Original investigation

A Genetic Epidemiological Mega Analysis of Smoking Initiation in Adolescents

Hermine H. Maes PhD1,2,3,4, Elizabeth Prom-Wormley MPH, PhD1,5, Lindon J. Eaves PhD1,2, PhD Soo Hyun Rhee PhD6, John K. Hewitt PhD6, Susan Young PhD6,7, Robin Corley PhD6, Matt McGue PhD8, William G. Iacono PhD8, Lisa Legrand PhD9, Diana R. Samek PhD8,9, E. Lenn Murrelle PhD10, Judy L. Silberg PhD1,2, Donna R. Miles PhD11, Richard M. Schienken MD12, Gaston P. Beunen PhD4, Martine Thomis PhD4, Richard J. Rose PhD13, Danielle M. Dick PhD2,14, Dorret I. Boomsma PhD15, Meike Bartels PhD16, Jacqueline M. Vink PhD15,16, Paul Lichtenstein PhD17, Victoria White PhD18, Jaakko Kaprio MD, PhD19,20, Michael C. Neale PhD1,2

1Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 2Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 3Massey Cancer Center, Virginia Commonwealth University, Richmond, VA; 4Department of Kinesiology, KU Leuven, Leuven, Belgium; 5Division of Epidemiology, Department of Family Medicine and Population Health, Virginia Commonwealth University, Richmond, VA; 6Institute for Behavioral Genetics, University of Colorado, Boulder, CO; 7Department of Psychiatry, University of Colorado, Denver, CO; 8Department of Psychology, University of Minnesota, Minneapolis, MN; 9Human Development and Family Studies Department, Auburn University, Auburn, AL; 10Venebio Group, LLC, Richmond, VA; 11Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 12Department of Pediatric Cardiology, Virginia Commonwealth University, Richmond, VA; 13Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN; 14Department of Psychology, Virginia Commonwealth University, Richmond, VA; 15Department of Biological Psychology, VU University, Amsterdam, The Netherlands; 16Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands; 17Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 18Centre for Behavioural Research in Cancer, Cancer Council Victoria, Melbourne, Australia; 19Department of Public Health and Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland; 20Department of Health, National Institute for Health and Welfare, Helsinki, Finland

Corresponding Author: Hermine H. Maes, PhD, Virginia Institute for Psychiatric and Behavioral Genetics, PO Box 980033, Richmond, VA 23298-0033, USA. Telephone: 804-828-8145; Fax: 804-828-8801; E-mail: hmaes@vcu.edu

Abstract

Introduction: Previous studies in adolescents were not adequately powered to accurately disentangle genetic and environmental influences on smoking initiation (SI) across adolescence. Methods: Mega-analysis of pooled genetically informative data on SI 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 ages 10 and 19, with 76,358 longitudinal assessments between 1983 and 2007, from 11 population-based twin samples from the United States, Europe, and Australia.
Results: Although prevalences differed between samples, twin correlations did not, suggesting similar etiology of SI across developed countries. The estimate of additive genetic contributions to liability of SI increased from approximately 15% to 45% from ages 13 to 19. Correspondingly, shared environmental factors accounted for a substantial proportion of variance in liability to SI at age 13 (70%) and gradually less by age 19 (40%).

Conclusions: Both additive genetic and shared environmental factors significantly contribute to variance in SI throughout adolescence. The present study, the largest genetic epidemiological study on SI to date, found consistent results across 11 studies for the etiology of SI. Environmental factors, especially those shared by siblings in a family, primarily influence SI 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 SI 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’s report.1 Tobacco kills nearly 6 million people each year, of whom more than five million are users and ex-users and more than 600,000 are nonsmokers exposed to secondhand smoke.2 In the United States, 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.4

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 19645 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 addiction6—which can occur quickly with smoking as few as 100 cigarettes7—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- to 19-year olds8 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.9 Recent reviews10,11 show heritability estimates for SI from 40% to 70% with family environmental influences more pronounced in adolescence than in adulthood.12–20 Furthermore, there is evidence for overlapping genetic and environmental risk factors between SI and nicotine dependence in adults21,22 suggesting that partly the same genes contribute to liability to SI and nicotine dependence. This evidence for a correlated liability of SI and nicotine dependence 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 factors.23 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.24 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 SI. 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: (1) estimate prevalence of SI across adolescence and test heterogeneity across samples; (2) estimate twin correlations for SI and test their equivalence across samples by age; (3) estimate contributions of genetic, shared environmental, and specific environmental factors to liability of SI at every age across adolescence; and (4) test for sex differences.

Methods

Subjects

Table 1 shows the main characteristics of the participating 11 studies (Supplementary 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. 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 Supplementary Appendix 1. Two studies (CardioVascular Twin Study and Leuven Longitudinal Twin Study) 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

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, and correlations to differ by zygosity and sample (Supplementary Table 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, sex, and sample using structural equation modeling.

Results
Descriptive Statistics
Data from 11 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 CardioVascular Twin Study and Leuven Longitudinal Twin Study. 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 co-twin, sex, and sample using structural equation modeling.


Two-letter abbreviations to be used in tables and graphs.

### Table 1. Years of Assessments, Ages of Participants, Number of Data Collection Waves, Number of Unique Individuals and Geographical Location of Participating Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Years</th>
<th>Ages</th>
<th>Wave</th>
<th>N#</th>
<th>Location</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Health</td>
<td>1994–</td>
<td>12–18, 13–19, 18–26, 24–32</td>
<td>1–4</td>
<td>1556</td>
<td>United States</td>
<td>US</td>
</tr>
<tr>
<td>CTS</td>
<td>1990–</td>
<td>13–18</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MASATS</td>
<td>1986–</td>
<td>8–16, 18–30, 22–32</td>
<td>1–6</td>
<td>2832</td>
<td>United States: VA</td>
<td>VA</td>
</tr>
<tr>
<td>CVT</td>
<td>1983–</td>
<td>9–17</td>
<td>1</td>
<td>1180</td>
<td>United States: VA</td>
<td>VA2</td>
</tr>
<tr>
<td>EFPTS: LLTS</td>
<td>1986–</td>
<td>10–16, 18</td>
<td>1–5</td>
<td>210</td>
<td>Belgium: EFPTS</td>
<td>BE</td>
</tr>
<tr>
<td>FTC: FinnTwin16</td>
<td>1991–</td>
<td>16, 17,18</td>
<td>1–3</td>
<td>14,279</td>
<td>Finland</td>
<td>FI</td>
</tr>
<tr>
<td>FTC: FinnTwin12</td>
<td>1997–</td>
<td>14, 17</td>
<td>2, 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTC: Old cohort</td>
<td>1975</td>
<td>18–19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NTR</td>
<td>1991–</td>
<td>13–22</td>
<td>1–8</td>
<td>13,425</td>
<td>the Netherlands</td>
<td>NL</td>
</tr>
<tr>
<td>STR: TCHAD</td>
<td>1993–</td>
<td>8–9, 13–14, 16–17, 19–20</td>
<td>1–4</td>
<td>2942</td>
<td>Sweden; STR</td>
<td>SW</td>
</tr>
<tr>
<td>ATR: AYATS</td>
<td>1988–</td>
<td>13–18</td>
<td>1–3</td>
<td>2888</td>
<td>Australia; ATR</td>
<td>AU</td>
</tr>
</tbody>
</table>

### Table 2. Number of Twins Assessed for Smoking Initiation (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

<table>
<thead>
<tr>
<th>Age</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>IA</th>
<th>UI</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>86</td>
<td>161</td>
<td>346</td>
<td>442</td>
<td>491</td>
<td>448</td>
<td>295</td>
<td>251</td>
<td>2520</td>
<td>1480</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>401</td>
<td>363</td>
<td>418</td>
<td>321</td>
<td>298</td>
<td>958</td>
<td>528</td>
<td>510</td>
<td>3797</td>
<td>2770</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>646</td>
<td>1763</td>
<td>105</td>
<td>535</td>
<td>1629</td>
<td>383</td>
<td>1024</td>
<td>1979</td>
<td>474</td>
<td>8538</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC</td>
<td>164</td>
<td>335</td>
<td>404</td>
<td>382</td>
<td>277</td>
<td>276</td>
<td>211</td>
<td>122</td>
<td>2171</td>
<td>2172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VA</td>
<td>565</td>
<td>636</td>
<td>619</td>
<td>755</td>
<td>884</td>
<td>755</td>
<td>829</td>
<td>660</td>
<td>183</td>
<td>5886</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VA2</td>
<td>476</td>
<td>826</td>
<td>178</td>
<td>712</td>
<td>542</td>
<td>40</td>
<td>434</td>
<td>305</td>
<td>72</td>
<td>3585</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>124</td>
<td>187</td>
<td>187</td>
<td>186</td>
<td>186</td>
<td>186</td>
<td>181</td>
<td>181</td>
<td>1542</td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL</td>
<td>4709</td>
<td>5729</td>
<td>718</td>
<td>487</td>
<td>675</td>
<td>389</td>
<td>421</td>
<td>473</td>
<td>1994</td>
<td>9698</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>10</td>
<td>132</td>
<td>494</td>
<td>399</td>
<td>1090</td>
<td>1410</td>
<td>1536</td>
<td>2089</td>
<td>1944</td>
<td>6036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SW</td>
<td>853</td>
<td>1352</td>
<td>43</td>
<td>924</td>
<td>1395</td>
<td>41</td>
<td>672</td>
<td>5280</td>
<td>2561</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>112</td>
<td>328</td>
<td>329</td>
<td>507</td>
<td>496</td>
<td>956</td>
<td>761</td>
<td>709</td>
<td>399</td>
<td>2869</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1165</td>
<td>2581</td>
<td>4029</td>
<td>4362</td>
<td>10854</td>
<td>5289</td>
<td>11911</td>
<td>18095</td>
<td>2506</td>
<td>39346</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>IA</th>
<th>UIZ</th>
<th>UPZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZm</td>
<td>268</td>
<td>282</td>
<td>246</td>
<td>220</td>
<td>380</td>
<td>1396</td>
<td>138</td>
<td>143</td>
<td>127</td>
<td>113</td>
<td>201</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MZf</td>
<td>613</td>
<td>729</td>
<td>440</td>
<td>457</td>
<td>529</td>
<td>2768</td>
<td>309</td>
<td>365</td>
<td>222</td>
<td>229</td>
<td>266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZm</td>
<td>1166</td>
<td>1226</td>
<td>642</td>
<td>728</td>
<td>0</td>
<td>3762</td>
<td>583</td>
<td>613</td>
<td>321</td>
<td>364</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZf</td>
<td>329</td>
<td>560</td>
<td>283</td>
<td>393</td>
<td>497</td>
<td>2062</td>
<td>179</td>
<td>301</td>
<td>159</td>
<td>209</td>
<td>286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZo</td>
<td>623</td>
<td>805</td>
<td>345</td>
<td>370</td>
<td>559</td>
<td>2702</td>
<td>312</td>
<td>403</td>
<td>173</td>
<td>185</td>
<td>281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MZm</td>
<td>288</td>
<td>308</td>
<td>146</td>
<td>164</td>
<td>242</td>
<td>1148</td>
<td>144</td>
<td>154</td>
<td>73</td>
<td>82</td>
<td>121</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MZf</td>
<td>613</td>
<td>729</td>
<td>440</td>
<td>457</td>
<td>529</td>
<td>2768</td>
<td>309</td>
<td>365</td>
<td>222</td>
<td>229</td>
<td>266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZm</td>
<td>1166</td>
<td>1226</td>
<td>642</td>
<td>728</td>
<td>0</td>
<td>3762</td>
<td>583</td>
<td>613</td>
<td>321</td>
<td>364</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZf</td>
<td>329</td>
<td>560</td>
<td>283</td>
<td>393</td>
<td>497</td>
<td>2062</td>
<td>179</td>
<td>301</td>
<td>159</td>
<td>209</td>
<td>286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZo</td>
<td>623</td>
<td>805</td>
<td>345</td>
<td>370</td>
<td>559</td>
<td>2702</td>
<td>312</td>
<td>403</td>
<td>173</td>
<td>185</td>
<td>281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MZm</td>
<td>288</td>
<td>308</td>
<td>146</td>
<td>164</td>
<td>242</td>
<td>1148</td>
<td>144</td>
<td>154</td>
<td>73</td>
<td>82</td>
<td>121</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MZf</td>
<td>613</td>
<td>729</td>
<td>440</td>
<td>457</td>
<td>529</td>
<td>2768</td>
<td>309</td>
<td>365</td>
<td>222</td>
<td>229</td>
<td>266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZm</td>
<td>1166</td>
<td>1226</td>
<td>642</td>
<td>728</td>
<td>0</td>
<td>3762</td>
<td>583</td>
<td>613</td>
<td>321</td>
<td>364</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZf</td>
<td>329</td>
<td>560</td>
<td>283</td>
<td>393</td>
<td>497</td>
<td>2062</td>
<td>179</td>
<td>301</td>
<td>159</td>
<td>209</td>
<td>286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DZo</td>
<td>623</td>
<td>805</td>
<td>345</td>
<td>370</td>
<td>559</td>
<td>2702</td>
<td>312</td>
<td>403</td>
<td>173</td>
<td>185</td>
<td>281</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MZm = monozygotic male; MZf = monozygotic female; DZm = dizygotic male; DZf = dizygotic female; DZo = dizygotic opposite sex.

---

![Prevalence for Males (boxes) & Females (circles) – Smoking Initiation](image)

**Figure 1.** Prevalence rates of smoking initiation (SI) by sex (males in boxes, females in circles), age and sample.
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 Supplementary eFigure 1, a and b, along with joint estimates after equating correlations across samples, for ages 12–19 (see Supplementary eFigure 2, a and 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 (Supplementary 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 showed 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 decreased from 1.0 to about 0.6 from ages 12 to 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 SI data across 11 primarily longitudinal twin datasets to obtain stable and reliable estimates of additive genetic, shared, and unique environmental contributions to liability of SI.
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. 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 11 samples, primarily of European descent, comprising five samples from the United States, four European, and one Australian, resulting in sample sizes ranging from ~4000 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 surveys.

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 epidemic. 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 competition. 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 drawn.

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 SI, 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 behavior. 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 smoking. As the majority of the samples in the current report were collected in the 1990s, 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 SI.

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. 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 behavior, it seems unlikely that assortment would account for the shared environment found here. Assuming a marital correlation for SI of 0.20, shared environmental contributions would be overestimated by 2% and 15% for MZ correlations of 0.85 and DZ correlations of 0.50 and 0.70 respectively. Correspondingly, heritability would be underestimated.

Fifth, while sex differences in genetic and environmental factors were nonsignificant in early adolescence, they were postpuberty, 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 factors.

Results from this study increase the evidence beyond suggestive 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 variation in consumption of cigarettes. The strongest association for SI was reported for SNPs in the BDNF gene. A handful of other SNPs have been found to be genome-wide significant for SI, but require replication. 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. 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,” evidence is accumulating that shared environmental factors contribute significantly to behavior in early adolescence, and especially for externalizing behaviors and substance use. These shared environmental effects may result from: parents and older siblings (including secondhand smoke and effects of assortative mating); peers, or social environment factors such as advertising controls, tobacco pricing, smoke-free regulations and tobacco availability.

Heritability of SI is significant even at young ages and jumps at ages 14–15, the transition from middle to high school in the United States, 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 teenagers. These results are consistent with a review of the effectiveness of family-based interventions to prevent children...
and adolescent from starting to smoke,60 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 promise.61 We also believe that providing personalized information—including genetic information—about smoking risks could improve smoking prevention.62 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 one or two 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.

Supplementary Material
Supplementary Appendix 1, eFigures 1 and 2, and eTables 1 and 2 can be found online at http://www.ntr.oxfordjournals.org

Funding
Data analysis supported by the National Institutes of Health (DA023109, DA024304 and DA018675). 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 (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 (OS twins) and DA036216. MATR: MASATS: Data collection was supported by the Templeton Foundation and the Keck Foundation (K01 DA000386). 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 RJR), and AA018755 and AA015416 to DD), the Academy of Finland Center of Excellence in Complex Disease Genetics (213506, 129680 to JK), and the Academy of Finland (100499, 205585, 118555, 141054, 265240, 263278 and 264146 to JK). NTR: Data collection was supported by Netherlands Organization for Scientific Research (NWO: 463-06-001, 451-04-034) and (MagW/ZonMW: 904-61-090, 985-10-002, ZonMw 912-10-020, ZonMw 904-61-193, 451-04-034, 400-05-717, ZonMw Addiction-31160008, ZonMw Middelgroot 911-09-032), Spinozopremie (SPI 56-464-14192); BMR-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: ATYS: 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. JMV is supported by the European Research Council (ERC) starting grant (284167) to PJ JMV.

Declaration of Interests
All authors reviewed and approved the final manuscript before its submission. JK has consulted for Pfizer on nicotine dependence in 2012–2014.

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

References


