The Norwegian Institute of Public Health Twin Study of Mental Health: Examining Recruitment and Attrition Bias

Kristian Tambs,[1,2] Torbjørn Rønning,[3] C. A. Prescott,[4] Kenneth S. Kendler,[2] Ted Reichborn-Kjennerud,[1,5] Svenn Torgersen,[6] and Jennifer R. Harris[3]

[1] *The Norwegian Institute of Public Health, Division of Mental Health, Oslo, Norway*
[2] *Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, United States of America*
[3] *The Norwegian Institute of Public Health, Division of Epidemiology, Oslo, Norway*
[4] *Department of Psychology, University of Southern California, United States of America*
[5] *Institute of Psychiatry, University of Oslo, Norway*
[6] *Department of Psychology, University of Oslo, Norway*

All Norwegian twin pairs born 1967–1974 and still living in Norway in 1992 were invited to a health questionnaire study (Q1). 2,570 pairs (65%) participated. These cohorts and the twin cohorts born 1967–1979 were invited to a new questionnaire study (Q2) in 1998. This time 3,334 pairs (53%) participated. Almost all pairs having participated in the 1998 study were invited to an interview study of mental health (MHS), taking place 1999–2004. 1,391 complete pairs (44%) participated. The questionnaire studies included extensive data on somatic health with fewer items on mental health and demography. Health-related and demographic information available from the Medical Birth Registry on all invited twins was applied to predict participation to the first study. A few registry variables indicating poor health predicted nonparticipation in Q1. Health information and demography from Q1 were tested as predictors of participation in the follow-up study (Q2). Monozygosity, female sex, being unmarried, having no children, and high education predicted participation, whereas few indicators of poor mental and somatic health and unhealthy lifestyle moderately predicted nonparticipation in Q2. No health indicators reported in Q2 predicted further participation. Standard genetic twin analyses of indicators of various mental disorders from Q2, validated by diagnostic data from the MHS, did not indicate differences in genetic/environmental covariance structures between participants and nonparticipants in MHS. In general the results show a moderate selection towards good mental and somatic health. Attrition from Q2 to the MHS does not appear to affect twin analyses of mental health related variables.

Keywords: recruitment bias, attrition, selection effect, mental health, twin studies

Bias resulting from nonresponse in epidemiological research and attrition in longitudinal studies may seriously limit generalizability or lead to false conclusions. Particular issues arise in twin studies due to the reliance on pair participation for deriving estimates of genetic and environmental variance components. Systematic selection towards higher or lower phenotypic values may occur at an *individual* level when the probability of participation is associated with particular trait values. For example, a personality trait such as conscientiousness may increase the probability that an individual participates independent of co-twin participation. Selection for factors that may correlate with the phenotypes under study, such as social background, or directly for the phenotypes, will shrink the phenotypic variance and can bias the estimates from twin studies (Martin & Wilson, 1982). Even more serious bias could result from pair-wise selection, in which the co-twin correlations systematically differ between the sample and the study population. It cannot be ruled out that twin pairs either more phenotypically different or more similar than usual are over-represented in twin studies. As argued by Lykken et al. (1988), selection towards co-twin similarity is perhaps more likely than selection toward differences. Some twins may be more preoccupied by being twins than others, perhaps because the co-twins are similar in many respects and therefore feel close, or perhaps because they feel close and therefore become similar. Although scarce, some evidence supports this expectation. Twins who heard about a twin study and who eagerly volunteered to participate without being invited, were more similar, within pairs, on personality, interests, and demography than were invited participants (Lykken et al., 1990).

Received 18 September, 2008; accepted 29 January, 2009.

Address for correspondence: Kristian Tambs, Division of Mental Health, The Norwegian Institute of Public Health, Post Box 4404 Nydalen, N-0403 Oslo, Norway. E-mail: kristian.tambs@fhi.no

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

It is well known that participation rates are greater among MZ than among DZ pairs (Lykken et al., 1978). As previously suggested, identical twins seem to be more invested in their identity as twins and therefore more willing to take part in twin studies (Kendler & Prescott, 2006). This might reflect a general tendency for co-twins who have more in common and are more similar to be more willing to participate compared to co-twins who have less in common. Since DZ twins vary more in phenotypic co-twin similarity than do MZ twins, such a selection effect directly driven by co-twin similarity might well affect data from DZ pairs more than MZ data. A nominally similar inflation of both MZ and DZ similarity will primarily inflate the estimate of common environment at the expense of the estimate of non-shared environment. A nominally stronger inflation of DZ than MZ correlations would lead to underestimated genetic effects and overestimated common environment, whereas the consequences will be reversed when there is a stronger inflation of the MZ correlation. Most models suggested to account for recruitment bias in twin studies, although highly sophisticated, have only specified selection on an individual level (e.g., Dominicus et al., 2006; Kendler and Holm, 1985; Martin & Wilson, 1982, Neale et al., 1989; Taylor, 2004), although at least one study modeled correlated ascertainment rates (Kendler and Eaves, 1989). Also, contrary to the variety of studies on modeling of selection effects and data simulation, few studies have investigated recruitment and attrition bias in real data materials.

This study uses data from a population based sample of Norwegian twins to explore factors influencing attrition and predicting participation in several phases of a longitudinal program of research that includes registry data collected at birth, general health questionnaires in adulthood, and an interview study of mental health and personality disorders. A wide array of measures, focusing on health in general and mental health in particular, were employed for predicting participation. In addition to examining sample selection as such, we tested for systematic differences between future participants and non-participants in the relative size of genetic and environmental variance components for mental health related variables. For these purposes we used birth registry data collected at the twins' birth and data from a questionnaire study, a follow-up questionnaire study, and a follow-up interview study. This article also serves as a general presentation of the Norwegian Institute of Public Health (NIPH) Mental Health Study (MHS).

Materials and Methods

Sample

The twins in the MHS are participants in a population-based program of twin research at the NIPH (Harris et al., 2002; 2006). All twins were identified through the national Medical Birth Registry of Norway (MBRN) which began with mandatory registration of all pregnancies from 16 weeks gestation in 1967. Standardized information is recorded for all births by attending midwives and physicians (Irgens et al., 2000). The NIPH twin program of research is based on cohorts born from 1967 through 1979. During this period there were 15,374 twin births in Norway, and the percentage of pairs for which both twins survived to age three ranged from 82% to 89% (Harris et al., 2002). Twins born from 1967 through 1974 who were at least 18 years of age were contacted in 1992 via a mail-out questionnaire (Q1). These twins were re-contacted for longitudinal follow-up using a second questionnaire (Q2) in 1998. At that time younger cohorts, born 1975 to 1979, were also recruited into the twin study and administered the Q2 questionnaire. Altogether, 5,864 twins (75% of those eligible) including 2,570 pairs (65%) responded to Q1, and 8,045 twins (63%) including 3,334 pairs (53%) responded to Q2. The longitudinal sample responding to both Q1 and Q2 included 4,430 twins and 1,725 pairs.

The interview study of mental disorders took place from 1999 to 2004. The majority of twins were examined at the Norwegian Institute of Public Health in Oslo, with the remainder examined at home or in localities close to their residence. All complete pairs from the Q2 study in which both twins had accepted to be contacted again for new studies (n = 3,153 pairs) were invited by mail. Due to technical problems an additional 68 pairs were accidentally drawn from twin pairs that had not completed Q2. Twins who did not respond were reminded once. Only pairs in which both twins were initially willing to participate were interviewed, but 19 single twins were lost for practical reason or they changed their minds after initially having accepted the invitation. Altogether 1,391 complete pairs (43.2%) and 19 single twins (0.6% pairwise), in total 2,801 twins aged 19-36 years were interviewed. In 531 pairs (16.5%) one twin agreed to participate (including some pairs where both twins initially agreed, but one re-decided before any of the twins were interviewed). In 1,280 pairs (39.7%) none of the twins agreed to participate. 181 pairs (5.6%) were not traceable due to wrongly registered address(es). The real loss due to wrong addresses is higher since not all non-received invitation letters were returned to the sender. In 0.8% of the pairs one or both twins actively refused or gave a reason for not participating. In the remaining pairs who did not participate none of the twins responded. Informed consent was obtained from all participants after complete description of the study. A flow chart of the NIPH questionnaire and mental health twin studies is shown in Figure 1.

Zygosity classification was initially based on questionnaire data using discriminant analyses. Blood samples were obtained for the majority of the interview sample, and zygosity of 676 like-sexed pairs was decided using 24 microsatellite markers. The results indicated a 97.5% correct original classification from

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

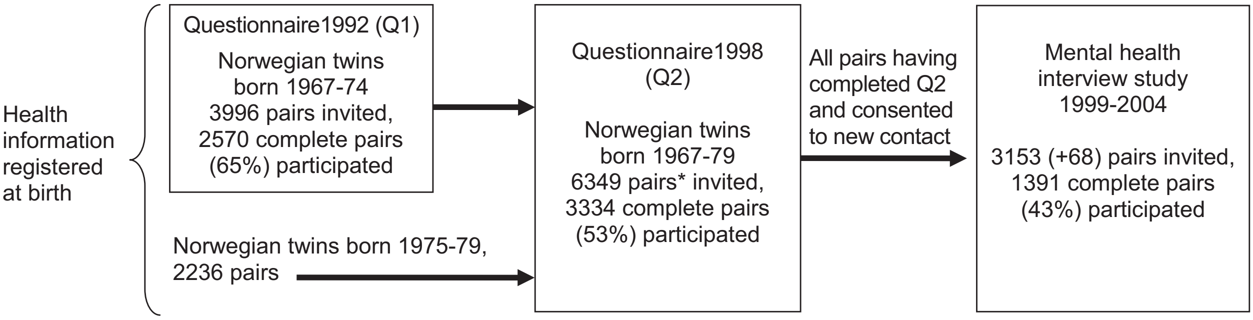
Health information registered at birth
Questionnaire1992 (Q1)
Norwegian twins born 1967-74
3996 pairs invited, 2570 complete pairs (65%) participated
Norwegian twins born 1975-79, 2236 pairs
Questionnaire1998 (Q2)
Norwegian twins born 1967-79
6349 pairs* invited, 3334 complete pairs (53%) participated
All pairs having completed Q2 and consented to new contact
Mental health interview study 1999-2004
3153 (+68) pairs invited, 1391 complete pairs (43%) participated

Figure 1

The Norwegian Institute of Public Health questionnaire and interview studies.

Note: * Approximately 100 more 1967–1974 pairs were invited to Q2 than to Q1 due to addresses unknown in 1992 but known in 1998.

questionnaire items alone (Harris et al., 2006). Correcting the originally misclassified pairs with DNA-based zygosity information, correct classification was estimated to be 98.0% in the Q2 sample and 99.1% in the interview sample (in which most of the classifications were based on DNA data). Further details about the Q1 and Q2 study design, recruitment procedures, zygosity classifications, sample sizes by birth cohort and questionnaire items are available elsewhere (Harris et al., 1995; 2002; 2006).

The project was approved by the Data Inspectorate and the Regional Committee for Medical Research Ethics of the Norwegian government. Informed consent was obtained after participants received a complete description of the study.

Measures

Measures From the Medical Birth Registry of Norway (MBRN)

Measures from the MBRN were used to predict participation after the first invitation to a questionnaire study (Q1 for cohorts born 1967–1974 and Q2 for cohorts born 1975–1979). The MBRN includes information on birth order, demographics about the parents, ICD-8 coded measures of maternal health before and during pregnancy, circumstances, interventions and complications related to the delivery, status of the newborn, gestational age, and birth length and weight. For the analyses conducted here, the most prevalent maternal diagnoses were identified and six dichotomous measures were created indicating a positive diagnosis for any of the following: preeclampsia, eclampsia, bleeding during pregnancy, kidney infection, other urinary tract infections and hyperemisis. Other information from the MBRN was used to code for interventions during delivery, including use of narcosis, morphine-like drugs, inhalation analgesics, epidural or spinal anesthesia, amniotomy, forceps, vacuum extraction and placenta extraction. Information on complications during labor included placenta previa, placenta abruption, dystocy, threatening asphyxia, obstructions and bleeding after birth. Information obtained from the MBRN about the newborn included sex, asphyxia, birthweight and gestational age. The latter two measures were dichotomized into two groups representing the highest 75% or lowest 25% of the respective variable distributions because only variation in the lower end was expected to be associated with later health outcomes.

Questionnaire Measures

With the exceptions described below, most of the measures derived from Q1 are identical to those derived from Q2. Demographic and pair measures included marital status (including nonmarried partnerships), number of children, and highest completed education. Contact with the co-twin is a composite measure based upon items inquiring about current and previous frequency, type, and quality of co-twin contact, described elsewhere (Tambs et al., 1995).

Physical health measures. *Body mass index* (BMI) was calculated from self-reports using the standard formula. Two other physical health measures, one from Q1 and one from Q2, were based on a health-history list (21 items in Q1 and 31 items in Q2) inquiring ‘Have you now, or have you ever had, any of the following illnesses or health problems?’ For analysis, the items were then grouped into dichotomously-coded categories reflecting diseases or health problems (six categories from Q1 and seven from Q2) indicating a positive health history of the following: atopic diseases, autoimmune diseases, neurological diseases, musculoskeletal pain, stomach/intestinal illness, infections, and hyperactivity (Q2 only). Atopic diseases comprised six items: hay fever, nettle rash, asthma, nickel allergy, atopic eczema, and childhood eczema. Autoimmune disease was based only on psoriasis in Q1, whereas in Q2 it also included Crohn’s disease, ankylosing spondylitis, and arthritis. Neurological diseases included migraine and epilepsy. Musculoskeletal pain was measured by: low back pain, neck and shoulder pain, headache, muscular pain, and (in Q2 only) fibromyalgia. Stomach/intestinal illness included two items: diarrhea/constipation/painful abdominal distension and intestinal ache. Infection refers mainly to upper respiratory tract infections associated with ear infections, tonsillitis, and sinusitis, but in Q2 an item inquiring about repeated bladder infections is also included. Hyperactivity was

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

measured with two items in Q2, hyperactivity and minimal brain dysfunction (MBD). *Self-rated health* was measured with a single item inquiring 'What is your health like, at present?' (coded on a scale rated from 1 = *poor* to 4 = *very good*). This single item measure of perceived health, and highly similar versions, have been widely used and it has been shown to have acceptable psychometric properties (e.g., Bardage et al., 2001; Krause & Jay, 1994). *Reading and writing problems* were assessed with two parallel questions asking 'do you have, or have you had, difficulties with 'reading' or 'writing', with three response categories for each item (major problems, minor problems, no problems).

Health related behaviors included alcohol (ab)use, smoking and exercise habits. Alcohol (ab)use was scored as an additive index based on items measuring the frequency of alcohol use in the last 14 days (five response categories from *never* to *more than 10*), whether the respondent had felt drunk within the last 14 days (*yes* or *no*), frequency of consumption of five or more alcohol units (six categories from *0–4 per year* to *six or more per week*), and whether there had been periods in life with too much drinking (*yes*, *perhaps*, *no*). The exercise scale was created from three items inquiring about exercise frequency per week (four response categories), intensity of exercise (three response categories) and the duration of exercise sessions (four response categories). These items were standardized and summed. Smoking status, coded as a smoker or nonsmoker, was based on two items related to current daily smoking and the number of cigarettes smoked per day.

Mental health and psychological wellbeing. Symptoms of anxiety and depression were measured using a five-item version (SCL-5) of the Hopkins Symptom Check List (SCL-25) (Hesbacher et al., 1980). The short-version correlates highl (r = .92) with the original instrument (Tambs & Moum, 1993). *Subjective well-being* (SWB) was assessed with seven items used and described elsewhere (Moum et al., 1990): life satisfaction, nervousness, fatigue, loneliness, sleeping problems, sadness, and use of tranquilizers. A SWB composite score was computed as the sum of the seven standardized items. The Q2 contained a few items related to phobias, panic disorder, obsessive–compulsive disorder and major depression. Endorsing certain combinations of symptoms, currently or ever, was scored as a positive response. A selection of those of the items used for the present purpose is described later. Finally the questionnaire included the Dysfunctional Personality Questionnaire, DPQ (Torgersen, 1980), consisting of 91 items selected to measure symptoms of personality disorders.

Interview

The interviews were carried out between June 1999 and May 2004. The assessments comprise a lifetime history of psychiatric disorders, including substance abuse (Axis I), and personality disorders (Axis II) as diagnosed by the DSM-IV. The interview typically lasted two hours and participants received compensation corresponding to $40 plus travel expenses to the interview. The interviews were usually administered face-to-face, but 231 interviews (8.3%) that for practical reasons could not be conducted face-to-face, were obtained by telephone. The interview consisted of two parts. In the first we used the Norwegian version of the computerized Munich-Composite International Diagnostic Interview (M-CIDI) (Wittchen & Pfister, 1997). This is a comprehensive structured diagnostic interview for the assessment of all the DSM-IV axis I and ICD-10 life time diagnoses, originally developed in 1993 by the World Health Organization in conjunction with the former United States Alcohol, Drug Abuse and Mental Health administration and later updated (WHO, 1997). The other part was the Norwegian version of the Structured Interview for DSM-IV Personality (SIDP-IV) (Pfohl et al., 1995). The SIDP-IV assesses all DSM-IV Axis II personality disorders including those listed in the DSM-appendix, and — not included in DSM-IV, but described in the DSM-IIIR appendix — 'self-defeating disorder'. The DSM-IV criteria are scored on a 4-point scale spanning *absent, sub-threshold, present* and *strongly present.*

Most interviewers were psychology graduate students completing the final part of their training or experienced psychiatric nurses. Interviewers were trained for the SIDP-IV by one psychiatrist and two psychologists with extensive previous experience with the instrument. They also received a standardized training program by teachers certified by the WHO and passed a user license test for the computerized CIDI. They were supervised during the data collection period.

Each twin in a pair was interviewed by different interviewers blind to the information obtained from the co-twin. To assess interrater reliability for the SIDP a subset of 70 audiotaped interviews were re-scored by two raters. The size of the subsample required a dimensional scoring of the personality disorders, counting numbers of endorsed full and sub-threshold criteria for each individual. The interrater polychoric correlations ranged from .86 to .94 for the cluster A disorders (Kendler et al., 2007), from .80 to .94 for cluster B disorders (Torgersen et al., 2008), from .87 to .97 for cluster C disorders (Reichborn-Kjennerud et al., 2007) and from .95 to .97 for appendix diagnoses (Czajkowski et al., 2008; Ørstavik et al., 2007).

Q2-based Indicators of Mental Health Diagnoses Validated by Interview Data

Items from the Q2 on symptoms of various anxiety disorders and major depression were used to generate questionnaire-based indices of the DSM-IV diagnoses. The indices were used to test for differences between Q2 responders who also participated in the mental health interview study versus those who did not. The Q2-based indices were validated against the interview-based DSM-IV diagnoses using scores from

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

twins who participated in both Q2 and the interview study. Those combinations of questionnaire items that maximally correlated with the diagnoses were chosen as indicators.

The interview derived DSM-IV diagnoses for *agoraphobia* and *social phobia* were highly intercorrelated (tetrachoric correlation = .66), and were, for the purpose of the validation of the questionnaire items, collapsed. A corresponding Q2 based index of the two disorders was scored positive if two of the following three items were endorsed: 'I feel a strong aversion when together with many people at a time, e.g., in stores, on the street, or in the movie theater', 'I often feel afraid when I travel alone by bus, streetcar or train', or 'I often feel a strong aversion when I go out to eat or drink with people, e.g., in a cafeteria, a café, or a restaurant'. The tetrachoric correlation (and asymptotic standard error) between the diagnosis and the questionnaire based indicator was .66 (.04). A single Q2 item, 'I'm very afraid of certain things, such as animals, heights, deep water, blood, or flying' correlated .50 (.03) with the diagnosis *specific phobia*. Another single item, 'I can suddenly become very afraid or panic without a reason', correlated .64 (.04) with DSM-IV *panic disorder* (including panic reaction). The questionnaire indicator of *obsessive–compulsive disorder,* based on positive endorsements of both 'I check and control everything much too often, for instance electric burners and locked doors' and 'I am often bothered by 'stupid' thoughts which keep coming back', correlated .45 (.09) with DSM-IV obsessive–compulsive disorder. One item about whether 'There have been periods in my life when I felt depressed, had sleeping problems and trouble concentrating' correlated .46 (.03) with DSM-IV *major depression*. The sum of the four alcohol (ab)use items correlated .56 (.03) (polychoric correlation) with an alcohol measure scored '1' for DSM-*IV alcohol abuse*, '2' for *alcohol dependence*, and '0' if unaffected.

SCL-5, the short-form version of SCL-25 which taps symptoms of anxiety and depression, is usually applied as a global indicator of *psychological distress*. A dichotomized version of the SCL-5 scores from Q2 was compared with the axis-I anxiety disorders listed above and with generalized anxiety disorder and with major depression. The tetrachoric correlations ranged from .32 to .52.

The 91 DPQ items in Q2 were used to generate indicators of 13 personality disorders, scored as continuous measures. In multiple linear regression analyses the questionnaire items were entered as predictors and each of the interview derived dimensionally scored personality disorders as dependent variables. The results were used to generate weighted sum-scores from the DPQ items which correlate maximally with the interview based scores. The Q2-based (continuous, but for the most part non-normally distributed) scores were recoded into ordinal variables with four categories (lower 50%, next 25%, next 15% and higher 10%). The polychoric correlations (and asymptotic standard errors) between the interview-based personality disorder (PD) scores and Q2-based PD indicators were, for *paranoid* .40 (.02), *schizoid* .41 (.02), *schizotypal* .42 (.02), *antisocial* .48 (.02), *histrionic* .46 (.02), *narcissistic* .39 (.02), *borderline* .54 (.02), *avoidant* .61 (.01), *dependent* .50 (.02), *obsessive–compulsive* .39 (.02), *self-defeating* .45 (.02), *passive–aggressive* .43 (.02), *depressive* .60 (.02), *cluster A* .42 (.02), *cluster B* .54 (.02), *cluster C* .56 (.01), and *any PD* .58 (.01). The interval between the Q2 and the interview ranged from a few months to 4 years. The observed correlations between interview- and Q2-based measures are therefore expected to be attenuated by changes in mental health during the period between the questionnaire and interview. Personality disorders are assumed to be relatively stable throughout life. Still, for *any PD* the correlation was .64 (.03) for twins interviewed within two years after the questionnaire study (*N* = 1282) as compared to .52 (.03) for twins interviewed more than two years after (*N* = 1471).

Analyses

To test for possible recruitment bias in the Q1 pairs, we compared some data obtained from single responders and pair-wise responders. We assume that mean and prevalence differences between participants and nonparticipants would produce differences between single and pair-wise responders as well, because the data from the single responders are likely to be highly correlated with the data from their non-responding co-twins (Neale & Eaves, 1993). For these comparisons we used the SCL-5 (as a continuous variable), the somatic index based on 21 Q1 disease items, and the indicator of co-twin contact.

Three series of logistic regression analyses were conducted to explore possible predictors of each stage of future participation. These analyses were performed using generalized estimating equations (GEE) (Liang & Zeger, 1986) in SAS. The GEE methodology adjusts for the statistical dependence between the co-twin data, which somewhat reduces the statistical power of the sample. Multiple imputation methods (Rubin 1976; 1987) were used to handle missing data for the multivariate logistic regression utilizing SAS 9.1.3 Service Pack 4. PROC MI was used for the imputation, generating five complete datasets. Each of these data sets was then analyzed with PROC GENMOD, and the results were combined using PROC MIANALYZE to generate valid statistical inferences. For practical reasons, we used SPSS for generating Q2-based indices of the DSM-IV diagnoses. Since multiple imputation is not available in SPSS, another method for treating missing values was used here. The SCL-5 items together with seven well-being items, all which correlate highly with SCL-5, were used for imputation of the SCL items in cases where five or less of the eleven items were missing. The imputation was conducted with SPSS MVA, EM estimation. The 91 PD items from Q2 were imputed with the same method,

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

using all the items as predictors, in cases with valid data for more than 75% of the items. The interview data were very close to complete and imputation was not required.

The first set of GEE regression analyses explored MBRN predictors of participation in the questionnaire study. The dependent variable was participation (yes/no) in the first questionnaire study to which the twins were invited (Q1 for the older cohorts and Q2 for the younger cohorts). The independent measures from the MBRN were sex, maternal age, and variables related to the mothers' and children's health. To test the effect of the MBRN data, each measure was tested separately in a regression model that adjusted only for age, sex, and whether invited to Q1 (scored 0, only invited to Q2 scored 1). A full model was then tested that included as predictors all the MBRN measures together with age, sex, and first invitation to Q1 or Q2.

The second set of GEE regression analyses investigated Q1 predictors of Q2 participation among the sub-sample of twins who were born before 1975 and thus eligible for both Q1 and Q2. The analyses were conducted similarly to those described above. In consecutive analyses Q2 participation was regressed on each of the measures while adjusting for background variables only (number of children, education, civil status, sex, age, and zygosity). Next, all the Q1 explanatory measures were entered simultaneously.

The final set of GEE regression analyses tested for Q2 predictors of participation in the MHS. These analyses were based on items measuring demographic factors, co-twin contact, lifestyle behaviors, physical health, mental health, and subjective well-being. As before, each predictor was entered one at the time, only adjusting for the background variables. Next, all the Q2 predictors were entered together with cohort (invited to Q1 and Q2 or only invited to Q2).

Finally, to explore the effect of attrition on the genetic and environmental variance structures, the variables from Q2 indicating, and validated against, mental health diagnoses were analyzed using standard univariate twin analyses with the raw data option in Mx (Neale et al., 1999). The pattern of twin correlations (data not shown) did not suggest the presence of dominance effects, and full ACE models with no sex-specific effects were fit to the Q2 data. The samples of MHS-participants included 220 MZM, 116 DZM, 440 MZF, 263 DZF, and 334 DZU pairs. The numbers of pairs of non-participants were 285 MZM, 243 DZM, 331 MZF, 357 DZF, and 569 DZU. Heterogeneity between the covariance structures in the two sets of data (MHS participants and non-participants) was analyzed by testing the difference in fit between nested models. The unconstrained model specified separate genetic and environmental parameters in the two groups and the constrained model specified these parameters to be equal across groups, yielding a chi-square difference test with 3 degrees of freedom. Due to sex differences in prevalence rates, thresholds were modeled separately for men and women but were constrained to be equal across zygosity and twin1-twin2. The following Q2 measures (described above) were analyzed: 1) alcohol (ab)use, 2) six indicators of axis-I disorders (agoraphobia/social phobia, specific phobia, major depression, panic disorder, obsessive-compulsive disorder, and SCL-5 anxiety/depression scores) as well as subjective well-being, 3) thirteen indicators of personality disorders, including two appendix diagnoses and one, self-defeating, not included in DSM-IV and, in addition, the aggregated variables 'any Cluster A', 'any Cluster B', any 'Cluster C', and 'any personality disorder'.

Results

Testing for Mean Differences

Single male Q1-responders scored 0.26 *SD* higher on SCL-5 than did male pair-wise responders ($t = 4.47$, $p < .001$, comparing single responses with pair means). There were no significant SCL-5 difference in females. Single male responders also scored 0.17 *SD* higher than did pairs on the 21 item summative somatic illness index ($t = 2.60$, $p = .01$), whereas a nonsignificant trend in females was in the opposite direction. There was no tendency for co-twin contact to differ between singles and pairs among either males or females ($p > .47$).

Predicting Participation by Previously Ascertained Health Information

Sex predicted participation. The proportion of men was 50.1% among the eligible twins and 45.3% among participants in the questionnaire studies. Overall participation rates were greater among first time invitations to a questionnaire study for the 2.5-page Q1 (73.4% response rate among first time invitees, twins born 1967–1974) than for the 8-page Q2 (58.4% response rate among first time invitees, twins born 1975-79).

Analyses of MBRN predictors of participation after first invitation (Q1 or Q2) revealed six significant 'crude effects' (only adjusting for age, sex, and which questionnaire was received at first invitation) shown in Table 1. Three of these measures (placenta previa, asphyxia and low birth weight) are considered health risk factors. Placenta previa (OR = 2.76, $p = .044$) predicted participation, whereas asphyxia (OR = .91, $p = .036$) and low birth weight (OR = .90, $p = .013$) predicted non-participation. Having a mother not being married and having older siblings (high parity) predicted non-participation. Results from the full model containing all the MBRN predictors revealed that four of the MBRN variables with significant 'crude' effects remained significant, the adjusted results also revealed reduced participation among males and among those receiving Q2 as their first questionnaire (Table 1).

Results from analyses testing Q1 predictors of participation in Q2 are shown in Table 2. The 'crude' results, adjusted for age, sex and zygosity only,

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

Table 1

Predictors from the Medical Birth Registry of Norway of Questionnaire Participation (Q1 and/or Q2)

Measure	Crude[A] OR[B]	(95% CI)	*p*	Adjusted OR[B]	(95% CI)	*p*
Male	—			0.64	(0.59, 0.69)	< .001
Invited to Q2, not to Q1	—			0.50	(0.45, 0.53)	< .001
Mother not married at birth of twins	0.66	(0.55, 0.79)	< .001	0.63	(0.52, 0.76)	< .001
Parity, OR per previous birth	0.95	(0.92, 0.98)	0.004	0.93	(0.89, 0.96)	< .001
Placenta previa	2.76	(1.03, 7.41)	0.044	4.28	(1.28, 14.36)	.018
Inhalation of analgetica	1.50	(1.06, 2.13)	.023			
Asphyxia	0.91	(0.83, 0.99)	.036			
Low birth weight (lower 25%)	0.90	(0.83, 0.98)	0.013	0.89	(0.81, 0.98)	.019

Note: [A] Adjusted by age, sex, and whether Q2 was the first invitation to a questionnaire study.

[B] Adjusted by age, sex, Q2 as first invitation, and all the remaining MBRN variables. Values greater than 1 indicate high participation.

Predictors not showing significant effects:

1. *Maternal health during pregnancy:* preeclampsia, eclampsia, bleeding during pregnancy, kidney infection, other urinary tract infection, hyperemesis.
2. *Circumstances related to labor and delivery:* breech presentation, early membrane rupture, ceasarian section, narcosis, morphine, epidural or spinal anestesia, amniotomy, forceps, vacuum extraction, placenta extraction, placenta abruption, long birth/ineffective labour, slow/poor development in birth, pelvis augusta and other obstructions, bleeding after birth.
3. *Status of the newborn:* Born after co-twin.

Table 2

Q1 Predictors of Participation in Q2

Measure	Crude[A] OR	(95% CI)	*p*	Adjusted OR	(95% CI)	*p*
Male	—			0.55	(0.48, 0.64)	< .001
Age, ratio per year	—			1.05	(1.01, 1.09)	.013
Dizygote	—			0.78	(0.66, 0.92)	.002
Married	0.82	(0.71, 0.94)	.005			
Having children	0.65	(0.54, 0.78)	< .001			
Education[B]	1.22	(1.17, 1.28)	< .001	1.18	(1.12, 1.23)	< .001
Autoimmune diseases	0.74	(0.54, 1.00)	.049			
Stomach/intestine illness				1.39	(1.07, 1.81)	.015
Low self-rated health[C]	0.90	(0.85, 0.96)	.001			
Alcohol consumption[C]	0.91	(0.85, 0.96)	.001			
Smoking	0.73	(0.64, 0.83)	< .001			
Exercise[C]	0.92	(0.86, 0.98)	.010			
Low subjective wellbeing[C]	0.84	(0.79, 0.89)	<.0001	0.91	(0.85, 0.98)	.019
Symptoms of anxiety and depression (SCL-5)[C]	0.87	(0.82, 0.92)	< .001			
Writing/reading problems	0.79	(0.69, 0.90)	.001			

Note: [A] Adjusted by age, sex, and zygosity

[B] Seven levels from public school to ≥ 4 years college

[C] z-transformed.

Predictors not showing significant effects: Contact with co-twin, BMI, atopic diseases, neurological diseases, musculoskeletal pain, infections.

revealed that being married and having children both predicted nonparticipation, whereas higher education predicted participation. The health related variables, including wellbeing, were all scaled in the direction of high scores reflecting poor health, and all were associated with nonparticipation. There were no significant effects of BMI, atopic diseases, neurological diseases, musculoskeletal pain or infections. Also there was no effect of co-twin contact, a variable of potential importance for genetic analyses of twin data. The significant ‘crude‘ effects of the lifestyle and health measures were moderate, with odds ratios typically close to 0.9 per *SD* change. Results from the full regression model indicated that participation was predicted by older age, female sex, higher education, monozygosity, high wellbeing, and suffering from stomach/intestine illness.

The selection effect of education, OR = 1.22 per unit in a 7-unit scale, adjusted by age, sex, and zygosity, is strong. To generate a supplementary and more easily interpretable expression of this effect size the participation rates for each education group were compared. Results showed that participation increased monotonically from 64% in the least educated group to 89% in the highest educated group.

The third set of analyses predicted participation in the MHS from Q2 measures. Among the 45 predictors, including 22 indicators of mental health, the ‘crude’ and multivariate analyses revealed that only old age and monozygosity predicted participation. Odds ratios from the multivariate analyses were 1.04 per year (95% CI: 1.00–1.08, *p* = .036) for age and 0.56 (95% CI: 0.49-0.64, *p* < .001) for dizygosity.

Testing for Differences in Heritability Analyses Between Participants and Nonparticipants

Analyses exploring whether the genetic and environmental variance structures differed between the MHS participants and non-participants were conducted for 25 variables, the ten ordinary DSM-IV personality disorders, two 'appendix disorders' and Self Defeating, Cluster A-C, and 'any personality disorder, five indicators of axis-I disorders, SCL-5, subjective wellbeing, and alcohol (ab)use. The results yielded no significant findings. Increases in the chi-square values after constraining the parameters to be equal ranged from 0.00 to 3.28 (*df* = 3), and the AIC values ranged from –2.72 to –6.00.

Discussion

The study explored to what extent recruitment bias and attrition may have affected the representativeness of our sample and the results of the biometric twin analyses. Results from comparisons between complete pairs and single responders indicate a clear but moderate selection effect in the first stage of the study, Q1, among male twins for symptoms of anxiety and depression (mean difference approximately ¼ *SD* between single and pairs) and somatic illnesses (difference ⅙ *SD*), but not among females. There was no difference between complete pairs and singles regarding co-twin closeness.

An array of factors could potentially influence response to broad-based health questionnaires and participation in a mental health interview. In our Norwegian, population-based program of twin research, socio-demographic factors — primarily female sex, higher education, and zygosity — were the most important predictors of participation. There was a clearly higher participation rate in the first questionnaire study (1992) than among twins invited for the first time to the second questionnaire study. This difference is probably reflecting both a time trend of decreased willingness to respond to questionnaire studies and less willingness to complete the much longer second, compared to the first, quite short, questionnaire.

Few health variables recorded at the twins' birth predicted participation, and the effects were weak. Whereas health disadvantages mostly predicted low participation, one, placenta previa, predicted high participation. Good somatic and mental health and a healthy life style reported in the first questionnaire (1992) quite moderately predicted participation in the next (1998). There are virtually no health selection effects between the second questionnaire and the interview study.

It is worthwhile noting that none of the analyses indicated selection towards low — or a priori more likely — high physical and emotional co-twin closeness. Such selection might have had particular consequences for phenotypic co-twin similarity and, thus, for results from biometric genetic analyses.

A series of quantitative genetic analyses did not show evidence of differences between interview study participants and nonparticipants in the genetic and environmental covariance structure for a broad range of mental health indicators.

Previous studies of bias in twin studies due to selection of demography and health related phenotypes are rather scarce. A longitudinal study examined sampling bias in the Australian twin cohort born from 1944 to 1963, using a reverse design to identify correlates of nonresponse at a previous occasion (Heath et al., 1998). All twins invited to participate in a questionnaire study were also recruited to a telephone interview conducted approximately 10 years later. Analyses of the interview data revealed findings highly similar to ours. Sociodemographic correlates, including education below university level, male sex, being a dizygotic twin and membership in the youngest birth cohort, showed the strongest effects explaining nonresponse to the earlier questionnaire study. The psychiatric measures yielded much more modest effects with nonresponse associated with a history of alcohol dependence, childhood conduct disorder, and social anxiety. The authors conclude that sampling biases are strongest for the sociodemographic measures and relatively minor for the psychiatric measures. Another Australian study of somewhat younger twins also showed results similar to ours (Heath et al., 2001). A British study of twin children showed around 0.2 *SD* lower mean scores for aggressive and delinquent behavior in a group whose parents responded to a questionnaire compared to nonresponder families (Taylor, 2004). Model fitting results showed lower genetic effect and higher shared environmental effect for aggressive behavior, and the opposite trend (increased genetic effect at the expense of shared environmental effect) for delinquent behavior, in responder families compared to the full sample. However, the data were not tested for differences in covariance structure between participant and nonparticipant families. The results are also not fully comparable with ours, examining effects of parental response rather than the twins' own responses, and because of the twins' young age. A study of Virginian twins aged from 8 to 16 years and their parents showed no selection towards high SES from the start of the study. Unlike with our sample, however, a selection took place during later phases of the study as families living in low-income communities dropped out of the study. This demographic selection had almost no effect on the prevalence estimates for parental mental disorders (Meyer et al., 1996). Likewise, another Virginian study of twin data, including a broad set of variables on demographic information and mental health, showed only moderate effects on participation in subsequent studies of the following variables: female sex,

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

higher education, older age, Protestant religious affiliation, and an absence of drinking problems (Kendler & Prescott, 2006). A Swedish study of selection bias in elderly twins (Simmons et al., 1997) concluded: 'The results of the present study suggest that although a selection bias may exist, it is neither pervasive nor large in population based samples' (p. 565). A Norwegian study of intelligence, comparing twins having participated or not participated in a questionnaire study, showed higher IQ among the participants, but there were no difference between participants and nonparticipants in covariance structure or heritability of IQ (Tambs et al., 1989). Likewise, our results, showing no recruitment bias for the genetic covariance structure for mental health, is in agreement with results from the Minnesota Twin Family Registry. These results showed no differences in co-twin correlations between twins responding after first invitation and twins only responding after offered incentives to participate (Lykken et al., 1990). Although differences in samples and phenotypes make comparisons difficult, our results seem to be consistent with most previous results.

Our results should not be interpreted in too much detail. Multiple testing may have resulted in a few results reaching significance by chance — perhaps in nonexpected direction — and which true effects that reached significance and which did not are to a large extent random. In general, however, strong selection over a broad range of somatic and mental health variables would have resulted in more and stronger effects than observed in our sample.

Another important limitation is related to the appropriateness of the medical variables from the MBRN used as predictors to the entry of the study. A lot of evidence exists for the impact of fetal factors on adult health (Barker, 1998), including some pertaining to mental health (Cheung et al., 2004; Thompson et al., 2001). There is also evidence of an association between low birth-weight and risk for epilepsy in males and with refractive disorders, chronic ear infections and stomach problems in women in our Q1 data (Harris et al., 1997), and with nearsightedness and minimal brain dysfunction (MBD) in the Q2 data (Grjibovski et al., 2005). Nonetheless, such observed associations are typically weak, and prenatal factors typically predict illness among people older than our twin cohort. This indirect and relatively insensitive way of testing associations between health and participation may well have left selection for health related factors undetected. The absence of evidence of differences in somatic and mental health problems between MZ and DZ twins is somewhat reassuring regarding undetected selection, however.

Perhaps the most important results derive from the biometric twin study of mental health indicators testing for differences in covariance structure between pairs who did and did not participate in the interview. One interpretation of these results is that attrition bias in our data affects prevalence rates but not the genetic and environmental variance estimates. However, this interpretation is not without caveats; the clearest selection effects for the mental health measures seem to have already occurred by Q2, and not very much selection appears to have taken place from Q2 to the interview. Furthermore, the variance component findings should be evaluated in light of power to reject equivalence in the variance components across the two groups. Simulation studies clearly demonstrate that only quite substantial differences in parameter estimates can be expected to be detected with our sample size (Neale et al., 1994). Rather than proof of absence of recruitment bias, the results — showing no trends of differences between covariance structures for a large number of phenotypes — should be understood as suggestive of no strong bias effect.

Regardless of the described limitations, the results indicate that twin studies of health are not strongly selected towards poor or strong health generally, and mental health specifically. No definite conclusion can be drawn regarding selection bias at the entry of the study, but we are confident that drop-outs during follow-ups do not seriously threaten the representativeness of our sample. The main question is whether individual or pair-wise selection affects the estimates from the genetic analyses. The tentative answer is no.

Acknowledgment

This work was supported by grants from The National Institute of Mental Health (MH-068643), NIH, and from funding in support of the twin program of research at the Norwegian Institute of Public Health including grants from the Norwegian Research Council, the Norwegian Foundation for Health and Rehabilitation, and the European Commission under the program 'Quality of Life and Management of the Living Resources' of the 5th Framework Program (no. QLG2-CT-2002-01254). We are very grateful to the twins for their participation.

References

Bardage, C., Isacson, D., & Pedersen, N. L. (2001). Self-rated health as a predictor of mortality among persons with cardiovascular disease in Sweden. *Scandinavian Journal of Public Health, 29,* 13–22

Barker, D. J. P. (1998). In utero programming of chronic disease. *Clinical Science, 95,* 115–128.

Cheung, Y. B., Ma, S., Machin, D., & Karlberg, J. (2004). Birthweight and psychological distress in adult twins: A longitudinal study. *Acta Paediatrica*, *93*, 965–968.

Czajkowski, N., Kendler, K. S., Jacobson, K. C., Tambs, K., Røysamb, E., & Reichborn-Kjennerud, T. (2008). Passive–aggressive (negativistic) personality disorder: A population-based twin study. *Journal of Personality Disorders, 22,* 109–122.

Dominicus, A., Palmgren, J., & Pedersen, N. L. (2005). Bias in variance components due to nonresponse in

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

twin studies. *Twin Research and Human Genetics, 9,* 185–193.

Grjibovski, A. M., Harris, J. R., & Magnus, P. (2005). Birthweight and adult health in a population-based sample of Norwegian twins. *Twin Research and Human Genetics*, *8*, 148–55.

Harris, J. R., Magnus, P., & Tambs, K. (2002). The Norwegian Institute of Public Health Twin Panel: A description of the sample and program of research. *Twin Research, 5,* 415–423.

Harris, J. R., Magnus, P., & Tambs, K. (1997). Does normal variation in birthweight confer susceptibility to health problems? A co-twin control study. *The Norwegian Journal of Epidemiology*, *7*, 41–47.

Harris, J. R., Magnus, P., & Tambs, K. (2006). The Norwegian Institute of Public Health Twin Program of Research: An update. *Twin Research and Human Genetics, 9,* 858–864.

Harris, J. R., Tambs, K., & Magnus P. (1995). Sex-specific effects for body mass index in the new Norwegian twin sample. *Genetic Epidemiology, 12,* 251–265.

Hesbacher, P. T., Rickels, R., Morris, R. J., Newman, H., & Rosenfeld, M. D. (1980). Psychiatric illness in family practice. *Journal of Clinical Psychiatry, 41,* 6–10.

Heath, H. C., Howells, W., Kirk, K. M., Madden, P. A., Bucholz, K. K., Nelson, E. C., Slutske, W. S., Statham, D. J., & Martin, N. G. (2001). Predictors of non-response to a questionnaire survey of volunteer twin panel: findings from the Australian 1989 twin cohort. *Twin Research, 4*, 73–80.

Heath, A. C., Madden, P. A., & Martin, N. G. (1998). Assessing the effects of cooperation bias and attrition in behavioral genetic research using data-weighting. *Behavior Genetics*, *28*, 415–27.

Irgens, L. M., Bergsjø, P., & Lie, R. T. (2000). The Medical Birth Registry of Norway. Epidemiological research and surveillance throughout 30 years. *Acta Obstetricia et Gynecologica Scandinavica, 79,* 435–439.

Kendler, K. S., & Eaves, L. J. (1989). The estimation of probandwise concordance in twins: The effect of unequal ascertainment. *Acta Geneticae Medicae et Gemellologiae, 38,* 253–270.

Kendler, K. S., & Holm, N. V. (1985). Differential enrollment in twin registries: Its effect on prevalence and concordance rates and estimates of genetic parameters. *Acta Geneticae Medicae et Gemellologiae, 34,* 125–140.

Kendler, K. S., Myers, J., Torgersen, S., Neale, M.C., & Reichborn-Kjennerud, T. (2007). The heritability of cluster A personality disorders assessed by both personal interview and questionnaire. *Psychological Medicine*, *37*, 655–65.

Kendler, K. S., & Prescott, C. A. (2006). *Genes, environment, and psychopathology: Understanding the causes of psychiatric and substance use disorders*. New York: Guilford Press.

Krause, N. M., & Jay G. M. (1994). What do global self-rated health items measure? *Medical Care, 32,* 930–942.

Liang, K. Y. & Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika, 73*, 13–22.

Lykken, D. T., Bouchard, T. J. Jr., McGue, M., & Tellegen, A. (1990). The Minnesota Twin Family Registry: some initial findings. *Acta Geneticae Medicae et Gemellologiae, 39*, 35–70.

Lykken, D. T., McGue, M., & Tellegen, A. (1988). Recruitment bias in twin research: The rule of Two-thirds reconsidered. *Behavior Genetics, 17*, 343–362.

Lykken, D. T., Tellegen, A. & DeRubeis, R. (1978). Volunteer bias in twin research – The rule of two-thirds. *Social Biology, 25*, 1–9.

Martin, N.G., & Wilson, S.R. (1982). Bias in the estimation of heritability from truncated samples of twins. *Behavior Genetics, 12*, 467–472.

Meyer J. M., Silberg, J. L., Simonoff, E., Kendler, K. S., & Hewitt, J. K. (1996). The Virginia Twin-Family Study of Adolescent Behavioral Development: assessing sample biases in demographic correlates of psychopathology. *Psychological Medicine, 26,* 1119–1133.

Moum, T., Næss, S., Sørensen, T., Tambs, K., & Holmen, J. (1990). Hypertension labelling, life events, and psychological well-being. *Psychological Medicine, 20*, 635–646.

Neale, M. C., & Eaves. L. J. (1993) Estimating and controlling for the effects of volunteer bias with pairs of relatives. *Behavior Genetics, 23*, 271–277.

Neale M. C., Eaves, L.J., Kendler, K. S., & Hewitt, J. K. (1989). Bias in correlations from selected samples of relatives: The effects of soft selection. *Behavior Genetics*, *19*, 163–9.

Neale M. C., Eaves, L. J., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavioral Genetics*, *24*, 239–58.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. (1999) *Mx: Statistical Modeling* (5th ed.). Richmond, VA: Department of Psychiatry, Virginia Commonwealth University.

Pfohl, B., Blum, N., and Zimmerman, M. (1995). *Structured Interview for DSM-IV Personality (SIDP-IV)*. Iowa City: University of Iowa, Department of Psychiatry. (Pamphlet)

Reichborn-Kjennerud, T., Czajkowski, N., Neale, M. C., Ørstavik, R. E., Torgersen, S., Tambs, K., Røysamb, E., Harris J. R., & Kendler, K. S. (2007). Genetic and environmental influences on dimensional representations of DSM-IV Cluster C personality disorders: A population-based multivariate twin study. *Psychological Medicine, 37*, 645–653.

Rubin, D. B. (1976). Inference and Missing Data. *Biometrika, 63,* 581–592.

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons, Inc.

Simmons, S. F., Johansson, B., Zarit, S. H.; Ljungquist, B., Plomin, R., & McClearn, G. E. (1997). Selection bias in samples of older twins? A comparison between octogenarian twins and singletons in Sweden. *Journal of Aging and Health, 9, 553–567.*

Taylor, A. (2004). The consequences of selective participation on behavioral-genetic findings: Evidence from simulated and real data. *Twin Research*, *5*, 485–504.

Tambs, K., Harris, J. R., & Magnus, P. (1995). Sex specific causal factors and effects of common environment for anxiety and depression in twins. *Behavior Genetics, 25,* 33–44.

Tambs, K. & Moum, T. (1993). How well can a few questionnaire items indicate mental health? *Acta Psychiatrica Scandinavica, 87,* 364–367.

Tambs, K., Sundet, J. M., Magnus, P., & Berg, K. (1989). No recruitment bias for questionnaire data related to IQ in classical twin studies. *Personality and Individual Differences, 10,* 269–271.

Thompson, C., Syddall, H., Rodin, I., Osmond, C., & Barker, D. J. P. (2001). Birth weight and the risk of depressive disorder later in life. *British Journal of Psychiatry, 179,* 450–455.

Torgersen, S. (1980). Hereditary-environmental differentiation of general neurotic, obsessive, and impulsive hysterical personality traits. *Acta Geneticae Medicae et Gemellologiae, 29*, 193–207.

Torgersen, S., Czajkowski, N., Jacobson, K., Reichborn-Kjennerud, T., Røysamb, E., Neale M. C., & Kendler, K. S. (2008). Dimensional representations of DSM-IV cluster B personality disorders in a population-based sample of Norwegian twins: A multivariate study. *Psychological Medicine, 38,* 1617–1625.

Wittchen, H.-U. & Pfister, H. (1997). *DIA-X Interviews [M-CIDI]: Manual fur Screening-Verfahren und Interview: Interviewheft Langsschnittuntersuchung [DIA-X-Lifetime]; Erganzungsheft [DIA-Xlifetime]; Interviewheft Querschnittuntersuchung [DIA-X 12 Monate]; Erganzungsheft [DIA-X 12 Monate]; PC-Programm zur Durchfuhrung des Interviews [Langs- und Querschnittuntersuchung]; Auswertungsprogramm.* Frankfurt: Swets & Zeitlinger.

World Health Organization. (1997). Composite International Diagnostic Interview (CIDI) Version 2.1. Geneva: Author.

Ørstavik, R. E., Kendler, K. S., Czajkowski, N., Tambs, K., & Reichborn-Kjennerud, T. (2007). The relationship between depressive personality disorder and major depressive disorder: A population-based twin study. *American Journal of Psychiatry, 164*, 1866–1872.

https://doi.org/10.1375/twin.12.2.158 Published online by Cambridge University Press