## ORIGINAL PAPER

Manfred M. Fichter · Norbert Quadflieg · Gabriele Kohlboeck · Horst Dilling · Thomas Niedermeier

# The Upper Bavarian longitudinal community study on psychopathology 1975–2004: 1. Methods and first results

■ **Abstract** *Objective* This paper presents background information on the methods and first results of the 25-years follow-up of the prospective longitudinal Upper Bavarian Study in the community. Longitudinal epidemiological studies which cover very long time spans require special methods. Issues concerning these requirements are discussed using design and experiences from the Upper Bavarian Study. *Method* Assessments focused on three time points: baseline survey, five-year follow-up, and 25-year follow-up. Self-rating scales as well as expert-rating interviews yielded data on a wide range of social and psychopathological risk factors and outcome measures. *Results* Of the 1,342 study participants in the original sample, 390 had died during the 25-year follow-up period. Participation rate was 88% of those alive. At all three time points, a total of 838 participants were interviewed. Data on this unselected sample over a quarter century will be presented in forthcoming papers. Because of their mobility young individuals from our original sample were more difficult to follow up. We analyzed in detail data of subjects interviewed at t3 (and earlier time points) as compared to data from subjects not traced or reached, and subjects who refused to participate at wave t3. We found no evidence that the long-term outcome was biased by drop-out due to gender or earlier mental illness. *Conclusion* Methodological possibilities and limitations concerning long-term epidemiological studies across decades are presented and discussed.

■ **Key words** community study · longitudinal study · long-term course · mental illness · psychopathology · outcome · prediction · chronicity

M.M. Fichter, MD, Dipl-Psych (✉) · N. Quadflieg, Dipl-Psych
G. Kohlboeck, PhD
Department of Psychiatry and Psychotherapy
University of Munich (LMU)
Nussbaumstr. 7
80336 Munich, Germany
Tel.: +49-8051/683-510
Fax: +49-8051/683-532
E-Mail: MFichter@Schoen-Kliniken.de

M.M. Fichter, MD, Dipl-Psych · T. Niedermeier, PhD
Klinik Roseneck
83209 Prien, Germany

H. Dilling, MD
Department of Psychiatry and Psychotherapy
University of Luebeck
23538 Luebeck, Germany

## Introduction

After World War II a number of psychiatric epidemiological studies in the general population in Europe and North America were carried out with the aim to identify and tabulate cases of mental illness in geographically defined populations (e.g. Lundby Study [17]; Stirling County Study [40]; Midtown Manhattan Study [55]). In and after this period, enormous strides have been made towards the standardization of diagnostic methods and towards providing a systematic nosology as exemplified in the contemporary revisions of the international classification of diseases (ICD [62]) and the Diagnostic and statistical manual of mental disorders (DSM [2]).

Population-based representative data on the prevalence and distribution of somatic and mental disorders, associated impairments, disabilities and handicaps and the identification of met and unmet needs for care are of central importance for healthcare policymakers and providers [31]. Studies on unselected representative cohorts mirror the mental health in a population. Risk factors associated with outcome can be studied, and several outcomes of one determinant can be assessed [47].

Most epidemiological studies have been cross-sectional and have focused on the point prevalence and/or the retrospectively assessed life time prevalence in general populations. There are only few prospective longitudinal epidemiological studies based

on community samples that have assessed the same individuals at two or more time points over decades. Such studies were the Midtown Manhattan Restudy [54], the Zurich Study [3], the Lundby Study [26], the Stirling County Study [44] and a Finnish Study by Lehtinen et al. [39]. Substantial and sometimes confusing disparities in the findings of these studies may be attributable to different assessment instruments, diagnostic systems, locations and birth cohorts. The Stirling County Study is one of these long-term longitudinal studies in the community that focused on depressive and anxiety disorders but no data on the course of other mental disorders (e.g. substance use disorders or psychoses) were reported. Little is known on course and outcome of certain aspects associated with mental disorders, such as severity of illness or degree of impairment.

The Upper Bavarian Study (UBS) is a long-term prospective study of a representative sample of 1,668 individuals in rural South Eastern Bavaria, Germany [12, 20, 21]. This sample has been studied over a period of 25 years addressing a wide range of outcomes related to mental health, social functioning and personal well-being.

During the last decades epidemiological research became increasingly more elaborated and committed to stricter scientific rules. First, the assessment procedures became increasingly elaborate and more precise. Second, with growing possibilities of complex computerized analyses it became possible to deal with large epidemiological data sets using sophisticated statistical analyses. Because of the increasing complexity of assessments we felt it useful to describe the assessment procedures used in this study in more detail here in this paper.

The major goals of this paper are (1) to provide background information on the concepts, design, procedures, instruments and samples of this study, and (2) based on this study, to reflect critically some methodological issues associated with such a long-term longitudinal study.

Additional papers will cover the course of depression (Fichter et al. in this issue), anxiety, substance use, and other disturbances in our sample from the general population.

## Methods

### Sample and design

*Wave 1 (t1)*: in 1975, random cluster samples were drawn from the community registers of the towns of Traunstein and Traunreut, and the village of Palling in Upper Bavaria. In Germany every person is obliged to be registered at his or her place of residence. The clusters represented the three main sectors in rural Bavaria: agriculture (village of Palling), service and administration (town of Traunstein with a high proportion of civil servants and administrative employees), and industry (town of Traunreut). Simple random samples were drawn from within these clusters. Initially, this resulted in a representative random sample of 1,668 adult community residents aged 15 and above. In the first assessment period or "wave" of the study (1975–1979 [11]), 1,536 of these 1,668 persons of the original sample were interviewed ("*representative 1970s sample*"). Only 132 individuals (6.8%) could not be reached or refused to participate.

*Wave 2 (t2)*: five years later (1980–1984), 1,342 individuals (83%) from the original sample of wave 1 participated in interviews [20, 21]. In addition, 44 individuals who had refused participation in wave 1 agreed to participate in wave 2. Eighty people (4.8%) had died in the interval, and 202 persons (12% of the original sample of 1,668) refused to participate or could not be located. Thus, 1,386 persons from the original sample of N = 1,668 were assessed at wave 2 ("longitudinal 1980s sample").

In addition to this longitudinal sample a new representative sample ("representative 1980s sample") was formed at the second wave assessment (t2). Study participants, who had moved away from the area, were replaced by an additional random sample, which in the meantime had moved into the area. This new representative sample included participants from the Upper Bavarian Children's Sample aged 15 years and older at t2 who had not been included in the adult sample of the 1970s. In the second wave, 1,666 of the representative sample of 1,979 persons (84.2%) were interviewed.

*Wave 3*: this assessment occurred 25 years after wave 1 (2001–2004). At that time 390 individuals (29.1%) of the longitudinal 1980s sample had died. Of the 952 persons still alive at the 25-year follow-up, 838 (88%) were interviewed. Sixty-one individuals (6.4%) refused participation, and 53 (5.5%) persons could not be traced or contacted. Of the 61 individuals who refused participation at wave 3, 22 (36.0%) gave no reason for refusal, 10 (16.4%) could be located, but did not respond to calls, letters, or a scheduled interview for at least three times, in 7 (11.5%) cases, the person's relative or friend refused to establish the contact, 9 (14.8%) reported having had negative experiences with previous surveys or were worried about the confidentiality of the study, 4 (6.6%) were extremely busy and no interview could be scheduled, 4 (6.6%) reported illness or handicaps, and 5 (8.2%) gave other reasons for refusal. Thus, 838 subjects [373 men (44.5%), 465 women (55.5%)] could be interviewed at all three time points ("*three wave longitudinal sample*").

Data were gathered prospectively during three time periods ("waves"). The first wave assessments were made at the original survey from 1975 to 1979, the second wave assessments from 1980 to 1984 and wave 3 assessments from 2001 to 2004.[1]

The answer to the question which of the described samples proves to be the most useful and meaningful one depends on the research question. One straightforward approach is the presentation of data based on the identical assessment instruments in individuals assessed at all three time periods. This approach permits detection of whether or not certain symptoms, syndromes or diagnoses have changed in their frequency of occurrence. Another way to analyze these data is to consider the expression of a certain variable (such as "depressive thoughts") at the first wave and to calculate its impact on health status 25 years later. This approach permits detection of variables which predict outcome at a much later time. To account for missing data a useful approach can be to combine results from waves 1 and 2, which occurred only 5 years apart, to predict outcome 20–25 years later. This approach was used for some analyses. For other analyses we compared the intensity or severity of the same variable across the three waves.

### Instruments

Having to select appropriate assessment instruments for a new wave in a very long-term longitudinal study leads to a research dilemma. To enable direct comparison of symptom severity over time, it is desirable to use the same instruments as in earlier waves.

---

[1] An assessment in the 1990s was approved by the reviewers but did not receive financing because of the exorbitant costs for the unification of East and West Germany.

However, definitions of mental disorders may have been revised and modified in the meantime to reflect new knowledge and interest, which was actually the case in our study. Outdated instruments cannot reflect new concepts, which are of interest to today's researchers. As an example, let us consider the semi-structured "Standardized Psychiatric Interview", SPI, described below [24]. Developed in the late 1960s, it was used in our first wave 10 years later. This structured interview was among the first to be designed specifically for community studies. A number of other standardized psychiatric interviews were developed in the 1970s, but they focused more on the reliable and valid assessment of psychotic symptoms or were designed for administration to psychiatric in- or outpatients (Inpatient multidimensional psychiatric scale, IMPS [41]), the Mental status schedule [52], and the Present state examination [35, 59]. The National Institute of Mental Health Diagnostic Interview Schedule (DIS) was developed in the late 1970s and a German translation of the DIS was therefore not available at the beginning of our study. To ensure continuity we again included the SPI-interview at t2 and t3 to allow for direct comparison over time on the basis of the same instrument. To account for up-to-date diagnostic procedures, we also included the CIDI (see below) at wave 3. The CIDI is currently the generally accepted standard for assessing mental illness in community studies.

Table 1 gives an overview of assessments used in the three waves of our study.

## Expert ratings

The "Standardized psychiatric interview" (SPI) was developed by Goldberg et al. [24]. It was translated into German by a bilingual researcher (Brian Cooper, unpublished), for the use in several larger epidemiological studies within the "Sonderforschungsbereich (Special Research Division) 116 Psychiatric Epidemiology at the University of Heidelberg". This interview should not be confused with the "General health questionnaire" (GHQ) by the same author David Goldberg.

Using the SPI-interview, information on specific symptoms and signs of illness is assessed and rated on operationalized scales for individuals from the community. The first part of the SPI documents an individual's subjective rating in ten areas ("somatic symptoms", "fatigue", "sleep disturbance", "irritability", "lack of concentration", "depression", "anxiety and worry", "phobias", "obsessions and compulsions", "depersonalization"). In addition, the interviewer explores the frequency, duration, and severity of

**Table 1** Assessments used in the three waves of the Upper Bavarian Study

|  | First wave 1975–1979 | Second wave (5-year follow-up) 1980–1984 | Third wave (25-year follow-up) 2001–2004 |
|---|---|---|---|
| *1. Psychopathology* | | | |
| 1.1 Standardized psychiatric interview (SPI [23]) | + | + | + |
| 1.2 Composite international diagnostic interview (CIDI [47, 61]) | − | − | + |
| 1.3 Structured clinical interview for DSM-IV (SCID [52]) | − | − | +[a] |
| 1.4 Complaint list [62] | + | + | + |
| 1.5 PERI demoralization scale [13] | − | + | + |
| 1.6 Munich alcoholism test (MALT [17]) | − | + | + |
| 1.7 Psychiatric diagnoses given on the basis of the classification system valid at each wave (ICD-8; ICDI-9; ICD-10) on a 5-point scale of severity (0 = none to 4 = very severe) | + | + | + |
| 1.8 Impairment associated with psychiatric illness [31] | − | + | + |
| *2. Somatic illness* | | | |
| 2.1 Documentation of somatic illnesses based on patients' self report according to the valid ICD diagnostic system for each wave (ICD-8; ICD-9; ICD-10) | + | + | + |
| 2.2 Munich Blood Pressure Program [55] | − | + | − |
| *3. Intermediate course of illness* | − | + | + |
| Somatic and psychiatric illness (worst status), and utilization of health services estimated for the interim 5-year interval (t2) and 20-year interval (t3) | n.a. | + | + |
| *4. Adverse events and coping* | | | |
| 4.1 PERI-life-event-scale [12] | − | +[b] | +[c] |
| 4.2 Bedford College life-event and difficulty schedule (LEDS [5]) | − | +[b] | +[c] |
| 4.3 Coping questionnaire [19] | − | + | − |
| 4.3 Social interview schedule (SIS [6]) | − | + | (+)[c] |
| 4.5 Questionnaire on social support (F-SOZU [22]) | − | − | + |
| *5.Use of medical and psychosocial services* | + | + | + |
| *6. Disability* | | | |
| 6.1 World Health Organization Disability Assessment Schedule II (WHODAS II [59]) | − | − | + |
| 6.2 Consortium to establish a registry for Alzheimer's disease assessment battery (CERAD [56]) | − | − | + |
| 6.3 Mini mental status examination (MMSE [21]) | − | − | + |
| 6.4 DSM-IV Axis V: global assessment of functioning [2] | − | − | + |
| *7. Satisfaction with Life* | | | |
| Satisfaction with life scale [10] | − | − | + |

+ Instrument was administered

− Instrument was not administered

n.a. Not applicable

[a] Assessment of psychotic symptoms only

[b] Only a sub-sample was assessed (not used in Palling)

[c] A modified version was administered

each symptom. Additional expert-ratings are performed for abnormalities of behaviors such as being "slow, lacking spontaneity", "being suspicious, defensive", "histrionic", and for abnormal mood ("depressed", "anxious, tense", "elated, euphoric", "flattened, incongruous"), and for perceptual and cognitive abnormalities ("excessive concern with bodily functions", "depressive thought content", "thought disorder, delusions, misinterpretations", "hallucinations", "intellectual impairment"). Each symptom is rated on a 5-point severity scale (0 = no impairment; 1 = slight impairment; no medical intervention required; 2 = marked impairment usually requiring general medical treatment; 3 = very distinct impairment, usually requiring outpatient psychiatric treatment; 4 = severe impairment, generally requiring inpatient psychiatric treatment). The 22 ratings for each individual are added and weighted to yield an overall index of general psychopathology. Following instructions of Goldberg et al. [24], the interviewer's expert-rating is weighted by 2 and the subjective rating by 1. Data based on objective symptoms and subjective symptom scores were also summarized in two separate scales. At both follow-ups (waves 2 and 3), the most severe symptom expression during the intervals (wave 2: 5 years; wave 3: 20 years) was assessed in addition to the current (last seven days) status, which also had been assessed at wave 1.

In order to establish interrater and intrarater reliability, we used videotapes of the SPI-interview with study participants to establish interrater and intrarater reliability. At the second cross-section the interrater reliability of the SPI was 0.86 (spearman rank correlations; range 0.74–0.96), and the intrarater reliability was 0.92 (range 0.83–0.96). Similar tapes were used to train the interviewers at the third wave 20 years later. The instructions on how to rate the SPI-psychiatric symptoms were identical at all three waves, and we made all conceivable effort to ensure that the assessments were conducted according to the same criteria at all three time periods. Training of the third wave interviewers was conducted by two psychiatrists who had been trained interviewers at wave 2. Each wave 3 interviewer assessed a total of five cases on video tapes after receiving detailed instructions on how to rate each SPI item. Ratings were compared to the consensus rating of the two interview trainers. The causes for discordant ratings were discussed thoroughly and informed an additional round of SPI-training. The new interviewers then taped ten new cases. The interrater reliability (weighted Kappa [7]) between the interview trainers and the eight clinical psychologists conducting the wave 3 interviews ranged from Kappa = 0.81 to Kappa = 0.85, with an overall mean Kappa of 0.83 (SD = 0.01).

The computer-assisted German version of the composite international diagnostic interview (M-CIDI [48, 60]) was applied after the SPI. The CIDI is a fully structured interview that yields diagnoses according to the tenth revision of the ICD-10 [61] and the fourth revision of the Diagnostic and statistical manual of mental disorders (DSM-IV-TR [2]). Disorders assessed were: mood disorders, psychotic symptoms (screening followed by the relevant sections of the SCID-I; see below), anxiety disorders; substance related disorders (omitting the introductory item on the willingness to answer frankly on this topic), and somatoform and eating disorders. Diagnostic information covered a 12-month time frame and age at first onset, course, duration and persistence of problems. If indications of psychotic symptoms were found, study participants also answered all questions from the psychosis section of the Structured clinical interview for DSM-IV (SCID-I [53]). In cases where the CIDI did not provide satisfying diagnostic results, additional diagnoses (e.g. depression not otherwise specified) were given after an intensive discussion in an expert consensus conference.

Alcohol consumption and criteria of alcohol abuse dependence were assessed using the Munich alcoholism test (MALT [18]). The MALT is a screening test of alcohol use disorders and consists of a self-rating questionnaire (MALT-S 26 items) and an expert-interviewer rating scale (MALT-F 7 items). The MALT-F comprises questions on the presence of hazardous alcohol intake (quantity and frequency of alcohol use, 2 items), alcohol related medical conditions (alcohol-induced hepatopathy, alcoholic polyneuropathy, delirium tremens, three items), foetor alcoholicus at the time of the interview, and friends or family seeking help for the person's alcoholism (one item). The self-rating questionnaire assessed diagnostic criteria of drinking behavior, alcohol-related psychological and social impairment and physiological symptoms.

At the first wave interviewers diagnosed psychiatric symptoms according to the eighth revision of the ICD [8]. At the second wave, psychiatric disorders were classified according to ICD-9 and DSM-III [1], to permit comparisons with more recent community studies. During the first wave, the interviewer (M.D. in psychiatric residence) also performed a severity rating on a five point scale ranging from: 0 = no diagnosis to 4 = severe symptoms. To obtain the same diagnostic classification, the ICD-9 diagnoses from the second wave were re-classified into ICD-8 categories, yielding eight different categories of mental disorders: (1) diagnoses of senile dementia, (2) other organic psychoses, (3) schizophrenia, (4) affective psychoses, (5) neurotic and psychosomatic disorders, (6) personality disorders, (7) alcohol or drug addiction, and (8) oligophrenia or mental retardation. Based on the information obtained in the personal interview, psychiatric diagnoses were given on the basis of the classification system valid at each wave. For the first two waves the number of psychiatric diagnoses was limited to two diagnoses for each individual. No limitation on the number of psychiatric diagnoses was imposed during the third wave. Interviewers then rated the severity of the CIDI-diagnoses and of the additional diagnoses approved in the expert consensus conference. Severity ratings for diagnoses targeted the 7 days preceding the interviews and the five-year and 20-year intervals (worst status) between waves. Severity was rated from 0 to 4 (0 = no symptoms, 1 = slight disturbance, usually no medical intervention required; 2 = marked disturbance and intervention by general practitioner considered advisable; 3 = severe disturbance, usually requiring outpatient psychiatric treatment, and 4 = very severe disturbance, usually requiring inpatient or intensive outpatient psychiatric treatment). Somatic illnesses and complaints were assessed in the interview using a check list and categorized according to ICD-8 (wave 1), ICD-9 (wave 2) or ICD-10 (wave 3) based on the participants' reports.

Additional questions targeted the use of medical and psychosocial services. The items addressed whether respondents sought help from a professional (specialist, general practitioner, hospital) for specific somatic and psychiatric disorders (psychiatrist, psychologist, psychotherapist, psycho-social services, self-help groups). The questions also covered the reasons for using health services, the duration of treatment, the type of provider and the clinical setting.

At wave 1, poverty and poor housing standards were assessed by questions concerning the characteristics of the person's housing situation, such as the ownership of the house or flat, the number of bedrooms, and the subjective satisfaction with the housing. Poverty and poor housing conditions were rated on a 5-point scale from 0 = no problems to 4 = very severe problems.

At wave 2 individuals from the towns of Traunstein and Traunreut [N = 1,070 (80% of the longitudinal 1980s sample)] were assessed using the Psychiatric Epidemiology Research Interview (PERI) Life Events Scale and PERI Chronic Difficulty Scale [13].[2] Acute events were labeled "events" (duration less than one month) and chronically stressful conditions "difficulties" (duration longer than one month). The study participants were asked about all life events and difficulties they had experienced in the 5-year interval between wave 1 and wave 2. Life events fell into ten categories: (1) schooling, 2) work matters, (3) relationship problems, (4) children, (5) family and relatives (death of the partner, of friends or important relatives), (6) housing situation (building a house, relocation), (7) legal or criminal matters, (8) financial situation, (9) medical illness/injury, and (10) other life events. Difficulties were grouped into five categories: health, exceptional events and crises, work, housing, marital issues, and family and social relationships. The interviewers rated events and difficulties as "1", very severe threat/difficulty to "4", no threat/difficulty based on the definitions

---

[2]For the village of Palling, the instrument was not yet available in German.

for the severity of life events and chronic difficulties as described in detail by Brown and Harris [5]. Thus, we used the list of life events and chronic difficulties from the Dohrenwend et al. instrument and performed the severity ratings on the basis of the thorough and detailed definitions of Brown and Harris.[3]

A sub-sample of 248 women between the ages of 20 and 65 was interviewed by an independent interviewer using the *Bedford College Life Event and Difficulty Schedule (LEDS* [5]). The LEDS is an extensive semi-structured interview designed to assess a wide variety of potential stressors. The women reported all events that had occurred in the 5 year interval preceding wave 2 by using a 110-event self-report checklist. Then each woman underwent a semi-structured interview in which the endorsed events were explored with specific questions tailored to the particular event. The information from the interview was presented to a panel of raters trained in the LEDS procedures for defining life events and rating dimensions of these stressors (German version [34]). A 4-point scale captured short- and long-term threats for each event, with a rating of "1" corresponding to the most severe threat and "4" for no severe threat.

The *Social interview schedule* (SIS [6]) employed at wave 2 assessed social maladjustment and dysfunction in the areas of housing, finances, occupation, social and leisure activities, and relationships. Answers on a 4-point scale in eight more detailed areas combined into three conceptual categories (objective material conditions, social management, and subjective satisfaction). Four overall SIS measures were defined based on the maximum severity (=worst expression) of a number of items as follows: the domain "Income and housing conditions" included items on satisfaction with housing, objective housing conditions, satisfaction with housework conditions, household income, management of income, satisfaction with income, and housework conditions. The domain "Work conditions or satisfaction with being a housewife, retired or jobless" consisted of SIS scales concerning work/study conditions, management of work/studies, satisfaction with work/studies, and satisfaction with other social roles (includes being unemployed, disabled, retired). The domain "social contacts" included SIS scales, opportunities for interaction with colleagues, quality of personal interaction with colleagues, satisfaction with personal interaction with colleagues, extent of social activities, satisfaction with social activities, opportunities for interaction with relatives, quality of interaction with relatives, satisfaction with interaction with relatives, and quality of domestic interaction. The domain "leisure time" consisted of opportunities for leisure and social activities, extent of leisure activities, and satisfaction with leisure activities. At third wave, SIS objective conditions of work and housework, income, opportunities for leisure and social activities were assessed again.

Assessments of family history at wave 2 included psychiatric illness (suicide, alcohol and drug consumption, and mental illness), psychiatric treatments, as well as seizures, hypertension and diabetes mellitus in family members. To study the prevalence of high blood pressure, we used the same method as the Munich Blood Pressure Program [56]. Physicians measured blood pressure three times during the wave 3 interview using a random-zero sphygmomanometer [19].

At the second and third waves, standards established by Jakubaschk et al. [32] guided the assessments of stress and impairment related to psychiatric disorders. Questions rated on a 5-point scale (0 = no to 4 = very severe strain/impairment) included psychological strain, impairment of social relations, ability to work, and disturbance of leisure activities.

At the third wave, special efforts were directed toward the collection of data on the participants' impairments and disabilities. The German version of the *World Health Organization Disability Assessment Schedule II* [16, 62] measured the degree of disability in six domains: understanding and communication, getting around, self-care, getting along with others, household and work activities, and participation in society. The WHODAS II employs a 5-point

rating scale for all items in which "1" indicates no difficulty and "5" indicates extreme difficulty or inability to perform the activity at all.

For the assessment of cognitive impairment, four subscales of the German version of the "Consortium to establish a registry for Alzheimer's disease assessment battery" (CERAD [57]) were administered: (1) categorical verbal fluency, (2) word list learning, (3) word list recall, (4) word list recognition. Higher scores indicate better performance on all CERAD subtests.

The *Mini Mental Status Examination* (MMSE [22]) measured cognitive deficits with scores ranging from 0 to 30, with higher scores indicating better general cognition. The *Global Assessment of Functioning Scale* (GAF; DSM-IV [2]), an interviewer rating of a person's current overall level of functioning, was also employed. The GAF rated psychological, social, and occupational functioning independently of impairment due to physical or environmental limitations. GAF scores range from 1 (lowest level of functioning) to 100 (superior functioning).

## Self-ratings

The *complaint list ('Beschwerdenliste')* by von Zerssen [63] assessed somatic, psychological, and psychophysiological symptoms on a scale ranging from 0 = no complaint to 3 = strong complaint. A higher total score indicates more or more severe complaints.

The *demoralization* self-rating scale of the Psychiatric Epidemiologic Research Instrument (PERI) by Dohrenwend et al. [14] was employed at wave 2 to assess feelings of helplessness, hopelessness, and self-derogation. Items are scored from 0 to 4; high total scores indicate a higher degree of demoralization. At wave 2, assessments also included a *coping questionnaire* (Fichter [20]—based on a coping questionnaire developed by Junk and Junk [33]).

At wave 3 assessments included a German self-rating questionnaire on *social support (F-SOZU* [23]). The short form of the F-SOZU (14 items) was applied to determine perceived social support, emotional and practical support as well as social integration. Items were rated on a 5-point scale with higher scores indicating more social support. The internal consistency of this instrument has been found to be reasonable, with Cronbach's α ranging between 0.79 and 0.94, and the validity was established in primarily non-clinical populations. The *Satisfaction With Life Scale* (*SWLS* [10]) was used as a measure of life satisfaction that is designed around the idea that one has to ask a person for an overall judgment of their life in order to measure the concept of "life satisfaction". Each person indicated their degree of agreement or disagreement on a 7-point Likert-type scale. The 5 items of the SWLS were selected from a pool of 48 items based on factor analyses. Scores range from 5 to 35, with higher scores indicating greater life satisfaction.

At all three waves, participants were asked about their former and current health status, educational and occupational status, social contacts, familial and personal history, alcoholism, and questions concerning work and leisure time for the 7 days preceding the interviews and for the five and 20 year intervals between waves. Social class was determined according to Kleining and Moore [38]: Classes I and II: upper class, upper and middle middle class (professional occupations, management); Class III: lower middle class (skilled occupations), Class IV: upper lower class (partially skilled occupations), Class V: lower lower class (unskilled occupations).

## ■ Procedure at third wave

All individuals received a letter explaining the aims of the study and inviting them to participate. Several days after sending the letter, each person was contacted by telephone, and after preliminary consent, an appointment was made for the personal interview. Interviews were conducted at the person's home after written informed consent had been obtained. For those unable or unwilling to participate in the long interview, a shorter interview

---

[3] Interviewers were trained for the LEDS interview by George Brown personally.

focusing on central research questions was available. During the visit at the person's home, he or she filled out the self-rating questionnaires. Reasons for not participating in the study were recorded, and individuals who had moved were located through the local community registers. The duration of the wave 3 interviews was between 2 and 3 h (mean: 2.14 h, SD: 49.5). The study was approved by the ethics committee of the medical faculty of the University of Munich (LMU) and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

## ◼ Statistical analysis

Mean with SD and frequencies of categorical variables are presented. Group comparisons were made using one-way analysis of variance (ANOVA) followed by post hoc Scheffé tests for continuous variables and the chi-square test (Fisher exact test in $2 \times 2$ tables with a minimum count of less than 5) for categorical variables. Missing data varied slightly between measures, and therefore numbers of cases are reported for each analysis. Logistic regression analysis was applied to identify predictors of refusal at the 25-year follow-up.

## Results

In the longitudinal sample for all three waves, the time interval from wave 1 to wave 2 was $5.0 \pm 0.3$ years

(5-year follow-up; mean + SD), and between waves 1 and 3 it was $25.4 \pm 0.8$ years (25-year follow-up). Mean age at wave 3 was $61.4 \pm 13.4$ years. At wave three 737 of 838 participants were assessed with the complete interview, 92 with the shortened interview, and 9 with a brief interview with basic questions only. Most participants (85.4%) were interviewed at their home, 5.3% were interviewed at the study's local office in the town of Traunstein, and 4.8% were interviewed by telephone with the questionnaires mailed due to geographical distance or the person's privacy concerns or lack of time. In addition, 4.5% of interviews were conducted either at the person's workplace, or at our research offices in Munich or Prien.

Table 2 gives an overview of the longitudinal sample's characteristics at the baseline assessment at t1.

At the first wave, 1.4% of this sample were 65 years or older, whereas almost half of the participants (47%) were below age 35. Gender distribution differed by age groups, and there were more men in the younger groups (15–44 years) than in the older age groups. Two thirds of the probands were married or co-habitating, and most had relatively few years of school education, about 11 to 12 years. Gender

**Table 2** Sample characteristics at first assessment (first wave)

| | Total | | Men | | Women | | Statistics | | |
|---|---|---|---|---|---|---|---|---|---|
| | First wave 1975–1979 | | $n = 373$ | | $n = 465$ | | $\chi^2$ test | | |
| | N | % | N | % | N | % | $\chi^2$ | df | P |
| Age group ($n = 838$) (years) | | | | | | | 19.9 | 5 | 0.001 |
| 15–24 | 190 | 22.7 | 92 | 48.4 | 98 | 51.6 | | | |
| 25–34 | 203 | 24.2 | 94 | 46.3 | 109 | 53.7 | | | |
| 35–44 | 208 | 24.8 | 107 | 51.4 | 101 | 48.6 | | | |
| 45–54 | 152 | 18.1 | 58 | 38.2 | 94 | 61.8 | | | |
| 55–64 | 73 | 8.7 | 19 | 26.0 | 54 | 74.0 | | | |
| 65–74 | 12 | 1.4 | 3 | 25.0 | 9 | 75.0 | | | |
| Family status ($N = 838$) | | | | | | | 16.1 | 3 | 0.001 |
| 0 Single | 223 | 26.6 | 108 | 48.4 | 115 | 51.6 | | | |
| 1 Married or co-habitating | 554 | 66.1 | 251 | 45.3 | 303 | 54.7 | | | |
| 2 Divorced or separated | 33 | 3.9 | 11 | 33.3 | 22 | 66.7 | | | |
| 3 Widowed | 28 | 3.3 | 3 | 10.7 | 25 | 89.3 | | | |
| Level of education ($N = 838$) | | | | | | | 63.1 | 2 | 0.000 |
| Low ("Grund-, Sonder-, Hauptschule") | 272 | 32.5 | 68 | 25.0 | 204 | 75.0 | | | |
| Middle ("Berufs-, Realschule") | 499 | 59.5 | 265 | 53.1 | 234 | 46.9 | | | |
| High ("Abitur, Studium") | 67 | 8.0 | 40 | 59.7 | 27 | 40.3 | | | |
| Occupational status ($N = 838$) | | | | | | | 232.6 | 8 | 0.000 |
| Entrepreneurs, self-employed | 36 | 4.3 | 24 | 66.7 | 12 | 33.3 | | | |
| Farmers | 14 | 1.7 | 10 | 71.4 | 4 | 28.6 | | | |
| Family workers | 34 | 4.1 | 0 | 0 | 34 | 100.0 | | | |
| Civil servants | 48 | 5.7 | 38 | 79.2 | 10 | 20.8 | | | |
| Employees | 204 | 24.3 | 85 | 41.7 | 119 | 58.3 | | | |
| (Blue Collar) Worker | 210 | 25.1 | 140 | 66.7 | 70 | 33.3 | | | |
| Retired | 29 | 3.5 | 9 | 31.0 | 20 | 69.0 | | | |
| Housewives | 147 | 17.5 | 0 | 0 | 147 | 100.0 | | | |
| In formal education/in military service | 116 | 13.8 | 67 | 57.8 | 49 | 42.2 | | | |
| Social class (Kleining-Moore) ($N = 837$) | | | | | | | – | 4 | ns |
| Upper class | 32 | 3.8 | 15 | 46.9 | 17 | 53.1 | | | |
| Upper middle class | 115 | 13.7 | 54 | 47.0 | 61 | 53.0 | | | |
| Lower middle class | 317 | 37.9 | 134 | 42.3 | 184 | 57.7 | | | |
| Upper working class | 312 | 37.3 | 140 | 44.9 | 172 | 55.1 | | | |
| Lower working class | 61 | 7.3 | 30 | 49.2 | 31 | 50.8 | | | |

distribution differed by family status, and men were more likely to be single, whereas more women were widowed. More women than men (75 vs. 25%) had low levels of education.

Concerning sociodemographic characteristics, few differences emerged between those from the longitudinal 1980s sample ($N = 1,342$) who were interviewed at t3 ($N = 838$) and those who could not be traced or reached ($N = 53$), who refused participation ($N = 61$) or who were deceased ($N = 390$) (see Table 3).

Males and females did not differ in rate of participation in the study nor in the causes for dropping out. Refusers differed from participants in family status and social class. From our data it seems that married men and women were less liable to refuse participation, whereas no clear picture emerged for social class. Individuals who could be traced or reached belonged to the younger age groups, were more often single and in formal education or military service. No differences between these groups emerged for the different categories of mental illness.

ANOVAs specified differences for a large number of severity ratings of mental illness, impairment associated with mental illness, social circumstances, work-related factors, (difficult) childhood experiences, family history of illness, use of medical services, and adverse living conditions assessed at t1 and t2 between participants, refusers, and those who could not be reached at the t3-follow-up. Very few significant differences emerged from these analyses (Table 4). Refusers showed higher "oligophrenia" severity scores and higher SPI Objective Psychopathology Scores than participants, with those not traced or not reached lying between these groups and not differing from either group significantly. Moreover, individuals who could not be traced or reached showed more difficulties regarding the SIS domain social contacts (significance levels not corrected for multiple testing because it would have erased the significance of these effects).

Comparing refusers to participants at t3 and including family status (t1), social class (t1), and SPI objective psychopathology score (worst t1–t2) as predictors logistic regression identified only SPI objective psychopathology score as significant predictor of refusal at t3 (Wald $= 4.77$; $P = 0.029$; odds ratio $= 3.12$ (CI $95\% = 1.12$–$8.65$); $R^2 = 0.023$). In additional analyses age, sex and oligophrenia were also included but had no influence on the prediction.

## Discussion

The *strengths* of our study consist in (1) the prospective longitudinal approach with three cross-sectional assessments, (2) a very long follow-up period of 25 years, (3) a large sample size of 838 community residents with assessments at all three cross-sections, (4) a reasonable diagnostic uniformity,

(5) the use of standardized/structured interviews and self-rating scales, and (6) a participation rate (88% at 25-year follow-up) that is higher than commonly found in follow-ups of this kind (typically: 70–80%; cf. Rothman and Greenland [51]). This is of relevance, because according to some authors non-responders have a strong tendency to distort the results in psychiatric population studies (e. g. Hagnell [27]). Reporting on the 5-year follow-up of our study, Fichter [20] found that individuals who refused to participate in the follow-up tended to have more disorders at the previous assessment. Our long-term results from the same study, however, indicated that this effect is no longer relevant after 20–25 years.

There are some *limitations* of our study: (1) due to the long interval of 20 years between the second and third assessments, we probably missed some relevant information. Thus, no further information (causes of death) is available on the 390 individuals who had died between the original survey and the third investigation. (2) As our sample has been followed over 25 years, our study participants were growing older and were no longer representative for the general population of Upper Bavaria at t3. A cohort which was originally selected to be representative of a population may develop into another direction than the population at large with respect to socio-economic stratum or other demographic characteristics [47]. (3) The sample investigated from 2000 to 2004 was an aging cohort. An under-estimation of prevalence of psychiatric disorders may derive from the fact that data could be gathered in surviving individuals only. Some studies found evidence for an elevated mortality risk in depressive subjects [43, 50, 58], and in our study some cases may have been lost to follow-up because they died before the assessment in the first decade of the twenty-first century. This important issue will be addressed in an additional paper. Finally, our original sample was drawn from a rural population. It has been shown that the prevalence of mental disorders is somewhat lower in rural areas [21]. Therefore, there might be limitations for generalizing our findings to metropolitan areas.

Individuals who were younger at the time of the second wave changed their residence more often than older individuals, making it harder to locate or reach the younger ones. In the 15-year follow-up of the epidemiologic catchment area (ECA) survey, Badawi et al. [4] came to the conclusion that refusal to participate is a less important source of bias than failure to locate individuals. The time between 1980 and 2000 in Germany was marked by a prospering economy and steadily increasing individual mobility. Individuals who were at the beginning of their productive life in 1980 certainly were those benefiting most from these changes. Although we cannot know for certain if drop-outs from our study occurred due to individual impairment, it seems more plausible to interpret the drop-outs as signs of health, mobility and produc-

**Table 3** Differences in sociodemographics and psychopathology at t1 and t2 in individuals interviewed and those not traced (B vs. C) and subjects who refused participation (B vs. D)

| | A N = 1,342 total sample t3 | B N = 838 Interviewed t3 | | C N = 53 Not traced, not reached t3 | | D N = 61 Refused t3 | | E N = 390 Deceased t3 | | Interviewed (B) vs. not traced (C) | | | Interviewed (B) vs. refused participation (B vs. D) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | N | % | N | % | N | % | N | % | χ²/Fisher exact test | P | df | χ²/Fisher exact test | P | df |
| Gender (N = 1,342) | | | | | | | | | | − | ns | 1 | − | ns | 1 |
| Male | 610 | 373 | 61.1 | 23 | 3.8 | 28 | 4.6 | 186 | 30.5 | | | | | | |
| Female | 732 | 465 | 63.5 | 30 | 4.1 | 33 | 4.5 | 204 | 27.9 | | | | | | |
| Age group t1 (N = 1,342) | | | | | | | | | | 34.9 | 0.000 | 5 | − | ns | 5 |
| 15–24 years | 243 | 190 | 78.2 | 31 | 12.8 | 13 | 5.3 | 9 | 3.7 | | | | | | |
| 25–34 years | 234 | 203 | 86.8 | 7 | 3.0 | 15 | 6.4 | 9 | 3.8 | | | | | | |
| 35–44 years | 255 | 208 | 81.6 | 6 | 2.4 | 14 | 5.5 | 27 | 10.6 | | | | | | |
| 45–54 years | 219 | 152 | 69.4 | 5 | 2.3 | 11 | 5.0 | 51 | 23.3 | | | | | | |
| 55–64 years | 186 | 73 | 39.2 | 3 | 1.6 | 5 | 2.7 | 105 | 56.5 | | | | | | |
| 65–74 years | 170 | 12 | 7.1 | 1 | 0.6 | 3 | 1.8 | 154 | 90.6 | | | | | | |
| 75–94 years | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 35 | 100 | | | | | | |
| Family status t1 (N = 1,342) | | | | | | | | | | 22.8 | 0.000 | 3 | 9.35 | 0.025 | 3 |
| 0 Single | 311 | 223 | 71.7 | 30 | 9.6 | 19 | 6.1 | 39 | 12.5 | | | | | | |
| 1 Married | 847 | 554 | 65.4 | 22 | 2.6 | 32 | 3.8 | 239 | 28.2 | | | | | | |
| 2 Separated/ divorced | 58 | 33 | 56.9 | 1 | 1.7 | 4 | 6.9 | 20 | 34.5 | | | | | | |
| 3 Widowed | 126 | 28 | 22.2 | 0 | 0 | 6 | 4.8 | 92 | 73.0 | | | | | | |
| Occupational status t1 (N = 1,342) | | | | | | | | | | 28.3 | 0.000 | 8 | − | ns | 8 |
| Entrepreneurs, self-employed | 55 | 36 | 65.5 | 0 | 0 | 3 | 5.5 | 16 | 29.1 | | | | | | |
| Farmers | 24 | 14 | 58.3 | 0 | 0 | 1 | 4.2 | 9 | 37.5 | | | | | | |
| Family workers | 45 | 34 | 75.6 | 1 | 2.2 | 1 | 2.2 | 9 | 20.0 | | | | | | |
| Civil servants | 60 | 48 | 80.0 | 2 | 3.3 | 0 | 0 | 10 | 16.7 | | | | | | |
| Employees | 271 | 204 | 75.3 | 10 | 3.7 | 16 | 5.9 | 41 | 15.1 | | | | | | |
| (Blue Collar) Worker | 296 | 210 | 70.9 | 14 | 4.7 | 20 | 6.8 | 52 | 17.6 | | | | | | |
| Retired | 233 | 29 | 12.4 | 3 | 1.3 | 2 | 0.9 | 199 | 85.4 | | | | | | |
| Housewives | 212 | 147 | 69.3 | 3 | 1.4 | 10 | 4.7 | 52 | 24.5 | | | | | | |
| Education/military service | 146 | 116 | 79.5 | 20 | 13.7 | 8 | 5.5 | 2 | 1.4 | | | | | | |
| Social Class (Kleining–Moore) t1 (N = 1,341) | | | | | | | | | | − | ns | 4 | 10.6 | 0.031 | 4 |
| Upper Class | 45 | 32 | 71.1 | 1 | 2.2 | 1 | 2.2 | 11 | 24.4 | | | | | | |
| Upper middle class | 179 | 115 | 64.2 | 8 | 4.5 | 6 | 3.4 | 50 | 27.9 | | | | | | |
| Lower middle class | 520 | 317 | 61.0 | 18 | 3.5 | 25 | 4.8 | 160 | 30.8 | | | | | | |
| Upper working class | 477 | 312 | 65.4 | 18 | 3.8 | 18 | 3.8 | 129 | 27.0 | | | | | | |
| Lower working class | 120 | 61 | 50.8 | 8 | 6.7 | 11 | 9.2 | 40 | 33.3 | | | | | | |

Note: Percentages are row percent

ns = Not significant

**Table 4** Comparison of data of subjects interviewed at t3 as compared to those with missing data at t3 (not traced or reached, refused) based on data of these subjects from earlier waves of assessment (t1 or t2)

| | Interviewed t3 N = 838 | | Not traced, not reached t3 N = 53 | | Refused t3 N = 61 | | ANOVA df = 2 | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | F | P |
| **1. Psychiatric disorders and psychopathology t1 or t2** | | | | | | | | |
| Severity[1] of "senile dementia or other organic psychiatric disorders" | 0.03 | 0.29 | 0.02 | 0.14 | 0.11 | 0.58 | na | ns |
| Severity of "schizophrenia" | 0.02 | 0.24 | 0 | 0 | 0 | 0 | na | ns |
| Severity of "affective psychoses" | 0.04 | 0.33 | 0.02 | 0.14 | 0.03 | 0.26 | na | ns |
| Severity of "neurotic-psychosomatic disorders" | 0.87 | 1.07 | 0.75 | 1.07 | 0.89 | 1.16 | na | ns |
| Severity of "personality disorders" | 0.14 | 0.45 | 0.13 | 0.48 | 0.21 | 0.49 | na | ns |
| Severity of "alcoholism" | 0.16 | 0.54 | 0.11 | 0.42 | 0.13 | 0.50 | na | ns |
| Severity of "oligophrenia" | 0.04[a] | 0.30 | 0.08[a, b] | 0.55 | 0.16[b] | 0.64 | 3.77 | 0.023 |
| SPI[2] subjective psychopathology score | 0.42 | 0.37 | 0.43 | 0.38 | 0.45 | 0.38 | na | ns |
| SPI objective psychopathology score | 0.21[a] | 0.21 | 0.22[a, b] | 0.22 | 0.29[b] | 0.27 | 4.21 | 0.015 |
| SPI total psychopathology score | 0.81 | 0.71 | 0.84 | 0.75 | 0.96 | 0.76 | na | ns |
| Total score complaint list | 13.73 | 10.09 | 14.87 | 1.41 | 14.71 | 9.93 | na | ns |
| Subjective health (0 = very good- 4 = very poor[3]) | 1.41 | 0.80 | 1.34 | 0.73 | 1.43 | 0.85 | na | ns |
| Number of suicide attempts lifetime (retrospectively up to t2) | 0.04 | 0.47 | 0.10 | 0.30 | 0.03 | 0.18 | na | ns |
| Total score PERI[4] Demoralization scale t2 | 17.41 | 3.63 | 20.52 | 17.42 | 18.10 | 15.98 | na | ns |
| **2. Severity of psychiatric illness** | | | | | | | | |
| Severity of psychiatric illness[5] last 7 days t1 or t2 | 0.82 | 0.91 | 0.72 | 0.97 | 1.00 | 1.03 | na | ns |
| Severity of psychiatric illness last 12 months t1 or t2 | 0.99 | 1.04 | 0.77 | 1.05 | 1.16 | 1.11 | na | ns |
| Severity of psychiatric illness 5 years t1-t2[6] | 1.09 | 1.12 | 0.92 | 1.14 | 1.25 | 1.16 | na | ns |
| **3. Alcoholism** | | | | | | | | |
| MALT-F[7] sum score of objective alcoholism criteria (1–7) t2 | 0.22 | 1.30 | 0.08 | 0.56 | 0.07 | 0.52 | na | ns |
| MALT-S sum score of subjective Items | 0.25 | 1.28 | 0.49 | 1.88 | 0.15 | 0.82 | na | ns |
| MALT-F and MALT-S sum score | 0.44 | 2.24 | 0.57 | 2.10 | 0.22 | 1.32 | na | ns |
| MALT average alcohol consumption per day in ml t2 | 18.52 | 28.39 | 10.20 | 21.50 | 19.41 | 24.80 | na | ns |
| **4. Social factors** | | | | | | | | |
| Poor housing situation (anamnesis) (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = very severe) t1 | 0.09 | 0.41 | 0.17 | 0.58 | 0.08 | 0.33 | na | ns |
| Poverty, objective (anamnesis) (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = very severe) t1 | 0.14 | 0.53 | 0.09 | 0.40 | 0.23 | 0.64 | na | ns |
| Total score of social strain (anamnesis) t1 | 1.03 | 2.27 | 1.21 | 2.37 | 1.48 | 3.04 | na | ns |
| Prestigescore Treiman t2 | 43.92 | 10.77 | 42.65 | 10.33 | 41.77 | 9.44 | na | ns |
| **5. Work related factor** | | | | | | | | |
| Total score of job strain t1 or t2 | 9.53 | 5.81 | 11.03 | 6.34 | 9.76 | 6.58 | na | ns |
| Limitation in working ability by somatic symptoms last 7 days (0 = no limitation to 4 very severe limitation) t2, | 0.31 | 0.81 | 0.19 | 0.56 | 0.51 | 0.98 | na | ns |
| **6. Childhood factors (retrospective)** | | | | | | | | |
| Conduct problems in childhood and adolescence, objective (anamnesis) (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = very severe) t1 | 0.47 | 1.15 | 0.34 | 1.04 | 0.44 | 1.12 | na | ns |
| Illness related strain in childhood and adolescence, objective (anamnesis) (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = very severe) t1 | 1.58 | 2.70 | 1.19 | 3.06 | 1.54 | 2.77 | na | ns |

**Table 4** Continued

| | Interviewed t3 N = 838 | | Not traced, not reached t3 N = 53 | | Refused t3 N = 61 | | ANOVA df = 2 | |
|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | F | P |
| Pathogenic events in childhood and adolescence (war, lack of money, relocation) objective (anamnesis) (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = very severe) t1 | 0.64 | 1.09 | 0.32 | 0.85 | 0.72 | 1.23 | na | ns |
| Relationship with mother in childhood (0 = very good, 1 = good, 2 = moderate, 3 = poor, 4 = very poor) t2 | 0.86 | 0.66 | 0.72 | 0.70 | 1.01 | 0.72 | na | ns |
| Relationship with father in childhood (0 = very good, 1 = good, 2 = moderate, 3 = poor, 4 = very poor) t2 | 0.95 | 0.76 | 1.10 | 1.04 | 1.05 | 0.70 | na | ns |
| Quality of relationship of parents in childhood (0 = very good, 1 = good, 2 = moderate, 3 = poor, 4 = very poor) t2, | 1.15 | 0.83 | 1.16 | 0.91 | 1.27 | 0.84 | na | ns |
| Financial situation in childhood (0 = very good, 1 = good, 2 = moderate, 3 = poor, 4 = very poor) t2, | 2.05 | 0.80 | 1.98 | 0.73 | 2.05 | 0.78 | na | ns |
| 7. Family history | | | | | | | | |
| Illness related strain and other problems in the family objective (anamnesis) (0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = very severe) t1 | 0.24 | 0.86 | 0.36 | 1.09 | 0.18 | 0.81 | na | ns |
| 8. Use of medical services | | | | | | | | |
| Number of outpatient visits last 12 months t1 | 5.36 | 10.46 | 3.60 | 4.42 | 4.18 | 5.35 | na | ns |
| Number of visits of family doctors/general practitioners last 5 years t2 | 4.83 | 8.77 | 3.95 | 4.64 | 3.54 | 4.37 | na | ns |
| Number of visits of medical specialists last 5 years t2 | 6.33 | 10.43 | 6.82 | 7.37 | 7.60 | 14.26 | na | ns |
| 9. Adverse living conditions and coping | | | | | | | | |
| Total score coping (10 Items from 0 = none to 4 severe coping problem) t2 | 10.78 | 5.93 | 10.90 | 1.02 | 11.98 | 6.51 | na | ns |
| SIS[8] domain income and housing conditions (severity 0 = none to 3 = severe difficulties or dissatisfaction) t2 | 1.51 | 0.77 | 1.56 | 0.77 | 1.53 | 0.70 | na | ns |
| SIS domain work conditions or satisfaction with being housewife, retired, or jobless (severity 0 = none to 3 = severe difficulties or dissatisfaction) t2 | 1.19 | 0.63 | 1.16 | 0.60 | 1.17 | 0.63 | na | ns |
| SIS domain Social contacts (severity 0 = none to 3 = severe difficulties or dissatisfaction) t2 | 1.44[a] | 0.67 | 1.79[b] | 0.71 | 1.56[a, b] | 0.79 | 6.63 | 0.001 |
| SIS domain leisure time (severity 0 = none to 3 = severe difficulties or dissatisfaction) t2 | 1.23 | 0.70 | 1.33 | 0.78 | 1.32 | 0.71 | na | ns |

ns = Not significant; na not applicable
t1 = first wave (1975–1979), t2 = second wave (1980–1984)
a, b, c = Different letters indicate a significant difference between groups (Scheffé test; 5%)
[1]Maximum severity (0–4) of ICD-8 diagnoses at t1 (last 7 days and last 12 months) or t2 (last 7 days and last 5 years)
[2]Maximum severity (0–4) of standardized psychiatric interview psychopathology scales at t1 or t2 (last 7 days before interview)
[3]Subjective health assessed at t1 or t2 (Maximum = worst expression)
[4]PERI = the psychiatric epidemiologic research instrument
[5]Maximum of severity (0–4) of the two psychiatric diagnoses assessed at t1 or t2
[6]Maximum of severity (0–4) of the two psychiatric diagnoses of t1 (last 12 months), t2 (last 12 months) and 5 year interval between t1 and t2
[7]MALT = Munich alcoholism test; F = Expert-rating; S = Self-rating
[8]SIS = Social interview schedule

tivity. Our main efforts to reassess individuals of the study focused on the county of Traunstein in Upper Bavaria. Individuals who meanwhile had moved further away in Germany could be located. Those who moved abroad naturally were very difficult to locate.

Although a wide range of psychiatric disorders and psychopathological measures and severity scores was examined, we found no differences between participants and non-participants of the 25-year follow-up. Other studies are inconclusive on this issue. The National Comorbidity Survey [36] provided evidence that people who were missing initially tended to be more ill. In the Stirling County Study [28], failure to fill out self-reports was predicted by the treating physicians' reports of psychiatric disorders. According to these reports, the individuals who could not be interviewed were more likely to suffer from a psychiatric disorder than those who were interviewed successfully. Results of the Zurich Study [15] showed that drop-out had become more likely in persons with extreme (high or low) Symptom Checklist 90-R scores.

On the other hand, our results are consistent with recent findings from long-term studies of the general population who failed to observe any associations between psychopathology and refusals. In the Epidemiologic Catchment Area Study [4], psychopathology had no influence on refusal once adjusted for individual variables (age, gender, race, and education). Specific mental disorders, suicide behavior, comorbidity, and drug use did not influence refusal. Similarly, in the NEMESIS study [25], none of the psychiatric disorders were positively associated with refusal after adjustment for sociodemographics. Results from the US National Comorbidity Survey Replication [37] also confirmed that psychopathology of non-respondents did not differ from that of respondents when comparing diagnostic stem question predictors (mood disturbances, anxiety, substance abuse problems, and problems with impulse control). The same results were confirmed by the Lundby Study [47] where drop-out was not associated with rate of mental disorder.

Our analyses showed that the younger study participants below age 35 were harder to trace and reach for interview. We made extensive efforts to reach every person. It may be advisable for future epidemiological community studies to plan a systematic over-sampling to enable useful analysis of this group.

The Upper Bavarian Study (1975–2004) is one of the few longitudinal epidemiological studies characterized by repeated assessments of the same unselected representative population. By using the same instrument at all three time points we were able to study the longitudinal course of psychiatric symptoms as well as risk factors associated with poor outcome. In comparing results over time, Murphy et al. [45] argued that there are several difficulties posed by the very fact that time progresses. Diagnostic criteria and the methods of applying them may change, as may the views on which criteria need to be applied [30]. Murphy et al. [42] suggested that in conducting longitudinal research, the passage of time, which permits an examination of historical trends and the prospective study of individuals, is usually accompanied by advances in research technology. The authors argued that the field of psychiatric epidemiology has experienced a thrust forward in terms of the conceptual models, the development of new interviewing schedules for implementing these models, and the design of computer programs for the analysis of such information. While recognizing the importance of the improvements embodied in these newer approaches, continuity of concepts and methods that would allow to conduct a longitudinal study must also be considered.

There is a recent discussion concerning the impairment associated with the prevalence of mental illness. Insel and Fenton [29] criticized the high numbers of study participants with a lifetime history of mental disorder (46.4%) in the National Comorbidity Study [37]. Kessler reported that 40.4% of the cases could be classified as "mild". Moreover, the WHO World Mental Health Survey Consortium [9] commented on the importance of recognizing that many mental disorders are mild and self-limiting, and they argued that this was not a focus of the DIS or CIDI surveys, which had been designed to estimate prevalence rather than severity. The high prevalence estimates in these surveys raised concerns that even the richest of countries could not afford to treat all people with mental disorders. Motivated by this concern, investigators performed secondary analyses of two US surveys [36, 49], which concluded that up to half of the current (12 months) mental disorders were of mild severity [46]. Results from the World Mental Health Survey [9] (The WHO World Mental Health Survey Consortium) indicated that the percentages with a milder disorders varied substantially across countries from 33.1% in Columbia to 80.9% in Nigeria. To cover both prevalence and relevance of mental disorders, severity was rated for all diagnoses in our study.

Whereas this paper focused on the methodology of epidemiologic studies, forthcoming papers (e.g. on the course of depression, Fichter et al., in this issue) report on the results of this unique long-term longitudinal community study in Upper Bavaria.

# References

1. American Psychiatric Association (APA) (1980) Diagnostic and statistical manual of mental disorders (DSM III). American Psychiatric Association, Washington, DC

2. American Psychiatric Association (APA) (2000) Diagnostic and statistical manual of mental disorders, 4th edn (text revision). American Psychiatric Association, Washington, DC

3. Angst J, Dobler-Mikola A, Binder J (1984) The Zurich study—a prospective study of depressive, neurotic and psychosomatic syndromes. I. Problem, methodology. Eur Arch Psychiatr Neurol Sci 234:13–20

4. Badawi MA, Eaton WW, Myllyluoma J, Weimer LG, Gallo J (1999) Psychopathology and attrition in the Baltimore ECA 15-year follow-up 1981–1996. Soc Psychiatry Psychiatr Epidemiol 34:91–98

5. Brown GW, Harris T (1978) Social origins of depression. Psychol Med 8(4):577–588

6. Clare AS, Cairns VE (1978) Design, development and use of standardized interview to assess social maladjustment and dysfunction in community studies. Psychol Med 8:589–604

7. Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46

8. Degkwitz R, Helmchen H, Kockott G, Mombour W (1967) ICD-8. Internationale Klassifikation Psychiatrischer Krankheiten 8. Revision. Eighth Revision of the International Classification of Diseases, Section V; World Health Organization

9. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, Angermeyer MC, Bernert S, de Girolamo G, Morosini P, Polidori G, Kikkawa T, Kawakami N, Ono Y, Takeshima T, Uda H, Karam EG, Fayyad JA, Karam AN, Mneimneh ZN, Medina-Mora ME, Borges G, Lara C, de Graaf R, Ormel J, Gureje O, Shen Y, Huang Y, Zhang M, Alonso J, Haro JM, Vilagut G, Bromet EJ, Gluzman S, Webb C, Kessler RC, Merikangas KR, Anthony JC, Von Korff MR, Wang PS, Brugha TS, Aguilar-Gaxiola S, Lee S, Heeringa S, Pennell BE, Zaslavsky AM, Ustun TB, Chatterji S, WHO World Mental Health Survey Consortium (2004) Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 291(21):2581–2590

10. Diener E, Emmons RA, Larsen RJ, Griffin S (1985) The satisfaction with life scale. J Pers Assess 49:71–75

11. Dilling H, Weyerer S (1984) Psychische Erkrankungen in der Bevölkerung bei Erwachsenen und Jugendlichen. In: Dilling H, Weyerer S, Castell R (eds) Psychische Erkrankungen in der Bevölkerung. Enke Verlag, Stuttgart, pp 1–122

12. Dilling H, Weyerer S (1984) Prevalence of mental disorders in the small-town–rural region of Traunstein (Upper Bavaria). Acta Psychiatr Scand 69:60–79

13. Dohrenwend BS, Krasnoff L, Askenasy AR, Dohrenwend BP (1978) Exemplification of a method for scaling life events: the peri life events scale. J Health Soc Behav 19(2):205–229

14. Dohrenwend BP, Levav I, Shrout PE (1980) Screening scales from the Psychiatric Epidemiology Research Interview (PERI) (unpublished data)

15. Eich D, Ajdacic-Gross V, Condrau M, Huber H, Gamma A, Angst J, Rössler W (2003) The Zurich study: participation patterns and Symptom Checklist 90-R scores in six interviews, 1979–99. Acta Psychiatr Scand 108(suppl 418):11–14

16. ESEMeD/MHEDEA (2000) Investigators (2004) sampling and methods of the European study of the epidemiology of mental disorders/ESEMeD) project. Acta Psychiatr Scand 109(suppl 420):8–20

17. Essen-Möller E (1956) Individual traits and morbidity in a Swedish rural population. Acta Psychiatr Neurol Scand Suppl:100

18. Feuerlein W, Ringer C, Küfner H, Antons K (1977) Diagnose des Alkoholismus: per Münchner Alkoholismustest (MALT). MMW Munch Med Wochenschr 119:1275–1282

19. Fichter MM, Weyerer S (1982) Prävalenz des hohen Blutdruckes in einem ländlichen Gebiet. MMW Munch Med Wochenschr 124:753–756

20. Fichter MM (1990) Verlauf psychischer Erkrankungen in der Bevölkerung. Springer, Heidelberg

21. Fichter MM, Narrow WE, Roper MT, Rehm J, Elton M, Rae DS, Locke BZ, Regier DA (1996) Prevalence of mental illness in Germany and the USA: comparison of the Upper Bavarian study and the epidemiologic catchment area program. J Nerv Ment Dis 184:598–606

22. Folstein MF, Folstein SE, McMugh PR (1975) "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12:189–198

23. Fydrich T, Sommer G, Brähler E (2002) F-SozU—Fragebogen zur sozialen Unterstützung. In: Brähler E, Schumacher J, Strauß B (eds) Diagnostische Verfahren in der Psychotherapie. Hogrefe, Göttingen, pp 150–153

24. Goldberg DP, Cooper P, Eastwood RR, Kedward MB, Shepherd M (1970) A standardized psychiatric interview for use in community surveys. Br J Prev Soc Med 24:18–23

25. Graaf RD, Bijl RV, Smit F, Ravelli A, Vollebergh WAM (2000) Psychiatric and sociodemographic predictors of attrition in a longitudinal study. The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Am J Epidemiol 152(11):1039–1047

26. Hagnell O, Lanke J, Rorsman B, Öjesjö L (1982) Are we entering an age of melancholy? Depressive illness in a prospective epidemiological study over 25 years. Psychol Med 12:279–289

27. Hagnell O (1989) Repeated incidence and prevalence studies of mental disorders in a total population followed during 25 years. The Lundby Study, Sweden. Acta Psychiatr Scand 348(suppl):61–77

28. Horton NJ, Laird NM, Murphy JM, Monson RR, Sobol AM, Leighton AH (2001) Multiple informants: mortality associated with psychiatric disorders in the Stirling County Study. Am J Epidemiol 154(7):649–656

29. Insel TR, Fenton WS (2005) Psychiatric epidemiology. It's not just about counting anymore. Arch Gen Psychiatry 62:590–592

30. Isohanni MK (2001) Administrative aspects in longitudinal studies: how to navigate on a stormy and dangerous ocean? Acta Psychiatr Scand 104(1):1–3

31. Jacobi F, Wittchen HU, Holting C, Sommer S, Lieb R, Hofler M, Pfister H (2002) Estimating the prevalence of mental and somatic disorders in the community: aims and methods of the German National Health Interview and Examination Survey. Int J Methods Psychiatr Res 11(1):1–18

32. Jakubaschk J, Klug J, Weyerer S, Dilling H (1978) Bedarf and Behandlungsbedürftigkeit—Überlegungen zur psychiatrischen Versorgung. Psychiatr Prax 5:203–211

33. Junk J, Junk U (1981) Entwicklung eines Messinstrumentes zur epidemiologischen Untersuchung der psychosozialen Belastung durch Lebensereignisse (Life Events) und des Coping-Potentials. Unpublished diploma theses, University of Heidelberg

34. Katschnig J (1986) Life event and psychiatric disorders. Cambridge University Press, Cambridge

35. Kendell RE, Everitt B, Cooper JE, Sartorius N, David ME (1968) Reliability of the present state examination. Soc Psychiatr 3:123–129

36. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS (1994) Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 51(1):8–19

37. Kessler RC, Berglund P, Chiu WT, Demler O, Heeringa S, Hiripi E, Jin R, Pennell BE, Walters EE, Zaslavsky A, Zheng H (2004) The US National Comorbidity survey Replication (NCS-R): design and field procedures. Int J Methods Psychiatr Res 13(2):69–92

38. Kleining G, Moore H (1968) Soziale Selbsteinstufung (SSE). Ein Instrument zur Messung sozialer Schichten. Kolner Z Soz Sozialpsychol 20:502–552

39. Lehtinen V, Lindholm T, Veijola J, Vaisanen E, Puukka P (1991) Stability of prevalences of mental disorders in a normal population cohort followed for 16 years. Soc Psychiatry Psychiatr Epidemiol 26(1):40–46

40. Leighton AH (1959) My Name Is Legion: the Stirling County Study of psychiatric disorder and sociocultural environment, vol 1. Basic Books, New York

41. Lorr M, Klett CJ, McNair DM, Lasky JJ (1963) Inpatient multidimensional psychiatric scale (IMPS). Consulting Psychologists Press, Palo Alto

42. Murphy JM, Neff RK, Sobol AM, Rice JX, Olivier DC (1985) Computer diagnosis of depression and anxiety: the Stirling County Study. Psychol Med 15:99–112

43. Murphy JM, Monson RR, Olivier DC, Sobol AM, Leighton AH (1987) Affective disorders and mortality. A general population study. Arch Gen Psychiatry 44(5):473–480

44. Murphy JM, Laird NM, Monson RR, Sobol AM, Leighton AH (2000) A 40-year perspective on the prevalence of depression from the Stirling county Study. Arch Gen Psychiatry 57:209–215

45. Murphy JM, Horton NJ, Laird NM, Monson RR, Sobol AM, Leighton AH (2004) Anxiety and depression: a 40-year perspective on relationships regarding prevalence, distribution, and comorbidity. Acta Psychiatr Scand 109:355–375

46. Narrow WE, Rae DS, Robins LN, Regier DA (2002) Revised prevalence estimates of mental disorders in the United States: using a clinical significance criterion to reconcile 2 surveys' estimates. Arch Gen Psychiatry 59:115–123

47. Nettelbladt P, Bogren M, Mattison C, Öjesjö L, Hagnell O, Hofvendahl E, Toraker P, Phugra D (2005) Does it make sense to do repeated surveys?—The Lundby Study, 1947–1997. Acta Psychiatr Scand 111:444–452

48. Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablensky A, Pickens R, Regier DA, Sartorius N, Towle LH (1989) The composite international diagnostic interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 45:1069–1077

49. Robins LN, Regier DA (1991) Psychiatric disorders in America: the epidemiologic Cathment Area Study. The Free Press, New York

50. Rorsman B, Hagnell O, Lanke J (1982) Mortality in the Lundby study. Natural death in different forms of mental disorder in a total population investigated during a 25-year period. Neuropsychobiology 8(4):188–197

51. Rothman KJ, Greenland S (1998) Modern epidemiology, 2nd edn. Lippincott-Raven, Philadelphia

52. Spitzer RL, Fleiss JL, Endicott J, Cohen J (1967) Mental Status Schedule: properties of factor analytically derived scales. Arch Gen Psychiatry 16:479–493

53. Spitzer RL, Williams JBW, Gibbon M (1985) Instructional manual for the structured clinical interview for DSM-III-R (SCID, 7/1/1985 revision). Biometrics Research Department, New York State Psychiatric Institute, 722 West, 168th Street, New York, NY 10032. (German editing: Wittchen HU, Zaudig M, Schramm E, Sprengler P, Mombaur W, Klug J, Horn R "Strukturiertes Klinisches Interview für DSM-III-R" (SKID). Interviewheft und Manual. Beltz, Weinheim)

54. Srole L (1975) Measurement and classification in sociopsychiatric epidemiology: Midtown Manhattan Study (1954) and Midtown Manhattan Restudy (1974). J Health Soc Behav 16:347–364

55. Srole L, Langner T, Michael S, Opler MK, Rennie TAC (1962) Mental health in the Metropolis: the Midtown Manhattan Study. McGraw-Hill, New York

56. Stieber J, Döring H, Keil U (1982) Häufigkeit, Bekanntheitsund Behandlungsgrad des hohen Blutdrucks in einer Großstadtbevölkerung der Bundesrepublik Deutschland; Ergebnisse der Münchner Blutdruckstudie I. MMW Munchn Med Wochenschr 124:747–752

57. Thalmann B, Monsch AU (1997) CERAD—Neuropsychologische Testbatterie: Vorläufige Normen. Memory Clinic, Basel

58. Tsuang MT, Woolson RF (1977) Mortality in patients with schizophrenia, mania, depression and surgical conditions. A comparison with general population mortality. Br J Psychiatry 130:162–166

59. Wing JK, Birley JLT, Cooper JE, Graham P, Isaaes AD (1967) Reliability of a procedure for measuring and classifying "Present Psychiatric State". Br J Psychiatry 113:499–551

60. Wittchen HU, Pfister H (1997) Diagnostisches Expertensystem für Psychische Störungen DIA-X. Swets Test Services, Frankfurt

61. World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders. Clinical descriptions and diagnostic guidelines. World Health Organization, Geneva

62. World Health Organization (2000) World Health Organization: Disability Assessment Schedule WHODAS II Phase 2 field trials. Health services research. World Health Organization, Geneva, pp 678–687

63. Zerssen von D, Koeller DM (1976) Die Beschwerdenliste (BL). Beltz, Weinheim