

ORIGINAL PAPER

R. F. Ferdinand · F. C. Verhulst

Psychopathology in Dutch young adults: enduring or changeable?

Accepted: 7 September 1994

Abstract This study reports on stability and change in emotional and behavioral problems in young adults over a 2-year time span. A sample of 528 18- to 22-year-olds from the general population was assessed using the Young Adult Self-Report (YASR) on two occasions. Stability coefficients for the total problem score of the YASR were 0.63 for males and 0.75 for females. Forty-nine percent of the subjects who were initially classified as deviant were still deviant at follow-up. Of all YASR syndromes, the highest stability was for the Anxious/Depressed scale.

Introduction

This study investigated the 2-year course of psychopathology in young adults from the general population. Young adulthood is an important developmental period in which individuals are expected to take care of themselves more and more, making their own decisions and applying the social and occupational skills they acquired in adolescence. In a retrospective design, Bland et al. (1988) have found that most DSM-III (American Psychiatric Association 1980) psychiatric disorders, assessed with the Diagnostic Interview Schedule (Robins et al. 1981), have their first symptoms in young adulthood. Therefore, knowledge of the course of behavioral and emotional problems in young adults is needed to determine the need for intervention and prevention.

To assess the course of psychopathology, general population studies are needed because results of clinical studies cannot be generalized. First, clinical samples can be confounded by referral biases. Second, individuals with more than one disorder are more likely to be referred to mental health services than individuals with

only one disorder (Berkson 1946). Third, treatment may have a beneficial effect on the course of psychopathology in clinical samples. These factors may hamper the generalizability of clinical studies concerning the developmental course of psychopathology.

To our knowledge, one previous study has assessed the course of psychopathology in young adults from the general population. Vollrath and Angst (1989) found that of 88 21-year-olds who initially met DSM-III criteria for major depression, 44 (50%) were still depressed 2 or 7 years later, while 6 of 37 subjects (16%) who initially fulfilled criteria for panic attacks received a diagnosis of panic attacks at follow-up. However, these findings cannot be generalized because the sample followed up by Vollrath and Angst (1989) was not randomly drawn from the general population, but predominantly consisted of high scorers on the Symptom Checklist (Derogatis et al. 1973).

The present study assessed the stability and change in psychopathology over a 2-year period in 528 young adults, initially aged 18 to 22 years, from the general population. On two occasions, subjects were assessed with the Young Adult Self-Report (YASR; Achenbach 1990), a self-report questionnaire that covers a broad range of behavioral and emotional problems.

Methods**Instruments**

The YASR is a self-report questionnaire for ages 18–30 years. It was derived from the Child Behavior Checklist (CBCL), a parent questionnaire for 4- to 18-year-olds (Achenbach 1991a), and has roughly the same format, except that items are worded in the first person. Furthermore, items pertaining to childhood problems are replaced by items pertaining to adults' functioning.

The first part of the YASR comprises 14 competence items. The second part contains 110 problem items, covering emotional and behavioral problems during the previous 6 months, and 15 socially desirable items. The scoring format is 0 = not true, 1 = somewhat or sometimes true, and 2 = very true or often true. A total problem score is derived by summing the responses of each problem

F. C. Verhulst (✉) · R. F. Ferdinand
Sophia Children's Hospital/Erasmus University Rotterdam,
Department of Child and Adolescent Psychiatry,
Dr. Molewaterplein 60, NL-3015 GJ Rotterdam,
The Netherlands

item. The reliability and validity of the YASR have been reported elsewhere to be good (Achenbach et al. 1994, Ferdinand and Verhulst 1994, Ferdinand et al. 1994a, Witznitzer et al. 1992).

For adolescents aged 11–18 years, Achenbach (1991b) has constructed eight “cross-informant” syndromes that are similar for the parent, teacher (TRF; Teacher’s Report Form; Achenbach 1991c), and self-report (Youth Self-Report; YSR; Achenbach 1991d) versions of the CBCL. The eight cross-informant syndromes are: ‘Withdrawn,’ ‘Somatic Complaints,’ ‘Anxious/Depressed’ (together constituting the ‘Internalizing’ group of syndromes), ‘Delinquent Behavior,’ ‘Aggressive Behavior’ (together constituting the ‘Externalizing’ group of syndromes), ‘Social Problems,’ ‘Thought Problems,’ and ‘Attention Problems.’ The cross-informant syndromes, which were initially designed for younger ages, are applicable to YASR scores of subjects aged 18–25 years (Ferdinand and Verhulst 1994, Ferdinand et al. 1994a).

Population

The present study was part of an ongoing longitudinal study. The original sample consisted of children aged 4–16 years drawn in 1983 from the Dutch province of Zuid-Holland. This province encompasses both highly urbanized and rural areas. Using municipal birth registers that list all residents, we drew a random sample of 100 children of each sex and age of Dutch nationality ($n = 2600$). Two small municipalities out of a total of 86 declined to participate. Of the parents of the 2447 target children aged 4–16 years who were reached, the parents of 2076 children (84.8%) completed a CBCL on their child (Verhulst et al. 1985). The first assessment was designated time 1.

The present study stemmed from follow-ups of the original sample at 2-year intervals. Subjects who were aged 12–16 years at time 1 were asked to complete YASRs in 1989 (time 4), and 1991 (time 5). Subjects whose parents declined to participate at time 1 were not contacted.

At time 4, subjects aged 18–22 years received a letter explaining the study. They were then visited by an interviewer who asked them to complete the YASR (Ferdinand and Verhulst 1994). Twenty-three subjects were aged 23 when the interview was performed.

At time 5, subjects aged 20–24 years were reassessed with the YASR. After receiving the introductory letter, subjects were contacted by telephone to make an appointment (Ferdinand and Verhulst 1994).

At the first assessment in 1983, 765 parents completed the CBCL. In 1989, 36 subjects were not contacted because their parents had refused to cooperate at some point during the first 6 years of follow-up. The 528 young adults (240 males and 288 females) who completed YASRs at time 4 and time 5 formed 69.0% of the subjects on whom we obtained parent information at time 1.

To investigate selective attrition, we compared drop outs ($n = 237$) and remainers ($n = 528$) with respect to their 1983 CBCL total problem scores and their parents’ socioeconomic status (SES). The CBCL total problem score was computed in the same way as the YASR total problem score. The SES of the parents at initial assessment was assessed via a six-point scale of parental occupation (Van Westerlaak et al. 1975), with ‘1’ = lowest and ‘6’ = highest SES. Drop outs and remainers did not differ significantly in the initial CBCL total problem score (t -test; n.s.). However, the mean SES for drop outs (SES = 3.4) was slightly though significantly ($P = 0.01$) lower than that for remainers (SES = 3.6).

The finding that drop outs did not belong to a group of especially problematic individuals supported the representativeness of the sample in the present study. The lower SES for drop outs versus remainers indicated that remainers may have been subjected to somewhat more favorable environmental circumstances.

Results

Total problem scores

Changes in mean scores

Mean total problem scores, averaged across time 4 and time 5, were 26.51 (SD = 15.54) for males and 32.95 (SD = 19.27) for females. To assess the effects of sex, age, and time of assessment, problem scores were compared by multivariate analysis of variance (MANOVA) with a two sex (240 males versus 288 females) \times three age groups (18–19 years, $n = 209$; 20–21 years, $n = 202$; 22–23 years, $n = 117$, at time 4) between-subjects, and two times of assessment within-subjects factorial design.

Using a significance level of $P < 0.01$, we applied the following criteria suggested by Cohen (1988) for judging the effect sizes: effects accounting for 1.0–5.9% of variance are considered small, 6.0–13.8%, medium, and greater than 13.8%, large. The MANOVA revealed no effect for time of assessment. Across the two times of assessment, females scored significantly higher than males. This effect was small (4% of variance). No significant age effects or interactions were found.

Stability coefficients

Pearson correlations or stability coefficients for total problem scores for both sex groups and three age groups, as well as average stability coefficients, are presented in Table 1. We used Cohen’s (1988) criteria to judge the magnitude of stability coefficients: stabilities between 0.10 and 0.30 are considered small, between 0.30 and 0.50, medium, and greater than 0.50, large. All stability coefficients were large.

The stability of total problem scores was significantly higher for females ($r = 0.75$) than males ($r = 0.63$; $z = 2.62$; $P < 0.01$). Analyses for separate age groups revealed significant sex differences in stabilities for 18- to 19-year-olds ($z = 2.68$; $P < 0.01$) and 20- to 21-year olds ($z = 2.16$; $P < 0.05$), but not for 22- or 23-year-olds. No significant differences were found between the stabilities for different age groups.

Categorical continuity and change

To investigate individual changes in problem behavior across time, a categorical approach was applied. Across both times of assessment, we computed cumulative frequency distributions of the total problem score for each sex separately. Subjects scoring above the 90th percentile score (P90) of the frequency distribution were classified as deviant.

Because changes from above P90 to just below P90 can hardly be regarded as meaningful, we applied the 50th percentile (P50) of the cumulative frequency distribution as a cut off point below which individuals were

Table 1 Stability coefficients (*r*) between time 4 and time 5 total problem and syndrome scores by sex and age at time 4

	Males				Females			
	Average <i>n</i> = 240	Age 18–19 <i>n</i> = 91	Age 20–21 <i>n</i> = 92	Age 22–23 <i>n</i> = 57	Average <i>n</i> = 288	Age 18–19 <i>n</i> = 118	Age 20–21 <i>n</i> = 110	Age 22–23 <i>n</i> = 60
Withdrawn	0.61	0.61	0.63	0.57	0.65	0.59	0.70	0.62
Somatic Complaints	0.40 ^{a, b}	0.38	0.39	0.51	0.54 ^{a, b}	0.58	0.49	0.54
Anxious/Depressed	0.67	0.56	0.72	0.67	0.70	0.71	0.71	0.65
Social Problems	0.39	0.29 ^{a, b}	0.56	0.38*	0.51	0.54 ^{a, b}	0.56	0.32*
Thought Problems	0.30 ^a	0.35 ^a	0.24*, a, b	0.32 ^{a, b}	0.53 ^a	0.65 ^a	0.37 ^{a, b}	0.08 ^{a, b}
Attention Problems	0.61	0.63	0.63	0.54	0.64	0.60	0.70	0.55
Delinquent Behavior	0.48	0.57	0.42	0.42	0.49	0.53	0.48	0.40
Aggressive Behavior	0.58	0.57	0.54	0.73	0.66	0.62	0.67	0.71
Internalizing	0.64	0.59	0.67	0.67	0.71	0.73	0.73	0.64
Externalizing	0.57	0.60	0.49	0.67	0.66	0.62	0.67	0.67
Total problem score	0.63 ^a	0.63 ^a	0.60 ^a	0.69	0.75 ^a	0.77 ^a	0.76 ^a	0.68

All correlations significant at $P < 0.001$ except * $p < 0.01$ and n. s. = not significant

^a Significant sex difference

^b Not significant after correction for chance findings

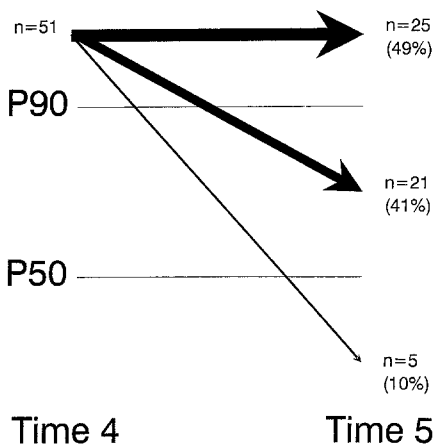


Fig. 1 The course of problem behaviors in individuals who scored above the 90th percentile (P90) at time 4

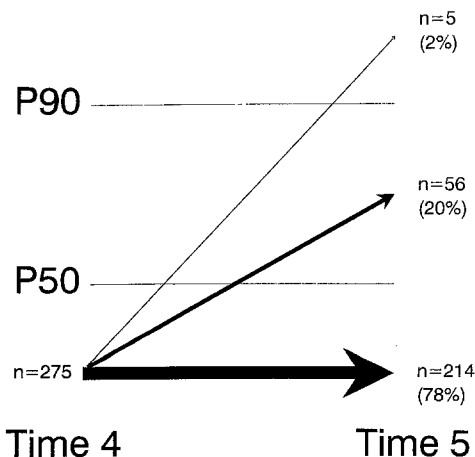


Fig. 2 The course of problem behaviors in individuals who scored below the 50th percentile (P50) at time 4

considered functioning well. The use of the 50th and the 90th percentiles enabled us to identify individuals whose functioning improved or worsened substantially across time.

Figure 1 shows the course of problems in 51 subjects (23 males, 28 females) who scored above P90 at time 4. It can be seen that 25 (49 %) still scored above P90 at time 5, while 21 (41 %) scored between P50 and P90, and 5 (10 %) scored below P50.

Figure 2 shows the pathways of the 275 subjects who scored below P50 at time 4. At time 5, 214 (78 %) still scored below P50, whereas 56 (20 %) scored between P50 and P90, and 5 (2 %) scored above P90.

Syndrome scores

Changes in mean scores

To assess the effects of sex, age, and time of assessment, scale scores were compared by multivariate analysis of variance (MANOVA) with a two sex (240 males versus 288 females) \times three age groups (18–19 years, 20–21 years, or 22–23 years at time 4) between-subjects, and two times of assessment within-subjects factorial design. Applying a $P < 0.01$ protection level (Sakoda et al. 1954), we controlled for chance findings by correcting for the number of comparisons. Effect sizes were judged according to Cohen’s criteria (1988).

Females scored significantly higher than males on the Withdrawn (3%), Somatic Complaints (16%), Anxious/Depressed (5%), Attention Problems (3%), and Internalizing (8%) scales, while males scored significantly higher on the Delinquent Behavior (2%) scale. The effects for the Withdrawn and Delinquent Behavior scales were most likely to be chance findings. There were no significant effects for age and time of assessment. No significant interactions were found.

Stability coefficients

Table 1 shows the stability coefficients for the YASR syndromes. Stabilities for separate ages, as well as the average stabilities for each entire sex group, are presented. All average stability coefficients could be regarded large, except stabilities for Somatic Complaints ($r = 0.40$), Social Problems ($r = 0.39$), and Thought Problems ($r = 0.30$) in males, and Delinquent Behavior ($r = 0.49$) in females, which were medium.

Following the method described by Steiger (1980) to determine differences between statistically dependent correlations, we did not find significant differences between the stabilities for Internalizing versus Externalizing scores, except for 20- to 21-year-old males. However, in accordance with Sakoda et al. (1954), this was regarded a chance finding.

Sex differences were determined for the entire sample and for separate age groups. After correction for chance findings (Sakoda et al. 1954), using a $P < 0.05$ protection level, we found a higher average stability coefficient for the Thought Problems scale for females than for males ($z = 3.19$, $P < 0.001$). This difference reflected a difference for the age group 18–19 years ($z = 2.89$, $P < 0.005$). After correction for chance findings, we found one age effect. In females, the stability of the Thought Problems scale was significantly higher for ages 18–19 versus 22–23 ($z = 4.29$, $P < 0.001$).

Discussion

Total problem scores

Mean total problem scores did not change significantly across the 2-year time span. Furthermore, mean total problem scores were not affected by age. This indicated that, across the entire sample, the extent of self-reported problems in young adults did not change with the subjects' increasing age.

Two-year stabilities for YASR total problem scores were high in males ($r = 0.63$) and females ($r = 0.75$), indicating that individual problem scores were considerably stable across time. Furthermore, the stability of total problem scores did not differ significantly between the three age groups, which indicated that the stability of problems in younger subjects was similar to that in older subjects.

Stability coefficients provide information on the stability of problems for the entire scoring range, and not on the stability of extreme levels of psychopathology. We therefore tracked individuals across time at a categorical level. Of subjects who were initially classified as deviant, 49 % were still deviant 2 years later, while only 10 % scored in the well-functioning range at follow-up. This indicated that extreme levels of behavioral and emotional problems tended to persist to a considerable degree.

Of the 275 subjects who scored below P50 at time 4, 78 % still scored below P50 2 years later, whereas only

2 % scored above P90 at time 5. In other words, the majority of the subjects who scored in the well-functioning range at initial assessment still functioned well 2 years later, whereas only a minority had moved to the deviant range.

Across both assessments, total problem scores were higher for females than males. Furthermore, the stability of total problem scores was significantly higher for females ($r = 0.75$) than males ($r = 0.63$). In other words, in comparison with males, females reported somewhat higher levels of problems and were also more likely to retain these problems across time.

Internalizing versus Externalizing scores

To determine whether the larger tendency of problems to persist in females than in males was associated with specific types of psychopathology, we compared the stabilities of the Internalizing and Externalizing scale scores. Internalizing problems reflect individuals' internal distress, while externalizing problems represent conflicts with other people and their expectations of the individual.

Mean Internalizing and Externalizing scores did not change significantly across the two assessments. Furthermore, no age effects were found. Hence, the level of Internalizing and Externalizing problems remained stable with increasing age. The stabilities of the Internalizing and Externalizing scales were high, and did not differ significantly after correction for chance findings. This indicated that both types of problems tended to persist to a similar degree.

While the stabilities of the Internalizing and Externalizing scales were similar for both sexes, females scored significantly higher than males on the Internalizing, but not on the Externalizing scale. This allows for two explanations. First, internalizing problems may occur more frequently in females than in males. This was supported by the finding that the rate of referral to mental health services between time 4 and time 5 was approximately 4 times higher in females than in males, while subjective need for help, without actually receiving mental health services, was reported approximately 3 times more often by females than males (Ferdinand and Verhulst 1994).

Second, females may show a greater propensity to report internalizing problems than males, while the actual degree of internalizing problems is similar for both sexes. However, the findings of Ferdinand et al. (1994b) in 131 young adults from the general population suggests that self-reported internalizing problems are related to clinical judgments of functional impairment in daily life to a similar degree in both sexes. In their study, no significant sex difference was found in the correlations (mean r across both sexes = 0.64) between the Internalizing YASR scale and the Global Assessment of Functioning scale (GAF scale; American Psychiatric Association 1987). Furthermore, no significant sex difference was found for mean GAF scores.

Syndrome scores

We did not find significant differences between mean time 4 and time 5 syndrome scores. Furthermore, no age effects were found. While mean group scores did not change across time, we found substantial stabilities for most YASR scales. This indicated that most types of problems tended to persist to a considerable degree.

The stability of the Thought Problems scale for 18- to 19-year-old females ($r = 0.65$) was significantly higher than that for males ($r = 0.35$). Further investigations are needed to determine whether this finding can be generalized, or was caused by a cohort effect.

After correction for chance findings, females scored higher than males on the following syndromes: Somatic Complaints, Anxious/Depressed, and Attention Problems. The only large difference between males and females was for Somatic Complaints (16%), which suggested that differences in the prevalence of psychopathology between males and females predominantly occurred in the area of physical complaints.

A developmental perspective

We compared our findings with those of Verhulst and Van Wattum (1993) who determined the 2-year stability of YSR total problem scores in adolescents aged 11–16 years at the first assessment. The total problem scores of the YSR and the YASR are similar measures of psychopathology (Ferdinand et al. 1994a). Verhulst and Van Wattum (1993) have found stabilities of 0.54 among males ($n = 275$) and 0.70 ($n = 305$) among females. These stabilities did not differ significantly from the stabilities of total problem scores for males (0.63) and females (0.75) in the present study (males: $z = 1.54$; females: $z = 1.28$).

Conclusions

In summary, the stability of self-reported behavioral and emotional problems in young adults across a 2-year time span was considerable. Almost half of the subjects who were classified as deviant initially still scored in the deviant range at follow-up. These findings indicated that problems in young adulthood do not tend to disappear spontaneously, which argues against a wait-and-see policy and for research on the efficacy of interventions.

Acknowledgements This study was financially supported by grants from the Dutch National Fund for Mental Health (National Fonds Geestelijke Volksgezondheid) and the Sophia Foundation for Medical Research. The authors are grateful to Marianne C. Kasius, for her comments on the manuscript.

References

- Achenbach TM (1990) The Young Adult Self-Report. University of Vermont, Department of Psychiatry, Burlington, Vt.
- Achenbach TM (1991a) Manual for the Child Behavior Checklist/4–18 and 1991 Profile. University of Vermont, Department of Psychiatry, Burlington, Vt.
- Achenbach TM (1991b) Integrative guide to the 1991 CBCL/4–18, YSR, and TRF Profiles. University of Vermont, Department of Psychiatry, Burlington, Vt.
- Achenbach TM (1991c) Manual for the Teacher's Report Form and 1991 Profile. University of Vermont, Department of Psychiatry, Burlington, Vt.
- Achenbach TM (1991d) Manual for the Youth Self-Report and 1991 Profile. University of Vermont, Department of Psychiatry, Burlington, Vt.
- Achenbach TM, Howell CT, McConaughy SH, Stanger C (1995) Six-year predictors of behavioral and emotional problems in a national sample: III. Transitions to early adult syndromes. *J Am Acad Child/Adolesc Psychiatr* (in press)
- American Psychiatric Association (1980) Diagnostic and statistical manual of mental disorders, 3rd edn. Washington, DC
- American Psychiatric Association (1987) Diagnostic and statistical manual of mental disorders, 3rd edn revised. Washington, DC
- Berkson J (1946) Limitations of the application of fourfold table analysis to hospital data. *Biometrics* 2: 47–53
- Bland RC, Newman SC, Orn H (1988) Age of onset of psychiatric disorders. *Acta Psychiatr Scand* 77: [Suppl 338]: 43–49
- Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale, NJ
- Derogatis LR, Lipman RS, Covi L (1973) The SCL-90: an outpatient psychiatric rating scale. *Psychopharmacol Bull* 9: 13–28
- Ferdinand RF, Verhulst FC (1994) The prediction of poor outcome in young adults: comparison of the Young Adult Self-Report, the General Health Questionnaire, and the Symptom Checklist. *Acta Psychiatr Scand* 89: 405–410
- Ferdinand RF, Verhulst FC, Wiznitzer M (1994a) Continuity and change of self-reported problem behaviors from adolescence into young adulthood. *J Am Acad Child Adolesc Psychiatry* (in press)
- Ferdinand RF, Reijden M van der, Verhulst FC (1994b) Assessment of the prevalence of psychiatric disorder in young adults: different approaches. *Br J Psychiatry* (in press)
- Robins LN, Helzer JE, Croughan J et al (1981) NIH Diagnostic Interview Schedule: Version III (May 1981). National Institute of Mental Health, Rockville, Md.
- Sakoda JM, Cohen BH, Beall G (1954) Test of significance for a series of statistical tests. *Psychol Bull* 51: 172–175
- Steiger JH (1980) Tests for comparing elements of a correlation matrix. *Psychol Bull* 87: 245–251
- Verhulst FC, Wattum PJ van (1993) Two-year stability of self-reported problems in an epidemiological sample of adolescents. *Acta Psychiatr Scand* 87: 322–328
- Verhulst FC, Akkerhuis GW, Althaus M (1985) Mental health in Dutch children: I. A cross-cultural comparison. *Acta Psychiatr Scand* [Suppl] 323: 72
- Vollrath M, Angst J (1989) Outcome of panic and depression in a seven-year follow-up: results of the Zurich study. *Acta Psychiatr Scand* 80: 591–596
- Westerlaak JH Van, Kropman JA, Collaris JWM (1975) Beroepenklapper. Instituut voor toegepaste sociologie (Manual for occupational level. Institute for Applied Sociology) Nijmegen, The Netherlands
- Wiznitzer M, Verhulst FC, Van den Brink W, Koeter M, Ende J van de, Giel R, Koot HM (1992) Detecting psychopathology in young adults. A comparison of the Young Adult Self-Report, the General Health Questionnaire and the Symptom Checklist as screening instruments. *Acta Psychiatr Scand* 86: 32–37