Explaining low mortality among US immigrants relative to native-born Americans: the role of smoking

Laura Blue1* and Andrew Fenelon2

1Office of Population Research, Princeton University, Princeton, NJ, USA and 2Department of Sociology and Population Studies Center, University of Pennsylvania, Philadelphia, PA, USA

*Corresponding author. Office of Population Research, Wallace Hall (Floor 2), Princeton University, Princeton, NJ 08544, USA. E-mail: lblue@princeton.edu

Accepted 11 January 2011

Background In many developed countries, immigrants live longer—that is, have lower death rates at most or all ages—than native-born residents. This article tests whether different levels of smoking-related mortality can explain part of the ‘healthy immigrant effect’ in the USA, as well as part of the related ‘Hispanic paradox’: the tendency for US Hispanics to outlive non-Hispanic Whites.

Methods With data from vital statistics and the national census, we calculate lung cancer death rates in 2000 for four US subpopulations: foreign-born, native-born, Hispanic and non-Hispanic White. We then use three different methods—the Peto–Lopez method, the Preston–Llei–Wilmoth method and a novel method developed in this article—to generate three alternative estimates of smoking-related mortality for each of the four subpopulations, extrapolating from lung cancer death rates. We then measure the contribution of smoking-related mortality to disparities in all-cause mortality.

Results Taking estimates from any of the three methods, we find that smoking explains >50% of the difference in life expectancy at 50 years between foreign- and native-born men, and >70% of the difference between foreign- and native-born women; smoking explains >75% of the difference in life expectancy at 50 years between US Hispanic and non-Hispanic White men, and close to 75% of the Hispanic advantage among women.

Conclusions Low smoking-related mortality was the main reason for immigrants’ and Hispanics’ longevity advantage in the USA in 2000.

Keywords Health status disparities, minority health, smoking, lung neoplasms, mortality, statistics as topic

Introduction In many developed countries, including Australia,1 Canada,2 Germany3 and the USA,4 immigrants appear to outlive native-born residents, with lower death rates at most or all ages. Scholars have been puzzled by this ‘healthy immigrant’ (or ‘healthy migrant’) effect—sometimes also called the ‘immigrant paradox’—as it runs counter to an otherwise persistent trend for richer and better-educated populations to
live longer. In the USA, for example, immigrants are not only less well educated, less wealthy and more likely to live in poverty,5 but also often have poorer access to health care. For similar reasons, scholars have been perplexed by the ‘Hispanic paradox’: a tendency for US Hispanics to live longer5 and healthier lives than non-Hispanic Whites.

There are several hypotheses to explain immigrants’ life expectancy advantage. Kennedy et al.7 found evidence for migrant self-selection—that people who are relatively healthy find it easier to settle in new countries. There is also some evidence for selection in return migration—a so-called ‘salmon bias’ effect, where unhealthy migrants are more likely than healthy ones to return to their countries of origin,8 although Turra and Elo9 suggest this can explain only some fraction of migrants’ and Hispanics’ longevity advantage in the USA. Finally, immigrants may maintain healthier lifestyles in their host country than native-born residents. A study of adolescent behaviour10 shows that California minority immigrants, for example, eat more fruit and vegetables and drink less soda than the state’s non-Hispanic White adolescent population. Foreign-born adults in the USA also have lower rates of obesity than native-born Americans.11

We suggest that smoking habits may contribute to immigrants’ relative good health, at least in the USA. Immigrants and Hispanics are less likely to be current smokers12,13 or former smokers13 than native-born US Whites. Although the differences in smoking prevalence are only ~2 percentage points among men and 6 among women, smoking-attributable mortality is affected not only by current behaviour, but also by past behaviour. People dying of smoking-related diseases today—almost always people in middle age or older—most likely began smoking as teenagers or young adults in the 1950s through 1970s, a period of relatively heavy tobacco use in the USA. Figures from the National Center for Health Statistics (NCHS) put smoking prevalence in the mid-1960s at 51% among adult men and 33% among adult women.14

Using individual-level data, Denney et al.15 show that Hispanics’ survival advantage is attenuated once controls for smoking status are introduced. However, such studies may still fail to capture the full impact of smoking if—as reports from the National Health Interview Survey suggest13—Hispanic smokers are typically lighter smokers than other smokers are. Indirect methods may yield more robust estimates of smoking-attributable mortality.

Methods

We estimate the contribution of smoking-attributable mortality to all-cause mortality disparities observed in 2000 between foreign-born US residents and native-born Americans, and between US Hispanics and non-Hispanic Whites.

We use three different methods to generate, for each analysis, three alternative estimates of smoking-attributable mortality. All three methods use the death rate from lung cancer as a marker of accumulated smoking exposure within a population. This approach is supported by a large research literature that shows smoking to be the main source of variation in lung cancer mortality among populations.16–18 Our three methods are the Peto–Lopez method,19,20 which extrapolates smoking-related mortality from lung cancer mortality based on risks observed among smokers in the Cancer Prevention Study II; the newer Preston–Glei–Wilmoth (PGW) method,21 which instead extrapolates smoking-related deaths from correlations between lung cancer mortality and mortality from other causes across 20 high-income countries; and a novel indirect estimation technique, outlined below. While all three methods make broadly similar assumptions, they allow assumptions to be relaxed in different ways.

Calculation of lung cancer mortality

For all three methods, we first calculated all-cause mortality for the year 2000, tabulating age-specific death rates by sex for four subpopulations: foreign-born, native-born, Hispanic and non-Hispanic White (Figure 1). We took rate numerators (deaths) from the Multiple Cause-of-Death Public-Use Microdata files, available from NCHS. These data include dece- dents’ race, ethnicity and place of birth, as recorded on each person’s death certificate. From a total of 2,407,193 deaths in 2000, we dropped 363 entries without a recorded age and 14,378 with missing place of birth. We followed the death certificate classification of ‘Hispanic’ and ‘non-Hispanic White’. We considered anyone born outside the 50 US states and the District of Columbia to be ‘foreign-born’, and anyone born inside to be ‘native-born’. We took our rate denominators (population counts) from the US 2000 Census 5% Public Use Microdata Sample Files. These data, too, provide individual-level records on race, ethnicity and place of birth. Using these same data sources, we then calculated age- and sex-specific death rates in 2000 for lung cancer only, using as our numerator the number of deaths indicated on death certificates as attributable to lung cancer [International Classification of Diseases (ICD)-10 codes C33–C34].

A new method to estimate smoking-attributable mortality

We have developed a new method to estimate smoking-attributable mortality. Our method builds on the familiar assumption that lung cancer mortality is a reliable marker of smoking exposure in a population. Not all lung cancers are caused by smoking, and lung cancer is not the only deadly smoking-related condition. However, if we know both the proportion of lung cancers caused by smoking (\(P\)) and the
proportion of smoking-related mortality that is lung cancer mortality \((Q)\), we can estimate total smoking deaths \((D_S)\) as follows:

\[
D_S = \frac{P \cdot D_L}{Q}
\]

where \(D_L\) is the number of lung cancer deaths.

A great deal of epidemiological work has focused on the two proportions that we call here \(P\) and \(Q\). \(P\) is simply the attributable risk, the proportion of lung cancer mortality that would not have occurred in the absence of smoking:

\[
P = \frac{(M_L - M_L^*)}{M_L}
\]

where \(M_L\) is the observed population lung cancer death rate, and \(M_L^*\) is the lung cancer death rate among members of that population who have never smoked. It can be difficult to get precise \(M_L^*\) values by age and sex, since lung cancer deaths are rare among never-smokers. However, Michael Thun et al.\(^{22}\) have recently produced such estimates, pooling together rates from both US Cancer Prevention Studies, the Nurses’ Health Study, the Women’s Health Study and other major US and non-US trials and cohort studies. We use the pooled estimates among never-smoker Whites (both in the USA and abroad) as our baseline risk for lung cancer mortality in the absence of smoking. We then calculate \(P\) by age and sex for each subpopulation (see Supplementary Appendix A for values, available as supplementary data at IJE online).

We calculate \(Q\) from Centers for Disease Control and Prevention (CDC) tabulations of smoking-related mortality from 1997 to 2001.\(^{23}\) During those years, the CDC estimates that lung cancer caused 32% of all smoking-attributable deaths among men and 29% among women, with the bulk of deaths due to other conditions, including heart disease, chronic obstructive pulmonary disease and cancer of other organs. In our calculations, therefore, we take \(Q\) as 0.32 for men and 0.29 for women. For simplicity, we assume these values do not vary by age.

**Sensitivity analyses**

The Peto–Lopez and PGW methods assume that excess lung cancer deaths and other smoking-attributable deaths are related in the same way in all populations. Our new method allows this assumption to be relaxed. To test robustness, we use the new method to calculate six different scenarios in addition to our principal estimates.

- **Low estimate**: we allow \(P\) and \(Q\) to vary by subpopulation. \(P\) is calculated as in the principal analysis. Among native-born Americans and non-Hispanic Whites \(Q = 0.29\) for females and \(Q = 0.32\) for males, as observed in the population at large. Among immigrants and Hispanics, however, \(Q = 0.24\) for females and \(Q = 0.27\) for males—arbitrarily set values that allow immigrants and Hispanics to be unusually susceptible to non-lung cancer smoking deaths.

- **High estimate**: we allow \(P\) and \(Q\) to vary by subpopulation. \(P\) is calculated for native-born Americans and non-Hispanic Whites as in the principal analysis; for immigrants and Hispanics, it is calculated from the never-smoker lung cancer death rates among Asians compiled by Thun et al.\(^{22}\) Among native-born Americans and non-Hispanic Whites, \(Q = 0.29\) for females and \(Q = 0.32\) for males as observed in the population at large. Among immigrants and Hispanics, \(Q = 0.34\) for females and \(Q = 0.37\) for males—arbitrarily set values that allow immigrants and Hispanics to be unusually resistant to smoking-related deaths from causes other than lung cancer.

- **Salmon bias correction**: \(P\) and \(Q\) are the same as those from ‘low estimate’. We account here for the possibility that observed deaths among immigrants or Hispanics are only some fraction of true deaths, as sick immigrants may leave the country to die.
Data from the Surveillance, Epidemiology and End Results (SEER) Program\textsuperscript{24} show that the ratio of lung cancer mortality to lung cancer incidence is higher for non-Hispanics (0.90 for men; 0.76 for women) than for Hispanics (0.79 for men; 0.57 for women). This could be a sign of Hispanics’ selective migration, or it could be that Hispanics have lower case severity (perhaps because they smoke less). Here, we assume the former, and inflate lung cancer death rates for immigrants and Hispanics by the full difference in mortality-to-incidence, multiplying by 1.14 for men and by 1.34 for women. [Data on mortality-to-incidence exist only by ethnicity, and not by place of birth. The mortality-to-incidence ratio is 14\% higher for non-Hispanic men than Hispanic men (0.90/0.79 = 1.14) and 34\% higher for non-Hispanic women relative to Hispanic women (0.76/0.57 = 1.34).]

- $\frac{1}{2}$ Salmon bias correction: $P$ and $Q$ are those from ‘low estimate’. Lung cancer death rates among immigrants and Hispanics are inflated by one-half of the difference in SEER mortality-to-incidence ratios observed between Hispanics and non-Hispanic Whites. For foreign-born and Hispanic women, lung cancer death rates are inflated by multiplying by $0.5 \times (0.76/0.57 - 1) + 1 = 1.16$; for men, by $0.5 \times (0.90/0.79 - 1) + 1 = 1.07$.

- Age-varying $Q$: we allow the proportion of smoking-attributable deaths caused by lung cancer to vary by age. We use the PGW method\textsuperscript{21} and national data for the USA in 2000 (described above) to calculate new estimates of $Q$, by sex, for each 5-year age group.

- Strong adjustment for possible confounding: taking $Q$ from the CDC assumes that all estimated excess deaths among smokers are directly caused by smoking. Here, we assume only half of smokers’ excess non-lung cancer deaths are smoking-attributable. That is, $Q = 0.29/(0.29 + [1 - 0.29] \times 0.5) = 0.45$ for women and $Q = 0.32/(0.32 + [1 - 0.32] \times 0.5) = 0.48$ for men.

## Results

In 2000, native-born Americans had substantially higher lung cancer death rates than US immigrants, and non-Hispanic Whites had substantially higher rates than US Hispanics (Figure 2). Using three alternative methods, we estimate, as a result, that total smoking-related mortality was also higher among native-born Americans and non-Hispanic Whites than among US immigrants and Hispanics. For each method, we present detailed mortality estimates by age and sex in Supplementary Appendix B (available as supplementary data at IJE online).

![Figure 2] US lung cancer death rates (log scale) in 2000
We use Arriaga’s method\textsuperscript{25} to decompose differences in life expectancy at age 50 years into two components: a component due to smoking and a component due to other factors. Taking estimates from any of the three methods, we find that smoking accounts for at least 50\% of migrants’ advantage in life expectancy at 50 years among men and at least 70\% among women. Smoking explains $>75\%$ of the difference in life expectancy at 50 years between US Hispanic and non-Hispanic-White men, and close to 75\% of this difference among women (Figure 3).

Table 1 shows sensitivity-analysis results. In most of the scenarios, smoking still accounts for $>50\%$ of the differences in life expectancy at 50 years. In one instance, smoking is estimated to explain $>100\%$; this means that, in the absence of smoking, male life expectancy would be higher among non-Hispanic Whites than among Hispanics.

**Discussion**

We find that low mortality from smoking is the main reason for immigrants’ and Hispanics’ longevity advantage in the USA in 2000. While previous studies have documented lower smoking prevalence among US immigrants compared with native-born Americans, and among Hispanics compared with non-Hispanic Whites, to our knowledge none of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Differences in life expectancy at age 50 years and proportion explained by smoking, by nativity and ethnicity. PL = Peto–Lopez method; PGW = Preston–Glei–Wilmoth method; BF = Blue–Fenelon method, developed in this article.}
\end{figure}
Table 1 Sensitivity analysis (from the new method to estimate smoking-attributable mortality): proportion of the difference in life expectancy at 50 years explained by smoking

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Foreign-native</th>
<th>Hispanic-non-Hispanic White</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal estimate</td>
<td>Females (%)</td>
<td>Males (%)</td>
</tr>
<tr>
<td></td>
<td>71.02</td>
<td>58.27</td>
</tr>
<tr>
<td>Low estimate (P and Q vary by subpopulation)</td>
<td>61.75</td>
<td>47.77</td>
</tr>
<tr>
<td>High estimate (P and Q vary by subpopulation)</td>
<td>84.90</td>
<td>79.02</td>
</tr>
<tr>
<td>Low estimate with strong salmon-bias correction</td>
<td>43.84</td>
<td>38.42</td>
</tr>
<tr>
<td>Low estimate with ½ salmon-bias correction</td>
<td>52.79</td>
<td>43.10</td>
</tr>
<tr>
<td>Age-varying Q (with Q from PGW)</td>
<td>66.12</td>
<td>46.77</td>
</tr>
<tr>
<td>Strong adjustment for possible confounders</td>
<td>45.77</td>
<td>38.84</td>
</tr>
</tbody>
</table>

these earlier studies calculates the contribution of smoking to US mortality disparities.

Like other demographers and epidemiologists,19–21,26 we believe lung cancer mortality is the most reliable marker of a population’s smoking behaviour. Mortality data are available by age and sex in many developed countries, whereas detailed data on smoking prevalence, duration and intensity generally are not. To date, studies using indirect methods have typically found that smoking explains a greater proportion of longevity differences than studies using direct methods. For example, Jha et al.27 (using the Peto–Lopez method) found that smoking explains more than half of the social gradient in mortality in England and Wales, whereas English longitudinal studies generate estimates closer to one quarter.28,29

Crude survey measurement of smoking habits may explain part of the discrepancy. Surveys can suffer from inaccurate self-reports, from non-response rates that leave the sample of respondents unrepresentative of the population at large, and—where participants are followed over time—by unrecorded changes in smoking behaviour. It is also difficult to measure smoking intensity with precision; survey respondents are usually asked to categorize their consumption into large bins, such as ‘more than 20 cigarettes per day’ or ‘former smoker’.

In this article, we develop a new method to estimate smoking-attributable deaths. Our technique is similar to the widely accepted Peto–Lopez method, although it is somewhat simpler; it ignores age variation in the cause-of-death distribution. Our main advantage, however, is the possibility of straightforward sensitivity checks. We can relax assumptions about P and Q, whereas the Peto–Lopez and PGW methods both assume that never-smoker lung cancer risks will not vary across populations (as they may with differences in genetic susceptibility or in exposure to carcinogens like asbestos or smoke from cooking fires). In practice, our new estimates of smoking-related mortality appear quite similar to those generated using existing methods.

We consider there to be two main sources of uncertainty in our results. First, all three methods used in this article assume that death certificates (and thus the NCHS Multiple Cause-of-Death data) contain complete and correct information on decedents’ birthplace, race, ethnicity and cause of death. As noted above, very few death records (<1%) are missing entries for these variables. Our assumption is nonetheless complicated by the fact that race and Hispanic status are identified differently in the Census (where they are given by self-report) than they are in death data (where reports are made by a third party). However, although race and ethnicity misclassification on death certificates may be more common among immigrants and Hispanics than among native-born non-Hispanics,30 Arias et al.31 show that the magnitude of this bias is probably not great. Furthermore, we see no obvious reason that these classification errors should also vary systematically and greatly by lung cancer as a cause of death. As a result, errors in birthplace and ethnicity recorded on death certificates may indeed cause us to overestimate both the immigrant paradox and the Hispanic paradox, but, crucially, they should not greatly bias our estimates of the proportion of these disparities that are caused by smoking—provided that the direction of the disparities is correct. We are confident that it is. Cohort study results,32 death data from Social Security9 and Medicare,30 and new US life tables by Hispanic origin6 are unlikely to suffer the same misclassification errors. All show a clear Hispanic or immigrant advantage.

Secondly, for the new estimation technique, there is some uncertainty in estimates of the key inputs, P and Q. For simplicity, to calculate P, we have assumed all US subpopulations would have identical lung cancer death rates in the absence of smoking. We chose to approximate this rate with the never-smoker lung cancer death rates of Whites, compiled by Thun et al.22 Whites have the lowest rates of any of the race in these authors’ global review. Using White rates therefore generates the most conservative estimates of the proportions of the immigrant paradox and the Hispanic paradox that are due to smoking. If we assumed instead that immigrants had the baseline lung cancer risk of either Blacks or Asians in Thun
et al.’s\textsuperscript{22} analysis (as we do in the ‘high estimate’ of the sensitivity analysis), we would estimate that smoking explains a larger proportion still.

For our estimates of $Q$, the proportion of smoking-related deaths that are lung cancer deaths, we import figures from the CDC. Although we believe these figures are among the best available, they are based on relative risks observed in cohort studies of smoking. Past cohorts were rarely representative of the US population, and it remains almost impossible to control fully for confounders that correlate smoking with diseases not actually caused by smoking.\textsuperscript{33} Furthermore, we cannot be certain that the cause-of-death distribution of smoking deaths is constant across subpopulations in which smoking-related mortality has not been studied, and calculations using the PGW method suggest this distribution does vary by age, violating the assumption of age-invariant $Q$. We consider several possible sources of error in our sensitivity analysis. Ultimately, however, the input figures in these alternative scenarios are to some degree arbitrary.

As it is unclear which assumptions about smoking-related mortality may be most accurate, in this article we have estimated the contribution of smoking to mortality disparities using three different methods. In addition, we have conducted sensitivity analysis for six alternative scenarios, relaxing assumptions about never-smoker lung cancer mortality, the cause-of-death distribution of smoking-attributable deaths, possible confounding and salmon bias. Many of these scenarios are no doubt highly implausible. Nevertheless, our estimates of the contribution of smoking to mortality disparities are very similar across all three methods, and are only moderately sensitive to the assumptions changed in the sensitivity analysis. As a result, though one may doubt the precision of figures we use to estimate smoking-related mortality, we remain quite confident in our conclusion: smoking is likely the major cause of America’s immigrant paradox and the related Hispanic paradox.

**Supplementary Data**

Supplementary data are available at IJE online.

**Funding**

This work was supported by the National Institutes of Health (5T32 HD 007163 and 5T32 HD 007242–28).

**Acknowledgements**

We thank Samuel H. Preston and two anonymous referees for their comments and guidance.

**Conflict of interest:** None declared.

**KEY MESSAGES**

- In the USA, foreign-born residents have higher life expectancy than native-born Americans, and Hispanics have higher life expectancy than non-Hispanic Whites, despite socioeconomic disadvantages among the long-lived populations.
- We hypothesize that smoking may account for some of these observed life expectancy differences by nativity and ethnicity.
- We use three different indirect methods—the Peto–Lopez method, the Preston–Glei–Wilmoth method and a novel method outlined in this article—to estimate smoking-related mortality in 2000 for four US subpopulations: foreign-born, native-born, Hispanic and non-Hispanic White.
- Using any of the three indirect estimation techniques, we find that smoking explains more than half of the difference in life expectancy at 50 years between foreign-born US residents and native-born Americans, and close to three-quarters of the difference in life expectancy between US Hispanics and non-Hispanic Whites.

**References**