cook J, Burd L. Preliminary report on construction and validation of a pediatric sleep disturbance questionnaire. Perceptual Motor Skills 1990;70:259-267.
68. Kump K, Whalen C, Tishler PV, Browner I, Ferrette V, Strohl KP, Rosenberg C, Redline S. Assessment of the validity and utility of a sleep-symptom questionnaire. Am J Respir Crit Care Med 1994;150:735-41.
69. Brener ND, Collins JL, Kann L, Warren CW, Williams BI. Reliability of the youth risk behavior survey questionnaire. Am J Epidemiol 1995;141(6):575-80.
70. Katz JN, Phillips CB, Poss R, Harrast JJ, Fossel AH, Liang MH, Sledge CB. The validity and reliability of a total hip arthroplasty outcome evaluation questionnaire. J Bone Joint Surg 1995;77(10):1528-34.
71. Gauthier-Gagnon C, Grisé M-C. Prosthetic profile of the amputee questionnaire: validity and reliability. Arch Phys Med Rehabil 1994;75:1309-14.
72. Kawas C, Segal J, Stewart WF, Corrada M, Thal LJ. A validation study of the dementia questionnaire. Arch Neurol 1994;51:901-6.
73. Coker S, Roger D. The construction and preliminary validation of a scale for measuring eating disorders. J Psychosom Res 1990;34(2):223-31.
74. Sno HN, Schalken HFA, DeJonghe F, Koeter MWJ. The inventory for déja vu experiences assessment. Development, utility, reliability, and validity. J Nerv Ment Dis 1994;182(1):27-33.
75. Pomerleau CS, Carton SM, Lutzke ML, Flessland KA, Pomerleau OF. Reliability of the Fagerstrom tolerance questionnaire and the Fagerstrom test for nicotine dependence. Addict Behav 1994;19(1):33-9.
76. Smith MF, Nathan RG, Wayner DS, Mitnick NC. Comparative validity of two hearing loss screening questionnaires. J Fam Pract 1992;35(4):411-6.
77. Mizes JS. Construct validity and factor stability of the anorectic cognitions questionnaire. Addict Behav 1991;16:89-93.
78. Andersson G, Melin L, Lindberg P, Scott B. Development of a short scale for self-assessment of experiences of hearing loss. Scand Audiol 1995;24:147-54.
79. Higson JM, Haggard MP, Field DL. Validation of parameters for assessing obscure auditory dysfunction - robustness of determinants of OAD status across samples and test methods. Br J Audiol 1994;28:27-39.
80. Kushi LH. Gaps in epidemiologic research methods: design considerations for studies that use food-frequency questionnaires. Am J Clin Nutr 1994;59(suppl):180S-4S.

81. Andersen LF, Nes M, Lillegaard IT, Sandstad B, Bjoerneboe G-EAa, Drevon CA. Evaluation of a quantitative food frequency questionnaire used in a group of Norwegian adolescents. Eur J Clin Nutr 1995;49:543-54.
82. Beer-Borst S, Amadò R. Validation of a self-administered 24-hour recall questionnaire used in a large-scale dietary survey. A Ernaehrungswiss 1995;34:183-9.
83. Shearin EN, Russ MJ, Hull JW, Clarkin JF, Smith GP. Construct validity of the three-factor eating questionnaire: flexible and rigid control subscales. Int J Eating Disorders 1994;16(2):187-98.
84. Männistö S, Virtanen M, Mikkonen T, Pietinen P. Reproducibility and validity of a food frequency questionnaire in a case-control study on breast cancer. J Clin Epidemiol 1996;49(4):401-9.
85. Riley MD, Blizzard L. Comparative validity of a food frequency questionnaire for adults with IDDM. Diabetes Care 1995;18(9):1249-54.
86. Kendall A, Olson CM, Frongillo EA. Validation of the Radimer/Cornell measures of hunger and food insecurity. J Nutr 1995;125:2793-2801.
87. Hittner JB. Factorial validity and equivalency of the alcohol expectancy questionnaire tension-reduction subscale across gender and drinking frequency. J Clin Psychol 1995;51(4):563-76.
88. Groenbaek M, Heitmann BL. Validity of self-reported intakes of wine, beer and spirits in population stuGdies. Eur J Clin Nutr 1996;50:487-90.
89. Jones TV, Lindsey BA, Yount P, Soltys R, Farani-Enayat B. Alcoholism screening questionnaires: Are they valid in elderly medical outpatients? J Gen Intern Med 1993;8(Dec):674-8.
90. Bohn MJ, Krahn DD, Staehler BA. Developing and initial validation of a measure of drinking urges in abstinent alcoholics. Alcohol Clin Exp Res 1995;19(3):600-6.
91. Liskow B, Campbell J, Nickel EJ, Powell BJ. Validity of the CAGE questionnaire in screening for alcohol dependence in a walk-in (triage) clinic. J Stud Alcohol 1995;56:277-81.
92. Williams BTR, Drummond DC. The alcohol problems questionnaire: reliability and validity. Drug Alcohol Depend 1994;35:239-43.
93. Adelekan ML, Gowers S, Singh D. Reliability and validity of the Westminster substance use questionnaire among Lancashire adolescents. Drug Alcohol Depend 1994;34:191-9.
94. Gulick EE, Escobar-Florez L. Reliability and validity of the smoking and women questionnaire among three ethnic groups. Public Health Nurs 1995;12(2):117-26.
95. Ahlborg GA. Validity of exposure data obtained by questionnaire. Two examples from occupational reproductive studies. Scand J Work Environ Health 1990;16:284-8.
96. Chornick NL, Yocom CJ. NCLEX job analysis study: questionnaire development. J Nurs Educ 1995;34(3):101-5.
97. Broersen JPJ, Weel ANH, van Dijk FJH, Verbeek JHAM, Bloemhoff A, van Duivenbooden JC. The atlas of health and working conditions by occupation. 2. A comparison with the "atlas of health and working conditions in the construction industry". Int Arch occup Environ Health 1995;67:337-42.
98. Ahrens W, Jöckel K-H, Pohlabeln H, Bolm-Audorff U, Iwatsubo Y, Brochard P, Berrino F, Orlowski E. Assessment of exposure to asbestos in a case-control study of lung cancer: comparison of supplementary questionnaires and exposure check-list. Occupational Hygiene 1996;3:125-36.
99. Orlowski E, Pohlabeln H, Berrino F, Ahrens W, Bolm-Audorff U, Grossgarten KH, Iwatsubo Y, Jöckel K-H, Brochard P. Retrospective assessment of asbestos exposure.

II. At the job level: Complimentary of job-specific questionnaire and job exposure matrices. Int J Epidemiol 1993;22(Suppl. 2):96-105.
100. Pols MA, Peeters PHM, Bueno-de-Mesquita HB, Ocké MC, Wentink CA, Kemper HCG, Collette HJA. Validity and repeatability of a modified Baecke questionnaire on physical activity. Int J Epidemiol 1995;24(2):381-8.
101. Lebowitz MD, Quackenboss J, Soczek ML, Kollander M, Colome S. The new standard environmental inventory questionnaire for estimation of indoor concentrations. J Air Pollut Assoc 1989;39:1411-19.
102. Jantunen M, Jaakola JJK, Krzyzanowski M. (eds). Assessment of exposure to indoor air pollutants. Copenhagen, WHO. Regional Office for Europe, 1997 (WHO Regional Publications, European Series, No. 78 (under publication).
103. DeBruin A, Picavet HSJ, Nossikov A. Health interview surveys. Towards international harmonization of methods and instruments. Copenhagen, WHO Regional Office for Europe, 1996 (WHO Regional publications, European Series, No.58).