Reliability of life event assessments: test—retest reliability and fall-off effects of the Munich interview for the assessment of life events and conditions

Hans-Ulrich Wittchen, Cecilia Ahmoi Essau, Heidemarie Hecht,
Wolfgang Teder and Hildegard Pfister

Max Planck Insitute for Psychiatry. Kraepelinstr. 10, 8000 Munich 40, F.R.G.

Summary

This paper presents the findings of two independent studies which examined the test-retest reliability and the fall-off effects of the Munich Life Event List (MEL). The MEL is a three-step interview procedure for assessing life incidents which focuses on recognition processes rather than free recall. In a reliability study, test–retest coefficients of the MEL, based on a sample of 42 subjects, were quite stable over a 6-week interval. Stability for severe incidents appeared to be higher than for the less severe ones. In the fall-off study, a total rate of 30% fall-off was noted for all incidents reported retrospectively over an 8-year period. A more detailed analysis revealed average monthly fall-off effects of 0.36%. The size of fall-off effects was higher for non-severe and positive incidents than for severe incidents. This was particularly evident for the symptomatic groups. Non-symptomatic males reported a higher overall number of life incidents than females. This was partly due to more frequent reporting of severe incidents. The findings of the fall-off study do not support the common belief that the reliability of life incident report is much worse when the assessment period is extended over a period of several years as compared to the traditional 6-month period.

Key words: Munich Life Event List; Life events; Recall; Fall-off effects

Introduction

One important and controversial methodological problem in life event research is the question of the reliability of methods used to assess life

events, life conditions, and their impact on the individual (Brown, 1974; Rabkin and Struening, 1976; Paykel, 1983; Schroeder and Costa. 1984; Katsching, 1987). Final judgements about this topic are still mainly hampered by the fact that the few studies which have examined this question in more detail used different strategies for the assessment of life events as well as quite remarkably different approaches for measuring reliability. Furthermore, studies vary considerably with regard to selection of subjects. statistics used. and selection of target variables derived from the respective interviews. Assessment strategies range from simple self-report checklists (e.g., Holmes and Rahe's Schedule of Recent Experience. 1967) to more detailed self-report questionnaires (Dohrenwend et al., 1978) and extremely elaborate interview and rating procedures such as the Bedford College Life Event Schedule (LEDS) (Brown, 1974).

Reliability measures include studies examining the test-retest concordance between global scores from a life event questionnaire (e.g., the so-called life change units), as well as examinations of the instrument concordance with reports from 'significant others' and, more often, analyses of the 'falloff' effect. The 'fall-off' effect is defined as a decrease in the number of events reported by a subject as the period of time lengthens between the occurrence and the assessment of those events. As this effect was found to occur even in relatively short time intervals of 6 months (Jenkins et al., 1979) it was interpreted as an indicator of serious recall problems. Assuming that the accuracy of recall drops further. researchers were generally recommended to not extend the assessment period for more than a year. However, empirical evidence for this assumption is missing. Furthermore. a considerable variation among studies has to be mentioned in this context. Studies using checklists ascertained fall-off effects of about 445%' per month (Jenkins et al., 1979: Monroe. 1982). More complex approaches using systematic probing reported fall-off rates of only 9% in a 6-month period (Paykel, 1983). and Brown and Harris (1982). using the LEDS, found rates of about 1-3s per month. These findings strongly suggest that - beside other well-known variables such as the severity of an event (Paykel, 1983) - the extent of the fall-off effect might depend on the method of data collection. There seems to be a growing consensus that poorly standardized interviews as well as self-report questionnaires are markedly less reliable than the more elaborate interview methods, such as the one by Brown (1974). As Paykel (1983) recently summarized in a review: 'Careful interview techniques appear to produce reliable and valid information. Self report inventories have generated much productive research but should no longer be regarded as acceptable' (p. 92). The superiority of elaborate interviews compared to questionnaires for assessing life events is comprehensible within the realm of psychological research on recall (retrieval) processes for past events. Although not specifically referring to types of events as measured in life event approaches, memory research has indicated that the retrieval process for past events might be improved by the so-called memory aids (Harris, 1978; Sunderland et al.. 1983). The few available data from psychological memory research suggest that the reliability and validity of recall for everyday and non-severe events drops most quickly during the first few days after such events have happened. After this initial rapid drop, however. the quality of the rate of recall does not seem to deteriorate further to any marked degree. Rather. it seems to remain quite stable even if long periods of time of up to 15 years are analyzed (Sunderland et al.. 1983). This is particularly the case if memory aids are used and if the assessment strategy uses 'recognition' instead of free recall techniques. Memory aids typically refer to assessment interventions in which the interviewer specifically mentions the relevant situations to be assessed. their context variables. their clustering in a critical time period as well as cues and marked points which are easily remembered, such as birthdays, the date of a disaster of major societal events. Thus, we can assume for retrospective life incident assessments over longer time spans that the more concrete one describes specific incidents and the more thoroughly one places the respective incidents in proper chronological order. and the better and more comprehensive their particular 'context variables' are specified, the more the subject's accuracy of recall will be improved. The specific questions and probes used in some of the

more complex life event interviews (e.g., the LEDS) could be judged as more or less systematic specific memory aids which have the potential of stimulating a more accurate retrieval process.

Because of the inherent disadvantages of complex approaches such as the one by Brown (i.e., a time-consuming interview, long training of interviewers, complicated rating procedures, etc.), the search continues for more viable compromises between elaborate life event interviews and self-report questionnaires. Additionally, it has to be acknowledged that the format and content of the traditional life event scales and interviews are generally not suitable for the retrospective assessment of longer time spans, for example, over several years. As life event instruments usually exclusively aim at examining the occurrence of stressful, especially severe events that possibly play a causal role in the onset of the disorder shortly before the interview (mostly 6 months), their applications for long-term studies (which are also interested in evaluating positive incidents and a general characteristic of psychosocial changes) often seem inappropriate.

Based on these considerations, we have developed and tested (Maier-Diewald et al., 1983; Dehmel and Wittchen, 1984; Teder, 1984) a complex method for the retrospective assessment of positive and negative life events (events are defined as changes in everyday life) and life conditions ('chronic events' which occur for a duration of at least 3 months) for assessment periods of several years. Throughout this paper, the term 'incident' will be used to indicate the combination of all life events and conditions.

Briefly, the Munich Interview for the Assessment of Life Events and Conditions (MEL) is a three-step procedure that uses a very detailed and specific description of life incidents (questionnaire- like to stimulate recognition) where the subject rates the presence of incidents in yearly intervals. The use of a life chart, which mentions anchor incidents and the date of their occurrence, is intended to facilitate the general timing of incidents in the assessment period chosen. This first step is followed by an interview about the incidents and their context; furthermore, incidents not yet mentioned as well as the time order of incidents are checked carefully. This includes systematic probe questions spelled out in the manual as well as memory aids (for more detailed description see Maier-Diewald et al., 1983). Incidents which do not appear on the MEL but are remembered during this part of the MEL interview are also coded.

The purpose of this paper is to present the results of two independent studies which attempted to examine the test-retest reliability and the fall-off effects of the MEL. Another purpose is to explore the effects of sex, the presence or absence of a mental disorder, as well as the severity and the nature of incidents on recall.

Methods

The test-retest study (study A)

Twenty-three males and 19 females with no mental disorder were examined twice with an average of 6 weeks (mean = 42.7 days, SD = 6.3) between the test and the retest examination. Nineteen subjects belonged to the age group 25-30 years, 13 to the age group 31-35 years, and 10 to the age group 36-40 years. The mean age was 31.6 years (SD = 6.3) 52% were married, and 64% were employed at the time of the interview. Furthermore, it should be noted that almost half of the sample was recruited from students who had just completed their examinations. Thus this sample is not representative, and neither are the number and kind of incidents mentioned. All investigations were carried out by one of the authors (W.T.).

The fall-off study (study B)

Data from the Munich Follow-up Study (MFS) were used for the examination of the fall-off effects. Since this study has been described in detail elsewhere we will limit the method section to some of the most important issues (Wittchen et al., 1985; Wittchen, 1987; Wittchen and Bronisch, 1988; Wittchen and von Zerssen, 1985, 1988). The MFS includes a general population sample to study fall-off effects in a representative sample as well as a cohort of former psychiatric inpatients of the Max Planck Institute for Psychiatry, aged 25-64 years at the time of the follow-up interview. The average follow-up period for both groups was 7 years (Table 1).

TABLE 1

SOCIODEMOGRAPHIC CHARACTERISTICS OF PATIENTS AND CASES

		Representative sample of the adult population (MFS)				Former psychiatric inpatients neuroses (ICD 300.0; 300.2; 300.4) ($n = 76$)		Former psychiatric inpatients psychoses (ICD 295.0; 296.0; 297.0) ($n = 97$)	
		Non-symptomatic cases ($n = 321$)		Cases ($n = 123$)					
		N	%	N	%	N	%	N	%
Sex	Male	180	56.0	35	29.0	40	53.0	46	47.0
	Female	141	44.0	88	72.0	36	47.0	51	53.0
Age	Mean	44.25		45.75		39.47		39.54	
	SD	9.74		9.29		8.13		9.47	
	25–35	65	20.0	19	16.0	26	34.0	36	37.0
	36–45	121	37.0	46	37.0	33	43.0	36	37.0
	46–55	78	25.0	36	29.0	15	20.0	18	19.0
	> 55	57	17.0	22	18.0	2	3.0	7	7.0
Marital status	Single	25	8.0	9	7.0	33	43.0	42	43.0
	Married	269	84.0	92	75.0	32	42.0	40	41.0
	Div./Wid./Sep.	27	8.0	22	18.0	11	14.0	15	16.0
Occupational status	Employed	238	74.0	75	61.0	51	67.0	48	50.0
	Unemployed	83	26.0	48	39.0	25	33.0	49	51.0

The patient group consisted of 76 neurotic subjects, former inpatients with depressive (ICD: 300.4) or anxiety disorders (ICD: 300.2/300.0) (subgroup neurosis), and 97 former inpatients with schizophrenia (295.X) schizoaffective disorders (295.7) or affective psychosis (296.X) (subgroup psychosis). At the time of their re-examination, 42% of these patients still had marked psychopathological symptoms, and another 28% had slight symptoms as measured by the German version of the Inpatient Multidimensional Psychiatric Rating Scale (Hiller et al., 1986). Forty-six percent of these patients were either in an out- or inpatient treatment in the 6-month period preceding the interview.

From the general population sample of the original 501 subjects (76.3% completion rate) who were successfully reinterviewed in Wave II, 444 subjects who had a complete data set were used for the following analyses; 123 of these subjects had a lifetime DSM-III disorder as ascertained by the Diagnostic Interview Schedule (DIS, Robins et al.. 1981; German version: Wittchen and Rupp, 1981). Results will be reported separately for the 321 healthy subjects and the 123 lifetime 'cases' with a DSM-III disorder (Wittchen and Burke, 1988; Wittchen and von Zerssen, 1988). In addition to the life event assessment for all patients and subjects, a rather comprehensive psychosocial and psychopathological evaluation strategy was used at the time of the follow-up interview. The MEL interview itself took a substantial part of the time during the whole follow-up reevaluation (average 65 min).

Description of the MEL interview

The structure of the MEL is described in Table 2. The MEL includes 85 life incidents that are spelled out explicitly including time and frequency criteria. Furthermore, an additional open question is included for each social role area.

The whole procedure is mainly oriented towards the recognition process. Therefore, it is divided into three steps. In step I of the procedure, the interviewer and the subject read every single 'item' on the questionnaire-like list together. Each of the items is explicitly formulated, defining criteria for both duration and severity. In this step, the subject has to decide whether or not the incident has been present during the assessment period. In our study, we asked for a period of 7 years. The subject is instructed to go back and forth on the list, whenever he/she recalls incidents in the sections which have been filled

TABLE 2

THE MUNICH LIFE EVENTS LIST (MEL)

A strategy to assess life events and social conditions		Social role areas and rating dimensions	
Step I:	Identification of life events and social conditions (in the last 7 years) by means of specific items, using a list of social situations as well as a life chart. The term 'incident' is used to indicate the combination of all life events and social conditions	Social role areas:	Education/profession/household (26); Partner/marriage (14); Children (10); Parents/relatives (6); Social life/leisure time activities (8); Deaths (4); Housing (4); Health (6); Others (7) (n = 85 incidents)
Step II:	Semi-structured interview to ascertain and date each incident (life chart)	Rating dimensions:	(A) Subjective Degree of 'burden' (five-point scale)
Step III:	Rating of the incidents' impact on the subject (subjective and objective ratings)		Degree of positive or negative impact (five-point scale)
The MEL offers an additional Step IV, in which the interviewer asks for the three most important incidents in the assessment period, and a rating for their past and present impact on the person's life.			(B) Objective Gain vs. loss; Responsible vs. non-responsible; Positive vs. negative; Controllable vs. not controllable; Desirable vs. undesirable; Independent vs. dependent; Adjustment score (1=10); Degree of threat (0–5); etc.

out. Furthermore, the subject is free to ask the interviewer whenever a problem or a difficulty of judgement occurs. This step is facilitated by a rough chronology. This means that an incident should be marked by a cross in the year in which it occurred according to the subject. Furthermore, well-known key incidents, such as important birthdays and other 'valid' landmarks are offered to facilitate a more accurate timing ('memory aids').

Once the number of specific incidents has been obtained, step 11 follows. This step consists of a semi-structured interview designed to assess context variables of the life incidents indicated on the list. In an attempt to assist the subjects date each incident as accurately as possible, memory aids for timing the incidents are given. There are three levels of timing: by date, month, or by 3-month period. In addition, the interviewer codes interrelated incidents that have clustered together according to the rules spelled out in the manual. Step III then determines the impact of each incident on the respondent. For each incident, both 'subjective' (by the patient) and 'objective' ratings (by the observers) are coded. Subjective ratings are limited to two dimensions (degree of distress on a five-point scale and a bipolar positive-negative scale) that have been derived from a factor analytic trial study (Dehmel and Wittchen. 1984). Objective ratings include more options which are specified in Table 2. For this paper we only report the subjective rating for severity which correlated substantially with the objective rating ($r = 0.67$). The reliability of the objective ratings will be reported in a separate paper. It should be noted that the subjective ratings could not necessarily be regarded as the same as severity ratings used by Brown and his associates.

Analysis

As a measure of concordance between test and retest interviews we used percentage agreement and kappa coefficients including all 85 items of the MEL. To examine fall-off effects, the frequency of events per year was analyzed. To test the effects for significance, we did not use regression analytic methods because data were not normally distributed. Instead, the Successive Difference Variance test by Neumann et al. (cited in Sachs, 1969) was used. This method takes into account the variance of each sample and ascertains a significant chronological trend effect if consecutive values are more alike than the more distant ones. These calculations were performed for the MEL overall values as well as for the scores referring to specific role areas.

Results

Test-retest analysis (study A)

The 42 subjects reported an overall number of 1053 incidents during the test and 972 during the retest interview; a total of 161 incidents have been included in either the test or retest interview. Women reported significantly more incidents than men ($P < 0.05$). Although there was a steady decrease in the number of incidents reported from year 1 (210 in the test, 189 in the retest interview)

TABLE 3

NUMBER OF INCIDENTS AND CONCORDANCE RATES BY NUMBER OF YEARS BEFORE THE INTERVIEW

	Years before the interview							
	8	7	6	5	4	3	2	1
Number of incidents in:								
Test interview	82	86	116	123	126	151	159	210
Retest interview	75	73	111	115	129	136	144	189
Agreement (%)	98.03 [a]	96.59	94.79	93.36	92.82	92.10	93.90	93.36
Kappa	0.91	0.84	0.82	0.78	0.77	0.77	0.83	0.85

[a] Kappa and overall percentage agreement were calculated using a cross-tabulation of the presence and absence of all 85 incidents over the 8 years.

All κ values were significant ($P < 0.01$).

TABLE 4a

RELIABILITY OF DATING OF LIFE INCIDENTS

			Test interview					
			By year	Three-month periods				Total
				1	2	3	4	
Retest interview	By year		196	9	3	2	4	214
	Three-	1	5	108	6	1	2	120
	month	2	3	9	128	11	1	152
	periods	3	4	5	16	91	7	123
		4	1	5	5	8	100	119
	Total		209	136	158	113	114	728

Percentage agreement: 84%; $\kappa = 0.83$ ***.

before the interview to year 7 (82 in the test and 75 in the retest interview), the concordance rates were very high. Table 3 shows the number of incidents reported for each year in the test and the retest interview in addition to the percentage agreement and kappa coefficients. All kappa coefficients were significant.

There were no significant differences with regard to the reliability coefficients between male and female subjects. The overall agreement between test and retest interviews was 95.49%; $_k$ was 0.85. The differential analysis of agreement for each 'social role area' revealed perfect concordance for the most severe incidents, such as deaths or extremely threatening incidents within the family, and lower agreement for less severe incidents such as those prominent in the MEL areas for social Contacts (90.5%, $_k = 0.74$). Financial events (94.6%, $_k = 0.78$) also showed a high concordance.

The data in Table 3 reveal substantial differences in the number of incidents between year 1 and year 8. Incidents elicited in year 8 were only 39% of those named in year 1. This, however, could not be regarded as a pure 'fall-off' rate because it is confounded heavily with a cohort effect. As the majority of the subjects were either students or professionals who had just completed university/ school examinations, these subjects can be considered as entering a new stage of life with the consequence that they experienced a high number of MEL positive and negative incidents. Thus, there is a real effect because many incidents clustered 2-3 years before the interview was con-

TABLE 4b

RELIABILITY OF SEVERITY OF LIFE INCIDENTS

		Test interview					
		Non-severe 1	Slightly severe 2	Moderately severe 3	Markedly severe 4	Extremely severe 5	Total
Retest interview	1	213	50	2	–	–	266
	2	35	159	16	4	1	215
	3	9	33	56	18	2	118
	4	3	12	30	27	6	78
	5	1	4	3	27	42	77
Total		261	258	107	76	52	754

Percentage agreement: 65%; $\kappa = 0.56$ ***.

ducted as compared to the period before that. Therefore, we refrain from further discussion of this finding.

To test the accuracy of dating, we analyzed agreement for the same year and for the 3-month periods over the whole time span covered by our interview. No separate reliability test for day or week of the incidents was performed because the subjects could not accurately date 46% of the incidents to the level of days or weeks, but only to the respective 3-month periods. Table 4a summarizes the number and the degree of disagreements between the two interviews. Generally, agreement was high for both time frames, with more disagreements for the 3-month periods. Concordance for agreement in the occurrence of incidents within the same 3-month period varied only slightly between $_k$ values of 0.83 for the most distant year (year 8) and 0.84 in the year the interview was conducted.

With regard to the stability of the subjective severity ratings (degree of stress experienced as measured on a five-point scale), considerably lower agreement coefficients were found. Percentage agreement was 65%, the $_k$ value 0.56 ($P < 0.01$).

Agreement coefficients were significantly higher for incidents which were rated as 'markedly' and 'extremely' severe. The $_k$ value for less severe incidents in the test interview was 0.42, for the more severe incidents 0.65. As results for the other rating dimensions were found to be similar, they are not reported separately here. Surprisingly, no indications were found for a drop in reliability of subjective severity ratings with regard to more distant incidents.

In order to study the reliability of dating and severity, only those incidents which were indicated in the test as well as in the retest interview were chosen for analysis. Therefore, we had to exclude 161 incidents (including life conditions which have occurred with a duration of more than 3 months, thus counting more than once). Furthermore, 26 respondents in either the test or the retest interview were not able to indicate the exact date for some incidents and thus these respondents had to be dropped from the analysis as well, which re-

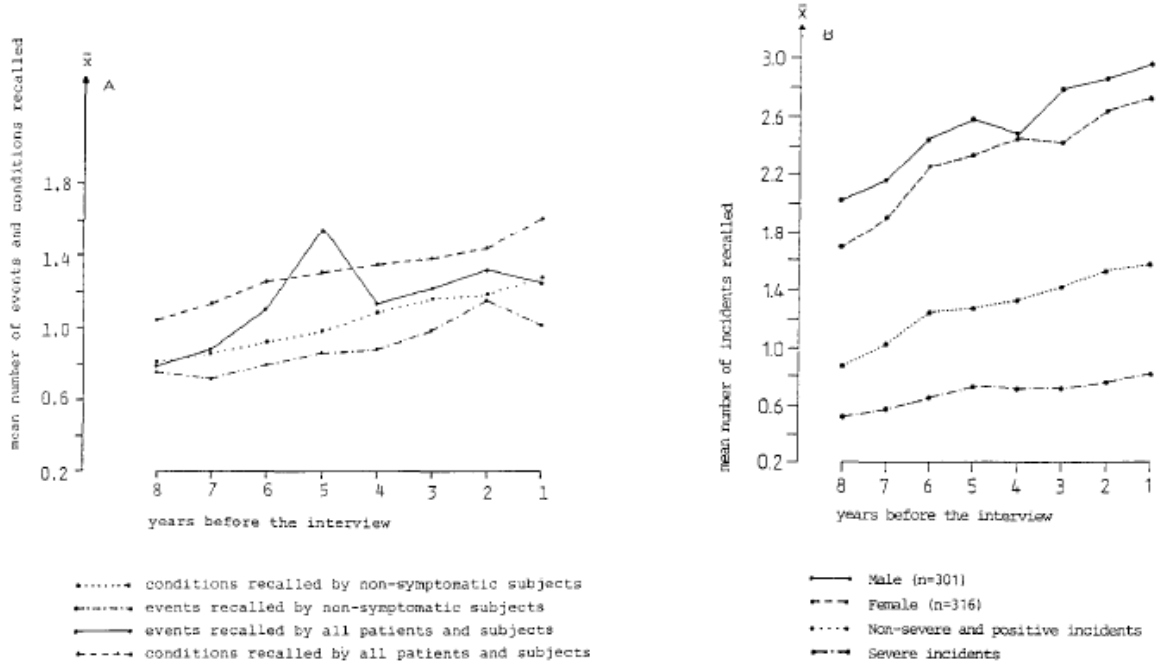

Fig. 1. (A) Mean number of events and conditions recalled by non-symptomatic subjects ($n = 321$), and by all patients and subjects ($n = 617$). (B) Mean number of incidents recalled by all male and female subjects and sub-divided according to their severity rating.

sulted in a total of 728 incidents for the dating analysis and 754 incidents for the severity analysis.

Fall-off effects (study B)

The decline in the number of life incidents reported with increase in time from the interview was almost linear and significant. For the whole g-year period, there was a drop of almost 30% for incidents.

In order to determine the impact of life incidents on the subjects, they were asked to make a subjective rating of the stress impact of each life incident which they experienced on a five-point scale. Only subjective and not objective ratings were considered because one's judgement of the stressfulness of an incident is very subjective. The following subjective ratings were used: 1 = not severe, 2 = slightly severe, 3 = moderately severe, 4 = markedly severe and 5 = extremely severe. For the subsequent analyses, ratings of 1-2 were grouped as not severe and ratings of 3-5 were classified as severe. An important finding was that the size of drop-off effects was greater for the non-severe and positive incidents (38.7%) than for the severe incidents (26.3%) (see Fig. 1B). This drop was quite moderate. The more detailed analysis for monthly time intervals revealed an average drop-off effect of 0.36% per month. In comparison with previous findings (Paykel, 1980; Schmid et al., 1981) our monthly fall-off rate was the lowest. For example, Brown and Harris (1982) found a monthly fall-off rate of l-3%, Paykel (1980) l%, and Schmid et al. (1981) 10%.

A separate analysis was done (see Fig. 1A) to examine whether or not events and conditions drop in the same manner during the entire assessment period. The result indicated that a fall-off rate of events and conditions occurred in a similar manner as one moved back in time from the interview. Thus, in subsequent analyses events and conditions were combined to give a total score of incidents.

Although males were found to have a significantly higher mean of severe, as well as non-severe and positive, incidents (Fig. IB), none of the sex differences in subsequent analyses on other variables proved to be significant.

Separate analyses were also made to examine

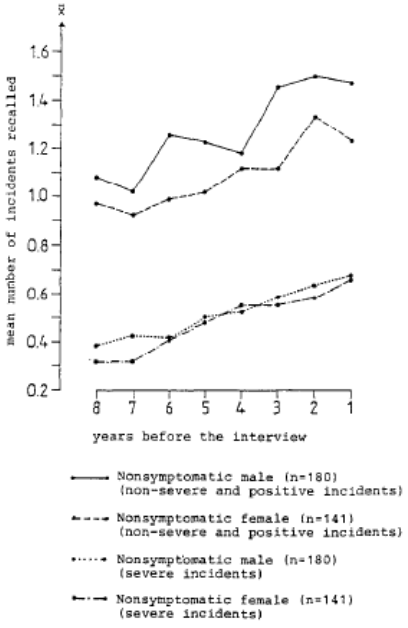

Fig. 2. Mean number of incidents recalled by the non-symptomatic group.

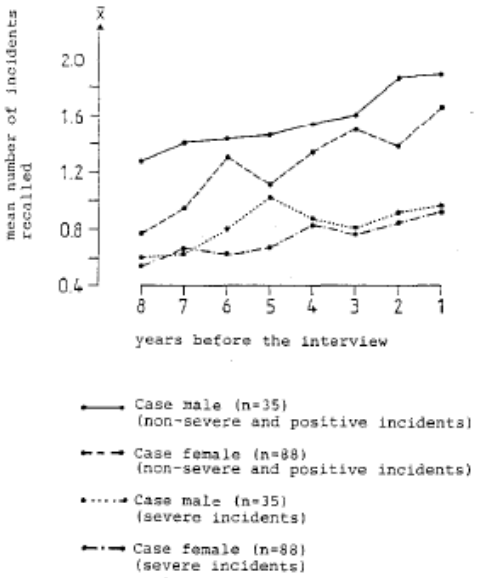

Fig. 3. Mean number of incidents recalled by the case group.

the fall-off rates of both severe and non-severe and positive incidents for males and females (separately) in each of the four groups. As shown in Fig. 2, both non-symptomatic males and females reported a significantly higher mean of non-severe and positive incidents than severe ones. The fall-off rates of non-severe and positive and severe incidents were quite apparent as one moved away in time from the interview. However, while the fall-off rate for non-severe and positive incidents showed peaks and troughs in magnitude, the fall-off rate for severe incidents was almost linear.

Compared to the female cases, males reported a slightly higher mean of severe as well as non-severe and positive incidents (Fig. 3). This result was similar to the one found in the non-symptomatic group. The drop was almost linear for all kinds (severe, non-severe and positive) of incidents.

When analyzing the patient's data, neurotic males reported significantly higher means of nonsevere and positive incidents than neurotic females. The same finding was obtained for the severe incidents (Fig. 4). A significant drop in severe as well as in non-severe and positive incidents was noted for males and females. All groups showed much more variation with a rather high and rather low mean score of incidents as compared to the cases and the non-symptomatic subjects. There is no apparent explanation for the drop 3 years prior to assessment in the female neurotic group.

Males with psychosis also proved to have a higher variability of mean values than cases and non-symptomatic subjects. Furthermore, they had higher means of severe incidents than females (Fig. 5). Fall-off effects for incidents were significant for both males and females. Means of nonsevere and positive incidents were relatively similar in males as well as in females.

To investigate further the finding of higher mean scores for males, an additional analysis was done to examine social role areas and types (events versus conditions) of life incidents experienced by non-symptomatic males and females. It has to be remembered that the MEL definition of chronic difficulties (i.e., conditions) does not overlap with the LEDS definition. For a MEL incident to be qualified as chronic, it must be present for at least 3 months. Table 5 indicates that non-symptomatic males generally reported more life conditions in nearly all the social role areas. This finding seems

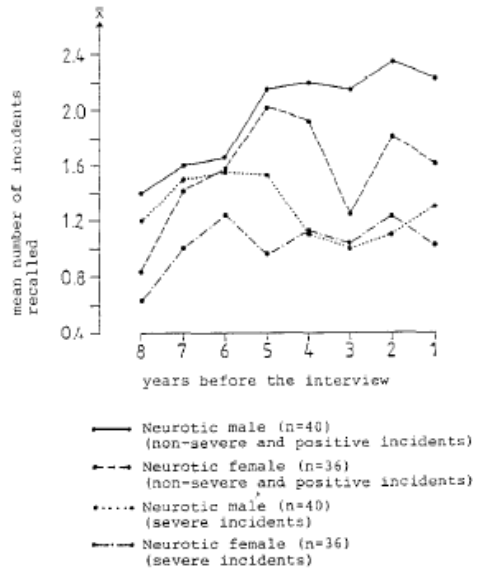

Fig. 4. Mean number of incidents recalled by the neurotic group.

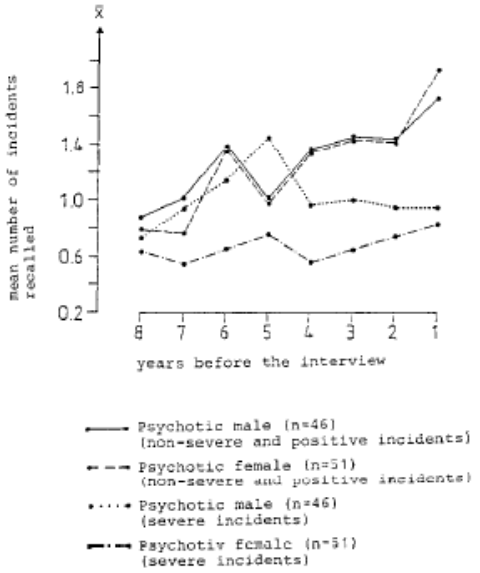

Fig. 5. Mean number of incidents recalled by the psychotic group.

TABLE 5

NUMBER OF INCIDENTS RECALLED BY NON-SYMPTOMATIC SUBJECTS ACCORDING TO SEX AND AGE

Social area	Young male ($n_1 = 95$) < 44 years			Old male ($n_2 = 85$) > 44 years			Young female ($n_3 = 78$) < 44 years			Old female ($n_4 = 63$) > 44 years		
	Ev.	Con.	Total	Ev.	Con.	Total	Ev.	Con.	Total	Ev.	Con.	Total
Education	85	8	93	8	1	9	25	–	25	1	–	1
Profession/household	221	468	689	94	596	690	140	444	584	56	351	407
Partner/marriage	135	797	932	27	314	341	69	424	493	14	344	358
Children	120	351	471	71	322	393	89	439	528	54	107	161
Parents	18	184	202	1	156	157	12	261	273	4	59	63
Social life	128	662	790	81	579	660	90	541	631	52	251	303
Death	52	–	52	78	–	78	60	–	60	64	–	64
Housing	92	150	242	23	56	79	53	127	180	13	38	51
Financial	62	36	98	24	19	43	30	55	85	19	31	50
Law	12	11	23	4	20	24	5	5	10	2	6	8
Health	78	359	437	90	483	573	74	279	353	71	447	518
MEL total	1003	3026	4029	501	2546	3047	647	2575	3222	350	1634	1984

Ev. = event, Con. = condition.

to be related to the subject's age. For example, non-symptomatic males under 44 years reported the highest number of life incidents, followed by females under 44, males over 44 years and females over 44 years. The most frequent incidents reported by young males (under 44 years) include those items related to partner/marriage, followed by social life, profession/ household, and children. For older females (over 44 years), the most frequent incidents reported are those related to profession/ household. The most frequent incidents reported by young females (under 44 years) belonged to the social role areas of social life, profession/ household, children, and partner/marriage. For older females (over 44 years), higher rates of life conditions were reported in the social role areas of health/illness, followed by profession/ household, partner/ marriage, and social life.

TABLE 6

NUMBER AND PERCENTAGE OF INCIDENTS WHICH WERE SEVERE IN EACH OF THE SOCIAL ROLE AREAS RECALLED BY NON-SYMPTOMATIC SUBJECTS ACCORDING TO SEX AND AGE

Social area	Young male ($n_1 = 95$) < 44 years		Old male ($n = 85$) > 44 years		Young female ($n_3 = 78$) < 44 years		Old female ($n_4 = 63$) > 44 years	
	N	%	N	%	N	%	N	%
Education	37	39.7	3	33.3	4	16.0	–	–
Profession/household	198	40.5	270	39.1	132	22.6	47	11.5
Partner/marriage	284	30.5	43	12.6	125	25.4	58	16.2
Children	62	13.2	48	12.2	185	35.0	73	45.3
Parents	14	6.9	43	27.4	71	26.0	42	66.6
Social life	53	6.7	63	9.5	49	7.8	52	17.2
Death	35	67.3	47	60.3	47	78.3	41	64.1
Housing	110	45.5	43	54.4	106	58.9	43	84.3
Financial	41	41.8	28	65.1	13	15.3	33	66.0
Law	10	43.5	13	54.2	2	20.0	4	50.0
Health	246	56.3	439	76.6	270	76.5	384	74.1
MEL total	1090	27.1	1040	34.1	1004	31.2	777	39.2

TABLE 7

MEAN NUMBER OF SEVERE AND NON-SEVERE INCIDENTS EXPERIENCED BY SUBJECTS IN THE FOUR GROUPS

Group	Years before the interview															
	8		7		6		5		4		3		2		1	
	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS	S	NS
Non-symptomatics ($n = 312$)	0.36	1.03	0.38	0.98	0.41	1.13	0.49	1.14	0.54	1.16	0.58	1.31	0.62	1.43	0.66	1.38
N – S	0.67		0.60		0.72		0.65		0.62		0.73		0.81		0.72	
Cases ($n = 123$)	0.56	0.93	0.64	1.08	0.67	1.34	0.79	1.21	0.84	1.39	0.79	1.53	0.86	1.53	0.93	1.72
N – S	0.37		0.44		0.67		0.42		0.55		0.74		0.67		0.79	
Psychotics ($n = 97$)	0.66	0.84	0.72	0.89	0.88	1.39	1.05	1.25	0.75	1.16	0.80	1.41	0.84	1.42	0.88	1.82
N – S	0.18		0.17		0.51		0.20		0.41		0.61		0.58		0.94	
Neurotics ($n = 76$)	0.93	1.13	1.28	1.51	1.46	1.63	1.22	2.14	1.11	2.07	1.04	1.72	1.16	2.06	1.17	1.93
N – S	0.20		0.23		0.17		0.92		0.96		0.68		0.93		0.76	

S = Severe incidents.

NS = Non-severe and positive incidents.

N – S = Difference between non-severe and positive, and severe incidents.

More importantly, Table 6 indicates that the higher overall rate of incidents in males was also due to more severe incidents. Severe incidents were reported mostly by males under 44 years, followed by males over 44 years, and females under the age of 44 years. Older females (over 44 years) reported the least number of severe incidents. Among young males, items related to death (67.3%) and health/illness (56.3%) were regarded as severe, whereas among old males items related to health/illness (76.6%) and financial situation (65.1%) were evaluated as severe. For young females, the items that were considered severe are those related to the social role areas of health/ illness (76.5%) and death (78.3%). A similar finding was also obtained in young males. The incidents which were rated as severe by old females belonged mainly to the social role areas of housing (84.3%) and health/illness (74.1%).

The rates of severe as well as non-severe and positive incidents were also compared for subjects in all four groups. The non-symptomatics showed a steady decrease in the number of severe incidents, whereas the symptomatics showed a modest drop. For non-severe and positive incidents and conditions, the reverse was found in that greater drops were obtained for symptomatics than for non-symptomatics.

Discussion

Our two independent studies examined the test-retest reliability (study A) and the fall-off effects (study B) of the MEL, to investigate, besides other issues, the widely held view that a subject's accurate recalling of incidents is limited to just 6 months.

Based on a sample of 42 normal subjects (study A), the test-retest coefficients of the MEL were surprisingly stable over a 6-week interval. The differential analysis of agreement for each 'social role' area showed perfect concordance for the most severe incidents and lower agreement for less severe incidents. With regard to the accuracy of dating, a high agreement was found for the year and for the 3-month period, although more disagreements were found for the 3-month period. With regard to the stability of the subjective ratings, considerably lower agreement coefficients were found. Although the test-retest study revealed a steady and overall significant decrease in the number of incidents reported from the first to the seventh year (before the interview), the concordance rates between test and retest interview were similar for recent and for more distant incidents. This study suggests that although normal subjects tend to 'forget' quite a number of incidents over the years (fall-off effect), their reporting of incidents over a 6-week period remains rather accurate.

With regard to the specific analysis of fall-off effect in study B, based on a random population sample, one important finding was that in total only a 30% fall-off rate was noted for all incidents for the whole g-year period. Fall-off effects found in the non-severe and positive incidents (38.7%) were greater than in the severe ones (26.3%). This finding seemed to support the notion that the salience of the events for the respondents plays a crucial role in recall (Brown and Harris, 1982) and is in accordance with findings by Surtees et al. (1986) and other authors who have examined this issue for shorter periods of time (Paykel, 1980).

The fall-off effect at an average rate of 0.36% per month was fairly low in comparison with other findings. For example, for less severe events, Brown and Harris (1982) found a monthly fall-off rate of about l-3%, and Paykel (1980) 1%. Both these studies used semi-structured interviews for their data collection. Studies using self-report techniques produced even higher fall-off rates. For example, monthly fall-off rates of 445% and 4% were found by Jenkins et al. (1979) and Uhlenhuth et al. (1977) respectively. Methods of data collection used could have contributed to the different findings in some of the older life event studies (Brown and Harris, 1982). Unlike many other studies, our three-step procedure is mainly oriented towards the recognition process. For example, the initial step of the MEL procedure is facilitated by a rough chronology and during the second step a semi-structured interview is carried out with the use of memory aids. One apparent advantage of having an interview as opposed to self-administered questionnaires is that the interviewer has the opportunity of probing the subject's responses. This may consequently elicit a better recall of events over long

periods of time. Our findings in study B suggest that it is the method of data collection that is responsible for the observed unreliability in reporting life incidents and not necessarily the subject's memory per se.

Differential findings were obtained for the different clinical and non-clinical subgroups in study B. As one moves away from the interview, a rather modest drop in the number of severe incidents is obtained among the symptomatics, whereas the non-symptomatics show a steady decrease (falloff). However, with regard to non-severe and positive incidents, the reverse is found with greater drops for symptomatics than for the controls. There ia no easy explanation for this finding. Symptomatics have actually experienced life events in a different frequency as well as in a different pattern over the 7-year assessment period. We could speculate that the symptomatic subjects focus more on the negative incidents, which consequently have a great impact on their life and probably on their recall, whereas the non-symptomatics focus on the good side of life. This might be interpreted as the two groups' differences in attributional styles toward life incidents. If our finding of differential attributional styles proves to be adequate, this might have some important implications for future research in this field. If our assumption about attributional styles and symptomology (e.g., depressed mood) is correct. Further research should examine more carefully the types, the sequence, and the cluster of events over a certain assessment period in order to learn more about possible systematic distortions affecting recall.

Regardless of the possible explanation of why non-symptomatic subjects differ in their recalling of non-severe and positive as well as in severe incidents, the present findings indicate that the assessment of past severe events and conditions can be extended up to 7 years in symptomatic subjects without necessarily showing stronger falloff rates than studies which were limited to shorter assessment periods of 6 months or 1 Fear. Our results, however, strongly suggest that ignoring onset dating would probably worsen reliability and fall-off effects, as the retrieval process for past life incidents would not he stimulated any more sufficiently. The central role of dating as a tool for enhancing the accuracy of recall and diminishing the degree of fall-off effects should be studied further in the future.

References

Brown, G.W. (l974) Meaning. measurement and stress of life events. In: B.S. Dohrenwend and B.P. Dohrenwend (Eds.). Stressful Life Events. Their Nature and Effects. John Wiley. New York. NY.

Brown, G.W. and Harris, T. (1982) Fall-off in the reporting of life events. Sot. Psychiatry 17, 23.

Dehmel, S. and Wittchen, H-U. (1984) Anmerkungen retrospektiver Erfasung von Lebensereignissen und Lebensbedingungen bei Verlaufsuntersuchungen – Bewertung und Vergessen. Z. Klin. Psychol. 13, 88-110

Dohrenwend, B.S., Krasnoff, L., Askenasy, A.R. and Dohrenwend, B.P. (1978) Exemplification of a method for scaling life events: the Peri Life Events Scale. J Health Soc Behav. 19, 205-29.

Harris, J. E. (1978). External memory aids. In M. M Gruneberg & P. E. Morris (Eds.), Practical aspects of memory (172-179). London: Academic Press.

Hiller, W. , von Zerssen, D., Mombour, W., Wittchen, H.-U. (1986). IMPS - Inpatient Multidimensional Psychiatric Scale. Eine multidimensionale Skala zur systematischen Erfassung des psychopathologischen Befundes.Beltz-Test. Weinheim.

Holmes, T.H., Rahe, R.H.(1967) The Social Readjustment Rating Scale. J Psychosom Res. 11,213-218.

Jenkins, C.D., Hurst, M.W., Rose, R.M.(1979) Life changes. Do people really remember? Arch Gen Psychiatry. 36, 379-84.

Katschnig, H. (1986). Life events and psychiatric disorders: controversial issues. Cambridge University Press, Cambridge

Maier-Diewald, W., Wittchen, H.-U., Hecht, H., & Werner-Eilert, K. (1983). Die Münchner Ereignisliste (MEL) - Anwendungsmanual. München: Max-Planck-Institut für Psychiatrie, Klinische Psychologie und Epidemiologie.

Monroe S.M. (1982) Life events and disorder: event-symptom associations and the course of disorder. J Abnorm Psychol. 91(1),14-24.

Paykel E.S. (1983) Methodological aspects of life events research. J Psychosom Res. 27(5),341-52.

Rabkin JG, Struening EL.(1967) Live events, stress, and illness. Science. 194(4269),1013–1020.

Robins LN, Helzer JE, Croughan J, Ratcliff KS. (1981) National Institute of Mental Health Diagnostic Interview Schedule. Its history, characteristics, and validity. Arch Gen Psychiatry. 38(4),381–389.

Sachs, L. (1969) Statistische Auswertungsmethoden. Springer-Verlag Berlin Heidelberg New. York

Schmidt, I., Scharfetter, C. and Binder, J. (1981) Lebensereignisse in Abhängigkeit von Soziodemographischen Variablen. Social Psychiatry, 16, 63-68

Schroeder, D. H. & Costa, P. T. (1984). Influence of life event stress on physical illness: Substantive effects or methodological flaws? Journal of Personality and Social Psychology, 46, 853-863.

Sunderland, A., Harris, J.E., Baddeley, A.D. (1983) Do laboratory tests predict everyday memory? A neuropsychological study. Journal of Verbal Learning and Verbal Behavior, 22, 341-357

Surtees PG, Miller PM, Ingham JG, Kreitman NB, Rennie D, Sashidharan SP. Life events and the onset of affective disorder. A longitudinal general population study. J Affect Disord. 1986 Jan-Feb;10(1):37-50.

Teder. W. (1984) Reliabilität von Life-Event-Daten und Gedächtnisleistung. Diplomarbeit. Universitat Mannheim.
Teder. W. and Wittchen, H.-U. (1986) Zur Reliabilität einer strukturierten mehrstufigen Interviewmethode zur Lebensereignissen und Lebenslagen (Münchner Ereignisliste. MEL) - Test Retest Reliabilität. Unpublished manuscript.
Uhlenhuth EH, Haberman S J, Balter MD, Lipman RS (1977) Remembering life events. In: Strauss JS, Babigian HM, Rolf M.(Eds.) The origins and course of psychopathology: methods of longitudinal research. Plenum Press, New York
Wittchen, H.-U. (1987). Chronic difficulties and life events in the long-term course of affective and anxiety disorders: Results from the Munich-Follow-up Study. In M. C. Angermeyer (Ed.), From social class to social stress. New developments in psychiatric epidemiology (pp. 176-196). Berlin: Springer.
Wittchen, H.-U., & Rupp, H.-U. (1981). Diagnostic Interview Schedule (DIS) Version II (deutsche Übersetzung). München: Max-Planck-Institut für Psychiatrie, Klinische Psychologie und Epidemiologie.
Wittchen, H.-U., & Zerssen, D. v. (1985). The longterm outcome of treatment of psychiatric disorders with special regard to chronicity of symptom and social functioning. In T. K. Sluss & M. Kramer & B. M. Gruenberg & B. Cooper (Eds.), The chronically mentally ill: An international perspective (pp. 225-244). Cambridge: University Press.
Wittchen HU, Bronisch T. (1988) Use, abuse, and dependence of alcohol in West Germany — Lifetime and 6-month prevalence in the Munich Follow-up Study. In: J. Helzer and G. Canino (Eds.). Cross-National Comparison of Rates of Alcoholism. Oxford University Press. Oxford.
Wittchen, H.-U. and Burke, J.D. (1988) DIS/DSM-111 prevalence rates of mental disorders in the US and Germany. In preparation.
Wittchen. H.-U. and von Zerssen, D. (1988) Verläufe behandelter Depressionen und Angststörungen. Eine klinisch-psychiatrische und epidemiologische Untersuchung. Springer, Berlin. Heidelberg, New York. Tokyo.
Wittchen, H.-U., Semler, G. & Zerssen, D. v. (1985). A comparison of two diagnostic methods - clinical ICD diagnoses vs. DSM-III and research diagnostic criteria using the Diagnostic Interview Schedule (version 2). Archives of General Psychiatry, 42, 677-684.