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 decrease 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.