

3 400 words

2 tables & 2 figures

mega_smoking_twins

A Genetic Epidemiological Mega Analysis of Smoking Initiation in Adolescents.

Hermine H Maes^{1,2,3,4,*} PhD, Elizabeth Prom-Wormley^{1,5}, PhD, Lindon J Eaves^{1,2}, PhD,
Soo Hyun Rhee⁶, PhD, John K Hewitt⁶, PhD, Susan Young^{6,16}, PhD, Robin Corley⁶, PhD,
Matt McGue⁷, PhD, William G Iacono⁷, PhD, Lisa Legrand⁷, PhD, Di Samek^{7,17}, PhD,
E Lenn Murrelle⁸, PhD,
Judy L Silberg^{1,2}, PhD, Donna Miles⁹, PhD,
Richard M Schieken¹⁰, MD,
Gaston P Beunen⁴, PhD, Martine Thomis⁴, PhD,
Richard J Rose¹¹, PhD, Danielle M Dick², PhD,
Dorret I Boomsma¹², PhD, Meike Bartels¹², PhD, Jacqueline M Vink¹², PhD,
Paul Lichtenstein¹³, PhD,
Victoria White¹⁴, PhD,
Jaakko Kaprio¹⁵, MD, PhD, Michael C Neale^{1,2}, PhD

05/22/2016

¹ Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond VA

² Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond VA

- 3 Massey Cancer Center, Virginia Commonwealth University, Richmond VA
- 4 Department of Kinesiology, KU Leuven, Leuven Belgium
- 5 Division of Epidemiology, Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond VA
- 6 Institute for Behavioral Genetics, University of Colorado, Boulder CO
- 7 Department of Psychology, University of Minnesota, Minneapolis MN
- 8 ELM International Research Consultants, LLC, Richmond VA
- 9 Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill NC
- 10 Department of Pediatric Cardiology, Virginia Commonwealth University, Richmond VA
- 11 Department of Psychological and Brain Sciences, Indiana University, Bloomington IN
- 12 Department of Biological Psychology, VU University, Amsterdam the Netherlands
- 13 Karolinska Institutet, Stockholm, Sweden
- 14 Cancer Council Victoria, Melbourne Australia
- 15 University of Helsinki, Hjelt Institute, Dept of Public Health & Institute for Molecular Medicine (FIMM); & National Institute for Health and Welfare, Dept of Mental Health and Substance Abuse Services, Helsinki, Finland
- 16 Department of Psychiatry, University of Colorado, Denver CO
- 17 Human Development and Family Studies Department, Auburn University, Auburn AL

Address correspondence to Dr H H Maes, Virginia Institute for Psychiatric and Behavioral Genetics, PO Box 980003, Richmond, VA, 23298-0003, USA. Voice: +804 828 8145. Fax: 804 828 8801. E-mail: hmaes@vcu.edu

Running Head: Genetic Epidemiology of Smoking

Keywords: smoking, twin studies, adolescence

Submitted to: Nicotine & Tobacco Research

Accepted Manuscript

ABSTRACT

Introduction. Previous studies in adolescents were not adequately powered to accurately disentangle genetic and environmental influences on smoking initiation across adolescence.

Methods. Mega-analysis of pooled genetically informative data on smoking initiation was performed, with structural equation modeling, to test equality of prevalence and correlations across cultural backgrounds, and to estimate the significance and effect size of genetic and environmental effects according to the classical twin study, in adolescent male and female twins from same-sex and opposite-sex twin pairs (N=19 313 pairs) between age 10 and 19, with 76 358 longitudinal assessments between 1983 and 2007, from 11 population-based twin samples from the US, Europe and Australia

Results. Although prevalences differed between samples, twin correlations did not, suggesting similar etiology of smoking initiation across developed countries. The estimate of additive genetic contributions to liability of smoking initiation increased from approximately 15% to 45% from age 13 to 19. Correspondingly, shared environmental factors accounted for a substantial proportion of variance in liability to smoking initiation at age 13 (70%) and gradually less by age 19 (40%).

Conclusions. Both additive genetic and shared environmental factors significantly contribute to variance in smoking initiation throughout adolescence. The present study, the largest genetic epidemiological study on smoking initiation to date, found consistent results across 11 studies for the etiology of smoking initiation. Environmental factors, especially those shared by siblings in a family, primarily influence smoking initiation variance in early adolescence, while an increasing role of genetic factors is seen at later ages, which has important implications for prevention strategies.

IMPLICATIONS

This is the first study to find evidence of genetic factors in liability to smoking initiation at ages as young as 12. It also shows the strongest evidence to date for decay of effects of the shared environment from early adolescence to young adulthood. We found remarkable consistency of twin correlations across studies reflecting similar etiology of liability to initiate smoking across different cultures and time periods. Thus familial factors strongly contribute to individual differences in who starts to smoke with a gradual increase in the impact of genetic factors and a corresponding decrease in that of the shared environment.

INTRODUCTION

Smoking remains a serious public health problem. Briefly, tobacco smoking is associated with increased morbidity, mortality, personal and public cost, even after 50 years since the first Surgeon General report ¹. Tobacco kills nearly six million people each year, of whom more than five million are users and ex-users and more than 600 000 are non-smokers exposed to second-hand smoke ². In the US, smoking is responsible for 69% and 74% of all cancer deaths and 69% and 61% of deaths from cardiovascular disease in female and male smokers, respectively ³. Up to half of current users will eventually die of a tobacco-related cause ².

According to the Surgeon General's report on 'Preventing Tobacco Use Among Youth and Young Adults', "...evidence is suggestive that tobacco use is a heritable trait, more so for regular use than for onset". The expression of genetic risk for smoking among young people may be moderated by small group and larger social-environmental factors ⁴. The editorial from JAMA in 1964 ⁵ seems like it could have been written today. Its statements "Why some teenagers smoke and others do not is not fully understood" and "...reduction or elimination of cigarette smoking can only be achieved if today's nonsmokers never start" remain true. Given tobacco use and addiction ⁶ — which can occur quickly with smoking as few as 100 cigarettes ⁴ — almost always begins before age 18, efforts must be directed toward adolescents and even younger children. Even though substantial reductions in smoking rates have occurred in some countries ⁷, the number of smokers worldwide is still increasing. That the largest reduction in daily smoking between 1980 and 2012 was among 15-19 year-olds ⁸ is encouraging but gains are still modest.

Twin studies have consistently found a significant genetic component to the liability to smoking initiation (SI) and nicotine dependence (ND)⁹. Recent reviews^{10,11} show heritability estimates for SI from 40 to 70% with family environmental influences more pronounced in adolescence than in adulthood¹²⁻²⁰. Furthermore, there is evidence for overlapping genetic and environmental risk factors between SI and ND in adults^{21,22} suggesting that partly the same genes contribute to liability to SI and ND. This evidence for a correlated liability of SI and ND makes it more important to study SI as a necessary stage to nicotine addiction^{23,24}.

Previous studies of SI in adolescence have been unable to accurately assess the role of shared environmental factors in the development of smoking behavior for several reasons. First, most studies were underpowered for estimating shared environmental influences in the presence of genetic factors²⁵. Thus while evidence for familial resemblance may be strong, sample sizes are often too small to distinguish between shared environmental and genetic factors. Second, given the need for large samples for genetic studies of binary traits, data from different ages and cohorts are often combined, which can overestimate contributions of shared environmental factors for traits correlated with age^{23,26}. This problem is exacerbated by low prevalence of SI in early adolescence, reducing power of individual studies. We use prevalence here to refer to lifetime prevalence of having initiated smoking.

In this report, we attempt to address these concerns by performing a mega-analysis by pooling data from available adolescent prospective longitudinal twin studies with data on smoking initiation. Substantial sample sizes are available for all ages throughout adolescence which allows, for the first time, familial resemblance of SI to be separated into genetic and shared environmental factors. Our aims are to: i) estimate prevalence of SI across adolescence and test heterogeneity across samples; ii) estimate twin correlations for SI and test their equivalence across samples by age; iii) estimate contributions of genetic, shared environmental and specific

environmental factors to liability of SI at every age across adolescence; and iv) test for sex differences.

METHODS

Subjects

Table 1 shows the main characteristics of the participating eleven studies (see online material). We approached investigators who had published results on the genetic epidemiology of SI in adolescence by the start of the project, all of whom agreed to share individual anonymized data with us. We also had access to local samples with SI data, and a publicly available nationally representative sample. Samples are organized by continent, starting in North America, followed by Europe and Australia, alphabetically by abbreviation. Participating studies were approved by their respective human subjects protection committees. Inclusion criteria were availability of population-based adolescent twin data on smoking.

Measures

Data were collected via questionnaire or personal interview. For the purposes of this report, we focused on ever use of tobacco, thus including those who have experimented with tobacco by trying just one or a few cigarettes. We use the term 'smoking initiation' as it has been widely used in genetic epidemiologic studies of smoking behavior. *Smoking initiation* (SI) was coded 0/1 and defined according to responses to questions like "Have you ever smoked cigarettes or tried any form of tobacco?". Exact wording of questions, and coding of answers is presented in Appendix 1. Two studies (CVT & LLTS) asked 'Have you ever smoked at least 100 cigarettes in your life?' thus requiring a higher threshold for SI.

Statistical Approach

Structural equation models were fit to the twin data, in order to estimate the proportions of variance of additive genetic (A), shared (C) and unique (E) environmental factors contributing to individual differences in liability to SI, using the statistical package OpenMx^{27,28}. In brief, greater similarity of monozygotic (MZ) than dizygotic (DZ) twins implicates genetic factors, whereas DZ similarity greater than half that of MZ suggests shared environment²⁹ (see online material for further detail).

RESULTS

Descriptive Statistics

Data from eleven samples were analyzed for SI at each age from 10 to 19. Sample sizes are provided by sample, age and zygosity in Table 2. Prevalence rates for SI and their standard errors are shown in Figure 1 by age, sex and sample, with a mean estimate across all samples. Few children (<1%) had initiated smoking by age 10. From age 11 onwards SI rates increased almost linearly to age 19, by which time ~60% of adolescents/young adults have smoked at least one cigarette. There was considerable variability across samples at each age; rates were consistently higher in AYATS compared than all others, and lower in CVT and LLTS. Variability likely stems from differences in cultural background or in wording between assessments. We formally tested equality of prevalences by twin order, zygosity, sex of cotwin, sex and sample using structural equation modeling.

Model Assumptions Testing

Models were fitted by maximum likelihood to the combined SI data, separately at each age from 12 to 19 years, allowing prevalence of SI to differ by twin order, zygosity, sex of co-twin, sex and sample, and correlations to differ by zygosity and sample (online eTable 1). Prevalences of SI

at ages 10-11 were too low to permit meaningful analysis. Model assumptions, including equality of prevalences across twin order (model T1), zygosity (model T2) and sex of co-twin (model T3) were met across all samples at each age. However, prevalences for males were higher than those for females at all ages between 12 and 19, and significantly so at ages 16, 18 and 19 (model T4). Furthermore, prevalences could not be constrained to be equal across samples at any age (model T5).

We further tested whether twin correlations could be equated across all eleven samples. Tests were performed separately for monozygotic (MZ, model C1) and dizygotic (DZ, model C2) twins, for all same sex twins (model C3) and including opposite sex twins (model C4). Correlations could be equated across samples at most ages, except for MZ correlations at age 16 and DZ correlations at age 14 (eTable 1). When applying Bonferroni or false discovery rate corrections, no equality tests were significant, except at age 16, which was borderline significant. Testing equality of correlations and prevalences across datasets simultaneously (models CT1 & CT2) suggested that twin correlations could be equated across samples, when allowing for differences in prevalence by sex and sample. Twin correlations by zygosity for each sample are presented in eFigures 1a&b, along with joint estimates after equating correlations across samples, for ages 12-19 (see eFigures 2a&b for correlations by sample).

MZ correlations were consistently high during adolescence. Like-sex DZ correlations varied more and gradually decreased towards young adulthood. Opposite-sex DZ correlations showed a more pronounced decreasing trend. This pattern of correlations over adolescence is broadly consistent with a changing role of sources of SI familial resemblance from shared environmental to genetic factors. The nature and/or magnitude of these effects possibly differ by sex.

Genetic Analyses

Twin models were fitted to data from all samples at ages 12-19. DZ twin correlations for SI were invariably greater than half those of MZ twins, consistent with A, C and E factors contributing to individual differences in liability to SI. Based on results from testing model assumptions, prevalences for SI were equated across twin order, zygosity and sex of co-twin, but allowed to differ by sex and sample. Models were tested for heterogeneity across samples by equating variance components (ACE), genetic correlations (rg) and shared environmental correlations (rc) parameters (online eTable 2). Parameters could be equated across all samples at all ages, as indicated by more parsimonious fits of D1-D4 models compared to the corresponding S1-S4 models.

We fitted alternative models testing whether different genetic or different shared environmental factors contributed in males and females, and whether the magnitude of ACE contributions was the same across sexes. At ages 12, 13, 15 and 19, neither type of sex difference was significant. At remaining ages (14, 16, 17 and 18), models including different proportions and different types of genetic and shared environmental factors in males and females - by estimating correlations between them (rc) across sex - performed better suggesting sex differences in etiology of liability to SI at later ages. Neither genetic (model D5) nor shared environmental (model D6) parameters could be dropped from models at any age.

We present results for models with separate parameters for males and females. Estimates and confidence intervals for A, C, E and rc from the best fitting models, with parameters constrained across samples, are presented for ages 12-19 in Figure 2. Results show an increase in proportion of liability to SI explained by additive genetic factors from 15% to 45%, and a corresponding decrease in proportion of variance accounted for by shared environmental factors, with unique environmental factors explaining a small stable amount of variance in

liability to SI. Gender differences in factors influencing SI increased across adolescence: the shared environmental correlation between males and females decreases from 1.0 to about 0.6 from age 12–19, suggesting that in early adolescence environmental factors that increase similarity between twins are mostly the same in males and females, whereas in later adolescence, only about half are. Although the magnitude of genetic and environmental factors was allowed to differ by sex, the trend of increasing contributions of genetic factors and decreasing contributions of shared environmental factors was observed for both.

DISCUSSION

We set out to combine smoking initiation data across eleven primarily longitudinal twin datasets to obtain stable and reliable estimates of additive genetic, shared and unique environmental contributions to liability of smoking initiation across adolescence. Previous studies on individual samples and review articles indicated a trend towards increasing contributions of additive genetic factors and decreasing contributions of shared environmental factors from early to late adolescence and into young adulthood^{13,15-17,19,20,30-33}. However, no single study was large enough to estimate genetic and shared environmental variance components accurately at every age across adolescence, especially at younger ages where prevalence of smoking is relatively low which limits statistical power. We successfully combined data from eleven samples, primarily of European descent, comprising five samples from the US, four European and one Australian, resulting in sample sizes ranging from ~4 000 to >18 000 individuals at each age from 12 to 19 years.

We draw five main conclusions. First, smoking prevalence increased rapidly and almost linearly between ages 10 and 19 from zero to >60% of adolescents having tried smoking cigarettes. Furthermore there appeared to be substantial consistency in longitudinal trends as well as variability within each age in smoking prevalence by sample. There were significant sex

differences in prevalence, however, with higher rates in males than in females, consistent with epidemiological literature on smoking in adolescents from European, North-American and global surveys³⁴⁻³⁷.

Second, model fitting confirmed that prevalence of SI could not be equated across samples.

This could reflect differences in assessment of SI, availability, access and attitudes towards smoking across cultures and variation in stage of the tobacco epidemic³⁷. Assumptions of the classical twin method were met in that prevalences could be equated across twin order, zygosity and sex of co-twin. Differences in prevalence by zygosity could be interpreted as sibling cooperation or sibling competition³⁸. Our results suggest little role for sibling interaction, because prevalences were similar across zygosity. Prevalence of SI did not differ significantly if the twin was the same or opposite sex of their twin. While we did not have singletons in the analysis, these analyses indicate the absence of twin-specific effects on SI and support the generalization of our results to the populations from which they were drawn³⁹.

Third, even though prevalences could not be equated across samples, twin correlations could. This suggests that etiology of liability to initiate smoking is broadly consistent across samples — of primarily European descent — collected on three continents. To our knowledge, this is the first report to show this notable similarity in twin correlations across cultures for smoking in adolescence. It suggests that similar etiological factors operate within families in high-income countries of mostly European ancestry. The implication is that preventive measures found to be effective in one such country would likely work well in others. Although twin correlations, and thus heritability of smoking initiation, could be equated across samples collected on different continents, this does not imply that social factors at the school, state, population composition level do not moderate aspects of smoking behavior⁴⁰⁻⁴². However, this type of genotype by environment effect appeared stronger for regular smoking than for initiation. Furthermore,

changing policies across time, such as the Surgeon General's Report about the effects of smoking on health, has also been shown to affect the magnitude of genetic influences on regular smoking⁴³. As the majority of the samples in the current report were collected in the nineties, we did not take genotype by cohort effects into account. Future analyses, however, should investigate whether macro and micro environmental factors cause variation in liability to smoking initiation.

Fourth, at each age across adolescence both additive genetic and shared environmental factors contribute significantly to variance in liability to SI. Consistent with prior literature on SI from early to late adolescence, the influence of additive genetic factors appears to increase while that of the shared environment decreases^{10,31}. Although at first sight these results may not appear novel, the current study is the first to be sufficiently powered to detect a contribution of additive genetic factors of 20% of the variance in early adolescence, and a modest contribution of shared environmental factors in later adolescence, thus stressing the importance of both sources throughout adolescence. While assortative mating could mimic effects of shared environmental influences, and given significant spousal correlations for smoking behavior¹³, it seems unlikely that assortment would account for the shared environment found here. Assuming a marital correlation for SI of .20, shared environmental contributions would be overestimated by 2% and 15% for MZ correlations of .85 and DZ correlations of .50 and .70 respectively. Correspondingly, heritability would be underestimated.

Fifth, while sex differences in genetic and environmental factors were non-significant in early adolescence, they were post-puberty, suggesting that to some degree environmental factors that contribute to SI liability in males differ from those in females. Models allowing for different shared environmental factors across sex fitted marginally better than those allowing different

genes across sex, although power is limited in comparing these alternative explanations, even with current sample sizes. While it seems plausible that girls are exposed to partially different environmental factors than boys, we cannot exclude sex-specific genetic factors. A possible source of such genetic differences between boys and girls across adolescence would be differential rates of maturation, which is known to be partly influenced by genetic factors⁴⁴.

Results from this study increase the evidence beyond suggestive⁴ that smoking behavior, and in particular SI in adolescence, is a heritable trait. However, genome-wide studies of smoking behavior have only identified some genes underlying SI in contrast to major findings underlying variability in consumption of cigarettes⁴⁵⁻⁴⁸. The strongest association for SI was reported for SNPs in the BDNF gene⁽⁴⁶⁾. A handful of other SNPs have been found to be genome-wide significant for SI, but require replication^(49, 50). Given the broad age range of individuals in the large scale consortia, selective attrition by genotype due to smoking related mortality could have obscured signals for SI⁽⁵¹⁾. Furthermore, evidence for shared environmental contributions, especially in early adolescence, is strong. Even though a recent review of behavioral genetics research suggests that 'most environmental effects are not shared by children growing up in the same family'⁽⁵²⁾, evidence is accumulating that shared environmental factors contribute significantly to behavior in early adolescence, and especially for externalizing behaviors and substance use^(53, 54, 55). These shared environmental effects may result from: parents and older siblings (including second-hand smoke and effects of assortative mating); peers^{56,57}; or social environment factors such as advertising controls, tobacco pricing, smoke-free regulations and tobacco availability⁽⁵⁸⁾.

Heritability of SI is significant even at young ages and jumps at ages 14-15, the transition from middle to high school in the US, which therefore appears to be a critical period to target

prevention measures. Most individual differences in SI in adolescence are accounted for by familial factors, although they shift from being more shared environmental to more genetic. It would seem justified to target prevention at the whole family, rather than solely at teenagers⁵⁹. These results are consistent with a review of the effectiveness of family-based interventions to prevent children and adolescent from starting to smoke⁶⁰, which report a moderately positive effect of high intensity programs that address family functioning. Furthermore, office-based interventions by pediatric providers engaged in delivering prevention and cessation counseling to both patients and parents/caregivers show great promise⁶¹. We also believe that providing personalized information - including genetic information - about smoking risks could improve smoking prevention⁶². Finally, prevention efforts might be especially effective if targeted at children with both high genetic and environmental risk, as they are at greatest risk of nicotine addiction. Our results suggest an increase in risk for children having 1 or 2 parents who smoke which could be considered when evaluating the cost-benefit ratio of targeted (families, high-risk children) versus whole population intervention campaigns.

In summary, this study showed that even though substantial differences exist in prevalence of SI across samples, etiology of SI liability is markedly similar across different populations of Western European descent.

Limitations

This study should be interpreted in the context of four potential limitations. First, items used to query participants about their SI differed across studies. While most samples included questions about lifetime SI, two samples only recorded SI when participants had smoked at least 100 cigarettes. This difference likely accounts for differences in prevalence across studies. A re-analysis excluding these samples showed almost identical results; the largest difference was <.1 change in A and C estimates at ages 12-13 (results available upon request).

Second, results may not generalize to the entire US population as over 95% of participants were of European ancestry. Results from analyses limited to European ancestry twins were very similar to those that included twins of other ethnicities (results available upon request), thus strengthening generalizability of results. Furthermore, available samples are from high-income countries, so results may not generalize to low- and middle-income countries. Third, sample sizes at different ages differed, which affected power to detect certain effects, including sex differences at younger ages. Fourth, the current study only included data on twins, thus limiting to three the sources of variance to be estimated. Future modeling including other relatives such as parents and siblings would allow estimation of the effects of assortative mating, parent-child environmental transmission, and the action and interaction of additional types of genetic and environmental factors.

ACKNOWLEDGEMENTS

We thank Dr Amir Toor for his helpful comments on the manuscript.

FUNDING

Data analysis supported by the National Institutes of Health [DA025109, DA024304 and DA018673].

Add Health: This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J Richard Udry, Peter S Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant [P01-HD31921] from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due to Ronald R Rindfuss and Barbara Entwisle for assistance in the original design.

Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

LTS-CTS: Data collection was supported by [DA011015].

MFTS: Data collection was supported by [DA05147 (male MFTS twins), AA09367 (female MFTS twins), DA013240 (ES twins) and DA036216].

MATR: MASATS: Data collection was supported by the Templeton Foundation and the Keck Foundation.

VTSABD-YAFU-TSA: Data collection was supported by [MH045268, MH068521, DA024413, and Virginia Tobacco Settlement Foundation grant 8520012].

CVT: Data collection was supported by [HL031010].

EFPTS: LLTS: Data collection was supported by Research Fund K.U.Leuven [OT/86/80], the National Bank of Belgium, the Fund for Medical Research (Belgium) [3.0038.82, 3.0008.90, 3.0098.91] and [NATO 860823].

FTC: Data collection in the Finnish twin cohorts was supported by National Institute of Alcohol Abuse and Alcoholism [AA-08315, AA-12502, AA-00145, and AA-09203 to R J Rose], the Academy of Finland Center of Excellence in Complex Disease Genetics [213506, 129680 to JKaprio], and the Academy of Finland [100499, 205585, 118555, 141054, 265240, 263278 and 264146 to JKaprio].

NTR: Data collection was supported by Netherlands Organization for Scientific Research (NWO) and [MagW/ZonMW: 904-61-090, 985-10-002, ZonMw 912-10-020, ZonMw 904-61-193, 480-04-004, 463-06-001, 451-04-034, 400-05-717, ZonMw Addiction-31160008, ZonMw Middelgroot 911-09-032], Spinozapremie [SPI 56-464-14192]; BBMRI–NL [184.021.007]: Biobanking and Biomolecular Resources Research Infrastructure and from European Science Council (ERC) Genetics of Mental Illness [230374] and Beyond the genetics of addiction [284167]. VU University: Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA).

STR: TCHAD: Data collection supported by the Swedish Council for Working Life and Social Research [2004-0383] and the Swedish Research Council [2004-1415].

ATR: AYATS: Data collection in the Australia twin cohorts was supported by National Health and Medical Research Council grants, Victorian Health Promotion Foundation, and the NHMRC Australian Twin Registry. Prof J Hopper established the cohort and received the funding grants.

Competing interests: none. All authors reviewed and approved the final manuscript before its submission. J Kaprio has consulted for Pfizer on nicotine dependence in 2012-2014.

REFERENCES

1. USDHHS. The health consequences of smoking-50 years of progress. A Report of the Surgeon General. 2014.
2. WHO. Tobacco or Health: A Global Status Report. *World Health Organization Geneva*. 2012.
3. Jha P, Ramasundarahettige C, Landsman V, et al. 21st-century hazards of smoking and benefits of cessation in the United States. *N Engl J Med*. 2013;368(4):341-350.
4. USDHHS. Preventing tobacco use among youth and young adults. A Report of the Surgeon General. 2012(Reports of the Surgeon General.).
5. Childish habit. *JAMA*. 2014;311(2):205.
6. USDHHS. Preventing tobacco use among young people. A report of the Surgeon General. Executive summary. *MMWR Recomm Rep*. 1994;43(RR-4):1-10.
7. Ng M, Freeman MK, Fleming TD, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. *JAMA*. 2014;311(2):183-192.
8. Vink JM, Boomsma DI. Interplay between heritability of smoking and environmental conditions? A comparison of two birth cohorts. *BMC Public Health*. 2011;11:316.
9. Sullivan PF, Kendler KS. The genetic epidemiology of smoking. *Nicotine Tob Res*. 1999;1 Suppl 2:S51-57; discussion S69-70.
10. Maes HH, Neale MC. Genetic modeling of tobacco use behavior and trajectories. In: Swan GE, Baker TB, Chassin L, et al., eds. *NCI Tobacco Control Monograph Series 20: Phenotypes and Endophenotypes: Foundations for Genetic Studies of Nicotine Use and Dependence*: US Department of Health and Human Services, National Institutes of Health; 2009.
11. Rose RJ, Broms U, Korhonen T, Dick DM, Kaprio J. Genetics of Smoking Behavior. In: Kim Y, ed. *Handbook of Behavior Genetics*. New York: Springer; 2009:411-432.
12. Langinvainio H, Kaprio J, Koskenvuo M, Tarkkonen L. Structural analysis of smoking, alcohol use, and personality factors in MZ and DZ twin pair relationships. *Prog Clin Biol Res*. 1981;69 Pt C:23-35.
13. Boomsma DI, Koopmans JR, Van Doornen LJ, Orlebeke JF. Genetic and social influences on starting to smoke: a study of Dutch adolescent twins and their parents. *Addiction*. 1994;89(2):219-226.
14. Koopmans JR, Slutske WS, Heath AC, Neale MC, Boomsma DI. The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav Genet*. 1999;29(6):383-393.
15. Maes HH, Woodard CE, Murrelle L, et al. Tobacco, alcohol and drug use in eight- to sixteen-year-old twins: the Virginia Twin Study of Adolescent Behavioral Development. *J Stud Alcohol*. 1999;60(3):293-305.
16. Han C, McGue MK, Iacono WG. Lifetime tobacco, alcohol and other substance use in adolescent Minnesota twins: univariate and multivariate behavioral genetic analyses. *Addiction*. 1999;94(7):981-993.
17. McGue M, Elkins I, Iacono WG. Genetic and environmental influences on adolescent substance use and abuse. *Am J Med Genet*. 2000;96(5):671-677.
18. Hopfer CJ, Stallings MC, Hewitt JK. Common genetic and environmental vulnerability for alcohol and tobacco use in a volunteer sample of older female twins. *J Stud Alcohol*. 2001;62(6):717-723.
19. White VM, Hopper JL, Wearing AJ, Hill DJ. The role of genes in tobacco smoking during adolescence and young adulthood: a multivariate behaviour genetic investigation. *Addiction*. 2003;98(8):1087-1100.

20. Rhee SH, Hewitt JK, Young SE, et al. Genetic and environmental influences on substance initiation, use, and problem use in adolescents. *Arch Gen Psychiatry*. 2003;60(12):1256-1264.
21. Maes HH, Sullivan PF, Bulik CM, et al. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol Med*. 2004;34(7):1251-1261.
22. Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. *Behav Genet*. 2005;35(4):397-406.
23. Neale MC, Harvey E, Maes HH, Sullivan PF, Kendler KS. Extensions to the modeling of initiation and progression: applications to substance use and abuse. *Behav Genet*. 2006;36(4):507-524.
24. Neale MC, Roysamb E, Jacobson K. Multivariate genetic analysis of sex limitation and G x E interaction. *Twin Res Hum Genet*. 2006;9(4):481-489.
25. Sullivan PF, Eaves LJ. Evaluation of analyses of univariate discrete twin data. *Behav Genet*. 2002;32(3):221-227.
26. Neale MC, Martin NG. The effects of age, sex, and genotype on self-report drunkenness following a challenge dose of alcohol. *Behav Genet*. 1989;19(1):63-78.
27. Boker S, Neale M, Maes H, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika*. 2011;76(2):306-317.
28. *Mx: Statistical Modelling (6th ed)* [computer program]. Richmond, VA: Department of Psychiatry; 2006.
29. Neale MC, Cardon LR. *Methodology for genetic studies of twins and families*. Dordrecht, The Netherlands: Kluwer Academic Publishers BV; 1992.
30. Kaprio J, Hammar N, Koskenvuo M, et al. Cigarette smoking and alcohol use in Finland and Sweden: a cross-national twin study. *Int J Epidemiol*. 1982;11(4):378-386.
31. Hopfer CJ, Crowley TJ, Hewitt JK. Review of twin and adoption studies of adolescent substance use. *J Am Acad Child Adolesc Psychiatry*. 2003;42(6):710-719.
32. Kaprio J. Genetic epidemiology of smoking behavior and nicotine dependence. *COPD*. 2009;6(4):304-306.
33. Baker JH, Maes HH, Larsson H, Lichtenstein P, Kendler KS. Sex differences and developmental stability in genetic and environmental influences on psychoactive substance consumption from early adolescence to young adulthood. *Psychol Med*. 2011;41(9):1907-1916.
34. Harrell JS, Bangdiwala SI, Deng S, Webb JP, Bradley C. Smoking initiation in youth: the roles of gender, race, socioeconomic, and developmental status. *J Adolesc Health*. 1998;23(5):271-279.
35. WHO Europe. The European Tobacco Control Report 2007. 2007.
36. Okoli C, Greaves L, Fagyas V. Sex differences in smoking initiation among children and adolescents. *Public Health*. 2013;127(1):3-10.
37. Lopez AD, Collishaw NE, Piha T. A descriptive model of the tobacco epidemic in developed countries. *Tobacco Control*. 1994;3:242-247.
38. Eaves L. A model for sibling effects in man. *Heredity (Edinb)*. 1976;36(2):205-214.
39. Andrew T, Hart DJ, Snieder H, et al. Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Res*. 2001;4(6):464-477.
40. Boardman JD. State-level moderation of genetic tendencies to smoke. *Am J Public Health*. 2009;99(3):480-486.
41. Boardman JD, Blalock CL, Pampel FC, et al. Population composition, public policy, and the genetics of smoking. *Demography*. 2011;48(4):1517-1533.
42. Boardman JD, Saint Onge JM, Haberstick BC, Timberlake DS, Hewitt JK. Do schools moderate the genetic determinants of smoking? *Behav Genet*. 2008;38(3):234-246.

43. Boardman JD, Blalock CL, Pampel FC. Trends in the genetic influences on smoking. *J Health Soc Behav.* 2010;51(1):108-123.
44. Eaves L, Silberg J, Foley D, et al. Genetic and environmental influences on the relative timing of pubertal change. *Twin Res.* 2004;7(5):471-481.
45. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet.* 2010;42(5):436-440.
46. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010;42(5):441-447.
47. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al. Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet.* 2010;42(5):448-453.
48. Loukola A, Hallfors J, Korhonen T, Kaprio J. Genetics and smoking. *Curr Addict Rep.* 2014;1(1):75-82.
49. Vink JM, Smit AB, de Geus EJ, et al. Genome-wide association study of smoking initiation and current smoking. *Am J Hum Genet.* 2009;84(3):367-379.
50. Yoon D, Kim YJ, Cui WY, et al. Large-scale genome-wide association study of Asian population reveals genetic factors in FRMD4A and other loci influencing smoking initiation and nicotine dependence. *Hum Genet.* 2012;131(6):1009-1021.
51. Taylor AE, Morris RW, Fluharty ME, et al. Stratification by smoking status reveals an association of CHRNA5-A3-B4 genotype with body mass index in never smokers. *PLoS Genet.* 2014;10(12):e1004799.
52. Plomin R, DeFries JC, Knopik VS, Neiderhiser JM. Top 10 Replicated Findings From Behavioral Genetics. *Perspect Psychol Sci.* 2016;11(1):3-23.
53. Buchanan JP, McGue M, Keyes M, Iacono WG. Are there shared environmental influences on adolescent behavior? Evidence from a study of adoptive siblings. *Behav Genet.* 2009;39(5):532-540.
54. Burt SA. Rethinking environmental contributions to child and adolescent psychopathology: a meta-analysis of shared environmental influences. *Psychol Bull.* 2009;135(4):608-637.
55. Korhonen T, Latvala A, Dick DM, et al. Genetic and environmental influences underlying externalizing behaviors, cigarette smoking and illicit drug use across adolescence. *Behav Genet.* 2012;42(4):614-625.
56. Dick DM, Rose RJ, Viken RJ, Kaprio J. Pubertal timing and substance use: associations between and within families across late adolescence. *Dev Psychol.* 2000;36(2):180-189.
57. Rose RJ, Viken RJ, Dick DM, et al. It does take a village: nonfamilial environments and children's behavior. *Psychol Sci.* 2003;14(3):273-277.
58. Do E, Maes H. Narrative review of genes, environment, and cigarettes. *Ann Med.* 2016:1-15.
59. Hudziak JJ, Bartels M. Genetic and environmental influences on wellness, resilience, and psychopathology: A family-based approach for promotion, prevention and intervention. In: Hudziak JJ, ed. *Developmental Psychopathology and Wellness.* Washington DC: American Psychiatric Publishing, Inc.; 2008:267-286.
60. Thomas RE, Baker PR, Thomas BC, Lorenzetti DL. Family-based programmes for preventing smoking by children and adolescents. *Cochrane Database Syst Rev.* 2015;2:CD004493.
61. Pbert L, Farber H, Horn K, et al. State-of-the-art office-based interventions to eliminate youth tobacco use: the past decade. *Pediatrics.* 2015;135(4):734-747.
62. Vandenberg DJ, Schlomer GL, Cleveland HH, et al. An Adolescent Substance Prevention Model Blocks the Effect of CHRNA5 Genotype on Smoking During High School. *Nicotine Tob Res.* 2015.

Figure 1: Prevalence rates of SI by sex (males in boxes, females in circles), age and sample

Figure 2: Estimates of proportions of variance of additive genetic (A estimates), shared environmental (C estimates), specific environmental (E estimates) factors, and of the correlation between male and female shared environmental factors (R_c estimates) to liability of SI by sex and age

Accepted Manuscript

Table 1: Years of assessments, ages of participants, number of data collection waves, number of unique individuals and geographical location of participating studies

	years	ages	wave	N#	location	&
Add Health	1994-	12-18, 13-19, 18-26, 24-32	1-4	1556	US	US
LTS CTS	1992-	12-13,17-19 13-18	2 1	3166	US: Colorado	CO
MFTS	1990-	11-12,14-15,17-18, 20-21	1-6	4137	US: Minnesota	MN
MASATS	1995-1997	11-18	1	2211	US: VA, NC	NC
VTSABD-YAFU-TSA	1986-2007	8-16, 18-30, 22-32	1-6	2832	US: VA	VA
CVT	1983-1993	9-17	1-5	1180	US: VA	VA2
EFPTS: LLTS	1986-1999	10-16,18	1-8	210	Belgium: EFPTS	BE
FTC: FinnTwin16 FTC: FinnTwin12 FTC: Old cohort	1991-1997 1997-2004 1975	16,17,18 14,17 18-19	1-3 2,3 1	14279	Finland	FI
NTR	1991-	13-22	1-8	13425	the Netherlands	NL
STR: TCHAD	1993-	8-9,13-14,16-17,19-20	1-4	2942	Sweden: STR	SW
ATR: AYATS	1988-1996	13-18	1-3	2888	Australia: ATR	AU

Add Health = National Longitudinal Study of Adolescent Health

LTS = Colorado Longitudinal Twin Sample; CTS = Colorado Twin Registry Community Twin Sample

MFTS = Minnesota Family and Twin Studies

MATR: MASATS = Mid-Atlantic Twin Registry: Mid-Atlantic School Age Twin Study

VTSABD-YAFU-TSA = Virginia Twin Study of Adolescent Behavioral Development - Young Adult Follow Up - Transitions to Substance Use

CVT = Medical College of Virginia CardioVascular Twin Study

EFPTS: LLTS = East Flanders Prospective Twin Survey: Leuven Longitudinal Twin Survey

FTC = Finnish Twin Cohort

NTR = Netherlands Twin Registry

STR: TCHAD = Swedish Twin Registry: Twin Study of Child and Adolescent Development

ATR: AYATS = Australian Twin Registry: Australian Young Adult Twin Study

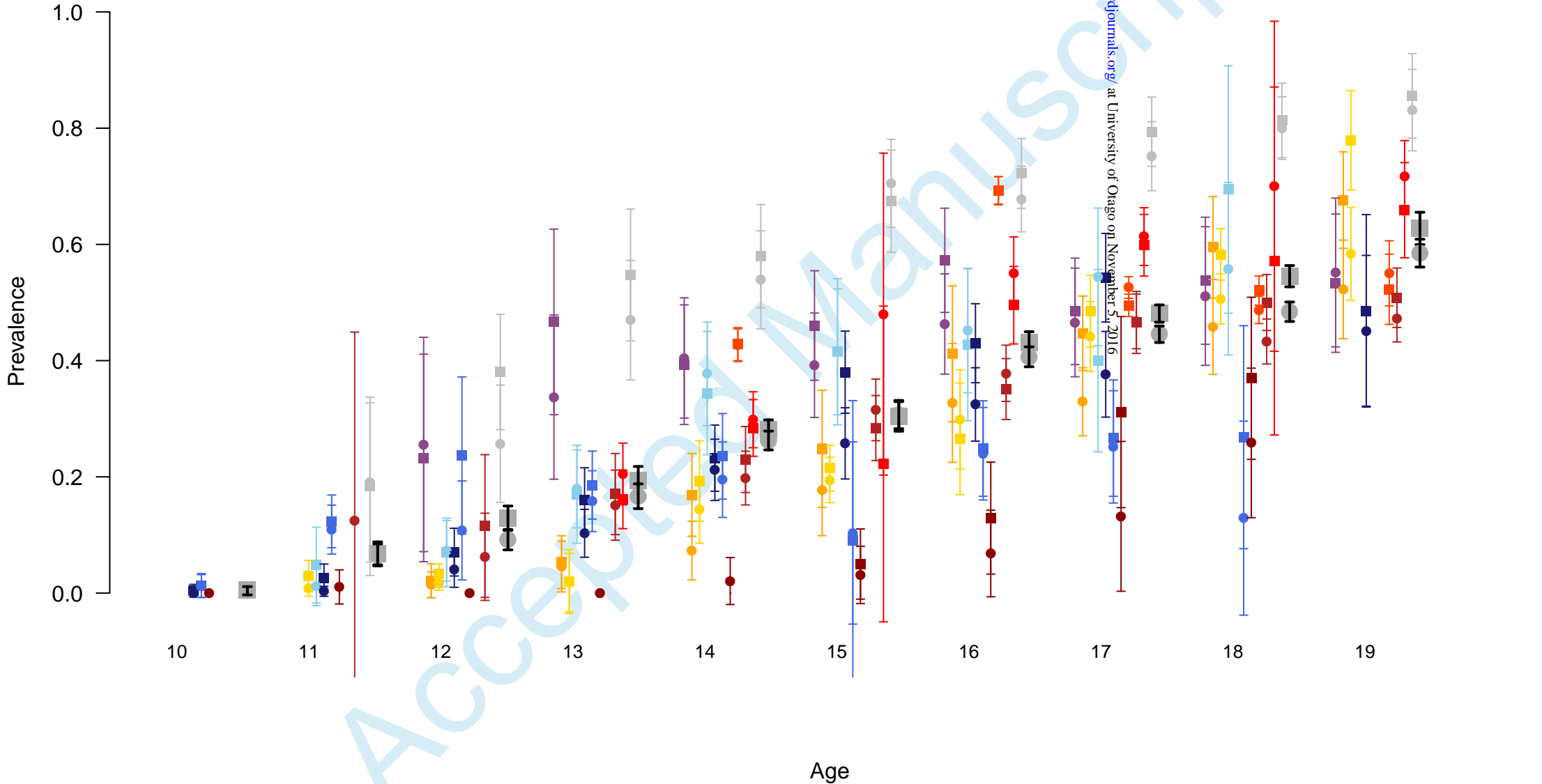
& two-letter abbreviations to be used in tables and graphs

Table 2: Number of twins assessed for SI across age and sample (top), including number of individual assessments (IA), number of unique individual twins per sample (UI) between the ages of 10 and 19, and by zygosity and sample (bottom), including number of unique individuals (UIZ) and pairs of twins (UPZ) with known zygosity per sample.

Age	10	11	12	13	14	15	16	17	18	19	IA	UI
US			86	161	346	442	491	448	295	251	2520	1480
CO			401	363	418	321	298	958	528	510	3797	2770
MI		646	1763	105	535	1629	383	1024	1979	474	8538	4128
NC		164	335	404	382	277	276	211	122		2171	2172
VA	565	636	619	755	884	755	829	660		183	5886	2710
VA2	476	826	178	712	542	40	434	305	72		3585	1148
BE	124	187	187	186	186	196	181	114	181		1542	210
FI					4709		5729	10683	6490	1133	28744	13262
NL		10	132	494	993	1090	1410	1536	2089	1944	9698	6036
SW				853	1352	43	924	1395	41	672	5280	2561
AU		112	328	329	507	496	956	761	709	399	4597	2869
Total	1165	2581	4029	4362	10854	5289	11911	18095	12506	5566	76358	39346
	MZm	MZf	DZm	DZf	DZo	UIZ	MZm	MZf	DZm	DZf	DZo	UPZ
US	268	282	246	220	380	1396	138	143	127	113	201	722
CO	613	729	440	457	529	2768	309	365	222	229	266	1391
MI	1166	1226	642	728	0	3762	583	613	321	364	0	1881
NC	329	560	283	393	497	2062	179	301	159	209	286	1134
VA	623	805	345	370	559	2702	312	403	173	185	281	1354
VA2	288	308	146	164	242	1148	144	154	73	82	121	574
BE	42	44	42	40	42	210	21	22	21	20	21	105
FI	1828	2241	2436	2393	3542	12440	919	1125	1235	1208	1775	6262
NL	939	1535	793	1009	1646	5922	494	816	421	551	913	3195
SW	475	515	337	388	754	2469	241	258	170	195	387	1251
AU	574	751	395	441	704	2865	288	376	199	225	356	1444
Total	7145	8996	6105	6603	8895	37744	3628	4576	3121	3381	4607	19313

Prevalence for Males (boxes) & Females (circles) – Smoking Initiation

US CO MN NC VA VA2 BE FI NL W AU All



from <http://nr.oxfordjournals.org/> at University of Otago on November 5, 2016

Accepted Manuscript

