Int J Health Econ Manag.
https://doi.org/10.1007/s10754-018-9255-y

RESEARCH ARTICLE

Genetic variation in health insurance coverage

George L. Wehby[1,2,3,4,5] · **Dan Shane**[1]

Received: 25 August 2017 / Accepted: 3 November 2018

Abstract

We provide the first investigation into whether and how much genes explain having health insurance coverage or not and possible mechanisms for genetic variation. Using a twin-design that compares identical and non-identical twins from a national sample of US twins from the National Survey of Midlife Development in the United States, we find that genetic effects explain over 40% of the variation in whether a person has any health coverage versus not, and nearly 50% of the variation in whether individuals younger than 65 have private coverage versus whether they have no coverage at all. Nearly one third of the genetic variation in being uninsured versus having private coverage is explained by employment industry, self-employment status, and income, and together with education, they explain over 40% of the genetic influence. Marital status, number of children, and available measures of health status, risk preferences, and prevention effort do not appear to be important channels for genetic effects. That genes have meaningful effects on the insurance status suggests an important source of heterogeneity in insurance take up.

Keywords Health insurance · Genetic variation · Health determinants · Risk taking

JEL Classification D1 · D82 · I10 · I12 · I13

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10754-018-9255-y) contains supplementary material, which is available to authorized users.

✉ George L. Wehby
george-wehby@uiowa.edu

Dan Shane
dan-shane@uiowa.edu

1 Department of Health Management and Policy, University of Iowa, 145 N. Riverside Dr., 100 College of Public Health Bldg., Room N250, Iowa City, IA 52242-2007, USA

2 Department of Economics, University of Iowa, Iowa City, IA, USA

3 Department of Preventive and Community Dentistry, University of Iowa, Iowa City, IA, USA

4 Public Policy Center, University of Iowa, Iowa City, IA, USA

5 National Bureau of Economic Research, Cambridge, MA, USA

Introduction

There is increasing interest in exploring genetic influences on healthcare utilization (True et al. 1997; Wehby et al. 2015, 2017). Health insurance status is a key determinant of access to and use of health services. Is it possible that genes play a significant role in whether someone obtains health insurance? While no previous work directly investigates the link between genetic variation and health insurance status, a burgeoning literature reviewed below has suggested connections between genes and several pathways linked to insurance status including health status, human capital, and risk preferences. Coverage decisions in the US are increasingly shifting directly to individuals, highlighting the need to more thoroughly understand the individual-level determinants of insurance status. Almost nothing is known about whether or not genetic factors are relevant for explaining the variation in health insurance status across the population and by how much.

We provide the first evaluation of the extent to which genes might explain variation between individuals in having health insurance coverage or being uninsured using a twin design and a national sample of US twins that is the only national US twin sample with data on health insurance coverage. We also investigate possible mechanisms related to human capital, labor market participation, household formation, health, and preferences for risk and prevention. Understanding whether there is a meaningful genetic link to health insurance decisions could open the door toward future research to better understand the underlying behavioral mechanisms which may ultimately inform how to more optimally design approaches to influence the behavioral side of such decisions.

Conceptual framework

Health insurance coverage is influenced by several individual-level traits such as a person's preferences for risk and prevention, health status which determines need for healthcare services, and human capital which affects access to private coverage via labor market outcomes (employment, earnings) as well as one's knowledge about insurance needs and efficiency in obtaining insurance. These traits may in turn be related to genes, leading to gene-driven variation in insurance coverage status through those channels (Fig. 1). The evidence for genetic influence on those traits has largely been based on twin studies similar to ours, but recent work has also employed molecular variation on population-based samples of unrelated individuals.

Beginning with preferences, individuals who are more risk averse and discount the future less are more likely to obtain insurance coverage. Genes can explain 14–20% of differences in risk aversion (Cesarini et al. 2009; Benjamin et al. 2012) and as much as 30–50% of the variation in discounting the future (Anokhin et al. 2011). Cronqvist and Siegel (2015), Cesarini et al. (2010), and Kuhnen and Chiao (2009) also show that genetic variation can explain between one quarter and one third of financial risk taking behavior. Cronqvist and Siegel (2015) further note that genetic differences in saving behavior likely reflect genetic differences between individuals in time preferences or self-control. This genetic influence is also supported when studying risky behaviors including smoking and alcohol dependence, which have been shown to have a strong genetic etiology that can explain as much as 50% of their variation (Boardman et al. 2010; Maes et al. 2004; Stacey et al. 2009; McGue et al. 2013).

There is also evidence that an individual's preferences over preventing disease are influenced by genes. For example, over one third of the variation in beliefs about whether personal actions can modify heart disease and cancer risks may be explained by genetic effects (Wehby

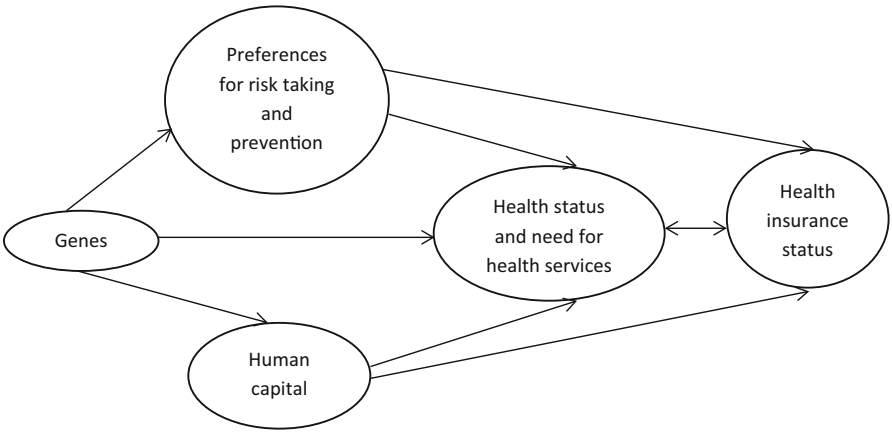

Fig. 1 Conceptual channels between genes and health insurance coverage. *Notes* The potential channel through household formation discussed above is omitted from the figure for brevity

et al. 2015). This genetic influence is more prominent when examining prevention effort (about 50% of the variation explained by genetic differences; Wehby et al. 2015). Consistent with those findings, there is evidence of strong genetic influence on key prevention activities such as participation in exercise (40–70% of variance explained by genetic effects; Moor et al. 2011; Stubbe et al. 2006), seeking preventive care such as having a preventive check-up (40% explained by genetic effects; Wehby et al. 2015) or women's participation in cancer screening programs (37–66% explained by genetic effects; Treloar et al. 1999), as well as health indicators tied to lifestyle such as body mass index (between 40 and 90% of the variation in body mass index is linked to genes; Elks et al. 2012; Carlsson et al. 2013; Boardman et al. 2015). Other personal traits specifically measured in relation to health decisions including self-efficacy and beliefs about healthcare effectiveness have also been examined but not found to be influenced by genes (Wehby et al. 2015).

The influence of health status on health insurance status, particularly through adverse selection, is well recognized (Cutler and Zeckhauser 1998; Cutler et al. 2008) and suggests another link between genes and insurance. There is ample evidence that genes affect health status which can generate indirect effects on the need and demand for insurance coverage. As direct evidence of adverse selection based on genetic risk, Oster et al. (2010) find that individuals carrying the genetic mutation for Huntington's disease are more likely to obtain long-term care insurance. The effects of genes on coverage through health status are likely much broader, however, as genes influence the incidence and severity of chronic physical and mental health diseases such as diabetes, hypertension, major depression, and asthma, among others; at least a third of the variation in each of those conditions can be explained by genetic effects (e.g. Agarwal et al. 2005; Carlsson et al. 2013; Kendler et al. 2006; Thomsen et al. 2010). Wehby et al. (2015) find that as much as half of the variance in the number of chronic conditions can be explained by genetic differences. Consistent with the connection to chronic diseases, genes explain an important fraction of the variation in use of prescription drugs (~40%, Wehby et al. 2015), seeking treatments for chronic conditions including mental health, joint problems, hearing problems (42–56%, True et al. 1997), and self-reported health status (Romeis et al. 2000, 2005; Wehby et al. 2015). Furthermore, polygenic scores for several chronic conditions have been associated with various measures of functional health and wellbeing (Wehby et al. 2018).

Genetic effects on human capital accumulation, which in turn affects labor market participation and income, represents an additional channel for genetic variation in insurance coverage. Human capital can reduce information costs, resulting in more effective search for and enrollment in health plans. Empirically, educational attainment is one of the key indicators of human capital that is also linked to insurance coverage (Cutler and Lleras-Muney 2006; Fang et al. 2008). Several studies suggest that genes can explain between a third and two thirds of the variation in educational attainment (Branigan et al. 2013; Boardman et al. 2015), raising the possibly of an indirect genetic effect on insurance through educational attainment and other forms of human capital accumulation. Interestingly however, there does not appear to be an important genetic effect on a measure of one's knowledge about own health (Wehby et al. 2015). Another potential connection with labor market participation is through self-employment, since genetic differences may explain nearly half of the variation in self-employment status (generally considered a proxy for entrepreneurship) (van der Loos et al. 2013) and since most employed individuals gain coverage through their employers.

A weaker indirect effect may occur through household formation given that individuals can obtain insurance coverage through their spouse or may be more likely to obtain family coverage if they have children. There is some evidence of genetic variation in marriage status at younger ages (20–40) but not older ages (Trumbetta et al. 2007) as well as on fertility indicators such as number of births and age at first child (Kosova et al. 2010). As detailed below, we investigate whether these potential channels explain observed genetic variation in insurance coverage.

Methods

Data

We employ data from the National Survey of Midlife Development in the United States (MIDUS I; Brim et al. 2017). The MIDUS included a national sample of 957 twin pairs in 1995 and 1996 from which data on several socioeconomic and health indicators were obtained. The sample includes an age range from 25 to 74 years. The sample declines to 907 pairs with data on identical/non-identical status and consistent data on age/birth year between the twins. In our first analysis, we make no restrictions on the type of health insurance (public or private) and focus on uninsured versus any insurance. The main analytical sample excludes 55 pairs because of missing data on coverage for one or both twins resulting in a total of 769 twin pairs, including 307 identical twin pairs and 462 non-identical twin pairs (268 same-sex pairs and 194 different-sex pairs).

In an additional model, we focus on private insurance versus uninsured among individuals who are not age-eligible for Medicare. Our expectation is that genetic variation is greater in this group when focusing on choice of private coverage compared to the main analysis. Therefore, we exclude individuals aged 65 or older (94% of whom had Medicare) and individuals younger than 65 who had Medicaid or military insurance (only 47 pairs). This second analysis tests the robustness of our results to availability and demand of public coverage. This analysis includes 583 twin pairs, with 240 identical pairs and 343 non-identical pairs (190 same-sex pairs and 153 different-sex pairs). The number of publicly insured individuals younger than 65 in this sample is too small to allow for an analysis comparing uninsured versus public coverage only. However, we also examine genetic variation in uninsured versus any coverage among individuals younger than 65 as an additional model, adding back the small sample of publicly insured individuals.

Empirical strategy

The outcome of interest in this study is having health insurance coverage or not. Privately insured individuals can have employment-based coverage (from their or their spouses' employer or union) or may obtain individual (non-group) coverage on their own. Publicly insured individuals include those with Medicare, Medicaid, or military insurance. Individuals without coverage in this analysis are those without any health care coverage plan. As mentioned above, we first examine uninsured versus private or public coverage then focus on uninsured versus private coverage.

We implement a twin comparison model to decompose the variation in having health insurance coverage or not across genetic and environmental differences. Identifying the overall genetic and environmental variation in a trait using this approach is based on the fact that identical twins share all genetic variants, while non-identical twins share on average half of the genetic variants, but also assumes that identical twins do not share a more similar household environment than non-identical ones (commonly referred to as the equal environment assumption). This assumption still holds if the greater similarity in the household environment for identical twins is driven by their similar genes. Identical versus non-identical twin status in the MIDUS I has been shown not to be related to household demographic and socioeconomic indicators including race, family financial status, family history of moving to new neighborhoods during childhood, and maternal education (Wehby et al. 2015). Other assumptions in the basic twin model include parents not being genetically related, additive genetic effects (the effect of each copy of a genetic variant can be summed linearly), and no or minimal interactions between genetic and environmental factors, although some of these assumptions can be relaxed. For instance, we estimate below both additive and non-additive genetic effects. Gene-by-environment interactions limit the generalizability of the estimated genetic variation with changing environments. For example, it is possible that the Affordable Care Act (ACA) has modified genetic variation in coverage. In our case, we recognize that our estimates are specific to the insurance policy environment at the time when our data were collected (1995–1996). Thus, our estimates can be viewed as a reference for evaluating changes in genetic variation in coverage in later periods. Despite concerns about some of its assumptions (e.g. Charney and English 2013; Domingue et al. 2014; Burt and Simons 2014), the basic model of comparing twins has been shown to be generally robust (Barnes et al. 2014).

We follow a regression based approach to decompose the variation in being uninsured into three sources of variation: genes, environment shared between twins, environment unique to each twin. Specifically, our main estimation follows a generalized mixed model estimated by a probit function with random effects and innovated by Rabe-Hesketh et al. (2008) based on the following error decomposition model:

$$I_{ij} = v + A_{ij} + D_{ij} + C_{ij} + E_{ij}. \tag{1}$$

In Eq. (1), the uninsured status (0/1) of twin i of twin-pair j is a function of the overall uninsured rate plus several error components including additive genetic effects (A), dominant genetic effects (D), shared environmental effects (C), and unique environmental effects (E). Each component is assumed to be normally distributed with a mean zero and a constant variance. Each genetic variant consisting of a DNA base pair referred to as a single-nucleotide-polymorphism (SNP) can have two copies (alleles) which can be the same or different. In an additive model, the genetic effect from a variant is the linear sum of the genetic effects of the two alleles. In contrast, dominant effects result from interactions between the two copies of a genetic variant (making their joint influence not simply the sum of their individual effects).

Shared environmental factors (C) in this context can be the household environment relevant to insurance coverage, while unique environmental factors (E) can include differences in availability and characteristics of employer-sponsored insurance between twins and in characteristics of the local individual (non-group) insurance markets.

The covariance matrix of A within and across twin pairs is a function of the constant variance multiplied by a matrix of weights based on genetic similarity (100% between identical twins and 50% between non-identical twins): diagonal entries are 1; off-diagonal terms are 1 for identical twins and 0.5 for non-identical twins in the same families and 0 across twins from different families (see further details in Rabe-Hesketh et al. 2008). Similarly, the covariance matrix of D involves a similar weighting matrix, except that the off-diagonal terms are 0.25 for non-identical twins following genetic theory. For C, the diagonal terms and off-diagonal terms for twins from the same family are 1 in the covariance weighting matrix and 0 otherwise. Finally for E, all off-diagonal terms are 0. In the linear model, the covariance of I is then simply the sum of the covariances of these four components: A, D, C, and E.

Of course, identifying all four components simultaneously is impossible. Therefore, one has to impose restrictions on one or more of these components in order to separate these sources of variation. In the first specification, we assume all genetic effects to be additive (A) and assume no genetic dominance (this model is commonly referred to as the ACE model). In the second specification, we add the restriction that shared environmental effects are null and only estimate additive genetic effects and unique environmental effects (AE model). In the third specification, we relax the hypothesis of all genetic effects being additive and separate genetic effects into both additive and dominant components (ADE model), while continuing to assume null shared environmental influence. Details on the parameterizations for these specifications are included in Rabe-Hesketh et al. (2008). From all these models, we estimate the proportion of the variance in being uninsured that is due to genetic effects; this parameter is commonly referred to as "genetic heritability" or h^2, which is defined as follows under each of the three specifications mentioned above (standard errors obtained via the delta-method):

$$ACE: h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_C^2 + \sigma_E^2} \tag{2a}$$

$$AE: h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2} \tag{2b}$$

$$ADE: h^2 = \frac{\sigma_A^2 + \sigma_D^2}{\sigma_A^2 + \sigma_D^2 + \sigma_E^2} \tag{2c}$$

In its basic form, we estimate the regression without any covariates. We then add gender as a covariate to account for opposite-sex non-identical twins. Next, we explore several of the potential channels for genetic variation in health insurance status noted above. First, we examine education as a proxy for human capital and indicator for time discounting preferences. Next, we add indicators for full- and part-time employment, self-employment, and household income quintiles to capture availability and affordability of insurance. In alternative models, we include indicators for the work industry of employed individuals based on industry code aggregations in MIDUS I (professional services, manufacturing, agriculture/mining/construction, and other industries); we also estimate a model including an aggregated socioeconomic status index (SEI) capturing occupational status based on income and education and developed in MIDUS I. Next, we add marital status to examine household

formation and potential availability of insurance through a spouse. We then control for health status by including the number of chronic health conditions and self-rated health in order to capture individual health needs and account for the potential of adverse selection. Next, we add excessive alcohol consumption, smoking, and use of illicit drugs and medications on one's own (excluding normal use of pain killers such as for a headache) as indicators of risk preferences. Finally, we control for prevention preferences (overall and for heart disease and cancer) and overall prevention effort. While these variables may not uniquely capture the hypothesized channels, all are conceptually relevant for insurance status. When adding covariates, we estimate all models for the sample with complete data on all covariates to ensure no changes in sample composition which includes 680 twin pairs when including both private and public coverage and individuals older than 65 and 538 twin pairs when focusing on private coverage versus uninsured among individuals younger than 65. Table 1 shows summary statistics for the analytical sample with complete data on all covariates when including both private and public coverage (supplementary Table S1 shows summary statistics for individuals younger than 65 years and excluding publicly covered individuals and their co-twins).

Results

Our analysis suggests genes as an important source of variation in being uninsured. This is first revealed in a simple comparison of correlations of being uninsured or not between identical and non-identical twins. The correlation in uninsured versus private/public coverage between identical twins was twice that among fraternal twins (0.438 vs. 0.216). The difference is more prominent when focusing on uninsured versus private coverage among individuals younger than 65 years (0.578 among identical twins vs. 0.238 among fraternal twins).

In Table 2, we present the variance proportions accounted for by genetic and shared environmental effects based on the generalized linear mixed model without controlling for any covariates. Across the three genetic models, ACE, AE, and ADE, we observe similar results for genetic variation. As much as 43% of the variance in uninsured versus private/public insurance is accounted for by genetic effects. As expected, genetic variation is more pronounced when focusing on individuals younger than 65 years and excluding public coverage because of the greater choices that individuals face in those circumstances; genes account for 50% of the variation in being uninsured versus having private coverage. The estimate of insurance variation explained by shared environment based on the ACE model was practically 0, suggesting that the remaining variation is explained by environmental factors unique to each twin, and that the AE model is a preferred model to the ACE model. Overall, genetic variation from the ADE model combining additive and dominant genetic effects is virtually the same as that of the ACE or AE models.

Next, we evaluate the extent to which selective demographic, socioeconomic, and health characteristics that vary between twins explain the observed genetic variation by introducing these as covariates into the model. Because we use individuals with complete data on all covariates for these regressions, we first re-estimate the variance proportion accounted for by genetic effects without adjusting for any covariates in this specific sample. We find overall similar estimates (46% of variance in uninsured vs. private/public coverage and 50% of variance in private coverage vs. uninsured among individuals younger than 65 explained by genetic effects). Then, we first test the sensitivity of the main estimates to controlling for sex. Next, we successively add the following variables: (1) education as a proxy for human capital; (2) employment indicators and income which influence availability/affordability of

Table 1 Sample description for individuals with complete data on all variables excluding individuals with public coverage and their co-twins and individuals 65 years and older

Variable	Description	Mean	SD
Uninsured[a]	0/1 indicator for being uninsured relative to having private health insurance	0.084	0.277
Male	0/1 indicator for male vs. female	0.451	0.498
Less than high school[b]	0/1 indicator for less than high school graduate	0.089	0.285
Some college[b]	0/1 indicator for some college	0.305	0.461
College graduate[b]	0/1 indicator for college graduate	0.305	0.461
Employed	0/1 indicator for employed individuals	0.660	0.474
Full-time employment[c]	0/1 indicator for working 40 or more hours per week on average in main job	0.512	0.500
Part-time employment[c]	0/1 indicator for working less than 40 h per week on average in main job	0.148	0.355
Professional services[d]	0/1 indicator for working in professional services	0.188	0.390
Manufacturing[d]	0/1 indicator for working in manufacturing	0.123	0.328
Agriculture/mining/construction[d]	0/1 indicator for working in agriculture, forestry, fishery, mining or construction	0.041	0.199
Other industry[d]	0/1 indicator for working in other industries (e.g. transportation, community/public utility, trade, finance, real estate, repair)	0.308	0.462
Self-employed[c,d]	0/1 indicator for self-employed individuals	0.138	0.345
SEI[e]	Aggregated index for occupational status based on income and education across occupations as developed in MIDUS I (0 assigned for unemployed or not self-employed)	31.190	19.985
Income	Total household income	60,780	47,426
Married	0/1 indicator for being married versus unmarried	0.757	0.429
Children	Number of biological children	1.939	1.519
Chronic conditions	Number of chronic conditions	2.097	2.161
Health status	Self-reported health on a scale from 1 to 10	7.718	1.496

Table 1 continued

Variable	Description	Mean	SD
Smoker	0/1 indicator for being a current smoker	0.200	0.400
Excess alcohol 1–2 times[f]	0/1 indicator for using alcohol excessively 1–2 times in past 12 months	0.143	0.351
Excess alcohol ≥ 3 times[f]	0/1 indicator for using alcohol excessively 3 times or more in past 12 months	0.079	0.270
Drug use	0/1 indicator for using illicit drugs or medications on one's own (excluding normal use of over-the-counter pain medications)	0.125	0.331
Responsible	Response on a 7-category scale from strongly disagree to strongly agree to the following statement: "Keeping healthy depends on things that I can do"	6.462	0.922
Prevention effort	Response on a 7-category scale from strongly disagree to strongly agree to the following statement: "I work hard at trying to stay healthy"	5.646	1.264
Heart disease prevention	Response on a 7-category scale from strongly disagree to strongly agree to the following statement: "There are certain things I can do for myself to reduce the risk of a heart attack"	6.699	0.696
Cancer prevention	Response on a 7-category scale from strongly disagree to strongly agree to the following statement: "There are certain things I can do for myself to reduce the risk of getting cancer"	6.008	1.205

Descriptive statistics are shown for 1360 individuals (680 twin pairs) with complete data on all variables

[a]The uninsured rate in this group is 8.9 when including individuals with information on insurance regardless of missing data on covariates

[b]Reference is high school graduate

[c]Reference is unemployed or not self-employed with the model including full-time, part-time, and self-employment as the labor market indicators

[d] reference is unemployed or not self-employed with the model including the industry indicators (as coded in MIDUS I) for employed individuals and an indicator for self-employment

[e]SEI is included as the only labor market indicator when included as a covariate (i.e. the employment and self-employment indicators are omitted)

[f] question asks about number of times in past 12 months the person had larger amounts of alcohol or used alcohol for a longer period than intended

Table 2 Proportion of variance of uninsured accounted for by genetic effects and shared environment under different genetic models

Model	Genetic variation			Shared environment	Model log likelihood
	Estimate	SE	95% CI	Estimate	
Panel A: Uninsured versus private/public coverage					
ACE	0.432	0.14	0.157–0.707	0.000	−456.51
AE	0.432	0.12	0.171–0.625	–	−456.51
ADE	0.434	0.136	0.167–0.702	–	−456.51
Panel B: Uninsured versus private coverage for adults < 65 years excluding public coverage					
ACE	0.497	0.217	0.073–0.922	0.000	−376.32
AE	0.500	0.122	0.213–0.686	–	−376.32
ADE	0.504	0.138	0.234–0.775	–	−376.32

The proportion of the uninsured variance due to unique environment variance is one minus the genetic heritability minus the shared environment proportion in the ACE model
ACE indicates the model decomposing variation in coverage over additive genetic effects, environmental effects shared between twins, and environmental effects unique to each twin. AE indicates the model decomposing variation in coverage over additive genetic effects and environmental effects unique to each twin
ADE indicates the model decomposing variation in coverage over additive and dominant genetic effects as well as environmental effects unique to each twin

insurance; (3) household formation including marital status and having biological children which influence demand for family-based coverage; (4) health status including number of chronic conditions and self-reported health; (5) risk preferences captured by excessive alcohol consumption, smoking, and use of drugs and medications on one's own (excluding normal use of pain killers); and (6) prevention preferences and overall prevention effort. We limit this estimation to the AE model due to the insignificance of shared environment effect as explained above and the similarity of genetic variation between the AE and the ADE model. Furthermore, the log-likelihood of the three models (without covariates) is virtually the same, making the AE model more appealing than the other two models since it requires estimating 2 instead of 3 parameters (Rabe-Hesketh et al. 2008).

The estimates of the uninsured variance accounted for by genetic effects under the various sets of control variables are shown in Table 3 both for uninsured versus private/public coverage and then uninsured versus private coverage only among individuals younger than 65 years. Controlling for sex has no effect on genetic variation. Education has a slight effect, reducing genetic variation by 6–9%. Adding employment and income explains an additional 33% of the genetic effect on coverage (whether uninsured vs. public/private coverage or uninsured vs. private coverage among individuals younger than 65) in the model that includes industry indicators for employed individuals. This specification including education, employment industry indicators, self-employment indicator, and income quintile indicators explains the most of the genetic effect, including nearly 38% of the genetic variation in uninsured versus private/public coverage and 43% of genetic variation in uninsured versus private coverage among individuals younger than 65. Including the aggregated SEI as the only labor market indicator does not explain additional genetic variation. Surprisingly, adding the other covariates—marital status, number of biological children, number of chronic health conditions, self-rated health, the behavioral indicators for risk preferences (excessive alcohol, smoking, and illicit drugs and medications on one's own excluding normal use of pain killer), and

Table 3 Proportion of uninsured variance accounted for by genetic effects under different control variables

Control variables	Uninsured versus private/public coverage			Uninsured versus private coverage for adults < 65 years excluding public coverage		
	Estimate	SE	95% CI	Estimate	SE	95% CI
No covariates: Sample with complete data on all covariates	0.463	0.127	0.178–0.660	0.499	0.130	0.194–0.694
+ sex	0.467	0.126	0.182–0.662	0.501	0.129	0.196–0.695
+ education	0.441	0.131	0.152–0.646	0.454	0.138	0.148–0.664
+ employment, household income quintiles						
Indicators for full-time employment, part-time employment, self-employment	0.328	0.160	0.039–0.588	0.321	0.174	0.022–0.600
Industry indicators for employment, indicator for self-employment[a]	0.288	0.165	0.018–0.564	0.288	0.180	0.008–0.584
+ marital status, number of biological children	0.322	0.17	0.024–0.599	0.316	0.197	0.005–0.624
+ number of chronic conditions, self-rated health status	0.328	0.172	0.026–0.603	0.316	0.196	0.006–0.622
+ current smoker, excessive alcohol, drug use	0.322	0.173	0.023–0.600	0.319	0.197	0.006–0.626
+ preferences for prevention, overall prevention effort	0.309	0.174	0.018–0.591	0.291	0.197	0.002–0.606

All estimations are based on the samples with complete data on all covariates. Covariates are added sequentially into the regression retaining the previous ones. Estimates obtained from an AE model (additive genetic effects plus environmental effects unique to each twin) including covariates

[a]Heritability was 0.364 when including industry codes for both employment and self-employment and 0.354 when using the SEI as the only labor market indicator (0.298 and 0.357 when comparing private coverage to uninsured and excluding individuals 65 years and older)

preferences for prevention and overall prevention, does not explain any additional genetic effects on either specification of insurance types.

Conclusions

In the first work to explore whether genes are relevant for the variation in health insurance coverage status in the US, we analyze a twin model and find that over 40% of the variation in whether adults have health insurance coverage or not can be explained by genetic influences. The genetic effect is stronger when focusing on uninsured versus private coverage among individuals younger than 65 years. This is expected because individuals have discretion in obtaining private coverage or not whereas individual choice is much less important given the nearly universal availability of Medicare for individuals aged 65 and older. In terms of underlying channels, we can explain nearly 40% of the genetic effects based on differences in education, employment industry, self-employment status, and income, with the employment and income indicators explaining nearly one third of the variation on their own. Marital status, number of biological children, health status, and indicators for risk taking and disease prevention as well as overall prevention effort did not explain any of the genetic influence when added to the specification that included education, employment, and income.

The unexplained genetic variation can still be related to the pathways illustrated in our conceptual framework above but characterizing it requires better measures of health status, human capital, and preferences over risk and time discounting. The included measures for health status indicators and risky behaviors do not mediate any of the genetic effects on coverage in our models. This may partly be due to measurement errors in those variables particularly in the risky behaviors that we use (excessive alcohol, smoking, illicit drugs and use of medications on one's own excluding normal use of painkillers). We lack better proxies for risk preferences in this dataset. However, self-rated health status is overall considered a reasonable measure of health and self-report of chronic conditions is generally concordant with objective data sources such as claims (Wolinsky et al. 2014). That those health status measures explain none of the genetic effect suggests that any adverse selection in insurance that is related to health is largely not linked to genetic factors.

Our work indicates that genes were an important contributor to explaining the variation in private health coverage status in the US population prior to the ACA. The genetic effects appear to be partly expressed through economic traits, especially employment and income. To the extent that individuals have a choice in having health coverage or differences in access to insurance depending on their employment and income—in contrast to a scenario of automatic universal coverage (such as in Canada or the United Kingdom)—genetic factors operating through such channels may continue to result in a proportion of the population being without health coverage. The reduced genetic variation when adding public coverage into our analysis and the evidence that employment and income explain nearly one third of the genetic variation supports this conclusion.

Of course, declining uninsured rates due to the Medicaid expansions and the individual and employer mandates of the ACA would be expected to reduce genetic influence on insurance decisions. We are unable to directly quantify the ACA effect on genetic influence because we lack data on a nationally representative sample of twins after the key provisions of the ACA related to the Medicaid expansions and the individual mandate had been implemented. However, several factors would suggest that genetic influence on coverage may still be relevant even after the ACA. Not all states have expanded their Medicaid programs; to date, 33 states and DC expanded their Medicaid programs, 3 states are considering expansion, and

14 states have not expanded and are not considering expansion (KFF 2018a). Furthermore, several states either have obtained, are in the process of obtaining, or are considering applying for federal waivers to tie Medicaid eligibility to work status, which may reduce take-up of Medicaid coverage. Also, the 2017 Tax Cut and Jobs Act has repealed the individual mandate penalty of the ACA (effective January 2019), with the CBO projecting an increase in the number of uninsured individuals by 4 million in 2019 and 13 million in 2027 as a result (assuming no other changes; CBO 2017). In addition, there is little evidence of large changes in the employer-sponsored coverage following the ACA and no evidence of a meaningful change in the uninsured rate among the non-elderly after 2015, with a 10.7% uninsured rate in 2017 (KFF 2018b). As a whole, this evidence suggests again that genetic factors continue to partly explain differences in coverage status after the ACA, and highlights the importance of collecting data on coverage among twins after the ACA to re-estimate genetic variation.

An implication of our results is that the policy environment in the mid-1990s did not dissipate the large influence of intrinsic personal traits on private health insurance coverage. Until the recent ACA driven healthcare reform, there had been little change in the policy environment regarding access to private coverage and private insurance markets that would have dramatically modified genetic variation in coverage. Our results therefore may serve as a benchmark to examine whether genetic variation in coverage choices have changed over the past 20 years.

Our findings highlight the value of future research to examine the genetic mechanisms that influence health coverage choices and the intermediary behavioral or health channels. Identifying these mechanisms may improve our understanding of how health insurance policy changes such as the ACA impact coverage choices and inform future adjustments to the way insurance policies are implemented. An example of interventions that may eventually be informed by such future research is whether to have health insurance be an opt-out choice rather than opt-in, similar to ongoing experiments in saving for retirement. Characterizing genetic influences could also prove useful in understanding differential individual responses to insurance availability or mandates. Improvements in understanding the decision making process to seek or not seek health insurance may further lead to beneficial gains in the way individuals are informed and educated about related programs such as retirement programs, annuities, life insurance, etc.

We employ a regression-based model of the basic twin design which has been shown to be generally robust to the assumptions it makes. It is important to note, however, that genetic influence in this design may be over or under-estimated depending on a range of factors including gene-environment correlations and interactions and the extent of genetic correlations between spouses, although empirical estimates available for some political traits do not suggest major biases (Verhulst 2013). Further, the higher the genetic similarity between spouses, the greater the underestimation bias in twin studies. Spouses may have positive genetic correlation on average across the pathways through which genetic effects on insurance coverage may develop (Fig. 1). While it is impossible for us to sign the bias or infer its magnitude, the key point is that there is no reason to expect a priori that the twin model would over-estimate the genetic influence on health insurance coverage.

Even though we study a national sample of twins, the results may not be nationally generalizable. To assess the representativeness of the twin sample, we compare key characteristics to a nationally representative sample of individuals within a similar age range (25–74) from the March Supplement of the 1995 Current Population Survey (Supplementary Table S2). The two samples were close on mean age (~45 years) and employed to population ratio (~65 vs. 68%). The uninsured rate was lower in the twin sample (9% vs. 14%) and education (~30% vs. 24% college graduates), household income ($59,000 vs. $49,000), and number

of children (2 vs. 0.9) were higher. Also, the twin sample had larger proportions of Whites (95% vs. 84%), married individuals (74% vs. 66%), and females (56% vs. 49%). Because of the lower variability in coverage and higher socioeconomic status overall in the twin sample, our estimates may represent a lower bound of genetic variation in a more diverse and representative sample.

Finally, given the evidence of genetic variation in healthcare use and health (e.g. Wehby et al. 2015, 2017, 2018), our results suggest genes as another important source of unobservable heterogeneity that may result (if ignored) in selection bias when investigating the effects of insurance on health and healthcare use. To illustrate this potential bias, we examined associations between uninsured status and self-reported health within twin pairs separately for identical and non-identical twins. We did this by regressing self-reported health (on a scale from 1 to 10) on private coverage versus uninsured status using OLS including twin-pair fixed effects separately for the two twin groups and clustering standard errors at the twin-pair level. We found that coverage was associated with a 9% (relative to mean) increase in health status among non-identical twins, but had a practically null association among identical twins. This ad hoc analysis provides an example of how genetic variation can be an important confounder for examining the relationship between insurance and health or healthcare use in analyses that focus on "association estimates" without an exogenous source of coverage variation, and indicates that direct adjustment for observable confounders is insufficient for removing that bias. Collecting data on a large and nationally representative sample of twins may, therefore, be useful for studying insurance effects on health and healthcare use. Furthermore, as national surveys begin to collect genetic information and as genetic variants associated with health coverage decisions are characterized in future research, researchers may be able to directly adjust for such genetic effects in general (i.e. non-twin) samples.

References

Agarwal, A., Williams, G. H., & Fisher, N. D. (2005). Genetics of human hypertension. *Trends in Endocrinology and Metabolism, 16*(3), 127–133.

Anokhin, A. P., Golosheykin, S., Grant, J. D., & Heath, A. C. (2011). Heritability of delay discounting in adolescence: A longitudinal twin study. *Behavior Genetics, 41*(2), 175–183.

Barnes, J. C., Wright, J. P., Boutwell, B. B., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., et al. (2014). Demonstrating the validity of twin research in criminology. *Criminology, 52*(4), 588–626.

Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Guðnason, V., et al. (2012). The promises and pitfalls of genoeconomics. *Annual Review of Economics, 4,* 627.

Boardman, J. D., Blalock, C. L., & Pampel, F. C. (2010). Trends in the genetic influences on smoking. *Journal of Health and Social Behavior, 51*(1), 108–123.

Boardman, J. D., Domingue, B. W., & Daw, J. (2015). What can genes tell us about the relationship between education and health? *Social Science and Medicine, 127,* 171–180.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social Forces, 92*(1), 109–140.

Brim, O. G., Baltes, P. B., Lachman, M. E., Markus, H. R., Shweder, R. A., Marmot, M. G., et al. (2017). *Midlife in the United States (MIDUS 1), 1995–1996*. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2017-11-16. https://doi.org/10.3886/icpsr02760.v12.

Burt, C. H., & Simons, R. L. (2014). Pulling back the curtain on heritability studies: Biosocial criminology in the postgenomic era. *Criminology, 52*(2), 223–262.

Carlsson, S., Ahlbom, A., Lichtenstein, P., & Andersson, T. (2013). Shared genetic influence of BMI, physical activity and type 2 diabetes: A twin study. *Diabetologia, 56*(5), 1031–1035.

Cesarini, D., Dawes, C. T., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *Quarterly Journal of Economics, 124,* 809–842.

Cesarini, D., Johannesson, M., Lichtenstein, P., Sandewall, Ö., & Wallace, B. (2010). Genetic variation in financial decision-making. *The Journal of Finance, 65*(5), 1725–1754.

Charney, E., & English, W. (2013). Genopolitics and the science of genetics. *American Political Science Review, 107*(02), 382–395.

Congressional Budget Office (CBO). (2017). *Repealing the individual health insurance mandate: An updated estimate*. https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/53300-individualmandate.pdf. Accessed June 20, 2018.

Cronqvist, H., & Siegel, S. (2015). The origins of savings behavior. *Journal of Political Economy, 123*(1), 123–169.

Cutler, D., Finkelstein, A., & McGarry, K. (2008). Preference heterogeneity and insurance markets: Explaining a puzzle of insurance. *American Economic Review: Papers and Proceedings, 98*(2), 157–162.

Cutler, D. M., & Lleras-Muney, A. (2006). *Education and health: Evaluating theories and evidence* (No. w12352). National Bureau of Economic Research.

Cutler, D. M., & Zeckhauser, R. J. (1998). Adverse selection in health insurance. *Forum for Health Economics & Policy, 1*(1), 1–31. https://doi.org/10.2202/1558-9544.1056.

Domingue, B. W., Fletcher, J., Conley, D., & Boardman, J. D. (2014). Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences, 111*(22), 7996–8000.

Elks, C. E., Den Hoed, M., Zhao, J. H., Sharp, S. J., Wareham, N. J., Loos, R. J., et al. (2012). Variability in the heritability of body mass index: A systematic review and meta-regression. *Frontiers in endocrinology, 3,* 29.

Fang, H., Keane, P. M., & Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy, 116*(2), 303–350.

Kaiser Family Foundation (KFF). (2018a). *Status of state action on the medicaid expansion decision*. https://www.kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-the-affordable-care-act/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D. Accessed June 20, 2018.

Kaiser Family Foundation (KFF). (2018b). *Uninsured rate among the nonelderly population, 1972–2017*. https://www.kff.org/uninsured/slide/uninsured-rate-among-the-nonelderly-population-1972-2017/. Accessed June 20, 2018.

Kendler, K. S., Gatz, M., Gardner, C. O., & Pedersen, N. L. (2006). A Swedish national twin study of lifetime major depression. *American Journal of Psychiatry, 163*(1), 109–114.

Kosova, G., Abney, M., & Ober, C. (2010). Heritability of reproductive fitness traits in a human population. *Proceedings of the National Academy of Sciences, 107*(suppl 1), 1772–1778.

Kuhnen, C. M., & Chiao, J. Y. (2009). Genetic determinants of financial risk taking. *PLoS ONE, 4*(2), e4362.

Maes, H. H., Sullivan, P. F., Bulik, C. M., Neale, M. C., Prescott, C. A., Eaves, L. J., et al. (2004). A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychological Medicine, 34*(07), 1251–1261.

McGue, M., Zhang, Y., Miller, M. B., Basu, S., Vrieze, S., Hicks, B., et al. (2013). A genome-wide association study of behavioral disinhibition. *Behavior Genetics, 43*(5), 363–373.

Moor, M. M., Willemsen, G., Rebollo-Mesa, I., Stubbe, J., Geus, E. C., & Boomsma, D. (2011). Exercise participation in adolescents and their parents: Evidence for genetic and generation specific environmental effects. *Behavior Genetics, 41,* 211–222.

Oster, E., Shoulson, I., Quaid, K., & Dorsey, E. A. (2010). Genetic adverse selection: Evidence from long-term care insurance and Huntington disease. *Journal of Public Economics, 94,* 1041–1050.

Rabe-Hesketh, S., Skrondal, A., & Gjessing, H. K. (2008). Biometrical modeling of twin and family data using standard mixed model software. *Biometrics, 64,* 280–288.

Romeis, J. C., Heath, A. C., Xian, H., Eisen, S. A., Scherrer, J. F., Pedersen, N. L., et al. (2005). Heritability of SF-36 among middle-age, middle-class, male-male twins. *Medical Care, 43*(11), 1147–1154.

Romeis, J., Scherrer, J., Xian, H., Eisen, S., Bucholz, K., Health, A., et al. (2000). Heritability of self-reported health status. *Health Services Research, 35*(5 Part 1), 995–1010.

Stacey, D., Clarke, T. K., & Schumann, G. (2009). The genetics of alcoholism. *Current Psychiatry Reports, 11*(5), 364–369.

Stubbe, J. H., Boomsma, D. I., Vink, J. M., Cornes, B. K., Martin, N. G., Skytthe, A., et al. (2006). Genetic influences on exercise participation in 37,051 twin pairs from seven countries. *PLoS ONE, 1*(1), e22.

Thomsen, S. F., Van Der Sluis, S., Kyvik, K. O., Skytthe, A., & Backer, V. (2010). Estimates of asthma heritability in a large twin sample. *Clinical and Experimental Allergy, 40*(7), 1054–1061.

Treloar, S. A., McDonald, C. A., & Martin, N. G. (1999). Genetics of early cancer detection behaviours in Australian female twins. *Twin Research, 2*(01), 33–42.

True, W. R., Romeis, J. C., Heath, A. C., Flick, L. H., Shaw, L., Eisen, S. A., et al. (1997). Genetic and environmental contributions to healthcare need and utilization: A twin analysis. *Health Services Research, 32,* 37–53.

Trumbetta, S. L., Markowitz, E. M., & Gottesman, I. I. (2007). Marriage and genetic variation across the lifespan: Not a steady relationship? *Behavior Genetics, 37*(2), 362–375.

van der Loos, M., Rietveld, C. A., Eklund, N., et al. (2013). The molecular genetic architecture of self-employment. *PLoS ONE, 8*(4), e60542. https://doi.org/10.1371/journal.pone.0060542.

Verhulst, B. (2013). Gene-environment interplay in twin models. *Political Analysis, 21*(3), 368–389.

Wehby, G. L., Domingue, B. W., & Boardman, J. D. (2015). Prevention, use of health services, and genes: Implications of genetics for policy formation. *Journal of Policy Analysis and Management, 34*(3), 519–536.

Wehby, G. L., Domingue, B. W., Ullrich, F., & Wolinsky, F. D. (2017). Genetic predisposition to obesity and medicare expenditures. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 73*(1), 66–72.

Wehby, G. L., Domingue, B. W., & Wolinsky, F. D. (2018). Genetic risks for chronic conditions: Implications for long-term wellbeing. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 73*(4), 477–483.

Wolinsky, F. D., Jones, M. P., Ullrich, F., Lou, Y., & Wehby, G. L. (2014). The concordance of survey reports and medicare claims in a nationally representative longitudinal cohort of older adults. *Medical Care, 52*(5), 462–468.