Methodology in longitudinal studies on psychological effects of predictive DNA testing: a review

R Timman, T Stijnen, A Tibben

In the last two decades predictive testing programs have become available for various hereditary diseases, often accompanied by follow-up studies on the psychological effects of test outcomes. The aim of this systematic literature review is to describe and evaluate the statistical methods that were used in these follow-up studies. A literature search revealed 40 longitudinal quantitative studies that met the selection criteria for the review. Fifteen studies (38%) applied adequate statistical methods. The majority, 25 studies, applied less suitable statistical techniques. Nine studies (23%) did not report on dropout rate, and 18 studies provided no characteristics of the dropouts. Thirteen out of 22 studies that should have provided data on missing values, actually reported on the missing values. It is concluded that many studies could have yielded more and better results if more appropriate methodology had been used.

In fifteen years ago, predictive genetic testing became available for hereditary diseases with onset later in life. Since then, quantitative and qualitative psychological follow-up studies have shown that predictive testing offers several benefits. Although most psychological research on genetic testing has focused on the tests for Huntington’s disease (HD) and the BRCA1 and BRCA2 mutations, clinical experience has provided evidence that the findings can be extrapolated to other diseases with similar inheritance patterns. Research included the characteristics of tested individuals, why they make certain choices, their understanding of the information conveyed by the test, and how they adjust after the test results. Generally, people are able to make informed decisions and they can cope with the test results. However, little is known about those who did not apply for predictive testing, or those who withdrew from the follow-up studies. The latter individuals who drop out of studies may bias the findings. Obviously, careful statistical modelling is required to provide reliable conclusions, but also to make optimal use of the data. This has led to growing interest in the methodological quality of the studies. Although sophisticated software enables almost everyone to carry out the most exotic of statistical and epidemiological analyses, the methodological implications are too often incompletely understood. This review aims to investigate the statistical methods used in the studies of psychological effects of DNA testing for genetic diseases.

METHODS
Search methods and inclusion criteria
First, the reviews of Broadstock et al., Meiser and Dunn, and Duisterhof et al. provided insight into the studies that have been published. In addition, we used the databases MEDLINE and PsycLIT from 1988 onwards. The key words used for the searches were: Huntington’s disease, HBOC, BRCA*, FAP, familial adenomatous polyposis, HNPCC, colon cancer, SCA, or spinocerebellar ataxia.

Follow-up studies encompass the prediction of psychological adjustment or the course of adjustment over time. Firstly, studies are included if they analyse quantified measures statistically. Secondly, studies should have investigated the psychological effects of a genetic test outcome for hereditary diseases. Thirdly, studies should have a longitudinal, not a cross-sectional or retrospective design.

Categorisation strategy for longitudinal research methods
This review categorises and discusses articles according to the longitudinal methods that are used. Methods that could be valuable for longitudinal research such as structural equation models were not observed and hence are not discussed. For a more complete overview of longitudinal research we refer to Bijleveld et al. Statistical methods can be classified as less adequate for two reasons. First, when the measurement level of the variable does not correspond with the level that is required for the method that is used. Second, when the study includes more than three waves, but the method used is only suitable for two waves at most. We distinguished in this review five issues of interest: (1) the measurement level of the variables, (2) the number of measurement moments (waves), (3) dropout handling, (4) missing value handling, and (5) the groups and subgroups that are studied. For classifying the studies we considered the measurement level (continuous or discrete) and the number of waves (two wave or multiple wave studies).

A longitudinal effect study must contain a pre-test baseline measure, and one or more post-test measures. Depending on the number of waves, several effects can emerge over the course of time. With two waves, a linear effect can be found. With multiple waves a quadratic effect can be found, which implies that the linear effect between baseline and first post-test follow-up is
not continued at a later follow-up. In studies with more than three waves, cubic effects can be found, analogous to a third power polynomial. We will categorise firstly according to measurement level: continuous and discrete; and secondly according to longitudinal capabilities: higher order time effects, only linear time effects, and no time effects.

**Dropout analyses and reporting**

Dropout, caused by subjects who do not return for follow-up measurements, is a serious problem in virtually all longitudinal studies. Dropout can invalidate the findings of a study when dropouts have characteristics or psychological outcomes different from the persons who remain in the study.

**Missing value handling and reporting**

Related to the dropout problem, but different in nature, is the handling of missing measurement points. Dropouts are subjects who participated at the start of the study, but do not return for follow-up after a certain time point, for instance after the disclosure of a test result. We refer to missing values when subjects did not complete questionnaires at a certain time point, but did return for follow-up at other time points. It should be noted that this problem emerges only in longitudinal studies with multiple waves.

**RESULTS**

The search was carried out in November 2003. The first search resulted in 41,840 references. This search was then combined with the keywords: genetic* and psychol*, which decreased the number of references to 585. From the abstracts it was concluded that 86 of these could be useful for this study.

Forty-six articles did not meet our inclusion criteria as:

- fourteen were qualitative studies. Although some of these studies used quantitative measures, the data were not statistically analysed and thus the studies were classified as qualitative for our purpose.7–20
- ten studied psychological effects, but not of a genetic test outcome.21–30
- thirteen studies were not longitudinal but cross-sectional or retrospective.31–43
- three did not investigate psychological outcomes of genetic testing.44–46
- six were reviews on effects of (genetic) testing.47–50

Studies that could be classified in more than one category were mentioned in the first of the exclusion criteria involved. Forty articles met the inclusion criteria.51–60 (table 1)

**Categorisation of methods**

**Methods for continuous outcome variables and three or more waves**

General linear mixed models (GLMM) is also referred to as random effect modelling, random coefficient regression modelling, mixed models, or multilevel regression analysis. This method, used in five studies,1 49 52 54 64 has three advantages: (1) incomplete cases can be analysed, (2) different time spans for individuals between the waves can be handled, and (3) the method allows control for confounding variables, for example age or the number of children, by entering these as covariates into the model equation. Missing data that are dependent on the observed outcome variables and other observed characteristics can be dealt with by including these in the analysis.51–60

**Repeated measures analysis of variance**

Eight studies used repeated measures analysis of variance on continuous data with multiple waves.51 52 54 64 65 75 81 86 Repeated measures analysis of variance has three main advantages: (1) it is relatively easy to perform, (2) several outcome variables can be analysed simultaneously, and (3) confounding variables can be included as covariates. The disadvantage, that only complete cases can be analysed, can be reduced by imputing the missing values of incomplete cases. A second, but less serious problem is that the time spans between the waves must be equal for each participant. Five studies did not report on quadratic or higher order time effects, which suggests that they have not made optimal use of repeated measures analysis.51 59 75 80 86 Three studies have used missing value imputation,51 59 81 which is discussed in the section on missing data. One study75 used SPSS MANOVA on discrete variables, which must be considered as less adequate.

**Methods for continuous outcome variables and two waves**

**Repeated measures analysis of variance**

Three studies53 57 58 used this highly appropriate method for analyses on two waves.

**Regression analysis, (multiple) linear regression analysis, sequential or hierarchical regression analysis**

The follow-up outcome variable is defined as the dependent variable in the regression equation. The baseline scores, DNA test outcome, and other variables (gender, age, education) are defined as independent variables. Two studies59 60 used this method adequately on two waves. Four studies61 62 69 73 did not include the baseline measure in the analysis, which is less adequate when there are baseline differences. One of these63 did not include the baseline measure because of missing baseline data. Four studies60 71 79 81 used this analysis for multiple waves, which is less adequate, because only two by two comparisons can be made.

**Analysis of variance or covariance of change scores**

In principle this method is the same as regression analysis. Two studies66–68 used it to analyse multiple waves, which is not optimal. One of these68 did not use the baseline score as a covariate, and differences were analysed at baseline in separate t-tests.

**Logistic regression**

By using the DNA test outcome as the dependent variable, logistic regression can be used for analysing continuous outcome variables. Baseline, follow-up scores, and confounding variables can be entered as independent variables. This method is unconventional since the role of the determinant (DNA test outcome) and the outcome (psychological test) are interchanged. The advantage of this method is that very few requirements are posed on the independent variables. Two studies64 65 used logistic regression in this way for their multiple wave study.

**Paired samples t-test**

With two time points, the paired samples t-test yields the same solution as a repeated measure ANOVA. However, no comparisons for change in course of time between groups can be made. One study72 used it for two time points, and they analysed the differences between groups separately. It is better to use one integrative method, so that interaction effects can be revealed. Two studies66 67 used this technique less adequately for multiple waves and for more than one group.

**Methods for continuous outcome variables that are not longitudinal**

**t-test for independent samples**

Brandt et al.90 used this method for their study comprising of eight waves. Probably because the sample size was relatively
small compared to the number of waves, repeated measures analysis of variance would have yielded invalid results. As more advanced methods were not common in 1989, there is no reason to object to this method. Lawson et al. used this test to analyse baseline characteristics of persons who had an adverse event after the prediagnostic test for HD.

Methods for discrete outcome variables and three or more waves

Friedman's test for ordinal data
This test is regarded as the non-parametric equivalent of a one sample repeated measures design. It neither reveals differences between groups, nor can it reveal quadratic, cubic, or higher order time effects. If a significant time effect within a group is revealed, pair-wise comparisons between waves must be analysed separately for significance. Two studies used this procedure for their continuous data. This could be adequate if the variables could not be successfully transformed to normality. Neither study reported on the distribution of the variables.

Methods for discrete outcome variables and two waves

Logistic regression
Logistic regression is an appropriate method of analysis when the outcome variable is dichotomous. One study

Table 1

<table>
<thead>
<tr>
<th>Study</th>
<th>Disease</th>
<th>Outcome variables</th>
<th>Statistical method</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continous variables, more than two waves</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Codori (1997)</td>
<td>HD</td>
<td>BDI, BHS</td>
<td>GLMM</td>
<td>+</td>
</tr>
<tr>
<td>Meiser (2002)</td>
<td>BRCA</td>
<td>IES, BDI, STAI, MBSS</td>
<td>GLMM</td>
<td>+</td>
</tr>
<tr>
<td>Codori (2003)</td>
<td>FAP</td>
<td>SCL-90, BDI</td>
<td>GLMM</td>
<td>+</td>
</tr>
<tr>
<td>Almqvist (2003)</td>
<td>HD</td>
<td>SCL-90, BDI, GWB</td>
<td>GLMM</td>
<td>~</td>
</tr>
<tr>
<td>Timman (in press)</td>
<td>HD</td>
<td>IES, BHS</td>
<td>GLMM</td>
<td>~</td>
</tr>
<tr>
<td>Wiggins (1992)</td>
<td>HD</td>
<td>IES, BHS, GHQ</td>
<td>Repeated measures MANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Tibben (1994)</td>
<td>HD</td>
<td>SCL-90, BDI, WAIS-R, HVLT, SCWT, WCST, QNE</td>
<td>Repeated measures MANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Campbellonic (1996)</td>
<td>HD</td>
<td>IES, BHS</td>
<td>Repeated measures MANOVA</td>
<td>+</td>
</tr>
<tr>
<td>Dudakdolཏ Willis (1998)</td>
<td>HD, FAP, HBOC</td>
<td>IES</td>
<td>Repeated measures MANOVA</td>
<td>+</td>
</tr>
<tr>
<td>Akten-Callan (2001)</td>
<td>HNPPC</td>
<td>STAI, IAS</td>
<td>Repeated measures ANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Ladder (2002)</td>
<td>BRCA</td>
<td>IES, SCL-90, HADS, BIS</td>
<td>Repeated measures ANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Decruyenaere (2003)</td>
<td>HD</td>
<td>STAI, BDI, SCL-90, IES, HOS, MWWI</td>
<td>Repeated measures ANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Tercyak (2001)</td>
<td>STAI</td>
<td>Regression</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>Schwartz (2002)</td>
<td>BRCA</td>
<td>IES, HSCL</td>
<td>Regression, ANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Decruyenaere (1996)</td>
<td>HD</td>
<td>STAI, BDI, MWWI, UCL</td>
<td>Paired &amp; test, regression</td>
<td>~</td>
</tr>
<tr>
<td>Friedman (1999)</td>
<td>BRCA</td>
<td>IES, POMS-SF</td>
<td>Logistic regression</td>
<td>~</td>
</tr>
<tr>
<td>Oostrom (2003)</td>
<td>BRCA</td>
<td>HADS, IES, CWS, ODCFS</td>
<td>Logistic regression</td>
<td>~</td>
</tr>
<tr>
<td>Horowitz (2001)</td>
<td>HD</td>
<td>IES, BDI</td>
<td>ANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Brandi (1999)</td>
<td>HD</td>
<td>STAI, SCL-90, BDI, BHS</td>
<td>Test for independent samples</td>
<td>~</td>
</tr>
<tr>
<td>Lerma (1999)</td>
<td>HNPPC</td>
<td>CES-D</td>
<td>Logistic regression</td>
<td>~</td>
</tr>
<tr>
<td>Michel (2001)</td>
<td>FAP</td>
<td>IES, STAI, HADS, RCBS, HOS, LOT</td>
<td>Friedmann’s test for ordinal data</td>
<td>~</td>
</tr>
<tr>
<td>Broadstock (2000)</td>
<td>HNPPC</td>
<td>IES, STAI, GHQ, CWS</td>
<td>Friedmann’s test for ordinal data</td>
<td>~</td>
</tr>
<tr>
<td>Aho (1997)</td>
<td>SCA</td>
<td>STAI, SDS</td>
<td>Kruskal-Wallis</td>
<td>~</td>
</tr>
<tr>
<td>Quaid (1995)</td>
<td>HD</td>
<td>SCL-90, BDI, BHS, MSI</td>
<td>Mann-Whitney U test</td>
<td>~</td>
</tr>
<tr>
<td>Lawson (1996)</td>
<td>HD</td>
<td>SCL-90, BDI, QSS, adverse events</td>
<td>No longitudinal analyses performed</td>
<td>~</td>
</tr>
<tr>
<td>Codori (1996)</td>
<td>HD</td>
<td>SCL-90, BDI, CDI, RADS, RCAS, CBCL</td>
<td>Repeated measures ANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Ladder (2001)</td>
<td>BRCA</td>
<td>IES, HADS</td>
<td>Repeated measures ANOVA</td>
<td>~</td>
</tr>
<tr>
<td>Kirkwood (2003)</td>
<td>STAI, MWWI</td>
<td>Regression</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>Metteau (1997)</td>
<td>STAI, GHQ, 12 dichotomatic questions</td>
<td>Regression, logistic regression</td>
<td>~</td>
<td></td>
</tr>
<tr>
<td>Crayle (1997)</td>
<td>BRCA</td>
<td>IES, STAI</td>
<td>Regression</td>
<td>~</td>
</tr>
<tr>
<td>Decruyenaere (1999)</td>
<td>HD</td>
<td>STAI, BDI, MWWI</td>
<td>Regression</td>
<td>~</td>
</tr>
<tr>
<td>Smith (1999)</td>
<td>BRCA</td>
<td>IES, STAI</td>
<td>Regression</td>
<td>~</td>
</tr>
<tr>
<td>Ritvo (2000)</td>
<td>BRCA</td>
<td>STAI, CES-D, LOT, MOSS</td>
<td>Regression</td>
<td>~</td>
</tr>
<tr>
<td>Grosfeld (2000)</td>
<td>MEN</td>
<td>IES, STAI, SCL-90</td>
<td>Paired &amp; test</td>
<td>~</td>
</tr>
<tr>
<td>Lerma (1999)</td>
<td>BRCA</td>
<td>IES, CES-D</td>
<td>ANOVA on difference scores, regression</td>
<td>~</td>
</tr>
<tr>
<td>Wood (2000)</td>
<td>BRCA</td>
<td>IES, HSCL</td>
<td>Wilcoxon’s sign rank/rank sum test</td>
<td>+</td>
</tr>
<tr>
<td>Taylor (1997)</td>
<td>HD</td>
<td>SCL-90, BDI, BHS</td>
<td>Fisher’s exact test</td>
<td>~</td>
</tr>
</tbody>
</table>

Discrete variables, more than two waves

Akten-Callan (2000) | HNPPC | 5-answer categories questions | Repeated measures ANOVA, paired & tests | ~ |

*BRCA, breast cancer (BRCA1 and BRCA2); CF, cystic fibrosis; FAP, familial adenomatous polyposis; HBOC, hereditary breast and ovary cancer; HD, Huntington disease; HNPPC, hereditary non-polyposis colon cancer; LF, Li-Fraumeni cancer syndrome; MEN, multiple endocrine neoplasia; SCA, spinocerebellar ataxia.
†BDI, Beck Depression Inventory; BHS, Beck Hopelessness Scale; BIS, Body Image Scale; BSI-53, Brief Symptom Inventory; CBCL, Child Behaviour Checklist; CDI, Children’s Depression Inventory; CES-D, Centre for Epidemiological Studies, Depression; CWS, Cancer Worry Scale; GHQ, General Health Questionnaire; GWB, General Well Being scale; HADS, Hospital Anxiety and Depression Scale; HOS, Health Orientation Scale; HSC, Hopkins Symptom Checklist; HVLT, Hopkins Verbal Learning Test; IAS, Illness Attitude Scale; IES, Impact of Event Scale; LOT, Life Orientation Test; MBSS, Miller Behavioural Style Scale; MWWI, Minnesota Multiphasic Personality Inventory; MOSS, Medical Outcomes Study Support Scale; MSI, Marital Satisfaction Inventory; ODCFS, Openness to Discuss Cancer in the Family Scale; POMS-SF, Profile of Mood States-Short Form; QNE, Quantified Neurological Examination; RADS, Reynolds Adolescent Depression Scale; RCBS, Rutter Child Behaviour Scale; RCAS, Revised Children’s Manifest Anxiety Scale; SCL-90, Symptom Checklist-90; SCWT, Stroop Color and Word Test; SDDS, Social Desirability Scale; SSS, Social Support Questionnaire; STAI, State Trait Anxiety Inventory; UCL, Utrechtse Coping List; WAIS-R, Wechsler Adult Intelligence Scale-Revised; WCST, Wisconsin Card Sorting Test.
‡ & test, regression adequate; ~ less adequate.
dichotomised the continuous outcome variable, and performed logistic regression analysis which is less efficient. The authors also report that measures were taken at three time points, but they barely touched on the third wave in the result section.

Wilcoxon signed ranks test in combination with Wilcoxon rank sum test
These are non-parametric equivalents to a paired samples t-test and a t-test for independent samples. One study77 used these tests for one group and two waves, on continuous variables that were not normally distributed. This is a reasonable alternative when variables cannot be transformed to normality, though no interactions can be analysed.

Methods for discrete outcome variables that are not longitudinal
Kruskal-Wallis H test
This is a non-parametric equivalent to one-way ANOVA. One study59 seem to have used this test to compare three groups with respect to differences between the follow-up measure and baseline. Analyses for each of the three follow-up measurements were performed separately.

Mann-Whitney U test
This test is equivalent to a Kruskal-Wallis H test, but restricted to comparing only two groups. One study63 used this test on continuous data for multiple waves because of a small sample size. Carriers were compared with non-carriers at each time point separately. Another study62 used this test to analyse baseline characteristics of persons who had an adverse event after the prediagnostic test for HD.

Fisher’s exact test
One study45 used Fisher’s exact test for analysing the difference between the number of people who had an increase and those who had a decrease since baseline with regard to certain outcome variables. When continuous variables are treated in this way, much information may be lost.

Dropout analysis and reporting
In this review we differentiate between individuals lost to follow-up (that is, dropouts) and those for whom data are incomplete as a consequence of missing time points. Incompleteness of data within questionnaires because participants did not answer all questions is not discussed here. In general, questionnaire manuals provide rules for handling this problem. Moreover, this is not a specific issue of longitudinal designs.

We divided the studies into four groups: (1) baseline differences analysed and found, (2) baseline differences analysed but not found, (3) dropout rate reported, but no analysis for differences reported, and (4) no mention of dropout (table 2):

(1) thirteen studies reported differences between dropouts and participants who returned for follow-up questionnaires.1 3 5 8 48 63 64 66 70 78 83-86 90
(2) ten studies reported that differences between dropouts and persons participating in follow-up were analysed, but no differences were found.11 64 80 71 75 77 81 82 87
(3) eight studies reported the dropout rate, but did not analyse possible characteristics of drop-outs.50 54 56 59 61 67 70 72
(4) nine studies reported neither on dropout analyses, nor on dropout rate.52 55 62 69 71 74 79 One study57 seemed to claim that there were no dropouts at all, though from the text it can be inferred that there must have been between one and 18 dropouts. And one study60 reported that unfortunately no baseline records of dropouts were kept.

Missing value handling and reporting
Twenty-eight studies included three or more waves, which made these studies vulnerable to missing values. Five studies used GLMM for analysis, and one study performed no longitudinal analysis.57 From the remaining 22 studies information is needed about how they dealt with missing values. Three studies imputed missing values before performing repeated measures analysis of variance. One72 used singular regression imputation, which is considered inferior to multiple imputation.51 58 59 one27 used mean substitution, which is generally insufficient, and one43 did not report which method they used. In 10 studies participants with missing time points were excluded from the analyses84 54 56 66 70 71 78 86 88 and nine studies did not report on the handling of missing time points at all.50 52 53 59 64 76 79 83 87

The groups and subgroups that are studied
In 35 studies carriers were compared to non-carriers. Several of these studies also included other groups: people with an uninformative test outcome,50 51 57 62 65 69 73 74 79 one27 people who refrained from testing,57 58 66 partners,16 48 1 parents of individuals tested from testing,57 58 66 and people who had had a spinalcord attack.10 One study compared unaffected Li-Fraumeni, unaffected BRCA1 tested individuals and women who were carriers of BRCA1 mutations.73 One study66 compared high and average risk groups of BRCA 1/2 mutation negatives. One study62 compared parents with children tested for the MEN2 gene: all children positive, all negative, and mixed. Two studies included only one group.50 79

Accuracy of reporting
Some studies reported in an incomplete or unclear fashion. Sometimes the size of the study group and inclusion criteria remained unclear, or no actual p-values were reported. In other studies the presented p-values were different from values that could be calculated from the tables. Sometimes, the number of participants inferred from df or χ²-values was not in accordance with the reported number of participants.

DISCUSSION
The aim of this review was to describe the methodology and statistics of psychological follow-up studies on effects of predictive genetic testing. Fifteen studies were found to have applied more or less adequate statistical methods. The majority of the studies, however, applied statistical techniques that were less suitable or less efficient for the data that were available to the researchers. We evaluated studies on five issues: (1) the measurement level of the variables, (2) the number of waves, (3) dropout handling, (4) missing value handling, and (5) the groups and subgroups that were studied.

The measurement level of the variables
Generally, most studies used variables with a continuous measurement level. Many statistical methods that are appropriate for this measurement level require normal distribution. If not, an attempt can be made to transform it to normality.92 93 If transformation is not successful, a non-parametric test should be used, as is prescribed for discrete variables. Generally, parametric tests are more efficient than non-parametric tests.94 For this reason it is recommended that a parametric test is used whenever permitted. Thirty-one studies used a parametric test on continuous data, seven studies used a non-parametric test on continuous data, one study used a parametric test on discrete data, and one study did not perform any longitudinal analysis.
Dropout analyses and reporting

Eighteen studies gave no evidence of having performed any analysis on the characteristics of dropouts. These differences should include outcome variables and all biographical measures that have been assessed. We favour the suggestion of Moher who reports that a flow diagram should be provided with the number of participants in any condition and any moment, and that reasons for these numbers should be given. Only two studies in this review actually provided such a flow chart.

Sample size

An important characteristic of a study is the number of participants. The costs and efforts needed to conduct a large sample study will be higher than a small sample study. Obviously, a study with a large sample will reveal more (significant) effects. Although sample size issues are not within the scope of this review, it should be noted that a minimum sample size is needed to perform quantitative analyses. Our scope is to review the effectiveness and correctness of the methodology used, which is an issue independent of the sample size.
A model for study design and analysis

The type of study for examining the effects of testing for late onset genetic diseases is a longitudinal design in which one or more post-test measures are compared with a pre-test measure. Admittedly, there is not one ideal study design as it depends on the resources the researcher has access to, the subjects who can and want to participate, and the questions that are to be answered. We provide some recommendations. The study should include a baseline measure, an intervention such as a genetic test outcome, and follow-up measures. Generally it is concluded from previous research that test results have a large impact directly after the test, but measures taken some time after baseline measurement. Timman et al suggested that the genetic test outcome does have long term effects. For this reason it is recommended that a study be continued over several years. For an initial study report, for example when baseline and the first follow-up are undertaken, analysis can be done with repeated measures analysis of variance. For reports on subsequent follow-ups, GLMM is recommended. Although GLMM can handle incomplete cases, all efforts should be made to avoid missing values and retain as many participants in the study as possible.

In many articles no rationale was given for the analytical approach used. Sometimes the method used for the analyses was not clearly described, and we had to infer the analysis from the reported results. In some cases this may have meant that a different procedure was used than we inferred. The use of an inadequate method can result in incorrect conclusions, but more often it can result in a failure to find a significant effect. To determine whether studies would have found different results if a more adequate method had been used, one needs to reanalyse the data. In a number of circumstances we performed the method described on our own data, and we compared the results of the various analyses. It is beyond the scope of this article to report on this extensively. Often the use of a non-parametric test where a parametric test could be used, does not lead to a dramatic loss of power efficiency. Mostly the power of a non-parametric test is about 95% compared to an F-test when conditions for this test are met. In some circumstances, however, for example when distributions are dichotomised before analysing, this power can drop dramatically to a lower 63%.

In this review, we found a number of studies that used sound methods and reported their findings and dropout handling in an excellent way. Our purpose is to present these methods and ways of reporting to all researchers in the field of psychological effects of genetic testing. Using an inadequate technique can cause loss of information, for example when distributions are dichotomised before analysing, this power can drop dramatically to a lower 63%.

ACKNOWLEDGEMENTS

We thank Vicky Kraver for correcting our use of the English language.

Authors’ affiliations
R Timman, A Tibben, Department of Neurology, Leiden University Medical Centre, Leiden, The Netherlands, and Department of Medical Psychology and Psychotherapy, Erasmus University Medical Centre, Rotterdam, The Netherlands
T Stijnen, Department of Epidemiology and Biostatistics, Erasmus University Medical Centre, Rotterdam, The Netherlands
A Tibben, Centre of Human and Clinical Genetics, Leiden University Medical Centre, Leiden, The Netherlands

Conflict of interest: none declared.

REFERENCES


