A twin-family study of smoking behavior

Jacqueline M. Vink

ISBN: 90-9018584-4
Printed by: Universal press
Cover: Eric van Rossum
Lay-out: Jacqueline Vink

VRIJE UNIVERSITEIT

A TWIN-FAMILY STUDY OF SMOKING BEHAVIOR

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit Amsterdam
op gezag van de rector magnificus
Prof. Dr. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Psychologie en Pedagogiek
op maandag 15 november 2004 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Jacqueline Mignon Vink

geboren te Hoogeveen

Promotor:	Prof. Dr. D.I. Boomsma Department of Biological Psychology, Vrije Universiteit Amsterdam
Co-promotor:	Dr. A.H.M. Willemsen Department of Biological Psychology, Vrije Universiteit Amsterdam
Reading committee:	Prof. Dr. R.C.M.E. Engels Institute of Family and Child Care Studies, University of Nijmegen Prof. Dr. J.J. Hudziak Department of Psychiatry, University of Vermont, Burlington, US Prof. Dr. P.E. Slagboom Section Molecular Epidemiology, Leiden University Medical Centre Prof. Dr. A.B. Smit Department of Molecular and Cellular Neurobiology, Vrije Universiteit Amsterdam Prof. Dr. J.P. Vandenbroucke Department of Clinical Epidmiology, Leiden University Medical Centre Dr. J.P.W. Vermeiden IVF-centre, VU University Medical Centre

This work was supported by the Netherlands Organization for Scientific Research (NWO 985-10-002) and ZonMW/NIDA (3100.0038).

Financiel support by the Netherlands Heart Foundation and Stivoro for the publication of this thesis is gratefully acknowledged.

De verstand-houding

Wij weten het niet
maar verschillen wel
van mening over
wat nu de vraag is

Onzekerheden
worden beantwoord
met andere en
er door vervangen

De waarheid is niet
wat bekend is maar
het onbekende
en ondenkbare

Wetenschap is een
verstand-houding die
geloof hecht aan dat
wat onzeker is

Jan Slothouber,
Uit: "Tegen Beter Weten",
Amsterdam 2000, VU-podium

Table of Contents

Chapter 1:	General introduction and research design	1
Chapter 2:	Estimating non-response bias in family studies: application to mental health and lifestyle	15
Chapter 3:	The association of current smoking behavior with the smoking behavior of parents, siblings, friends and spouses	27
Chapter 4:	Smoking status of parents, siblings and friends: predictors of regular smoking? Findings from a longitudinal twin-family study	41
Chapter 5:	The Fagerström Test for Nicotine Dependence in a Dutch sample of daily smokers and ex-smokers	55
Chapter 6:	Heritability of smoking initiation and nicotine dependence	67
Chapter 7:	Gene finding strategies	81
Chapter 8:	Linkage analyses of smoking initiation and quantity in Dutch sibling pairs	99
Chapter 9:	Summary and discussion	115
Samenvatting	(Dutch summary)	125
References		129
List of publications		141
Dankwoord		145

Chapter 1

General introduction and research design

Introduction

How can we explain the individual differences in smoking behavior? Some individuals never initiate smoking, some start but do not become nicotine dependent, while others become highly dependent. Explanations at different levels are possible and include upbringing, peer influence and genetics.

In 1997, Judith Koopmans used data from the Netherlands Twin Register (www.tweelingenregister.org) to write her PhD-thesis entitled: "The Genetics of Health-related Behaviors" (Koopmans, 1997). These health-related behaviors included sport participation, alcohol use and smoking. For smoking the results indicated a small genetic influence on initiation and a larger genetic influence on quantity smoked (number of cigarettes per day). Familial resemblance for smoking behavior could not be attributed to cultural transmission: children do not imitate the smoking behavior of their parents but genetic similarities are responsible for resemblance in smoking behavior of parents and children.

In the present thesis the individual differences of smoking behavior and nicotine dependence are further investigated. The familial association is examined, not only between parents and offspring but also between siblings, friends and spouses. In addition, the influence of smoking family members and friends on the uptake of regular smoking is predicted using a longitudinal design. Familial associations can be due to shared genetic influences or to shared environmental factors. Koopmans already explored whether genetic and/or shared environmental influences played a role in smoking initiation and quantity. In the present thesis I extended the study to nicotine dependence. After obtaining evidence for heritability for a trait, the next step is identification of the chromosomal regions involved in smoking behavior. This thesis concludes with a linkage analyses of both smoking initiation and nicotine dependence.

1. Smoking behavior

Smoking is a complex behavioral trait. Several, possibly associated, dimensions of smoking behavior may be distinguished: smoking initiation, quantity (number of cigarettes per day) and nicotine dependence.

Smoking initiation

Smoking initiation is often measured by asking whether or not a person has ever smoked. In the Dutch population 65.4% of the males of 15 years and older and 54.8% of the females of 15 years and older has ever smoked. The percentage current smokers, ex-smokers and never-smokers is shown in Figure 1.1 (STIVORO 2002).

The prevalence of smoking increases from adolescence to adulthood. At later ages (>64 years) the percentage current smokers decreases. For men the percentage individuals who ever smoked is higher in the older age groups while for women the percentage is highest in the middle age groups (Table 1.1).

The prevalence of smoking increases from adolescence to adulthood. At later ages (>64 years) the percentage current smokers decreases. For men the percentage

individuals who ever smoked is higher in the older age groups while for women the percentage is highest in the middle age groups (Table 1.1).

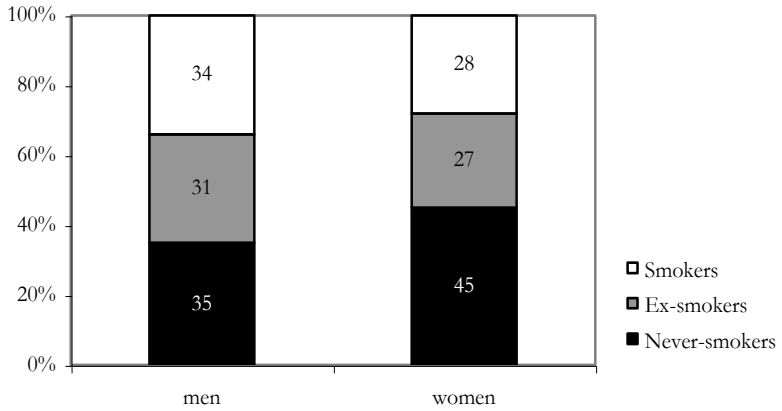


Figure 1.1 Percentage smokers, ex-smokers and never-smokers in the Dutch population (n=18212). Stivoro – rookvrij. Roken, de harde feiten. Volwassenen 2002.

Quantity

After smoking is initiated some smokers become regular smokers while others do not smoke on a regular basis. The 'chippers' are smokers who do not smoke daily and smoke less than 5 cigarettes per day. Among the daily smokers there is a large individual variety in the quantity smoked. Some regular smokers smoke 5 cigarettes per day while others smoke 30 cigarettes per day or more. In 2002, the average for the Dutch population of 15 years and older was 20.5 cigarettes per day. The amount of cigarettes smoked is reflected by the nicotine levels in blood. Measures of blood nicotine levels in groups of smokers indicate that these levels are stable from day to day within individuals (Winger *et al.*, 1992).

Table 1.1 Percentage smokers in the Netherlands in age-groups in 2001.
Source: Stivoro 2002.

	% current smoker		% ever smoked	
Age:	Men	Women	Men	Women
15-19	27	28	30	32
20-34	39	31	50	46
35-49	40	35	68	66
50-64	31	26	81	60
>64	18	15	86	51

Nicotine Dependence

The number of cigarettes smoked per day is often used as proxy for the extent of nicotine dependence. A more direct way to determine nicotine dependence is to use structured interviews like the DSM-IV or, alternatively, self-report measures of

nicotine dependence such as the Fagerström Tolerance Questionnaire (FTQ) (Colby *et al.*, 2000). The FTQ was developed in 1978 (Fagerström, 1978) and a revised version was published in 1991: the Fagerström Test for Nicotine Dependence (FTND), (Heatherton *et al.*, 1991). The FTND consists of 6 items and produces a score ranging from 0 to 10 with higher scores indicating more nicotine dependence. The FTQ and FTND scores are correlated with biochemical measures of dependence like plasma nicotine, cotinin-levels in plasma and urine and expired CO (Heatherton *et al.*, 1991). In several countries, FTND scores ranged from 1.84 to 4.30 in population samples of current smokers. (Fagerström *et al.*, 1996; Etter *et al.*, 1999; John *et al.*, 2003). FNTD scores for the Dutch population are currently unknown, chaper 5 of this thesis presents the first available FTND data in a Dutch population.

2. Research design and participants

In this thesis, individual differences in the dimensions of smoking behavior are investigated using longitudinal survey data and DNA data. For my study I have used existing data on smoking and also collected new data.

The existing data consist of:

1. Survey data from the longitudinal twin-family study on health, personality and lifestyle collected in 1991, 1993, 1995 and 1997.
2. DNA data collected in the Netherlands Twin-family Study of Anxious Depression (NETSAD).

The new data collection for the present thesis can be summarized in two parts:

1. Survey data collected in 2000 as part of the longitudinal twin-family study of health, personality and lifestyle (including a telephone interview in 2001).
2. DNA and additional smoking data collected in the Netherlands Twin-family Study of Smoking (NETSMOK).

In the next section an overview is presented of the research design and participants. First the survey study is described: the existing survey data from 1991, 1993, 1995 and 1997 are introduced and the data collection in the 2000 survey. The phenotypes on smoking assessed in all 5 data waves are presented. Next, the two DNA studies are introduced: the existing data from the NETSAD study and the new data collected in the NETSMOK study are described.

Existing survey data collected in 1991-1997

In 1991 the Netherlands Twin Register (NTR) started a large-scale twin/family study on health, personality and lifestyle. Families of adolescent twins were recruited in 1990 by asking city councils in the Netherlands for addresses of twins aged 13-22 years old. In later years, additional volunteer twin families also participated in the study. Data were collected by mailed surveys in 1991, 1993, 1995 and 1997 (Boomsma *et al.*, 2002). Twins were invited at all occasions, parents in 1991, 1993 and 1995 and siblings in 1995 and 1997. All surveys included questions on lifestyle (smoking, alcohol use,

exercise), personality (e.g. sensation seeking, neuroticism, extraversion) and physical and mental health (e.g. general health, depression).

New survey data collected in 2000

A fifth questionnaire was sent in 2000 to twins, their siblings and to spouses of the twins aged between 25 and 30 years old. In May 2000, questionnaires were sent to 13724 twins/triplets and 2889 siblings (appendix 1: letter). In July 2000 a reminder was send to the non-respondents (appendix 2). Additional smoking data were collected with telephone interviews in 2001.

In total, questionnaires were send to 14288 twins/triplets and 3665 siblings from 7223 families. The average family size was 2.48 (SD 0.99). At the end of the data collection, 4609 twins/triplets and 1474 siblings from 3178 families had completed a questionnaire booklet (Figure 1.2). Twins aged between 25 and 30 years old received an additional questionnaire for their spouse. Of the 5629 twins aged between 25 and 30 years old, 1564 completed a questionnaire themselves. Of the 1564 twins who completed a questionnaire, 1116 reported to have a spouse and 686 spouses completed the survey (61.5%). Furthermore, 21 spouses of twins aged younger than 25 or older than 30 completed a questionnaire. They entered the study because twins not aged between 25-30 called to ask if their spouse could complete a survey.

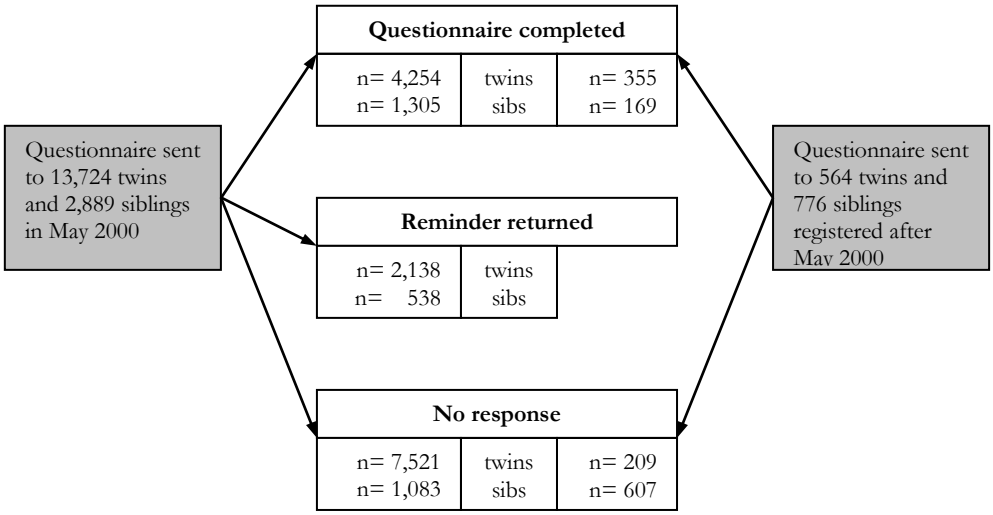


Figure 1.2 Overview of the number of participants for the 2000 survey. In total, questionnaires were completed by 4,609 twins (4,254 + 355), 1,474 siblings (1,305 + 169) and 708 spouses (not shown in figure). The reminder was returned by 1.949 twins and 501 siblings who were not willing to complete the questionnaire. Finally, 7730 twins and 1690 siblings did

not return the questionnaire or the reminder (for different reasons like moved to another address, not interested anymore, died or reason unknown).

The overall response rate for the fifth questionnaire was 32.3% for the twins and 40.2% for the siblings. Some participants were registered since the beginning of the study in 1991 by asking city councils in The Netherlands for addresses of twins aged 13-22 years, while other participants registered themselves in the past years. Table 1.2 shows the response rates for the different groups. For the twins the response rate is highest for newly registered twins (63 %), moderately high for twins who had already completed at least one other questionnaire in the longitudinal study (47.3 %) and low for twins who were registered between 1991 and 1997 but had returned none of the previous questionnaires in the longitudinal study (17.4 %). Response rates were highest (51.6%) for the siblings who had already returned a questionnaire before 1998, that is the 1995 and/or 1997 survey. Response rates were low for siblings who were registered before 1998 but had not returned the previous questionnaires (25.7%). Response rates were also low for the newly registered siblings (23.4%). However, large differences were found between newly registered siblings of families who were already registered before 1998 (19.8%) and newly registered siblings of families who were registered from 1998 onward (49.4%). The first group is registered by asking the parents for addresses of additional siblings while the second group registered themselves.

Table 1.2. Response rates of the 2000 survey for twins and siblings. The group of participant registered before 1998 was divided in participants who completed one or more earlier surveys and participants who did not complete one or more earlier surveys (surveys were sent to twins in 1991, 1993, 1995 and 1997 and to siblings in 1995 and 1997).

	Twins		Siblings	
	n returned q / n total q sent	Response rate (%)	n returned q / n total q sent	Response rate (%)
Registered before 1998, not completed other surveys	1307 / 7501	17.4	221 / 859	25.7
Registered before 1998, completed at least one other surveys	2941 / 6214	47.3	1091 / 2115	51.6
Registered after 1997	361 / 573	63.0	162 / 691	23.4
Total	4609 /14288	32.3	1474 / 3665	40.2

To maximize the response rate, a reminder was sent to all individuals (n=11167) who had not returned a questionnaire in July 2000. The reminder contained a reply card with a question on the willingness to participate. The reply card was returned by 32% of the individuals (n=3596). Most individuals who returned the reply card were not willing to participate in the survey study (Table 1.3). Almost 70% of the subjects who promised to complete the survey actually returned it. Of the subjects who asked for a new survey 51% returned a completed questionnaire and surprisingly, 1% of the subjects who indicated they would not complete a survey this time have send back a

completed booklet. A small group returned the reminder without providing an answer but 57% of those subjects completed the 2000 survey.
Because smoking was an important theme in the 2000 survey, the reply-card also contained a question on current smoking status (smoker /ex-smoker /non-smoker). In total, 2676 individuals returned the reply card but did not complete the 2000 questionnaire, and of those 2676 individuals, 2473 individuals answered the question on smoking. Those data were used to investigate whether a response bias occurred for smoking, as described in chapter 2.

Table 1.3. In total, 3596 reminders were returned. The reminder contained a question on their willingness to participate. In the second column the number of individuals who returned a reminder for each answer category is shown (n reminder). The last column shows the number of individuals who completed the 2000 survey within each answer category (n 2000 survey).

Answer on reminder:	N reminder	N 2000 survey
I will complete the questionnaire soon	742	520
I did not receive/lost the questionnaire, please send me one	540	273
I'm not willing to complete a questionnaire this time	2132	23
No answer	182	104
Total	3596	920

Table 1.4 Cross-sectional participation (1991, 1993, 1995, 1997, 2000). n.i. denotes that Ss were not invited to participate

	1991	1993	1995	1997	2000
Fathers	1439	1775	1572	n.i.	n.i.
Mothers	1607	1921	1688	n.i.	n.i.
Twins	3386	4227	3413	3234	4610
Siblings	n.i.	n.i.	1481	1518	1474
Spouses	n.i.	n.i.	n.i.	n.i.	708
	6432	7923	8154	4752	6792

Table 1.5 Longitudinal participation (1991, 1993, 1995, 1997, 2000):

	Twins	Siblings	Fathers	Mothers	Spouses	Total
1x	3237	1207	747	785	708	6684
2x	1719	943	1016	1097	-	4775
3x	1491	460	670	746	-	3367
4x	1213	-	-	-	-	1213
5x	574	-	-	-	-	574
	8234	2610	2433	2628	708	16613

Overview of longitudinal participation 1991 - 2000
Twin pairs were invited to participate in all waves of data collection. Parents were invited in 1991, 1993 and 1995, siblings were included in the assessments in 1995, 1997 and 2000, and spouses only in 2000. Table 1.4 shows the number of participants for the five surveys. Most individuals participated more than once, which is shown in

Table 1.5. Not all individuals are registered since the beginning of the study, and therefore have not had the opportunity to reach the maximum number of participations (5 times for twins, 3 times for parents and siblings). In total, 16613 individuals participated in the study. Table 1.6 shows the distribution of the birth cohorts for all individuals who participated at least once in the survey study (1991, 1993, 1995, 1997 or 2000). Year of birth is unknown for 0.3% of the participants. In 2000, most twins (71%) were aged between 20 and 30 years old.

Table 1.6. Representation of the total sample of twins, siblings, parents and spouses in different birth cohorts.

Born between	Twins		Siblings		Parents		Spouses		Total	
	N	%	N	%	N	%	N	%	N	%
1980-1988	703	8.5	276	10.6	-	-	9	1.3	988	6.0
1970-1979	5845	71.0	1179	45.2	-	-	455	65.5	7479	45.1
1960-1969	795	9.7	748	28.7	4	0.1	218	31.4	1765	10.7
1950-1959	497	6.0	218	8.4	1195	23.7	13	1.9	1923	11.6
1940-1949	252	3.1	122	4.7	3174	63.1	-	-	3548	21.4
1930-1939	117	1.4	49	1.9	606	12.0	-	-	772	4.7
<= 1929	24	0.3	15	0.6	53	1.1	-	-	92	0.6
	8233	100	2607	100	5032	100	695	100	16567	100

Table 1.7 Overview of the smoking data that were collected in the longitudinal twin-family study on health, personality and lifestyle in 1991, 1993, 1995, 1997 and 2000.

	1991	1993	1995	1997	2000
Self-report					
Ever smoked	x	x	x	x	x
Smoked last 12 months	-	x	x	-	-
Smoked last 4 weeks	-	x	x	-	-
Frequency	-	x	x	x	x
Quantity	x	x	x	x	x
Age of onset	x	x	x	x	x
Craving	-	-	-	-	x
Fagerström Test for Nicotine Dependence	-	-	-	-	x
Number of quit attempts	-	-	-	-	x
Report on others					
Smoking status parents	x	x	x	x	x
Smoking status co-twin	x	x	x		-
Smoking status siblings	-	x	x	-	-
Smoking status friends / best friend	-	x	x	-	-
Smoking status spouse	-	-	-	-	x

Smoking data collected in the survey study

The survey study is a large scale longitudinal study. Data on smoking were collected in all waves. Table 1.7 gives an overview of the smoking variables that were assessed at the five waves (1991, 1993, 1995, 1997 and 2000). The participants not only answered

questions on their own smoking behavior but also on the smoking behavior of their family members and friends.
The percentage current smokers, ex-smokers and never-smokers in the 2000 sample are presented in figure 1.2a (women) and 1.2b (men). In general, the percentage smokers is low in the younger age-groups and increases with age. Of the subjects aged 15 years and older, 23.8% of the women were current smokers and 29.8% of the men were current smokers. Those percentages are slightly lower than the percentages reported by STIVORO for the Dutch population (n=18.212). STIVORO reports that 27% of the Dutch women aged 15 years and older and 33% of the Dutch men aged 15 years and older are current smokers in 2001.

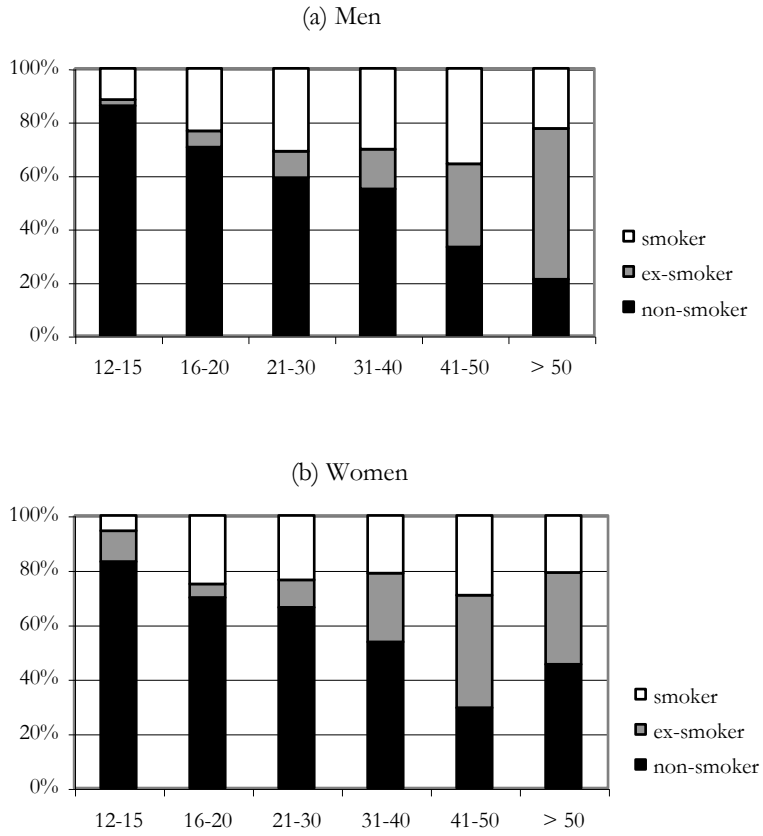


Figure 1.2 The percentage current smokers, ex-smokers and never-smokers in the 2000 survey: (a) men aged 15 years and older and (b) women aged 15 years and older.

Telephone interviews

In 2001, additional data on smoking behavior including nicotine dependence as measured by the FTND were obtained by telephone interviews (appendix 3).

Additional family members were called when only 1 family member participated in the questionnaire study and participants of the 2000 survey were called to obtain data on the stability of the FTND. In total, telephone interviews were completed for 110 individuals (56 twins and 54 siblings), most individuals (n=79) also completed the 2000 survey.

NETSAD study
The survey data of 1991, 1993, 1995 and 1997 were used to select the most informative families for a linkage study of anxious depression (Boomsma *et al.*, 2000). For each individual a genetic factor score for anxious depression was computed based on a genetic multivariate analysis of anxiety, depression, neuroticism and somatic anxiety. If at least two offspring formed an extremely concordant or discordant sibling pair for anxious depression, the entire family, including parents and all additional siblings, were asked for a DNA sample for genotyping (MZ twin pairs were treated as a single offspring). As some families consisted of more than two siblings, this selection procedure resulted in a non-random sample from the entire empirical distribution, not merely from its tails. In total, 2624 individuals from 563 families were approached and 1975 individuals from 479 families (643 parents and 1332 offspring) returned a buccal swab for DNA isolation. DNA was isolated the section Molecular Epidemiology of the Leiden Universitary Medical Centre (headed by prof Dr. E. Slagboom). For 917 individuals a complete genome scan was carried out by the Centre for Medical Genetics in Marshfield (research.marshfieldclinic.org/ genetics/). These genotype data were used for a linkage analyses of smoking initiation and quantity.

NETSMOK study
Families were selected for genetic linkage and association studies of smoking behavior based on data from the longitudinal survey study (1991, 1993, 1995, 1997 and 2000). Based on their answers twins and siblings were classified as 'nicotine dependent' (ND) or 'light smoker' (LS). A person was classified as ND when the FTND score was 6 or higher (FTND was only included in 2000 survey), when cigarette consumption was more than 20 cigarettes per day (based on answers in the 5 surveys), when smoking continued during pregnancy (women only, 2000 survey) or when the subject had tried to quit smoking more than 3 times (2000 survey). A subject was classified as LS when the person ever smoked or tried smoking but never smoked more than 5 cigarettes per day (based on answers in the 5 surveys). In total, 10.844 twins and siblings participated at least once in the longitudinal survey study, smoking data were available for 10.584 participants (8.089 twins and 2494 siblings) from 4.392 families. Of those, almost 10% was classified as ND (n=1.040) and 37% was classified as LS (n=5.666). The remaining 53% of the sample was unclassified and included both never smokers and participants who smoked more than 5 cigarettes per day but less than 20.
Families were selected when at least 2 siblings were both ND or at least 1 siblings was LS and 1 sibling was ND (MZ twin pairs were treated as single offspring). Using these criteria 388 families were selected for the linkage study and the entire family (twin, siblings and parents) was invited to participate in the study (appendix 4 and 5). In

total, 1.933 individuals were asked to provide a buccal swab for DNA isolation in the period January to July 2002 (appendix 6 and 7). Response rate was approximately 52%; a buccal swab was returned by 1.014 participants from 303 families. The individuals who participated in this study also completed an informed consent (appendix 8) and a questionnaire (appendix 9) on their smoking behavior, including the FTND. DNA was isolated from the buccal swabs at the section Molecular Epidemiology of the Leiden Universitary Medical Centre and the DNA samples of 1008 participants were sent to Prof. Dr. K. Kendler and colleagues of the Common Wealth University of Richmond for assessment of candidate genes.

Table 1.8 Overview of the papers presented in this thesis.

Ch	Main issue	Main Method	Data	Participants
2	Response bias	Comparison cooperative and non-cooperative families (χ^2 and anova)	Survey 2000 (incl. reminder)	Twins, siblings
3	Familial association smoking	Relative Risk ratio	Survey 1991, 1993, 1995, 1997 and 2000	Twins, siblings, parents, spouses
4	Predictors uptake of regular smoking	Logistic regression	Survey 1993 and 1995	Twins, siblings, parents
5	Fagerström Test for Nicotine Dependence	Chronbach alpha, test-retest correlations	Survey 2000, NETSMOK	Twins, siblings, spouses
6	Heritability smoking initiation and nicotine dependence	Modelfitting with Mx (liability models)	Survey 1991-2000, telephone interview, NETSMOK	Twins
7	Gene finding strategies	Overview literature	-	-
8	Linkage for smoking initiation and quantity	Sib pair QTL analyses in Mx	Survey 1991-2000, NETSAD	Twins, siblings

3. Outline of the thesis

In this thesis data on smoking behavior of twins and their family members of the Netherlands Twin Registry are analyzed to unravel the etiology of individual differences in smoking behavior. Chapter 2 evaluates whether a response bias occurred using a new approach: when the variable of interest has a familial component, data from respondents can be used as proxy for the data from their non-responding family members. In chapter 3 it is investigated to what extent smoking behavior of family members (parents, older/ younger/ same-age siblings) and friends (most friends, best friend, spouse) elevates an individual's risk to smoke. The study described in chapter 3 was cross-sectional. The goal of the study described in chapter 4 was to analyze whether the variables that are cross-sectionally associated with smoking behavior also predict the uptake of regular smoking over a two-year period.

In addition to smoking behavior of family and friends, other factors may influence the uptake of regular smoking. Therefore, we explored whether other factors add to the prediction of regular smoking.
When the uptake of smoking is established some individuals become nicotine dependent and others do not. To investigate the extent of dependence the Fagerström Test for Nicotine Dependence was used. In Chapter 5 the reliability of this test is described for both smokers and ex-smokers. Nicotine dependence can only be assessed in individuals who have initiated smoking but not every person who initiates smoking becomes dependent. Chapter 6 explores whether smoking initiation and nicotine dependence are part of the same continuum or whether they represent two independent dimensions. After identification of the correct model, the relative contribution of genetic and environmental factors to initiation and nicotine dependence is estimated. The next step after obtaining evidence for significant heritability is to identify chromosomal regions involved in smoking behavior, either by linkage or association approaches. The different approaches are described in chapter 7. In chapter 8, the linkage results from a complete genome scan on smoking initiation (ever/never smoked) and number of cigarettes smoked per day in a sample of dizygotic (DZ) twin and sibling pairs are reported. Table 1.8 presents a summary of the main issues, methods, data and participants of the papers in this thesis.

Chapter 2

Estimating non-response bias in family studies: application to mental health and lifestyle

Jacqueline M. Vink, Gonneke Willemsen, Janine H. Stubbe, Christel M. Middeldorp, Rozemarijn S.L. Ligthart, Kim Baas, Hanneke Dirkzwager, Eco de Geus, Dorret I. Boomsma (2004) Estimating non-response bias in family studies: applications to mental health and lifestyle. European Journal of Epidemiology 19(7); 623-630

Abstract
Non-response to mailed surveys reduces the effective sample size and may introduce bias. Non-response has been studied by 1) comparison to available data in population based registers, 2) directly contacting non-respondents by telephone or single-item reply cards, and 3) longitudinal repetition of the survey. The goal of this paper was to propose an additional method to study non-response bias: when the variable of interest has a familial component, data from respondents can be used as proxy for the data from their non-responding family members. This approach was used with data on smoking, alcohol consumption, physical activity, coffee- and tea-use, education, body mass index, religion, burnout, life events, personality and mental health in large number of siblings and DZ-twins registered with the Netherlands Twin Register. In addition, for smoking behavior, we also used the second strategy by sending a reply card.
Results show that scores of members from less cooperative families or incomplete twin pairs tended to be more unfavorable than the scores from highly cooperative families or complete twin pairs. For example, family members from less cooperative families cycled less often and scored higher on anxious depression and neuroticism. For smoking, both the results of the reply card and the results of the additional method suggested a higher percentage smokers among the non-respondents but this was only significant with reply card method. In general, differences between highly/less cooperative families and complete/incomplete DZ-twins were small. Results suggest that, even for studies with moderate response rates, data collected on health, personality and lifestyle are relatively unbiased.

Introduction

Mailed surveys are widely used to collect data on health and lifestyle in large populations. In Europe, response rates to mailed surveys vary from 52 to 95 percent (Hupkens *et al.*, 1999). Non-response to mailed questionnaires reduces the effective sample size and therefore the statistical power of the study. Moreover, survey results will be biased by non-participation if refusal to participate is not distributed randomly, and is either directly or indirectly related to the traits under study. Although studies usually recognize the risk of response bias, they are often unable to quantify the degree of bias.

Studies quantifying response bias may use different methods to obtain information on the non-respondents. First, when access is available to population based registers like health insurance databases, utilization databases or population registers, it is possible to compare respondents with non-respondents with regard to the information provided by the registers. In general, studies using this method (Bergstrand *et al.*, 1983; Etter and Perneger, 1997; Reijneveld and Stonks, 1999) have shown differences between non-respondents and respondents; for example, non-respondents had lower annual incomes, more sickness benefit days and were more often unmarried. A drawback of this method is that the response bias can only be examined with regard to the available- often rather general- characteristics in population registers and can not indicate the degree of response bias regarding the more specific characteristics of interest in comprehensive survey studies.

A second method to quantify response bias is obtaining this specific information by contacting the non-respondents themselves either by telephonic interview or by sending a reply card. A study that used a telephonic interview to obtain information on the non-respondents showed statistically significant differences between respondents and non-respondents for smoking status, hazardous alcohol consumption and lack of vigorous activity (Hill *et al.*, 1997; Barchielli and Balzi, 2002). Although such a telephonic interview provides valuable information on non-response, there will always remain a group of non-respondents who either can not be reached by phone or will be unwilling to participate.

Longitudinal studies provide a third source of information on non-response by allowing the comparison of respondents and non-respondents at later follow-up, using the information obtained at the start of the study. Most of those studies have found small or no differences between respondents and non-respondents (Jacobsen and Thelle, 1988; Macera *et al.*, 1990; Heath *et al.*, 2001; Loon *et al.*, 2003). Subjects who repeatedly returned a questionnaire tended to be married, non-smokers and more physically active than those who returned it only once. However, a possible problem with these studies is that they are not based on random samples; the original study population at the first measurement itself may already have been a selected sample.

Here we propose an additional method to obtain information on non-respondents which is based on family and/or twin designs. We will concentrate on general demographic (education), lifestyle variables (smoking, alcohol use, physical activity, coffee- and tea-consumption, religious practice) and personality/mental health (body mass index, burnout, problem behavior, neuroticism). These variables are familial, that

is family members resemble each other for those characteristics (Koopmans *et al.*, 1994; Koopmans and Boomsma, 1996; Aarnio *et al.*, 1997; Boomsma *et al.*, 1999; Eaves *et al.*, 1999; Geus De *et al.*, 2003; Vink *et al.*, 2003a; Vink *et al.*, 2003b). Therefore, data from non-respondents will be correlated with the data from the respondents and data from responding family members thus will offer information on the non-respondents. We illustrate this approach with data of twins and their siblings collected in 2000 in a survey study on health, personality and lifestyle of the Netherlands Twin Register.

Acting on the idea that health, lifestyle and personality of the non-responding subjects is reflected by the values obtained on the responding family members we first investigate whether the answers on health, lifestyle and personality variables are different for siblings from highly cooperative families (more than 80 percent of the family members participated) compared to siblings from less cooperative families (less than 80% of the family members participated). With this method the non-response bias is estimated using information of the responding family members. In addition, we compared data from DZ twins from complete pairs (both twins completed a questionnaire) with data from DZ twins from incomplete twin pairs (co-twin did not participate in the survey study). Data of dizygotic (DZ) twins were used because DZ twins share on average 50 percent of their genes, just like siblings. However, DZ twins are a select group and may have some distinct features in common which are not generalized to a singleton population (e.g. same age).

Methods

Participants

This study is part of an ongoing twin family study on health-related behavior of the Netherlands Twin Register (NTR) that assesses families with adolescent and (young) adult twins every two/three years since 1991 (Boomsma *et al.*, 2002). For the present study, the data from the 2000 survey were used.

In May 2000, questionnaires were sent to 13724 twins/triplets and 2889 siblings. In July 2000 a reminder was send to the non-respondents. Because smoking was an important theme in the 2000 survey, the reminder contained a pre-stamped reply-card with a question on their smoking status (smoker /ex-smoker /non-smoker). The reply card was returned by 2676 persons (2138 twins and 538 siblings) who were not willing to complete the questionnaire. The question on smoking behavior on the reply card was answered by 2473 of the 2676 non-respondents.

Twins and siblings who registered after May 2000 also received a questionnaire (n=564 twins and n=776 siblings) but not a reminder. Twins registered themselves, while most siblings were recruited by asking their mother for their addresses. In total, questionnaires were send to 14288 twins/triplets and 3665 siblings from 7223 families. The average family size of the families that were invited to complete a questionnaire was 2.48 (SD 0.99). At the end of the data collection, 4609 twins/triplets and 1474 siblings from 3178 families had completed a questionnaire booklet (Figure 1.2). For the same sex twins, zygosity was based on questionnaire data or, when available, on DNA typing (zygosity based on DNA was available for 26.1% of the same sex twins).

Agreement between zygosity based on questionnaire data and zygosity based on DNA results was 98 percent. For the opposite sex twin pairs, zygosity is known (DZ) based on their sex. The average family size was 1.91 (SD .94). The triplets (41 persons from 22 families), the half-siblings (n=27) and adoption siblings (n=5) were excluded from the analyses.

Data analyses

The percentage smokers were compared for respondents and non-respondents using χ^2-square tests.

Familial correlations were calculated for all dependent variables. For the categorical traits tetrachoric correlations were calculated with a treshold model on raw data using MX (Neale *et al.*, 1999). The correlations between DZ twins and siblings were constrained to be equal to estimate the familial correlation. For the continuous variables intraclass correlations were calculated from an ANOVA analyses using all DZ twins and sibships (Falconer and Mackay, 1996a).

Data from respondents of highly cooperative families were compared with data from respondents of less cooperative families. For each family, the number of respondents was divided by the number of family members who were asked to complete a questionnaire. When less than 80 percent of the family members participated, the family was marked as 'less cooperative family' and when 80-100 percent of the family members participated the family was marked as 'highly cooperative family'. The dataset contained 1,099 families with at least 1 additional sibling. For some of these families more than one additional siblings participated (on average 1.3 sibling per family). From each family one sibling who completed a questionnaire was chosen and scores of the respondents of highly cooperative families (n=444) were compared to the scores of the respondents of less cooperative families (n= 655).

The mean age of the individuals from highly cooperative families was 30.2 (SD 11.7) and 31.2 (SD 9.8) for the individuals from less cooperative families. Furthermore, 39% of the individuals from incomplete pairs were males and 43% of the individuals from complete pairs were males.

In addition, data from twins from complete DZ twins pairs (both twins completed a questionnaire) were compared with data from twins from incomplete DZ twin pairs (their co-twin did not participate in the questionnaire survey). Data of the monozygotic (MZ) twins were excluded. MZ twins are a special group of siblings who, in contrast to DZ twins or singleton siblings, are genetically identical. When a trait is influenced largely by genetic influences, a MZ twin will be more similar than a DZ twin or singleton sibling. To be able to use the results of this study as example for other large family studies (without twins) we selected the DZ twins and singleton siblings for those analyses. DZ twins share on average 50 percent of their genes just like singleton siblings, but have, in contrast to other siblings, exactly the same age and are more likely to share similar environmental influences. The DZ twin sample contained 1,498 individuals from complete twin pairs and 772 individuals from incomplete twin pairs. The mean age of individuals from incomplete DZ twin pairs was one year lower than the mean age of individuals from complete DZ twin pairs

(respectively, 28.8 (SD 8.9) and 29.8 (SD 11.4), p=0.022). Furthermore, 33.6% of the individuals from incomplete pairs were males and 32.8% of the individuals from complete pairs were males (p=.339).
For both comparisons statistical significance was assessed by χ^2 test for categorical variables, by Mann-Whitney test for ordinal variables and by ANOVA for continuous variables. Because multiple comparisons were performed we considered the chance of a type I error. To protect against this error, a Bonferroni correction was used by dividing the significance level by the number of comparisons (.05/28 = .002). For a comparison to be considered significant it must have a significance level of .002 or less. Statistical analyses were performed using SPSS 11.0 for windows.

Variables
The following variables obtained in the questionnaire were explored in this study:
Smoking: "Did you ever smoke?" was recoded to ever smoked (yes) and never smoked (a few times to try, no). "How often do you smoke now?" was recoded to non-smokers (never smoked, never smoked regularly, quitters) versus smokers (I smoke once a week or less, I smoke more than once a week but not every day, I smoke daily). "Do you think you'll smoke next year?" (definitely not, probably not, I don't know, probably yes, definitely yes) was analyzed as numerically ordered variable.
Alcohol consumption: "Have you ever drunk alcohol?" was recoded to ever drunk alcohol (yes) versus never drunk alcohol (a few times to try, no). "How often do you drink alcohol?" was recoded to more than once a week (once a week, several times a week, daily) versus less than once a week (I do not drink alcohol, once a year or less, a few time a year, once a month). When at least 2 questions were answered with 'yes' on the CAGE (a 4 item questionnaire to detect alcohol problems) the person was classified as possibly having alcohol problems (yes, no).
Coffee- and tea-use: How many cups of coffee/ tea do you drink a day? (number of cups/ day).
Physical activity: "Do you participate in sports regularly?" (no, yes) and "Do you cycle regularly?" (no, yes). "During the last 6 months, how often have you been physical active for more than 20 minutes?" (never, less than once a month, once a month, 2-3 times a month, 1-2 times a week, 3 times a week or more often) was analyzed as numerically ordered variable.
Religion: "Are you an active member of a religious communion?" (yes I am an active member, I am religious but not a member of a religious communion, no I am not religious) was recoded to 'religious practice' (yes for active members versus no for non-members and not religious persons).
Burnout: A log transformation was used on the scores of the Dutch version of a five-item subscale (emotional exhaustion) of the Maslach Burnout Inventory-General Survey (Schaufeli *et al.*, 1996).
Body mass index: A combined measure of height and weight: weight in kg / (height in m)^2.
Education: Subjects were divided in 3 groups: low, medium and high education.

Personality: The subscales neuroticism, somatic anxiety, test attitude and extraversion subscales from the Amsterdamse Biografische Vragenlijst (ABV) (Wilde, 1970), and the subscales anxious depression, withdrawn, somatic complaints, thought problems, attention problems, intrusive behavior, aggressive behavior, rule breaking behavior of the Young Adult Self Report (YASR) (Achenbach, 1997) translated and validated for the Dutch population by Verhulst et al. (Verhulst *et al.*, 1997).

Results

The overall response rate for the 2000 survey was 32.3 percent for the twins and 40.2 percent for the siblings. Our database consisted of Dutch twin families which were recruited in different ways; a large part of the twin families were recruited by asking city councils in the Netherlands for addresses of twin-families, while other twins volunteered to register throughout the study period. The response rates for the survey varied across various subsets of twins. Newly registered twins who volunteered to register and twins who have participated in other waves were more likely to complete a questionnaire. It is important to note that addresses in our database were most up to date for those groups.

Estimating the non-response bias for smoking using a reply card to the non-respondents

To investigate whether a response bias occurred for smoking, the smoking status of the non-respondents was compared with the smoking status of the respondents. Smoking status was available for 6,016 respondents (4,566 twins and 1,450 siblings), as answers on smoking behavior were missing or contradictory for 43 twins and 24 siblings. Smoking status was also available for 2,473 non-respondents (1,971 twins and 502 siblings) who returned the reply card. As shown in table 2.1, the percentage smokers was higher in the non-respondents groups compared to the respondents-groups. A χ^2-square test showed that those differences were significant (p=.000 for men and p=.016 for women).

Table 2.1 Number and percentage smokers for non-respondents of the 2000 survey (1949 twins and 501 siblings) and respondents of the 2000 survey (smoking status was known for 4566 twins and 1450 siblings).

	Twins			Siblings		
	N smok	N total	% smok	N smok	N total	% smok
Male respondents	445	1,505	29.6	148	580	25.5
Male non-respondents	377	995	37.9	90	268	33.6
Female respondents	730	3,061	23.8	196	870	22.5
Female non-respondents	256	954	26.8	61	233	26.2

Familial correlations for mental health and lifestyle variables

To explore to what extent the variables are familial, the intraclass correlations (continuous data) and tetrachoric correlations (categorical data) were calculated. The lifestyle variables like smoking, alcohol consumption, physical activity and coffee- and tea-use showed familial correlation ranging from .21 to .44. A high familial correlation

was found for religious practice (.77). The personality and mental health variables showed a somewhat lower familial correlation ranging from .11 to .20 (Table 2.2). All correlations were statistically significant.

Table 2.2 Familial correlation between DZ twins and singleton siblings. For the categorical variables the tetrachoric/polychoric correlations are calculated. For the continuous data the intraclass correlations are presented.

	Familial correlation	p-value
Ever smoked (yes /no)	.44	.000
Current smoker (yes /no)	.33	.000
Smoke next year? (5 categories)	.26	.000
Ever tried alcohol (yes/no)	.39	.000
Regular alcohol use: > than once a week (yes /no)	.33	.000
Possible alcohol problems CAGE (yes /no)	.21	.000
Coffee-use (mean n of cups a day)	.38	.000
Tea-use (mean n of cups a day)	.25	.000
Regular sports participation (yes /no)	.25	.000
Regular cycling (yes /no)	.21	.000
Physical activity (6 categories)	.13	.000
Body mass index	.38	.000
Education (3 categories)	.45	.000
Actively religious (yes /no)	.77	.000
Burnout	.17	.000
N life events in past 5 years	.26	.000
Anxious depression	.20	.000
Withdrawn	.12	.000
Somatic complaints	.16	.000
Thougth problems	.11	.000
Attention problems	.15	.000
Intrusive	.11	.000
Aggressive behavior	.20	.000
Rule breaking behavior	.12	.000
Neuroticism	.19	.000
Somatic anxiety	.18	.000
Test attitude	.18	.000
Extraversion	.13	.000

Estimating the non-response bias using information of the co-twin and siblings

The scores of 655 siblings from a less cooperative family (less than 80 percent of the family members participated) were compared with scores of 444 siblings from a highly cooperative family (80-100 percent of the family members participated). The scores for lifestyle variables like smoking, alcohol consumption, physical activity, coffee- and tea-use and personality scores seemed more unfavourable for individuals from less cooperative families, but after Bonferroni correction the differences were not significant (Table 2.3).

The DZ twin sample consisted of 772 individuals from incomplete pairs and 1.498 individuals from complete twin pairs. The scores of the incomplete DZ-twins were somewhat more unfavourable than the scores of the complete DZ-twin pairs (Table

2.3). Individuals from incomplete twin pairs cycled significantly less often than individuals from complete twin pairs. They also had significantly higher scores for anxious depression, somatic complaints and neuroticism. The other variables (smoking behavior, alcohol consumption, physical activity, coffee- and tea-use, burnout score, body mass index and education) showed the same trend, namely a more unfavourable score for the individuals from incomplete pairs compared to individuals from complete pairs, but differences were not significant (after Bonferroni correction).

Table 2.3 Comparison between individuals from less-cooperative families (<80% of the family members that were asked to complete a questionnaire participated) and highly cooperative families (80-100% of the family members that were asked to complete a questionnaire participated) and comparison between twins from incomplete DZ twin pairs (co-twin did not complete a questionnaire) and DZ twins from complete twin pairs (both twins completed a questionnaire) . I= values of the respondents from incomplete twin pairs, C= values of the respondents from complete twin pairs. L= values of the respondents from less cooperative families, H= values of the respondents from highly cooperative families. Comparisons are significant if $p < .002$ (Bonferroni correction).

	Siblings			DZ-twins		
	L	H	p-value	I	C	p-value
% ever smoked	46.2	42.4	.222	48.2	44.5	.096
% smoker	24.8	22.1	.313	30.7	27.2	.081
Smoke next year?	-	-	.717	-	-	.048
% ever tried alcohol	92.6	90.7	.257	92.3	90.0	.066
% regular alcohol use (> than once a week)	42.2	40.7	.638	37.1	43.2	.005
% more than once been drunk	56.0	52.2	.221	55.6	51.0	.041
% alcohol problems (CAGE)	12.7	8.6	.034	9.7	9.5	.850
Coffee-use (mean n of cups a day)	3.2	2.9	.323	2.6	2.8	.090
Tea-use (mean n of cups a day)	2.7	2.6	.104	2.6	2.4	.108
% sports participation	58.3	60.8	.408	57.4	59.2	.399
% regular cycling	63.0	65.8	.351	57.0	65.9	.000
How often > 20 minutes physical active?	-	-	.131	-	-	.240
Body mass index	24.0	23.6	.080	22.9	23.0	.677
Education	-	-	.455	-	-	.516
% religious (active)	24.2	24.0	.942	18.3	23.6	.004
Burnout (mean score)	93.7	94.6	.549	94.5	92.3	.056
N life events in past 5 years	1.11	1.01	.471	1.40	1.27	.011
Anxious depression (mean score)	5.5	5.0	.057	5.9	5.1	.000
Withdrawn (mean score)	2.9	2.6	.091	2.7	2.6	.049
Somatic complaints (mean score)	3.1	2.8	.092	3.1	2.7	.000
Thought problems (mean score)	.58	.51	.987	.58	.48	.097
Attention problems (mean score)	4.3	4.0	.081	4.5	4.1	.004
Intrusive (mean score)	3.0	2.7	.024	2.8	2.7	.178
Aggressive behavior (mean score)	6.0	5.5	.038	6.0	5.6	.039
Rule breaking behavior (mean score)	3.4	3.4	.933	3.4	3.1	.063
Neuroticism (mean score)	49.0	45.9	.038	52.0	47.7	.000
Somatic anxiety (mean score)	18.1	17.1	.002	18.1	17.5	.024
Test attitude (mean score)	37.0	37.5	.376	37.1	37.9	.045
Extraversion (mean score)	58.4	58.7	.708	60.4	60.0	.606

Discussion

The goal of this paper was to examine an alternative method for determining response bias in a survey on health, personality and lifestyle, using data from responding family members as a proxy for non-responding family members. Response rates are dependent on a large variety of factors, such as a monetary incentive, short questionnaires, personalized questionnaires or letters, stamped return envelopes, contacting persons before sending the questionnaire and providing non-respondents with a second copy of the questionnaire (Edwards *et al.*, 2002).

The overall response rate of 33.9 percent in our study may underestimate the actual response rate, as questionnaires were sent to everyone in our database, regardless of earlier participation. Probably, a substantial percentage of the non-respondents are moved to another address. At present, a study is carried out on a sample of non-respondents of a next survey (sent in November 2002 /March 2003) which will illustrate the percentage of non-respondents who moved to another address. The exact number of non-respondents due to change in address can best be determined by linking our address database to governmental address databases. However, at the time of writing, Dutch legislation does not allow for use of the official population register to check and update the addresses of our database. Stang (2003) concluded that we should not uncritically use the response proportion as an indicator of the likelihood of non-response bias because there is not always a connection between low response proportions and non-response bias (Stang, 2003). It is more important to investigate the severity of response bias. Response bias can be explored by different methods like 1) comparison to available data in population based registers, 2) directly contacting the non-respondents by telephone or single-item reply cards, and 3) longitudinal repetition of the survey.

For smoking, we used the second strategy by sending a pre-stamped reply card with a single question on current smoking status to our non-respondents. The results showed a significantly higher percentage smokers among the non-respondents compared to the respondent for both men and women. In most other studies current smoking was also more prevalent among non-respondents (Hill *et al.*, 1997; Kotaniemi *et al.*, 2001; Loon *et al.*, 2003) or late respondents (Korkeila *et al.*, 2001). The disadvantage of sending a reply card is that only a few questions can be asked. Furthermore, there is a group non-respondents who neither responds to the invitation to complete the survey nor returns the reply card.

The data collected in twins and their singleton siblings offered a unique opportunity to estimate lifestyle and mental health of non-respondents by the values of their responding family members. Correlations showed that the variables on lifestyle, personality and mental health are familial, that is family members resemble each other for those characteristics. Correlations were higher for lifestyle variables (like smoking, alcohol consumption and physical activity) than for the personality and mental health variables. The higher the familial correlation the better the value of the responding family members may be used as a proxy for the non-responding family members. However, all correlations were significant, and represent the lower bound of the

estimates as correlations between same-sex siblings and between siblings close in age are expected to be even higher.
The values of singleton siblings from less cooperative families seemed somewhat more unfavorable than the values of singleton siblings from highly cooperative families although those differences were not statistically significant after Bonferroni correction.
Data on DZ twins are of special interest. Those twins share on average 50 percent of their genes just like singleton siblings, but have, in contrast to other siblings, exactly the same age. Results showed that individuals from incomplete DZ twin pairs have more unfavourable scores than individuals from complete DZ twin pairs which suggests that the non-responding co-twin is also likely to have an unfavourable lifestyle or lower mental health.
For smoking, both the results of the reply card and the results of the comparison between complete/incomplete twin pairs or highly/less cooperative families suggested a higher percentage smokers among the non-respondents. The differences found with the reply card were statistically significant while the results of the other method were not. Results on non-response bias are probably most trustworthy when collecting data of the non-respondents themselves (by telephone interview or reply-card). A limitation of the reply-card approach is the lack of information on completely non-cooperative subjects (subjects who did not respond to the invitation to complete the survey but also did not respond to additional requests for information). The approach to obtain information on non-respondents using the values of their responding family members offers additional information.
In conclusion, the specific composition of our database with twins and their singleton siblings, offered a unique opportunity to estimate lifestyle and mental health of non-respondents by the values of their responding family members. In general, results showed the scores of members from less cooperative families or incomplete twin pairs tended to be lower than scores from highly cooperative families or complete twins but differences were small. These results suggest that, even for studies with moderate response rates, data collected on health, personality and lifestyle are only mildly biased.

Chapter 3

The association of current smoking behavior with the smoking behavior of parents, siblings, friends and spouses

Jacqueline M. Vink, Gonneke Willemsen, Dorret I. Boomsma.(2003). The association of current smoking behavior with the smoking behavior of parents, siblings, friends and spouses. Addiction 98: 923-931.

Abstract
The aim of this study was to examine the association of current smoking behavior of adolescents and young adults with the smoking behavior of their parents, siblings, friends and spouses.
The study was performed using survey data from a large twin-family sample, the association between smoking behavior of participants and their family members, friends and spouses was investigated by calculating the relative risks (RR). To disentangle sex- and age-differences, calculations were carried out separately for males and females and for 3 different age groups: 12-15, 16-20 and 21-40 years old.
Results showed that the smoking behavior of the participants was significantly influenced by the smoking behavior of parents, siblings and friends, but all RRs decreased with age. No differences in RR were found between having older or younger smoking siblings. Within each age group, the RR to smoke when having a smoking friend was comparable to the RR to smoke when having a smoking same-age and same-sex sibling. For the older participants the RR to smoke was higher for monozygotic (MZ) twins with a smoking co-twin than for dizygotic (DZ) twins with a smoking co-twin. Most findings were sex-dependent: same-sex smoking family members influenced smoking behavior more than opposite-sex family members. The significant association of the smoking behavior of spouses decreased with age which suggests that assortment for smoking is based on similarity at the time dating began.
The results highlight the importance of both social and genetic influences on smoking behavior, with genetic influences increasing with the age of the participant.

Introduction

Using data from a large twin-family sample from the Netherlands Twin Register we explored whether the relative risk to be a smoker when having smoking parents, smoking siblings or smoking friends is different for young adolescents (aged 12-15 years) who are legally not allowed to buy tobacco, older adolescents (16-20 years) and adults (21-40 years). Several studies have shown associations between smoking behavior of adolescents and that of their family members and friends (Jensen and Overgaard, 1993; Wang *et al.*, 1995; Distefan *et al.*, 1998; O'loughlin *et al.*, 1998; Whiters *et al.*, 2000; Alexander *et al.*, 2001; Bauman *et al.*, 2001; Boyle *et al.*, 2001).

Most of these studies investigated smoking behavior in adolescents and young adults aged 12 to 18 years. Whether the effect of smoking family members or friends still influences smoking behavior after this age is not clear. West *et al.*, (1999) included 16 to 23 year olds and concluded that the period from late adolescence to early adulthood is still of importance for the uptake of regular smoking. In our study of smoking behavior adolescents and young adults were included and we hypothesized that, in line with earlier studies (Bauman *et al*, 2001, Wang *et al*, 1995), the relative risk to smoke when family or friends smoke will remain stable across all age groups.

When analyzing data from family members we distinguish between the influence of parents and the influence of siblings. In general, studies that include the smoking behavior of siblings have shown that having siblings who smoke increases an adolescents' risk of smoking two to fourfold (Jensen and Overgaard, 1993; Wang *et al.*, 1995; O'loughlin *et al.*, 1998; Moran *et al.*, 2000; Whiters *et al.*, 2000; Boyle *et al.*, 2001). Two studies concluded that having older smoking siblings increase an subjects' risk of being a smoker (Jensen and Overgaard, 1993, Wang et al, 1995). Although those studies have included older siblings, no studies have directly compared the association between subjects' smoking behavior and having younger or older smoking siblings. If adolescents and young adults imitate the smoking behavior of people in their environment we expect that the relative risk to smoke is higher when older siblings are smokers than when younger siblings are smokers. In contrast, if the association between the adolescents' smoking behavior and the smoking behavior of siblings is caused by genetic influences the relative risk to be a smoker should be equal for having younger or older smoking siblings.

The study of Wang *et al* (1995) included both siblings and friends and reported that having smoking friends formed a higher risk for adolescents' smoking than having smoking older siblings. This was also found by Jensen and Overgaard (1993) but not by Swan *et al.* (1990). However, siblings usually are a few years older or younger than the proband while friends are often more similar in age. Inclusion of same-age siblings, i.e. twins, may solve a problem regarding the comparison of sibling and friends. In this study, data are available for dizygotic twins, who share the same proportion of genes as other siblings. We investigated whether the influence of friends is higher than the influence of same age siblings by comparing the relative risk to smoke when having a smoking same age sibling, i.e. dizygotic twins, with the relative risk to smoke when having a smoking best friend.

If age differences are important for influencing smoking behavior it is also likely that parents who smoke have a lower impact on smoking behavior than siblings who smoke. Generally, studies have found a rather low but significant influence of parental smoking on the smoking behavior of adolescents (Green *et al.*, 1991; Meijer *et al.*, 1996; Distefan *et al.*, 1998; Boyle *et al.*, 2001). The risk ratios in those studies ranged from 1.6 to 2.1 when both parents smoked and from 1.4 to 2.2 when at least one parent smoked. Therefore we expect low relative risks for smoking when parents smoke. On the other hand, the genetic resemblance of parents and offspring is the same as among full siblings. So if genetic similarity explains familial resemblance, no differences between the relative risk to smoke when having smoking parents and the relative risk to smoke when having smoking siblings is expected, unless different genes are expressed in the parental and the offspring generation.

In contrast to the dizygotic (DZ) twins who share on average 50% of their genes, monozygotic (MZ) twins are genetically identical. From twin studies, it has been well established that there is a familial aggregation for smoking behavior and that a large part of this aggregation is due to genetic factors (Heath and Madden, 1995; Sullivan and Kendler, 1999). The genetic influence on smoking behavior is likely to depend on the age composition of the sample. Genetic factors seem less important in younger age cohorts than in older cohorts (Heath and Madden, 1995). So we hypothesize that the differences between the risk to smoke when having a MZ smoking co-twin and having a DZ smoking co-twin increases with age.

The effect of smoking family members and friends may differ for men and women. Wang *et al.* (1995) found higher odds ratios for adolescents' smoking behavior for same-sex smoking siblings or friends than for opposite-sex smoking siblings or friends. In the present study is explored whether the relative risk to be a smoker when having smoking brothers or sisters is different for males and females (e.g. brother-brother versus brother-sister).

Similarly, studies have reported that maternal, and not paternal, smoking significantly influenced adolescent smoking (Hover, 1988; Brenner and Scharrer, 1996; Herlitz and Westholm, 1996) although the opposite has also been reported (Shamsuddin and Abdul Harris, 2000). Only a few studies have included both the sex of the parent and the sex of the participant (Swan *et al.*, 1990; Wang *et al.*, 1995). The study of Wang *et al.* (1995) did not show significant influences of smoking mothers or smoking fathers on their sons and daughters aged 14-18 years while the study of Swan *et al.* (1990) showed a significant risk to smoke for females when mother smoked, but not when father smoked. For males, smoking behavior was not associated with that of either parent. Because the two studies that investigated the association between smoking father and mother separately for males and females have found different results, the present study also investigated the relative risk to smoke for males and females in relation to maternal and paternal smoking.

Spouses or long-term partners form a special group of 'best friends' who may have a significant influence on each other's smoking behavior. (Price and Vandenberg, 1980) reported spouse similarity for several variables including current smoking. In a large survey among women in the U.S., the odds that a women's spouse was a current

smoker was 5.5 times greater if she was a current smoker than if she did not smoke (Ogden *et al.*, 1997). Moreover, in a Dutch study, the correlation between husband and wife for current smoking status (r=0.43) was larger than for smoking history (r=0.18), (Boomsma *et al.*, 1994a). In contrast, Graham and Der (1999) concluded that partner's smoking status was not a predictor of tobacco consumption among women (Graham and Der, 1999). Similarity between spouses could be due to several factors; assortment may be based on similarity at the time dating began or phenotypes may have converged during the years of marriage because of reciprocal influences or shared living conditions (Price and Vandenberg, 1980). Price *et al* (1981) suggested that for smoking the convergence of phenotype was the most likely explanation. To explore this theory we have a unique dataset of twins aged 25-30 years and their spouses as well as a dataset of the twins' parents aged 30-45 years and 46-65 years. Assuming that the duration of the relationship for the younger spouses is shorter than for the older spouses we would expect higher relative risks for the older age groups if convergence of phenotype is the correct explanation.
In short, this paper investigates if smoking behavior of family members (parents, older/younger/same-age siblings) and friends (most friends, best friend, spouse) elevates an individual's risk to smoke as a function of age and sex.

Participants and Methods

Longitudinal study on personality and health related behaviors:

In 1991 the Netherlands Twin Register (NTR) started a large-scale twin-family study on personality and health related behaviors. Addresses of twins were obtained from City Council Registries as described in Boomsma *et al.* (2000) and Koopmans *et al.* (1999). New twins are recruited throughout the study period. Since 1991, the twin-families received a questionnaire every 2 or 3 years. Twins were asked to participate every time (1991, 1993, 1995, 1997 and 2000), parents only in 1991, 1993 and 1995 and siblings only in 1995, 1997 and 2000 (Table 3.1). Sample selection and response rates are described in detail in (Koopmans *et al.*, 1994; Boomsma *et al.*, 2000). Some individuals participated once, whereas others participated on multiple occasions. If subjects participated more than once, the answers on the first questionnaire they completed were used in the analysis. For each relative risk calculation, data were only included if family members or spouses participated in the survey at the same time.
The surveys contained items on health, lifestyle and personality (e.g. alcohol consumption, smoking, physical exercise, general health and personality). To analyze smoking behavior in twins, their siblings, parents, friends and spouses, two questions were selected from the questionnaires. The first question was 'Did you ever smoke?'. This question had 3 answer categories: no, just a few times to try, yes. The second question was 'How often do you smoke now?' with the answer categories: I have never smoked, I quit smoking, I smoke less than once a week, I smoke more than once a week but not every day, I smoke daily. Based on the answers to these questions, participants were classified as current smokers or non-smokers (combining never smokers and ex-smokers). The classification was cross-validated using other questions like "How many cigarettes do you smoke per day?". To investigate the

influence of age, twins were divided in 3 age groups: 12-15 years (young adolescents, legally not allowed to buy tobacco in the Netherlands), 16-20 years (older adolescents) and 21-40 years (adults).

Table 3.1. Number of participants in the longitudinal twin-family study of the Netherlands Twin Register. Fathers = fathers of the twins, mothers = mothers of the twins, MZ = monozygotic twins, DZ = dizygotic twins, Zyg unkn = zygosity of twins unknown, siblings = additional siblings, besides the twins, spouses = spouses of twins.

	1991	1993	1995	1997	2000
Fathers	1483	1778	1573	-	-
Mothers	1623	1919	1685	-	-
MZ twins	1304	1576	1382	1371	2196
DZ twins	2091	2309	2022	1760	2380
Zyg unkn	-	-	11	12	35
Siblings	-	-	1482	1434	1475
Spouses	-	-	-	-	706
Total	6501	7582	8155	4577	6592

Twin-pairs: To calculate the relative risk to be a smoker when having a smoking co-twin, the data of the questionnaires of 1991 (1679 twin pairs), 1993 (984 new twin pairs), 1995 (5 new twin pairs), 1997 (450 new twin pairs) and 2000 (480 new twin pairs) were used. In total, 3598 twin pairs (=7196 subjects) were included in the analyses.
Twins and Siblings: Siblings of twins were asked to participate in the survey for the first time in 1995. To calculate the relative risk to be a smoker when having full siblings who smoke, the data of the twins and their siblings were used from the surveys in 1995 (1714 families), 1997 (618 new families) and 2000 (711 new families). In total, questionnaire data were available for 3043 families (7906 participants). Due to incomplete questionnaires, smoking status was known for 7828 individuals (twins and siblings). For 1501 families, no additional siblings (beside the twins) participated. For the other families, the mean number of additional siblings was 1.34 and the mean age difference with the twins was 3.5 years (SD 2.2). To determine the relative risk to be a smoker as a function of having older or younger siblings who smoke, we determined whether the selected participants (twins and siblings) did have older brother(s), older sister(s), younger brother(s) and younger sister(s) and if at least one of those siblings was a smoker.
Parents: The biological parents of the twins were included in the study in 1991 (1499 fathers and 1642 mothers), 1993 (908 new fathers and 987 new mothers) and 1995 (40 new fathers and 13 new mothers). The questions on smoking were answered by 2447 fathers and 2633 mothers. The mean age of the fathers was 47.3 years (SD 5.4) and of the mothers 45.1 years (SD 5.0). For the calculations of the relative risk to smoke when having a smoking spouse, the parents were divided into two age-groups: 30-45 and 46-65 years.
Peers / best friend: In the 1993 questionnaire twins were asked if their best friend smoked. The answer categories were: never-smoker, ex-smoker, smokes sometimes,

smokes 1-10 cigarettes/day, smokes more than 10 cigarettes/day. The question was answered by 2772 of the 3884 twins (71.4%) who completed a questionnaire in 1993. The remaining 1112 participants missed the question because it was the last in a series of questions on the smoking behavior of father, mother, co-twin, 5 brothers and 5 sisters. When the twins reported that their best friend was a never-smoker or an ex-smoker, the best friend was classified as non-smoker and when the twins reported that their best friend smoked sometimes, smoked 1-10 cigarettes a day or more than 10 cigarettes a day, the best friend was classified as smoker. The relative risk to be a smoker when having a smoking best friend compared to having a non-smoking best friend was calculated.
In addition, the twins were also asked in 1993 how many of their friends were regular smokers. The answer categories were: no one, a few friends, half of the friends, most friends, all friends. The question was answered by 3828 participants. The relative risk of being a smoker was calculated for subjects who reported that most or all of the friends were regular smokers compared to subjects who reported no, a few or half of their friends were smokers. Mean age when the twins filled in the questionnaire was 17.34 (SD 3.07) for the males and 17.39 (SD 3.15) for the females.
Spouses: In 2000 all twins aged between 25-30 years were asked if they had a partner and, if so, if their partner was willing to fill in a questionnaire. In total 706 partners completed the questionnaire. The mean age of the male spouses was 30.1 (SD 4.3) and of the female spouses 26.1 (SD 3.9).

Statistical analyses:
The prevalence of smoking was calculated using SPSS 10.0 for windows. The relative risk (RR) was used to summarize associations between smoking status of the participants and parental/sibling/friend/spouse smoking. The RR was calculated as the ratio of the percentage smokers with smoking family members/friends to the percentage smokers with non-smoking family members/friends. The RR and the 95% confidence intervals were calculated using Epi-Info 2000 version 1.1.

Results

Prevalence of smoking
For the prevalence of smoking, the following general trends were found in the data of twins and siblings: the prevalence of current smoking was higher for males than for females and higher for older participants than younger participants (Table 3.2A). For the parents, the prevalence of smoking was higher in the 30-45 year old group then in the 46-65 year old group for both mothers and fathers (Table 3.2B).

Age-differences
We first investigated whether the relative risk to smoke when having smoking family members and when having smoking friends was stable across the 3 age groups. Results showed that having smoking family members and having smoking friends significantly elevated the relative risk to smoke in all age groups. However, the relative risk to smoke when mother smoked, when siblings smoked or when friends smoked

was elevated more for the young adolescents (12-15 years) than for the older adolescents and adults. Across all age groups, the relative risk to smoke when both parents smoke was lower or comparable to the relative risk to smoke when having smoking siblings. Having smoking siblings (other than co-twin) presented a lower risk than having smoking friends (Table 3.3A).
The risk to be a smoker was calculated separately for having a smoking older brother, older sister, same age DZ brother, same age DZ sister, younger brother and younger sister. As the minimum age to participate in the questionnaire study was 12 years, almost no data were available for younger siblings of participants aged 12-15. No differences were found between the influences of having older or younger smoking siblings. In most age groups the relative risk to smoke when having smoking same age siblings (DZ co-twin) seemed somewhat higher than the relative risk to smoke when having younger or older siblings, although these differences were not significant.

Table 3.2 Prevalence of smoking. (a): Prevalence of smoking for siblings and twins (data collected in 1995-2000). For each age group, the percentage of smokers (% smok), the total number of participants (N tot) and the mean age is shown. MZM = monozygotic males, DZM = dizygotic males, DOS males = males of a dizygotic opposite sex twin, brother = additional brother of a twin, MZF = monozygotic females, DZF= dizygotic females, DOS females = females from a dizygotic opposite sex twin, sister = additional sister of a twin. (b): Prevalence of smoking for parents and spouses (data of parents collected in 1991-1995 and data of spouses collected in 2000). The percentage smokers (% smok), the total number of participants (n tot) and the mean age (mn age) is shown.

(a)	12-15 years			16-20 years			21-40 years		
	% smok	N tot	mn age	% smok	N tot	mn age	% smok	N tot	mn age
MZM	8.2	97	14.4	25.3	368	18.1	31.4	344	24.9
DZM	18.8	64	14.6	30.7	267	18.0	41.4	295	24.9
DOS males	11.1	171	14.5	32.0	250	18.2	41.9	315	25.0
Brother	19.8	81	13.6	30.1	236	18.3	39.6	535	26.2
MZF	11.1	171	14.4	22.5	543	18.1	28.1	734	26.8
DZF	8.7	104	14.4	26.5	332	18.1	30.0	523	25.7
DOS females	11.3	80	14.5	29.6	253	18.2	35.2	341	25.0
Sister	10.8	83	13.7	29.5	261	18.3	34.3	636	26.7

(b)	% smok	N tot	mn age
30-45 year old fathers	41.4	1040	42.9
46-65 year old fathers	36.1	1200	50.8
30-45 year old mothers	33.7	1413	41.9
46-65 year old mothers	25.6	894	49.9
Male spouses (20-53 years)	34.6	413	30.1
Female spouses (18-47 years)	22.6	257	26.1

Results showed a strong association between being a smoker and having smoking friends. For example when 12-15 year old girls have a smoking best friend the relative risk to smoke was almost 17 times higher compared to girls having a non-smoking best friend. For participants aged 16 years or older the RR to smoke when having smoking friends or a smoking best friend was lower than for the younger participants but still significant. The relative risk to smoke when having a smoking same age sibling was not significantly different from the relative risk to smoke when having smoking friends or having a smoking best friend.
In MZ twins (genetically identical), the risk to be a smoker when their co-twin smokes was high. The RR to smoke when the co-twin smokes was significantly higher for MZ twin compared to DZ twins in the older age groups but not in the youngest group.

Sex-differences
In addition to the effect of age, sex-differences were investigated. Table 3.3A summarizes the influence of smoking parents. For males, both having a smoking father and having a smoking mother elevated the risk to smoke. In contrast, for females, the relative risk to smoke was not significantly influenced by a smoking father for the participants aged 12-15 and 21-40 years while a smoking mother significantly elevated the risk to smoke in all age groups.
A trend for sex differences was found for the relative risk to smoke when having smoking siblings. For example, for the 12-15 year old females, having a smoking same-sex sibling of the same age elevated the risk to smoke 16 times while having a smoking opposite-sex sibling of the same age elevated the risk to smoke 'only' 6.5 times. However, the confidence intervals showed this difference was not significant. The trend that the relative risk to smoke is higher when same-sex siblings smoke than when opposite-sex siblings smoke was also seen for younger and for older siblings. For both men and women, this effect was strongest in the younger age groups.
Data on smoking behavior of the spouses of twins aged 24-31 years showed that the RR to smoke was about 3 times higher when having a smoking spouse compared to having a non-smoking spouse (Table 3.3B). Furthermore, data on smoking were available for spouses 30-46 years and 46-65 years old spouses. The risk to smoke when having a smoking spouse decreased from the younger age group to the older age groups for men as well as for women.

Discussion
Most studies have investigated risk ratios for smoking in young adolescents, aged between 12 and 18. The present study also included older participants up to 40 years of age. Our data showed that the relative risks to smoke when family members or friends smoke were still significant in adults. However, the relative risks were clearly higher for the young adolescents compared to the adults. Within each age group the relative risk to smoke is highest when having smoking friends, somewhat lower when having smoking younger/older siblings and lowest when having smoking parents.

Table 3.3 Relative risk to smoke when having smoking family members, friends, and spouses. (a) RR when having smoking siblings, friends or parents. For each age-group the relative risk (RR) and the 95% confidence interval (95% CI) is shown. dz = dizygotic, mz = monozygotic. (b) Relative risk to smoke when having a smoking spouse. The mean age of the participant, the spouse, the relative risk (RR), the 95% confidence interval (95% CI) and the total number of participant is shown.

(a)	12-15 years			16-20 years			21-40 years		
Smoking:	RR	95% CI	N	RR	95% CI	N	RR	95% CI	N
Men:									
older brother(s)	5.23	2.18-12.55	136	3.04	2.04-4.53	283	1.51	1.14-2.00	289
dz-twin brother	13.80	4.57-41.66	146	3.33	2.22-5.00	237	1.93	1.08-3.43	82
mz-twin brother	22.4	8.53-58.82	195	10.82	6.50-18.00	258	4.02	2.45-6.60	132
younger brother(s)	-	-	16	3.40	2.38-4.87	249	1.72	1.36-2.19	481
older sister(s)	2.42	1.02-5.72	119	1.78	1.34-2.36	284	1.64	1.28-2.11	327
dz-twin sister	6.74	2.63-17.32	285	2.58	1.68-3.94	214	1.67	0.74-3.74	55
younger sister(s)	-	-	7	2.29	1.34-3.90	117	1.16	0.85-1.59	263
best friend	10.22	4.75-21.97	328	3.87	2.94-5.09	508	2.10	1.52-2.92	208
most/all friends	7.53	4.15-13.04	540	2.82	2.34-3.40	832	2.06	1.63-2.60	303
father	2.00	1.24-3.23	893	1.35	1.12-1.63	1285	1.41	1.14-1.74	374
mother	2.46	1.55-3.90	939	1.41	1.18-1.69	1412	1.24	1.01-1.52	415
both parents	3.06	1.66-5.64	623	1.58	1.24-2.01	841	1.62	1.21-2.16	222
Women:									
older brother(s)	3.58	1.10-11.60	140	1.66	1.11-2.48	306	1.35	1.05-1.73	392
dz-twin brother	6.47	2.63-15.90	285	2.39	1.54-3.70	211	1.60	0.83-3.10	54
younger brother(s)	-	-	21	1.96	1.07-3.58	145	1.67	1.17-2.37	247
older sister(s)	6.67	2.16-20.61	130	2.21	1.61-3.02	372	2.22	1.74-2.84	453
dz-twin sister	16.13	7.61-34.18	184	3.64	2.47-5.37	284	2.97	1.80-4.91	188
mz-twin sister	25.60	10.49-62.48	271	13.39	8.11-22.12	377	4.32	3.05-6.11	227
younger sister(s)	-	-	16	3.06	2.08-4.50	254	2.35	1.82-3.03	632
Best friend	14.93	8.54-26.12	552	5.23	3.92-6.98	800	2.73	1.90-3.92	343
Most/all friends	16.65	10.00-27.71	662	3.61	2.95-4.43	1023	2.61	1.96-3.48	417
father	1.38	0.85-2.25	1018	1.75	1.43-2.14	1434	1.18	0.91-1.53	407
mother	2.35	1.51-3.66	1106	1.80	1.50-2.17	1607	1.52	1.19-1.93	446
both parents	2.16	1.19-3.93	728	2.35	1.80-3.08	918	1.51	1.05-2.17	240

(b)	Age participant Mean (range)	Age spouse Mean (range)	RR	95% CI	N
Men	27.4 (24-31)	25.9 (18-47)	2.90	2.07-4.05	252
	42.9 (30-46)	41.6 (32-54)	2.43	2.11-2.80	1035
	50.9 (46-65)	47.9 (34-63)	1.87	1.62-2.16	1167
Women	27.4 (24-31)	30.1 (20-46)	3.58	2.42-5.30	417
	42.9 (30-46)	44.4 (30-61)	2.71	2.31-3.18	1374
	50.9 (46-65)	51.6 (38-73)	2.04	1.62-2.56	854

Most previous studies did not distinguish between younger or older siblings (Maziak and Mzayek, 2000; Moran *et al.*, 2000; Whiters *et al.*, 2000; Boyle *et al.*, 2001), or only included older siblings (Wang et al, 1995; Jensen and Overgaard, 1993). It is likely that, when younger siblings imitate the smoking behavior of older siblings, the RR to be a smoker when having smoking older siblings is higher than the RR to be a smoker

when having younger siblings. Boyle *et al.* (2001) used the youngest sibling as dependent variable in a logistic regression analysis and concluded that the dominant influence of substance use behavior appeared to be from older siblings to younger siblings. In the present study, the RR to be a smoker was not different for having smoking younger siblings and having smoking older siblings. The younger siblings of the adults (21-40 years) are likely to already have completed the smoking initiation process and therefore differences may not be evident in this sample. However, the younger siblings of the 16-20 year olds may not have completed the smoking initiation process. Still no differences were observed between having smoking younger or older siblings in this younger group. As the minimum age to participate in the questionnaire study was 12 years, almost no data were available for younger siblings of participants aged 12-15. Although the results do not necessarily exclude imitation, they suggest imitation is not the only explanation for the similarity in smoking behavior of siblings. Other mediating factors for smoking behavior, which include genetic factors, thus play a role.

The strongest test for genetic influences on smoking behavior is the comparison of the degree of similarity of smoking behavior in MZ and DZ twin pairs. As shown, in the youngest group the RR to smoke when having a smoking MZ co-twin did not differ significantly from having a smoking DZ co-twin. In contrast, for the older participants results showed differences between MZ and DZ twins indicating genetic influences on smoking behavior. Studies in genetically informative samples (mostly twin studies but also adoption and family studies) have shown that genetic factors are important for smoking behavior, but seem less important in younger age cohorts (Heath and Madden, 1995; Sullivan and Kendler, 1999). This is supported by the fact that in the older groups the RRs to smoke were higher for MZ twins with a smoking co-twin than for DZ twins with a smoking co-twin.

A considerable risk to be a smoker was found when most or all friends were smokers or when the best friend was a smoker. This finding is in line with results from other studies (Hover, 1988; Jensen and Overgaard, 1993; Sasco *et al.*, 1993; Meijer *et al.*, 1996; Moran *et al.*, 2000; Alexander *et al.*, 2001). An individual can select his/her own friends in contrast to his/her family members. Similarity of smoking behavior within a peer group can be due to 'influence': individuals imitate the behavior of the peer group to fit in, or 'selection': individuals choose friends whose behavioral patterns are similar to their own (Engels *et al.*, 1997). Possibly, adolescents with a certain genetic predisposition actively seek out certain environmental experiences that may increase their risk for the development of some behaviors (Scarr and Mccartney, 1983). Data from a longitudinal study among Dutch secondary school students demonstrated that although influence and selection processes both contributed to peer group homogeneity, the largest part of similarity in smoking status had to be attributed to selection (Engels *et al.*, 1997). The similarity of friends can be used as an example of an active genotype-environment (GE) correlation which occurs when a particular genotype is associated with the selection or creation of a particular environmental circumstance (Rowe, 2002). Madden *et al.*, (2002) described model-fitting techniques to resolve the contribution of selective friendship and reciprocal peer environmental

influences to peer resemblance using cross-sectional data from pairs of siblings or twins and their peers. They included a reciprocal environmental influence of the behavior of one peer on the behavior of a sibling and vice versa. A major limitation of the approach is the assumption that each sibling chooses friends independently from his/her brother or sister. Therefore, methods to examine peer similarity in behavior may not be valid for samples that include a large number of siblings reporting on the behavior of the same friend(s), such as may be observed in the case of siblings who are very close in age or with samples of twins.

Attempts to explain adolescent cigarette smoking often consider the smoking behavior of friends as the most important factor, more important than smoking siblings. However, friends are usually of the same age while siblings are not. Unique to this study is the possibility to compare the relative risk to smoke when having smoking friends with the relative risk to smoke when having smoking same age siblings (DZ twins). We found that within each age group the relative risk to smoke when having a smoking friend is comparable with having a smoking same-age and same-sex DZ sibling. Those results suggest that age of the other person is just as important as age of the participant himself. Future studies should take into account the age-difference between the participant and their siblings and friends.

The importance of age-differences is in line with the relatively low risks we found for having smoking parents. A study of Osler *et al.*, (2001a) found adoptees' smoking behavior was not associated with adoptive or biological parental smoking whereas the adoptees' smoking behavior was associated with their full-siblings smoking behavior. The authors suggest this finding supports a genetic influence on smoking within the same generation. In line with the study of Olser *et al.* (2001), we have found a lower relative risk to smoke in parent-offspring pairs than in sibling pairs. Although it is possible that genes get switched on or off with increasing age or that different genes are operating at different ages, it is also possible that the results are due to different environmental or social influences in the two generations. This hypothesis can be tested with an extended parent-twin design, including young twins, their parents, and twins of the same age as the parents, as described by Snieder *et al.*, (1997).

The present study also investigated whether the relative risk to smoke for males and females is different for same-sex and opposite-sex family members. Sex-dependent influences of smoking parents were found. Our results are in line with Swan *et al.*, (1990) who reported for women a significant relative risk to smoke when mother smoked and no significant risk ratio when father smoked. The only other study that distinguished between both sex of the parent and sex of the participant did not find a significant association for smoking behavior (Wang *et al.*,1995).

Similar to the association between subjects' smoking and parental smoking, the association with sibling' smoking tended to be sex-dependent. In line with results of a study of Wang *et al.*, (1995), the RR to be a smoker was elevated more when having same-sex smoking siblings than when having opposite-sex smoking sibling(s), especially for the participants aged between 12 and 20.

Regrettably, data on the sex of the friends were not available. However, the best friend is often of the same sex. It is likely that, in line with results from smoking family

members, same sex friends influence smoking behavior differentially than opposite sex friends.
Spouses can be considered as a special type of opposite-sex 'best friend'. We included 3 adult groups with data of spouses. For 24-31 year old participants, the RR to smoke was about 3 times higher when having a smoking spouse compared to having a non-smoking spouse. Because Price *et al.*, (1981) suggested the convergence of phenotype was the most likely explanation for smoking behavior, we expected the relative risks to smoke to be lowest in the youngest group and highest in the oldest group. The opposite was found: having a smoking spouse was a larger risk factor for smoking in younger than in older age groups, for men as well as for women. Assuming that the duration of the relationship for the younger spouses is shorter than for the older spouses, our data suggest that assortment for smoking is based on similarity at the time dating began.

Limitations of the study:
As relative risks were calculated for having one or more smoking siblings, compared to having only non-smoking sibling(s), it was not possible to include the age of the sibling in the univariate analyses; some individuals have more than one sibling.
The data on smoking from twins, siblings and parents were self-reported. Self-reported smoking status is usually considered a satisfactory way to classify smoking in epidemiological studies. The smoking behavior of friends was reported by the twins. A studies of Bauman and Fisher (1996) described that perceptions of friends' smoking behavior were more strongly correlated than actual reports to adolescent smoking behavior. It seems that people project their own smoking behavior to their friends' smoking behavior. Although results of this study suggest the effect is not large, it may be that RR for friends' smoking is slightly overestimated.
In this paper, the association between subjects' smoking behavior and the smoking behavior of family and friends is investigated using univariate statistics. Univariate analyses do not adjust for other variables as would be possible with multivariate statistics. However, in multivariate statistics, if a particular variable correlates with another variable that has explanatory power, it may be excluded from the model although it correlates with smoking on the univariate level. Therefore, we first investigated those variables at a univariate level. For a smaller sample, longitudinal data are available and multivariate longitudinal analyses are currently prepared for a next paper.
Overall, for all age groups, the smoking behavior of parents and siblings increased the chances of smoking in the participants. The smoking behavior of friends and spouses was also associated with the smoking behavior of the participants. The results highlight not only the importance of social and genetic influences, with genetic influences increasing with the age of the participant, but also show that most findings are sex and age-dependent. Possibly, different mechanisms for smoking behavior occur in different age groups and in males and females.

Chapter 4

Smoking status of parents, siblings and friends: predictors of regular smoking? Findings from a longitudinal twin-family study

*Jacqueline M. Vink, Gonneke Willemsen, Rutger C.M.E. Engels, Dorret I. Boomsma (2003) Smoking status of parents, siblings and friends: predictors of regular smoking? Findings from a longitudinal twin-family study. Twin Research **6**(3): 209-217.*

Abstract
The relationship between regular smoking behavior and the smoking behavior of parents, siblings and friends was investigated using data from the Netherlands Twin Register. Cross-sectional analyses on data of 3906 twins showed significant associations between smoking behavior of the subject and smoking behavior of co-twin, additional brothers, parents of the same sex as the participant and friends. Those variables, together with age, explained 47% of the variance in smoking behavior. Longitudinal analyses of data from 2397 twins, who, in 1993, reported never to have smoked (regularly), showed that uptake of regular smoking two years later was predicted by having a smoking co-twin, smoking same-sex siblings, smoking mother and smoking friends. Males are, in contrast to females, at later age still vulnerable to take up regular smoking. The variables explained 21% of the variance. Sport participation, alcohol use, boredom susceptibility and neuroticism significantly added to the predictive value of this model. Including those additional factors increased the explained variance to 30%, and subsequently adding experimental smoking behavior further increased the explained variance to almost 50%. In summary, having smoking family members, friends, lifestyle and personality factors are important predictors for the uptake of regular smoking. However, the experimental smoking behavior of the subject is equally important.

Introduction

Research consistently shows associations between adolescents' smoking behavior and smoking behavior of parents, siblings and friends. In general, smoking behavior of parents and smoking behavior of adolescents' is weakly associated, with risk ratios ranging from non-significant to 4.0 (Green *et al.*, 1991; Jensen and Overgaard, 1993; Sasco *et al.*, 1993; Wang *et al.*, 1995; Brenner and Scharrer, 1996; Herlitz and Westholm, 1996; Meijer *et al.*, 1996; Swan *et al.*, 1997; Distefan *et al.*, 1998; Maziak and Mzayek, 2000; Moran *et al.*, 2000; Shamsuddin and Abdul Harris, 2000; Whiters *et al.*, 2000; Bauman *et al.*, 2001; Boyle *et al.*, 2001). Most of these studies did not investigate the influence of smoking father or mother separately for both sexes. Smoking siblings increases an adolescents' risk of smoking two to fourfold (Jensen and Overgaard, 1993; Wang *et al.*, 1995; Swan *et al.*, 1997; O'loughlin *et al.*, 1998; Maziak and Mzayek, 2000; Moran *et al.*, 2000; Whiters *et al.*, 2000; Boyle *et al.*, 2001). Two studies disentangled the influence of brothers and sisters but did not find large differences (Swan *et al.*, 1990; Maziak and Mzayek, 2000). Only one study also took the sex of the participant in account and reported for both males and females a significant association between smoking behavior and having smoking same-sex siblings (Wang *et al.*, 1995). In a univariate cross-sectional analysis on our own data from Dutch adolescents and young adults, we also observed that same-sex smoking family members influenced smoking behavior more than opposite-sex family members (Vink *et al.*, 2003a) .

Having smoking friends increased the risk to become a smoker 2 to almost 20 times (Mcneill *et al.*, 1988; Jensen and Overgaard, 1993; Sasco *et al.*, 1993; Wang *et al.*, 1995; Herlitz and Westholm, 1996; Swan *et al.*, 1997; Moran *et al.*, 2000; Alexander *et al.*, 2001; Bauman *et al.*, 2001). A large study of Wang *et al.* (1995) found that the odds ratio (OR) for smoking was higher when having a smoking same-sex best friend than when having a smoking opposite-sex best friend (Wang *et al.*, 1995). In contrast, Swan *et al.,* (1990) reported that a significant association was found between subjects' smoking behavior and having smoking opposite-sex friends, not with having smoking same-sex friends (15).

Most studies described above are univariate, only four multivariate studies have simultaneously included smoking behavior of family members (parents and siblings) and peers (best friend or most friends). Three of these studies found that having smoking friends was the best predictor of smoking behavior of adolescents (Jensen and Overgaard, 1993; Wang *et al.*, 1995; O'loughlin *et al.*, 1998). In contrast, McNeill *et al.* (1988) showed that the odds ratio for smoking siblings was slightly higher than for smoking peers, but peers were defined as the number of smokers in school year and may not reflect the friends of the individual (Mcneill *et al.*, 1988).

In view of the study results summarized above, it is tempting to conclude that much of the variance in smoking behavior can be explained by the smoking behavior of friends and family members. However, most of these studies have used a cross-sectional design. Cross-sectional studies are useful in suggesting hypotheses and can rule out possible causes when relationships between variables are not found. A longitudinal design may provide insight in the causal mechanisms. A study by Engels *et*

al. (1999) assessed how associations between possible explanatory variables and smoking onset depend on the use of cross-sectional versus prospective designs. A set of variables was used which consisted of smoking-related beliefs and attitudes, self-efficacy and future intentions, socio-demographic factors as well as smoking behavior of parents, best friend and peer-group. The cross-sectional analyses showed strong associations between smoking behavior of the participant and the smoking behavior of best friend and parents. However, over a period of 3 and 5 years, respectively 14% and 8% of the variance in change of smoking status from non-smoking to regular smoking could be predicted by the model variables. Smoking behavior of peers was excluded from this model (Engels *et al.*, 1999).

In general, longitudinal studies on smoking have reported a small or non-significant influence of smoking parents, siblings and friends, but none of these studies explored the influence of smoking family members and friends separately for males and females (Mcneill *et al.*, 1988; Oygard *et al.*, 1995; Distefan *et al.*, 1998; West *et al.*, 1999).

The goal of the present study was to analyze whether the variables that are cross-sectionally associated with smoking behavior also predict the uptake of regular smoking. This study is unique because, in addition to including the smoking behavior of parents (father and mother), siblings (brothers and sisters) and friends, the importance of genetic factors were assessed by comparing data from monozygotic (MZ) and dizygotic (DZ) twins. A higher association between smoking behavior of MZ twins compared to DZ twins may indicate genetic influences on smoking behavior because MZ twins are genetically identical while DZ twins share, on average, 50% of their DNA. We looked at 3 age-groups: 12-15 years (legally not allowed to buy tobacco), 16-20 years old and 21-25 years old. Analyses were carried out separately for males and females to explore sex-differences. Based on the literature we expected a high association between adolescents' smoking behavior and having smoking family members and friends in a cross-sectional analyses. To examine whether the same variables are involved in uptake of regular smoking we carried out a longitudinal analysis. In addition to smoking behavior of family and friends, other factors may influence the uptake of regular smoking. Lifestyle variables such as sport participation, alcohol use and religion have all been associated with smoking behavior (Tyas and Pederson, 1998). There is also consistent evidence for an association between smoking and depression, anxiety or neuroticism (Dierker *et al.*, 2002). The present study will explore whether these factors add to the prediction of regular smoking, when family and friends' smoking behavior are already included in the model.

Methods

Participants

This study is part of a longitudinal questionnaire study of the Netherlands Twin Register (NTR) that assesses families with adolescent and young adult twins every two/three years since 1991 (Boomsma, 1998; Koopmans *et al.*, 1999). For this paper, data from the 1993 and 1995 surveys were used. Sample selection and response rates are described in detail in Koopmans *et al.,* (1994) and Boomsma *et al.,* (2000) (Koopmans *et al.*, 1994; Boomsma *et al.*, 2000). Both in 1993 and 1995, subjects

received a questionnaire booklet that contained personality inventories, items about health, life-style (including smoking, alcohol use and exercise), socio-economic status and family structure. The questionnaire of 1993 also contained a question on the smoking behavior of parents, siblings and friends. For the cross-sectional analyses, the study sample consisted of 3906 twins who participated in 1993; 669 MZ males, 532 DZ males from same-sex twin-pair, 535 DZ males from opposite-sex twin-pairs, 955 MZ females, 671 DZ females from same-sex twin-pair and 544 DZ females from opposite-sex twin-pairs. For the longitudinal analyses, 2397 twins who participated both in 1993 and in 1995 were included. This sample consisted of 399 MZ male twins, 328 DZ same-sex male twins, 309 males from opposite-sex twin-pairs, 636 MZ female twins, 400 DZ female twins and 326 females from opposite-sex twin-pairs. In 1993, the mean age of these twins was 17.7 (SD 3.2) for the females and 17.8 (SD 3.1) for the males.

Smoking status of the twins, siblings, parents and friends

The questionnaires contained 3 questions on smoking initiation: "Did you ever smoke a cigarette?", "Did you smoke during the last 12 months?" and "Did you smoke during the last 4 weeks?". The answer categories were: no, a few times to try, yes. Another question was "How often do you smoke now?" with the answer categories: I have never smoked, I have quit smoking, I smoke less than once a week, I smoke several times a week but not every day, I smoke daily. Participants also reported the number of cigarettes they smoke per day or per week. Based on their answers participants were classified as never-smokers, experimental smokers, regular smokers or ex-smokers in 1993 and 1995. The never-smokers and the experimental smokers were classified as non-regular smokers and the ex-smokers were excluded from the logistic regression analyses. Answers were checked for consistency across questions. Because the sample consisted of twins, 6 groups were created: having a non-smoking MZ co-twin, having a smoking MZ co-twin, having a non-smoking same-sex DZ co-twin and having a smoking same-sex DZ co-twin, having a non-smoking opposite-sex co-twin and having a smoking opposite-sex co-twin (Heath *et al.*, 1998).

In 1993, the twins were asked to report if their father, mother, co-twin, brother(s), sister(s) were non-smokers, ex-smokers, smoked sometimes, smoked 1-10 cigarettes a day or smoked more than 10 cigarettes a day. Based on these answers, parents and siblings were classified as non-smokers or regular smokers in 1993. Parents received a survey themselves and were also classified as regular smokers or non-regular smokers based on their self-reported data using the same criteria as for the twins. For 3165 fathers and 3497 mothers both self-reported data and reports from their children were available; 97% of the answers from father and twin and 98% of the answers from mother and twin were in agreement. When available, parent self-report data were used to classify each parent as non-smoker or regular smoker, otherwise the smoking status reported by their children was used. Furthermore, in 1993 the twins were asked how many of their friends were regular smokers. The answer categories were: no one, a few friends, half of the friends, most friends, all friends. The question was answered by 3828 participants.

Additional variables

Additional variables that were explored in the longitudinal analyses were: alcohol use (less than weekly alcohol use, more than 1 glass a week), regular sport participation (yes, no), religion (no religion, religious but not actively participating, religious and actively participating), tea- and coffee-use (no, more than 1 cup a day). In addition, the scales of the Amsterdamse Biografische Vragenlijst, (Neuroticism, Somatic anxiety, Extraversion and Test attitude) (Wilde, 1970), the scales of the Zuckerman Sensation Seeking Questionnaire (Boredom susceptibility, Disinhibition, Experience seeking and Thrill and adventure seeking) (Zuckerman, 1971) and the Beck Depression Inventory (Beck *et al.*, 1961) were included. For the personality variables, the 30% highest scores were classified as 'high', the 30% lowest scores as 'low' and the remaining (40%) as 'medium'. For the Beck Depression Inventory, the 30% highest scores were classified as 'high', and the remaining (70%) as low.

Data Analyses

Binary logistic regression analyses were carried out separately for male and female twins. Nagelkerk's R^2 was used to index the explained variance.

1. Cross-sectional analyses were carried out on the 1993-data to evaluate the association between the participants' smoking behavior and the smoking status of family members and friends. Ex-smokers were excluded from the analyses. The dependent variable was the smoking status of the twin (regular smoker: yes/no).
2. Longitudinal analyses were performed to explore whether the smoking behavior of family members and friends in 1993 predicted uptake of regular smoking in 1995. To focus on uptake of regular smoking, only never smokers or experimental smokers in 1993 were included in the analyses. The dependent variable was smoking status in 1995 (regular smoker: yes/no).
3. Next, a longitudinal analysis was performed to explore whether, in addition to smoking family and friends, other factors like alcohol use, regular sport participation, religion, tea-use, coffee-use, personality and depression significantly predict the uptake of regular smoking. The significant variables from the first longitudinal analyses were entered in the model at the first step. At the second step, the influence of additional variables was explored using the forward conditional method. Interaction effects between age and the other variables were examined but excluded from the model when not significant.

Results

Prevalence of smoking

Table 4.1 shows that for both males and females, the percentage regular smokers increased from the young adolescents, legally not allowed to buy tobacco (12-15 years), to the older adolescents (16-20) and young adults (21-25 years). Furthermore, for all age-groups the percentage regular smokers was higher in 1995 than in 1993.

Table 4.1 Smoking status in 1993 and 1995 for 3 age groups separately for males and females (n, %). Smoking status categories are: never smoked, tried smoking (but never smoked regularly), regular smoker and ex-smoker.

Smoking	12-15 years		16-20 years		21-25 years		Total	
phase	1993	1995	1993	1995	1993	1995	1993	1995
Males								
never smoked	501	240	555	241	166	72	1222	553
	89.1%	73.6%	65.0%	46.1%	53.5%	38.5%	70.8%	53.4%
tried smoking	42	43	99	90	35	38	176	171
	7.5%	13.2%	11.6%	17.2%	11.3%	20.3%	10.2%	16.5%
regular smoker	16	36	193	172	98	67	307	275
	2.8%	11.0%	22.6%	32.9%	31.6%	35.8%	17.8%	26.5%
ex-smoker	3	7	7	20	11	10	21	37
	0.5%	2.1%	0.8%	3.8%	3.5%	5.3%	1.2%	3.6%
Total:	562	326	854	523	310	187	1726	1036
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
Females								
never smoked	594	297	753	350	292	135	1639	782
	87.6%	65.0%	72.6%	54.5%	68.5%	51.5%	76.6%	57.5%
tried smoking	50	88	119	127	40	48	209	263
	7.4%	19.3%	11.5%	19.8%	9.4%	18.3%	9.8%	19.3%
regular smoker	32	62	157	152	86	61	275	275
	4.7%	13.6%	15.1%	23.7%	20.2%	23.3%	12.8%	20.2%
ex-smoker	2	10	8	13	8	18	18	41
	0.3%	2.2%	0.8%	2.0%	1.9%	6.9%	0.8%	3.0%
Total:	678	457	1037	642	426	262	2141	1361
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Cross sectional associations

Table 4.2 presents the cross-sectional associations between the subjects' smoking behavior and that of their parents, co-twin, additional siblings and friends. For males, smoking was associated with having a smoking father not mother, while for females smoking was associated with having a smoking mother not father. A significant association was found for having smoking brothers, not having smoking sisters, for both males and females. The twins' smoking behavior was significantly associated with that of their co-twin. Using participants with a non-smoking MZ co-twin as the reference group, highest odds ratios were found for having a smoking MZ co-twin, followed by having a smoking same-sex DZ co-twin and finally having a smoking DZ opposite-sex co-twin. The adolescents' smoking behavior was strongly associated with the smoking behavior of friends, with odds ratios being higher when most or all friends smoked than when half of the friends smoked. Compared to the youngest group, the risk to be a regular smoker was higher for the participants aged 16-20 and 21-26 years. Interaction effects between the different age-groups and the other variables in the model were not significant and therefore excluded from the model. The efficacy of the model in explaining smoking behavior was 47% for both males and females (Table 4.2).

Table 4.2 Cross-sectional association (odds ratio and 95% CI) between smoking behavior of twins and smoking behavior of their father, mother, co-twin, additional brothers or sisters and friends in 1993. Odds adjusted simultaneously for the other factors. Age-group was also included in the analyses. R square and number of participants is shown at the bottom of the table. Dependent variable is regular smoking (yes/no) of participant in 1993.

Factors:	Category:	Males		Females	
		OR	95% CI	OR	95% CI
Father	Non-smoker in 1993	1		-	
	Smoker in 1993	**1.48**	1.06 - 2.07	-	
Mother	Non-smoker in 1993	-		1	
	Smoker in 1993	-		**1.53**	1.06 - 2.21
Co-twin	Non-smoking MZ co-twin in 1993	1		1	
	Smoking MZ co-twin in 1993	**10.41**	5.60 -19.37	**21. 82**	11.70- 40.71
	Non-smoking DZ ss co-twin in 1993	1.56	1.26 - 3.04	1.37	0.81 - 2.32
	Smoking DZ ss co-twin in 1993	**8.70**	4.83 -15.67	**11.87**	5.96 - 23.66
	Non-smoking DZ os co-twin in 1993	**2.39**	1.47 - 3.87	**2.13**	1.23 - 3.68
	Smoking DZ os co-twin in 1993	**8.28**	4.23 -16.20	**6.48**	3.54 - 11.86
Brothers	Non-smoking brother(s) in 1993	1		1	
	Smoking brother(s) in 1993	**2.56**	1.47 - 4.47	**2.65**	1.56 - 4.50
	No additional brother(s)	**1.65**	1.12 - 2.43	1.29	0.85 - 1.95
Sisters	Non-smoking sister(s) in 1993	-		-	
	Smoking sister(s) in 1993	-		**-**	
	No additional sister(s)	-		-	
Friends	No one / a few smoking in 1993	1		1	
	Half of the friends smoke in 1993	**3.26**	2.13 - 4.99	**4.93**	3.05 - 7.97
	Most / all friends smoke in 1993	**7.25**	4.97 - 10.58	**10.22**	6.83 - 15.28
Age of	12-15 years	1		1	
subject	16-20 years	**6.46**	3.35 - 12.44	**2.73**	1.57 - 4.77
in 1993	21-25 years	**10.44**	5.19 - 21.02	**3.99**	2.15 - 7.38
R^2	*Nagelkerk R square*	*0.468*		*0.473*	
N	*Number of participants*	*1541*		*1910*	

- = variable did not enter the model, OR in bold = category is significantly associated with regular smoking behavior.

Table 4.3 Changes in smoking behavior (smoking status in 1993 and 1995).

	Smoking status in 1995 →			
Smoking status in 1993 ↓	never smoked/ tried smoking	regular smoker	ex-smoker	Total
never smoked/ tried smoking	1769 86.8 %	218 10.7 %	51 2.5 %	2038 100 %
regular smoker		332 92.2 %	28 7.7 %	360 100 %
Total	1769 73.7 %	550 22.9 %	78 3.3 %	2397 100 %

Changes in smoking behavior

Most of the subjects who did not smoke or only tried smoking in 1993 were still not smoking regularly in 1995, 10% took up regular smoking and 2.5% became ex-smokers. Most of the regular smokers in 1993 were also regular smokers in 1995, only 7.7% quit smoking (Table 4.3).

Table 4.4. Longitudinal association (odds ratio and 95% CI) between smoking behavior of twins in 1995 and the smoking behavior of their father, mother, co-twin, additional brothers or sisters and friends in 1993. Odds adjusted simultaneously for the other factors. Age-group was also included in the analyses. R square and number of participants is shown at the bottom of the table. Dependent variable is regular smoking (yes/no) of participant in 1995.

Factors:	Category:	Males		Females	
		OR	95% CI	OR	95% CI
Father	Non-smoker in 1993	-		-	
	Smoker in 1993	-		-	
Mother	Non-smoker in 1993	-		1	
	Smoker in 1993	**-**		**1.92**	1.21 - 3.04
Co-twin	Non-smoking MZ co-twin in 1993	1		1	
	Smoking MZ co-twin in 1993	**4.58**	1.40 - 14.99	**11.64**	3.34 - 40.58
	Non-smoking DZ ss co-twin in 1993	1.60	0.84 - 3.04	**1.90**	1.08 - 3.37
	Smoking DZ ss co-twin in 1993	3.03	0.90 - 10.19	**7.72**	2.44 - 24.48
	Non-smoking DZ os co-twin in 1993	**3.11**	1.69 - 5.70	**2.39**	1.32 - 4.30
	Smoking DZ os co-twin in 1993	3.24	0.81 - 12.89	1.27	0.39 - 4.14
Brothers	Non-smoking brother(s) in 1993	1		-	
	Smoking brother(s) in 1993	**3.61**	1.56 - 8.29	-	
	No additional brother(s)	1.05	0.63 - 1.73	-	
Sisters	Non-smoking sister(s) in 1993	-		1	
	Smoking sister(s) in 1993	-		**3.61**	1.39 - 9.42
	No additional sister(s)	-		1.62	0.98 - 2.69
Friends	No one / a few smoking in 1993	1		1	
	Half of the friends smoke in 1993	**4.79**	2.09 - 10.97	**2.70**	1.35 - 5.37
	Most / all friends smoke in 1993	**13.13**	5.62 - 30.68	**9.39**	4.65 - 18.98
Age of subject in 1993	12-15 years	-		1	
	16-20 years	-		0.67	0.41 - 1.11
	21-25 years	-		**0.38**	0.18 - 0.80
R^2	*Nagelkerk R square*	*0.210*		*0.218*	
N	*Number of participants*	*770*		*1068*	

- = variable did not enter the model, OR in bold = category is significantly associated with regular smoking behavior, ss = same-sex, os = opposite sex, MZ = monozygotic, DZ = dizygotic

Table 4.4 shows the results of the longitudinal regression analyses, performed to assess the effect of the smoking behavior of parents, siblings and friends in 1993 on the uptake of regular smoking behavior in 1995. Non-smoking females, not males, with a smoking mother have a higher risk to become a regular smoker 2 years later. Having a smoking father did not influence the uptake of regular smoking both for

males and females. Regular smoking in 1995 was predicted by having smoking same-sex siblings (other than co-twin) in 1993, but not by having smoking opposite-sex siblings. Using participants with a non-smoking MZ co-twin as the reference group, highest odds ratios were found for having a smoking MZ co-twin. Having a smoking same-sex DZ co-twin formed a higher risk to take up regular smoking than having a non-smoking same-sex DZ co-twin. This pattern was not found for having a smoking/non-smoking opposite-sex DZ co-twin. Having smoking friends in 1993 significantly predicted transition to regular smoking in 1995, for both males and females. Age entered the model for females only, with lower odds of becoming a regular smoker for subjects aged 21-25 year. The total model explained 21% and 22% of the variance for males and females, respectively (Table 4.4).
Next, the logistic regression was repeated with the significant variables of the first longitudinal logistic regression analyses entered at the first step and the additional variables sport participation, alcohol use, coffee- and tea-use, religion, depression and personality scores entered at the second step. The odds to be a regular smoker in 1995 was significantly higher for twins who used alcohol, had a high boredom susceptibility score or a high neuroticism score in 1993 and significantly lower for twins who exercised regularly in 1993. The model explained 31% of the variance for males and 30% of the variance for females (Table 4.5).
Earlier smoking status may account for some of the variance in changes in smoking status at a later wave (see West *et al.,* 1999). Table 4.3 shows that 218 subjects who did not smoke in 1993 became regular smokers in 1995. Of these subjects 138 (63%) had already experimented with smoking in 1993. In a final step we added the 1993-smoking behavior to the model. This showed that the chance to take up regular smoking was 15 and 23 times higher for respectively males and females who already experimented with smoking in 1993 compared to participants who never smoked in 1993. The explained variance increased markedly to 47% for males and 49% for females.

Discussion

Cross-sectional analyses

Smoking behavior of the subject is associated with smoking behavior of parents of the same sex as the participant, brothers, co-twin and friends. In general, those results are in line with most of the literature on this topic (Mcneill *et al.*, 1988; Green *et al.*, 1991; Brenner and Scharrer, 1996; Whiters *et al.*, 2000). Compared to other cross-sectional studies to smoking, this study is unique because it included having a smoking MZ/DZ co-twin. Results showed a genetic influence on smoking behavior since the OR for having a smoking MZ co-twin was higher than the OR for having a smoking DZ co-twin. The genetic influences seem sex-dependent because the OR for having a smoking DZ same-sex co-twin was higher than the OR for having a smoking opposite-sex co-twin. Results also showed that age was an important factor; the risk to be a regular smoker was significantly higher for 16-20 and for 21-25 year old participants than for 12-15 year old participants.

Table 4.5 Additional predictors of uptake of regular smoking in 1995. Never-smokers and experimental smokers in 1993 were selected. At the first step the significant variables from the first longitudinal analyses were entered in the model. At the second step alcohol use, regular sport participation, religion, tea- and coffee-use, personality and depression were entered. Odds adjusted simultaneously for the other factors.

Factors:	Category:	males		females	
		OR	95% CI	OR	95% CI
Mother	Non-smoker in 1993	X		1	
	Smoker in 1993	X		**1.89**	1.17 - 3.06
Co-twin	Non-smoking MZ co-twin in 1993	1		1	
	Smoking MZ co-twin in 1993	**6.46**	1.81 - 22.99	**13.95**	4.06 - 47.93
	Non-smoking DZ co-twin in 1993	1.54	0.79 - 3.01	**2.00**	1.10 - 3.64
	Smoking DZ co-twin in 1993	2.88	0.77 - 10.83	**10.93**	3.11 - 38.41
	Non-smoking DZ os co-twin in 1993	**2.32**	1.23 - 4.39	**2.54**	1.38 - 4.68
	Smoking DZ os co-twin in 1993	1.82	0.44 - 7.57	1.69	0.48 - 5.95
Brothers	Non-smoking brother(s) in 1993	1		X	
	Smoking brother(s) in 1993	**4.40**	1.86 - 10.39	X	
	No additional brother(s)	0.98	0.58 - 1.66	X	
Sisters	Non-smoking sister(s) in 1993	X		1	
	Smoking sister(s) in 1993	X		**2.86**	1.06 - 7.23
	No additional sister(s)	X		1.63	0.97 - 2.75
Friends	No one / a few smoking in 1993	1		1	
	Half of the friends smoke in 1993	**4.49**	1.88 - 10.72	**2.86**	1.35 - 6.05
	Most / all friends smoke in 1993	**10.38**	4.19 - 25.71	**7.30**	3.35 - 15.87
Age	12-15 years	X		1	
	16-20 years	X		**0.37**	0.20 - 0.70
	21-40 years	X		**0.27**	0.11 - 0.63
Sport Participation	No	1		1	
	Yes	**0.49**	0.29 - 0.84	**0.60**	0.37 - 0.98
Alcohol use	No alcohol use/ < than 1 glass a week	1		1	
	More than 1 glass /week	**1.66**	1.00 - 2.76	**3.43**	1.94 - 6.07
Boredom susceptibility	Low	1		1	
	Medium	**2.62**	1.25 - 5.49	1.07	0.56 - 2.03
	High	**3.66**	1.71 - 7.85	**2.42**	1.33 - 4.42
Neuroticism	Low	1		1	
	Medium	**1.95**	1.08 - 3.55	1.54	0.80 - 2.96
	High	**2.65**	1.35 - 5.19	**2.31**	1.20 - 4.45
R^2	*Nagelkerk R square*	*0.308*		*0.300*	
N	*Number of participants*	*755*		*1068*	

X = variable not included in the analyses as it was not significant in the previous analysis (Table 4.4), OR in bold = category is significantly associated with regular smoking behavior, ss = same-sex, os = opposite-sex, MZ = monozygotic, DZ = dizygotic

Longitudinal analyses with smoking family and friends

The longitudinal analyses showed that having a smoking co-twin, having smoking same-sex siblings and having smoking friends were predictors of the uptake of regular smoking for both males and females. Having a smoking father was not a significant predictor of transition to regular smoking while having a smoking mother was a significant predictor for females only. Two longitudinal studies found that mother's smoking, not father's smoking, predicted the transition from non-smoking to regular smoking (Oygard *et al.*, 1995; Engels *et al.*, 1999) while two longitudinal studies did not find an independent effect of parental smoking on uptake of regular smoking (Distefan *et al.*, 1998; West *et al.*, 1999). The influence of parental smoking thus seems to be small or insignificant, but when an effect is found having a smoking mother seems to be more important than having a smoking father.

The analyses showed that having smoking same-sex siblings (co-twin or other siblings), not having smoking opposite-sex siblings, significantly predicts transition to regular smoking two years later. Two longitudinal studies that included sibling smoking did not distinguish smoking brothers or sisters (Oygard *et al.*, 1995; West *et al.*, 1999). As far as we know, our study is the first one that investigates the influence of smoking brothers and sisters in a longitudinal analysis separately for males and females. The finding that only the smoking same-sex sibling(s) predicted the uptake of regular smoking indicates that different mechanisms for the uptake of regular smoking behavior occur in males and females. Possibly, the environmental influences of family and friends are different for males and females, it is also possible that different genes operate in males and females. The transition to regular smoking was significantly predicted by having a smoking co-twin. If genetic factors are important for smoking behavior, it is expected that participants with a smoking MZ co-twin have a high genetic liability for smoking themselves because MZ twins are genetically identical. DZ twins share in general only 50% of their genes, it is therefore expected that participants with a smoking DZ co-twin have a lower risk of having the same high genetic liability for smoking themselves. If sex-differences are important it is expected that having a smoking same-sex DZ co-twin forms a higher risk than having a smoking opposite-sex DZ co-twin. Our results showed that, compared to having a non-smoking MZ co-twin, odds ratios for having a smoking MZ co-twin were higher than for having a smoking DZ co-twin, suggesting that genetic factors are involved in the transition to regular smoking. Heath *et al.* (1998) used the same approach in a sample of twins and their results were also consistent with a significant genetic influence on smoking (Heath *et al.*, 1998). Those findings are in line with the classical twin studies using different approaches like structural equation modeling. Those studies have shown that regular tobacco use is largely heritable (Heath and Madden, 1995; Sullivan and Kendler, 1999).

The risk of taking up regular smoking two years later is 9 and 13 times higher, for respectively females and males, when most or all friends smoke compared to participants with no or just a few smoking friends. This finding is in line with a study that reported that uptake of smoking in the next 4 years was predicted by having a best friend who smokes (Distefan *et al.*, 1998) but in contrast with two longitudinal studies

that did not find friends' smoking a significant predictor for the uptake of regular smoking (Oygard *et al.*, 1995; Engels *et al.*, 1999). One of those studies explored the uptake of regular smoking in a 3- and 5-year period (Engels *et al.*, 1999) and the other used an 8- and 10-year interval (Oygard *et al.*, 1995). Another study reported that friends' smoking at age 15 increased the likelihood of uptake up to 10 times over the next year, but did not extend to later years. When those participants had smoking friends at age 18, they were three times as likely to become a regular smoker over a 3-year period (West *et al.*, 1999). This suggest that the age of the participant could be important but different results could also be due to the duration of the follow-up period. Possibly the association with smoking friends is higher when the period is shorter. This could be caused by the fact that individuals select a peer group with similar smoking behavior and friendships may change when the smoking behavior becomes dissimilar (Bauman and Ennet, 1996). It is important not to overlook the possibility that the selection mechanism could be based on the genotype of the subject. Similarity of friends' behavior might be caused by an active genotype-environment (GE) interaction that occurs when a particular genotype is associated with the selection or creation of a particular environmental circumstance (Rowe, 2002).

It should be noted that self-reported data were used to determine the smoking status of twins and parents, but the smoking status of siblings and friends was reported by the twins. Bauman *et al* (1996) described that perceived reports of friends' drug use were more strongly correlated than actual reports to adolescent drug use (Bauman and Ennet, 1996; Bauman and Fisher, 1996). Although according to the authors the effect is not very large, the reported OR for friends' and siblings' smoking in our study might have been overestimated. However, we do not expect this effect to be very large because we found a large agreement between parent self-report and the reports of the children on the smoking behavior of their parents.

For females, the odds to take up regular smoking is significantly lower for 21-25 year olds. For males, no differences between the three age-groups were found. This suggests that women, if regular smoking behavior is not established before age 20, have a low chance of taking up regular smoking. In contrast, even at later age males are still vulnerable to take up regular smoking.

Additional longitudinal analyses

The variables in the cross-sectional analyses explained 46% of the variance in males and 47% in females (Nagelkerk R^2) while the same variables in the longitudinal analyses explained 21% of the variance in males and 22% in females. These results imply that although having smoking siblings and friends significantly predicted the uptake of regular smoking, other factors are involved in the transition from non-smoking to regular smoking. A study of Engels *et al.* (1999) also showed high correlations between explanatory variables and smoking in a cross-sectional analyses while the explained variance in a longitudinal design was much lower.

We found that in addition to having smoking family members and friends, alcohol use, high boredom susceptibility and high neuroticism scores significantly predicted the

uptake of smoking two years later while sport participation was a protective factor for the uptake of regular smoking. Adding those variables increased the explained variance slightly from 21% to 31% for males and from 22% to 30% for females. The explained variance increased noticeably to 47% for males and 49% for females when the smoking behavior of the p articipant in 1993 was added. It seems that, although other factors are important for the uptake of regular smoking, it is most important whether participants have already experimented with smoking. To prevent adolescents and young adults from regular smoking it is important to keep adolescents from experimentation with smoking in the first place.

In summary, the uptake of regular smoking can only be predicted by a wide variety of genetic and environmental factors such as smoking family members, smoking friends, personality and lifestyle. This study has shown that having a smoking co-twin, having smoking same-sex siblings, having smoking friends and, for females only having a smoking mother, significantly predicts the uptake of regular smoking two years later.

Sport participation, alcohol use, boredom susceptibility and neuroticism significantly added to the predictive value of this model. However, subsequently including the 1993-smoking behavior of the subject increased the explained variance markedly.

Chapter 5

The Fagerström Test for Nicotine Dependence in a Dutch sample of daily smokers and ex-smokers

A short version of this chapter will be published as:
Jacqueline M. Vink, Gonneke Willemsen, A. Leo Beem, Dorret I. Boomsma. The Fagerström Test for Nicotine Dependence in a Dutch sample of daily smokers and ex-smokers. Addictive Behaviors (in press)

Abstract
The FTND showed a reasonably high internal consistency for both 1.378 daily smokers and 1.058 ex-smokers of a survey study of the Netherlands Twin Register. Mean FTND scores were higher for smokers than for ex-smokers. Male smokers scored significantly higher than female smokers, which was due to the item 'number of cigarettes per day'. Nicotine dependence level was associated with educational attainment of the father (for smokers only) and type of cigarette but not with age or sex. FTND score was highly correlated with the maximum number of cigarettes smoked (even after excluding item 'number of cigarettes per day' from FTND) but FTND score did not correlate with number of quit attempts (except for male ex-smokers) and showed a low correlation with age of first cigarette. Test-retest correlations of the Dutch FTND were high for both smokers (n=151) and ex-smokers (n=194). In general, the performance of the FTND in ex-smokers was comparable with the performance of the FTND in smokers. These findings suggest the FTND to be a valuable tool for studies of nicotine dependence in large epidemiological samples.

Introduction

Both for clinical practice and for research on smoking it is useful to have a measure of the degree of nicotine dependence which can be used in large epidemiological samples. To determine nicotine dependence, structured interviews like the DSM-IV can be used or, alternatively, self-report measures of nicotine dependence such as the Fagerström Tolerance Questionnaire (FTQ) (Colby *et al.*, 2000). The FTQ was developed in 1978 (Fagerström, 1978) and a revised version was published in 1991: the Fagerström Test for Nicotine Dependence (FTND), (Heatherton *et al.*, 1991). This revised version was created because the FTQ was multi-factorial and had a low internal consistency. The FTND items all loaded on a single factor and the internal consistency of the FTND was increased to .61 (Heatherton *et al.*, 1991). The FTND consists of 6 items and produces a score ranging from 0 to 10 with higher scores indicating more nicotine dependence. Fagerström *et al.* (1996) compared FTND data of treatment and population studies. Mean FTND scores ranged from 5.15 to 6.55 in treatment samples while lower FTND scores, ranging from 3.07 to 4.30, were found in population samples of current smokers. More recent studies have found comparable or even lower mean scores in population based samples, ranging from 1.84 to 3.2 (Etter *et al.*, 1999; John *et al.*, 2003).

In general, the prevalence of smoking is higher in men than in women, and men consistently score higher on nicotine dependence than women (Fagerström *et al.*, 1996; John *et al.*, 2003). Prevalence of smoking is also higher among participants with a low educational attainment compared to participants with a higher educational attainment (Escobedo and Peddicord, 1996; Malmstadt *et al.*, 2001; Osler *et al.*, 2001b) and more educated smokers tend to be less nicotine dependent (Gallus *et al.*, 2002). However, lifetime risk of nicotine dependence did not vary with the total number of years of education (Breslau *et al.*, 2001) indicating that the association between nicotine dependence and educational attainment is not unequivocal. While the prevalence of smoking increases with age for participants between 12 and 40 years old (Vink *et al.*, 2003a), FTND scores were stable in a German population of smokers aged 20 to 59 (John *et al.*, 2003).

Another interesting factor is whether persons who smoke light cigarettes are less dependent compared to persons who smoke regular cigarettes. Etter *et al.* (2003) showed that smoking mild, light or ultralight (versus regular) cigarettes was associated with a lower FTND score. A number of smoking history indicators are associated with the FTND score. For example, increased nicotine dependence is related to early age of smoking initiation (Everret *et al.*, 1999; Lando *et al.*, 1999) and the FTND score correlates low wit the duration of smoking (Horn *et al.*, 2003; John *et al.*, 2003).

There are few longitudinal studies that collected FTND data but the test-retest correlations in those studies were high, ranging from .85 in a study of the French FTND (Swiss sample, 7 months apart) to .88 in a study of the English FTND (American sample, 15 days apart), (Pomerleau *et al.*, 1994; Etter *et al.*, 1999).

As far as we know, there are no publications on the performance of the FTND in ex-smokers. Both for clinical practice and for research studies, it might be useful to have a measure of the degree of nicotine dependence for all participants who ever smoked (independent of their current smoking status). For example, genetic studies assume that

there is an underlying liability for nicotine dependence. If there is an underlying (genetic) liability for nicotine dependence then exclusion of the ex-smokers can cause bias and decreases the sample size unnecessary in, for example, family studies of (genetic) influences on nicotine dependence.
We have collected data with the FTND questionnaire in both smokers and ex-smokers in order to compare its performance for both groups. We first describe the distribution and internal consistency of the FTND for daily smokers (n=1.378) and ex-smokers (n=1.058) aged 16 years and older. Then, we investigate the mean FTND scores for different educational attainments of the father (low, medium, high) and for type of cigarette (low, medium or high nicotine cigarettes). We obtain the correlation between current age and FTND score. Finally, we obtain correlations between FTND score and smoking history indicators (i.e. age of first cigarette, number of quit attempts etc) and test-retest correlations for sub-samples of smokers and ex-smokers who completed the FTND twice. All analyses are carried out taking sex into account.

Methods

Participants

As part of a longitudinal survey study of the Netherlands Twin Register smoking data were collected in 6.792 participants in 2000 (Boomsma *et al.*, 2002; Vink *et al.*, 2003a; Vink *et al.*, 2003b). There were 3.939 (58.5%) non-smokers, 1.732 (25.7%) current smokers and 1.058 (15.7%) ex-smokers. Smoking status could not be determined for 63 participants. FTND data were available for 1.700 smokers who completed the FTND on their current smoking behavior. Of the 1.700 smokers, only the 1.397 daily smokers were included in the analyses. Participants who smoked less than once a week (n=103) and participants who smoked several times per week but not every day (n=200) were excluded from the analyses. After excluding participants below 16 years, FTND data were available for 584 men (mean age 30.3, SD 9.0) and 794 women (mean age 30.6, SD 10.4) who were daily smokers and completed the FTND. FTND data were also available for 1.035 of the 1.058 ex-smokers who reported on the period they smoked the heaviest. The sample of 1.058 ex-smokers consisted of 368 men (mean age 38.6, SD 14.3) and 690 women (mean age 37.1, SD 11.8).

Test-retest sample FTND

A sub-sample of 606 participants took part in both in the questionnaire study of 2000 and in a genetic linkage study of ND. Families were selected for the linkage study based on data from the longitudinal survey study. Questionnaires were send in 1991, 1993, 1995, 1997 and 2000. Most subjects participated more than once. Based on their answers twins and siblings were classified as 'nicotine dependent' (ND) or 'light smoker' (LS). A person was classified as ND when the FTND score was 6 or higher (FTND was only included in 2000 survey), when cigarette consumption was more than 20 cigarettes per day (based on answers in the 5 surveys), when smoking continued during pregnancy (women only, 2000 survey) or when the subject had tried to quit smoking more than 3 times (2000 survey). A subject was classified as LS when the person ever smoked or tried smoking but never smoked more than 5 cigarettes

per day (based on answers in the 5 surveys). In total, smoking data were available for 10.584 participants (8.089 twins and 2494 siblings) from 4.392 families. Of those, almost 10% was classified as ND (n=1.040) and 37% was classified as LS (n=5.666). The remaining 53% of the sample was unclassified and included both never smokers and participants who smoked more than 5 cigarettes per day but less than 20.
Families were selected when at least two siblings were both ND or at least one sibling was LS and one sibling was ND (MZ twin pairs were treated as single offspring). Using these criteria 388 families were selected for the linkage study and the entire family (twin, siblings and parents) was asked for a DNA sample. In total, 1.933 subjects were asked to provide a buccal swab for DNA isolation (January - July 2002). Response rate was approximately 52%; a buccal swab was returned by 1.014 participants from 303 families. The subjects who participated in this study also completed an informed consent and a questionnaire on their smoking behavior, including the FTND.
The average interval between the two FTND measurements was 1.8 years (SD .25). Of the 606 participants who took part both in the 2000 survey and in the genetic linkage study of ND, 183 participants were daily smokers, 39 participants were non-daily smokers, 173 participants were non-smokers, 214 participants were ex-smokers and for 7 participants the smoking status was unknown when they completed the 2000 questionnaire. Of the 183 daily smokers at the 2000 questionnaire, 151 (38 males, 113 females) were still smokers when participating in the DNA study (7 participants became non-daily smokers and 25 participants became ex-smokers). Of the 214 ex-smokers at the 2000 survey, 194 were still ex-smokers when participating in the DNA study (11 participants became daily smokers and 9 participants became non-daily smokers). Only participants who reported to be daily smokers at both measurements (FTND available for n=143) or who reported to be ex-smokers at both measurements (FTND available for n= 181) were included in the analyses for the test-retest correlations.

Measures
We used the Dutch version of the Fagerström Test for Nicotine Dependence. The same scoring system was employed as described in Heatherton *et al.* (1999). However, we allowed more answer categories for the question "how many cigarettes a day do you smoke?". The lowest category (10 or less cigarettes per day) was split into three categories (less than 1, 1-5 and 6-10 cigarettes per day) because the 2000 survey was also completed by (younger) participants who smoke less than once a week and participants who smoked several times per week but not every day.
Other variables considered in the analyses were:
-"At what age did you smoke your first cigarette" and "At what age did you start smoking regularly?" with answer categories 11 years or younger, 12-13 years, 14-15 years, 16-17 years, 18 years or older, never.
-"How many years did or do you smoke?"
-"How many times did you seriously try to quit?"

-"What is the longest quitting period?" with answer categories: maximum 1 week, maximum 1 month, maximum 6 month, maximum 1 year, longer than 1 year.
-Maximum number of cigarettes smoked per day: Participants (current smokers and ex-smokers) were classified as: never smokers, never smoked regularly, < 1 cigarettes per day, 1-5 cigarettes per day, 6-10 cigarettes per day, 11-20 cigarettes per day, 21-30 cigarettes per day and more than 30 cigarettes per day. The classification was constructed by taking the answers to all 5 surveys into account.
- Education level of the father: low (4 years high school), medium (high school and some years of university or polytechnic school) and high (university or polytechnic degree) education. Education of the father was used as indicator of socio-economic group.

Statistical analyses
All statistical analyses were performed using SPSS 11.5 for windows. The internal consistency of the FTND was assessed by Cronbach's alpha. To investigate the test-retest reliability, correlations between the two measurements were obtained using the Pearson-Lawley correction for selected samples (Pearson, 1903; Lawley, 1943). This correction method is used because the sub-sample which completed the FTND twice was selected for a gene-finding study of ND on the basis of scores on the first measurement. Therefore, the variances, covariances and correlations of the FTND scores in the selected sample will in general be different from those that would be obtained had the FTND been administered twice in the entire unselected sample. Using this correction method, which requires the validity of some commonly made assumptions in regression analysis, it is possible to estimate the correlation for the total sample using information from the total and selected sample.

Results
Cross-sectional analyses of the FTND for daily smokers
Figure 5.1 shows the distribution of the FTND score for smokers and ex-smokers in the total sample. For smokers, the mean score of the FTND was significantly higher for men than for women; respectively 3.02 and 2.77 (t=2.02, df=1376, p=.041). Post hoc analysis showed that men reported significantly higher values on the item 'number of cigarettes per day' (Mann-Whitney test; p=.006), scores on the other 5 items were not different for men and women. For ex-smokers, the FTND scores in men were also higher than the FTND scores in women, but differences were not statistically significant (respectively 2.22 and 1.97, t=1.72, df=1044, p=.085). Mean FTND scores of ex-smokers were lower than the mean FTND scores of smokers. The internal consistency of the FTND was reasonably high with Chronbach's alpha of .65 for male smokers, .69 for female smokers, .66 for male ex-smokers and .71 for female ex-smokers.
Correlations between FTND score and age were low for female smokers (r=.09) and not significant for male smokers and ex-smokers. Table 5.1 presents the FTND scores for type of cigarette and educational attainment of the father. The mean FTND scores was were highest when educational attainment of the father was low, somewhat

lower when educational attainment of the father was moderate and lowest when educational attainment of the father was high. Furthermore, FNTD scores were low for persons who smoked light cigarettes, somewhat higher for persons who smoked moderate nicotine cigarettes and highest for persons who smoke high nicotine cigarettes.

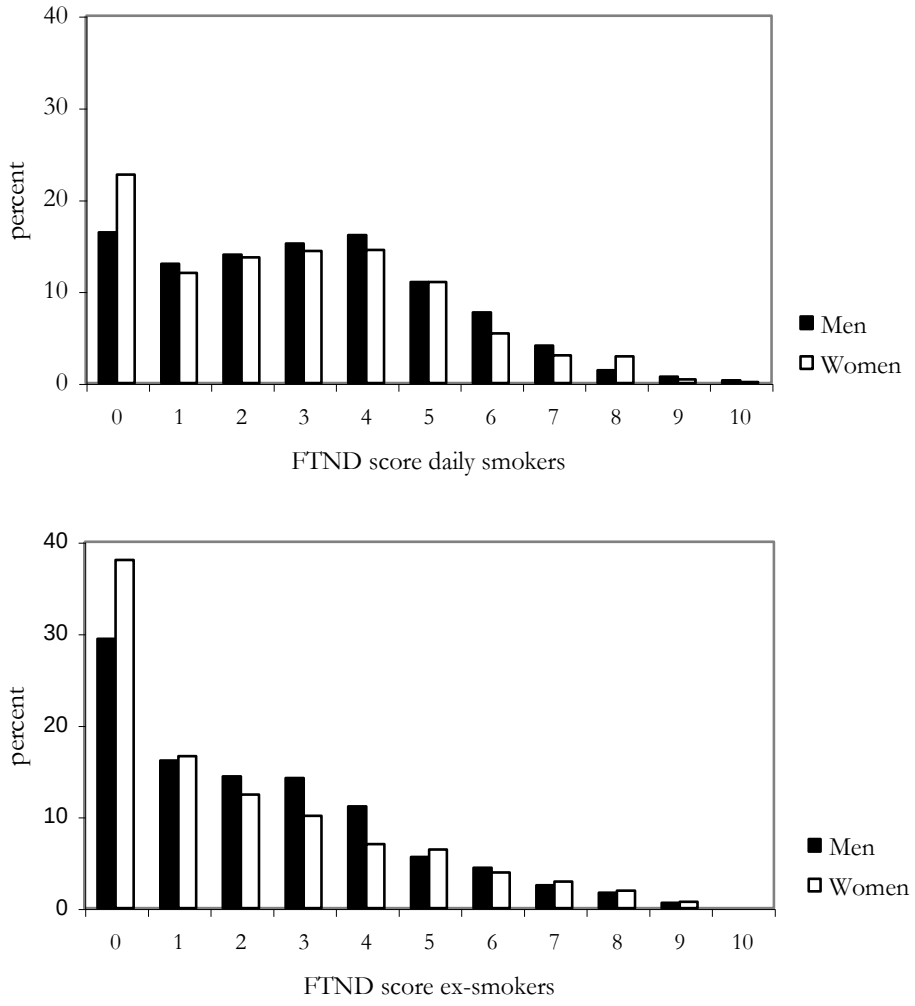


Figure 5.1: Distribution of the FTND score for daily smokers and for ex-smokers.

Table 5.1 Mean FTND score for male smokers, male ex-smokers, female smokers and female ex-smokers divided in: 1. educational attainment of the father (low, medium, high) and 2. type of cigarette (light, moderate or high nicotine cigarettes). Mean=mean FTND score, SD=Standard deviation of the mean FTND score, N = number of participants who completed the FTND

	Male smokers		Female smokers	
	Mean (SD)	N	Mean (SD)	N
Smokers				
Low educational attainment father	3.40 (2.14)	144	3.53 (2.36)	190
Medium educational attainment father	3.04 (2.19)	333	2.62 (2.21)	446
High educational attainment father	2.44 (2.26)	99	2.20 (1.95)	142
Low nicotine cigarettes	2.57 (2.32)	89	2.18 (2.16)	291
Moderate nicotine cigarettes	2.99 (2.18)	371	3.11 (2.22)	483
High nicotine cigarettes	3.44 (2.13)	123	3.53 (2.72)	17
Ex-smokers				
Low educational attainment father	2.41 (2.10)	184	2.05 (2.48)	375
Medium educational attainment father	2.32 (2.36)	60	2.04 (2.14)	139
High educational attainment father	1.69 (1.97)	77	1.79 (2.17)	101
Low nicotine cigarettes	1.47 (1.71)	51	1.31 (2.03)	218
Moderate nicotine cigarettes	2.22 (2.13)	265	2.26 (2.25)	454
High nicotine cigarettes	3.11 (2.34	44	3.67 (2.90)	12

For smokers, combining age, sex, educational attainment of the father and type of cigarette in an ANOVA analyses showed that educational attainment of the father (F= 8.02, p=.000) and type of cigarette (F=9.67, p=.000) were significantly associated with the FTND score. Age and sex were not associated with FTND score and no significant interaction effects were found. For ex-smokers, the same ANOVA analyses resulted in a significant association for the type of cigarette (F=11.49, p=.000) and a significant interaction effect between the type of cigarette and educational attainment of the father (F=11.49, p=.001). No main effects for age, sex or educational attainment of the father were shown.

Table 5.2 shows that most correlations between FTND score and smoking history variables were significant but rather low. The only exception was the maximum number of cigarettes per day which correlated highly with FTND score, even when the item 'number of cigarettes per day' was excluded from the FTND score. Around 25-30% of the smokers and ex-smokers experimented with smoking before age 14. More than one-third started smoking when 16 years or older. For all groups, age of first cigarette and age when becoming a regular smoker were negatively correlated with the FTND score. The smokers smoked on average 13 years (men 12,7 years, SD 9.1 and women 13,0 years, SD 9.3) while the ex-smokers smoked on average for 10 to 12 years (men 11.8 year, SD 10.2 and women 10.2 years, SD 7.6) before quitting. The total number of years smoked was moderately correlated with the FTND score and correlations were higher for ex-smokers than for smokers. Some smokers never tried to quit smoking (29.5% of the males and 26.6% of the females) and approximately one-fourth of the smokers reported that their longest quitting period was at most 1 week (23.1% of the male smokers, 26.6% of the female smokers). Only a small percentage succeeded in quitting for more than 1 year (9.7% of the male smokers and

12.9% of the female smokers). The FTND score was negatively correlated with the longest quitting period. For ex-smokers, the mean time between quitting smoking and completing the questionnaire was 7.5 years (SD 9.43) for men and 6.8 years (SD 8.25) for women. For men only, the time between quitting smoking and completing the questionnaire was negatively correlated with their FTND score. For smokers, the mean number of quit attempt was 1.6 (SD 1.8) for males and 1.8 (SD 2.1) for females. The FTND score was not correlated with the number of quit attempts in smokers. The male ex-smokers tried to quit smoking 1.5 times (SD 2.8) before they succeeded and the female ex-smokers tried 2.0 times (SD 6.4). Only for male ex-smokers, a low, negative correlation was found between the number of quit attempts and the FTND score.

Table 5.2 Correlation among FTND scores of daily smokers and smoking history.

	Smokers		Ex-smokers	
	Males	Females	Males	Females
Age first cigarette	-.19**	-.08*	-.11**	-.17**
Age regular smoking	-.26**	-.19**	-.20**	-.22**
Number of years smoked	.10*	.17**	.19**	.35**
Number of quit attempts	-.03	-.06	-.23**	-.08
Longest quitting period	-.14**	-.17**	n.a.	n.a.
Number of months since quitting	n.a.	n.a.	-.20**	-.07
Maximum number of cigarettes smoked per day	.70**	.66**	.79**	.75**
Maximum number of cigarettes smoked per day (with FTND score without number of cigarettes/day)	.51**	.50**	.60**	.59**

* Correlation is significant at the P<.05 level (two tailed)
** Correlation is significant at the P<.01 level (two tailed)
n.a. = not applicable

Test-retest sample

In the test-retest sample, 183 participants were daily smokers when completing the 2000 questionnaire. Of the 183 participants, 26 had quit smoking at the second measurement. For those 26 participants, the mean FTND score at the 2000 questionnaire was 2.35 (SD 2.5) while for the smokers who were still daily smokers at the second measurement, the mean FTND score at the 2000 questionnaire was 3.81 (SD 2.5). Furthermore, in the test-retest sample, 214 participants were ex-smokers when completing the 2000 questionnaire. For the ex-smokers in 2000 who became daily smokers at the second measurement (n=11), the mean FTND score at the 2000 questionnaire was 4.00 (SD 2.5) while the mean FTND score at the 2000 questionnaire for the ex-smokers who were still ex-smokers at the second measurement was 2.72 (SD 2.4).

For the test-retest analyses of the FTND, the participants who were daily smokers at both measurements (n=151) and the participants who were ex-smokers at both measurements (n= 180) were selected. The mean FTND score of the first measurement was not significantly different from the mean FTND score of the second measurement (Table 5.3).

The test-retest correlations between the FTND scores at both measurements with Pearson-Lawley correction were .70 for male smokers, .83 for female smokers, .91 for male ex-smokers and .83 for female ex-smokers. However, the correlations did not differ

much from the regular Pearson product-moment correlations (.72 for male smokers, .85 for female smokers, .92 for male ex-smokers and .86 for female ex-smokers).
The time between the two measurements varied from a half year to two and a half years. The mean difference was one year and 9 months. No decrease or increase in FTND score was observed as function of the time between the first and second measurement for men or women, ex-smokers or smokers.

Table 5.3 Paired t-test to compare mean FTND score at measurement 1 with mean FTND score at measurement 2. The table presents the mean FTND scores, number of participants, t-value, degrees of freedom (df) and p-value.

	FTND score measurement 1	FTND score measurement 2	N participants	t	df	Sig (2-tailed)
Male smokers	4.66	4.74	38	-.292	37	.772
Female smokers	3.39	3.64	113	-1.88	112	.063
Male ex-smokers	3.00	2.76	58	1.95	57	.056
Female ex-smokers	2.61	2.40	122	1.80	121	.075

Discussion
The mean FTND scores in the sample of daily smokers were comparable to other recent population studies (Fagerström *et al.*, 1996; Etter *et al.*, 1999; John *et al.*, 2003). Studies kkwith smokers seeking cessation assistance have reported higher FTND scores (Fagerström *et al.*, 1996), but it is likely that subjects seeking cessation assistance are more nicotine dependent than the general population. The finding that males scored higher on the FTND than females was also in line with other studies (Fagerström *et al.*, 1996; John *et al.*, 2003). Interestingly, post-hoc analyses showed that the gender differences were limited to one of the six items: the number of cigarettes per day. Studies have shown that cigarette nicotine dose may be less important for the reinforcing effects of smoking for women compared to men (Perkins, 1999; Perkins *et al.*, 2002). Less is known with regard to the performance of the FTND in populations of ex-smokers. Our results showed that the mean FTND scores of ex-smokers who reported on the period they smoked the heaviest were not significantly different for men and women and were lower than the mean scores of smokers. Fagerström *et al* (1996) quoted a written communication on two surveys that also showed former daily smokers to have lower dependence levels than current daily smokers. It is likely that those smokers who succeed in smoking cessation are the ones that are less nicotine dependent.
Our data showed a reasonably high internal consistency for the Dutch version of the FTND in daily smokers, which is in line with the study of Heatherton *et al.* (1991) and later studies who reported internal consistencies varying from .56 to .71 (Heatherton *et al.*, 1991; Payne *et al.*, 1994; Pomerleau *et al.*, 1994; Dijkstra and Tromp, 2002). No studies have investigated the internal consistency of the FTND in ex-smokers. Our results showed that the internal consistency of the FTND for ex-smokers was comparable with the internal consistency of the FTND for smokers.
Several factors are associated with the FTND score, both in smokers and ex-smokers. Educational attainment of the father was significantly associated with the FTND score. An Italian study showed that more educated smokers were less dependent (Gallus *et al.*,

2002). A study investigating the lifetime risk of nicotine dependence did not find an association with the total number of years of education (Breslau *et al.*, 2001) although it should be noted that they used different measures. Furthermore, FTND score was associated with the type of cigarette. Etter *et al* (2003) also found that smoking mild, light or ultra light versus regular cigarettes was associated with a lower Fagerström Dependence score (Etter *et al.*, 2003). It should be noted though, that smokers who use reduced nicotine yield products can increase the amount of nicotine extracted from each cigarette by taking more puffs per cigarette or inhaling a larger puff volume more deeply into the lungs (Thun and Burns, 2001) and that those aspects are not measured by the FTND. Remarkably, we did not find a correlation between FTND score and age (except a very low correlation for smoking women). A study with current smokers aged 20 years and older also concluded that the FTND score was independent of age (John *et al.*, 2003).
The correlations between FTND and smoking history variables were significant but low. The lower the age of smoking the first cigarette and the age when starting regular smoking as well as the higher the number of years smoked, the higher the FTND score. This is in line with two studies that have reported that early age of smoking initiation is related to more frequent current smoking, daily smoking and more dependent smoking (Everret *et al.*, 1999; Lando *et al.*, 1999) and with two studies that have found correlations between FTND score and number of years smoked ranging between .09 and .17 (Horn *et al.*, 2003; John *et al.*, 2003). Our results showed stronger correlations between FTND score and number of years smoked for ex-smokers than for smokers. However, this could be due to the age-difference between the smokers and the ex-smokers; mean age of the smokers was around 30 while the mean age of the ex-smokers was around 38 years.
For both smokers and ex-smokers, the FTND score was highly correlated with the maximum number of cigarettes smoked per day, and although correlations were somewhat lower when the item 'number of cigarettes per day' was removed from the FTND score, they were still relatively high. Thus, part of the variance in FTND score is explained by the maximum number of cigarettes smoked per day (25-26% for smokers and 34-36% for ex-smokers), but this also indicates that the FTND measures additional aspects of nicotine dependence. Etter *et al* (1999) suggested that the FTND does not address all aspects of nicotine addiction, such as unsuccessful attempts to quit smoking. Our results confirm this suggestion for smokers and female ex-smokers but for male ex-smokers a correlation of -.23 was found. A moderate, negative correlation was found for daily smokers between the longest quitting period and the FTND score. Results suggest that it is harder to quit for persons who are more nicotine dependent. For ex-smokers, the time between quitting smoking and completing the questionnaire was negatively correlated with the FTND score for men only. Possibly, men who quit long ago were less nicotine dependent while men who quit a short time ago were more nicotine dependent and may be at higher risk to start smoking again. Another explanation is that men change the perception of their smoking behavior as the time between quitting smoking and completing the questionnaire increases. For women, other aspects might be involved in quitting smoking, such as being pregnant.
In conclusion, although the strength of the correlations for smoking history and FTND score differ slightly between smokers and ex-smokers, the direction of the correlations is

the same. This suggests that the FTND questionnaire measures more or less the same in smokers and ex-smokers.
Results from the test-retest study confirm this observation. Test-retest correlations of the Dutch FTND were high, both in current smokers and ex-smokers. The test-retest correlations are in line with two other studies that have reported test-retest correlations for smokers (Pomerleau *et al.*, 1994; Etter *et al.*, 1999). As far as we know, no reports of test-retest correlations in samples of ex-smokers are published. Our results showed that the test-retest correlations for ex-smokers were comparable with the test-retest correlations for smokers.
The time between the two measurements ranged from a half year to two and a half years. No correlation was found between the absolute difference of the FTND scores and the amount of time between administration. This finding suggest that the FTND can also be used in follow up studies with longer follow up periods.
Interestingly, the test retest sample showed that smokers in 2000 who became ex-smokers at the next measurement had lower FTND scores in 2000 than the smokers who were still smokers at the next measurement. The ex-smokers in 2000 who became smokers at the next measurement had higher FTND scores in 2000 than the ex-smokers who were still ex-smokers at the second measurement. This is in line with our findings described above that ex-smokers are less nicotine dependent compared to smokers.
This study was carried out with data from twins and siblings. Collecting data on twins offers the opportunity to investigate the heritability of ND, which will be addressed in a next step. To be sure that the smoking prevalence in twins was the same as in singletons we compared both groups and did not find significant differences in prevalence of smoking. A study of non-response bias in our sample (of twins and siblings) showed a significantly higher percentage smokers among the non-respondents when smoking status was obtained using a reply-card. However, when another method was used (comparison between complete/incomplete twin pairs or highly/less cooperative families), differences were not significant (Vink *et al.*, 2004-b).
In this paper, data on smokers and ex-smokers were presented. It should be noted that for the ex-smokers FTND data were retrospectively obtained.
In conclusion, our data suggest sufficient internal consistencies and test-retest reliabilities for the Dutch version of the FTND in daily smokers. The performance of the FTND in ex-smokers was comparable with the performance of the FTND in smokers. Together, these findings suggest the FTND is a reliable questionnaire to measure nicotine dependence in smokers as well as ex-smokers. The FTND can be a valuable tool as a measure of the degree of nicotine dependence, although it should be noted that it may not address all aspects of nicotine addiction.

Chapter 6

Heritability of smoking initiation and nicotine dependence

Jacqueline M.Vink, G. Willemsen, D.I. Boomsma. Heritability of smoking initiation and nicotine dependence. Behavior Genetics (under revision)

Abstract
In contrast to other aspects of smoking behavior, little attention has been paid to the genetics of nicotine dependence. In this paper, three models (single liability dimension, independent liability dimension and combined model) have been applied to data on smoking initiation and nicotine dependence (n=1572 Dutch twin pairs, mean age 30.5). Results showed that a combined model best fitted the data. This model postulates an initiation dimension and a dependence dimension, but those dimensions are not independent. For both males and females, individual differences in smoking initiation were explained by genetic factors (44%), shared environmental influences (51%) and unique environmental influences (5%) while the nicotine dependence dimension was only influenced by genetic factors (75%) and by unique environmental factors (25%). The substantial impact of genetic factors on nicotine dependence emphasizes the need for further research to localize and identify specific genes and pathways involved in nicotine dependence.

Introduction

It is well established that genetic factors contribute to individual differences in smoking behaviour (Heath and Madden, 1995; Sullivan and Kendler, 1999; Hopfer *et al.*, 2003; Li *et al.*, 2003). Several, possibly associated, dimensions of smoking behavior may be distinguished, e.g. smoking initiation, number of cigarettes smoked per day and nicotine dependence (Mayhew *et al.*, 2000).

Nicotine dependence can only be assessed in individuals who have initiated smoking but not every person who initiates smoking becomes nicotine dependent. Also, not every individual who never initiated smoking can be assumed to score zero on the dependence dimension. For genetic research, it is important to investigate whether smoking initiation and nicotine dependence are part of the same continuum or whether they represent two independent dimensions. An incorrect definition of the phenotype could possibly lead to biased estimates of the genetic and environmental factors (Heath and Martin, 1993; Heath *et al.*, 2002). If the same genetic and environmental factors determine whether or not a person initiates smoking and how dependent a person becomes, then exclusion of non-smokers could lead to truncation of the distribution and as a consequence to biased estimates of the heritability. Conversely, if the determinants of smoking initiations are independent of the determinants of nicotine dependence, then inclusion of non-smokers in the analyses of dependence may confound two traits with different modes of inheritance (Heath and Martin, 1993; Heath *et al.*, 2002).

As far as the authors know, only one study has addressed the heritability of nicotine dependence. Kendler *et al* (1999) investigated the relationship between smoking initiation and nicotine dependence by using a model that estimates the correlation between the liability to smoking initiation and the liability to nicotine dependence, given smoking initiation. Results showed that while the majority of genetic risk factors for nicotine dependence were shared with smoking initiation, a distinct set of familial factors solely influenced the risk for nicotine dependence. Kendler *et al.* (1999) reported that genetic factors contributed to a total of 72% of the variance in liability to nicotine dependence and the remaining variance was explained by unique environmental factors. However, this study was performed in women only.

The number of cigarettes (quantity) is often used as proxy measure for nicotine dependence and both phenotypes are highly correlated (Vink *et al.*,). In a study of the heritability of smoking initiation and quantity in adolescent twins we considered three models: (a) a single liability model which assumes that the same genetic and environmental risk-factors influence initiation and quantity, (b) an independent liability model, which assumes independent initiation and persistence dimensions, each determined by separate genetic and environmental risk factors, (c) a combined model, which assumes two dimensions; an initiation dimension and a quantity dimension, but those two dimensions are not independent of each other (Koopmans *et al.*, 1999; Vink *et al.*, 2004-a). Results showed the combined model to be the best fitting model. Koopmans et al. (1999) reported that 39% of the total variance in smoking initiation was explained by genetic influences and 86% of the total variance in quantity (number of cigarettes per day) was explained by genetic factors.

In this paper, the three different models described by Koopmans *et al.* (1999) are applied to data on smoking and nicotine dependence in a Dutch twin sample aged 30.5 years (SD 11.9). After identification of the correct liability model, the relative contribution of genetic and environmental factors to initiation and nicotine dependence is estimated.

Methods

Subjects

This study is part of an ongoing twin/family study on health-related behavior of the Netherlands Twin Register (NTR) that assesses families with adolescent and young adult twins every two to three years since 1991 (Boomsma *et al.*, 2002). Data were collected by mailed surveys in 1991, 1993, 1995, 1997 and 2000. The surveys contained items on health, personality and lifestyle (a.o. smoking behavior). For this paper, the data from the 2000 survey (which included the Fagerström Test for Nicotine Dependence (FTND) for the first time) were used. Completed questionnaires were returned by 4610 twins/triplets. In 2001, additional data on smoking and FTND were obtained by telephone interviews (n=56 twins/triplets): additional family members were called when only 1 family member participated in the questionnaire study and participants of the 2000 survey were called to check the stability of the FTND. Data on smoking collected with questionnaires completed by participants in a study of the genetics of nicotine dependence were also included (n=426 twins/triplets). In total, 4672 twins participated at least once. Data of both smokers and ex-smokers were included (Vink *et al.*, in press). Smokers completed the FTND on their current situation while ex-smokers completed the FTND on the period they smoked the heaviest. If subjects participated more than once the highest FTND score was used for the analysis. For 24 persons smoking data were missing and for 208 additional persons known to have initiated smoking, FTND data were not available. The remaining 4440 persons were classified as never-smoked, low dependent (highest FTND score 0-2), moderate dependent (highest FTND score 3-5) and highly dependent (highest FTND score >= 6). Data were only included when smoking data (smokers or ex-smokers) of both twins were available, so analyses were performed using the smoking data of 1572 twin pairs.

Zygosity was based on questionnaire data, or when available, on DNA typing. For 29.8% of the same sex twin pairs information on their zygosity was available based on DNA polymorphisms. Agreement between zygosity based on questionnaire data and zygosity based on DNA results was 96%. There were 238 monozygotic male (MZM) twin pairs, 125 dizygotic male (DZM) twin pairs, 630 monozygotic female (MZF) twin pairs, 288 dizygotic female (DZF) twin pairs and 291 dizygotic opposite sex (DOS) twin pairs who had complete data on smoking and zygosity. The mean age was 30.5 years (SD 11.9).

Measures

Smoking initiation: subjects were classified as never smokers when they reported they never smoked or when they tried smoking but never smoked regularly. Data on smoking behavior were collected longitudinally (1991, 1993, 1995, 1997, 2000) and 61% of the twins participated more than once. The answers to all surveys were taken into account.

Nicotine Dependence: To measure the degree of nicotine dependence the Fagerström Test for Nicotine Dependence (FTND) was used (Heatherton *et al.*, 1991). The FTND consists of 6 items and produces a score ranging from 0 to 10 with higher scores indicating more nicotine dependence. The FTND includes items like 'How soon after you wake up do you smoke your first cigarette?' and 'Do you find it difficult to refrain from smoking in places where it is forbidden?'.

Genetic analysis

To investigate the inheritance of smoking behavior, the trait was considered to have an underlying, continuous liability. The variation of the liability is both genetic and environmental in origin (Falconer and Mackay, 1996a). Thresholds divide this normal liability distribution into discrete categories. We consider three models for the relationship between smoking initiation and nicotine dependence (figure 6.1).

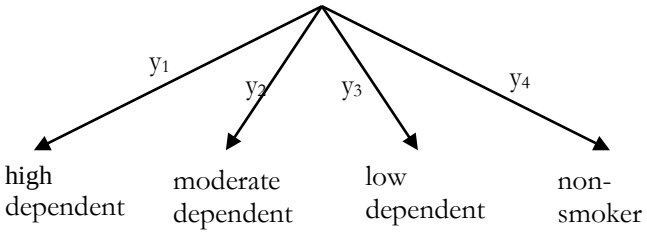
Single liability dimension model (SLD):

The single liability dimension (SLD) model postulates that the liability to smoking behavior is unidimensional and is normally distributed. Under this model the same genetic and environmental factors predispose to smoking initiation and to nicotine dependence.The underlyinnormal distribution is divided by thresholds into discrete categories which, in the case of the SLD model, corresponds to the observed categories. The probability that an individual falls in one of the four categories is given by y_1, y_2, y_3 and y_4 and can be calculated by integrating a standardized normal distribution between the corresponding threshold values. The model predicts that the co-twins of nicotine dependent participants are more likely to be nicotine dependent than the co-twins of non-smokers.

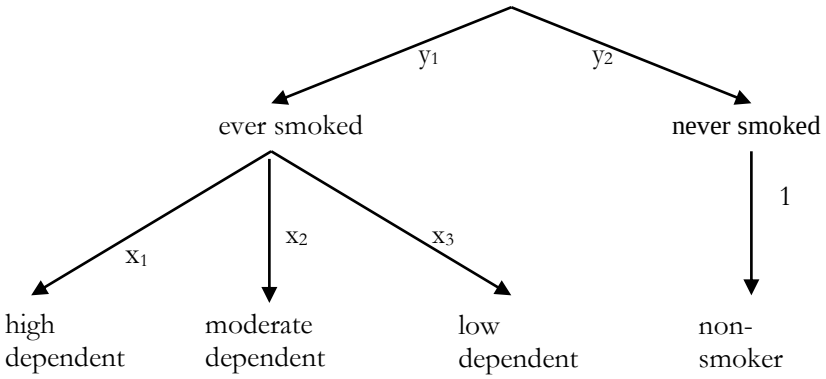
Independent liability dimension (ILD) model:

The independent liability dimension model assumes two independent liability dimensions for smoking initiation and nicotine dependence. The initiation dimension determines the probability that an individual initiates smoking (y_1) or never starts smoking (y_2). Individuals falling below the threshold are predicted to be smokers. The nicotine dependence dimension determines whether an individual becomes highly dependent (x_1), medium to low dependent (x_2) or very low dependent (x_3). Taking smoking initiation into account, the probabilities that an individual becomes highly dependent, moderate dependent or low dependent are y_1x_1, y_1x_2 and y_1x_3, respectively. The probability that an individual remains a non-smoker is y_2. The ILD model predicts that the co-twin of a twin who never smoked is more likely to abstain from smoking. Also, if the co-twin initiated smoking while the twin never smoked, the co-twin will not, on average, be less nicotine dependent as the co-twin of a nicotine dependent twin.

(a) Single Liability model:



(b) Independent liability model:



(c) Combined model:

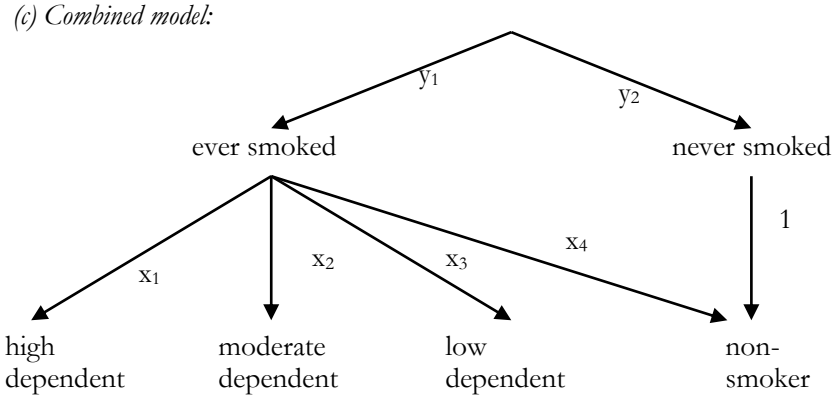


Figure 6.1 Schematic representation of (a) the single liability model, (b) the independent liability model and (c) the combined model for smoking initiation and nicotine dependence. x_i and y_i are the probabilities that an individual falls in one of the categories.

Combined model (CM):
The combined model includes features of both the SLD and the ILD models. Like the ILD, it postulates the existence of separate initiation and dependence dimensions. It also allows the possibility that there are some genetic and environmental risk-factors which influence both smoking initiation and nicotine dependence. Under the combined model, there are two different routes to non-smoker. Thus, under this model the co-twin of a twin who initiated smoking is more likely to become a non-smoker than under the ILD model.

Model fitting
Smoking behavior in the first twin was cross-classified with smoking behavior in the second twin, resulting in 4x4 contingency tables for each zygosity group; monozygotic males (MZM), dizygotic males (DZM), monozygotic females (MZF) and dizygotic opposite sex twins (DOS), (see also table 2). The three models were fitted to the five contingency tables by methods of maximum likelihood with the structural equation modeling package Mx (Neale *et al.*, 1999). The thresholds were allowed to be different for males and females.
Under the SLD model one twin correlation for each zygosity group (5) and three thresholds for males and three thresholds for females were estimated, giving in total 11 parameters to be estimated. Under the ILD, separate twin correlations for the initiation and dependence dimensions were estimated for each zygosity group. For the initiation dimension one threshold for males and one for females was estimated. There was no 'non-smoker' category for the dependence dimension, leaving two thresholds for males and two for females to be estimated. This means, 16 parameters were estimated under the ILD model. Using the CM, the same parameters were estimated as in the ILD model, except for the thresholds: in the dependence dimension three thresholds for males and three for females were estimated because non-smokers were also included in the dependence dimension. So under the CM, 18 parameters were estimated.
The predicted probabilities for a twin pair under the three models are presented in Table 6.1. Under the SLD model, y_{11} denotes the probability that both twins are highly nicotine dependent, y_{12} denotes the probability that the first twin is highly dependent and the second twin is moderate dependent on nicotine, and so on. Under the CM and ILD model y_{11} denote the probability that twins both initiated smoking, y_{22} denotes the probability that both twins not initiated smoking, and y_{12} and y_{21} denote the probabilities that twins are discordant for smoking initiation. The conditional probabilities that both twins are highly dependent on nicotine, the first twin is highly dependent and the second twin is moderate dependent is represented by x_{11}, x_{12} etc. Under the CM there are two routes to 'non-smoker'. For example, $y_{11}x_2x_4$ gives the probability that both twins are smokers on the initiation dimension (y_{11}) and the first twin is moderate to low dependent on nicotine while the second twin is a non-smoker on the dependence dimension, $y_{21}x_{1.}$ gives the probability that the first twin is a smoker on the initiation dimension while the second twin is a non-smoker on the initiation dimension (y_{21}), and the first twin is highly dependent on nicotine ($x_{1.}$).

The three models were fitted to the data, estimating separate polychoric correlations for each zygosity group. The goodness-of-fit of nested models was assessed with the likelihood-ratio statistic, this statistic is distributed as χ2-square.

Table 6.1 Predicted probabilities for a twin pair under the single liability dimension (SLD), the independent liability dimension (ILD) and the combined model (CM).

		twin 2 →			
twin 1 ↓	model	FTND >=6	FTND 3-5	FTND 0-2	Non-smoker
FTND >=6	SLD	y_{11}	y_{12}	y_{13}	y_{14}
	ILD	$y_{11}x_{11}$	$y_{11}x_{12}$	$y_{11}x_{13}$	$y_{12}x_{1.}$
	CM	$y_{11}x_{11}$	$y_{11}x_{12}$	$y_{11}x_{13}$	$y_{12}x_{1.} + y_{12}x_{14}$
FTND 3-5	SLD	y_{21}	y_{22}	y_{23}	y_{24}
	ILD	$y_{11}x_{21}$	$y_{11}x_{22}$	$y_{11}x_{23}$	$y_{12}x_{2.}$
	CM	$y_{11}x_{21}$	$y_{11}x_{22}$	$y_{11}x_{23}$	$y_{12}x_{2.} + y_{11}x_{24}$
FTND 0-2	SLD	y_{31}	y_{32}	y_{33}	y_{34}
	ILD	$y_{11}x_{31}$	$y_{11}x_{32}$	$y_{11}x_{33}$	$y_{12}x_{3.}$
	CM	$y_{11}x_{31}$	$y_{11}x_{32}$	$y_{11}x_{33}$	$y_{12}x_{3.} + y_{11}x_{34}$
non-smoker	SLD	y_{41}	y_{42}	y_{43}	y_{44}
	ILD	$y_{21}x_{.1}$	$y_{21}x_{.2}$	$y_{21}x_{.3}$	y_{22}
	CM	$y_{11}x_{41}$	$y_{11}x_{42}$	$y_{11}x_{43}$	$y_{22} + y_{11}x_{44} + y_{21}x_{.4} + y_{12}x_{4.}$

Under the SLD model, y_{jk} = the probability that a twin pair falls in the j,k-th category of smoking behavior. For example, y_{11} is the probability that both twin 1 and twin 2 fall in the first category (FTND >=6). Under the CM and ILD model, y_{jk} = the probability that a twin pair falls in the j,k-th category of the initiation dimension; x_{jk} = the probability that a twin pair falls in the j,k=th category of the nicotine dependence dimension; $x_{j.}$ = the probability that the first twin falls in the j=th category of the nicotine dependence dimension; $x_{.k}$ = the probability that the second twin falls in the k-th category of the nicotine dependence dimension.

Genetic models

The three models were fitted to the data, and for the model that gave the best description of the data, the twin correlations in liability were expressed as a function of genetic and environmental parameters based on the classical twin design (Neale and Cardon, 1992).

For both the initiation and the nicotine dependence dimensions, different genetic models were fitted. Sources of variation that were considered in modeling were additive genetic variation (A), shared environmental variation (C) and unique environmental variation that is not shared by twin pairs (E). Under the full model, both additive genetic and shared environmental factors contribute to resemblances between twins. Sex-differences were tested by allowing the magnitude of the genetic and environmental effects to be different for males and females and by allowing the correlation between the genetic factors in opposite-sex twins to be less than unity. For all models, different thresholds were estimated for males and females, allowing for differences in the prevalence of smoking between males and females.

Results

Table 6.2 shows concordance rates and the proportions of never-smokers, low dependent, moderate dependent and high dependent individuals for the first and

second twin in each zygosity group. Concordance is higher in MZ twins than in DZ twins. The three different models (SLD, ILD and CM) were fitted to the data. Table 6.3 shows the goodness-of-fit for each liability model. The ILD fitted somewhat better to the data than the SLD but the combined model gave the best description of the data. Therefore, the combined model was used when further investigating the genetic and environmental influences on nicotine dependence.

Table 6.2 Twin concordance (and proportions) for nicotine dependence for each zygosity group. FTND= score on Fagerström Test for Nicotine dependence (including both smokers and ex-smokers), MZM= monozygotic male twin pairs, DZM= dizygotic male twin pairs, MZF= monozygotic female twin pairs. DZF= dizygotic female twin pairs, DOS= dizygotic opposite sex twin pairs (males horizontal, females vertical).

		Twin 2				
Twin 1		FTND >= 6	FTND 3-5	FTND 0-2	Never smoked	*%*
MZM	FTND >= 6	1	8	2	2	*5.5*
	FTND 3-5	3	14	10	4	*13.0*
	FTND 0-2	2	6	33	9	*21.0*
	Never smoked	0	3	13	128	*60.5*
	%	*2.5*	*13.0*	*24.4*	*60.1*	*n=238*
DZM	FTND >= 6	2	3	1	3	*7.2*
	FTND 3-5	2	5	4	4	*12.0*
	FTND 0-2	3	3	17	8	*24.8*
	Never smoked	0	4	14	52	*56.0*
	%	*5.6*	*12.0*	*28.8*	*53.6*	*n=125*
MZF	FTND >= 6	14	9	5	3	*4.9*
	FTND 3-5	10	35	19	9	*11.6*
	FTND 0-2	2	18	81	40	*22.4*
	Never smoked	2	9	39	335	*61.1*
	%	*4.4*	*11.3*	*22.9*	*61.4*	*n=630*
DZF	FTND >= 6	7	3	5	3	*6.2*
	FTND 3-5	2	10	17	14	*14.9*
	FTND 0-2	4	6	28	19	*19.8*
	Never smoked	4	12	31	123	*59.0*
	%	*5.9*	*10.8*	*28.1*	*55.2*	*n=288*
DOS	FTND >= 6	2	3	8	6	*6.5*
	FTND 3-5	3	13	19	16	*17.5*
	FTND 0-2	1	11	21	33	*22.7*
	Never smoked	8	6	23	118	*53.3*
	%	*4.8*	*11.3*	*24.4*	*59.5*	*n=291*

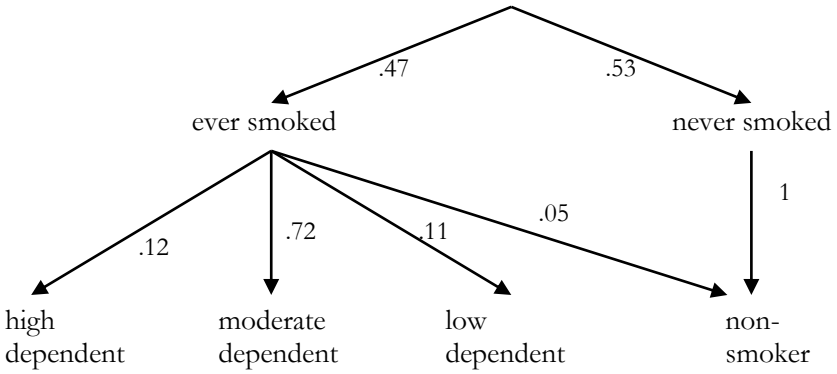
Table 6.3 Goodness-of-fit of the single liability dimension (SLD), the independent liability dimension (ILD) and the combined model (CM) to the data on smoking.

	df	χ^2	p	AIC
SLD	64	132.40	< .001	4.40
ILD	59	87.87	.009	-30.13
CM	57	61.90	.306	-52.10

Df= degrees of freedom, AIC=χ^2- 2df, this is a measure of the parsimony of the model, al lower AIC indicates a more parsimonious model.

Under the combined model, an individual can be a non-smoker due to genetic and/or environmental factors that influence the smoking initiation dimension or because the individual is low on the nicotine dependence dimension. The predicted marginal probabilities for smoking under the full combined model are represented in figure 6.2. The full model allows for sex-differences. Under the full model, the probability of becoming a highly dependent smoker in the total sample is 6% for males (.12*.47) and 5% for females (.09*.55). The probability of being a non-smoker while smoking is initiated is 2.3% in males (0.05*0.47) and 7.1% in females (0.13*0.55) while the probability of being a non-smoker when having never initiated smoking is 53% for males and 45% for females.

(a) males:



(b) females:

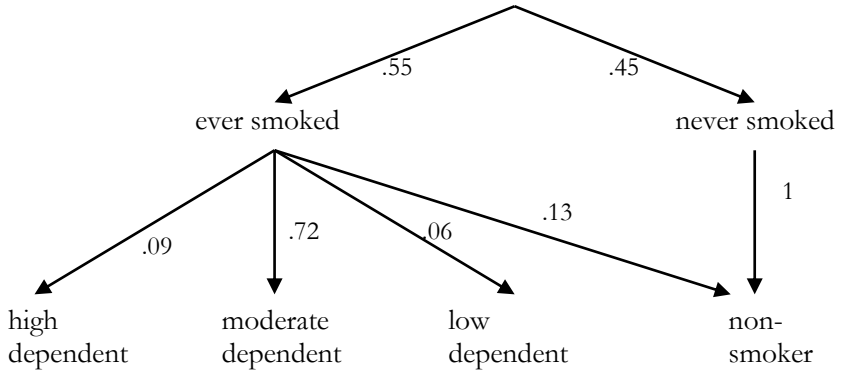


Figure 6.2 Estimated probabilities under the full combined model in (a) males and (b) females.

Table 6.4 shows the estimated polychoric correlations for each zygosity group for the initiation and dependence dimension under the full combined model. For smoking initiation, the correlations between MZ twins were somewhat higher than the correlations between DZ twins suggesting both genetic and shared environmental influence on smoking initiation. For the dependence dimension, the difference between the correlations in MZ pairs and the correlations in DZ pairs is somewhat larger for females than for males, suggesting that genetic factors are possible more important for females.

Table 6.4 Estimated polychoric twin pair correlations with 95% confidence intervals for the initiation and dependence dimensions under the full combined model.

	Initiation		Nicotine dependence	
	R	95% CI	R	95% CI
MZM	.96	.83 -1.00	.61	.23 - .83
DZM	.75	.40 - .98	.50	.23 - .80
MZF	.94	.81 - .99	.80	.67 - .89
DZF	.75	.26 -1.00	.48	.06 - .77
DOS	.70	.38 -1.00	.32	-.16 - .67

Different genetic models were fitted both to the initiation dimension and to the nicotine dependence dimension under the combined model. Sources of variation that were considered in modelling were additive genetic variation (A), shared environmental influences (C) and a unique environmental influence (E) that is not shared by family members. The full model permitted sex-differences by allowing the magnitude of the genetic and environmental effects to be different in males and females and by allowing the correlation between the genetic factors in opposite-sex twins to be less than one. The results are shown in Table 6.5. The first model is a full model with an additive genetic factor (A), shared environment (C) and a unique environmental factor (E) for both dimensions. Constraining A, C and E to be equal for both sexes in the initiation dimension improved the fit of the model (model 2). Removing additive genetic factors or shared environmental factors from the initiation dimension gave a significant reduction in the goodness of fit of the model (model 3 and 4 respectively). The initiation dimension was best described by an ACE model without sex differences (model 2). For the nicotine dependence dimension, the full model could be reduced to an AE model without sex differences (model 6). Overall, the best fitting model was an ACE model without sex-differences for the initiation dimension and an AE model without sex-differences for the nicotine dependence dimension (model 8).

The best fitting model showed that individual differences in smoking initiation could be explained by genetic influences (44%), by shared environmental factors (51%), and by unique environmental factors (5%). The nicotine dependence dimension was largely influenced by genetic factors (75%) and the remaining variance was explained by unique environmental factors (25%).

Tables 6.5 Model fitting results for a combined model with smoking initiation and nicotine dependence (best fitting model is given in boldface).

	initiation	nicotine dependence	χ^2	df	P	AIC
1.	full	full	61.89	57	.306	-52.11
2.	ACE	full	61.90	60	.408	-58.10
3.	AE	full	77.22	61	.079	-44.78
4.	CE	full	72.29	61	.153	-49.71
5.	full	ACE	65.76	60	.284	-54.24
6.	full	AE	66.62	61	.290	-53.38
7.	full	CE	82.71	61	.034	-37.29
8.	**ACE**	**AE**	**67.84**	**64**	**.348**	**-60.16**

Full= full model with sex-dependent effects; ACE=full model without sex differences; AE= additive genetic model; CE= shared environmental model; AIC=χ^2 – 2 df, this is a measure of the parsimony of the model, a lower value of AIC indicates a more parsimonious model.

Discussion

The present study simultaneously investigated the heritability of smoking initiation and nicotine dependence. First, three different mutifactorial threshold models were fitted to the data on smoking to explain smoking initiation and nicotine dependence. The SLD model was rejected indicating there is not one underlying continuum of liability to smoking. The ILD model also fitted poorly to the data indicating that the smoking initiation dimension and the nicotine dependence dimension are not independent. The combined model was the best fitting model. Under this model there are two routes to non-smoking: an individual can be a non-smoker due to genetic and/or environmental factors that influence the initiation dimension or because that individual is low on the nicotine dependence dimension. The predicted marginal probabilities for smoking were estimated under the full combined model which included sex-differences. Results showed that under the full combined model only a small proportion of the male twins were non-smokers due to the genetic and environmental risk factors which influence the nicotine dependence dimension, for the female twins this proportion was somewhat higher. Sex differences were tested in the genetic models which were fitted both to the initiation dimension and to the nicotine dimension. The sources of variation that were investigated were additive genetic variation (A), shared environmental influences (C) and a unique environmental influence (E)). Variation in the initiation dimension was best described by an ACE model, while variation in the nicotine dependence dimension could be described with an AE model.

For smoking initiation, 44% of the variation could be explained by genetic factors, 51% by shared environmental factors and 5% by unique environmental factors in both males and females.

Numerous twin studies to the genetic and environmental contribution to smoking initiation have been reported in the literature (reviewed by (Heath and Madden, 1995; Hopfer *et al.*, 2003))). Given the differences in age, sex, smoking measures and statistical models used in each study it is hard to compare these estimates directly across different studies. Li *et al.* (2003) selected six studies of smoking initiation in

adults for a meta-analyses. Results indicated that the parameter estimates for h^2, c^2 and e^2 were .37, .49 and .17 for male adults and .55, .24 and .16 for female adults, respectively (Li *et al.*, 2003). The heritability estimates in the present study are in line with the results of the meta-analysis of Li *et al.* (2003). The heritability estimates in a review paper of Sullivan and Kendler (1999) were somewhat higher. Sullivan and Kendler (1999) calculated the weighted means of 10 studies to smoking initiation and reported that the weighted mean heritability was 56%, the weighted mean shared environmental influence .24 and the weighted mean individual-specific environmental influence .20 (Sullivan and Kendler, 1999).

In the literature, relatively little attention is paid to the genetics of nicotine dependence. Our results show that in both sexes 75% of the variation in nicotine dependence was explained by genetic factors. The remaining variance was explained by unique environmental factors. Those findings closely resemble those of Kendler *et al.* (1999) who reported, in a sample of female twins, that genetic factors contributed to a total of 72% of the variance in liability to nicotine dependence and the remaining variance is explained by unique environmental factors. Model fitting results in the present paper showed that the differences in heritabilities between males and females were not significant.

Several other measures, like smoking persistence or quantity, are often used as proxy for nicotine dependence. In a larger, partly overlapping sample, we have found for the maximum number of cigarettes per day that the parameters h^2, c^2 and e^2 were 51%, 30% and 18%, respectively (Vink *et al.*, 2004-a). Li *et al.* (2003) performed a meta-analysis for smoking persistence (including studies to persistence, quantity, dependence and regular use) and found that the parameters h^2, c^2 and e^2 were 59%, 8% and 37% for male adults and 46%, 28% and 24% for female adults, respectively (Li *et al.*, 2003). In a review study of Sullivan and Kendler (1999) the weighed mean heritability for proxy measures of nicotine dependence was 67%, the weighted mean environmental influence was 2% and the weighted mean individual-specific influence was 31%. Most studies have found small or no influences of shared environmental factors, which is in line with the results of the present study.

A limitation of the present study is that the data of the incomplete twin pairs were excluded from the analyses. It is possible that selection bias plays a role; those individuals who are most nicotine dependent may have refused to participate. A recent study comparing smoking behavior in complete and incomplete twins showed a somewhat higher percentage current smokers and ever smokers in the incomplete twins compared to the complete twin pairs but these differences were small and not significant (Vink *et al.*, in press). It is therefore unlikely that the exclusion of incomplete twin pairs significantly influenced the results.

While the CM model provided the best fit to the data, the heritability estimates did not differ much between the three models. When fitting an AE-model without sex-differences the estimates for a^2 was .75 for the CM model, while these estimates were .71 for the ILM (not-smokers are not taken into account when analyzing nicotine dependence), and .80 for the SLM (not-smokers are seen as scoring low on nicotine dependence). The existence of two overlapping dimensions is supported by a recent

linkage study in another, partly overlapping, sample. On chromosomes 6 and 14, LOD-scores of 3.0 and 1.7 respectively were found for smoking initiation. We obtained a LOD-score of 2.0 on chromosome 3 for quantity (number of cigarettes per day). Interestingly, an overlapping peak on chromosome 10 was found for both smoking initiation (LOD-score 1.9) and quantity (LOD-score 2.3), (Vink *et al.*, 2004-a). Those results confirm the findings in this paper; smoking initiation and nicotine dependence are two dimensions which are not independent of each other. Further research is needed to localize and identify the specific genes involved in both the smoking initiation dimension and the nicotine dependence dimension.

Chapter 7

Gene finding strategies

Jacqueline M. Vink and Dorret I. Boomsma (2002). "Gene finding strategies." Biological Psychology ***61****: 53-71.*

Abstract
Both linkage and association methods have been used to localize and identify genes related to behaviour and other complex traits. The *linkage* approach (parametric or non-parametric) can be used for whole genome screens to localize genes of unknown function. The parametric linkage approach is very effective for locating single-gene disorders and is usually based on large family pedigrees. The non-parametric method is useful to detect quantitative trait loci (QTLs) for complex traits and was originally developed for sib pair analyses. Genetic *association* studies are most often used to test the association of alleles at a candidate gene with a disease or with levels of a quantitative trait. Allelic association between a trait and a marker can be studied in a case-control design, but because of possible problems due to population stratification, within-family designs have been proposed as the optimal test for association.

Introduction

Behavioural genetic studies (including twin studies) have shown that genetic influences often contribute to individual differences in behaviour. Behavioural traits are complex, reflecting the aggregate effect of multiple, possibly interacting genetic and environmental determinants. Molecular genetic methods have been applied to complex and quantitative traits trying to identify genes responsible for the moderate to high heritabilities seen for behavioural traits (e.g. Gayan and Olson 1999; Gershon 2000; Faraone and Doyle 2001), but with the availability of relatively cheap and easy DNA marker typing, many more molecular genetics studies of behavioural traits can be expected in the next few years.In this respect, the completion of the human genome sequence will be valuable in locating and identifying genes involved in human behaviour (International Human Genome Sequencing Consortium 2001; International SNP Map Working Group 2001; Peltonen and McKusick 2001; Venter *et al.* 2001). This article reviews strategies for gene finding in humans, especially linkage and association methods. The gene finding strategies that we will discuss have been mostly applied to disease phenotypes. They are, however, increasingly applied to behaviour in a broader sense. We will illustrate gene finding for behavioural traits with examples of phenotypes taken from various research fields such as addiction and personality. and psychophysiological traits considered to be risk markers or risk factors for disease. We will discuss EEG power as one of the examples, and other papers in the special issue of Biological Psychology will provide a number of further examples (Busjahn *et al.*, 2002; Porjesz *et al.*, 2002; Snieder *et al.*, 2002). Variations and extensions of linkage and association methods are summarized and combined linkage and association tests are introduced as a tool for testing for genuine associations, as well as for fine mapping of broad linkage regions.

Linkage

Genes contribute to variation in both normal behaviour and behavioural disorders (Sullivan and Kendler 1999; Plomin and Crabbe 2000; Plomin *et al.* 2000; Bouchard and Loehlin 2001). Some disorders have a simple Mendelian mode of transmission in which a specific mutation confers the certainty of developing the disorder, in other words a single gene is responsible for the disorder. Many single gene diseases and disorders are listed in full in the "Mendelian Inheritance in Man" (McKusick 1998) and its freely available online version (www.ncbi.nlm.nih.gov/omim; updated every day). A general strategy to find genes for Mendelian traits is called classical linkage and is based on Fisher's theory of likelihood inference (Fisher 1918). It is referred to as being parametric or model-based because an explicit genetic model for the disease or trait locus has to be provided. Classical linkage analysis models the distance between a DNA marker locus and a putative disease locus in small numbers of large multigenerational families (pedigrees) consisting of both affected and unaffected family members. It is the method of choice for the genetic mapping of single-gene diseases, especially when these diseases are rare. Classical linkage requires that a model for the disease or trait locus is specified a priori, in terms of allelic frequencies, penetrance and mode of action (recessive or dominant). Complete penetrance implies

that all individuals with a high-risk genotype (genotype dd in the case of a recessive disorder and genotypes Dd and DD in the case of a dominant disorder) will develop the disorder. If there are individuals with a high-risk genotype who do not develop the disease, then the penetrance of the genotype is said to be incomplete. Individuals without a high-risk genotype who develop a disorder that is phenotypically indistinguishable from the genetic form, are called phenocopies (Sham 1998).
In linkage analysis a number of DNA markers of known location, evenly dispersed throughout the entire genome, are measured in individuals from multiple generations. DNA markers can be mutations in a single base pair (Single Nucleotide Polymorphisms, SNPs) or a variable number of repeats of two or more base pairs (microsatellites), as described in more detail by (Slagboom and Meulenbelt, 2002). They need not to be part of a functional gene – they are just landmarks of known location in the genome. For each DNA marker, evidence for linkage is derived using statistical procedures that trace the co-segregation of the trait (and thus in many instances the gene) and a specific variant of the DNA marker along familial lineages in extended pedigrees. The genetic distance between a marker locus of known position and a disease/trait locus (of unknown position) is estimated by observing the segregation of the marker locus in a pedigree together with the disease status/trait (figure 7.1). The second law of Mendel states that the inheritance of one gene is not affected by the inheritance of another gene (law of independent assortment). This law applies if two loci are on different chromosomes or are far apart on the same chromosome, because recombination between the loci will prevent alleles from being transmitted together. The closer two loci are on the same chromosome, the less likely crossovers during meiosis will be and the fewer recombinants will be observed in the offspring. Starting with the known position of a marker locus, it can thus be tested whether another locus is genetically close (linked) by counting the number of recombinations that occurred between both loci in a given number of meioses. The probability that two alleles at different loci on the same chromosome are derived from different parental chromosomes (i.e. recombinant) is called the recombination fraction. The recombination fraction ranges from $\theta=0$ (tight linkage) to $\theta=0.5$ (no linkage). If the loci are tightly linked, alleles from both loci are always inherited together in a pedigree. The recombination fraction can be taken as a measure of the genetic distance, or map distance, between gene loci. The unit of measurement is 1 map unit or 1 centimorgan (cM), corresponding approximately to a recombination fraction of 1%.
In parametric linkage analysis, it is standard practice to summarize the results of a linkage analysis in the form of a *LOD score* function (Morton 1955). LOD score stands for the logarithm of the odds that the locus is linked to the trait and indicates the strength of the linkage (figure 7.2). LOD scores are expressed according to the following equation:

$$\text{LOD score} = {}^{10}\log \frac{\text{Likelihood of the observed genotypes given } \theta \text{ is less than 0.5 (linkage)}}{\text{Likelihood of the observed genotypes given } \theta = 0.5 \text{ (no linkage)}}$$

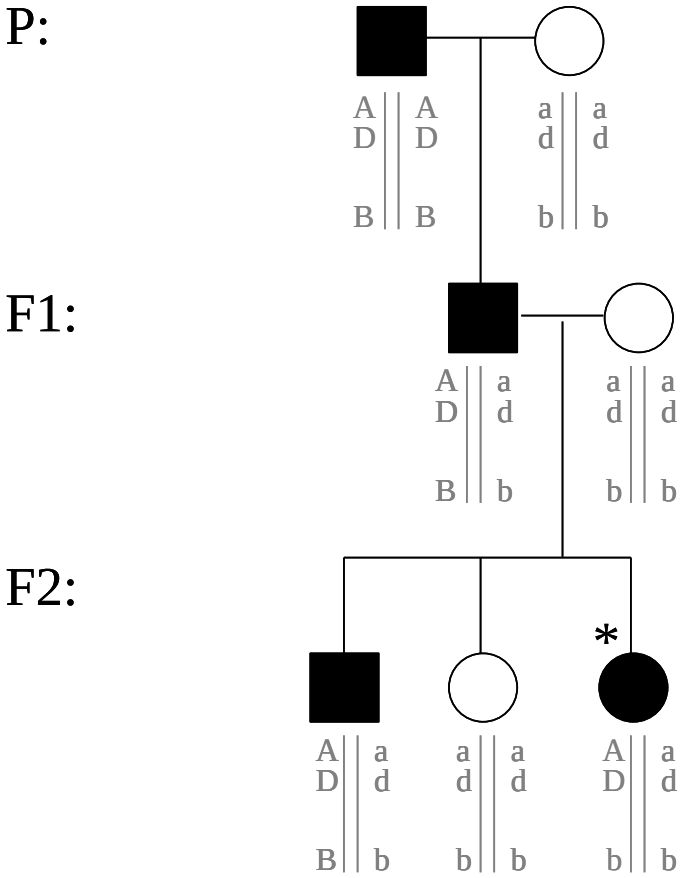


Figure 7.1 linkage and recombination. Artificial family with assumed disease locus. P = parents, F1 = first generation offspring, F2= second generation offspring, square = male, circle = female. The disease locus carries a variation which is reflected in *D* or *d*; when having allele *D*, the individual is affected and having allele *d,* the individual is unaffected. The affected individuals are shown in black and the unaffected individuals in white. There are two markers; marker 1 with polymorphism *A* or *a* and marker 2 with polymorphism *B* or *b*. Both parents are doubly homozygous: the father is homozygous for the *A* and *B* marker (haplotype AB), and the mother is homozygous for the *a* and *b* variant of the marker (haplotype ab). The son in the first generation has received marker *A* and *B* from the father and *a* and *b* from the mother (who newly enters the pedigree at F1). In the F2 generation recombination has occurred (individual is marked with an *), this individual carries haplotype Ab. The other two individuals are non-recombinant. The recombination fraction is the number of recombinations divided by the total number of meioses is, so when recombination occurs in 1 of the 100 meioses, the recombination fraction is 1%. In this example, full linkage will be found between marker A and the disease/trait (they always co-segregated together), but less so between marker B and the disease/trait (they co-segregate in only 2 of 3 meisoses).

Evidence for linkage is said to be present when the maximal LOD-score exceeds a pre-defined threshold, which depends on the size of the genome and the number of markers (Lander and Kruglyak 1995). The LOD-score is a function of the unknown recombination fraction θ. It is customary to plot LOD score against different recombination fractions in order to obtain an impression of the relative support for different values of the recombination fraction and thus the distance between the marker and the disease locus. The chromosomal region surrounding a marker with a significantly high LOD-score under the optimal recombination fraction will be selected for fine-mapping, which is essentially a repetition of the same procedure but now with many additional markers concentrated in the area of interest on a single chromosome. If the region containing the putative gene is sufficiently small, the DNA in the entire region is sequenced in full to find genetic variants (polymorphisms). The next step could be an association study (described below). The entire process from significant LOD scores to the actual allelic variants is usually summarized as *'positional cloning'*. The circa 1500 disease genes now listed in the Online Mendelian Inheritance in Man catalogue have largely been detected by this process.

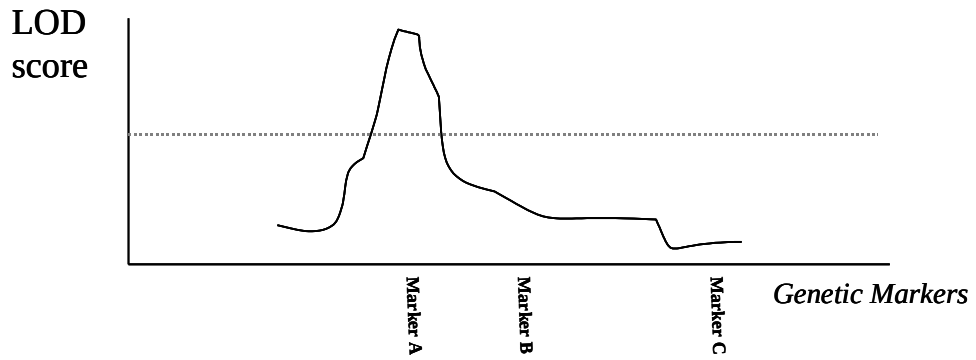


Figure 7.2 LOD-score. A significant LOD-score will be found if the locus of a marker (in this example marker A) is linked to the trait. The higher the LOD-score the tighter the linkage. The chromosomal region surrounding a marker with a significantly high LOD-score will be selected for fine-mapping. Because the disease locus significantly co-segregates with marker A, the position of the disease locus is probably (relatively) close to the locus of marker A

The localization of a locus for the human low-voltage EEG on chromosome 20q is an early example of a classical linkage approach. The inter individual variability of the human EEG is largely determined genetically (Beijsterveld and Boomsma 1994, Vogel and Motulsky 1997, (Beijsterveldt Van and Baal Van, 2001)). Some EEG variants were shown to follow a simple mode of inheritance. In the case of the low-voltage EEG the familial transmission pattern was found to follow an autosomal dominant mode of inheritance. Steinlein et al. (1992) studied a total of 22 blood and serological markers (as a proxy for the underlying polymorphisms) and 73 DNA markers (restriction fragment length polymorphisms =RFLPs) in 17 families with 191 individuals. The markers were distributed over all autosomal chromosomes. The frequency of the low-

voltage EEG allele was taken to be 0.02. An autosomal dominant mode of inheritance with full penetrance was assumed. Linkage analysis carried out for all families pooled together yielded no significant evidence for linkage. This "null" finding provided a nice example of a nasty complexity that may arise in linkage analyses when variations in recombination fraction occur because of the existence of multiple disease loci. A marker that is close to a particular disease locus will demonstrate linkage in families where the disease is caused by alleles at that locus. In other families, in which the disease is caused by alleles at other loci, the marker will show no linkage with the disease. This heterogeneity is known as locus heterogeneity. The results of the Steinlein study provided evidence for locus heterogeneity with respect to the low-voltage EEG variant. One of the markers, the CMM6 (D20S19), localized on the distal part of chromosome 20q showed linkage in some of the families (maximum LOD score 3.13 and recombination fraction 0) and exclusion of linkage in the other families. In short, two types of families were found: with and without linkage to chromosome 20q. Within the first type of family the autosomal dominant inherited low-voltage EEG is determined by a gene located close to the highly polymorphic marker CMM6 on chromosome 20q. In the second type of family this phenotype is caused by another gene, or genes, located elsewhere (Anokhin et al. 1992; Steinlein et al. 1992).

Non-parametric linkage

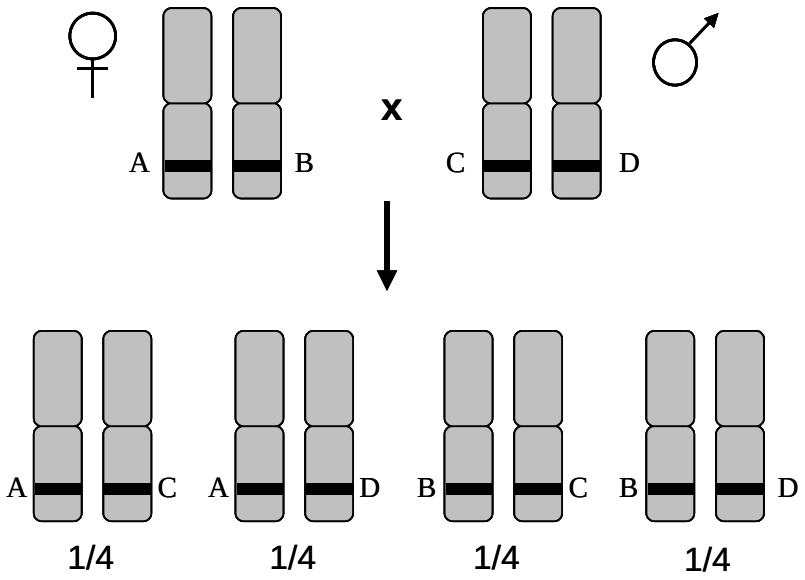
Most complex traits are multifactorial, i.e. they are influenced by a number of different genes, environmental factors, their possible interactions, and possibly a third source of variation that consist of nonlinear epigenetic processes (Molenaar et al. 1993). Traits that are influenced by the developmental interplay of many genes and environmental factors are usually quantitative traits, and each of the genes that influence such quantitative traits is called a polygene. The chromosomal region (or locus) where such a polygene can be found is called a quantitative trait locus (QTL). Typically, the word 'quantitative' is used when 'continuous' is meant, and variation in the phenotype shows a normal distribution. However, for some quantitative traits the scale of measurement can also be discrete. In the case of a binary disease phenotype (affected / unaffected) the penetrance, or probability of being affected, is often transformed to a probit (or logit), giving rise to what is called the 'liability' to disease. This liability can be thought of as the underlying vulnerability to the disease and is treated as a continuous phenotype (Falconer and Mackay 1996; Elston 2000).

To detect QTLs, non-parametric or model free linkage analysis uses a similar linkage concept as described above, but unlike parametric linkage, no explicit model of the disease is required for this type of genomic search. Non-parametric methods were originally developed for sibling pairs but have been extended to general pedigrees. In this kind of analysis, several hundreds of DNA markers are obtained from siblings and (optimally) their parents and allele sharing between siblings (or other relatives) is investigated. There are two definitions of allele sharing, identity-by-state (IBS) and identity-by-descent (IBD). Two alleles of the same form (i.e. having the same DNA sequence) are said to be IBS. If, in addition to being IBS, two alleles are descended

from the same ancestral allele, then they are said to be IBD. Full siblings both receive an allele from the father and an allele from the mother. Let the variable D_f be 1 when both siblings have received the same paternal allele, and 0 otherwise. Similarly, let the variable D_m be 1 if the two siblings have received the same maternal allele, and 0 otherwise. The total IBD value of the sibling pair, D, is defined as the sum of D_f and D_m and therefore can be 0,1,2 with the probabilities ¼, ½, ¼ respectively (figure 7.3). Linkage of a marker to a QTL implies that the differences in the trait between the relative pairs will be smaller if they share the same variant of the marker, obtained from the same ancestor (IBD), (Haseman and Elston 1972).
The original sib pair method proposed by Penrose, (1953) was based on the idea that linkage is supported if sibling pairs with two affected or two unaffected siblings are significantly more alike in terms of allele sharing at a marker locus compared with sibling pairs with just one affected member. This sib pair test was refined to give rise to the currently popular affected sib pair (ASP) method. Attention was focused exclusively on sibling pairs in which both members are affected, since such pairs are often more informative than unaffected sibling pairs, or sibling pairs with one affected and one unaffected member (Sham 1998).
The ability of the affected sib pair method to detect a disease susceptibility locus depends on the contribution the locus makes to family resemblance, which is often measured in terms of the increased risk to relatives of an affected proband as compared to the population prevalence (Risch 1990). For affected sib pair studies, this can be measured by the sibling risk ratio λ_s of the risk to a sib of an affected proband versus the population prevalence. This λ_s is an overall risk ratio that summarizes the collective effect of all the disease loci plus any other non-genetic familial resemblance; higher λ_s indicates stronger familial effects (Lathrop and Weeks 2000).
An example of non-parametric analyses of a dichotomous trait in affected sib pairs is the study of Straub et al. (1999). This is the first published report of a complete genome scan designed to detect genes that influence the risk of nicotine dependence. A genome scan using 451 DNA markers was conducted to identify chromosomal regions linked to nicotine dependence in a sample from Christchurch, New Zealand (201 affected sib pairs from 130 families). Non-parametric linkage scores (z_{all}) were obtained under the assumption of locus heterogeneity. The z_{all} statistic is a 'similarity statistic' for affected relatives, and is defined as the average of the possibilities that relatives are IBS. The best result was with marker D2S1326 on chromosome 2. Straub et al. also found a number of large chromosomal regions where many consecutive markers yielded small but positive z_{all} scores. Selected regions of chromosomes were further investigated by additional genotyping of the Christchurch sample and an independent sample from Richmond, Virginia (190 affected sib pairs from 91 families). For example, the analyses of the DNA markers on chromosome 2 in the Christchurch sample showed six positive z_{all} scores in a region over 19 cM. The best result marker D2S1326 (z_{all}=2.65, p=0.0011) was roughly in the middle of this region. In the Richmond sample there is a cluster of seven markers on chromosome 2 which all have positive z_{all} scores and the best result for the Richmond sample was marker D2S442 (z_{all} 1.05) which is located about 2 cM of the D2S1326 marker. Straub et al.

(1999) found regions on chromosome 2, 4, 10, 16, 17 and 18 that merit further study. However, they also concluded that when simply judged against the usual standards of linkage significance, none of the individual regions yielded strong evidence. It is probable that the size of the available sample provided only limited power to detect linkage. This illustrates that it is difficult to detect genes of small effect, or genes that are influencing risk in only a small proportion of the families (Straub et al. 1999).

(a)



(b)

Sibling 1:

Sibling 2:	AC	AD	BC	BD
AC	2	1	1	0
AD	1	2	0	1
BC	1	0	2	1
BD	0	1	1	2

Figure 7.3 (a): Graph showing the possible allele combination for children from a mother with allele A and B and a father with allele C and D. The chance for each combination (AC, AD, BC and BD) in the offspring is ¼. (b): Identical by descent. The probability that two siblings share 2 parental alleles (IBD=2) is 4/16 = ¼ . The probability that they do not share parental alleles is also 4/16 = ¼, but the probability that they share 1 parental allele is 8/16 = ½.

Statistical methods and power to detect linkage

Different methods can be used to calculate linkage. In general, the methods can be divided in regression analyses and maximum likelihood methods (variance component models). Haseman and Elston introduced in 1972 an elegant regression method to test for linkage for quantitative traits. Evidence that a marker is linked to the trait is obtained by regressing the squared trait difference between phenotypes of siblings on the proportion of marker alleles shared identical-by-descent (π) (Haseman and Elston 1972). A major drawback of this method is that it requires large numbers of sibs to detect significant evidence for linkage. The variance-component models were originally developed for the partitioning of phenotypic variation into genetic and environmental components from correlational data from pairs of relatives (usually twins), but have now been extended for QTL analysis (figure 7.4). With this approach not only the differences within sibships but also the differences between sibships can be modelled as a function of the QTL. Moreover, the full IBD distribution can be used in the estimation procedure and the model generalizes quite easily to larger sibships and to mulivariate phenotypes (Martin, Boomsma and Machin, 1997; Boomsma and Dolan 2000; Neale 2000, Dolan et al 1999, Fulker and Cherny 1996). Under assumptions of (multivariate) normality, the parameters depicted in Figure 4 can be estimated with maximum-likelihood methods, which are available in standard software packages such as Lisrel or Mx (Joreskog and Sorbom 1989, Neale *et al.* 1999).

A major problem of linkage analysis of complex traits with multiple contributing loci is the lack of statistical power. The main challenge thus is to develop linkage methods that have the highest statistical power to detect QTLs of small effect. The extensions of variance components methods listed above, such as multivariate approaches and testing larger sibships all lead to an increase in power. Visscher and Hopper (2001) compared the statistical power of linear regression and maximum likelihood methods to map QTLs for univariate traits from unselected sib pair data, and determined which methods are superior under which set of population parameters. Their derivations of statistical power for regression and maximum likelihood methods provide a simple way to compare alternative methods. If there are many covariates to be adjusted, a full maximum likelihood approach is recommended because regression methods have the drawback that they cannot perform multivariate analyses (Visscher and Hopper 2001).

An important factor for power is the magnitude of the heritability or the familial risk ratio. Risch (1990) describes the power to detect linkage as a function of the risk ratio λ_s (familial risk ratio) by using affected sibling pairs and assuming a fully informative marker and a recombination fraction of 0 between marker and QTL. For a sample of 200 affected sibling pairs the power to detect linkage is 0.4 when $\lambda_{s=}2$, while the power to detect linkage is > 0.9 when $\lambda_s= 4$ (Risch 1990a). In the same paper, Risch also shows that many of the power estimates are too optimistic if some of their common assumptions are violated. For most estimates it is assumed that the marker and the disease susceptibility locus are completely linked, and that markers are completely informative. Risch demonstrated the potentially damaging effect on the power to detect linkage when the distance (recombination fraction) between marker and QTL is large. For a sample of 300 affected sibling pairs and a λ_s of 3, the power to detect

linkage is 0.85 when the recombination fraction is 0, while the power is 0.4 and 0.15 when the recombination fraction is 0.05 and 0.10, respectively. Those problems, however, can be overcome to some extent by the use of multiple linked markers in *multipoint analyses*. The use of multipoint analyses requires prior knowledge of the relative positions of several marker loci in the chromosomal region of interest. The positions of the markers can then be fixed, while, in an iterative procedure, the putative position of the trait locus is varied from the one end of the region through the other end of the region. A LOD score is then calculated for each of the tested positions.

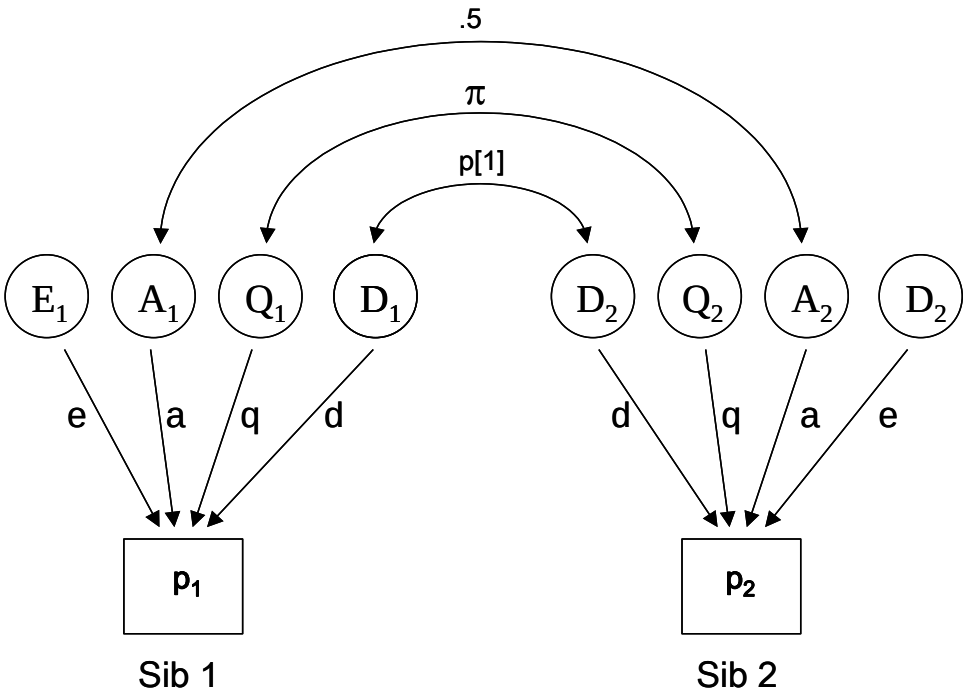


Figure 7.4. Path-diagram with QTL effect. Path-diagram with observed phenotypes (p1 and p2) in sib 1 and sib 2 represented by squares, and latent variables E (individual-specific environment), A (additive genetic background, Q (additive QTL effect) and D (non-additive QTL effect) represented by circles. The path coefficients of each latent variable on the observed phenotypes are estimated (a, q, d, e). The correlation between additive QTL effects equals the proportion of alleles shared IBD (π) and the correlation between non-additive QTL effects is p[1]: the probability that siblings share all alleles identical-by-descent. The significance of the QTL effect is tested by constraining the path from QTL to phenotype at zero and test if this leads to decrease in the goodness of fit statistic.

A genome-wide scan that provides a good illustration of non-parametric linkage analyses for a quantitative trait is a multipoint analysis for personality traits and a set of genetic markers. Cloninger *et al.* (1998) measured personality with the Tridimensional Personality Questionnaire and tested each of the four personality dimensions for linkage across all chromosomes. Genotyping was carried out on 987 individuals in 105 pedigrees, including 758 sibling pairs in 177 nuclear families. Multipoint variance components analysis was performed to estimate the genetic variance attributable to the QTL(s) linked to a genetic marker. Strong evidence was found that a genetic locus on 8p21-23 (marker D8S1106) accounted for most of the additive genetic variance in harm avoidance (anxiety-proneness vs risk taking), (LOD score=3.2, p=0.0006). Cloninger *et al.* (1998) also tested for epistatic interactions (possibility of interactions between alleles at different loci) and found strong evidence for epistasis between the locus on 8p and others on chromosome 18p, 20p and 21q (lod 5.1, p=0.000007).

Allelic association

Linkage is usually genome-wide, while association studies are limited to candidate genes or candidate regions. Furthermore, linkage analyses must be carried out in pedigrees (families and sibling pairs), while association can be performed at the population level. With allelic association studies an association between a disease and a specific allele can be detected in groups of unrelated cases (e.g. patients) and controls (e.g. healthy subjects). Association can be found either with functional genetic variants that have biological consequences related to disease, or with other variants that are in linkage disequilibrium with these variants. Linkage disequilibrium occurs when a marker allele (i.e. a SNP) and the QTL are so close on the chromosome that they co-segregate in the population over many generations of meiotic recombination. Association studies are similar in design to classic case-control studies in epidemiology. DNA is collected from all participants and the trait is compared across the various allelic variants of the DNA marker. Vice versa, frequencies of the various allelic variants may be compared in subjects with particular phenotypes, to detect an association between a particular allele and the occurrence of the phenotype. The advantage over linkage analysis is that association studies can detect the region of a QTL that has only very small effects on the trait (Risch and Merikangas 1997). Provided that either the selection of cases does not introduce population stratification or that the analyses properly control for such stratification, association studies provide a good complement to the linkage strategy. Screening the entire genome with association, however, requires huge numbers of markers (linkage requires only a few hundred markers) and is not currently feasible. Allelic association, therefore, has been used primarily with candidate genes.

Candidate genes

The ideal candidate gene has been shown to be functional: it influences the concentration of the (iso)form of a protein, its functionality or efficiency, or perhaps most importantly, its responsiveness to environmental factors triggering the

expression of the gene. The problem with a candidate gene approach for most complex traits is the potentially huge proportion of genes which can serve as candidates. Several strategies are possible to select an optimal set of candidate genes. First, genes that are part of physiological systems known to influence the trait can be tested as candidates. Secondly, genes or chromosomal regions that are known to influence the trait in animals can be tested as candidate genes (or regions) in humans. Candidate genes for smoking and nicotine dependence, for example, could be genes that are involved in dopamine activity (because the dopamine reward pathway plays a critical role in substance use) but also genes that are involved in nicotine metabolism and genes involved in personality (sensation seeking, neuroticism, depression). Using such candidates, several associations between the dopamine receptor genes and substance use have been reported. A significant effect was found for the dopamine transporter gene; individuals with a particular variant of this gene (SLC6A3-9) were significantly less likely to be smokers, especially if they also had a certain variant of the D2 dopamine receptor (DRD2-A2), (Lerman *et al.* 1999). The long form of the D4 receptor gene is more frequent in individuals with high quantity/frequency of drug use compared to controls (Vandenbergh *et al.* 2000). The results of a population-based association study of substance abuse and a microsatellite at the dopamine D5 receptor locus (DRD5) in a sample of European-American males and females found that the DRD5 locus is involved in the variation of substance abuse liability (Vanyukov *et al.* 1998). Duaux *et al.* (2000) reviewed molecular genetic studies in drug abuse; results of several association studies reported positive association between drug disorder and polymorphisms of several dopaminergic receptor genes (DRD1, DRD2, DRD3, DRD4). A problem with the candidate gene approach is that by looking for candidates among the pathways that we already know, we may still overlook the essential genes, because of our ignorance of other biological systems involved.

Within-family association studies

Another problem of association studies is the danger that a spurious association is found between the trait of interest and any locus that differs in allele frequency between subpopulations. This situation is illustrated by the 'chopstick gene' story described by Hamer and Sirota (2000). They describe a hypothetical study in which DNA markers were assessed in students who often used chopsticks and students who did not. One of the DNA markers showed a huge correlation to chopstick use. Of course this gene had nothing to do with chopstick use, but just happened to have different allele frequencies in Asians and Caucasians, who differ in chopstick use for purely cultural rather than biological reasons. Witte *et al.* (1999) have evaluated the asymptotic bias in relative risk estimates resulting from using population controls when there is confounding due to population stratification. The direction of the bias is what one would expect from the usual principles of confounding in epidemiology: if the allele frequencies and baseline risks are both higher in a population, the bias is positive; if different, the bias is negative. Case-control studies of genetic associations thus can lead to false positive as well as to false negative results.

To prevent significant findings due to population stratification, within-family association designs have been developed, because family members are usually well matched on a number of traits that could give rise to stratification effects (Spielman *et al.* 1993). Most available family based tests for association were initially developed for binary traits, such as the Transmission Disequilibrium association Test (TDT) and the Haplotype Relative Risk Test (HRR). Those tests usually collect DNA samples in affected individuals and their biological parents. Affected individuals must have received one or two susceptibility alleles from their parents. These alleles transmitted from parents to the affected individual can be viewed as a group of "case" alleles. The non-transmitted alleles from the parents can be considered as "control" alleles. In other words, those tests only needs affected individuals and their parents, no control group is required (Terwilliger and Ott, 1992).
In a different approach the effects of genotypes on phenotypic means are partitioned into between-family and within-family components, by comparing the association of alleles and trait values across siblings from different families to the association of alleles and trait values across siblings within the same family. Sibling pairs are by definition ethnically and racially homogeneous and any difference in trait scores between siblings of different genotypes at a candidate marker, therefore, reflect true genetic association. By partitioning the mean effect of a locus into a between and a within-sibship component, spurious associations due to population stratification and admixture are controlled for (Abecasis *et al.* 2000,Fulker *et al.* 1999). An early example is the study of Lesch *et al* (1996) that demonstrated that the observed associations between a polymorphism in the serotonin transporter gene (5-HTTLPR) and personality are the result of genetic transmission rather than population stratification (Lesch *et al.* 1996). The study population included 459 siblings from 210 independent families, of which 78 sibling pairs from 61 independent families had discordant 5-HTTLPR genotypes (one or two copies of the short form versus homozygous for the long form). The difference in personality scores between siblings with the long form and siblings with the short form of the 5-HTTLPR genotype was statistically significant. Most importantly, highly comparable results were obtained by population-based or across-pedigrees analyses.
The literature on family-based methods rapidly grows. Some methods extend the original tests to accommodate multi allelic markers, variable pedigree constellations, multiple loci, and quantitative traits. Family-based association studies are comprehensively described by Zhao (2000) and Schulze and McMahon (2002).

Fine mapping
To detect which candidate gene in a linkage region is the causal gene, Fulker *et al.* (1999) introduced a systematic approach for the simultaneous analysis of both association and linkage for quantitative traits in sib pairs. If significant linkage is detected while also modelling association, the putative locus modelled in the association is not the functional gene. If linkage evidence vanishes when simultaneously modelling association, the marker may be the QTL itself (or in very strong linkage disequilibrium with it). A simultaneous test for linkage and association

can be carried out using multipoint IBD information to model sib pair covariances (test of linkage) and a decomposition of the mean phenotype into allelic effects (test of association) between and within families. The novel joint analyses of both linkage and association is made possible by a statistical approach unified by the use of maximum likelihood and of a common biometrical model for the simultaneous analysis of means and covariance matrices. Cardon and Abecasis (2000) evaluated the behaviour of the association and linkage parameters in the model of Fulker *et al.* which may facilitate fine-mapping studies of complex traits that aim to localize QTLs by assessment of association with many markers in a candidate region of interest. An extension to the method of Fulker *et al.* is proposed by Posthuma *et al.* (2002), and allows the use of (variable) sibship sizes greater than two, the estimation of additive and dominance association effects, and the use of multiple alleles. These extensions can be implemented without parental genotypes but are most powerful when these genotypes are available.

Animal models in genetics

The oldest technique in behavioural genetics with animals is that of artificial selection. Mice (or other animals) are selected on their scores on tests for open field activity, behaviour in a maze or their behavioural response for e.g. sensitivity, tolerance, dependence and preference for alcohol or nicotine. Extreme scoring animals are mated and selection lines are created for high scoring animals and for low scoring animals. If such selection is possible, this proves that the trait is influenced by genetic factors (Crabbe *et al.* 1999). High and low scoring strains will differ at loci that influence the trait on which selection was based. In contrast inbred strains are created by repeated matings between brothers and sisters. Within an inbred strain, all same-sex animals are essentially monozygotic twins and have two identical copies of a single allele at each locus. Crabbe (2002) describes a study in which different inbred strains of mice were offered a choice between a bottle filled with tap water and one containing alcohol. The differences among strains in preference for alcohol far exceeded the within-strain differences; suggesting that these preferences have a genetic basis. Many different inbred strains of mice are available for genetic mapping experiments. Initial identification of QTLs involves examining many individual animals and correlating the possession of specific alleles at genetic markers with the degree of quantitative trait expressed (Crabbe *et al.* 1999). QTLs in mice have been found for several drug-sensitivity genes. For example, QTL analyses revealed that several genetic markers in inbred mice were associated with ethanol consumption levels, including markers for the D2 dopamine receptor (Philips *et al.* 1994; Buck *et al.* 2000). Another opportunity to evaluate the roles of gene products in animals is the genetic engineering approach. In this approach, mice that are made to lack ('knock-out'), under-express or over-express specific genes are studied. Several studies in knock-out mice have demonstrated the effects of specific genes on behavioural responses to drugs. For example, knocking out a serotonin receptor gene in mice leads to increased alcohol consumption and to increased vulnerability to cocaine (Rocha *et*

al. 1998). Rubinstein *et al.* (1997) found supersensitivity to alcohol, cocaine, and methamphetamine in mice whose dopamine D4 receptor was knocked out.
A new development is the use of expression arrays or so called 'gene chips'. Thousands of individual gene sequences can be bound to tiny chips (glass plates). When a sample of DNA or RNA is applied, those genes actively express in the sample, bind to their embedded ligand and the resulting interaction is visualized. At least 6000 mouse brain DNA probes are available on chips, and can be used to study gene expression under different conditions of e.g. environmental exposure, thus identifying genes (Crabbe 2002). For example, Freeman *et al.* (2001) used expression arrays to identify cocaine-regulated genes by comparing the gene-expression in rats treated with cocaine versus control rats. The findings suggest altered expression of genes with a number of different functions in the rat hippocampus after cocaine administration (a.o. induction of potassium channel 1.1, protein tyrosine kinase 2), (Freeman *et al.* 2001). Because the homologous region of murine genes in the human genome is often known, the genes/regions identified in mouse studies (or other animal studies) can be regarded as plausible candidate genes/regions in human genetic studies (Picciotto *et al.* 2000).

Discussion
The major strength of linkage is that it is systematic in the sense that a few hundred DNA markers can be used to scan the entire genome. In contrast, allelic association with a quantitative trait can only be detected if a DNA marker is the QTL itself or very close to it, so tens of thousands of DNA markers would need to be genotyped to scan the entire genome On the other hand, association studies can detect QTLs with only small effects on the trait, whereas linkage may not. Linkage and association analysis are therefore fully complementary approaches; association studies can be used as an approach to isolate a susceptibility gene in a region that has first been identified by linkage. Evidence for a substantial genetic contribution, in terms of the sibling recurrence ration (λ_s) for all-or-none traits and heritability (h^2) or sibling correlation (r) for the quantitative traits, is a prerequisite for embarking on gene mapping studies. Nonetheless, even for confirmed heritable complex diseases, linkage and association approaches have met with limited success so far.
Altmuller *et al* (2001) reviewed 101 whole genome scans of complex human disease, which were found by a systematic Medline search. These linkage studies were compared with regard to design, method and relative 'success'. Most studies (66.3%) did not show significant linkage (using the criteria of Lander and Kruglyak, 1995) and the results of studies of the same disease were often inconsistent. Altmuller *et al.* concluded that no single study design consistently produces more-significant results. The only factors independently associated with increased study success were 1) an increase in the number of individuals studied and 2) studies of subjects drawn from only one ethnic group (Altmuller *et al.* 2001; Guo 2002). An efficient method to realize the power of gene detection in large samples, is to phenotype a large samples and to select a subgroup of the most informative families for genotyping. Selection and genotyping of extremely discordant and concordant sibling pairs can increase the

power to detect linkage without the need to genotype the entire sample from which the extreme pairs were drawn (Risch and Zhang 1995; Dolan *et al.* 1999; Abecasis *et al.* 2001). Several studies of anxiety and depression have described such methods for selection of extreme discordant and concordant sib pairs (Boomsma *et al.* 2000; Kirk *et al.* 2000; Martin *et al.* 2000). The limited success of linkage studies has led to the proposal that genetic association studies may offer a better alternative to find genes for complex traits in humans. A review of meta-analysis studies on genetic associations in human disease by Ioannidis *et al.* (2001), shows that this is currently hardly the case. Of the 36 traits considered (based on a total of 370 studies) only 8 traits/diseases showed statistically significant associations in the meta-analysis (a.o. ischaemic stroke/ECE, bladder cancer/NAT2). In 8 other traits, the first study reached statistical significance and subsequent research did not disagree with the results; however, in only 4 of the 8 traits formal statistical significance was found for the genetic association at the end of the meta-analysis. When an initial study suggests a stronger genetic effect than is found in subsequent studies, this can be caused by sampling bias (the most prominent findings represent an extreme sample and associations may be less extreme in new studies), by publication bias, by inflation of the size of a genetic effect (if based only on a single study with impressive results) and by a large statistically uncertainty in the first study (Ioannidis et al 2001; Vieland 2001). We can concur with Ioannidis and colleagues that association studies require cautious replication – and we believe that to apply to all gene findings either linkage or association methods.

We have focused in this review mostly on the methodology of linkage studies employing sib pairs designs, which can also be used as a tool in association studies by decomposing the association effect into a between and within-families component. Other designs, such as linkage and association studies in large pedigrees or isolated populations may be less feasible for the complex traits studied in the fields of human behaviour and psychophysiology. More distant relatives or isolated populations are useful to detect rare susceptibility genes, whereas closer relatives, such as siblings are required for studies of common diseases and traits.

Chapter 8

Linkage analyses of smoking initiation and quantity in Dutch sibling pairs

Jacqueline M. Vink, A. Leo Beem, Danielle Posthuma, Michael C Neale, Gonneke Willemsen, Kenneth S Kendler, P. Eline Slagboom, Dorret I. Boomsma. (2004) Linkage analyses of smoking initiation and quantity in Dutch sibling pairs. The Pharmacogenomics Journal 4; 274-282

Abstract
The heritability of smoking initiation (SI) and number of cigarettes smoked (NC) was determined in 3657 Dutch twin pairs. For SI a heritability of 36% was found and for NC of 51%. Both SI and NC were also significantly influenced by environmental factors shared by family members. The etiological factors that influence these traits partly overlap. Linkage analyses were performed on data of 536 DZ twins and siblings from 192 families, forming 592 sibling pairs. Results suggested QTLs on chromosome 6 (LOD = 3.05) and chromosome 14 (LOD = 1.66) for SI and on chromosome 3 (LOD = 1.98) for NC. Strikingly, on chromosome 10 a peak was found in the same region for both SI (LOD = 1.92) and for NC (LOD = 2.29) which may partly explain the overlapping etiological factors for SI and NC.

Introduction

Large scale population based twin and family studies have shown that genetic factors contribute to individual differences in smoking behavior (Heath and Madden,1995; Hopfer *et al.*, 2003; Sullivan and Kendler, 1999; Li *et al.*, 2003). Several different, possibly correlated, dimensions of smoking behavior can be distinguished: smoking initiation, number of cigarettes smoked per day and nicotine dependence (Mayhew *et al.*, 2000). Koopmans *et al.* (1999) investigated the heritability of smoking initiation (SI) and the number of cigarettes smoked per day (NC) in adolescent Dutch twins by considering a single liability model, an independent liability model and a combined model. The combined model best described the data and showed that 39% of the variance in SI and 86% of the variance in NC was explained by genetic influences. Kendler *et al.* (1999) found that liabilities to SI and nicotine dependence (ND) were substantially correlated but not identical and that heritable factors played an important role in both SI and in ND.

The next step after obtaining evidence for significant heritability is to identify chromosomal regions involved in smoking behavior, either by linkage or association approaches (Vink and Boomsma, 2002). Both human and animal studies have explored candidate genes for smoking behavior. Association studies point to dopamine receptor genes, dopamine transporter genes, cytochrome P450 and serotonergic genes (Walton *et al.*, 2001; Batra *et al.*, 2003). Association studies have relatively high statistical power, and can detect quantitative trait loci (QTLs) with only small effects. A possible disadvantage of the candidate gene approach is that the focus is on known pathways, which may lead us to overlook genes that are etiologically important, because of our ignorance of other biological systems involved. In contrast, linkage analysis will identify chromosomal regions that harbor known and unknown genes, although the statistical power to identify such regions is relatively low.

Linkage studies for smoking are, at present, sparse and those that were performed have used different definitions of smoking behavior. Using smoking data collected in the Collaborative Study on the Genetics of Alcoholism (COGA), the most promising linkage results were reported for chromosome 6, 9, and 14 using single point sibling pair analysis (Bergen *et al.,* 1999), and for chromosome 3, 4, 5, 6, 9, 15 and 17 when applying a multipoint variance component method (Duggirala *et al.,* 1999). Smoking in these analyses was defined as ever having smoked daily for a month or having smoked more than 100 cigarettes during one's lifetime (Bergen et al., 1999) and as having smoked more than zero cigarettes/day for at least a year (Duggiurala et al., 1999). Using the same dataset but focusing on heavy smoking (more than 20 cigarettes per day for at least 6 months), LOD scores greater than one were found on chromosomes 5, 9, 11 and 21 (Bierut *et al.*, 2004). Using data from two different populations, Straub *et al.*, 1999 examined linkage for nicotine dependence, defined as a score of 7 or over on the Fagerström Tolerance Questionnaire (Fagerstrom, 1978). A genome scan was performed in a subsample of genotyped individuals from Christchurch (New Zealand) and confirmation of the results was sought by genotyping additional Christchurch subjects and linkage in an independent sample from Richmond (USA). For six of the most positive regions found in the first genome scan, located on chromosomes 2, 4,

10, 16, 17 and 18, replication was found. Thus, in the different studies, peaks have been found on most chromosomes. These results may reflect differences in populations but may also reflect the fact that different genes are involved in different aspects of smoking behavior.
In this paper we simultaneously examine smoking initiation and quantity smoked using longitudinal data from twins and siblings. Quantity smoked is often used as proxy measure for nicotine dependence and both phenotypes are highly correlated. Phenotypic data were collected in a study on health related behavior of the Netherlands Twin Register (Boomsma *et al.*, 2002). First we fit a single liability, an independent liability and a combined model (Heath and Martin, 1993; Heath *et al.*, 2002) to phenotypic data for SI and NC. From the model that best describes the data, heritability estimates for SI and NC are obtained. Next, we report the results from a complete genome scan on SI (ever/never smoked) and NC in a subsample of dizygotic (DZ) twin and sibling pairs.

Methods

Subjects:

This study is part of an ongoing twin family study on health-related behavior in participants of the Netherlands Twin Register. Addresses of twin families were obtained from City Councils in 1991 and 1993. In later years, additional volunteer twin families also participated in the study. Surveys were mailed to twin families in 1991, 1993, 1995, 1997 and 2000 (Boomsma *et al.*, 2002). A sixth wave of data collection is in progress. Twin pairs were invited to participate in all waves, and parents were invited in 1991, 1993 and 1995. Siblings of twins were included in the assessments since 1995 and spouses since 2000. Each survey collected data on smoking, other lifestyle factors, health, personality and psychopathology.
Data on smoking behavior were available for 8.039 twins and 2.529 siblings. Marker data were available for families selected for a linkage study of anxious depression (Boomsma *et al.*, 2000). Selection of extreme sibling pairs for anxious depression (according to EDAC design) was based on a composite score that included data on depression, neuroticism and anxiety. Selection of families took place at two occasions. The first selection used data from the first 4 surveys, the second selection from the 2000 survey. In this paper we use genome scan data from families from the first selection (Boomsma *et al.*, 2002).
If at least two offspring formed an extremely concordant or discordant sibling pair for anxious depression, the entire family, including parents and any additional siblings, were asked for a DNA sample for genotyping (MZ twin pairs were treated as a single offspring). As some families consisted of more than two siblings, this selection procedure resulted in a (nonrandom) sample from the entire empirical distribution, not merely from its tails.

DNA collection and Genotyping

The selected subjects were asked to provide a buccal swab for DNA isolation (Meulenbelt *et al.*, 1995). Of the subjects selected for the QTL study (n=2.724),

around 72% (n=1962) returned a buccal swab and 917 subjects were genotyped over the entire genome. Selection of the first set of families for genotyping was based on family size (larger sibships) and on amount of DNA. Genotyping was conducted by the Marshfield Laboratory. For this scan the 10 cM spaced microsatellite screening set 10 (Yuan *et al.*, 1997) was used with few alternative markers. On the autosomes 379 markers were measured. Pedigrees were checked for Mendelian errors with the program Unknown (Schäffer, 1996) and pedigree relationships in the entire sample with the GRR program (Abecasis *et al.*, 2001). Mendelian errors were removed by assigning missing values to the marker scores if the errors appeared incidental. One subject with an excessive error rate, two subjects with uncertain identities and two families for which apparent problems could not be resolved were removed from the analysis. This left a total of 896 subjects (606 siblings and 290 parents) from 215 families. A subset of 212 families contained 2 or more offspring, in which both parents were genotyped in 121 families, one parent in 43 families and no parent in 48 families (two families contained 2 parents and 1 offspring and 1 family contained 1 parent and 1 offspring).
Likelihoods for recombinations were checked using the program Merlin (Abecasis *et al.*, 2002). Excessive recombinations were observed for 5 markers indicating potential problems. Those markers were not included in the final analyses: two markers on chromosome 1 (D1S468-AFM280we5 and D1S1627-ATA25E0); two markers (D11S1985-GGAA5C04 and D11S2006-GATA46A12) in a group of five very closely or identically mapped markers on chromosome 11; and one marker on chromosome 20 (8; D20S159-UT1307). For all other recombination problems the data were cleaned using Merlin's default procedure. As a result of cleaning, 57 genotypings in 46 families were set to missing; for two subjects two marker scores were set to missing. For the linkage analyses, sibling pairs were selected for whom more than 50% of the markers were typed successfully. In total, successful genotyping data were available for 536 offspring and 278 parents from 192 families from which 592 sibling pairs were formed. Marker distances were assigned from the Decode map if available. For markers not mapped by Decode, the original distance provided on the Marshfield website (Broman *et al.*, 1998) was transformed by linear interpolation from adjacent markers with known Decode map values (Kong *et al.*, 2002).

Phenotype:
Data on smoking behavior were collected in every survey (1991, 1993, 1995, 1997, 2000) and most subjects participated more than once. The phenotypes were constructed by taking the answers to all surveys into account. The surveys contained several questions on smoking: "Did you ever smoke a cigarette?", "Did you smoke during the last 12 months?" and "Did you smoke during the last 4 weeks?". The answer categories were: no, a few times to try, yes. Furthermore was asked "How many years did/do you smoke?" and another question was "How often do you smoke now?" with the answer categories: I have never smoked regularly, I have quit smoking, I smoke less than once a week, I smoke several times a week but not every day, I

smoke daily. Participants also reported the number of cigarettes they smoke per day or per week.
For the simultaneous model fitting of smoking initiation (SI) and 'maximum number of cigarettes per day' (NC) subjects were classified as never smokers (never smoked, or tried but never smoked regularly), 1-5 cigarettes per day, 6-10 cigarettes per day or more than 10 cigarettes per day (both smokers and ex-smokers).
In the linkage analyses of SI, subjects were classified as never smokers (never smoked, or tried but never smoked regularly) or ever smokers (including current smokers and ex-smokers). In the linkage analyses of NC, subjects (current smokers and ex-smokers) were classified as: never smoked regularly, < 1 cigarettes per day, 1-5 cigarettes per day, 6-10 cigarettes per day, 11-20 cigarettes per day, 21-30 cigarettes per day and more than 30 cigarettes per day.

Genetic model fitting:
Three different threshold models were fitted to the data: independent liability, single liability and a combined model (described in Koopmans *et al.*, 1999). The independent liability model assumes two independent liability dimensions for SI and NC. The single liability model postulates that the liability to smoking behavior is uni-dimensional and normally distributed with 4 categories (> 10 cigarettes/day, 6-10 cigarettes/day, 1-5 cigarettes/day, non-smoker). The independent liability model postulates two independent liability dimensions for initiation and quantity (> 10 cigarettes/day, 6-10 cigarettes/day, 1-5 cigarettes/day, non-smoker) that are each determined by completely separate genetic and environmental factors. The combined model includes features of both models. It consists of an initiation and a quantity dimension. Under the combined model there are two different routes to being a non-smoker: an individual can be a non-smoker due to genetic and/or environmental factors that influence the SI dimension or because the individual is low on the quantity dimension. The smoking behavior of the first twin was cross-classified with the smoking behavior of the second twin, resulting in 4x4 contingency tables for each zygosity group. Contingency tables were available for 595 monozygotic male (MZM), 476 dizygotic male (DZM), 1011 monozygotic female (MZF), 644 dizygotic female (DZF) and 931 dizygotic opposite sex twin pairs (DOS). Models were fitted to the contingency tables by maximum likelihood with Mx (Neale *et al.*, 1999).
Sources of variation that were considered in modeling the variation in liability to SI and NC were additive genetic variation (σ^2_a), shared environmental variation (σ^2_c) and unique environmental variation not shared by family members (σ^2_e). Sex-differences in variance components were tested by allowing the magnitude of the genetic and environmental effects to be different for males and females. For all models, different thresholds were estimated for males and females, allowing for sex differences in the prevalence of smoking.

Genotyping and IBD estimation:
If a sibling pair receives the same chromosomal segment from a parent in a certain region of the genome, the pair is said to share the parent's alleles in that region

identical by descent (IBD). Since offspring receive their alleles from two parents, a pair can share 0, 1 or 2 alleles IBD. IBD status is not always unambiguously known and has to be estimated using the specific allele pattern across chromosomes of two or more siblings and parents. The IBD status is usually estimated for a number of markers with (approximately) known location along the genome and is then used as the measure of genetic similarity. The estimate of the proportion of alleles shared identical by descent is referred to as $\hat{\pi}$, and is obtained as:

$$\hat{\pi}_{ijk} = 0 \text{ x } p_{(IBD=0)_{ijk}} + 0.5 \text{ x } p_{(IBD=1)_{ijk}} + 1 \text{ x } p_{(IBD=2)_{ijk}}$$

where $\hat{\pi}_{ijk}$ is the estimated proportion of alleles shared IBD between sib j and k for the i-th family, and $p_{(IBD=0)_{ijk}}$, $p_{(IBD=1)_{ijk}}$ and $p_{(IBD=2)_{ijk}}$ are the probabilities that sib j and k share 0, 1 or 2 alleles, respectively, conditional on the marker information. The probabilities of sharing zero, one or two alleles IBD at every 7.5 cM (Haldane map) over the genome were estimated with the program Merlin (Abecasis *et al.*, 2002).

Linkage analyses:
Linkage to a putative QTL was assessed by variance components analyses. We selected the sibling pairs for whom more than 50% of the markers were successfully measured (592 sib pairs). The average number of missing markers was 34 (SD=58), which is 4.5% of the total number of markers measured.
A genome scan for SI was carried out in 592 siblings pairs (536 individuals) for whom both phenotypic and marker data were available. Different thresholds were estimated for males and females, allowing for sex differences in the prevalence of smoking initiation.
For the NC linkage analyses the never smokers were excluded. A genome scan for NC was carried out in 351 sibling pairs (424 individuals) for whom both phenotypic data and marker data were available and included also 763 MZ and 878 DZ twin pairs with only phenotypic data. Effects of sex and age were included.
Linkage analyses were performed with variance components analyses using Mx (Neale *et al.*, 1999). Estimates of the variance component associated with a putative QTL at or near a locus are commonly obtained from either of two approaches of modeling the contribution of the QTL to the covariance among sib pairs. The two approaches are the $\hat{\pi}$ approach and the mixture approach (Neale, 2000). In the $\hat{\pi}$ approach , the covariance due to the marker or trait locus for a sib pair is modeled as a function of the $\hat{\pi}$ of the sib pair. In the mixture model, the likelihood for each sib pair is computed as the weighted sum of the likelihoods of the three models (for IBD=0, IBD=1 and IBD=2) where the weights are the probabilities that the pair is IBD 0, 1 or 2. Apart from these variance components methods for linkage analyses, other statistical methods for conducting a QTL linkage analysis have been proposed, most notably regression methods. The results presented in this paper are obtained with the $\hat{\pi}$ approach (Amos, 1994).

For the dichotomous trait SI a threshold model with one threshold was used. The model assumes an underlying liability to SI that is a function of genetic and environmental factors. Subjects are affected if they cross a threshold (Falconer and Mackay, 1996). Different thresholds were estimated for males and females, allowing for sex differences in prevalence of SI.
For NC the means were modeled according to the formula:

$$y_{ij} = \mu + ß_1 \, Age_{ij} + ß_2 \, Sex_{ij} + e_{ij},$$

where y_{ij} is the observed phenotype for sibling j in the i-th family, μ denotes the grand mean, $ß_1$ represents the regression coefficient for age, $ß_2$ represents the female deviation, age_{ij} and sex_{ij} represent the age and sex (male=0 and female=1) respectively of sib j from the i-th family, and e_{ij} represents the residual term that is not explained by the fixed effects of age and sex.
For SI and NC, the variance in liability and the phenotypic variance of the residual term, respectively, were decomposed into additive genetic variance (σ^2_a), shared environmental variance (σ^2_c), variance due to non-shared environmental influences (σ^2_e), and variance due to the QTL (σ^2_q) (Fulker and Cherny, 1996). The variance-covariance matrix for pairs j, k of the i-th family, Ω_{ijk} is given by:

$$\sigma^2_a + \sigma^2_c + \sigma^2_q + \sigma^2_e \text{ if } j = k \text{ and by } 0.5\sigma^2_a + \sigma^2_c + \hat{\pi}_{ijk}\,\sigma^2_q \text{ if } j \neq k.$$

The analyses also included the phenotypic data from MZ and DZ twin pairs for whom no genotypic data were available to allow the distinction between background additive genetic and other familial effects (1596 MZ and 1943 DZ twin pairs for SI, 763 MZ and 878 DZ twin pairs for NC). For the twin pairs who were not genotyped, covariances were modeled as $\sigma^2_g + \sigma^2_c$ for MZ pairs and $0.5\sigma^2_g + \sigma^2_c$ for DZ pairs, where $\sigma^2_g = \sigma^2_a + \sigma^2_q$. Significance of genetic variation due to the QTL was evaluated by the likelihood ratio test, from which the LOD score can be calculated by dividing the test statistic χ^2 by 2ln10 (~4.6), (Sham, 1998). In addition to the $\hat{\pi}$ approach for which the results are reported in this paper, linkage analyses were also carried out using a mixture distribution model and a regression approach in Merlin. The three methods yielded similar results.

Results

Table 1 shows the distribution for smoking behavior in the genotyped sample and in the total sample. In the genotyped sample, approximately 57% of the subjects never smoked (regularly) while in the total sample 50% never smoked (regularly). The genotyped sample contained more heavy smokers than the total sample (Table 1). This is probably due to the fact that the average age (when reporting the maximum number of cigarettes per day) in the genotyped sample (DZ twins and siblings) was higher (28.3 years, SD 13.4) than in the total sample (24.7 years, SD 11.1).

Table 8.1 Distribution of smoking behavior in genotyped sample (n=642) and total sample (n=10623).

Smoking behavior:	Genotyped sample		Totalsample	
	N	%	N	%
Never smoked (regularly)	219	49.7	6006	56.6
Less than 1 cigarette per day	41	6.4	548	5.2
1-5 cigarettes per day	54	8.4	917	8.6
6-10 cigarettes per day	56	8.7	1038	9.8
11-20 cigarettes per day	113	17.6	1502	14.1
21-30 cigarettes per day	42	6.5	506	4.8
More than 30 cigarettes per day	17	2.6	106	1.0
Total	642	100	10623	100

Tables 8.2a Model fitting results for a combined model with smoking initiation and maximum number of cigarettes smoked per day (best fitting model is given in boldface).

	initiation	nicotine dependence	χ^2	df	P	AIC
1.	full	full	75.96	57	.047	-38.04
2.	ACE	full	81.49	60	.034	-38.51
3.	AE	full	101.45	61	.000	-20.55
4.	CE	full	97.20	61	.002	-24.79
5.	full	ACE	78.80	60	.052	-41.20
6.	full	AE	84.96	61	.023	-37.04
7.	full	CE	98.86	61	.002	-23.14
8.	**ACE**	**ACE**	**78.98**	**63**	**.084**	**-47.01**

Full= full model with sex-dependent effects and a correlation between shared environmental factors in opposite sex twins (r_c) that is allowed to be less than 1; ACE=full model without sex differences; AE= additive genetic model; CE= shared environmental model; AIC=χ^2 – 2 df, this is a measure of the parsimony of the model, a lower value of AIC indicates a more parsimonious model.

Table 8.2b. Proportion of the total variance in smoking initiation and maximum number of cigarettes smoked per day that is explained by additive genetic factors (h^2), shared environmental influences (c^2) and unique environmental influences (e^2) under the best fitting model.

	h^2	c^2	e^2
Smoking initiation	.36	.56	.07
Max n cigarettes	.51	.30	.18

Three models were fitted to the phenotypic data on SI and NC: single liability, independent liability model and a combined model (Koopmans *et al.*, 1999). The combined model gave the best description of the data. Under the combined model, several alternative explanations for familial resemblance in SI and NC were evaluated. Results are shown in table 8.2a. For both SI and NC the most parsimonious model included genetic, shared environmental and unique environmental factors without sex differences. Table 8.2b depicts the parameter estimates. For SI, 36% of the variance in liability was explained by genetic factors and 56% by shared environmental factors. The remaining variance was explained by non-shared environmental factors (7%). For

NC, 51% of the variance was explained by genetic, 30% by shared environmental and 18% by non-shared environmental factors.
Because both smoking initiation and quantity were heritable traits, we explored both phenotypes in linkage analyses. For SI the highest LOD scores (> 1.5) were found on chromosomes 6,10 and 14 (Figure 8.1). For NC, the highest LOD scores (LOD > 1.5) were found for chromosomes 3 and 10 (figure 8.2).

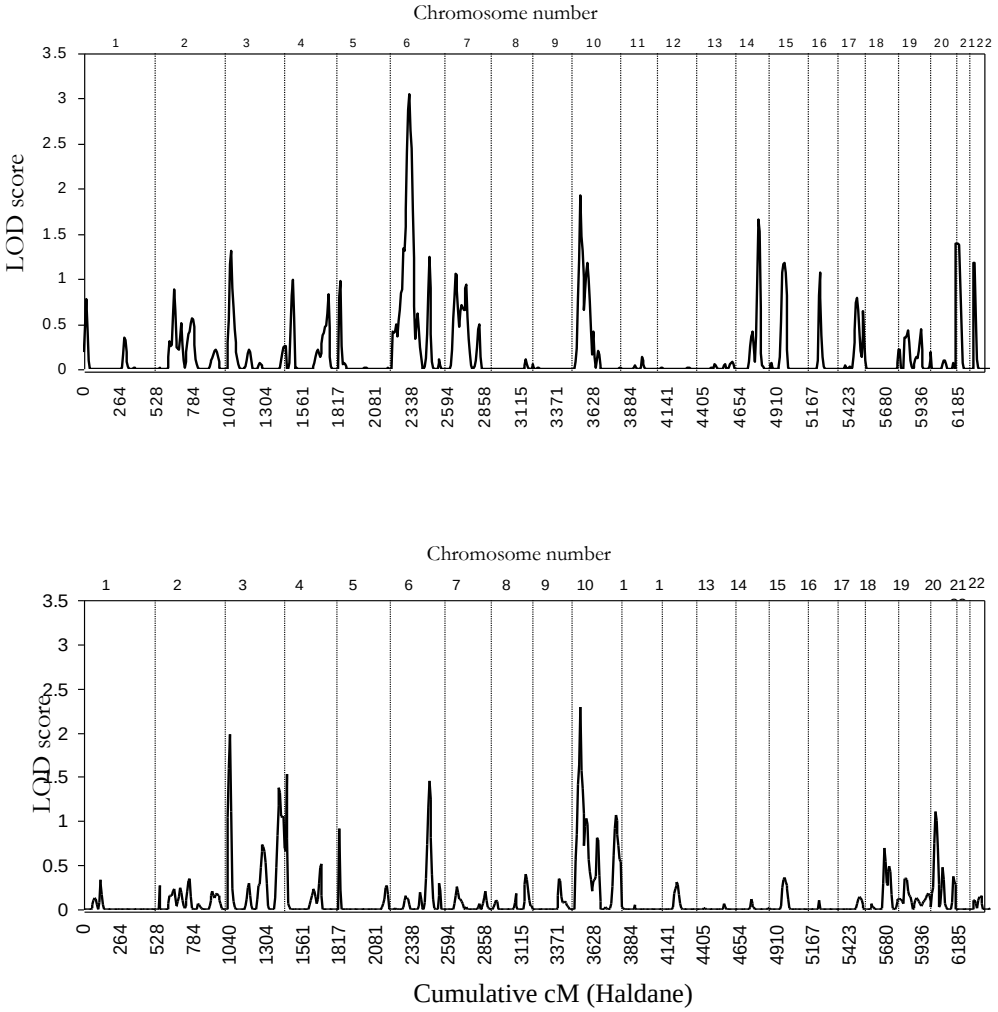
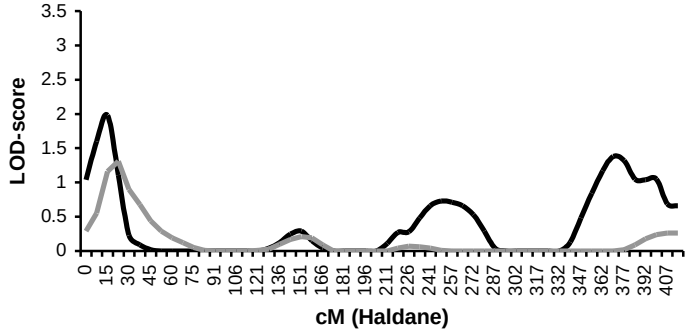


Figure 8.1 and 8.2: LOD scores across the genome for phenotype 'smoking initiation' (above) and 'maximum number of cigarettes per day' (under). The cumulative Haldane centiMorgans

are shown on the x-axis and LOD-score is shown on the y-axis. Chromosome number is shown at the top of the figures.

Chromosome 3



Chromosome 6

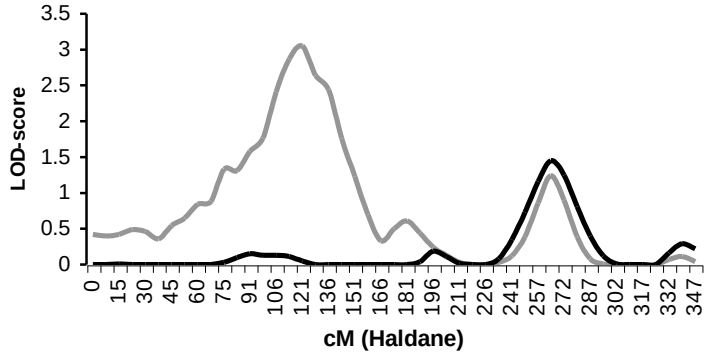


Figure 8.3 and 8.4 Linkage results for chromosome 3 and 6. Distance in Haldane cM is shown along the x-axis and the LOD-scores along the y-axis. The grey line represents the results for the phenotype 'smoking initiation' and the black line represents the results for the phenotype 'maximum number of cigarettes smoked per day' .

Figures 8.3, 8.4, 8.5 and 8.6 show LOD score plots from the linkage analyses for SI and NC for chromosomes 3, 6,10 and 14. For SI a peak was found on chromosome 6, in the region from approximately 98.1 to 143.3 cM (Haldane's map) with the highest peak (LOD=3.05) at approximately 120.7 cM in the vicinity of markers D6S2410 and D6S1053. Another peak was found on chromosome 14 at approximately 143.3 cM, in the vicinity of markers Unk283 and D14S617. For NC a peak LOD score (> 1.5) was found on chromosome 3 in the region from approximately 7.5 to 15.1 cM with the highest peak (LOD = 1.98) at approximately 15.1 cM in the vicinity of markers

D3S3050 and D3S4545. No noteworthy peaks were seen for SI on this chromosome. Finally, on chromosome 10 there was a peak for both SI and NC in the same region (37.7 –45.3 cM). The highest LOD-scores for SI (LOD score = 1.92) and NC (LOD score=2.29) were found at approximately 37.7 cM in the vicinity of markers D10S1412 and D10S1430.

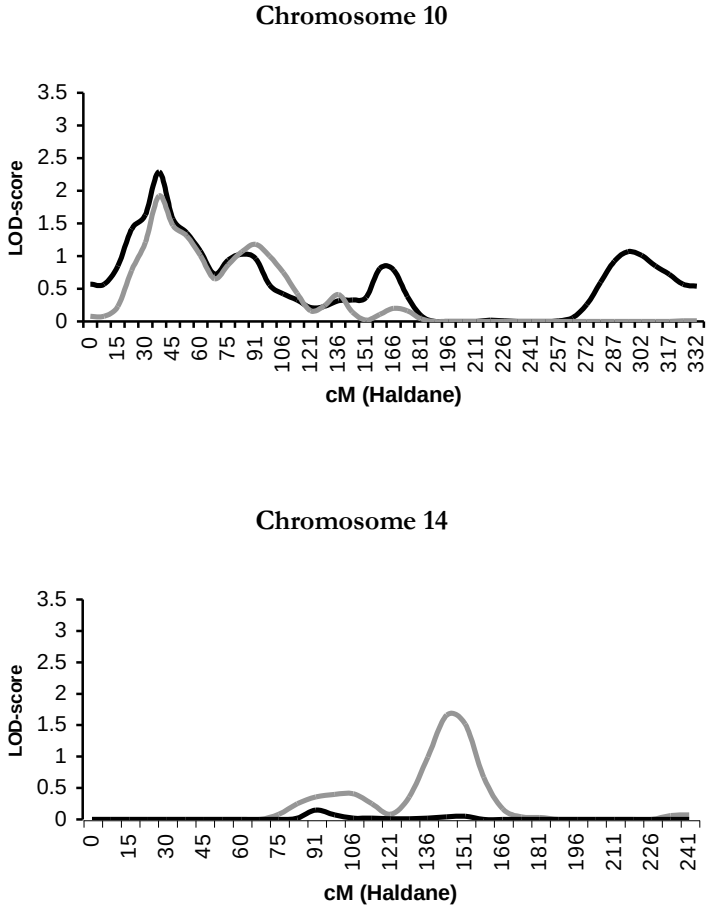


Figure 8.5 and 8.6 Linkage results for chromosome 10 and 14. Distance in Haldane cM is shown along the x-axis and the LOD-scores along the y-axis. The grey line represents the results for the phenotype 'smoking initiation' and the black line represents the results for the phenotype 'maximum number of cigarettes smoked per day' .

Discussion

Numerous twin studies of smoking behavior have been reported in the literature (reviewed by Heath and Madden (1995), Sullivan and Kendler (1999) and Hopfer *et al.* (2003)). Most studies found evidence for both genetic and shared environmental

contributions to familial resemblance in smoking behavior. We replicated this result in the combined genetic analyses of smoking initiation (SI) and number of cigarettes smoked (NC) and found the heritability for SI to be relatively low (36%). Based on 10 studies of SI, Sullivan and Kendler (1999) reported a weighted mean heritability of 56% (range from 33 to 79%). Li *et al.* (2003) obtained heritability estimates of 37% for adult males and 55% for females in a meta-analysis of SI. These estimates are somewhat higher than the heritability estimate in our study, which could be due to the fact that our sample is a relatively young one (on average 24.7 years).We observed a stronger genetic contribution to NC than to SI, as has consistently been reported others as well (e.g. Madden et al., 2004). Furthermore, we found a significant contribution of environmental factors shared by family members to variation in SI as well as in the quantity dimension. Our sample was large, which facilitates detection of shared environmental influences. What these influences consist of remains largely unknown. They may include the effects of socio-economic class (Barbeau *et al.*, 2004), religion (Koopmans *et al.*, 1999), social transmission, or the genetic effects of assortative mating (Eaves *et al.*, 1989). There is significant non-random mating for smoking behaviour (Boomsma *et al.*, 1994) and the estimate of shared environment may reflect this assortment (Willemsen, 2003).

After establishing the heritability for SI and NC, the next step was to localize chromosomal regions underlying these heritabilities by carrying out a genome scan. Linkage analyses showed peaks for SI on chromosomes 6 and 14 while for NC a peak on chromosome 3 was detected. For both SI and NC evidence for linkage was found on chromosome 10 at the same location. Those results suggest specific QTLs for SI on chromosome 6 and 14 and for NC on chromosome 3. Genetic factors common to both phenotypes are found on chromosome 10. This is in line with the model fitting results which suggested overlapping liabilities for SI and NC. The linkage results seem to argue against the suggestions made by Merikangas and Risch (2003) who questioned genetic studies of nicotine dependence. Our genome scan suggests it is possible to find evidence for linkage for smoking behavior. This evidence consists of QTLs common to SI and NC, as well as of QTLs which are unique to each phenotype. The unique QTLs which influence the quantity of cigarettes smoked become of importance only after an individual has crossed the threshold from non-smoker to smoker.

There are relatively few other linkage studies of smoking behavior. All available linkage results are summarized in Table 3. Both our study in Dutch twin families and the study by Straub *et al.* (1999) on nicotine dependence found positive results for chromosome 10. The multipoint analyses of Straub *et al.* suggested a large peak at 125 cM (Kosambi's map, 216 cM on Haldane's map) while our results suggest a large peak

at 38 cM (Haldane's map). Two studies found a peak on chromosome 3 (this study, Duggirala *et al.,* 1999). In a linkage study of substance dependence, Stallings *et al.* (2003) also found preliminary evidence for linkage to regions on chromosome 3 (LOD score 1.60) for the average number of dependence symptoms (i.e. the total symptom count across all classes of substances, including smoking, divided by the number of substances used more than five times). However, there is no overlap between the peaks of Duggirala *et al.* and Stallings *et al.* and our linkage signal on chromosome 3.

Table 8.3 Overview of positive results of linkage studies to smoking behavior. Only the most positive results are shown. The z-scores and p-values in the study of Straub et al were transformed to a χ^2 distribution which was divided by 2*ln10 to calculate the LOD-scores. The p-values from the affected sib-pair study of Bergen et al. were also transformed to a χ^2 distribution to calculate the LOD-scores.

Reference	Sample	Phenotype	Ch	Position (Kosambi)	Position (Haldane)	LOD
Straub *et al.* 1999	Christch	Nicotine Dependence	2	130-180	226-325	1.50
Vink *et al.* present study	NETAD	Quantity smoked	3	7-13	7-15	2.42
Duggirala *et al.* 1999	COGA	Ever smoked	3	105	176	1.71
Duggirala *et al.* 1999	COGA	Ever smoked	4	65	99	2.17
Bierut *et al.* 2004	COGA	Habitual smoking	5	119	204	1.12
Duggirala *et al.* 1999	COGA	Ever smoked	5	217	399	3.20
Duggirala *et al.* 1999	COGA	Ever smoked	6	63	95	1.10
Vink *et al.* present study	NETSAD	Smoking initiation	6	65-88	98-143	3.00
Bergen *et al.* 1999	COGA	Ever smoked	6	134-165	234-295	3.00
Duggirala *et al.* 1999	COGA	Ever smoked	9	0	0	1.14
Bierut *et al.* 2004	COGA	Habitual smoking	9	92/116/168	151/198/301	1.51
Bergen *et al.* 1999	COGA	Ever smoked	9	165-170	295-305	3.00
Vink *et al.* present study	NETSAD	Smoking initiation	10	24-39	30-53	2.28
Vink *et al.* present study	NETSAD	Quantity smoked	10	24-39	30-53	2.65
Straub *et al.* 1999	Christch	Nicotine Dependence	10	85-149	137-263	1.28
Bierut *et al.* 2004	COGA	Habitual smoking	11	87	141	1.64
Vink *et al.* present study	NETSAD	Smoking initiation	14	88	143	1.66
Bergen *et al.* 1999	COGA	Ever smoked	14	95-110	156-186	3.00
Duggirala *et al.* 1999	COGA	Ever smoked	15	135	236	1.97
Duggirala *et al.* 1999	COGA	Ever smoked	17	20	24	2.88
Bierut *et al.* 2004	COGA	Habitual smoking	21	0	0	1.51

CADD=family, twin and adoption studies of Colorado Center on Antisocial Drug Dependence.
COGA= Collaborative Study on the Genetics of Alcholism (analysis are based on sample from wave 1)
NETSAD= Netherlands Twin family Study of Anxious Depression
Christch = sample recruited in Christchurch, New Zealand. Inclusion criteria for a sibling pair included the presence of lifetime nicotine dependence.

The linkage result for smoking initiation on chromosome 6 was a replication of the LOD score of 1.10 reported by Duggirala *et al* (1999) for the same region. Bergen *et al* (1999) also reported positive findings for chromosome 6, though for another region. Both the present study and the study of Bergen *et al.* (1999) detected a peak on

chromosome 14, and although the signals were not located at exactly the same position, they were in the same region. It should be noted, that the location estimate of a linkage signal from a complex-trait may be many centiMorgans from the true disease locus (Roberts *et al.,* 1999).

Thus, there is some overlap between our results using the phenotypes SI and NC and the previous genome scans looking at smoking behavior and substance dependence. The different phenotypes in the studies are probably related, for example the phenotype 'maximum number of cigarettes per day' correlates highly with the score on the Fagerström Test of Nicotine Dependence (r=.66 to .70), (Vink *et al.*, submitted). However, they are not the same and each phenotype is likely to be influenced by multiple genes leading to a different picture of the genetic architectures of substance use or smoking behavior. Based on a reanalysis of genome scans on alcohol dependence, drug abuse and nicotine dependence, Uhl *et al.* (2002) report 15 regions which may harbor genes for substance abuse vulnerability, including regions on chromosome 10 and chromosome 3. These regions, however, do not overlap with the localization of the peaks in the present study.

To investigate the importance of chromosome 3, 6, 10 and 14 for SI and/or NC, the genome scan needs to be replicated in another sample using similar phenotypes. We have collected DNA samples in additional Dutch twin families selected when at least two siblings were nicotine dependent or when at least one sibling was nicotine dependent and one sibling was exposed to smoking but never smoked more than 5 cigarettes per day. We intend to carry out a genome scan in this selected sample.

If the positive results on chromosome 3, 6, 10 and 14 are replicated, candidate genes under the peaks can be considered for association analysis. Walton *et al.* and Batra *et al.* have summarized the current association studies of tobacco addiction. The strongest evidence linking particular alleles to nicotine addiction comes from studies on genetic variation in the dopaminergic system, in cytochrome P450 enzymes, the serotonin transporter gene and monoamine oxidase (Walton *et al.*, 2001; Batra *et al.*, 2003). None of these genes is located on chromosomes 3, 6, 10 or 14. However, most genes that play a major role in tobacco addiction are not yet known. Localizing and identifying the genes responsible for the linkage results on chromosomes 3, 6, 10 and 14 will help to unravel another fraction of the molecular basis of tobacco addiction.

Chapter 9

Summary and discussion

Summary and general discussion

The present thesis examined the genetics and epidemiology of smoking behavior in a large twin-family sample ascertained through the Netherlands Twin Register. In this final chapter the results and the implications for further research are discussed.

Response and non-response

When performing large, epidemiological studies using population based samples, one of the first questions is whether the sample is representative for the total population. Most studies of health and lifestyle use mailed surveys to collect data in large populations. In Europe, response rates to such surveys vary from 52 to 95%, with Dutch response rates at the lower end (Hupkens *et al.*, 1999). The overall response rate of the 2000 survey of the Netherlands Twin Register was 34% for twins and siblings. As described in the introduction, response rates differed between groups. For example, newly registered twins had a higher response rate than twins who were registered several years ago but returned none of the previous surveys. Some characteristics of those groups are compared to explore whether they are different (Table 9.1). The newly registered individuals have registered themselves while the addressed of twins registered before 1998 were obtained from city council registries and addresses of siblings registered before 1998 were obtained from the parents. In general, the newly registered individuals are older and are more often women.

Table 9.1. Response rate and characteristics of different groups for the 2000 survey

	Percentage Males	Mean age (SD)	Ever smoked
Twins registered before 1998, not completed other surveys	30.9%	30.0 (SD 11.2)	47.3%
Twins registered before 1998, completed at least one other survey	35.2%	28.0 (SD 9.8)	41.2%
Twins registered after 1997	22.5%	40.5 (SD 15.8)	48.5%
Siblings registered before 1998, not completed other surveys	45.9%	34.5 (SD 12.5)	61.4%
Siblings registered before 1998, completed at least one other survey	47.2%	30.5 (SD 10.4)	43.2%
Siblings registered after 1997	33.3%	38.3 (SD 13.7)	54.7%

The mean age of the participants of the 2000 survey is 30.1 years (SD 11.4). At this age life makes many demands on men and women: work, establishing a relationship, starting a family etc. These factors could influence the opportunity and willingness to participate. Furthermore, a substantial percentage of the non-respondents has probably moved to another address. Because the 2000 survey was not sent to parents of twins, those who changed address will not have been informed by their parents about the survey.

Non-response to mailed surveys reduces the effective sample size and may introduce bias. However, survey results will only be biased by non-participation if refusal to

participate is not distributed randomly, and is either directly or indirectly related to the traits under study. The results described in chapter 2 indicate that our data collected on health, personality and lifestyle are a reasonable reflection of the general population.

Familial association

Smoking behavior clusters in families. The results described in chapter 3 showed that the relative risk to smoke when having smoking family members or friends were clearly higher for young adolescents than for adults. Within each age group the relative risk to smoke was highest when having a smoking co-twin (especially for MZ twins) or smoking friends, somewhat lower when having smoking younger/older siblings and lowest when having smoking parents. In chapter 4 I analyzed whether the variables that were cross-sectionally associated with smoking behavior also predict the uptake of regular smoking. The uptake of regular smoking was predicted by having a smoking co-twin, smoking same-sex siblings, a smoking mother and smoking friends. Males are, in contrast to females, at a later age still susceptible to take up regular smoking.

Interestingly, both chapters showed that having smoking friends formed a high risk factor to smoke. Furthermore, same-sex siblings formed a higher risk to smoke than having opposite-sex siblings. Both the results in chapter 3 and 4 revealed a low influence of smoking parents. Previous studies of the NTR also showed that parental smoking behavior does not directly influence smoking behavior in their children. Resemblance between parents and offspring was completely accounted for by their genetic relatedness (Boomsma *et al.*, 1994b; Koopmans *et al.*, 1999). Some explanatory analyses were carried out using parent offspring data from the surveys 1991-2000. For smoking initiation (ever smoked), the best fitting model included genetic influences, shared environmental factors, unique environmental influence and cultural transmission. The cultural transmission coefficient was low indicating that children do not imitate the smoking behavior of their parents. The spouse correlation (correlation between fathers and mothers) was .24 which is in line with other studies (Boomsma *et al.*, 1994b; Koopmans *et al.*, 1999). This spouse correlation may reflect non-random mating for smoking behavior (Willemsen *et al.*, 2003).

Smoking and other traits

Analyses performed in chapter 4 indicated that in addition to having smoking family members and friends, high boredom susceptibility, high neuroticism scores, not participating in sports and alcohol use significantly predicted the uptake of regular smoking. Other studies have also found associations between smoking, alcohol use and other substances (Jarvis, 1994; Room, 2003). Our data of the 2000 survey showed that 5.2% of non-smokers never tried alcohol while only 1.9% of the smokers never tried. Furthermore, 67.5% of the non-smokers have regularly used alcohol while 85.6% of the smokers regularly drinks. For other substances the same pattern is found; smokers have more often tried soft drugs than non-smokers (47.6% versus 33.5%), smokers are more often regular users of soft-drugs (14.5% versus 1.4%) and

smokers have more often tried party-drugs (8.9% versus 1.9%). The prevalences of regular party-drugs use and of hard-drugs use were very low, but again smokers have more often used party- and hard-drugs than non-smokers.

Heritability of smoking and nicotine dependence

Smoking prevalence is lower in the younger age-groups than in the older ones but most individuals have established their smoking behavior when they are 20 years (Jefferis *et al.*, 2004). Therefore, the earlier surveys collected limited information on nicotine dependence, but the 2000 survey for the first time included comprehensive questions on nicotine dependence.

In the literature, relatively little attention is paid to the genetics of nicotine dependence. Other measures, like quantity smoked, are often used as a proxy for nicotine dependence. In our study, nicotine dependence was measured with the Fagerström Test for Nicotine Dependence (FTND) in both smokers and ex-smokers. The internal consistency of the FTND was reasonably high and results showed high test-retest correlations for both smokers and ex-smokers (chapter 5). As far as we know, no other studies measured the FTND in ex-smokers. As demonstrated in chapter 8, it is useful to have a measure of the degree of nicotine dependence for all participants who ever smoked (independent of their current smoking status) for research projects such as genetic epidemiological studies.

The degree of nicotine dependence can only be assessed in individuals who initiated smoking. Consequently, to analyze the dependence data we used models that simultaneously included smoking initiation and nicotine dependence to estimate the influence of genetic factors, shared environmental influences and unique environmental variance. For smoking initiation a heritability of 36 – 44% was found (chapter 6 and 8) and a significant contribution of environmental factors shared by family members to variation in SI (51-56%) was detected. The sample was large, which facilitates detection of shared environmental influences. What these influences consist of remains largely unknown. They may include the effects of socio-economic class (Barbeau *et al.*, 2004), religion (Koopmans *et al.*, 1999), social transmission or the genetic effects of assortative mating (Eaves *et al.*, 1989). The longitudinal survey study showed a heritability of 51% for the number of cigarettes smoked per day and .75 for nicotine dependence. Shared environmental influences significantly contributed to the variance in the number of cigarettes smoked per day (30%) but not to the variance in nicotine dependence (chapter 6 and 8).

Finding genes involved in smoking behavior

The next step after obtaining evidence for significant heritability is to identify chromosomal regions involved in smoking behavior, either by linkage or association approaches (chapter 7). The linkage approach can be used for whole genome screens to localize genes of unknown function. Genetic association studies are used to test the association of alleles at a candidate gene (or with SNPs in/near candidate genes) with a disease or with levels of a quantitative trait.

A linkage analyses was performed as described in chapter 8, with marker data from a twin-family sample. Results suggested QTLs on chromosome 6 (LOD = 3.05) and chromosome 14 (LOD = 1.66) for smoking initiation (SI). For number of cigarettes smoked per day (NC) a peak on chromosome 3 (LOD = 1.98) was detected. On chromosome 10 a peak was found in the same region for both SI (LOD = 1.92) and for NC (LOD = 2.29) which may partly explain the overlapping etiological factors for SI and NC. FTND data were only available for a smaller sample but additional FTND have recently been collected in the sixth NTR survey and will be used for linkage analyses to detect chromosomal regions involved in nicotine dependence.
The linkage peaks showed regions that enclose genes involved in smoking behavior. The regions found by the linkage analyses are still large and contain numerous genes. We explored which genes were located under the linkage peaks and found an interesting cluster of candidate genes under the linkage peak on chromosome 6. The peak encloses a cluster of genes encoding for the glutathione S-transferase alpha class genes. The cluster of alpha class genes is one of the eight classes that is identified for glutathione S-transferases. The alpha-class genes are the most abundantly expressed glutathione S-transferases in the liver. Genetic variations in the glutatione S-transferases can change an individual's susceptibility to carcinogens and toxins as well as affect the toxicity and efficacy of some drugs.
Another approach to find genes involved in smoking behavior is to perform an association study with a candidate gene. Both human and animal studies have explored candidate genes for smoking behavior. The most obvious candidate genes are genes influencing the metabolism of nicotine (like cytochrome p450), dopamine genes (including dopamine receptor genes, dopamine transporter genes and genes influencing the metabolism of dopamine), serotonergic genes and nicotine acetylcholine genes (Walton *et al.*, 2001; Batra *et al.*, 2003; Feng *et al.*, 2004; Sullivan *et al.*, 2004). Microarray and gene expression studies have been introduced into tobacco research (Li *et al.*, 2002). Using microarrays the expression pattern of thousands of genes can be monitored and differentiated through which it is a powerful tool to screen potential candidates for association studies. Konu *et al.* (2001) used this technique to study effects of nicotine in rats and identified several candidate genes that showed altered expression patterns after nicotine administration (Konu *et al.*, 2001). One of those genes is the Epac (exchange protein directly activated by camp) gene. The Epac gene is a rap1 guanine-nucleotide exchange factor involved in cAMP signal transduction pathway. Its downstream components include extracellular regulated kinase (ERK) and cAMP response element binding protein (CREB) that have also been suggested to be involved in nicotine dependence in mice (Brunzell *et al.*, 2003). The human Epac gene is located on chromosome 12. Chen *et al.* (written communication) tested the potential role that Epac plays in influencing the risk for smoking initiation and progression to nicotine dependence in a human sample. Three SNPs showed modest allele association with progression to nicotine dependence in their sample of US twins.
The DNA collected in the NETSMOK study of the Netherlands Twin Register (described in chapter 1) was used to investigate whether these results could be

replicated in a Dutch population. DNA was obtained from 1008 participants (from 302 families). SNPs for the EPAC gene were measured in the NETSMOK sample. The first analyses using the Quantitative Transmission Disequilibrium Tests (QTDT) (Abecasis *et al.*, 2000) showed no association between smoking initiation or FTND score with the three SNPs that were measured (rs757281, rs2074533 and rs2072115) in the NETSMOK sample. For the maximum number of cigarettes smoked per day an association was found with SNP rs2074533 (p= .0402). This is one of the three SNPs that showed positive results in the Richmond sample.

Usefulness of large-scale genetic studies of smoking behavior

In a paper in Science, Merikangas and Risch (2003) have questioned the usefulness of large-scale genetic studies of smoking behavior. They argued that diseases or traits appearing to be highly amenable to environmental modification should take low priority in genomic research. Indeed, the prevalence of smoking will decrease when tobacco becomes more expensive or when it is forbidden to smoke in public places (Lewitt, 1989). However, Fagerström *et al* (1996) showed that the lower the prevalence of smoking in a country, the higher the average dependence among those who smoke. For example, in the USA about 26% of the population is a smoker and the mean FTND score is 4.3 while in France approximately 36% of the population is a smoker and the mean FTND score is 3.4 (Fagerström *et al.*, 1996). It is likely that when smoking control policies reduce the availability of tobacco, the low dependent smokers are able to quit, leaving the highly dependent smokers in the population. The results in this thesis have shown that nicotine dependence is highly heritable (chapter 6). This suggest that although environmental influences may decreasse the prevalence of smoking, part of the population will remain nicotine dependent due to genetic factors. Merikangas and Rich (2003) further suggested that the public health research should focus on the social transmission of smoking as genetic influences play only a minor role. However, their arguments pertain more to smoking initiation than to number of cigarettes smoked or to nicotine dependence. In this thesis I decribe that these last two traits show substantial higher heritabilities than smoking initiation. The linkage study described in chapter 8 indicated that this heritability may be due to amplification of genetic effects which are common to SI and NC, as well as to contributions of QTLs which are unique to SI or NC. Recently, Science published a letter (Berrettini *et al.*, 2004) written by a large group of scientists who strongly disagree with the arguments of Merikangas and Risch (2003). In response to that letter, Merikangas and Risch (2004) have refered to my paper on the association of current smoking behavior with the smoking behavior of parents, siblings, friends and spouses (chapter 3 of this thesis). They argue that the greater concordance for substance abuse among non-biological than biological relatives (spouses, peers versus parents, siblings) demonstrates the importance of environmental factors in the development of substance use disorder. Indeed, the results in my study showed a considerable risk to be a smoker when most or all friends were smokers. However, it is possible that adolescents with a certain genetic predisposition actively seek out certain environmental experiences that increase their risk for the development of a

particular behavior, like smoking. The similarity of friends may be an example of an active genotype-environment correlation (Rowe, 2002). Furthermore, my study showed that the relative risk to smoke when having a smoking friend is comparable with the relative risk to smoke when having a same age sibling (DZ co-twin). In my study, the strongest test for genetic influences on smoking behavior was the comparison of the degree of similarity of smoking behavior in MZ and DZ twins. In the older groups, the relative risk to smoke was higher for MZ twins with a smoking co-twin than for DZ twins with a smoking co-twin, indicating genetic influences on smoking behavior. This finding was confirmed by the study described in chapter 4 (predictors of regular smoking).

In conclusion, both genetic and environmental factors influence smoking behavior. Environmental factors are more important for smoking initiation while genetic factors are more important for quantity and nicotine dependence. The linkage study has unraveled a small part of the molecular genetic mechanisms involved in smoking behavior. Hopefully, follow-up studies will shed light on the pathways by which some smokers become addicted and others not.

Samenvatting (Dutch summary)

Sommige personen beginnen nooit met roken, anderen proberen het wel maar worden geen dagelijkse rokers en sommige mensen worden zeer verslaafd. In dit proefschrift worden deze individuele verschillen in rookgedrag onderzocht met behulp van tweeling-familiedata uit het onderzoek naar gezondheid en leefgewoonten van het Nederlandse Tweelingen Register.

Data-verzameling en response rate
Sinds 1991 zijn er om de twee tot drie jaar vragenlijsten verstuurd naar tweelingen en hun familieleden. De vragenlijsten bevatten vragen over gezondheid, leefgewoonten (waaronder roken) en persoonlijkheid. In 2000 werd de vijfde vragenlijst verstuurd en dit keer werden naast tweelingen en hun broers en zussen ook de partners van 25 tot 30 jarige tweelingen uitgenodigd. In 2000 werden er 6792 vragenlijsten ingevuld. In totaal heeft 34% van de tweelingen en hun broers en zussen een vragenlijst ingevuld. In hoofdstuk 2 is onderzocht of deze groep een representatieve steekproef was. De mensen die niet mee wilden doen aan het vragenlijst onderzoek werden gevraagd om op een antwoordkaartje aan te geven of zij roker, ex-roker of niet-roker waren. De resultaten lieten zien dat de mensen die niet mee wilden doen iets vaker rokers waren dan de mensen die wel een vragenlijst wilden invullen. Vervolgens is een alternatieve methode geïntroduceerd waarbij de leefgewoonten van families waarvan bijna alle familieleden mee deden aan het onderzoek werden vergeleken met de leefgewoonten van families waarvan slechts een of enkele familieleden meededen aan het onderzoek. Deze laatste groep bleek in het algemeen iets ongezondere leefgewoonten te hebben (b.v. iets vaker rokers, iets minder vaak sporters). De verschillen waren echter klein en meestal niet significant.

Associatie met rookgedrag van vrienden en familieleden
In hoofdstuk 3 werd met behulp van alle vragenlijstgegevens (van 1991 tot 2000) gekeken naar de associatie tussen zelf roken en het hebben van rokende ouders, broers en zussen, vrienden of partner. De associatie blijkt het grootst te zijn bij het hebben van rokende vrienden en een rokende co-twin (vooral MZ-twin), iets lager voor rokende broers en zussen en het laagst bij rokende ouders. Voor jongens vormen rokende broers en een rokende vader een groter risico, terwijl voor meisjes rokende zussen en een rokende moeder de kans verhogen om zelf te gaan roken. Het risico om zelf te roken als familie en vrienden roken is groter voor de 12 tot 15 jarigen dan voor de oudere groepen. In hoofdstuk 4 werd met behulp van een longitudinale analyse onderzocht welke factoren voorspellen dat jongeren gaan beginnen met roken. Hievoor werden de jongeren die in 1993 niet rookten geselecteerd. Vervolgens werd onderzocht of het rookgedrag van familieleden en vrienden in 1993 een voorspeller was voor hun rookgedrag in 1995. De resultaten lieten zien dat vooral het hebben van een rokende tweelingbroer of -zus, rokende broers en zussen van hetzelfde geslacht, een rokende moeder en rokende vrienden de kans verhogen om later zelf te gaan roken. Daarbij is dee kans om te gaan roken is ook groter voor jongeren die weinig sporten, regelmatig alcohol drinken of hoog scoren op een neuroticisme schaal.

Erfelijkheid van roken en nicotine-verslaving
Als men eenmaal begonnen is met roken, dan raakt de ene persoon wel verslaafd aan nicotine en de andere niet. Vaak wordt het aantal sigaretten dat iemand per dag rookt gebruikt als indicatie voor de mate van nicotineafhankelijkheid. In dit proefschrift is, naast het aantal sigaretten per dag, gebruik gemaakt van een vragenlijst genaamd de Fagerström Test for Nicotine Dependence (FTND). Deze lijst bevat vragen als 'Vind u het moeilijk om niet te roken op plaatsten waar dat verboden is?'. De totale score varieert van 0 tot 10 waarbij een hogere score duidt op een hogere nicotineafhankelijkheid. De FTND wordt meestal afgenomen bij rokers. Wij hebben de FTND echter ook afgenomen bij de ex-rokers en we hebben gevraagd of ze de vragen willen beantwoorden over de periode dat ze het zwaarst rookten. Om te onderzoeken of erfelijke factoren van invloed zijn op nicotine afhankelijkheid is het namelijk belangrijk om de mate van verslaafdheid te meten ongeacht de huidige rookstatus. De resultaten beschreven in hoofdstuk 5 laten zien dat de interne consistentie van de FTND redelijk hoog was, zowel voor de rokers als de ex-rokers. Ook de test-hertestcorrelaties waren goed.
De mate van nicotineafhankelijkheid (FTND /aantal sigaretten per dag) kan alleen in kaart gebracht worden voor mensen die ooit gerookt hebben. Maar iemand die nooit gerookt heeft zou ook een bepaalde gevoeligheid voor verslaving kunnen hebben. Wel of niet beginnen met roken en de mate van verslaafdheid kunnen dus niet los van elkaar gezien worden. In hoofdstuk 6 en 8 zijn modellen beschreven die beginnen met roken en de mate van verslaving combineren. Resultaten laten zien dat de individuele variatie in beginnen met roken voor 36-44% wordt bepaald door erfelijke factoren. De gedeelde omgeving (opvoeding, school etc) blijkt echter nog belangrijker te zijn (51-56%). Als iemand eenmaal begonnen is met roken dan blijkt de individuele variatie in het aantal sigaretten voor 51% door erfelijke factoren bepaald te worden en voor 30% door gedeelde omgevingsinvloeden (hoofdstuk 8). De variatie in nicotine-afhankelijkheid (gemeten met de FTND) blijkt grotendeels door erfelijke factoren bepaald te worden (75%) en de gedeelde omgevingsinvloeden spelen hierbij geen rol (hoofdstuk 6).

Genen die betrokken zijn bij rookgedrag
Nadat is vastgesteld dat rookgedrag (voor een deel) beïnvloed wordt door erfelijke factoren is de volgende stap het in kaart brengen van de genen die betrokken zijn bij rookgedrag. Er zijn 2 manieren om naar genen te zoeken: 'linkage' en 'associatie'. Bij een linkageanalyse wordt het gehele genoom gescreend waarbij er interessante regio's gedetecteerd kunnen worden die mogelijk van belang zijn voor roken. Bij associatiestudies wordt er onderzocht of er een verband bestaat tussen genetische variatie in (of vlak bij) een kandidaatgen en rookgedrag. Beide methoden worden beschreven in hoofdstuk 7.
In hoofdstuk 8 staan de resultaten van een linkageanalyse voor beginnen met roken en het aantal sigaretten per dag. Voor beginnen met roken zijn er interessante regio's gevonden op chromosoom 6 en 14. Voor het aantal sigaretten per dag vonden we een gebied op chromosoom 3. Tenslotte werd er een interessante regio gevonden op

chromosoom 10 voor zowel beginnen met roken als het aantal sigaretten. In zo'n regio liggen veel verschillende genen, dus dit is slechts een eerste stap bij het zoeken naar genen die betrokken zijn bij rookgedrag. Vervolgstudies zullen meer inzicht geven in de mechanismen die betrokken zijn bij rookgedrag en verklaren waarom sommige rokers verslaafd raken en anderen niet.

References

Aarnio M, Winter T, Kujala UM and Kaprio J (1997). Familial aggregation of leisure-time physical activity -- a three generation study. Int J Sports Med. **18**(7): 549-56.

Abecasis GR, Cardon L and Cookson WOC (2000). A General test of association for quantitative traits in nuclear families. Am. J. Hum. Genet **662**: 279-292.

Abecasis GR, Cherny SS, Cookson OC and Cardon L (2002). Merlin-rapid analyses of dense genetic maps using sparse gene flow trees. Nature Genetics **30**: 97-101.

Abecasis GR, Cherny SS, Cookson WO and Cardon LR (2001). GRR: graphical representation of relationship errors. Bioinformatics **17**: 742-743.

Achenbach TM (1997). Manual for the Young Adult Self Report and Young Adult Behavior Checklist. Burlington, VT, University of Vermont, Department of Psychiatry.

Alexander C, Piazza M, Mekos D and Valente T (2001). Peers, Schools, and adolescent cigarette smoking. Journal of Adolescent Health **29**(1): 22-30.

Amos CI (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. Am. J. Hum. Genet **54**(3): 535-43.

Barchielli A and Balzi D (2002). Nine-year follow-up of a survey on smoking habits in Florence (Italy): higher mortality among non-responders. International Journal of Epidemiology **31**: 1038-1042.

Batra V, Patkar AA, Berrettini WH, Weinstein SP and Leone FT (2003). The genetic determinants of smoking. CHEST **123**(5): 1730-1739.

Bauman KE, Carver C and Gleiter K (2001). Trends in parent and friend influence during adolescence. The case of adolescent cigarette smoking. Addictive Behaviors **26**: 349-361.

Bauman KE and Ennet ST (1996). On the importance of peer influence for adolescent drug use: commonly neglected considerations. Addiction **91**(2): 185-198.

Bauman KE and Fisher LA (1996). On the measurement of friend behavior in research on friend influence and selection: findings from longitudinal studies on adolescent smoking and drinking. Journal of Youth and Adolescence **15**: 345-353.

Beck AT, Ward CH, Mendelson M, Mock J and Erbaugh J (1961). An inventory measuring depression. Arch. gen. Psychiatry **4**: 53-63.

Beijsterveldt van CEM and Baal van GCM (2001). Twin and family studies of the human electroencephalogram: a review and a meta-analysis. Biological Pschology **61**(1): 111-138.

Bergen AW, Korczak JF, Wieissbecker KA and Goldstein AM (1999). A genome-wide search for loci contributing to smoking and alcoholism. Genetic Epidemiology **17**(S1): S55-S60.

Bergstrand R, Vedin A, Wilhelmsson C and Wilhelmsen L (1983). Bias due to non-participation and heterogenous sub-groups in population surveys. Journal of Chronic Diseases **36**(10): 725-728.

Berrettini WH, Bierut LJ, Crowley TJ, J.F. C, Frascella J, Gelernter J, Hewitt JK, Kreek MJ, Lachman H, Leppert M, Li MD, Madden PAF, Miner C, Pollock JD, Pomerleau O, Rice JP, Rutter JL, Shurtleff D, Swan GE, Tischfield JA, Tsuang

MT, Uhl GR, M.M. V, Volkow ND and Wanke K (2004). Setting priorities for genomic research. Science **304**(4 june): 1445-1446 (letter).

Bierut LJ, Rice JP, Goate A, Hinrichs AL, Saccone NL, Foroud T, Edenberg HJ, Cloninger CR, Begleiter H, Conneally PM, Crowe RR, Hesselbrock V, Li TK, Nurnberger JIj, Porjesz B, Schuckit MA and Reich T (2004). A genomic scan for habitual smoking in families of alcholics: common and specific genetic factors in substance dependence. American Journal of Medical Genetics **124**(A): 19-27.

Boomsma DI (1998). Genetic analysis of cognitive failures (CFQ): a study of Dutch adolescent twins and their parents. European Journal of Personality **12**(5): 321-330.

Boomsma DI, Beem AL, Berg van den M, Dolan CV, Koopmans JR, Vink JM, Geus de EJC and Slagboom PJ (2000). Netherlands Twin Family Study of anxious depression (NETSAD). Twin Research **3**(4): 323-334.

Boomsma DI, Geus EJC, de, Baal GCM, van and Koopmans JR (1999). A religious upbringing reduces the influence of genetic factors on disinhibition: evidence for interactions between genotype and environment on personality. Twin Research **2**: 115-125.

Boomsma DI, Koopmans JR, Doornen v, L.J.P. and Orlebeke JF (1994a). Genetic and social influences on starting to smoke: a study of Dutch adolescent twins and their parents. Addiction **89**: 219-226.

Boomsma DI, Koopmans JR, Doornen vLJP and Orleke JF (1994b). Genetic and social influences on starting to smoke: a study of Dutch adolescent twins and their parents. addiction **89**: 219-226.

Boomsma DI, Vink JM, Beijsterveldt CEMv, Geus de EJC, Beem AL, Mulder EJCM, Riese H, Willemsen AHM, Bartels M, Berg van den M, Derks EM, Graaff SC, Kupper HM, Polderman JC, Rietveld MJH, Stubbe JH, Knol LI, Stroet T and Baal GCM (2002). Netherlands Twin Register: a focus on longitudinal research. Twin Research **5**(5): 401-406.

Boyle MH, Sanford M, Szatmari P, Merikangas K and Offord DR (2001). Familial influence on substance use by adolescents and young adults. Canadian Journal of Public Health **92**(3): 206-209.

Brenner H and Scharrer SB (1996). Parental smoking and sociodemographic factors related to smoking among German medical students. Eur. Journal of Epidemiology **12**(2): 171-176.

Breslau N, Johnson EO, Hiripi E and Kessler R (2001). Nicotine dependence in the united states; prevalence, trends and smoking persistence. Arc. gen. Psychiatry **58**: 810-816.

Broman KW, Murray JC, Sheffield VC, White RL and Weber JL (1998). Comprehensive human genetic map: individual and sex-specific variantion in recombination. Am. J. Hum. Genet **63**(3): 891-9.

Brunzell DH, Russell DS and Picciotto MR (2003). In vivo nicotine treatment regulates mesocorticolimbic CREB and ERK signaling in C57B1/6J mice. J. Neurochem. **84**(6): 1431-1441.

Busjahn A, Freier K, Faulhaber HD, Li GH, Rosenthal MR, Jordan J, Hoehe MR, Timmermann B and Luft FC (2002). Beta2 Adrenergic receptor gene variations and coping styles in twins. Biological Pschology **61**: 97-109.

Colby SM, Tiffany ST, Shiffman S and Niaura RS (2000). Measuring nicotine dependence among youth: a review of available approaches and instruments. Drug and Alcohol Dependence **59**(Suppl. 1): S23-S39.

Dierker LC, Avenevoli SA, Stolar M and Merikangas KR (2002). Smoking and depression: an examination of mechanisms of comorbidity. Am. J. Psychiatry **159**: 947-953.

Dijkstra A and Tromp D (2002). Is the FTND a measure of physical as well as psychological tobacco dependence? Journal of Substance Abuse Treatment **23**: 367-374.

Distefan JM, Gilpin EA, Choi WS and Pierce JP (1998). Parental influences predict adolescent smoking in the United States, 1989-1993. Journal of Adolescent Health **22**(6): 466-474.

Duggirala R, Almasy L and Blangero J (1999). Smoking behavior is under the influence of a major quantitative trait locus on human chromosome 5q. Genetic Epidemiology **17**(S1): S139-S144.

Eaves LJ, Eysenck HJ and Martin NG (1989). Genes, Culture and Personality. London, Academic Press.

Eaves LJ, Heath AC, Martin N, Maes HH, Neale MC, Kendler KS, Kirk K and Corey L (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. Twin Research **2**(2): 62-80.

Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R and Kwan I (2002). Increasing response rate to postal questionnaires: systematic review. Britisch Medical Journal **324**(18 MAY): 1-9.

Engels R, Knibbe RA, Drop MJ and de Haan YT (1997). Homogeneity of cigarette smoking within peer groups: Influence or selection? Health Education & Behavior **24**(6): 801-811.

Engels RCME, Knibbe RA and Drop MJ (1999). Predictability of smoking in adolescence: between optimism and pessimism. Addiction **94**(1): 115-124.

Escobedo LG and Peddicord JP (1996). Smoking prevalence in US birth cohorts: the influence of gender and education. Am. J. of Pub. Health **86**(2): 231-6.

Etter JF, Kozlowski LT and Perneger TV (2003). What smokers believe about light and ultralight cigarettes. Preventive Medicine **36**(1): 92-8.

Etter JF and Perneger TV (1997). Analysis of non-response bias in a mailed health survey. Journal of Clinical Epidemiology **50**(10): 1123-1128.

Etter JF, Vu Duc T and Perneger TV (1999). Validity of the Fagerström test for nicotine dependence and the heaviness of smoking index among relatively light smokers. Addiction **94**(2): 269-281.

Everret SA, Warren CW, Kann SD, Husten CG and Crosset LS (1999). Initiation of cigarette smoking and subsequent smoking behavior among U.S. high school students. Preventive Medicine **29**(5): 327-333.

Fagerström KO (1978). Measuring the degree of physical dependence to tobacco smoking with reference to individualization of treatment. Addicitive Behaviors **3**: 235-241.

Fagerström KO, Kunze M, Schoberberger R, Breslau N, Hughes JR, Hurt RD, Puska P, Ramstrom L and Zatonski W (1996). Nicotine dependence versus smoking prevalence: comparison among countries and categories of smokers. Tobacco Control **5**: 52-56.

Falconer DS and Mackay TFC (1996a). Quantitative genetics. Essex, Longman Group Ltd.

Falconer DS and Mackay TFC (1996b). Treshold Characters (chapter 18). Quantitative Genetics. Essex, Longman Group Ltd.

Feng Y, Niu T, Xing H, Xu X, Chen C, Peng S, Wang L, Laird N and Xu X (2004). A common haplotype of the nicotine acetylcholinge receptor alpha4 subunit gene is associated with vulnerability to nicotine addiction in men. Am. J. Hum. Genet **75**: 112-121.

Fulker DW and Cherny SS (1996). An improved multipoint sib-pair analysis of quantitative traits. Behavior Genetics **26**(5): 527-531.

Gallus S, Colombo P, Scarpino V, Zuccaro P, Apolone G and La Vecchia (2002). Smoking in Italy, 2002. Tumori **88**(6): 453-6.

Geus de EJC, Boomsma DI and Snieder H (2003). Genetic correlation of exercise with heart rate and respiratory sinus arrhytmia. Med. Sci. Sports Exerc. **35**(8): 1287-1295.

Graham H and Der G (1999). Patterns of predictors of tobacco consumption among women. Health Education Research **14**(5): 611-619.

Green G, Macintyre S, West P and Ecob R (1991). Like parent like child? Association between drinking and smoking behaviour of parents and their children. British Journal of Addiction **86**: 745-758.

Heath AC, Howells W, Kirk K, Madden PAF, Bucholz K, Nelson EC, Slutske WS, Statham DJ and Martin N (2001). Predictors of non-response to a questionnaire survey of a volunteer twin panel: findings from the australian 1989 twin cohort. Twin Research **4**(2): 73-80.

Heath AC and Madden PAF (1995). Genetic influences on smoking behavior. Behavior Genetic Approaches in Behavioral Medicine. New York, Plenum Press**:** 45-66.

Heath AC, Madden PAF and MArtin NG (1998). Statisitcal methods in genetic research on smoking. Statistical methods in medical research **7**: 165-186.

Heath AC, Martin N, Lynskey MT, Todorov AA and Madden PAF (2002). Estimating two-stage models for genetic influences on alcohol, tobacco or drug use initiation and dependence vulnerability in twin and family data. Twin Research **5**(2): 113-124.

Heath AC and Martin NC (1993). Genetic models for the natural history of smoking: evidence for a genetic influence on smoking persistence. Addictive Behaviors **18**: 19-34.

Heatherton TF, Kozlowski LT, Frecker RC and Fagerström KO (1991). The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. British journal of addiction **86**: 1119-1127.

Herlitz C and Westholm B (1996). Smoking and associated factors among young Swedish females. Scand J Prim Health Care **14**: 209-215.

Hill A, Roberts J, Ewings P and Gunnell D (1997). Non-response bias in a lifestyle survey. Journal of Public Health Medicine **19**(2): 203-207.

Hopfer CJ, Crowley TJ and Hewitt JK (2003). Review of twin and adoption studies of adolescent substance use. J. Am. Acad. Child Adolesc. Psychiatry **42**(6): 710-719.

Horn K, Fernandes A, Dino G, Massey CJ and Kalsekar I (2003). Adolescent nicotine dependence and smoking cessation outcomes. Addicitive Behaviors **28**: 769-776.

Hover SJ (1988). Factors associated with smoking behavior in adolescent girls. Addicitve Behaviors **13**: 139-145.

Hupkens CLH, Berg vdJ and Zee vdJ (1999). National health interview surveys in Europe: an overview. Health Policy **47**: 145-168.

Jacobsen BK and Thelle DS (1988). The Tromso Heart Study: responders and non-responders to a health questionnaire, do they differ? Scandinavian Journal of Social Medicine **16**(2): 101-104.

Jarvis MJ (1994). A profile of tobacco smoking. Addiction **89**: 1371-1376.

Jefferis BJ, Power C, Graham H and Manor O (2004). Changing social gradients in cigarette smoking and cessation over two decades of adult follow-up in a British birth cohort. J Public Health **26**(1): 13-8.

Jensen EJ and Overgaard E (1993). Investigation of smoking habits among 14-17-year-old boarding school pupils: factors which influence smoking status. Public Health **107**: 117-123.

John U, Meyer C, Hapke U, Rumpf HJ, Schumann A, Adam C, Alte D and Ludemann J (2003). The Fagerström test for nicotine dependence in two adult population samples - potential influence of lifetime amount of tobacco smoked on the degree of dependence. Drug and Alcohol Dependence **1**(6): in press.

Kendler KS, Neale MC, Sullivan P, Corey LA, Gardner CO and Prescott CA (1999). A population-based twin study in women of smoking initiation and nicotine dependence. Psychological Medicine **29**: 299-308.

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR and Stefansson K (2002). A high-resolution recombination map of the human genome. Nature Genetics **31**(3): 241-247.

Konu O, Kane JK, Barett T, Vawter MP, Chang R, Ma JZ, Donovan DM, Sharp B, Becker KG and Li MD (2001). egion-specific transcriptional response to chronic nicotine in rat brain. Brain Res **909**(1-2): 194-203.

Koopmans JR (1997). The genetics of health-related behaviors. Department of Biological Psychology. Amsterdam, Vrije Universiteit.

Koopmans JR and Boomsma DI (1996). Familial resemblances in alcohol use: Genetic or cultural transmission? Journal of Studies on Alcohol **57**: 19-28.

Koopmans JR, Doornen vLJP and Boomsma DI (1994). Smoking and sports participation. Genetic Factors in Coronary Heart Desease. UGe al., Kluwer**:** 217-235.

Koopmans JR, Slutske W, Heath AC, Neale MC and Boomsma DI (1999). The genectics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. Behavior Genetics **29**(6): 383-393.

Korkeila K, Suominen S, Ahvenainen J, Ojanlatva A, Rautava P, Helenius H and Koskenvuo M (2001). Non-response and related factors in a nation0wide health survey. European Journal of Epidemiology **17**: 991-999.

Kotaniemi JT, Hassi J, Kataja M, Jonsson E, Laitinen LA, Sovijarvi ARA and Lundback B (2001). Does non-responder bias have a significant effect on the results in a postal questionnaire study? European Journal of Epidemiology **17**: 809-817.

Lando HA, Thai DT, Murray DM, Robinson LA, Jeffery RW, Sherwood NE and Hennrikus DJ (1999). Age of initiation, smoking patterns, and risk in a population of working adults. Preventive Medicine **29**(6): 590-598.

Lawley DN (1943). A note on Karl Pearson's selection formulae. Proceedings of the Royal Society of Edinburgh **62**: 28-30.

Lewitt EM (1989). US tobacco taxes: behavioral effects and policy implications. Britisch Journal of Addiction **84**: 1217-1235.

Li MD, Cheng R, Ma JZ and Swan GE (2003). A meta-analysis of estimated and environmental effects on smoking behavior in male and female adult twins. Addiction **98**(1): 23-31.

Li MD, Konu O, Kane JK and Becker KG (2002). Microarray technology and its application on nicotine research. Mol. Neurobiol. **25**(3): 265-285.

Loon vAJM, Tijhuis M, Picavet HSJ, Surtees PG and Ormel J (2003). Survey non-response in the Netherlands: effects on prevalence estimates and associations. Ann Epidemiol **13**: 105-110.

Macera CA, Jackson KL, Davis DR, Kronenfeld JJ and Blair SN (1990). Patterns of non-response to a mail survey. Journal of Clinical Epidemiology **43**(12): 1427-1430.

Madden PAF, Bucholz K, Todorov AA, Grant JD and Heath AC (2002). The assesment of peer selection and peer environmental influences on behavior using pairs of siblings or twins. Twin Research **5**(1): 38-43.

Malmstadt JR, Nordstrom DL, Carty DC, Christiansen AL, Chudy M, Rumm PD and Remington PL (2001). Cigarette smoking in Wisconsin: the influence of race, ethnicity and socioeconomics. WMJ **100**(3): 29-33.

Mayhew KP, Flay BR and Mott JA (2000). Stages in the development of adolescent smoking. Drug and Alcohol Dependence **59**(S1): S61-S81.

Maziak W and Mzayek F (2000). Characterization of the smoking habit among high school students in Syria. European Journal of Epidemiology **16**: 1169-1176.

McNeill AD, Jarvis MJ, Stapleton JA, Russell MAH, Eiser JR, Gammage P and Gray EM (1988). Prospective study of factors predicting uptake of smoking in adolescents. Journal of Epidemiology and Community Health **43**: 72-78.

Meijer B, Branski D, Knol K and Keren E (1996). Cigarette smoking habits among schoolchildren. CHEST **110**: 921-926.

Merikangas K and Risch N (2003). Genomic priorities and public health. Science **302**(24): 599-601.

Merikangas K and Risch N (2004). Response to 'setting priorities for genomic research'. Science **304**: 1446-1447 (letter).

Meulenbelt I, Droog S, Trommelen GJM, Boomsma DI and Slagboom PJ (1995). High yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. American Journal of Human Genetic **57**: 1252-1254.

Moran A, Maguire N and Howell F (2000). Smoking and quitting among Irish teenage males. Irish Medical Journal **93**(9): 272-273.

Neale MC (2000). QTL mapping with sib-pairs: the flexibility of Mx. Advances in twin and sib-pair analysis. TD Spector, H Snieder and AJ MacGregor. London, Greenwich Medical Media Ltd.**:** 219-243.

Neale MC, Boker SM, Xie G and Maes HH (1999). Mx: Statistical Modeling. Richmond, VA 23298, VCU Box 900126.

Neale MC and Cardon L (1992). Methodology for genetic studies of twins and families. Dordrecht, the Netherlands, Kluwer Academic Publishers.

Ogden MW, Morgan WT, Heavner DL, Davis RA and Steichen TJ (1997). National incidence of smoking and misclassification among the U.S. married female population. Journal of Clinical Epidemiology **50**(3): 253-263.

O'Loughlin J, Paradis G, Renaud L and Gomez LS (1998). One-year predictors of smoking initiation and of continued smoking among elementary schoolchildren in multiethnic, low- income, inner-city neighbourhoods. Tobacco Control **7**(3): 268-275.

Osler M, Holst C, Prescott E and Sorensen TIA (2001a). Influence of genes and family environment on adult smoking behavior assessed in an adoption study. Genetic Epidemiology **21**: 193-200.

Osler M, Holstein B, Avlund K, Damsgaard MT and Rasmussen NK (2001b). Socioeconomic position and smoking behaviour in Danisch adults. Scand J Public Health **29**(1): 32-9.

Oygard L, Klepp KI, Tell GS and Vellar OD (1995). Parental and peer influence on smoking among young adults: ten-year follow-up of the Oslo youth study participants. Addiction **90**: 561-569.

Payne TJ, Smith PO, McGracken LM, McSherry WC and Antony MM (1994). Assessing nicotine dependence: a comparison of the Fagerström Tolerance Questionnaire (FTQ) with the Fagerström Test for Nicotine Dependence (FTND) in a clinical sample. Addicitive Behaviors **19**: 307-317.

Pearson K (1903). Mathematical contributions to the theory of evolution--XI. On the influence of natural selection on the variability and correlation of organs. Philosophical Transactions **200-A**: 1-66.

Perkins KA (1999). Nicotine discrimination in men and women. Pharmacology Biochemistry and Behavior **64**(2): 295-299.

Perkins KA, Jacobs L, Sanders M and Caggiula AR (2002). Sex differences in the subjective and reinforcing effects of cigarette nicotine dose. Psychopharmacology **163**: 194-201.

Pomerleau CS, Carton SM, Lutzke ML, Flessland KA and Pomerleau OF (1994). Reliability of the Fagerström tolerance questionnaire and the Fagerström test for nicotine dependence. Addictive Behaviors **19**(1): 33-39.

Porjesz B, Begleiter H, Wang K, Almasy L, Chorlian DB, Stimus AT, Kuperman S, O'Connor SJ, Rohrbaugh J, O. Bauer LO and al. e (2002). Linkage and linkage disequilibrium mapping of ERP and EEG phenotypes. Biological Pschology **61**: 229-248.

Price RA, Chen K-H, Cavalli - Sforza LL and Feldman MW (1981). Models of spouse influence and their application to smoking behavior. Social Biology **28**(1-2): 14-29.

Price RA and Vandenberg SG (1980). Spouse similarity in American and Swedish couples. Behavior Genetics **10**(1): 59-71.

Reijneveld SA and Stonks K (1999). The impact of response bias on estimates of health care utilization in a metropolitan area: the use of administrative data. International Journal of Epidemiology **28**: 1134-1140.

Roberts SB, MacLean CJ, Neale MC, Eaves LJ and Kendler KS (1999). Replication of linkage studies of complex traits: an examination of variation in location estimates. Am. J. Hum. Genet **65**: 876-884.

Room R (2003). Smoking and drinking as complementary behaviours. Biomedicine & Pharmacotherapy **58**: 111-115.

Rowe DC (2002). Assesing genotype-environment interactions and correlations in the postgenomic era. Behavioral Genetics in the postgenomic era. R Plomin, Defries, J.C., Craig, I.W., McGuffin, P. Washington, American Psychological Association**:** 71-86.

Sasco AJ, Pobel D, Benhaim V, Bruin de K, Stiggelbout A and Tuyns A (1993). Smoking habits in French adolescents. Epidemiology and Public Health **41**: 461-472.

Scarr S and McCartney K (1983). How people make their own environments: a theory of genotype - environment effects. Child Development **54**(424-435).

Schäffer AA (1996). Faster linkage analysis computations for pedigrees with loops or unused alleles. Hum. Hered. **46**(4): 226-35.

Schaufeli WB, Leiter MP, Maslach C and Jackson SE (1996). The MBI-General Survey. Maslach Burnout Inventory. C Maslach, SE Jackson and MP Leiter, Palto Alto, CA: Consutling Psychologists Press**:** 19-26.

Sham P (1998). Statistics in human genetics. New York, Oxford university press.

Shamsuddin K and Abdul Harris M (2000). Family influence on current smoking habits among secondary school children in Kota Bharu, Kelantan. Singapore Med Journal **41**(4): 167-171.

Slagboom PE and Meulenbelt I (2002). Organisation of the human genome and our tools for identifying disease genes. Biological Pschology **61**: 11-31.

Snieder H, Harshfield GA, Barbeau P, Pollock DM, Pollock JS and Treiber FA (2002). Dissecting the genetic architecture of the cardiovascular and renal stress response. Biological Pschology **61**: 73-95.

Snieder H, van Doornen LJ and Boomsma DI (1997). The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. American Journal of Human Genetics **60**(3): 638-50.

Stallings MC, Corley RP, Hewitt JK, Krauter KS, Lessem JM, Mikulich SK, Hyun Rhee S, Smolen A, Young SE and Crowley TJ (2003). A genome-wide search for quantititative trait loci influencing substance dependence in vulnerability in adolescence. Drug and Alcohol Dependence **70**(3): 295-307.

Stang A (2003). Nonresponse research - an underdeveloped field in epidemiology. European Journal of Epidemiology **18**: 929-931.

Straub RE, Sullivan PF, Ma Y, Myakishev MV, Harris-Kerr C, Wormley B, Kadami B, Sadek H, Silverman MA, Webb BT, Neale MC, Bulik CM, Joyce PR and Kendler KS (1999). Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample suggest that regions on chromosones 2, 4, 10, 16, 17 and 18 merit further study. Molecular Psychiatry **4**: 129-144.

Sullivan PF and Kendler KS (1999). The genetic epidemiology of smoking. Nicotine Tobacco Res. **1**: S51-S57.

Sullivan PF, Neale B, Oord van den E, Miles MF, Neale MC, Bulik CM, Joyce PR, Straub RE and Kendler KS (2004). Candidate genes for nicotine dependence via linkage, epistasis, and bioinformatics. Am. J. of Med. Gen Part B (Neuropsychiatric Genetics) **126B**: 23-36.

Swan AV, Creeser R and Murray M (1990). When and why children first start to smoke. International Journal of Epidemiology **19**(2): 323-330.

Swan GE, Carmelli D and Cardon LR (1997). Heavy consumption of cigarettes, alcohol and coffee in male twins. Journal of Studies on Alcohol **4**: 182-190.

Thun M and Burns DM (2001). Health impact of 'reduced yield' cigarettes: a critical assessment of the epidemiological evidence. Tobacco Control **10**(Suppl I): i4-i11.

Tyas SL and Pederson LL (1998). Psychosocial factors related to adolescent smoking: a critical review of the literature. Tobacco Control **7**: 409-420.

Uhl GR, Liu QR and Naiman D (2002). Substance abuse vulnerability loci: converging genome scanning data. TRENDS in Genetics **18**(8): 420-425.

Verhulst FC, Ende Jvd and Koot HM (1997). Handleiding voor de Youth Self-Report. Rotterdam, Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis / Academisch Ziekenhuis Rotterdam / Erasmus Universiteit Rotterdam.

Vink JM, Beem AL, Posthuma D, Neale MC, Willemsen G, Kendler KS, Slagboom PE and Boomsma DI (2004-a). Linkage analysis of smoking initiation and quantity in Dutch sibling pairs. The Pharmacogenomics Journal 4;274-282

Vink JM and Boomsma DI (2002). Gene finding strategies. Biological Pschology **61**: 53-71.

Vink JM, Willemsen G, Beem AL and Boomsma DI (in press). The Fagerström Test for Nicotine Dependence in a Dutch sample of daily smokers and ex-smokers. Addictive Behaviors.

Vink JM, Willemsen G and Boomsma DI (2003a). The association of current smoking behavior with the smoking behavior of parents, siblings, friends and spouses. Addiction **98**: 923-931.

Vink JM, Willemsen G, Engels R and Boomsma DI (2003b). Smoking status of parents, siblings and friends: predictors of regular smoking? Findings from a longitudinal twin-family study. Twin Research **6**(3): 209-217.

Vink JM, Willemsen G, Stubbe JH, Middeldorp CM, Ligthart RSL, Baas KD, Dirkzwager HJC, Geus de EJC and Boomsma DI (2004-c). Estimating non-response bias in family studies: application to mental health and lifestyle. European Journal of Epidemiology 19(7); 623-630

Walton R, Johnstone E, Munafò M, Neville M and Griffiths S (2001). Genetic clues to the molecular basis of tobacco addiction and progress towards personalized therapy. Trends in Molecular Medicine **7**(2): 70-76.

Wang MQ, Fitzhugh EC, Westerfield C and Eddy JM (1995). Family and peer influences on smoking behavior among american adolescents: an age trend. Journal of adolescent health **16**: 200-203.

West P, Sweeting H and Ecob R (1999). Family and friends' influences on the uptake of regular smoking from mid-adolescence to early adulthood. Addiction **94**(9): 1397-1411.

Whiters NJ, Low JL, Holgate ST and Clough JB (2000). Smoking habits in a cohort of U.K. adolescents. Respiratory Medicine **94**: 391-396.

Wilde GJS (1970). Neurotische labiliteit gemeten volgens de vragenlijst methode (the questionnaire method as a means of measuring neurotic instability). Amsterdam, van Rossen.

Willemsen G, Vink JM and Boomsma DI (2003). Assortative mating may explain spouses' risk of same disease. Britisch Medical Journal **326**: 396.

Winger G, Hofmann FG and Woods JH (1992). chapter 2, Tobacco and nicotine. A handbook on drug and alcohol abuse, Oxford University press**:** 22-38.

Yuan B, Vaske D, Weber JL, Beck J and Sheffield VC (1997). Improved set of short-tandem-repeat polymorphisms for screening the human genome. Am. J. Hum. Genet **60**(2): 459-60.

Zuckerman M (1971). Dimensions of sensation seeking. J Consult Clin Psychol **36**: 45-52.

List of publications

Published papers:

Vink JM, Willemsen G, Beem AL and Boomsma DI. The Fagerström Test for Nicotine Dependence in a Dutch sample of daily smokers and ex-smokers. Addictive Behaviors *in press.* (a short version of Chapter 5)

Vink JM, Willemsen G, Stubbe JH, Middeldorp CM, Ligthart RSL, Baas KD, Dirkzwager HJC, Geus de EJC and Boomsma DI. Estimating non-response bias in family studies: application to mental health and lifestyle. European Journal of Epidemiology, 2004, 19(7); 623-630 (Chapter 2)

Vink JM, Beem AL, Posthuma D, Neale MC, Willemsen G, Kendler KS, Slagboom PE, Boomsma DI. Linkage analysis of smoking initiation and quantity in Dutch sibling pairs. The Pharmacogenomics Journal, 2004, 4: 274-282 . (Chapter 8)

Willemsen G, Boomsma DI, Beem AL, **Vink JM,** Slagboom PE and Posthuma D. (2004). QTLS for height: results from a linkage analysis in Dutch sibling pairs. European Journal of Human Genetics 1-9.

Vink JM, Willemsen G, Engels RCME and Boomsma DI. Smoking status of parents, siblings and friends: predictors of regular smoking? Findings from a longitudinal twin-family study. Twin Research 2003; 6(3): 209-217. (Chapter 4)

Vink JM, Willemsen G and Boomsma DI. The association of current smoking behavior with the smoking behavior of parents, siblings, friends and spouses. Addiction 98: 2003; 923-931. (Chapter 3)

Willemsen G, **Vink JM**, and Boomsma DI. Assortative mating may explain spouses' risk of same disease. Britisch Medical Journal; 2003; 326: 396.

Weering van H, Schats R, McDonnell J, **Vink JM**, Vermeiden JPW, and Hompes P. The impact of the embryo transfer catheter on the pregnancy rate in IVF. Human Reproduction, 2002; 17(3): 666-670.

Vink JM and Boomsma DI. Gene finding strategies. Biological Psychology; 2002; 61: 53-71. (Chapter 7)

Boomsma DI, **Vink JM**, Beijsterveldt van CEM, Geus de EJC, Beem AL, Mulder EJCM, Riese H, Willemsen GAHM, Bartels M, Berg van den M, Derks EM, Graaff SC, Kupper HM, Polderman JC, Rietveld MJH, Stubbe JH, Knol LI, Stroet T, and Baal GCM. Netherlands Twin Register: a focus on longitudinal research. Twin Research, 2002; 5(5): 401-406.

Willemsen M, **Vink JM**, and Boomsma DI. Roken en erfelijkheid. Tijdschrift voor Gezondheidswetenschappen, 2001; 79: 79-89.

Vink JM, Groot AS, Kerkhof GA and Boomsma DI. Genetic analysis of morningness and eveningness. Chronobiology International, 2001; 18(5): 809-822.

Vink JM, Boomsma DI, and Willemsen M. De genetische achtergrond van individuele verschillen in rookgedrag. Handboek Verslaving, 2001; juli 2001(E 3300): 1-21.

Boomsma DI, Beem AL, Berg van den M, Dolan CV, Koopmans JR, **Vink JM**, Geus de EJC, and Slagboom PJ. Netherlands Twin Family Study of anxious depression (NETSAD). Twin Research, 2000; 3: 323-334.

Hoek RM, Smit AB, Frings H, **Vink JM**, Jong-Brink de M, and Geraerts WP. A new Ig-superfamily member, molluscan defence molecule (MDM) from Lymnaea stagnalis is down-regulated during parasitosis. Eur J Immunol, 1996; 26: 939-44.

Under review/submitted:

Vink JM, Willemsen G and Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behavior Genetics *under review*. (Chapter 6)

Lynskey M, **Vink JM,** Boomsma DI. Early onset of cannabis use and progression to other drug use in a sample of Dutch twins *(submitted)*.

Published abstracts:

Sadrzadeh S, **Vink JM**, Homburg R, Lambalk CB and Boomsma DI (2004). Pathogenesis of polycystic ovary syndrome due to genetic factors: a twin study. Twin Research **7**(4): 376.

Boomsma DI, **Vink JM**, Willemsen G, Posthuma D, Slagboom PE and Chen X (2004). Genetics of smoking and nicotine dependence. Twin Research **7**(4): 339-340.

Chen X, Che Y, **Vink JM**, Posthuma D, Neale MC, Kendler KS and Boomsma DI (2004). A study of the epac gene for smoking initiation and nicotine dependence with a Dutch twin sample. Behavior Genetics (in press).

Boomsma DI, Beem AL, van den Berg M, Dolan CV, Geus de EJC, Riese H, **Vink JM**, Willemsen AHM and Slagboom PE (2002). Netherlands twin-family study of anxious depression (NETSAD): Genetics of anxious depression in a selected sample of twins and siblings. Am J Med Genetics 2002; 114(7): 067.

Boomsma DI, Derks EM, Heidema G, **Vink JM**, Dirkzwager HC, Hudziak JJ, Verhulst FC, Bartels M, Rietveld MJH and Beijsterveldt van CEM. (2002). Parental, teacher and self-reported aggression in Dutch twin families: Longitudinal studies in young and adult twins. 32th annual meeting of the Behavior Genetic Association. Behavior Genetics 2002; 32(6): 460.

Vink JM and Boomsma DI. Relative risk of smoking parents and smoking siblings on smoking status. 31th annual meeting of the Behavior Genetic Association, Cambridge 2001 Behavior Genetics 2001; 31(5): 472-473

Vink JM and Boomsma DI (2001). The Fagerström Test for Nicotine Dependence (FTND) in a twin-family study. 31st annual meeting of the Behavior Genetic Association, Cambridge. Behavior Genetics 2001; 31(5): 473

Vermeiden JPW, Ceelen M, **Vink JM** and Oudejans CBM (2000). In-situ analysis of 11 beta-hydroxysteroid dehydrogenase type I and II mRNA in granulosa cells from IVF patients: potential use as cytoplasmic maturity marker. Human Reproduction 15(special issue A1): 120.

Dankwoord

Natuurlijk gaan de eerste zinnen van dit dankwoord uit naar alle tweelingen en hun familieleden die hebben meegedaan aan het onderzoek. In totaal hebben 6795 personen de 'vijfde' vragenlijst van 18 pagina's ingevuld. Daarnaast hebben ook nog eens meer dan 1000 mensen de moeite genomen om met behulp van wattenstaafjes een monduitstrijkje te maken en deze terug te sturen naar de VU. Daarvoor mijn hartelijke dank!
Mijn promotor prof. Dr. Dorret Boomsma wil ik heel hartelijk bedanken voor alle tijd, moeite en vertrouwen in mij en mijn proefschrift. Dorret, hoewel jouw opmerkingen, ideeen of suggesties meestal leiden tot meer werk of nieuwe analyses tillen ze een artikel altijd naar een hoger niveau. Bedankt, ik heb ontzettend veel geleerd! Natuurlijk heeft ook mijn co-promotor Dr. Gonneke Willemsen veel bijgedragen. Gonneke, bij jou kon ik altijd even binnenlopen om nieuwe ideeën te spuien, raad te vragen, te klagen of leuke resultaten te laten zien. Dank je wel! Voordat Gonneke de dagelijkse begeleiding op zich nam heeft Dr. Judith Koopmans dat het eerste jaar gedaan. Judith, van jou heb ik de 'geschiedenis' van het onderzoek naar gezondheid en leefgewoonten geleerd. In de jaren die volgden heb ik heel veel gehad aan de dingen die jij mij het eerste jaar uitgelegd hebt!
Juanita, Suzanne en Marijn: jullie hebben echt bergen werk verzet! Jullie hebben stapels post open gemaakt, mensen gebeld, dna-pakjes ingepakt en uitgepakt, bedankkaartjes verstuurd, wijzigingen verwerkt in Trapp, het is teveel om op te noemen. Bedankt voor al jullie hulp, maar ook voor de gezelligheid! Voor het DNA onderzoek (NETSMOK) hadden we extra hulp nodig om buizen te vullen met buffer (7.600), wattenstaafjes in te pakken (ruim 30.000), stickers uit te draaien (14.000), pakketjes te maken en te versturen (2.000), mensen te bellen, retour gekomen pakketjes uit te pakken etcetera: Sharon, Jeanine, Ryanne, Marilig, Danielle, Mena en Cathelijne ontzettend bedankt voor jullie hulp!
Ik wil alle collega's van de afdeling graag bedanken voor de prettige samenwerking en gezelligheid. Er zijn een paar collega's die ik in het bijzonder wil bedanken: Danielle en Caroline voor het beantwoorden van bijna alle mx-vragen, Leo voor alle statistiek-problemen, Therese om te overleggen over de complexe trapp-families, Hanneke voor het uitvoeren van de controles op de databestanden en Alexia voor de gezellige koffiepauzes (lang geleden, maar niet vergeten). Eric bedankt voor het ontwerpen van de kaft. Verder natuurlijk alle mede-aio's, in het bijzonder mijn kamergenoten Mireille en Christel, bedankt voor de gezelligheid en het delen van de dagelijkse AIO-perikelen en Janine, bedankt voor de vriendschap. De studenten Evelien en Laura hebben zich beide bezig gehouden met onderzoek naar verslaving; bedankt voor de leuke samenwerking. Eco, zonder jouw dagelijkse graai in de droppot vergezeld van het nodige gemopper (op het soort dropjes, de studenten van tegenwoordig of het leven in het algemeen) was het schrijven van mijn proefschrift niet hetzelfde geweest! Tenslotte het secretariaat, Alies, Evelien, Christina en natuurlijk Natascha ontzettend bedankt voor alle hulp.
Ook collega's van buiten de VU die betrokken zijn geweest bij mijn onderzoek wil ik graag bedanken. Prof. Dr. Eline Slagboom, Eka Suchiman, Nico Lakenberg en Bernd Brandt in Leiden, Prof. Dr. Rutger Engels in Nijmegen en Dr. Marc Willemsen van

Stivoro in Den Haag, allemaal hartelijk dank voor jullie bijdrage. I would like to thank Prof. Dr. Ken Kendler, Prof. Dr. Mike Neale, Dr. Sam Chen and their colleagues from the Common Wealth University of Richmond for their valuable contributions. I would also like to thank the reading committee for their time and effort.
Voordat ik als AIO ging werken bij de afdeling Biologische Psychologie, werkte ik bij het IVF-centrum van het VU-ziekenhuis. Mijn toenmalige baas Jan Vermeiden is degene die mij heeft begeleid bij mijn eerste stapjes op het pad der wetenschap. Jan, dank zij jouw enthousiasme en vooral eeuwige optimisme ben ik de wetenschap echt leuk gaan vinden!
Gelukkig is er nog een leven naast 'de wetenschap': Petrina, Luc, Marianne, Cindy, Marijn, Jeanine, Ryanne, Sharon, Sabine en Renata, allemaal op jullie eigen manier een bijzondere vriendschap, bedankt! Ook popkoor Together wil ik bedanken voor de gezellige en muzikale vrijdagavonden waar ik door saxofoon te spelen 'de wetenschap' even helemaal kon vergeten! En tenslotte de 'IVF-meiden': Lilian, Melanie, Debbie en Kayleen: dat er nog maar vele gezellige etentjes gaan komen!!
En als laatste natuurlijk mijn familie: Mam, toen ik 6 jaar oud was schreef je het volgende rijmpje in mijn poezie-album: *"Je was ons lieve schattekind, van pappa en mij, nu ik hier in jouw boekje schrijf is pappa niet meer daarbij. Maar ik hoop en pappa zou het ook hebben gezegd, dat voor jou in je leven nog veel goeds is weggelegd"*. Lieve mam, het 'goeds' wat ik heb bereikt, heb ik grotendeels te danken aan alles wat jij mij hebt meegegeven in mijn leven en daarvoor wil ik je in dit 'boekje' bedanken! Mijn andere familieleden: Loes (tevens paranimf) en Thijs, Kees en Siny, Tiddo, Oom Jan en natuurlijk de 'kids', bedankt voor jullie belangstelling! Tenslotte, Fia (ook paranimf), wat heb jij vaak verhalen moeten aanhoren over analyses die ik aan het doen was of artikelen die ik aan het schrijven was, waar ik in mijn enthousiasme graag over wilde vertellen. Bedankt dat je er altijd voor me was en bent!

Jacqueline